UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL AND APPLIED SCIENCES

Doctoral Training Centre in Complex Systems Simulation
School of Electronics and Computer Science

Network structure and dynamics of empirical multiplex systems

by

Massimo Stella

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

December 11, 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

Doctoral Training Centre in Complex Systems Simulation
School of Electronics and Computer Science

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

NETWORK STRUCTURE AND DYNAMICS OF EMPIRICAL MULTIPLEX SYSTEMS

by Massimo Stella

Interactions are important, since they can influence and shape a variety of real-world networked systems. Interactions can have a multi-relational nature, i.e. being of different type. Different layers of interactions can look very different from each other, motivating the need for developing multi-layer network models and metrics in network science.

This thesis aims at developing novel multiplex frameworks for the quantitative investigation of two real-world systems: (i) the web of relationships between words in the human mind and (ii) the ecological interactions among animals in ecosystems. Despite being different in nature, both the mental lexicon of words and ecosystems can be represented as a multiplex network, where nodes represent distinct entities (e.g. words or animal groups) interacting on different layers in different ways (e.g. words being semantically and/or phonologically similar; animal species eating or parasitising each other). In both the considered systems, interactions crucially determine function and dynamics of a variety of processes.

In the mental lexicon, individual interactions have been shown to influence both language acquisition and usage. In Part I of this thesis, I show that the structure of the phonological layer reflects constraints related to language use. I proceed by introducing the framework of multiplex lexical networks for quantifying, for the first time, the influence that phonology and semantics combined can have on (i) word acquisition of toddlers and (ii) word processing of adults. Results highlight phenomena that are not observable in single-layer networks. In toddlers word learning strategies based on the whole multiplex structure match empirical word learning significantly better than strategies based on individual layers, indicating that multiplexity is important for early word acquisition. At later ages the multiplex structure evolves by displaying an early, explosive emergence of a multiplex network core of words, which facilitates mental navigation and increases robustness against cognitive impairments.

The second part of this thesis focuses on ecosystems, where interactions encapsulated in food webs or host-parasite networks greatly influence species extinction. In Part II of this thesis, I introduce the framework of ecological multiplex or “ecomultiplex” networks for combining predator-prey and host-parasite contact interactions as two layers of a network representing trophic links in a given ecosystem. I show that host-parasite interactions can dramatically increase the susceptibility of ecosystems to a parasite pandemic compared to models based on single-layer trophic networks only. Results of the ecomultiplex model are tested against empirical findings from field work in Brazilian ecosystems, finding agreement between my theoretical results and empirical data. Furthermore, by considering the multi-relational nature of trophic interactions, I quantitatively show that generalist top predators might accelerate parasite spread rather than hampering it, thus providing a theoretical explanation to recent empirical findings.

Both in the mental lexicon and ecosystems, multiplexity influences structure, dynamics and function in ways not yet accounted for in the literature. This thesis aims to fill this gap by suggesting multiplex frameworks suitable for quantitative testing of empirical conjectures, while opening new modelling challenges at the interface of physics, network science and other disciplines.

Contents

List of Figures

List of Tables

UNIVERSITY OF
Southampton

Academic Thesis: Declaration Of Authorship

I, MASSIMO STELLA

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

[title of thesis] ....NETWORK STRUCTURE AND DYNAMICS OF EMPIRICAL MULTIPLEX SYSTEMS............

....

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Either none of this work has been published before submission, or parts of this work have been **published** as:

1. M. Stella, N. M. Beckage and M. Brede, Multiplex lexical networks reveal patterns in early word acquisition in children, Scientific Reports 7: 46730 (2017).
2. M. Stella, C. S. Andreazzi, S. Selakovic, A. Goudarzi and A. Antonioni, Parasite Spreading in Spatial Ecological Multiplex Networks, Journal of Complex Networks, 5 (3) (2016).
3. M. Stella and M. Brede, Patterns in the English language: phonological networks, percolation and assembly models, Journal of Statistical Mechanics, P05006 (2015).

Signed: Massimo Stella

Date: 30 September 2017

Acknowledgements

This thesis focuses on those systems where interactions and diversity are key factors. Indeed, my work itself was the result of several fruitful interactions with people from different backgrounds who need to be acknowledged.

Firstly, I would like to thank my family, my parents Italo and Ginny and my brother Pietro, for supporting me during the whole PhD process with assiduity and patience. Thanks for cheering me on and believing in me!

Another very important acknowledgement goes to my main supervisor, Markus Brede, who attentively provided me with the freedom to explore complexity science and expertly supported me during the whole PhD. Speaking about supervision, another very important acknowledgement goes to my external supervisor, Manlio De Domenico, who expertly supported and inspired me with his enthusiasm for science and multiplex networks in my final year of the PhD. Without my supervisors, I don't think I would have achieved what I did and I am really grateful for that.

In these four years, I was also lucky enough to meet and interact with many clever and passionate young researchers. I would like to acknowledge my amazing "Santa Fe" friends and collaborators Alberto, Cecilia, Sanja, Alireza, Francesca, Leto, Luis, Cole, Claire, Bernardo, Flavia, Brais, José, Sarah, Emilia and Matt, my PhD colleagues Sabin, Anastasia, Kostas, Garvin and Valerio, my "Kansas" friends Cynthia, Nichol, Trevor and Matt, and last but not least fellow complexity explorers and friends Jelena, Serafina, Marco, Sarah, Federico Botta, Federico Battiston, Iacopo, Elisa, Sandro, Maria, Saska and Laura. I learned a lot from all of you and I hope we will meet again at the next conference!

I would also like to thank for their kind support and feedback professors Nicole M. Beckage, Michael Vitevitch, Mason A. Porter, Alex Arenas, Michael Zock, Sonia Kéfi, Ramon Ferrer i Cancho, Jesus Gomez-Gardenes, Anxo Sanchez and Tom Lenaerts. A special thanks goes also to the faculty and staff of the DTC for Complex Systems Simulation: Seth Bullock, Jonathan Essex, Hans Fangohr, Chris Cave-Ayland, Jason Noble, James Dyke, Tracey Giovacchini and especially Nicki Lewin for their friendly assistance and attentive help with several aspects of my PhD. It was thanks to all of you and my supervisors if I managed to attend and organise several scientific events around the world during my PhD.

In the "complicated" word of complexity (and academia), I felt lucky to meet you all, learn from your ideas and get some of your enthusiasm for science. I hope this thesis reflects well enough how proud I am of all the above interactions and collaborations I pursued during the PhD.

Preface

This thesis is presented in the “Three-paper” format, following the formatting instructions[1] provided by the University of Southampton.

The “Three-paper” format allows to present the publishable or published work of a candidate as a series of papers, constituting the substantive work of a thesis together with Introduction and Conclusions sections.

Recurring themes, representation and quantitative methodologies are introduced at the beginning in an Introduction section and later summarised and discussed in a Conclusions section, for interpretation of this body of work as a whole. The thesis starts with an outline and a brief literature review which lays the motivation behind the approach used in the papers and the gaps in the scientific literature that they account for.

Papers constitute the main part ot this “Three-paper” thesis. The papers are written and reported in a self-contained way. Each one of them constitutes a Chapter. Terminology in each Chapter has been slightly modified for increased clarity (e.g. sentences like “In this paper, ...” or “In this manuscript, ...” have been changed into “In this Chapter, ...”). However, in order to adhere to the “Three-paper” format, the structure of sections in each paper has been preserved as in the original papers. For the same reason and also for easier consultation in supporting the results of the thesis, supplementary information is presented at the end of each Chapter rather than in appendices.

Papers are organised in two parts, each one introduced by a brief literature review. The work ends with a discussion of the results achieved in each paper and acknowledging possible research directions for future work.

---

[1] The pdf file with the University guide is available online goo.gl/uimV6k (Last Accessed: August 23 2017).

List of Publications

During my PhD I achieved a total of 9 publications: 4 as journal papers and 5 as proceedings.

The published journal papers are:

1. M. Stella, N. M. Beckage and M. Brede, Multiplex lexical networks reveal patterns in early word acquisition in children, Scientific Reports 7: 46730 (2017).
2. M. Stella, C. S. Andreazzi, S. Selakovic, A. Goudarzi and A. Antonioni, Parasite Spreading in Spatial Ecological Multiplex Networks, Journal of Complex Networks, 5 (3) (2016).
3. M. Stella and M. Brede, Patterns in the English language: phonological networks, percolation and assembly models, Journal of Statistical Mechanics, P05006 (2015).
4. M. Stella and M. Brede, A k-deformed model of growing complex networks with fitness, Physica A: Statistical Mechanics and its Applications, 407, 360-368 (2014).

The published proceedings are:

1. M. Stella and M. Brede, Mental Lexicon Growth Modelling Reveals the Multiplexity of the English Language, Complex Networks VII, Springer (2016).
2. M. Stella and M. Brede, Investigating the Phonetic Organisation of the English Language via Phonological Networks, Percolation and Markov Models, Lecture Notes in Computer Science, Springer (2015).
3. C. S. Andreazzi, A. Antonioni, A. Goudarzi, S. Selakovic and M. Stella, Disease spreading on ecological multiplex networks, Santa Fe Institute CSSS 2014 Proceedings (2014).
4. A. Antonioni, L. A. Martinez-Vaquero, N. Mathis, L. Peel and M. Stella, Dynamical game theory, Santa Fe Institute CSSS 2014 Proceedings (2014).
5. F. M. D. Marquitti, D. Wu, L. A. Martinez-Vaquero, M. Stella, A. Antonioni, C. Graebner and B. Krese, Persistence of pollination systems, Santa Fe Institute CSSS 2014 Proceedings (2014).

I also have 4 manuscripts in preparation or submitted to journal venues:

1. M. Stella, N. M. Beckage, M. Brede and M. De Domenico, Multiplex model of mental lexicon reveals explosive learning in humans, submitted (arXiv:1705.09731).
2. M. Stella, S. Selakovic, A. Antonioni and C. S. Andreazzi, Investigating multiple-host parasite spread and immunisation strategies through ecomultiplex networks, submitted (arXiv:1706.05121).
3. A. Antonioni, N. Mathis, M. Stella, L. Peel and L. A. M. Vaquero, Drunk Game Theory: individual perceptions in evolutionary games, in preparation.
4. J. Aguilar-Rodríguez, L. Peel, M. Stella, A. Wagner and J. Payne, The architecture of an empirical genotype-phenotype map, under review.

During my PhD, I also produced 20 contributions accepted in international conferences and provided 8 invited talks. My institution sponsored me for attending the Santa Fe Institute Complex Systems Summer School 2014.

I organised 2 scientific workshops and in 2016 became a member of the Steering Committee for the Winter Workshop on Complex Systems for young researchers.

In October 2016 I was elected as Council Member in the Complex Systems Society and in the same year I was accepted as an associate member in the Institute of Physics, Group of Nonlinear and Complex Physics.

In terms of international research mobility, I performed the following funded research visits:

1. Department of Computer Science, University of Kansas, 16 April 2017 - 16 July 2017 (3 months);
2. AlephSys Lab, Universitat Rovira i Virgili, Tarragona, 16 October-5 November 2016 (3 weeks);
3. Grupo Interdisciplinar de Sistemas Complejos, Universidad Carlos III de Madrid, 8-22 May 2016 (2 weeks);
4. Artificial Intelligence Lab, Vrije Universiteit, Brussels, 1-8 December 2015 (1 week).

Chapter 1

Introduction to the research theme: applications of multiplex networks

> More is different.
>
> *Science*
>
> Philip Warren Anderson

1.1 Motivation

Network science investigates how interactions among individual entities produce and influence complex phenomena. The last 20 years have seen an exploding interest in network science (Barabási, 2015; Newman, 2010; Caldarelli, 2007; Boccaletti et al., 2006), mainly due to the discovery that despite differences in nature, size and structure, many different types of networks follow a common set of fundamental principles. In this way, network science quickly arose as the ideal quantitative framework for dealing with a variety of real-world systems through one common formalism. Novel theoretical tools and increasingly available large-scale datasets quickly made complex networks an interdisciplinary tool bringing new discoveries and insights in the understanding of complex systems (Newman, 2010).

From a network perspective, this thesis focuses mainly on the recently developed field of multiplex networks (Battiston et al., 2017; De Domenico et al., 2013; Kivelä et al., 2014; Boccaletti et al., 2014), i.e. multi-relational networks where the same set of agents can interact differently across several network layers. The last five years have seen an exponential increase in the publications related to multiplex networks, with successful applications in fields as diverse as transportation systems (Morris & Barthelemy, 2012; De Domenico et al., 2014), social and online network communities (Szell et al., 2010), brain physiology and functioning (Battiston et al., 2017; De Domenico et al., 2016; De Domenico, 2017), game theoretic dilemmas (Gómez-Gardenes et al., 2012) and epidemiology (Saumell-Mendiola et al., 2012; Sanz et al., 2014; Salehi et al., 2015; De Domenico et al., 2016) among many others. All the above applications suggest the high relevance of multiplex networks in modelling real-world systems. Such relevance is mainly due to the ability of multiplex networks to capture multi-relational and overlapping interactions which are not accounted for in single-layer complex networks.

Considering more than one type of interactions at once can indeed give rise to novel structure and phenomena not observable in single-layer networks. From the structural side, a node disconnected in one network layer might be highly connected in another network layer, thus altering the local connectivity of the node itself. From the dynamical side, diffusion processes on multiplex networks can indeed proceed at much faster rates compared to what would happen in the individual layers, a phenomenon called "super-diffusion" and not observable in single-layer networks (Gomez et al., 2013; De Domenico et al., 2014, 2016; Solé-Ribalta, 2013). Different structural properties and different dynamics of multiplex networks translate into potentially novel insights about the function of real-world systems and their interdisciplinary interpretation.

Despite the potential for achieving additional insights about real-world systems, the multiplex network framework mainly remained confined to the physics community with applications missing even in those subfields that historically benefitted from network representations of complex systems such as cognitive network science and network ecology.

This thesis focuses on filling the above gap by suggesting novel applications of multiplex networks to the investigation of: (i) language structure and cognitive development and (ii) epidemiological processes in ecosystems.

1.2 Language and ecosystems: different systems under the umbrella of the same formalism

The main contribution of this thesis is the development of multiplex network models for the investigation of real-world systems such as language and parasites in ecosystems. Although the considered systems do differ in their nature, they can both be represented by a multiplex network. This thesis represents a coherent body of work when interpreted as a network modelling approach addressing this common multiplexity. Multiplexity does embody the central theme of this thesis, in particular related to the following general research questions:

1. Is a multiplex representation of the considered system appropriate? Is it inspired by the nature of the underlying system? Are there theoretical approaches that can quantify if the multiplex approach is appropriate?
2. What are the novel phenomena that cannot be observed in single-layer networks? Why are multiplex networks a better representation of the underlying system compared to single-layer network representations?
3. What is the interpretation of the observed multiplex network patterns? Do they agree with previous empirical findings? Can multiplexity offer an explanation of patterns already observed in the literature?

The above research questions are discussed in Part I for the investigation of language structure and cognitive development and in Part II for the investigation of parasite spread in ecosystems.

1.2.1 A new multiplex approach to language and cognition: multiplex lexical networks

Language is a complex system of hierarchical constructs (phonemes in words, words in sentences, sentences in corpora) aimed at achieving communication (Aitchison, 2012). The specific linguistic system investigated within the thesis is the Human Mental Lexicon (HML), which is the mental representation of words in the human mind (Aitchison, 2012). In detail, the HML represents a multi-level repository, similar to a dictionary, where concepts are stored together with the underlying linguistic data (Aitchison, 2012; Elman, 2004). Many empirical studies indicate that relationships among concepts in this mental structure impact cognitive phenomena such as language acquisition (Hills et al., 2010; Beckage et al., 2011; Carlson et al., 2014) and usage (Collins & Loftus, 1975; Vitevitch, 1997; Sadat et al., 2014; Vitevitch et al., 2014; Griffiths et al., 2007). Since in the HML relationships do matter, then a network representation of it is appropriate. These relationships have a multiplex nature: words can be similar in their sound and/or in their semantic meaning at the same time. As better explained in the Introduction of Part I of this thesis, such multi-relational structure of word-word similarities has been barely considered in the literature. Therefore, a multiplex network approach fills a gap in the literature investigating the mental lexicon.

Along this direction, the thesis starts from the investigation of linguistic constraints in phonological networks. Building on this, the phonological network is combined with other semantic layers within the general framework of multiplex lexical networks for investigating language acquisition in toddlers and adults.

1.2.2 A new multiplex approach to ecology: ecomultiplex networks

The specific ecological systems investigated within the thesis are ecosystems such as Brazilian forests, in which a parasite can spread through different hosts and via different interaction mechanisms, i.e. blood-exchanges and predator-prey interactions. Although historically predator-prey interactions received a great deal of attention in terms of food webs (Pimm, 1982; Caldarelli, 2007), it was only recently that the scientific community found empirical indications of parasites significantly altering the organisation of ecosystems (Lafferty et al., 2008; Dunne et al., 2013) and their robustness to extinction (Keeling & Rohani, 2008; Keeling & Eames, 2005; Dunne et al., 2013). In the last few years this has led to the recognition of a need for novel modelling approaches, extending the analysis of ecological interactions beyond food webs. It is in this gap that the thesis contributes by introducing the formalism of ecological multiplex networks as a suitable multi-relational representation of how parasites can spread in spatial ecosystems.

This ecological multiplex (or “ecomultiplex”) approach represents the second part of this thesis, agreeing in terms of tools and approach with the first part.

1.2.3 Commonalities between ecology and psycholinguistics: the ecological psychology approach

The mental representation of language and ecosystems do have in common a multi-relational aspect, whose representation through multiplex networks is appropriate and valuable for extending previous approaches. Interestingly, this commonality in terms of approaching both ecosystems and cognitive systems starting from basic interactions finds well established roots in the scientific literature from psychology. Although no networks are considered, already Barker (1968) underlined the importance of the environment and its context-dependency in the development of human cognition. Barker considered as "ecological" the relationship of an individual mind with the environment, which is considered as made of other minds and external factors as well such as language or social norms[1]. Normative behaviour is therefore the expected behaviour of the majority of a population. Considering the interactions of words in the mind with the external environment represents a vision that is indeed very close to the ecological one in biology (Pimm, 1982), where ecosystems identify animals interacting with an environment made by other animals and exogenous factors such as plants, geographical constructs, weather, bacteria and so on.

Ecological psychology identifies the interactions with the environment as the ones effectively influencing the human mind (Clark, 2008) as opposed to the *innatism* view (Chomsky, 1986), according to which a human is a black-box whose development emerges autonomously over time. According to the innatism hypothesis, children do not acquire the fundamentals of language over time but they are rather born with it and can then develop it spontaneously even without explicit teaching. According to the ecological psychology view, instead, language acquisition is indeed a process arising from the interactions with the environment in the form of perceived word-word similarities and other type of stimuli like auditory or visual. Therefore ecological psychology postulates that the linguistic or social environment effectively influences the human mind, analogously to how ecologists consider natural environments influencing the dynamics of animal populations in ecosystems. These themes will be developed in Chapters 5, 6, 8 and 9 of this thesis.

1.3 Thesis structure

Since multiplex networks are the main common theme of the whole thesis, the above Introduction continues with a brief review of network theory concepts and a concise literature review of multiplex networks and their applications. This review provides a setting for the interested reader not expert in multiplex networks. Hence, Chapter 2 is otherwise optional for readers with such an expertise.

The main research component of the thesis is subdivided in two Parts:

1. in Part I we focus on the investigation of language structure and cognitive development, introducing the formalism of multiplex lexical networks;

---

[1] Notice that in this context a "norm" is the identification of the average behaviour of the majority in a system of interacting agents.

2. in Part II we focus on the investigation of parasite spread in ecosystems, introducing the formalism of ecomultiplex networks.

Every Part starts with a brief literature review and it contains published or publishable research articles. The subsectioning in every Chapter reflects the original published papers, in agreement with the University guidelines for the Three-paper thesis format. Every Chapter is linked to the previous ones by a brief introductory paragraph, providing structure and setting of each research article within the larger scope of the thesis.

The research achievements of the thesis are summarised and discussed within the Conclusions, after Part II. The supplementary information for the published articles is reported at the end of the respective chapters rather than in appendices for better readability through easier consultation of supporting information.

1.4 Research achievements

This Section briefly comments on the main research achievements contained within each presented research article, as clustered as in the two Parts.

1.4.1 Research achievements relative to language structure and cognitive development

Using a multiplex representation of the mental lexicon for the investigation of linguistic patterns arising from network topology is the main aim of the first part of the thesis. This part starts with Chapter 3, which briefly introduces the revelant literature by: (i) summarising the importance of a network approach for investigating the mental lexicon, (ii) discussing both previous single-layer approaches and (iii) reviewing potential multi-layer approaches suggested in the literature. The following list summarises the main research achievements provided in Chapters 4-6 of this part of the thesis.

1. Chapter 4: In this chapter we provide a quantitative framework for the study of phonological networks (PNs) for the English language by carrying out principled comparisons to null models, either based on site percolation or network growth models. We recognise that phonological networks are embedded in a combinatorial space and question which network features are inherited from the embedding and which represent linguistic patterns instead. We find that artificial networks matching connectivity properties of the English PN are exceedingly rare: This leads to the hypothesis that the word repertoire might have been assembled over time by preferentially introducing new words which are small modifications of old words. Our null models are able to explain the "power-law-like" part of the degree distributions and generally retrieve qualitative features of the PN such as high clustering, high assortativity coefficient, and small-world characteristics. However, the detailed comparison to expectations from null models also points out significant differences, suggesting the presence of additional constraints in

word assembly. Key constraints we identify are the avoidance of large degrees, the avoidance of triadic closure, and the avoidance of large non-percolating clusters.

2. Chapter 5: Phonology is only one type of word similarities encoded in the mental lexicon. In order to also account for multi-relational word-word similarities, we introduce the framework of multiplex lexical networks. We use this framework for investigating strategies of word learning in toddlers guided by semantic and phonological features of words. We find that the characteristics of the multiplex topology yield proxies of the cognitive processes of acquisition, capable of capturing emergent lexicon structure. In fact, we show that the multiplex structure is fundamentally more powerful than individual layers in predicting the ordering with which words are acquired. Furthermore, multiplex analysis allows for a quantification of distinct phases of lexical acquisition in early learners: While initially all the multiplex layers contribute to word learning, after about month 23 free associations take the lead in driving word acquisition.

3. Chapter 6: In this Chapter we extend the multiplex lexical framework to the lexicon of older children, adolescents and adult English speakers. The model detects a distinct cluster of words that are more frequent, easier to recollect and stimulating mainly brain regions involved in concrete perception compared to those words outside of it. This cluster emerges suddenly around age 7 yrs through an explosive phase transition that is not matched by models of random word acquisition. Furthermore, random word attacks suggest that this cluster acts as a core of densely connected words, providing short cuts for navigating the lexicon and thus increasing its robustness to word failures in cognitive impairments such as anomia. Our quantitative findings on a large-scale approximation of the mental lexicon confirm previous psycholinguistic conjectures about the presence and cognitive advantages of a core in the mental lexicon and extend this work to consider multiple types of linguistic relationships known to be important in psycholinguistics.

Due to space reasons, additional results similar to the ones contained in the above Chapters but published in conference proceedings such as (Stella & Brede, 2016a) cannot be included in the thesis. In (Stella & Brede, 2016a) we used Markov processes up to third order for building artificial phonological networks and highlighting the presence of long range correlations among phonemes in real words that cannot be captured by lower order phoneme correlations. In (Stella & Brede, 2016b) we built a toy model of word acquisition considering the interplay between phonological and semantic networks over the whole developmental period from early childhood to adulthood. Despite the different models, data and measures, the results from (Stella & Brede, 2016b) agree with those exposed in Chapter 5 in underlining the importance of the multiplexity of words in early word learning.

1.4.2 Research achievements relative to parasite spreading in ecosystems

The ecological side of this thesis follows the need in network ecology for multi-relational modelling of multiple ecological interactions within a given ecosystem. Similar to the psycholinguistic approach, we use a multiplex network representation for investigating a dynamical process of interest for the represented system. This part starts with Chapter 7, which is a brief literature review summarising the importance of a network approach for investigating ecosystems, discussing previous single-layer approaches and multi-layer approaches suggested in the literature. The following list introduces each chapter, its main motivation and its scientific outcomes.

1. Chapter 8: Several parasites can be transmitted among their hosts through different mechanisms, each one giving rise to a network of interactions. Modelling these networked, ecological interactions at the same time is still an open challenge. We present a novel spatially embedded multiplex network framework for modelling multi-host infection spreading through multiple routes of transmission. Simulations show that the parasite spreads at faster rates when both predator-prey and parasite-host interactions are considered with equal rates. We also find that the ratio of vectors-to-hosts in the community crucially influences the infection spread as it regulates a percolation phase transition in the rate of parasite transmission. This indicates that the multiplex model provide a richer phenomenology in terms of parasite spreading dynamics compared to more limited mono-layer models. By immunising the same fractions of predator and prey populations, we show that the multiplex topology is fundamental in appropriately capturing the role that each host species plays in parasite transmission in a given ecosystem.
2. Chapter 9: In this Chapter we use the ecomultiplex framework for investigating how individual species can facilitate parasite spread, adopting an immunisation strategy similar to the one introduced in Chapter 8. Let us say there is a vaccine immunising species populations from a given parasite, however it is available only in modest quantity so that it is possible to immunise only $\phi \ll N$ of the N available populations in the ecosystem. Assuming that the same vaccine works for all the species, which specific species of a given ecosystem promote parasite spread most and should therefore undergo vaccination? Here we show that an ecological multiplex network representation of ecosystems can indeed provide immunisation protocols that outperform either immunisation of random species or other vaccination protocols based only on the biology of species (e.g. body mass or family type). We relate these results to topological features of species within the multiplex network. We also identify a new potential mechanism with which top-predators might facilitate parasite spread rather than containing it, as suggested in the mainstream literature.

Chapter 2

Overview of Network Science Concepts

2.1 Network representation of real-world systems

Complex systems usually exhibit a group of microscopically interacting units giving rise to unexpected macroscopic behaviour: social relationships dynamically shape institutions, social norms and fashion trends (Scott, 2012); gene interactions regulate how proteins get synthesised within the human body (Caldarelli, 2007); routers manage the flow of information over the hard wired structure of the Internet (Barabási, 2015); human settlements change over time and influence the layout of road networks and power-grids (Newman, 2010). One of the major milestones in complexity theory was the development of a comprehensive set of theoretical tools for the understanding of such networked systems: network theory.

Informally, complex networks are representations of a given networked systems, with nodes (vertices) representing the basic system elements and links (edges) between nodes symbolising pairwise relationships. However not every complex system can be meaningfully interpreted as a network of connections. The network paradigm in itself (Newman, 2010; Caldarelli, 2007; Barabási, 2015) is particularly convenient for representing those complex systems in which interactions or rather constraints give rise either to a distinct structure or to a particular process crucially depending on pairwise interactions. As an example, the process could be a walk visiting the whole network by crossing its links only once. Euler (1741) showed that such a walk was impossible in the city of Königsberg by representing it as a complex network, where nodes/vertices represented the city areas and connected by bridges. Euler's approach was groundbreaking in that it was the first one using the network representation of a real-world system for the investigation of a dynamical process concerning it.

Even though representing physical systems in terms of networks became widespread only in the last 25 years, the importance of this approach can be traced also in older theoretical works from discrete mathematics focused on graph theory. It was in this context that random graphs were proposed, in the late 50s, by the mathematicians Paul Erdös and Alfred Rényi (Erdős & Rényi, 1959). Their main idea was connecting a fixed number of nodes N by building each link independently of the others with fixed probability p. Erdős & Rényi (1959) wrote that "the evolution of graphs may be considered as a rather simplified model of the evolution of certain communication nets (railways, road or electric network systems, etc.) of a country or some unit" but never explicitly used random graphs as null models for comparison with real-world networks. Although quite commonly used as null models in network theory, random graphs proved to be unsuitable for shedding light on many of the empirical features exhibited by real world networks, e.g. inhomogeneous degree distributions or high local clustering of connections (Newman, 2003, 2010).

These phenomena were later interpreted (or sometimes rediscovered) with the advent of network theory in the 90s, mainly by: (i) building a network representation of a given system, (ii) identifying its topological features and (iii) coming up with possible mechanisms capable of explaining and reproducing those observed topological features and (iv) comparing the null model and the empirical network. It is in these terms, with the comparison against more or less general null models, that the network representation of a given real-world system advances our knowledge and interpretation of that system as well. The most famous work in network theory following this scheme is the work from Barabási & Albert (1999). The authors investigated the degree distributions $P(k)$ of real-world networks, i.e. the probability $P(k)$ of finding a node with k links. Barabási & Albert (1999) identified heterogeneous degree distribution in both the World Wide Web and the power grid,

$$P(k) \propto k^{-\gamma} \text{with } \gamma > 1. \tag{2.1}$$

Following to this observation of a power-law behaviour for $P(k)$, Barabási & Albert (1999) came up with a theoretical network model where nodes appeared one at a time step and connected preferentially to higher degree nodes. This *preferential attachment* mechanism could reproduce a power-law degree distribution, with a few nodes having very large degrees. However this mechanism could not reproduce another feature exhibited by the WWW, i.e. small-worldness (Watts & Strogatz, 1998). Watts & Strogatz (1998) showed that by randomly rewiring a lattice graph it is possible to obtain small-world networks, i.e. networks where on average every node can be reached by any other within a few network hops and which resemble complete graphs at a local level[1] (Barabási, 2015). Notice that the smallest number of links d_{ij} that have to be crossed for going from one node i to another node j is called shortest path length. Small-world networks have an average shortest path length $\langle d \rangle$ among all N nodes that is "small" in the sense that it scales as $\langle d \rangle \propto \log N$. While random graphs and small-world networks both possess this distance scaling, they greatly differ in their clustering. The extent to which the neighbourhood of a node i resembles a complete graph is quantified by the local clustering coefficient c_i, defined as the ratio between the number Γ_i of links effectively present between the k_i neighbours of i and the number given by $k_i(k_i - 1)/2$ of all possible links. In formulas, the local clustering coefficient of node i in a network with undirected links is given by:

$$c_i = \frac{2\Gamma_i}{k_i(k_i - 1)}. \tag{2.2}$$

The higher local clustering, the more connections a neighbourhood has compared to a complete graph with the same nodes. High clustering and small average shortest path lengths imply that small-world networks are significantly more clustered compared to random graphs with similar average shortest distance among nodes. Small-worldness is a feature displayed by many

---

[1] The local level indicated here refers to a node and its adjacent/neighbouring nodes.

real-world complex networks, such as the mapping of synonyms relationships in the English language (Sigman & Cecchi, 2002). Cognitive scientists interpret small-worldness as the sign for an efficient navigation of lexical entries within the mental representation of words from a given language (Beckage et al., 2011; Collins & Loftus, 1975; Sigman & Cecchi, 2002). Most food webs do not display small-world structure (Dunne et al., 2002b).

The appropriate network representation of a given complex system can provide insightful quantitative information encoded in its topology. Developing null models becomes fundamental for obtaining and interpreting this information. In the past twenty years, there have been outstanding advances in this direction, prompted by the fundamental contributions that statistical mechanics (Boccaletti et al., 2006) and computer science (Mitchell, 2009) provided to network theory. On the one hand, the automatisation of data gathering in several fields led to the creation of large, publicly available databases of various empirical networks. This data availability motivated researchers to overcome the discipline boundaries and to tackle complex systems as a whole. The already mentioned powerfulness of the network paradigm ultimately benefited from this. On the other hand, the increasing amount of computational power and the development of faster theoretical algorithms allowed to tackle bigger size complex networks, whose computational modelling was not possible before. Motivated by the synergy of these developments in mathematics, physics and computer science, complex networks arose as a proper interdisciplinary tool for the investigation of real-world phenomena.

Having such a powerful representation in one hand and a set of theoretical tools for extracting information in the other is a quite powerful way of shedding light on a variety of trans-disciplinary complex systems. It is this network-based interpretation of complex phenomena that represents the main theme of this thesis.

2.2 Multi-layer networks as interpretation tools for multiplex phenomena

The mainstream network approach in the 90s and early 00s has mostly focused in treating each element of a given system as one node and each relationship as a link between nodes (Boccaletti et al., 2006; Newman, 2010). However, the increasing amount of available empirical data highlighted the shortcomings of this approach, as it does not take into account the heterogeneity of the actual relationships and the multitude of possible inter-dependencies between network nodes. We do experience such diversity in many real-world complex systems. For instance, multiple transportation networks (Morris & Barthelemy, 2012) can overlap and cover the same metropolitan areas (e.g. the metro and the bus networks). Furthermore, also countries interact via various political and economic channels (Gallotti & Barthelemy, 2014), while proteins can participate in a regulatory network according to different mechanisms (De Domenico et al., 2015). Historically, it is in the context of sociology that the concepts of *multi-relational* or *multiplex* networks arose (Padgett & Ansell, 1993; Wasserman & Faust, 1994). Social networks represent the human interactions among a given set of people. However, these interactions are clearly context-dependent and some might be present or not according to the level they are observed at. For instance, two people might be friends on Facebook but not in real life, after all. In order to fully encapsulate the complexity of their social interactions one should label the distinct interactions and represent them on different network layers (see Figure 2.1). The resulting multi-relational network is an example of a *multiplex network*, i.e. a multi-layered network where nodes appear on every layer but they can be connected according to different topologies (Kivelä et al., 2014; Boccaletti et al., 2014; De Domenico et al., 2013; Mucha et al., 2010; Szell et al., 2010; Battiston et al., 2017) on a given layer. Multiplex layers represent links of different type, e.g. co-location, kinship, commercial or collaboration relationships between individuals in a social environment. Replicas of the same node across layers can be either formally connected by inter-layer links (full multiplex representation (De Domenico et al., 2013; Battiston et al., 2014)) or not (edge-coloured graph or multiplex network in the sense of the computational social sciences (Wasserman & Faust, 1994)).

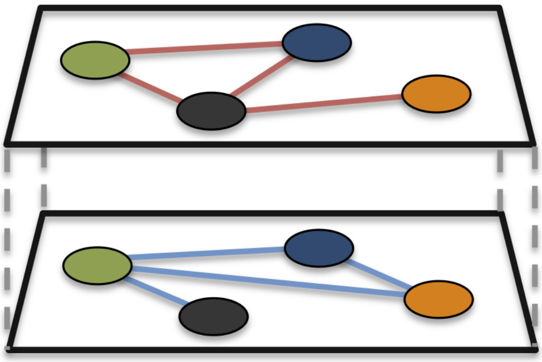


Figure 2.1: Top: A simple multiplex with only two network layers. Multiplex networks are multi-relational networks in which all the nodes are replicated on every layer.

In the last five years, a rather considerable amount of effort has been dedicated to building a comprehensive framework for the characterisation of multiplex networks, inside the broader panorama of interdependent networks (Buldyrev et al., 2010; Kivelä et al., 2014; Boccaletti et al., 2014; De Domenico et al., 2013; Battiston et al., 2017). The most exciting aspect of multiplex networks is that their layered structure gives rise to phenomena that would be unexpected in mainstream single-layer networks. This happens mainly because of three factors that are encoded in the multi-layered structure (Battiston et al., 2014; De Domenico et al., 2015; Kivelä et al., 2014; Battiston et al., 2017; De Domenico et al., 2014; Boccaletti et al., 2014): (i) the multiplex shortest path length, the (ii) presence of correlated or uncorrelated overlap and (iii) the potential reducibility of a set of layers by aggregating some or all of them. The following sections will introduce the above concepts, which will be later explored and related to the mental lexicon and ecosystems in the main parts of this thesis.

2.2.1 Multiplex geodesic distance, viable clusters and super-diffusion

Nicosia et al. (2013) and Battiston et al. (2014) define the multiplex shortest path distance $d_M(i,j)$ between two nodes i and j as the length of the shortest path $(i^{[\alpha]}, ..., j^{[\beta]})$ between node i and j considering also jumps between layers $\{\alpha, \beta, ...\}$. The possibility of transitioning between layers can dramatically change the average path distance of nodes within a given multiplex when compared to the individual layers. For instance, two nodes could be disconnected on layer α, so that their distance would be $d_\alpha(i,j) = \infty$, but adjacent in another one, so that $d_M(i,j) = 1$.

Therefore the presence of multiple layers adds more pathways for navigating the network from node to node. This increased connectivity does play a fundamental role in the analysis of the whole multiplex structure, as underlined later in the thesis (cf. Chapter 5 and 6).

The presence of multi-layer paths requires also a redefinition of connectedness. Single-layer networks are defined as connected if there exists at least one path connecting any two of their nodes. In this thesis we align with the definition of multiplex connectedness provided in (De Domenico et al., 2014): a multiplex network is connected if there exists at least one path connecting any two of its nodes allowing for jumps between layers. In other words, a multiplex network is connected if projecting all links onto one layer results in a connected network. The projection is also called "aggregate network" and it will play an important role in the following.

Notice that there is more than one definition of connectedness in multiplex networks. Rather than considering the aggregate network, connectedness could be imposed on each individual layer, thus identifying the largest set of nodes across layers that are mutually connected on each individual layer. This definition finds the *largest mutually connected component* of a multiplex network or its *largest viable cluster* (Baxter et al., 2016) or mutually connected giant component (Boccaletti et al., 2014) and it is a more restrictive definition compared to one by De Domenico et al. (2014). The largest viable cluster of a multiplex network can act as the counterpart of a densely connected network core in single-layer networks, as shown later in the thesis (cf. Chapter 6). Notice that viable clusters arise from the presence of more than one layer in multiplex networks. Their single-layer counterparts are largest connected components (in finitely sized networks) or giant components (in the limit of infinite size).

De Domenico et al. (2014) built on the concept of multi-layer paths by introducing several types of random walks on multiplex networks with heterogeneous transition probabilities between nodes. They demonstrated that the average time T_C required for a classic random walker to cover most of the nodes in a multiplex crucially depends on the jumping probabilities along inter-layer connections. However, T_C can be either bounded by the times necessary for covering each of the layers separately, i.e. $T_\alpha < T_C < T_\beta$ and the walk is said to be *intra-diffusive*, or it can be smaller than all of these times, i.e. $T_C < \min(T_\alpha, T_\beta, ...)$ and the walk is said to be *super-diffusive*. This super-diffusion phenomenon arises only in multiple layered networks and, in some instances, it can be related to the higher fragility displayed by some real-world multiplex networks to cascade failures (De Domenico et al., 2014; Boccaletti et al., 2014).

2.2.2 Overlap and multiplexity

The multi-layered structure can significantly alter the connectedness of nodes when compared to single-layer networks. In addition, intra-layer connections can be organised in a more or less correlated way among layers. Several attempts have been made in order to quantify these inter-layer dependencies (Min et al., 2014; Battiston et al., 2014; Nicosia & Latora, 2015; Bianconi, 2013). For instance, such dependencies can be measured via correlation metrics (Battiston et al., 2014), like the conditional probability $P(A_{ij}^{[\alpha]}|A_{ij}^{[\beta]})$ of finding a link at layer α given the presence of a link between the same nodes in layer β:

$$P(A_{ij}^{[\alpha]}|A_{ij}^{[\beta]}) = \frac{\sum_{ij} A_{ij}^{[\alpha]} A_{ij}^{[\beta]}}{\sum_{ij} A_{ij}^{[\beta]}}. \tag{2.3}$$

Gemmetto & Garlaschelli (2015) underlined that these approaches implicitly take into account random uncorrelated graphs, with homogeneous degree distributions. The authors remarked that this assumption of uniformity in the probability distributions strongly contrasts with the empirical evidence from real-world networked systems (Barabási, 2015; Kivelä et al., 2014; Boccaletti et al., 2014). With this motivation, Gemmetto & Garlaschelli (2015) introduced the notion of *multiplexity* as a suitable measure for inter-layer correlation in the form of edge overlap. In formulas,

$$M_{\alpha\beta} = \frac{2\sum_{i\neq j}\min\{A_{ij}^{(\alpha)}, A_{ij}^{(\beta)}\}}{L_{TOT}^{[\alpha]} + L_{TOT}^{[\beta]}}, \quad \mu_{\alpha\beta} = \frac{M_{\alpha\beta} - \langle M_{\alpha\beta}\rangle}{1 - \langle M_{\alpha\beta}\rangle}, \tag{2.4}$$

where $M_{\alpha\beta}$ is the *raw* multiplexity between layers α and β (i.e. the ratio of links that overlap between two layers (Bianconi, 2013)), $L_{TOT}^{[.]}$ is the number of edges in a given layer and $\mu_{\alpha\beta}$ is the *normalised* multiplexity, which is rescaled compared to a reference null model value $\langle M_{\alpha\beta}\rangle$. Gemmetto & Garlaschelli (2015) adopted $\mu_{\alpha\beta}$ for investigating the International Trade Network (ITN) and the European Airport Network (EAN). They reported that adopting an ensemble of random graphs as a null model in μ can lead to very different results compared to the adoption of a configuration model, which keeps into account also the heterogeneity in the degree distributions of real data. Gemmetto & Garlaschelli (2015) found that in the EAN, where different layers were characterised by different hubs, the multiplexity was practically absent, irrespective of the null model considered. Instead, in the ITN, where the same nodes tended to be hubs in all layers, the apparently strong inter-layer correlations are hugely reduced once the empirical degree distribution is considered. Their results imply that most dependencies are actually encoded in the correlated degree sequence of the ITN multiplex and not in the assortment of links. In other words, the rescaled multiplexity measure μ is capable of discriminating between edge overlap and inter-layer degree correlations Gemmetto & Garlaschelli (2015). Multiplexity was adopted in the structural analysis of real-world multiplex lexical networks for children in this thesis (cf. Chapter 5).

Multiplexity and degree-correlations can be fundamental for dynamics on multiplex networks. Lee et al. (2012) studied multiplexes consisting of two degree-correlated ER random graphs. They found that negative degree correlations actually increase the percolation threshold p_C, while positive correlations decrease it. Furthermore, they showed that $p_C \to 0$ when $r_{12} \to 1$, i.e. when each node has the same degree on both the layers. Further, Min et al. (2014) found that correlated degrees also affect the structural robustness of multiplexes. For maximally-correlated ER multiplexes ($r_{12} = 1$), the network structure was highly resilient to random failure. In contrast, anti-correlated ER multiplexes ($r_{12} = -1$) resulted robust against degree-based targeted attacks, but also quite vulnerable to random failure (Min et al., 2014).

2.2.3 Multiplex reducibility analysis

Having more layers does not always provide more topological information about interactions among a given set of nodes. As a trivial example, consider a fictional multiplex network where all layers are copies of a single one. In such case the multiplex representation would be inappropriate as considering different layers would be cumbersome while providing no additional information on the system. While individual layers can differ from each other in terms of their topology, they can also display rather high link overlap and thus contain redundant topological information. Therefore, what would be the optimal number of layers to consider in a tradeoff between having the most topological information available but the least link redundancy?

De Domenico et al. (2015) suggested a structural reducibility analysis based on the Von Neumann entropy for quantifying topological redundancy in a given multiplex network. The redundancy of paths on the whole multiplex network is measured compared to the aggregate network. In other words, structural reducibility quantifies how different the multiplex network is from its aggregate when layers are considered as distinct or when some or them are aggregated.

The structural reducibility analysis uses an entropic measure, based on the entropy of the spectrum of its single-layer and degree-normalised Laplacian $L = (D - A)/2K$, where D is a diagonal matrix having the node degrees on its main diagonal, K is the number of links in the network and A is the adjacency matrix ($a_{ij} = 1$ if nodes i and j are adjacent and 0 otherwise).

Let us denote with $\{\lambda_i^*\}_i$ the set of eigenvalues of the Laplacian of the aggregate network of a given multiplex network. Then the Von Neumann entropy H^* is defined as:

$$H^* = -\mathrm{Tr}(L^* \log_2 L^*) = -\sum_i \lambda_i^* \log_2 \lambda_i^*, \tag{2.5}$$

where Tr indicates the trace operator on matrices. The Von Neumann entropy H^* of the aggregated network encapsulates information about the network topology and it can be used as a reference value. De Domenico et al. (2015) suggest to quantify the difference between the aggregated network and the multiplex structure in terms of Von Neumann entropy. To this aim, the authors compute the Von Neumann entropy of the individual multiplex layers, let us denote the entropy of layer α with h_α. The authors then approximate the Von Neumann entropy of the whole multiplex network as the average of the entropies of the individual layers $H_M(C) = \sum_{\alpha=1}^{|C|} h_\alpha / |C|$ relative to the configuration C with $|C|$ different layers. This average relies on the assumption that link correlations across layers can be neglected. Notice also that $|C|$ can be either the number of the layers present in the original multiplex network or that the number of distinct layer after some of the original ones were aggregated together. Intuitively, the ratio H_M/H^* is upper bounded by 1 ($H_M = H^*$ when all the layers have been aggregated together). Therefore De Domenico et al. (2015) suggest to rescale it by 1 and define the following reducibility index $q(C)$ in the multiplex network:

$$q(C) = 1 - \frac{H_M(C)}{H^*}. \tag{2.6}$$

The above formula is simple and elegant: aggregating the multiplex layers in the configuration C leads to a certain average Von Neumann entropy, hence how is it different from the Von Neumann entropy of the aggregated multiplex? $q(C)$ ranges between 0 and a given maximum value. On the one hand, q is minimum when the aggregate network is equivalent to the multiplex in configuration C and this can happen if and only if $|C| = 1$ and all layers have been aggregated. This means that $q(C) \geq 0$. On the other hand, $q(\cdot)$ is maximum when the multiplex structure in configuration C is maximally different or the most distinguishable from the aggregated network. It is in this configuration that De Domenico et al. (2015) consider the multiplex structure as providing the most information about the multi-relational patterns within it.

Notice that $q(\cdot)$ is a relative measure: it cannot provide insights on the multiplex network having high or low entropy. $q(\cdot)$ is rather an entropy-based approach for defining how different a given multiplex configuration is compared to its aggregated counterpart and this is why the authors call $q(\cdot)$ a relative entropy. In general, $q(\cdot)$ can either increase or decrease when two layers are aggregated, depending on their connectivity patterns. More in detail, $q(\cdot)$ decreases when $H_M(C)$ increases. However, the increase in mean entropy ΔH_M is positive mainly in two cases: (i) when two layers with very different link densities are aggregated into one layer or (ii) when new structural patterns that were absent in the original layers emerge in the aggregated one. In both these two cases it is clear that the aggregation is not appropriate, and this is why configurations which minimise H_M and thus maximise the relative entropy $q(\cdot)$ are preferred, instead.

In the cases when $q(\cdot)$ is maximum but no aggregation was performed, i.e. layers were kept as separated, then the original multiplex configuration is considered being irreducible: performing any layer aggregation would imply a loss of structural information.

Notice that not all configurations have to be verified. Since the number of different configurations increases as the Bell number e^{e^M-1} for a multiplex with M layers, a brute-force exploration of each configuration would become impractical already from $M = 20$. Instead, De Domenico et al. (2015) consider a greedy-algorithm that explores the configurations where more similar layers are aggregated first. Similarity is considered via the Jensen-Shannon divergence, that is a function of the layers Von Neumann entropy (we refer to (De Domenico et al., 2015) for further details). Entropy-based measures such as the above one can also be used for clustering similar layers, as performed in (De Domenico & Biamonte, 2016), by using a generalised Jensen-Shannon divergence.

2.3 The added value of multiplex networks in modelling real world systems

As reported in the previous section, the multi-layered structure of multiplex networks makes them substantially different from single-layer networks. However, what is the added value of multiplex networks in terms of modelling? How can the multi-relational nature of multiplex network links be exploited for gaining insights about patterns in complex system in the real world? In the following, we review some of the main contributions multiplex network modelling has made in the last few years. We then tie-in the main contributions of this thesis, providing new multi-layer modelling approaches to ecological psychology and ecosystems.

2.3.1 Game theory on multiplex networks

The study of games on structured populations has proved quite a prolific research field, at the interface among statistical physics, evolutionary biology, and the computational social sciences (Weibull, 1997). In particular, evolutionary game theory on graphs studies the interplay between population structure and the adoption of precise strategies in prototypical dilemma situations (Gómez-Gardenes et al., 2012; Wang et al., 2014). For instance, consider two players playing with each other and receiving a certain payoff according to their respective strategies: cooperation or defection. When both players cooperate, they receive a payoff that is lower compared to when both of them defect. Also, if one defects and the other cooperates, the defector receives a higher payoff then the cooperator. This higher payoff is called the *temptation to defect* in the game-theoretic jargon (Weibull, 1997; Boccaletti et al., 2014). Therefore, even if cooperation is not the optimal strategy for both players in pairwise Prisoner's Dilemma-like games (like the one from the example), cooperation is actually observed in many human and biological systems (Weibull, 1997). This gap between phenomenology and theory motivated an extensive number of studies aimed at understanding the emergence and robustness of cooperation in structured populations (Weibull, 1997; Gómez-Gardenes et al., 2012; Wang et al., 2014). Network structure is, in fact, a pivotal element for the survival of cooperative strategies. In scale-free networks, hubs tend to act as "leaders", imposing their strategies to their (many) neighbours (Santos & Pacheco, 2005). This mechanism creates clusters of cooperators, shielding themselves from getting in contact with defection strategies. Such segregation phenomenon provides additional resilience to cooperative strategies and it is called *network reciprocity* (Santos & Pacheco, 2005; Weibull, 1997; Boccaletti et al., 2014). Single-layer random graphs do not display hubs and hence sustain cooperation less than scale-free single layer topologies. In (Gómez-Gardenes et al., 2012), Gómez et al. numerically studied the weak version of the Prisoner's Dilemma on a multiplex with up to $M = 20$ layered random graphs. They used a synchronous dynamics combining: (i) the adoption of local cooperative or defective strategies on each layer with the (ii) total payoffs of a node when playing with its neighbours on all the layers. Even for multiplex networks with ER sparse graphs, numerical simulations showed that the multiplex structure dramatically improves the resilience of cooperative strategies, compared to the case of single layer ER with the same average degree. In this way, multiplexity represents another possible approach to achieve network reciprocity without the need for scale-free topologies.

A different setup was tested in (Wang et al., 2014). The authors investigated a 2-layer multiplex (i.e. duplex) network with scale-free intra-layer topologies. One network, called the interaction layer, was used for the accumulation of payoffs, while the other layer served as the updating network for the evolution of strategies through imitation. The authors tested an evolutionary synchronous dynamics but they tuned the inter-layer degree correlations, generating assortative mixing by degree in one layer and disassortative mixing in the other. Their computational results showed that the evolution of cooperation on scale-free interdependent networks can be critically impaired by the presence of inter-layer degree correlations. The authors attributed this inhibition of cooperation to a drastic reduction of network reciprocity through the multiplex structure: clusters of cooperators can be exposed to defectors more systematically in the duplex network rather than in single-layer scale-free networks (Wang et al., 2014).

2.3.2 Brain multiplex networks

The human brain has been extensively investigated through the tools of network theory in the last few years (Bullmore & Sporns, 2009; De Domenico, 2017). One of the most prolific research project in this field is surely the Human Connectome Project (HCP), mapping the connectivity patterns in the human brain. HCP started in 2009 with a $30 million grant and it resulted in more than 65 international publications until this thesis was written (see www.humanconnectomeproject.org/ for more details).

The HCP studies the connectivity of the human brain in terms of (i) anatomical connectivity and (ii) functional connectivity (Sporns et al., 2004). Two brain regions are anatomically connected if they share at least one pathway through cerebral white matter. Pathways can be charted by means of brain dissection and therefore relate to the physical conformation of the human brain. On the other hand, two brain regions are functionally connected if they tend to be activated at the same time during given cognitive tasks. Functional connectivity is usually traced through functional magnetic resonance imaging (fMRI), which basically highlights intense metabolic processes (which correlate with neural activity). The strength of functional connectivity can also be empirically estimated by filtering and correlating fMRI signals at a given frequency (De Domenico et al., 2016).

Anatomical and functional connectivities give rise to the so-called connectome (Sporns et al., 2005; Towlson et al., 2013), i.e. a networked system where brain regions are connected according trough anatomical and functional patterns. While the pioneering studies in brain networks analysed individual snapshots of the connectome (Sporns et al., 2004, 2005), it is only recently that the interest gained by multi-layered networks in the last few years motivated a modelling of the connectome as a multiplex network.

In (Horn et al., 2014), the authors investigated the connectome of 19 healthy subjects between 21 and 31 years of age. The authors highlighted a statistically significant correlation in overlap of edges across the functional and the anatomical networks (Spearman Rho ~ 0.31). Peak values (Spearman Rho ~ 0.42) of positive correlation were found in the so-called Default Mode Network, a network that shows high activity values when people are at rest and low activity values when people are actively involved in cognitive tasks. This Default Mode Network contains also several connectome hubs, with high betweenness centralities. The Default Mode area of the brain, in fact, connects several regions of the front and back parts of the brain (Horn et al., 2014). The high edge overlap found in indicates that the anatomical profile of the brain and its functional repertoire seem to shape each other, in particular in highly central and well connected subregions of the human brain.

Battiston et al. (2017) investigated multiplex network motifs (i.e., overrepresented network subgraphs composed of only a few nodes) of the human connectome. In contrast with previous approaches, which consider motif analysis of single brain modalities, Battiston et al. (2017) are able to show further evidence for joint link correlations within human brain networks. These results are however not symmetric: The functional coordination of brain dynamics at rest is constrained by the topology of its anatomy while structural connections themselves might be necessary but not sufficient for the existence of functional correlations between two regions of the brain.

The work of (De Domenico et al., 2016) focused on functional connectivity only. In fact, the identification of connections in functional brain networks depends on the observed frequency range in the fMRI procedure Sporns et al. (2004). Quantitatively considering several different frequency bands simultaneously was the main achievement of (De Domenico et al., 2016). The authors identified each frequency component with one layer of a multiplex network, in which nodes remain brain regions. The authors showed that each layer in the functional multiplex network provided unique topological information, with strong disassortative mixing by degree among layers. This led to hubs in the multiplex network being in general different from those ones found by aggregating the measured signals as in previous studies. Also, the authors compared the functional multiplex networks of healthy and schizophrenic individuals. Interestingly, the layers of healthy individuals showed a higher dissimilarity compared to the network layers coming from schizophrenic individuals. Furthermore, multiplex hubs were located in the frontal cortices in healthy individuals, whereas in schizophrenic individuals multiplex hubs were found in posterior parts of the brain. The first of its kind, the work from De Domenico et al. (2016) showed how considering the multiplex centrality profile of brain regions at different frequencies allow to discriminate between control and schizophrenic groups of individuals more accurately than aggregated single functional layer approaches. A similar methodology based on multi-layer brain networks were used for detecting structural changes in patients with Alzheimer's Disease in (Guillon et al., 2017).

2.3.3 Versatility in real multiplex networks

Finding the most central nodes in complex networks is fundamental in a variety of real-world scenarios (Newman, 2010), e.g. finding the most fragile agents in failure cascades or pivotal disease spreaders in epidemics. While a plethora of network centrality measures has been suggested in single-layer networks (Caldarelli, 2007; Newman, 2010), the problem of centrality is still a relatively unexplored subfield in multiplex networks. It is only recently that (De Domenico et al., 2015) suggested a cohesive mathematical framework for finding versatile nodes in multilayer networks, i.e. those nodes playing a pivotal role in bridging together the different layered networks. The authors adopted the tensorial formulation of multi-layered networks (De Domenico et al., 2013) in order to generalise the eigenvector and the PageRank centralities for multi-layered networks. Compared to the definitions suggested by Halu et al. (2013) and Solá et al. (2013), the definition of (De Domenico et al., 2015) does not require to artificially combine the centralities of different layers through weighted averages with arbitrarily chosen weights.

In single-layer networks, eigenvector centrality is based on an iterative procedure giving to each node a centrality score that is the weighted sum of the scores of its neighbours (Newman, 2010). Therefore, the eigenvector centrality gives high centrality scores to vertices that are connected to many other well-connected vertices. In formulas, this iterative procedure translates into computing the eigenvector $\mathbf{u}$ related to the largest eigenvalue λ of the network adjacency matrix A (where the entry $a_{ij} = 1$ if i and j are connected, $a_{ij} = 0$ otherwise):

$$\mathbf{u}A = \lambda\mathbf{u}. \tag{2.7}$$

The entries of $\mathbf{u}$ are the eigenvector centralities of the network nodes. De Domenico et al. (2015) generalised the eigenvector problem of finding $\mathbf{u}$ into a tensorial formulation, where the leading eigentensor $\theta_{i\alpha}$ of the super-adjacency tensor $M^{i\alpha}_{j\beta}$ has to be computed. While $M^{i\alpha}_{j\beta}$ has entries in $\mathbb{R}^{N\times N\times M\times M}$ for a general multi-layered network with N nodes and M layers, in case of multiplex networks $M^{i\alpha}_{j\beta}$ can be flattened and recast into a super-adjacency matrix $\tilde{A}$. $\tilde{A}$ has the adjacency matrices of individual layers on its main diagonal and the inter-layer matrices as off-diagonal elements. For instance, for edge-coloured graphs, i.e. multiplex networks with no explicit inter-layer links, the inter-layer matrices have all elements equal to 0. This tensor-to-matrix mapping allows for a recasting of the tensorial problem of finding $\theta_{i\alpha}$ into an eigenvector problem for $\tilde{A}$.

De Domenico et al. (2015) explored also a variant of eigenvector centrality, Google's PageRank centrality, in the context of multi-layered networks . Page-rank centralities represent the likelihood that a random walker following links arrives at any particular node on the network, where the random walker jumps to a neighbour with rate r and teleports to any other node in the network with another rate r_0. In the context of interconnected multi-layer networks, teleportation might occur to any node replica in any layer.

De Domenico et al. (2015) applied their multiplex versions of eigenvector and PageRank centralities to the investigation of (i) a multiplex-like network of scientific reference on Wikipedia and the multiplex network of flight routes in European airports, among others. In the multiplex-like network of scientific co-citations nodes represented scientists and layers represented scientific disciplines (e.g. the layer of physics, the layer of chemistry, etc.). All scientists being X (e.g. chemists) were connected on the X-discipline (e.g. chemistry) layer if their Wikipedia web-pages contained a reference (e.g. chemist A references chemist B in his/her Wikipedia web-page). In the multiplex network of flight routes in European airports, nodes represent airports and layers represent different companies connecting the airports through flight routes. In these two very different systems, applying the single-layer eigenvector and PageRank centralities to the aggregated networks provided substantially different results from the versatile multiplex counterparts applied to the original multi-layer structures. Interestingly, in the scientific multiplex the PageRank versatility correlated positively with interdisciplinarity, i.e. highlighting the scientists with outstanding results in several, different scientific areas. Multiplex versatility highlighted trans-disciplinary nodes also in the analysis of the American Physics Society citation database (Omodei et al., 2017). In the airport multiplex network, versatility PageRank proved to be 150% better than PageRank (on the aggregate network) in finding the airports more prone to passenger traffic congestion. These differences indicate that the multiplex structure can provide more insights into some features exhibited by real-world systems.

Part I

Investigating language and cognition through multiplex lexical networks

A Foreword to Part I

The next three Chapters composing Part I of this thesis originated from work in collaboration with:

- my main PhD supervisor Markus Brede (University of Southampton, Southampton, UK);
- my external supervisor Manlio De Domenico (Universitat Rovira i Virgili, Tarragona, Spain);
- Nicole M. Beckage (University of Kansas, Lawrence, US);

The next three Chapters represent papers I published with the above co-authors. In all the papers I am the main author. More in detail, I conceived, developed and simplified most of the multi-layer and multiplex lexical framework presented in the papers, cleaned and processed the data, conceived the numerical experiments, analysed and interpreted the results, structured, wrote and revised most of the manuscripts reported here. I conceived the topic of multiplex linguistic networks for my PhD on my own initiative. My main supervisor M.B. assisted me in the design of the technical parts of all the next three papers and he also developed part of the simulation code for results in Chapter 4. I involved N.B. in the papers represented by Chapters 6 and 7 as an expert on cognitive network science. N.B. provided and cleaned the data for Chapter 6 and helped with the interpretation of the results in Chapters 6 and 7. I asked for M.D.D. to join my supervisory team as an expert on multilayer networks. M.D.D. helped with the technical and statistical techniques adopted in Chapter 7. All the co-authors contributed to writing and reviewing the manuscripts.

Chapter 3

A review of linguistic networks

> Language is the dress of thought.
>
> *The Life of Cowley*
> Samuel Johnson

One can imagine the human mental lexicon as a huge database where words are stored together with additional information, and in which words are related by specific correlations which ease navigation, e.g. words can be opposites or synonyms, might be pronounced in similar or dissimilar ways, be related to the same context area, etc. Following a connectionist approach (Aitchison, 2012; Griffiths et al., 2007) the HML can be interpreted as the representation of the biological patterns of synchrony/asynchrony among the 10^{10} neurones and 10^{14} synapses of the human brain (Baronchelli et al., 2013; Bullmore & Sporns, 2009). Individual aspects of the mental lexicon have been analysed in the previous literature even before the advent of network theory.

3.1 Single-layer network models

Historically, the idea of a network of semantic relationships between concepts is said to be traceable all the way back to Aristotle (Cong & Liu, 2014). However, the term "semantic network" dates back at least to Collins and Quillian, in the late 60s (Collins & Quillian, 1969). The intuition behind semantic networks is that nodes represent words while connections represent semantic relationships, possibly weighted according to (a more or less arbitrary) concept similarity measure.

3.1.1 Free Associations

Semantic relationships are a product of human cognition and cannot be fully measurable in the real world without either recurring to empirical data, i.e. experiments or language corpora (Aitchison, 2012; Beckage et al., 2011), or considering specific rules of similarities, i.e. defining synonymy or generalisation rules as in dictionaries (Miller, 1995; De Deyne & Storms, 2008). The

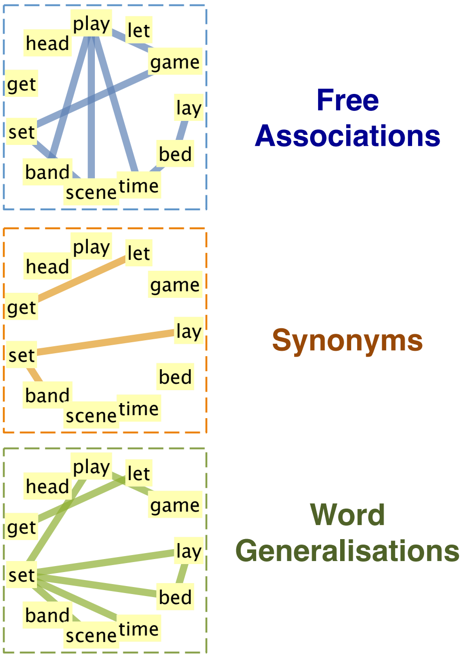


Figure 3.1: Examples of semantic networks based on free associations (Kiss et al., 1972), synonyms and word generalisations (Miller, 1995).

gathering of language corpora available with the advent of the computers motivated researchers to investigate a specific type of semantic network where words/concepts are connected to each other according to *free association norms* (Kiss et al., 1972; Aitchison, 2012). The term "norm" relates to empirical data used as a reference value for statistical abnormalities, in this context data that acts as an approximation for semantic relationships. In order to retrieve such norms, in the early 70s, Kiss et al. performed an extensive experiment involving over 700 British undergraduate students with different social and cultural backgrounds (Kiss et al., 1972). The gathered empirical data resulted in the Edinburgh Associative Thesaurus (EAT), namely a large scale set of free associations. These relationships were obtained by showing a given word (i.e. a *cue*) to several people and asking them to identify the first word coming to their minds (i.e. *the target words*) upon perceiving such stimulus. This linguistic task was repeated with a given number of subjects and severely incomplete or biased tasks were rejected. The resulting network contained more than 23.000 words and it displayed a power-law like in-degree distribution (Amancio et al., 2012). A subgraph of the dataset is reported in Figure 3.1. With a similar empirical procedure, another database about free associations but considering only 5.000 words was collected in the 90s, namely the Southern Florida Association (SFA) norms (Nelson et al., 2004).

Similarly to almost every other linguistic network, also the EAT and the SFA networks display the small-world feature. Applications of free association networks are quite recent in the scientific literature. Amancio et al. (2012) used the free associations for measuring the consistency of word usage in written texts and thus identifying authors based on their associations in scientific papers.

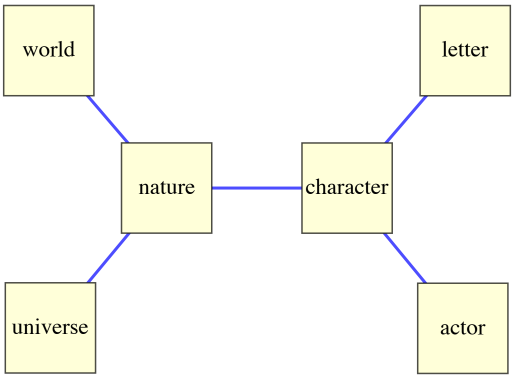


Figure 3.2: Polysemy refers to one word relating to a variety of context dependent meanings. Motter et al. suggested for the first time that polysemy was related to the small-world, scale-free topology of many semantic networks. Sigman and Cecchi showed that polysemy plays a fundamental role in creating shortcuts in semantic networks. These shortcuts ultimately increase the navigability of the mental lexicon network by decreasing the shortest path lengh distance between lexically different concepts. For instance, in the synonym network "universe" and "actor" are at distance 4 thanks to the polysemic link between "character" and "nature".

3.1.2 Linguistic relationships from dictionaries

The increasing number of digitally available dictionaries motivated the larger number of works focusing on rigid theoretical semantic dependencies rather than on word association (Beckage & Colunga, 2015). Motter et al. (2002) constructed a semantic network of synonyms by using the Moby thesaurus dictionary. Words appearing as synonyms in the dictionary were connected. The authors found that the resulting network exhibited small-worldness and attributed it to *polysemy*, which allows for several context-dependent links to be present at the same time among words having more than one meaning (i.e. being polysemous). In Figure 3.2, "character" as a complexion is synonym with "nature", while the same word as a graphic symbol is synonym with "letter".

The large-scale structure of semantic networks was extensively analysed also by Steyvers & Tenenbaum (2005). They analysed separately networks based on: (i) free associations, (ii) synonyms and (iii) generalisations. A word A is a generalisation of word B if the semantic area of A is wider than the one of B, e.g. "bird" is a generalisation of "dove". All the resulting semantic networks exhibited the small-world property. Furthermore, all the empirical networks showed strong correlations between word frequency, network degree and age of acquisition. In detail, Steyvers and Tenenbaum showed that higher frequency words tend to have more semantic connections and tend to be acquired at earlier stage of development too. Therefore, the authors showed an interaction among network topology, word frequency and word acquisition, with strongest effects evident on higher frequency words. These findings shed light on an interplay between network topology and language learning (Beckage et al., 2011; Steyvers & Tenenbaum, 2005).

It has been experimentally proved that semantic networks reflect patterns in word retrieval and memory tasks (Beckage et al., 2011; Collins & Loftus, 1975; De Deyne & Storms, 2008). Collins & Loftus (1975) showed that words closer on the network topology tend to be retrieved together

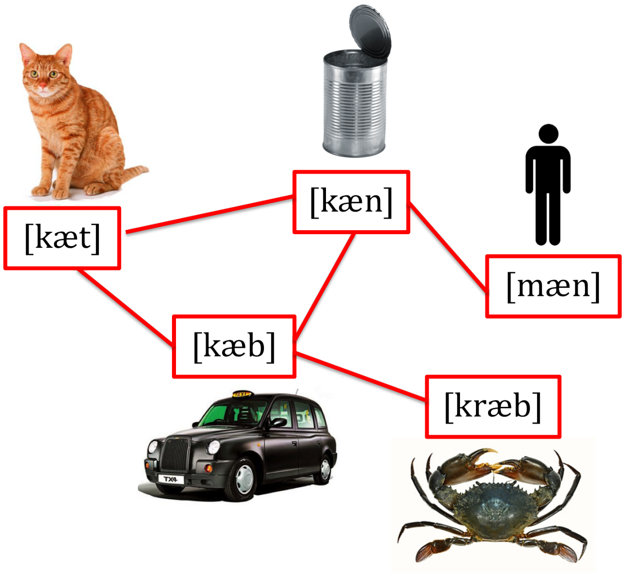


Figure 3.3: Example of a phonological network for the English language. Nodes represent words, which are connected together if they sound phonologically similarly (i.e. they differ by the addition, substitution or deletion of one phoneme only).

in shorter times. This finding indicates that the shortest path distance on the semantic levels matters for human search time in cognitive processing.

3.1.3 Phonological networks

Linguistic networks can be built also from sound similarities among words. Vitevitch (2008) built a phonological network (PN), based on the Merriam Webster Pocket Dictionary, with nodes representing phonetic word transcriptions and with links connecting phonologically similar words (i.e. words differing in the addition, deletion or substitution of one phoneme Luce & Pisoni (1998)). For instance, the two words “cat” and “can” would be connected in such a linguistic network (see Figure 3.3). Adopting such metric for the network construction led to a convenient equivalence between node degree and the so-called *phonological neighbourhood density* (PND) of a given word, i.e. the number of its phonologically similar words (Vitevitch, 2008; Sadat et al., 2014). By performing memory tasks and lexicon retrieval experiments, Vitevitch found that the network degree/PND correlated positively with speech errors such as malapropism[1] in high frequency words (Vitevitch, 1997). Additionally, in a later work, Vitevitch and Chan found that perceptual identification of words/concepts was harder for words in the phonological network with higher local clustering coefficient (i.e. percent of neighbours that are themselves connected) (Chan & Vitevitch, 2010). These two findings lead to a clear indication: the topology of a word in the phonological network significantly relates to its cognitive processing.

---

[1] A malapropism is a type of word speech error where a target word is erroneously substituted by a phonologically similar word but from a different semantic context.

3.2 Precursors of multiplex linguistic network models

A genuine multiplex approach constitutes a novelty in the field of computational linguistics, although precursors of multi-layer network studies were available in the literature (Liu & Cong, 2014; Ferrer-i Cancho, 2014; Martinčić-Ipšić et al., 2016) before the work from this thesis was published. This section briefly reviews such precursors.

The multi-relational semantic networks from Sigman and Cecchi cannot be considered a multiplex network (Sigman & Cecchi, 2002). In fact, they did analyse the structure of 66025 nouns in the WordNet database (Miller, 1995) by taking into account different semantic relationships but they also aggregated these relationships in single-layer or simplex networks. Sigman & Cecchi (2002) considered (i) *generalisations/specifications* (e.g. rose - is a type of - flower or vice versa); (ii) *antonymy* (e.g. good - is the opposite of - bad); (iii) *meronymy/holonymy* (e.g. display - is a part of - laptop); (iv) *polysemic relationships* (e.g. (department) head - can also mean - head (of a bird)). The authors aggregated different combinations of these relationships into single-layer networks and showed that the addition of polysemy creates shortcuts in the hierarchical structure of the generalisation tree, making semantically distant concepts closer on the network topology. See Figure 3.2 for an example. It has to be underlined that the work from Sigman and Cecchi dates almost ten years before the mathematics of multilayer networks. Therefore, their intuition about considering more than one type of semantic relationships is remarkable but it ultimately does not exploit the machinery behind multiplexes. Sigman and Cecchi simply aggregated connections but did not ask any research question about edge overlap, intra-layer correlations or multiplex structure as a whole.

A truly multi-relational network approach for modelling language was suggested by Liu & Cong (2014). Liu and Cong analysed the patterns of syntactic dependencies in the human mental lexicon at different scales of observation, ranging from the co-occurrences of characters in words, passing by co-occurrences of words in sentences and up to syntactic dependencies among words. The authors modelled these interactions as 4-layer network: (i) a predicate-argument network, (ii) a syntactic dependency network, (iii) a word co-occurrence network, and (iv) a Chinese-character co-occurrence network. All the layers were based on empirical data, i.e. sentences transcribed from the news of China Central Television. Let us explain the individual layers via an example: “John put an envelope on the table”. In layer (i), “John” is the argument of the predicate “put” hence a link between this two words is present. In layer (ii), “on” is a specification of “put” so a link between these two words is present. Words “envelope” and “on” co-occur hence a link between them is present on layer (iii). On layer (iv), nodes represent characters in a given word and are connected if they co-occur in at least one network. Inter-layer links are not explicitly considered, as the authors limit their network analysis to inter-layer overlap measures, which are also the main novelty of their approach. The three syntactic layers overlapped in almost 50% of their links, implying that almost half of the syntactic dependency relations in modern Chinese sentences are between adjacent words. All the four layers were separately analysed and found to be small-worlds and scale-free, with hubs in networks (ii) and (iii) being preferentially monosyllabic words, mainly prepositions. All the four layers exhibited disassortative mixing by degree: linguistic units with high degree tend to connect to those with low degree. This reflects the structure of syntactic dependencies, which associates general words or prepositions with more specific nouns. Although Cong and Liu applied for the first time a multi-layer approach to language modelling, their analysis focused only on syntactic relationships but did not condense together different aspects of the mental lexicon such as semantics or phonology. The co-occurrence layer in particular is of difficult psycholinguistic interpretation, as co-occurrences in adults have a limited interpretative power of cognitive phenomena (Zweig, 2015) and are rather considered informative only in terms of grammar relationships, which are more related to language structure and evolution rather than cognition Ninio (2014). A similar analysis on the dependency structure of language was performed also in English and Croatian by Martinčić-Ipšić et al. (2016). However, in both these two works the analysis remained purely topological: only correlations of the observed structure were quantified but the machinery of multi-layer measures was not used as a generative model for explaining linguistic patterns in additional data.

3.3 Conclusions and Research Questions

Within the networks community, multiplex networks constitute a novel and exciting research field, attracting increasingly more attention (Boccaletti et al., 2014; De Domenico et al., 2015; Kivelä et al., 2014; Battiston et al., 2017). The above literature review highlights those features of multiplex modelling that could be potentially insightful in the analysis of real-world multiplex systems such as words in the mind. Multiplexity, i.e. multi-relational interactions, is indeed present in the mental lexicon. In fact, words can simultaneously be related according to several semantic and phonological similarities like the ones previously reviewed.

The current scientific literature has a gap in terms of a multi-relational, multi-level modelling of real-world systems such as language. It is within this strongly interdisciplinary context that the current PhD project resides. This work aims to push the boundaries by providing novel insights into real-world multi-relational systems, such as language and cognition, by adopting a multilayered network framework.

Chapter 4

Patterns in the English language: Phonological Networks, Percolation and Assembly Models

Phonological networks (PN) represent sound similarities among words (Vitevitch, 2008). Empirical evidence suggested how words with higher degree and local clustering were more prone to confusability, thus highlighting a relationship between network topology and word processing (Chan & Vitevitch, 2010). However, even networks of random words were found to display patterns similar to the empirical PN for English (Gruenenfelder & Pisoni, 2009), thus limiting the interpretability of topological patterns from PNs. Through quantitative null models, this Chapter[1] indicates which PN features can indeed provide insights on real constraints affecting language and cognition and which are rather artefacts due to the PN's construction procedure.

Abstract

In this Chapter we provide a quantitative framework for the study of phonological networks (PNs) for the English language by carrying out principled comparisons to null models, either based on site percolation, randomization techniques, or network growth models. In contrast to previous work, we mainly focus on null models that reproduce lower order characteristics of the empirical data. We find that artificial networks matching connectivity properties of the English PN are exceedingly rare: this leads to the hypothesis that the word repertoire might have been assembled over time by preferentially introducing new words which are small modifications of old words. Our null models are able to explain the 'power-law-like' part of the degree distributions and generally retrieve qualitative features of the PN such as high clustering, high assortativity coefficient and small-world characteristics. However, the detailed comparison to expectations from null models also points out significant differences, suggesting the presence of additional constraints in word assembly. Key constraints we identify are the avoidance of large degrees, the avoidance of triadic closure and the avoidance of large non-percolating clusters.

---

[1] This Chapter follows the layout of the "Three-paper" PhD thesis in that it is an original, published research manuscript published as: *Stella, M., & Brede, M. (2015). Patterns in the English language: phonological networks, percolation and assembly models. Journal of Statistical Mechanics: Theory and Experiment, 2015(5), P05006.* In order to better present the content, the Chapter/paper is briefly introduced by a short summary. The Chapter follows the same structure of the original paper, in adherence to the guidelines of the "Three-paper" PhD thesis format. For a better presentation within the thesis, the words "paper" and "manuscript" were changed to Chapter.

4.1 Introduction

By constructing *phonological networks* Vitevitch (2008) also applied the network paradigm to modelling phonological patterns in English. In this construction nodes represent phonological transcriptions of words and edges indicate phonological similarity based on a similarity metric established in the field (cf., the phonological neighbourhood density (Fay & Cutler, 1977; Luce & Pisoni, 1998; Sadat et al., 2014)). The main motivation for a complexity approach to modelling the phonological structure of human language via network tools is that traditional psycholinguistic research has focused on local scale analyses to identify the role played by given lexical characteristics (e.g. word frequency, age of acquisition, word length) in determining the accuracy and speed of retrieval of a given word from the mental lexicon (Luce & Pisoni, 1998; Vitevitch, 1997). Although this approach has been valuable, a globally detailed understanding of structural patterns in the mental lexicon is still an open research question (Vitevitch et al., 2014) – a prime motivation for this study. For this purpose, the network approach is a suitable choice since it provides us with an established set of measures and tools to quantify local and global structural patterns.

Vitevitch's first analysis (Vitevitch, 2008) found the phonological network for $20,000$ English words to be disconnected, comprised of a giant component of almost 10^4 words, a variety of smaller-sized components (termed "linguistic islands"), and a very large number of isolated nodes (termed "hermit words"). Furthermore, the giant component exhibits the small-word property combined with high cliquishness, a rather high level of assortative mixing by degree, and a degree distribution that has been described as a power-law with cut-off (Vitevitch et al., 2012; Chan & Vitevitch, 2010; Arbesman et al., 2010b). These results have been confirmed for phonological networks constructed for various other languages, such as Spanish, Mandarin, Hawaiian, and Basque (Arbesman et al., 2010a,b). Building on these insights, Siew (2013) showed that the giant component of the English PN exhibits also a rich community structure, in which large communities are preferentially composed of short, frequent and highly connected words with low age of acquisition ratings. These findings strongly suggest that larger communities may actually be the first to form during the assembly of the mental lexicon, and therefore they may be essential in determining the final structure of the PN. This hypothesis is supported by the fact that phonological neighbours play a role in predicting the order of acquisition of nouns (Hills et al., 2010), and by the empirical evidence that late talkers tend to acquire semantically novel words relative to known words in a way that significantly alters the small-world property of the English PN of normal speakers (Beckage et al., 2011).

Analysing phonological networks it is important to realise that nodes (i.e. words) correspond to sequences of symbols (i.e. phonemes). Then the set of all possible combinations of symbols (together with the phonetic similarity metric) defines a space, of which the actual word repertoire is a subset. In the language of percolation (Stauffer & Aharony, 1991) one might speak of occupied and empty nodes, corresponding to words that are actually present in language and hypothetically possible but not realised words. As is the case for networks in more conventional Euclidean spaces (Newman, 2003) the topology of the underlying space also constrains the organisation of phonetic networks. To some extent this has been realised by Grunenfelder and Pisoni (Gruenenfelder & Pisoni, 2009) who carried out percolation style experiments similar to those of Mandelbrot (1953) to generate phonological pseudolexica, but restricted the study to very short words composed of between two and five phonemes. Corresponding phonological networks were found to retrieve some qualitative characteristics of the English PN (Gruenenfelder & Pisoni, 2009) such as high clustering and strong assortative mixing by degree and the authors suggested that peculiarities of this network, as stated by (Vitevitch, 2008), might be an artefact of the construction method. Whilst it is certainly true that the topology of the underlying space biases characteristics of the PN in the ways described by Grunenfelder and Pisoni, lack of quantitative agreement and the use of word length distributions that ignore longer (less connected) words make final conclusions difficult. Further, comparisons of higher order network statistics (as clustering or assortativity) are not necessarily compelling when lower order characteristics (such as the number of links or sizes of components or degree sequences) differ markedly. In contrast to Gruenenfelder & Pisoni (2009), a study across different languages Arbesman et al. (2010b) reiterated the original point of Vitevitch. As a result of these contradicting findings the main point: “Which characteristics of PNs are archetypical for organisations of words in language and which are mere artefacts of the construction method” remains unresolved.

In this chapter we develop a series of null models to carry out a principled analysis of the phonological network for English. Since our study of the English PN is based on a different database than previous work, we start by briefly reviewing some network properties. We then proceed by comparing the English PN to networks obtained from randomised sets of words. This naturally leads us to consider various types of percolation-style experiments that increasingly respect phonetic constraints. Comparisons to the English PN reveal significant differences in link counts and component distributions between lexicons of real words and pseudolexica. These differences hint at the presence of constraints on clustering and maximum degree in word assembly while pointing out that the power-law like part of the degree distributions observed in Arbesman et al. (2010a) appears as a natural consequence of the embedding space. We then refine these insights by using Monte Carlo Markov Chain (MCMC) like techniques to generate ensembles of words whose network representations have the same link counts as the original data set. The analysis reveals that differences in the sizes of giant components are particularly significant, suggesting the possibility of word assembly mechanisms that proceed by generating new words through small modifications of already existing words.

The question how likely it is to assemble pseudolexica that match link counts and component sizes of the English lexicon arises naturally. We next address this question by introducing various types of attachment models. Extending these models allows to construct ensembles of networks which match essential lower order statistics of the English PN. Hence a quantitative assessment of peculiarities of phonetic word organisation through network analysis becomes possible. The chapter concludes with a discussion of these results in the light of constraints that might have shaped the assembly of the human mental lexicon.

4.2 Empirical network construction and analysis

The construction of the PN for English adopted in this chapter is based on roughly $30,000$ English words using the database from Wolfram Research, a curated repository mainly based on Princeton University Cognitive Science Laboratory "WordNet 3.0." (Miller, 1995) and on

Oxford University Computing Service, British National Corpus, version 3 (Leech, 1992). Phonetic transcriptions in this database are given using the International Phonetic Alphabet (IPA). Before constructing the networks we remove any *supra-segmental* feature such as stress marks or accents and also remove all homophones, i.e. words with identical phonological transcriptions. Network construction then proceeds by associating the remaining words with nodes and connecting them whenever the respective words have edit distance one. It is worthwhile pointing out that the use of edit distance one to define connections is a to some extent arbitrary choice. Other choices are possible, but in the present study we follow earlier work (Luce & Pisoni, 1998; Vitevitch, 1997; Chan & Vitevitch, 2010) which has related this choice to other measures established in psycholinguistics.

Some network statistics for the resulting network are summarised in table 4.2. Comparison with the networks (Vitevitch, 2008; Gruenenfelder & Pisoni, 2009; Siew, 2013; Vitevitch et al., 2012) based on the smaller 20.000 words Hoosier Mental Lexicon (HML) (Aitchison, 2012) gives good quantitative agreement. For instance, in our database the giant component comprises 33% of the nodes (34% for the HML), the clustering coefficient CC is 0.21 (compared to 0.22) and the assortativity coefficient a is 0.70 (compared to 0.67 for the HML) (Newman, 2010). As expected mean geodesic path lengths are larger for our larger lexicon, i.e. $d = 7.71$ (whereas $d = 6.08$ for the HML). As already observed in previous works (Vitevitch, 2008; Gruenenfelder & Pisoni, 2009), the giant component of the PN for English has the small-world property, i.e. when compared to similar size random graphs, it exhibits a higher clustering coefficient and similarly low average shortest path length. Furthermore, on average, each linguistic island contains 2.49 ± 0.04 words, in agreement with the ~ 2.52 estimate from Gruenenfelder & Pisoni (2009). Also, the degree distribution of the giant component follows a power-law like behaviour with a cut-off, similar to the analysis of Arbesman et al. (2010b) for English and several other languages.

Altogether, our larger dataset is able to closely reproduce features of the English PN as constructed from smaller databases, and we find similar macro (degree distribution, assortative mixing by degree, average clustering coefficient, average path length) and micro (node degree and local clustering coefficient) characteristics as previous analyses.

4.3 Randomisation experiments

In order to investigate the features of the empirical English phonological network we next present a series of null models. The agenda for those is as follows. We first assume that system sizes are fixed and obtain reference models via randomisation of the existing dataset. The first step leads us to percolation-based approaches presented in section 4.3.1. These experiments could be seen as the generation of artificial word repositories generated from the original one through a shuffling procedure that preserves the word count. The approach can be refined by using MCMC like randomisation procedures (Snijders, 2002) which allow the generation of randomised ensembles that preserve additional constraints. In section 4.3.2 we apply this idea to generate ensembles of artificial words whose corresponding networks have exactly the same number of connections as the original data.

The randomisation approach is reminiscent of Exponential Random Graph Modelling (ERGM), which aims to identify the Hamiltonian that defines a network ensemble a given empirical network might be an example of (Newman, 2010). ERGM is particularly useful when one wants to create network models matching the features of empirical networks as closely as possible without specifying the details of the underlying network formation process (Fronczak, 2014). In the last decade ERGM has been applied widely in social network analysis (Robins et al., 2007), particularly for the analysis of reciprocity and transitivity in friendship and contagion networks. This method typically relies on the use of computationally demanding MCMC simulations combined with regression techniques. Applying ERGM models to the analysis of phonological networks could give interesting additional insights, but (i) computational costs and (ii) open theoretical questions on how to include constraints from spatial embedding (which has only been considered in Euclidean space quite recently by Daraganova et al. (2012)) and (iii) phoneme statistics make the randomisation approach the preferred choice for this study.

Following on from the randomisation analysis in Sec. 4.4 will consider if the English PN can be modelled via processes of word assembly over time. In the remainder of the chapter we will refer to a table that summarises network statistics for all reference models we present, cf. Table 4.2. To avoid confusion we also provide a table that gives an overview over all models, cf. Table 4.1.

4.3.1 Percolation

Let us introduce the set of all possible phoneme sequences $S = \cup_l L_l$, where $L_l = \{w_l\}$ is the set of all possible words $w_l = \{s_0, ..., s_l\}$, $s_i \in \mathcal{P}$ of length l and the set $\mathcal{P}$ is the set of all possible phonemes of a given language. For example, for our dataset of English words, we have $|\mathcal{P}| = 36$ phonemes and words up to length $l = 21$. Within a set L_l distances between words can be measured by the conventional Hamming distance, between words in different sets distances can be determined by the minimum number of phoneme additions, deletions are substitutions required to transform one word into another, the so-called edit distance $d_E(\cdot, \cdot)$, a generalisation of the Hamming distance (Ristad & Yianilos, 1998). By identifying nodes with possible phoneme sequences, i.e. the set S, and connecting nodes whenever their associated words have edit distance one, a substrate graph is defined, of which phonological networks are a subset. For a better understanding of this substrate graph it is useful to visualise it as a stacked set of *layers* of graphs composed of words of the same length, cf. Figure 4.1. In this way one can naturally distinguish between *intra-* and *inter-layer* connections which encapsulate additional information about phoneme organisation, on top of more conventional network measures.

The organisation of substrate graphs in such layers leads to an evident conclusion. Since the number of all possible words of given length grows exponentially $|L_l| = |\mathcal{P}|^l$, whereas the number of actual words of given length grows markedly less rapidly occupation densities of layers decrease exponentially with increasing word length l. Also, coordination numbers of nodes in L_l are given by $\kappa_l = (|\mathcal{P}| - 1)l$. Using the Bethe approximation as in (Stauffer & Aharony, 1991) as a rough estimate of percolation thresholds of individual layers, one thus expects to find giant components only in layers made up of shorter words, for which the percolation density threshold is exceeded. This points to the importance of the word length distribution $H(l)$ in determining properties of phoneme networks. Word length distributions with a bias for shorter words will naturally induce larger component sizes than word length distributions that account for relatively more long words. Hence, comparisons between pseudolexica and real datasets are only sensible if the

Notation	Type	Brief Explanation
type 0	Percolation Experiment	Percolation-like experiment with uniform random sampling of phonemes, which generates a pseudolexicon with the same word length distribution as the real data.
type 1	Percolation Experiment	The same as in type 0 but with additional constraints on the phoneme frequencies.
type 2	Percolation Experiment	The same as in type 1 but with additional constraints on the phoneme-phoneme correlations.
MCMC	Monte Carlo Markov Chain	Ensemble of randomized networks which are obtained from the English PN via word shuffling that enforces the empirical word length distribution and realistic phoneme statistics as in type 2 and additionally enforces a link constraint. The method generates an ensemble of networks with the same number of connections as the English PN.
grow gc (r)	Growing Network	The network is grown over time. At each timestep a new word is generated at random (as in type 2 experiments). If it does not receive any connection, then it is discarded with probability f. Words with degree larger than k_{max} are additionally suppressed.
grow gc (o)	Growing Network	The same as in grow gc (r) but here shorter words are generated first.
$(W_C,\, m_C),\, f$	Core-Periphery Models	The same as in grow gc (r) experiments but with an additional chance of rejection for words shorter than W_C with less than m_C links. These experiments enforce the generation of a network core, which is composed of preferentially short and highly connected words.
$(W_C, m_C),\, \epsilon,\, f$	Core-Periphery Models	The same as in $(W_C,\, m_C),\, f$ experiments but with an additional rejection probability ϵ of discarding words connecting to the giant component.

Table 4.1: Summary of all the null models implemented in this chapter, as reported in Table 4.2.

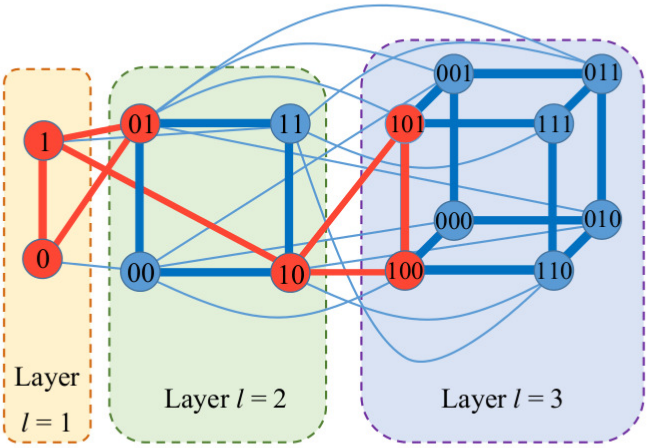


Figure 4.1: Visualisation of a substrate graph with a binary phonetic alphabet $\mathcal{P} = \{0, 1\}$. In this case, each layer is represented as a hypercube. Red nodes represent the actual words in a fictional binary language. Red links connect phonologically similar actual words. The other connections between layers have been omitted for a better visualisation.

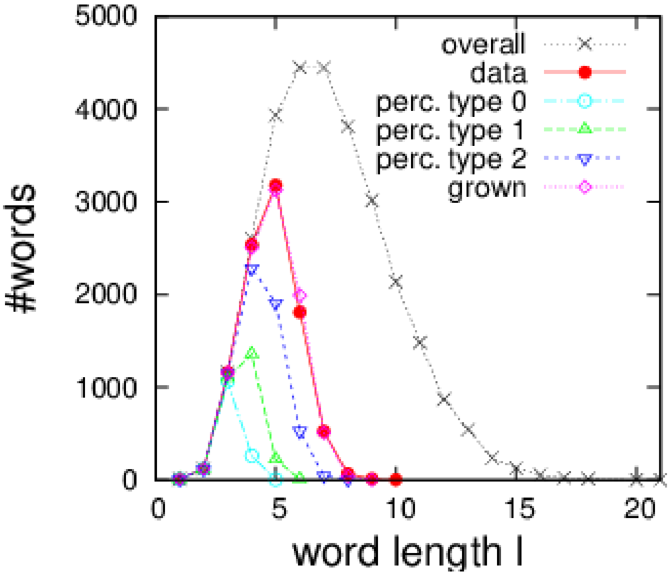


Figure 4.2: Word length distribution for English and distribution of word lengths in the giant component. Data for the word length distribution in the giant component are compared for the dataset of phonetic transcriptions of English words, type 0, 1, and 2 percolation experiments and the growth experiments introduced in Sec. 4.4 for $k_{\max} = 25$ and $f = 0.75$. Data points represent averages over at least 10 configurations.

Experiment	L	L_0	lr	gc	$k_{\max}$
English	38342	34896	2.46	9412	44
MCMC	38342	37845(10)	2.21(0.01)	7260(20)	88(1)
type 0	2840(50)	2490(50)	3.39(0.01)	1440(40)	18.4(0.8)
type 1	8370(30)	8210(30)	2.59(0.01)	2808(20)	46.8(0.7)
type 2	20420(30)	19820(30)	1.95(0.01)	6020(20)	52.7(0.8)
grow gc (r)	28890(20)	27740(20)	1.68(0.01)	9390(20)	43.1(0.7)
grow gc (o)	27260(20)	27110(20)	1.64(0.01)	9380(20)	42.2(0.9)
$(5, 4.2), f = 0.76$	35580(40)	34900(40)	2.17(0.01)	9380(40)	43.2(0.9)
$(5, 4.4)$, $\epsilon = 0.82$, $f = 0.962$	38320(30)	34860(30)	2.37(0.01)	9390(30)	43.8(0.4)

Experiment	CC	a	d	$d_{\max}$
English	0.207	0.707	7.71	33
MCMC	0.304(0.001)	0.41(0.01)	5.47(0.01)	20.3(0.4)
type 0	0.16(0.01)	0.59(0.01)	7.63(0.01)	22.5(0.5)
type 1	0.220(0.001)	0.46(0.01)	5.45(0.01)	16.1(0.3)
type 2	0.248(0.001)	0.45(0.01)	5.95(0.01)	18.9(0.3)
grow gc (r)	0.259(0.001)	0.46(0.01)	6.53(0.01)	18.8(0.3)
grow gc (o)	0.254(0.001)	0.47(0.01)	6.59(0.01)	19.0(0.4)
$(5, 4.2), f = 0.76$	0.258(0.001)	0.48(0.01)	6.35(0.01)	23.8(0.4)
$(5, 4.4)$, $\epsilon = 0.82$, $f = 0.962$	0.238(0.001)	0.55(0.01)	7.38(0.01)	36.8(0.4)

Table 4.2: Overview of characteristics calculated for the English PN, networks constructed from the various types of percolation experiments, link preserving randomization (MCMC), and networks grown by rejection sampling. L is the number of links, L_0 the number of links in the giant component, lr is the ratio of inter-layer over intra-layer links, gc is the giant component size, k_{max} is the maximum node degree, CC is the average local clustering coefficient, a is the assortativity coefficient, d is the mean geodesic distance and d_{max} is the network diameter. The networks grown by rejection sampling are those with word attachment ordered by word length (o) or at random (r) (cf. Sec. 4.4), and best fit core-periphery models that either match the size of the giant component and the number of links in it or additionally match the total number of links (cf. Sec. 4.4.2). A maximum degree constraint with $k_{\max} = 20$ and $\nu = 0.1$ was used. Averages over randomized ensembles are carried out by averaging over at least 10 ensemble members. Error bars are reported in a compact way: 16.1(0.3) is equivalent to 16.1 ± 0.3.

same word length distribution is used. Figure 4.2 gives the word length distribution for our dataset of phonetic transcriptions of English words.

A first reference case (which we refer to as type 0) to explore the organisation of words in real word repositories is given by models in which words are occupied at random. An appropriate model that respects the word length distribution might consider the union of subspaces $P_l \subset L_l$ constructed such that $H(l)$ unique words are chosen uniformly at random from L_l to make up P_l. The pseudolexicon constructed this way is associated with a phonetic network which is analysed in Table 4.2. In principle, the organisation of artificial words in S is similar to English. One finds a giant component, lexical islands, and an overwhelming majority of hermit words. The presence of the giant component for words of length $l \leq 4$ is consistent with the Bethe estimates for the percolation thresholds, in fact layers are clearly supercritical for words up to length three and clearly subcritical for larger l. Hence, every artificial PN assembled at random in this way will have a "core" of densely occupied shorter layers that enable the formation of a giant component, cf. also the word length distributions restricted to words within the giant component shown in Figure 4.2. However, the contrast in quantitative comparisons to the real PN is striking: Our type 0 percolation experiments result in a by far smaller giant component and more than a factor of 10 less links than observed in the data.

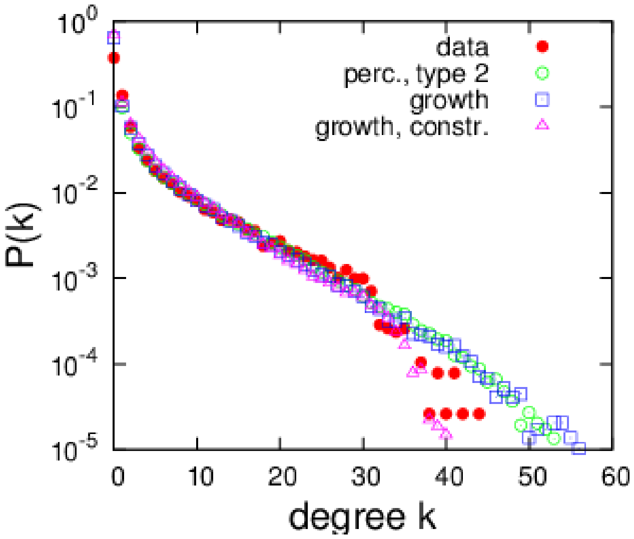


Figure 4.3: Degree distributions of the real PN and of artificial PNs, see legend. Data points for simulation experiments are averaged over at least 10 configurations.

Selecting words uniformly at random in the layers L_l assumes building phoneme sequences by sampling phonemes uniformly at random from the alphabet $\mathcal{P}$. In real language, however, phoneme usage is not uniform, but highly skewed (Siew, 2013). The above percolation experiments can be modified to account for such skewed phoneme frequencies. Instead of sampling words uniformly at random from layers, we construct $H(l)$ unique phoneme sequences of length l for each layer l by sampling phonemes from the phoneme frequency distribution determined for English. The resulting type 1 networks are analysed in table 4.2, and again agreement with the general structure of the English PN can be stated. Quantitative comparisons yield slightly larger giant components and link counts in the artificial networks of type 1 compared to type 0.

Real language incorporates correlations between phonemes at other levels (Gruenenfelder & Pisoni, 2009). Important among them are, e.g., consonant-vowel co-occurrence patterns in word production. To include such correlations, we develop a third type of percolation experiments (labelled type 2) in which phonemes are sampled from the real phoneme frequency distribution and empirically determined phoneme-phoneme correlations are respected when constructing artificial words, again in such a way that the resulting ensembles follow exactly the same word length distribution as the empirical data. Comparisons of artificial PNs pertaining to these ensembles are shown in table 4.2, again noting a closer match in giant component size and link counts with the English PN. Another important observation is that (as noted in the experiments of Gruenenfelder & Pisoni (2009)) all of the artificial networks are marked by high clustering, high assortativity coefficient and small average (chemical) distances, but quantitative comparisons do not yield a good match within error bounds.

It is, however, interesting to note that the low degree region ($k < 30$) of the degree distributions of all percolation experiments gives a very good match with the empirical data, in fact with the region that has previously been used to estimate a power-law dependence (Arbesman et al., 2010b). This is also the case with type 2 percolation, but even though these artificial PNs have significantly less links than the English PN already a heavier tail than observed for the real data is found. These observations support two conclusions: first, the power-law like region of the degree distribution results from the structure of the constraining space and from the decreasing word occupation density with word length so that no recourse to additional explanations is required (i.e. preferential attachment, as suggested in Vitevitch (2008)). Second, the English PN is characterised by a maximum degree cut-off such that words with large numbers of neighbours are suppressed in comparison to random sampling. As random sampling does not exhibit a similar cut-off, this cut-off does not result from constraints in the underlying space as speculated in Gruenenfelder & Pisoni (2009), but must be caused by an additional constraining influence in word repertoire formation.

One obvious way to continue modelling artificial word repertoires is by including higher order correlations, for instance by fitting higher order Markov processes to the empirical data. Such an approach will naturally lead to a better fit of network metrics, but offers relatively little explanatory power. Instead, we note that all percolation experiments yield giant components and link counts much lower than measured for the English PN – a consequence of the aforementioned higher order correlations in phonetic transcriptions of words. We next attempt to disentangle these observations of link count and component size. Are larger than expected component sizes just a consequence of larger link counts or do they represent additional peculiarities of the English PN?

4.3.2 Randomization and Markov Chain Monte Carlo methods

A standard way to explore peculiarities of a given network is via randomisation procedures that perform Monte Carlo steps designed to destroy correlations in the given network's architecture while enforcing constraints (Newman, 2010; Snijders, 2002). The method allows for the construction of ensembles of networks which preserve key characteristics of a given network, but are "random" in every other respect. Randomised networks constructed from such Markov Chain methods provide an important null model for network analysis; comparing a given network to such suitably randomised ensembles allows the identification of peculiarities of network structure. For this purpose one best proceeds by first randomising network structures subject to the lowest level constraint, subsequently adding constraints in the course of the analysis then allows increasingly more sophisticated insights into higher order correlations peculiar to the given network data, see, e.g., Milo et al. (2002) for applications of randomisation techniques in the context of motif detection.

In this section we explore Monte Carlo schemes that preserve lowest level network statistics while also enforcing spatial constraints, phoneme statistics, and word length distributions typical to the English PN. Proceeding systematically in this vein, one would first address the question of exploring the architecture of networks that have the same number of nodes as the given network, but are otherwise random (except for the above mentioned phoneme statistics, word lengths and spatial constraints). To construct such an ensemble, one could proceed as follows: (i) randomly select a word w in the current repertoire and remove it. (ii) Generate an artificial word $w\prime$ of the same length as the word w in such a way that phonemes are drawn from the phoneme frequency distribution corresponding to the original repertoire and such that phoneme-phoneme correlations as observed in the original repertoire are on average conserved. (iii) Accept the word w if it does not coincide with another word already in the repertoire. If $w\prime$ is not accepted,

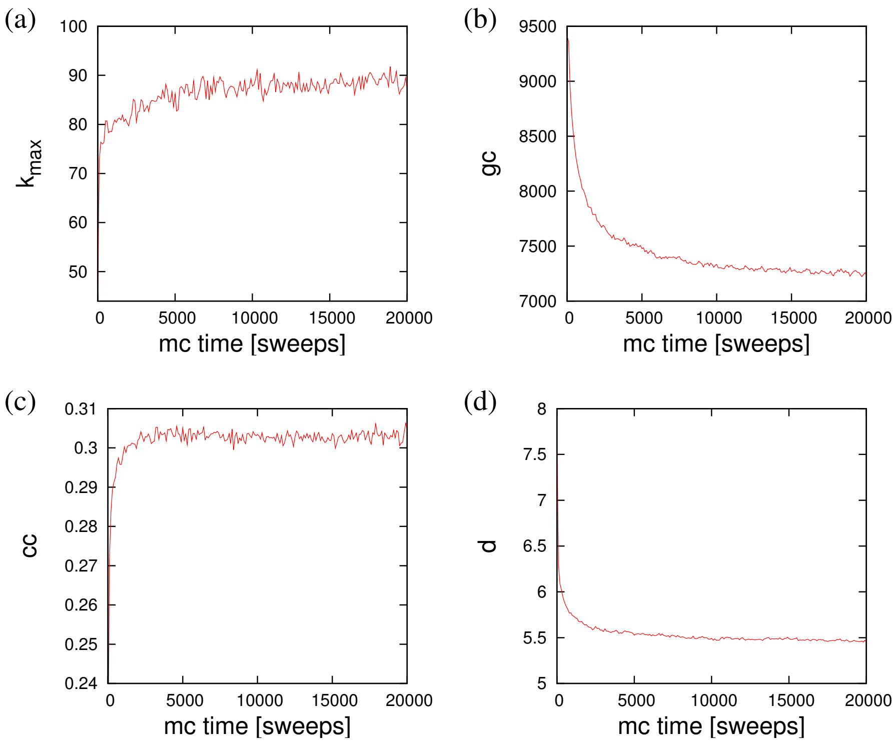


Figure 4.4: Change in (a) maximum degree, (b) size of the giant component, (c) clustering coefficient, and (d) average shortest path lengths with the number of randomization steps (measured in multiples of the network's size). The initial network is the English PN and then randomization steps that preserve the network's link density are performed. The curves correspond to averages over ten independent runs.

w is inserted again and we proceed with step (i) selecting another word w at random. It is straightforward to see that the suggested procedure will asymptotically result in ensembles as the ones discussed as "type 2 percolation" in the previous section, i.e. essentially random allocations of occupied and empty nodes in word space.

Increasing the order of preserved network statistics, we next analyse ensembles of networks that have the same number of nodes and links and phoneme and word lengths statistics as the given English PN, but are otherwise random. For this purpose we proceed as suggested above, but modify the acceptance step (iii) such that suggested words $w\prime$ are only accepted if the new suggested word does not already exist in the repertoire and if it has the same degree as the previous word w.

Figure 4.4 shows the change in some network characteristics versus the number of randomisation steps performed, measured in multiples of system size. The initial points in each panel reflect the data measured for the English PN, the following evolution shows how certain correlations peculiar to the English PN are gradually destroyed as the network is randomised. The simulation data illustrate that the method is computationally quite demanding, only after around 10,000 randomisation sweeps an equilibrium is reached. Next, characteristics of the ensemble of constrained networks with the same number of nodes and links as the English PN can be measured and averaged over the following 10,000 Monte Carlo updates.

Averages over some network measures that characterise the randomised ensemble are given in Table 4.2. Results essentially confirm earlier observations from percolation experiments: The English PN is found to have a substantially smaller tail of highly connected words, is less cliquish and has stronger assortative mixing by degree. Further, we observe that the English PN has a significantly larger and less densely connected giant component than expected. Larger than expected average path lengths and diameters appear as a consequence of the larger size of the giant component.

While confirming earlier results, comparisons to the ensemble of random PNs with the same link density also allows for additional conclusions. We find that the reduced size of the giant component and enhanced clustering and assortative mixing coefficients observed in the percolation experiments are not artefacts of reduced link counts, but hint at additional features characteristic of the English PN itself.

Our first steps of a randomisation analysis can be extended by fixing additional constraints which would allow deeper analysis of factors that contribute to the architecture of the English PN. We have undertaken some preliminary experiments in this direction, but computational costs become increasingly onerous as the rejection rate increases and the calculation of global network statistics at each Monte Carlo step can become very demanding. Instead of pursuing randomisation ideas subject to additional constraints, a comparison to ensembles of grown networks is a more flexible approach that sheds light on possible non-equilibrium features of repertoire formation. In particular, we can use it to explore what constraints are required in models of network growth to match important statistics of the English PN.

What we have observed so far is that without enforcing additional constraints randomisation-based null models can neither satisfactorily reproduce the size of the giant component nor core-periphery features of the empirical data set. These larger than expected connectivity properties hint at a process of repertoire formation in which words are assembled over time in such a way that preferentially such new words that connect to older words are added. Alternatively, one can interpret this as word assembly in a way that new words are preferentially formed by slightly modifying already existing word forms, as in word derivation (Aitchison, 2012). How likely is it that such a process could form PNs that are in quantitative agreement with the real data? Can the real data help to estimate constraints in such a process of word assembly? These are the questions we are going to explore in the remainder of the chapter.

4.4 Growing repertoires by rejection sampling

In this section we explore processes of word repertoire formation over time, i.e. growth models for the phonetic networks corresponding to the artificial repertoires. We start with the observation of the differences in the size of the giant components reported in the previous section. This motivates a first type of growth process, described in subsection 4.4.1, which aims at generating artificial networks with similar size of the giant component as the English PN. These models are then refined in subsection 4.4.2, in which we develop models that match further connectivity properties of the real data.

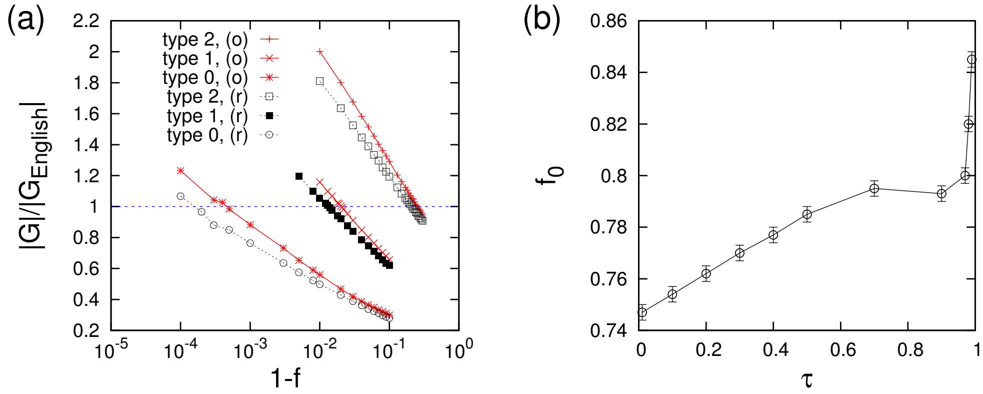


Figure 4.5: (a) Dependence of the relative size of the giant component of artificially grown word ensembles on the rejection probability f for both word length ordered and random attachment for artificially constructed words according to the percolation 0,1, or 2 rules. Data for degree constrained ensembles are not shown as data points are virtually identical. Data points represent averages over at least 10 networks. (b) Dependence of the probability f that allows to reach the same size of the giant component as in the real English PN on the order of attachment of words as measured by the parameter τ, cf. text.

Before proceeding it is worth pointing out that all our models cannot capture the real evolution of the English language over time. The main reason for this is that the framework lacks any realistic assumption about the role played by the influence of external languages and about the rate of language change. Nonetheless, the following section will investigate if models of growing networks can reproduce features of the English PN, and will explore what constraints in word assembly are likely to have played a role in repertoire formation.

4.4.1 Growth models

Consider a process of word assembly in which new words are added by suggesting new randomly sampled words of lengths drawn from a given word length distribution $H(l)$ and rejecting them with some probability

$$r = f + (1 - f)p_k \tag{4.1}$$

according to an acceptance criterion. The severity of the application of this criterion is tuneable by a parameter f and the probability p_k implements additional degree constraints explained below. New artificial candidate words can be generated according to the type 0, 1, or 2 percolation models detailed in the previous section, i.e. by uniform sampling from the set of all phonemes, sampling from the real phoneme distribution or additionally respecting phoneme-phoneme correlations.

To generate artificial ensembles that reach the same connectivity properties as the English PN we accept new words with probability r if they connect to at least one old word and reject them otherwise. In both cases a new artificial word is suggested until exactly as many words of length l have been accepted as present in the word length distribution $H(l)$ of the English repertoire. As the previous section has demonstrated the presence of an additional degree constraint in the English PN the second factor in Eq. (4.1) serves to suppress nodes of large degree. For a suggested word w we set

$$p_k = \exp(\nu(k_{\max} - k_w)) \prod_{n \in \mathcal{N}_w} \exp(\nu(k_{\max} - k_n)), \tag{4.2}$$

where $\mathcal{N}_w$ is the set of all neighbours of w, k_w the degree of w, and $k_{\max}$ and ν parametrise the cut-off behaviour for large degrees. Note that the major difference to established uniform attachment models (Callaway et al., 2001) is the constraint of the underlying space. Different from Callaway et al. (2001) new nodes in our model can have different degrees, depending on their location in the space S. The degree constraint expressed in Eq. (4.2) is essentially implemented in such a way that new nodes are likely to be rejected if any of their neighbours would reach too large a degree by the addition of the new node. Using the above framework additional constraints can easily be implemented as additional terms to Eq. (4.1), some of which we will discuss in more detail later.

By showing the dependence of average sizes of giant components of grown ensembles on the acceptance probability Figure 4.5(a) summarises some first simulation experiments. Stronger acceptance constraints (i.e. larger f) allows to construct networks with larger giant components and by estimating the crossing of simulation data with the line $|G|/|G_{\text{English}}| = 1$ one can find values for the rejection probability f such that the resultant ensembles will match the size of the giant component of the English PN. Comparing this crossing value f_0 for different growth procedures allows estimates about the relative likelihood of reproducing realistic features of the English PN using these procedures. For instance, with $f_0^{(\text{type 0})} \approx 1 - 10^{-4}$ generating realistic ensembles by uniform attachment proves excruciatingly difficult and even ignoring phoneme-phoneme correlations by only sampling from the real phoneme frequency distribution one still has $f_0^{(\text{type 1})} \approx 1 - 10^{-2}$, while inserting phoneme correlations yields $f_0^{(\text{type 2})} \approx 0.75$. As one would naturally expect, since $f_0^{(\text{type 0})} > f_0^{(\text{type 1})} > f_0^{(\text{type 2})}$ we observe that growing ensembles by suggesting new words that include more realistic phoneme statistics provides a more likely explanation of the real data. For this reason, and as we aim to construct word repertoires that respect lower order correlations in the real data we proceed with word generation method type 2 in all experiments presented below.

Words can be attached in different order and the order of attachment will generally influence the structure of the generated network. Panel (a) of figure 4.5 also compares attachment of words in random order (black symbols) and attachment ordered by word length, starting from the shortest words (red symbols). One generally finds $f_0^{(\text{ordered})} < f_0^{(\text{random})}$, i.e. attachment ordered by word length gives a more likely explanation of the data than random attachment. Similar experiments can also be carried out using other attachment criteria (e.g. accepting words if they connect to the giant component and rejecting them with probability f otherwise) or by determining crossing probabilities for other network statistics (e.g. the number of links). We have tested some of these alternatives and found similar qualitative results, but different numerical values of the estimated crossing probabilities f_0. Link attachment was chosen as the most suitable starting point for further exploration below.

In panel (b) of Figure 4.5 the attachment order is explored more systematically. For this purpose we introduce an additional attachment order parameter τ and construct lists of word lengths according to which words are attached in the following way. We start by setting $H' = H$. For a given place in the list, say t, a word will be allocated the smallest l bin of $H'(l)$ which has not been exhausted yet, i.e. for which $H'(l) > 0$. Then, we continue increasing l with probability τ. If a value of l was generated this way for which $H'(l) = 0$ we select the closest smaller value of l for which $H'(l) > 0$. Having thus determined the word length of the word which will be attached at step t we decrease $H'(l)$ by one and repeat the process until $t = N$ has been reached. The parameter τ is thus a measure of the order of attachment of new words. For $\tau = 0$ words are attached ordered exactly by word length, starting with the shortest words. For small τ attachment is generally ordered by word length, but there is a small chance that slightly longer words might sometimes precede shorter words. If $\tau = 1$ words are attached in reverse order, i.e. the longest words first and the shortest ones last. We find that in between for $\tau \approx 0.9$ the attachment order is approximately random.

Systematically varying τ we find that attachment orders that attach shorter words first are generally more likely explanations of the data than other word orders, cf. the monotonic trend in Figure 4.5b. This is compatible with the hypothesis in psycholinguistics that the mental lexicon is built starting preferentially with shorter words, that are also more frequent (Storkel, 2004; Hills et al., 2010; Aitchison, 2012; Siew, 2013). Consequently, in all experiments presented below we will assume almost perfectly ordered attachment of words only allowing for a small amount of disorder $\tau = 0.01$ to ensure for robustness.

Growing a network by attaching nodes over time introduces correlations which can alter the degree distribution in comparison to models that allocate links randomly in a fixed set of nodes. Figure 4.3 also compares the degree distributions of the grown ensembles with the real data. Again, as in the percolation experiments before, we note that the low degree part of the distribution is well matched by the networks associated with grown word ensembles, but growth also induces a heavier tail for large degrees. A good match with the data for the English PN can be obtained if one sets $k_{\max} = 25$ and $\nu = 0.1$, cf. the data for ensembles grown with constraints in Figure 4.3. Ensembles grown with constraints to match the size of the giant component of the English PN also have the same distribution of word lengths in the giant component as the English PN, cf. Figure 4.2.

Further analysis of the networks belonging to the grown ensembles is presented in table 4.2. Most prominently, one notes that all such networks have less links than the English PN, but also other network statistics show quantitative deviations whilst generally confirming qualitative observations about large clustering, high assortativity and the small world character of artificial PNs. One is thus lead to wonder which links are missing in the grown artificial PNs in comparison to the English PN. Figure 4.6 addresses this question by plotting average degree vs word length for the English PN and null models constructed by percolation or growth. As one would expect from our earlier arguments about percolation thresholds and the structure of the underlying space, links are densely concentrated on shorter words, leading to a clear core-periphery structure of the networks. Average nodes corresponding to word lengths larger than 5 or 6 have hardly any network neighbours, whilst nodes belonging to shorter words are very densely connected, having around 20 neighbours on average. Furthermore, a comparison between the null models and the data shows that mostly links within the core are not captured by our modelling yet. In particular, comparing percolation type 2 experiments and ensemble growth one notices that the link attachment constraint mostly adds links for words longer than four phonemes while leaving the average link count for shorter words almost unaltered. This is the case because short words almost always have a neighbour and will thus be preferentially accepted, such that

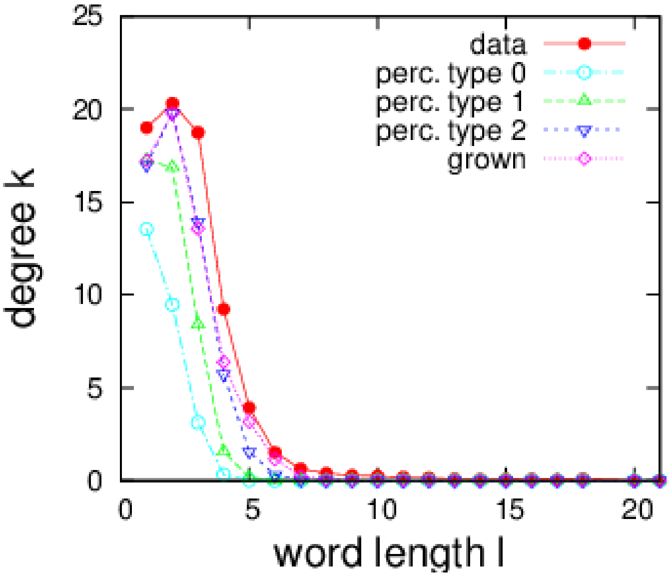


Figure 4.6: Average degree vs word length for the English phonetic network and various null models. For the simulation data averages over at least 10 independent runs have been taken.

the link attachment constraint only becomes effective in adding connections for longer words. Consequently, an improved null model will have to account for more links for short words. In the next section, we propose a family of core-periphery models aiming to generate artificial PNs with giant components and link counts matching the English PN.

4.4.2 Core periphery models

The main purpose of our core periphery (CP) models is to account for missing links in null models for short words relative to the data. Hence, we define these models by the iteration of the following steps: (i) construct an artificial (non duplicated) word according to the type 2 process, (ii) reject this word with probability

$$r = f + (1-f)p_k + (1-f)(1-p_k)C(w,k_w), \tag{4.3}$$

where, as in previous experiments, the factor f gives the tuneable rejection probability, p_k models the maximum degree constraint according to equation (4.1), and the additional factor $C(l_w, k_w)$ models a core-periphery constraint. For a suggested word w of length l_w and degree k_w we set

$$C(w,k_w) = \begin{cases} \delta & \text{if } l_w < W_C \text{ and } k_w < m_C \\ 0 & \text{if } l_w \geq W_C \end{cases}, \tag{4.4}$$

in which W_C models the core size (in terms of word length) and m_C [2] allows to tune the density of links in the core and δ is the strictness of the application of the core criterion for short words. Simply put, we suppress short words that do not have enough links, in order to counter the effect reported in Figure 4.6. Since there are rare configurations in which the core criterion cannot be

---

[2] Fractional values of m_C are interpreted probabilistically, i.e. with prob. $m_C - \lfloor m_C \rfloor$ we set the value to $\lceil m_C \rceil$ and to $\lfloor m_C \rfloor$ otherwise.

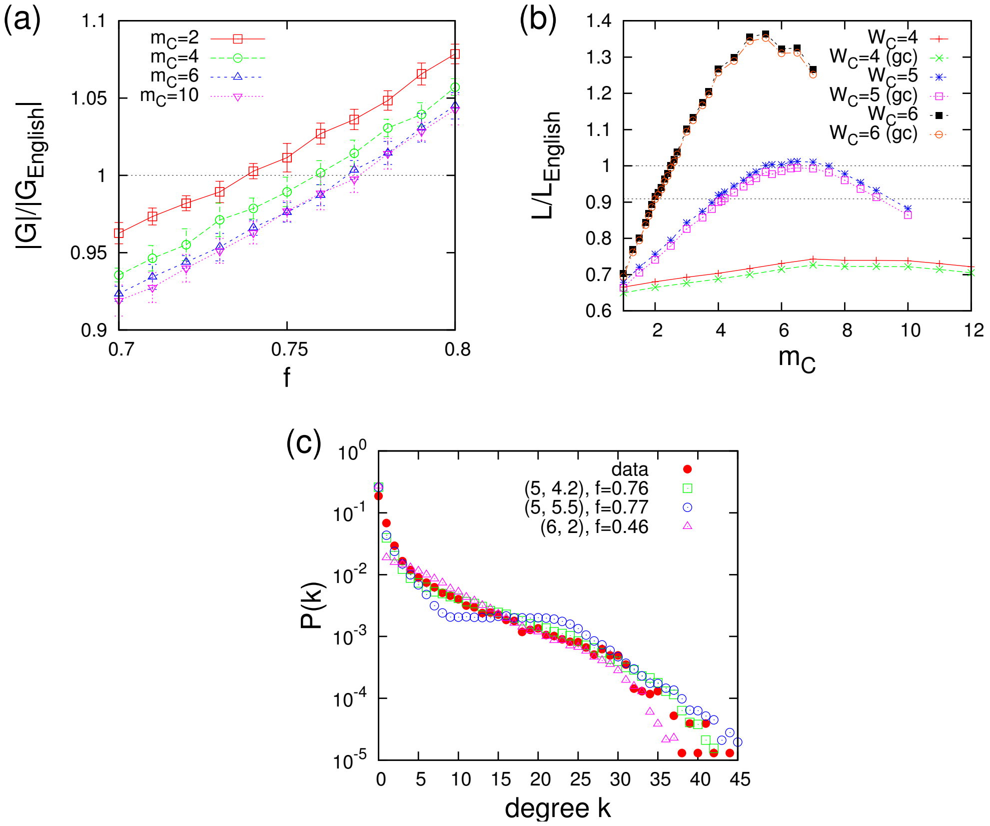


Figure 4.7: (a) Relative size of the giant component vs rejection probability for CP models with $W_C = 5$ and several choices of m_C. (b) Analysis of core periphery (CP) models. We note that for small core sizes $W_C \leq 4$ it is not possible to build networks with the same number of links as the real data, the same is true also for $W_C = 7$ irrespective of m_C. For $W_C = 5$ and $W_C = 6$, there are generally two intersection points. The following CP networks with same size of giant component and same number of links in the giant component as the real network emerge: $(W_C, m_C) = (5, 4.2), (5, 9)$ and $(6, 2)$. (c) Degree distributions of candidate CP networks compared to empirical data. If the core size is too large, deviations for low degrees occur (see (6,2)). Likewise, if m_C is too large relative to the core, nodes with $k < m_C$ are underestimated compared to the data. This only leaves the low m_C intersection point for core $W_C = 5$ networks as reasonable candidate models which replicate the size of the giant component, number of links, and the degree distribution of the English PN. Results represent averages over at least 10 configurations.

exactly met, we set $\delta = 0.99$ in all following experiments. The modified procedure allows us to tune the number of links in the core. Additionally, the parameter W_C allows to vary the size of the core, whereas m_C allows to tune the link density within the core. Comparison to the plot of average degrees versus word length in Figure 4.6 suggests that $4 \leq W_C \leq 6$ for English.

To explore which core-periphery models give a good description of the English PN we proceed as before. For each combination (W_C, m_C) a rejection probability f can be determined such that the respective CP network matches the size of the giant component of the English PN, cf. panel (a) of Figure 4.7 in which data for $W_C = 5$ and various values of m_C are analysed. Once this probability f has been determined link counts for links within and outside of the giant component can be compared for various core densities m_C, see panel (b) of Figure 4.7 which compares link count vs m_C dependencies for cores of various sizes. One notes that generally almost all links belong to the giant component. For better comparisons of properties of the giant component we determine intersection points of link counts in the giant component. As one would expect, link numbers increase at first when core connectivity m_C is increased. When large values of m_C are chosen the requirement for new nodes within the core to be accepted becomes very demanding and since the core requirement was implemented probabilistically increasingly more nodes are accepted without fulfilling it. This explains a reversal in trend, such that for each core size two intersection points at which CP networks of a certain core size match the number of links within the giant component of the English PN can be identified. This, however, is only the case if the core size W_C is large enough. Small cores are composed of too few words to allow for the addition of enough links to reach the required link number for comparison with the English PN, this is for instance the case for $W_C = 4$, cf. data in panel (b) of Figure 4.7. Following this argument four candidate parameter sets for comparison to the English PN are found, i.e. the low and high m_C intersection points for $W_C = 5$ and $W_C = 6$ (Figure 4.7).

When comparing the degree distributions of these networks to the English PN it becomes apparent that only the low m_C intersection point for $W_C = 5$ gives a reasonable match. If core connectivity is chosen too high relative to core size, too many high degree core nodes are generated while low degree nodes are underrepresented. The large degree cut-off then results in a degree distribution with a plateau (open circles in panel (c) of Figure 4.7). Similarly, if the core size is chosen too large (i.e. $W_C = 6$ or larger) not enough nodes of degree one or two are generated to allow for a good comparison to the English PN. Hence, only one ensemble of CP networks is identified which matches the size of the giant component, number of links in the giant component, and gives a good fit of the degree distribution of the English PN. Further analysis for this ensemble, constructed with $(W_C, m_C) = (5, 4.2)$ and $f = 0.76$, is presented in table 4.2. Comparing network statistics with the English PN, there is a large discrepancy between link counts within and outside of the giant component. Like for all percolation type networks also for all the grown network ensembles very few links connect nodes that do not belong to the giant component. Hence, comparisons of properties of the entire network are not yet reasonable, since a very large fraction of all nodes has significantly less connections than in the English PN.

The last observation motivates us to introduce a last set of null models in which the number of links in- and outside of the giant component can be tuned. To define such a variant of CP networks we add another term to Eq. (4.3) which now becomes

$$\begin{aligned} r = f + (1-f)p_k + (1-f)(1-p_k)C(w,k_w) + \\ + (1-f)(1-p_k)(1-C(w,k_w))R(w), \quad (4.5) \end{aligned}$$

where

$$R(w) = \begin{cases} \epsilon & \text{if } w \in G_t \\ 0 & \text{otherwise} \end{cases} \quad (4.6)$$

is an additional term that accounts for the rejection of new nodes if they link to the largest component G_t at iteration t of the assembly process. By tuning the probability ϵ in Eq. (4.6) we can construct ensembles of CP networks with relative fractions of links in the giant component lower than found in the models characterised by Eq. (4.3) which is retrieved for $\epsilon = 0$.

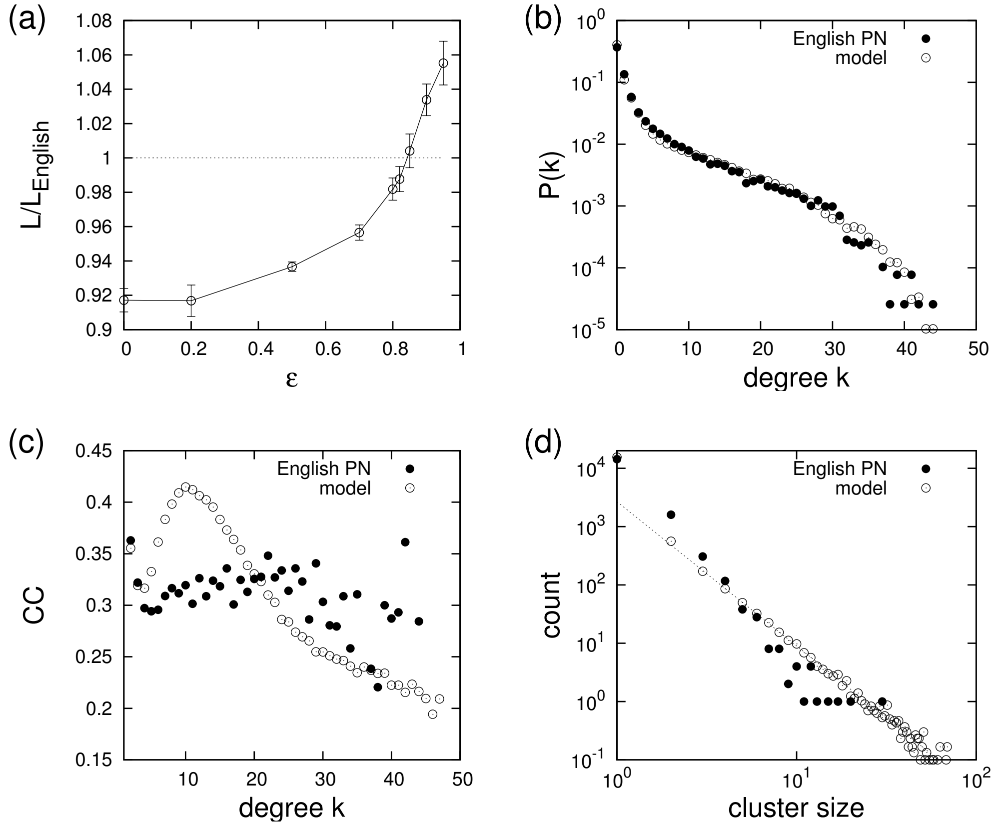


Figure 4.8: (a) Ratio of link counts of CP networks and the English PN vs. the parameter ϵ. The link count inside of and outside of the giant component are met for $\epsilon \approx 0.82$ (and $W_C = 5, m_C = 4.4, f = 0.962$). (b) Comparison of the degree distribution of the above network and the English PN. (c) Local clustering coefficient vs degree for the English PN and the CP networks. (d) Comparison of the component size histograms for the English PN and our model. Data points represent averages over at least 10 configurations.

In panel (a) of Figure 4.8 the ϵ-dependence of CP networks with $W_C = 5$ is explored. Data points in the figure are obtained in the following way. For fixed value of the parameter ϵ and given core connectivity m_C the parameter f for which the ensemble reproduces the size of the giant component of the English PN is determined. As in the analysis of the model described by Eq. (4.3), varying m_C the intersection points at which the networks give link counts within the giant component identical to the data in the English PN can be determined and average total link counts of these networks are than plotted. In this way we can identify the CP model which reproduces all criteria for a valid comparison to the English PN: These networks reproduce the size of the giant component, link counts within and outside of the giant component, and the degree distribution of the English PN (for the latter see panel (b)).

Relevant network statistics of this ensemble constructed with $(W_C, m_C) = (5, 4.4)$, $\epsilon = 0.82$, and $f = 0.962$ are given in table 4.2. As with all comparisons of null models along the way, we notice that whilst the null models suggest that phoneme networks should be highly clustered and highly assortative by degree, quantitative comparison yield that:

(i) The English PN is significantly less cliquish than what would be expected from the null model, i.e. the null model predicts a clustering coefficient of $CC = 0.238 \pm 0.004$ whereas $CC = 0.207$ for the English PN. In fact, quantitative comparisons of the dependence of clustering on degree can be carried out and show that lowly connected words in the English PN are part of less triangles than expected in the CP networks, whereas large degree nodes are part of more triangles than expected, cf. Figure 4.8(c). Due to peculiarities of the embedding space we cannot expect the $1/k$ dependence typical in preferential attachment models Barabási & Albert (1999).

(ii) The English PN has an assortativity coefficient significantly higher than expected, i.e. $a = 0.55 \pm .01$ is predicted by the null model and $a = 0.707$ is found for English.

(iii) Average path length and diameter of the English PN are (roughly) compatible with the expectations from the null model, i.e. the null model predicts $d = 7.38 \pm 0.3$ and $d_{\max} = 36.8 \pm 5.2$ whereas $d = 7.71$ and $d_{\max} = 33$ for English.

(iv) Also the arrangement of links into intra- and inter-layer connections are roughly compatible between the null model and the English PN, we find that the ratio of intra- to inter layer links is $lr = 2.37 \pm 0.07$ for the null model, whereas $lr = 2.46$ for English.

(v) The CP null model predicts significantly larger small clusters than found in English, see panel (d) of Figure 4.8. In fact, ignoring the topology of the underlying space, a model in which new nodes are accepted if they connect to old nodes with a certain probability, implements a preferential attachment mechanism as described by Barabási & Albert (1999) for degrees for cluster sizes. One thus expects a power law with exponent close to 3 for the distribution of small clusters and, since node additions can join clusters, lower exponents in the presence of constraints from an embedding space (Newman, 2010).

4.5 Conclusions and discussion

The English phonological network represents a snapshot of the organisational patterns of word pronunciation in the human mental lexicon. In the present study we started by recognising that English words are effectively a subset of the set of all possible words formed by all possible combinations of phonemes. The latter, i.e. the set of all possible words endowed with the edit distance as a metric, defines a high dimensional discrete space which can be visualised as a stack of structured sets of words of given lengths (which we call *layers*). Phonological networks are embedded into this space. We systematically explored how spatial characteristics influence word pronunciation patterns, thereby revealing characteristics of the English language.

Percolation experiments demonstrate that some features of the English PN are a consequence of the embedding space. Importantly, we find that the presence of a power-law like regime in the degree distribution arises also in pseudolexica constructed by random sampling, i.e. contrary to what was suggested in Vitevitch (2008), no additional attachment mechanism like preferential attachment needs to be invoked as an explanation. Furthermore, our percolation models highlight the presence of a maximum degree constraint on the PN that is not a direct consequence of the embedding space. This finding suggests the presence of a maximum number of phonologically similar words that can be associated and stored, i.e. it points to a constraint of word confusability (Sadat et al., 2014) in word repertoire formation. However, percolation experiments cannot reproduce connectivity properties of the English PN. In fact, all PNs associated with artificial repertoires constructed via percolation have substantially smaller link counts and sizes of the giant component than the real PN.

These insights can be refined by further comparisons to ensembles of networks with the same number of connections as the English PN. Randomisation experiments along these lines point out that the smaller than expected sizes of the giant components generated by percolation-like experiments are not only a consequence of reduced link counts. We conclude that the rather large size of the giant component is a characteristic of word organisation.

An explanation for this enhanced connectivity of the English PN is word repertoire formation through a process of constrained word assembly over time, in such a way that preferentially connected words are included. We systematically explore this idea by introducing a series of network growth models. We first focus on the sizes of giant components and consider models in which new words are rejected if not linked to older words. Quantitative analysis leads to three main conclusions. First, the growth models corroborate the findings of constraints on maximum degree in repertoire formation. Second, within the constraints of our models, word assembly ordered by word length is a likelier explanation of the data than random word addition, giving a quantitative basis to the hypothesis that language evolved from short to long words, similar to the language acquisition of children who tend to learn shorter words first (Aitchison, 2012; Carlson et al., 2011; Storkel, 2004; Vitevitch, 2008). Third, the analysis points towards a marked *core-periphery structure* of the English PN, suggesting that in the earlier stages of repertoire formation, preferentially such short words which are similar to (or can be derived from) multiple existing words have been assembled to the language, as already suggested in the psycholinguistic literature by Siew (2013). This latter finding inspires the introduction of core periphery (CP) network models.

CP network models of repertoire formation can reproduce the size of the giant component and the number of links within the giant component of the English PN. The English PN, however, has a far larger number of edges between nodes in smaller clusters than predicted, motivating the introduction of a last type of CP networks with tuneable link counts in- and outside of the giant component. These networks, finally, provide null models which retrieve the size of the giant component, link counts, and the degree distribution of the English PN, and hence a systematic comparison of higher order correlations in network structure becomes possible. Several additional features of network organisation are well-represented by expectations from these reference CP models: diameters and distances fall within the error bounds of prediction and the link organisation in and between the layers are in good agreement. In contrast to previous work, however, these comparisons point out that the English PN is less cliquish and more assortative than expected. The first is a feature that might point to further constraints in repertoire formation. Similar to the degree constraint, suppression of triangles might point towards a mechanism of word formation that under-represents words that are too similar to others.

Whilst our study has highlighted and explored some constraints likely at play in repertoire formation, other features of the English PN are not adequately captured or well enough understood by the models we presented. This applies to detailed patterns of cliquishness vs degree, the detailed statistics of smaller components or a better understanding of the very high assortative mixing by degree of the English PN. In the spirit of the first empirical analysis of Arbesman et al. (2010b) for Spanish, Hawaiian, Mandarin and Basque, our null models also enable a detailed comparison with other languages in future work. Are the same assembly mechanisms at play in all languages? How can differences in assembly be explained or related to cultural peculiarities? These are questions beyond the scope of a physics approach, but the methodology suggested here might enable linguists to explore them quantitatively.

Chapter 5

Multiplex lexical networks reveal patterns in early word acquisition in children

Phonological similarities are but one type of word similarities in the mental lexicon. Experiments with toddlers pointed out that phonology had an impact on word learning (Carlson et al., 2014) in toddlers. Other studies had identified semantic associations as influencing word acquisition at later developmental stages (Hills et al., 2009; Wiethan et al., 2014). An interplay between semantics and phonology during development had been conjectured (Dautriche et al., 2015; Wiethan et al., 2014) but not quantified within the same network approach. This Chapter[1] overcomes this limitation by introducing the framework of multiplex lexical networks.

Abstract

Network models of language have provided a way of linking cognitive processes to language structure. However, current approaches focus only on one linguistic relationship at a time, missing the complex multi-relational nature of language. In this work, we overcome this limitation by modelling the mental lexicon of English-speaking toddlers as a multiplex lexical network, i.e. a multi-layered network where $N = 529$ words/nodes are connected according to four relationship: (i) free association, (ii) feature sharing, (iii) co-occurrence, and (iv) phonological similarity. We investigate the topology of the resulting multiplex and then proceed to evaluate single layers and the full multiplex structure on their ability to predict empirically observed age of acquisition data of English speaking toddlers. We find that the multiplex topology is an important proxy of the cognitive processes of acquisition, capable of capturing emergent lexicon structure. In fact, we show that the multiplex structure is fundamentally more powerful than individual layers in predicting the ordering with which words are acquired. Furthermore, multiplex analysis allows for a quantification of distinct phases of lexical acquisition in early learners: while initially all the multiplex layers contribute to word learning, after about month 23 free associations take the lead in driving word acquisition.

---

[1] This Chapter follows the layout of the "Three-paper" PhD thesis in that it is an original, published research manuscript published as: *Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. Scientific Reports, 7:46730.* In order to better present the content, the Chapter/paper is briefly introduced by a short summary. The Chapter follows the same structure of the original paper, in adherence to the guidelines of the "Three-paper" PhD thesis format. For a better presentation within the thesis, the words "paper" and "manuscript" were changed to Chapter. The Supplementary Information of the original paper is reported at the end of the Chapter rather than in an appendix for increased readability and easier consultation with the main text.

5.1 Introduction

Language consists of a multi-level mapping of meanings onto words (i Cancho & Solé, 2003; Aitchison, 2012). In order to communicate, humans must learn how to use linguistic structures to express thoughts as words. The cognitive processes behind language learning may organise the components of language into a so-called *mental lexicon* (ML) (Aitchison, 2012). This lexicon can be described as a network structure of interacting lexical items (e.g. word representations). Empirical studies in psycholinguistics suggest that, rather than providing exact word definitions (as in common dictionaries), the ML stores word meanings as multi-relational or multiplex word patterns (Quillian, 1967; Collins & Quillian, 1969; Collins & Loftus, 1975; Borge-Holthoefer & Arenas, 2010; Baronchelli et al., 2013; Dautriche et al., 2015). How the multiplex organisation of the ML relates to and influences language learning is still poorly understood but new techniques related to multiplex networks (Wasserman & Faust, 1994; Kivelä et al., 2014; Boccaletti et al., 2014; De Domenico et al., 2013) allow us to explore patterns of word acquisition within the mental lexicon. We achieve this by constructing an edge-coloured multiplex network (Wasserman & Faust, 1994) based on relational features of phonology, semantics, and syntax. Going beyond the topological description of the multiplex network, we exploit the multiplex structure to predict normative acquisition of young children.

Previous literature on modelling language learning and use through network science has largely focused on single-layer representations of networks (Borge-Holthoefer & Arenas, 2010; Baronchelli et al., 2013; Beckage & Colunga, 2015). We build on these works, that strongly suggest that many cognitive constraints and mechanisms affecting the use of language can be explored through network science. In fact, experimental results have shown strong correlations between network structure and human performance in various language related tasks. Measurements of retrieval times (Collins & Quillian, 1969; Collins & Loftus, 1975; De Deyne & Storms, 2008), age of acquisition (Steyvers & Tenenbaum, 2005; Hills et al., 2009; Beckage et al., 2011), creativity (Kenett et al., 2014) and even semantic degradation due to ageing (Goñi et al., 2011) have been studied and modelled using single layer networks. Further, networks of phonological word similarities have highlighted an upper bound on the size of phonological neighbourhoods as well as a tendency to avoid local clustering (Vitevitch, 2008; Stella & Brede, 2015, 2016a). Experiments with adults have indicated that these constraints relate to word confusability in identification tasks (Vitevitch, 2008; Vitevitch et al., 2012).

While single layer networks reveal aspects of the structure of language and language-related cognitive processes, it is clear that this approach cannot offer a unified view of language that simultaneously accounts for phonological, semantic and syntactic aspects of language, as required by increasingly sophisticated experimental set ups (Dautriche et al., 2015). One approach that is capable of overcoming this limitation is to use multiplex network representations as originally introduced in the social sciences (Wasserman & Faust, 1994). Within this approach, we project the complexity of an individual's mental lexicon (ML) onto a multiplex network, which we call *Multiplex Lexical Network* (MLN). Our MLN is composed of multiple network layers, with nodes representing words and layers capturing different relationships between words. Multiplex networks (Wasserman & Faust, 1994) are a specific type of multi-layer networks (Kivelä et al., 2014; Boccaletti et al., 2014) where nodes represent the same set of items on all layers. In our MLN we do not consider inter-layer connections. Instead we focus on intra-layer relationships that have previously been shown to influence children's lexical learning in single-layer network studies: 1) free-associations (Nelson et al., 2004; Steyvers & Tenenbaum, 2005), 2) shared features (McRae et al., 2005; Hills et al., 2009), 3) co-occurrence in child directed speech (MacWhinney, 2000; Beckage et al., 2011) and 4) phonological similarity (Vitevitch, 2008; Wiethan et al., 2014). We quantitatively show that by understanding the structure of multiple layers of the ML we can gain further insight into human cognition as related to language acquisition in young toddlers.

The main advantage of the multiplex approach is that it allows for a more detailed representation and quantification of many real-world systems (De Domenico et al., 2016). This added model complexity provides additional insights into a system's structure and dynamics. In the last few years multiplex modelling has provided novel insights in areas as diverse as social balance in on-line platforms (Szell et al., 2010; Gómez-Gardenes et al., 2012), emergence and stability of multiculturalism (Battiston et al., 2016), congestion in transportation networks (De Domenico et al., 2016), and ecosystems in ecology (Pilosof et al., 2017; Stella et al., 2016), cf. (Kivelä et al., 2014; Boccaletti et al., 2014) for a review on multi-layer and multiplex networks.

The multiplex network approach relates to other works in psycholinguistics investigating language via multi-layer networks. Liu et al. (Liu & Cong, 2014) analysed Chinese as a multi-layer network, composed of syntactic and phonemic layers. They found that almost half of the syntactic dependency relations are between phonologically similar words. A similar analysis was performed for English and Croatian establishing language specific relationships between syntax and phonology (Martinčić-Ipšić et al., 2016). However, these studies consider only topological features of a given language representation. We extend this approach by explicitly considering the interaction of topological features with a cognitive dynamical process, i.e. for exploring language acquisition in young children.

Below, we consider the ability of structural network features to account for normative acquisition trajectories. We evaluate performance using empirical parent reports of toddlers' productive speech, aggregated to capture the order of word learning for children between 16 and 30 months (Dale & Fenson, 1996). Our aim is to assess the predictive power of a multiplex lexicon representation to capture emerging features and possible mechanisms of word acquisition in young children. As it currently stands, our model is solely descriptive: it only approximates importance of words from observational data rather than inferring it from a generative Bayesian framework (Xu & Tenenbaum, 2007). However, we demonstrate that our multiplex model is more accurate than any single layer representation in predicting the normative acquisition trajectories of young children. We find that the multiplex network can account for developmental trends, offering predictions and interpretability that analysis based on single layer networks or word specific measures such as frequency or length cannot.

5.2 Results

5.2.1 Multiplex network construction and analysis

We first construct a multiplex lexical network composed of four layers capturing i) free associations from the South Florida association norms (Nelson et al., 2004), (ii) feature sharing

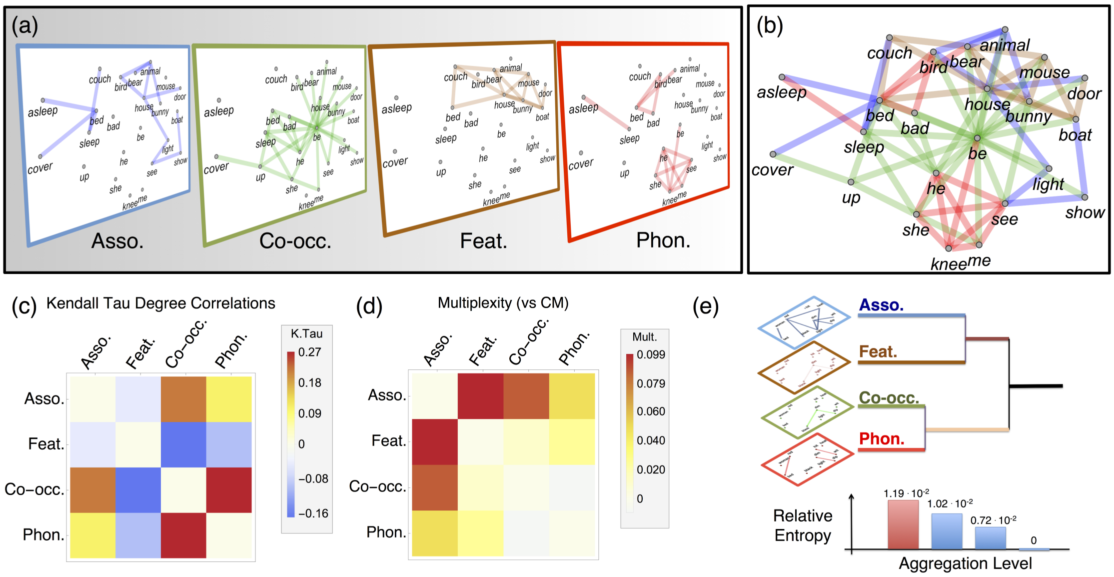


Figure 5.1: (a) Visualisation of part of the multiplex network representing a toddler's lexicon. (b) Visualisation of an example of the MLN as an edge coloured network. Both (a) and (b) include only 24 of the 529 nodes/words in the whole MLN. (c) Degree correlations across layers quantified by the Kendall Tau. (d) Multiplexity or edge overlap among different layers relative to expectations from configuration models. (e) Reducibility dendrogram for the MLN. The MLN is irreducible: its layers are so different from each other that aggregating any of them would lead to a loss of information about topological patterns.

from the McRae et al. dataset (McRae et al., 2005), (iii) co-occurrence in child-directed speech from the CHILDES dataset (MacWhinney, 2000), and (iv) phonological similarities from WordNet 3.0 (Vitevitch, 2008) (cf. Methods and SI Sect. 5.5.2). Hence the resulting MLN properties emerge from the assembly of independent measures rather than from a multiplex network design. Figures 5.1 (a) and (b) provide a visualisation of the links in each of the four MLN layers. Panel (a) treats each layer as separate whereas panel (b) describes the multiplex as an edge-coloured graph (Wasserman & Faust, 1994), in which different colours denote different relationships. As evident from (a) and (b), even if words might be disconnected on one layer, they could be connected on the MLN structure.

We report single-layer summary statistics and comparison to configuration models (i.e. random graphs preserving the degrees of an empirical network (Newman, 2010)) in Table 5.1. We focus on network features already analysed in previous works Motter et al. (2002); Sigman & Cecchi (2002); Steyvers & Tenenbaum (2005); Vitevitch (2008): degree, clustering, degree mixing, connectivity ratio and mean shortest path length. For instance, the degree of a node counts its connections Newman (2010) and on the phonological layer, it coincides with the size of its phonological neighbourhood Vitevitch (2008); Vitevitch et al. (2012). See Methods and SI Sect. 5.5.3 for precise definitions and interpretations.

Tab. 5.1 shows that the phonological layer has a smaller mean degree compared to other layers. Furthermore, apart from the association layer all other layers are highly fragmented. Their largest connected components include between 24.2% and 62.4% of the 529 words. These percentages quantify how many words could be reached from each other by navigating through links on individual layers. In contrast, the *multiplex network* is connected (De Domenico et al., 2014) – every pair of words is connected by a path that potentially exploits different relationship types. Compared to configuration models, except for the co-occurrence layer all other layers display higher clustering, non-zero assortativity and slightly larger path length in their largest connected component. In agreement with previous work (Gravino et al., 2012; Sigman & Cecchi, 2002; Vitevitch et al., 2014; Beckage & Colunga, 2015) these patterns suggest that the MLN layers display a core-periphery structure. From a cognitive perspective this feature might possibly facilitate navigation from word to word within cores (see SI Sect. S3 for further discussion). Layers also differ in their degree distributions, see SI Fig. S2: they are exponential-like for the synonym and phonological layers but much more heavy-tailed for the co-occurrence and association layers. Heavy tails indicate the presence of network hubs, i.e. words which are significantly more connected than average. These hub words play a central role for navigation through concept space (Sigman & Cecchi, 2002; Motter et al., 2002; Gravino et al., 2012) and could therefore be good candidates for earlier acquisition.

Is it appropriate to represent the ML in terms of four separate layers or can we get the same type of topological information with fewer layers? Structural reducibility analysis (De Domenico et al., 2015) can test if multiplex layers can be aggregated without losing information (see SI for details). Results reported in Fig. 5.1 (e) show that aggregation cannot be performed without information loss, demonstrating that the chosen 4-layer representation is in fact *irreducible* De Domenico et al. (2015). This is not to say that each layer plays a role in language acquisition, just that the information encoded in each layer is different from all other layers.

Before turning to modeling word acquisition, we investigate the MLN structure in more detail. Specifically, Fig. 5.1 (c) analyses the similarity between degrees of a word on different layers. We find that words in the feature and co-occurrence layers tend to have negative degree correlations (Kendall Tau $\kappa \approx -0.16, p < 0.0001$), indicating that hubs in one layer tend to have lower degrees in the other layers. This may suggest a type of semantic differentiation (Steyvers & Tenenbaum, 2005) in English in which words that have similar features tend to not be used in close proximity of each other, at least when speaking to young children. In contrast, the co-occurrence and phonological layers display positive degree correlations (Kendall Tau $\kappa \approx 0.27, p < 0.0001$): in the children's lexicon, words having many co-occurrences also have larger phonological neighbourhoods, further supporting the idea that phonological similarities and co-occurrences in child directed speech influence each other (MacWhinney, 2000; Wiethan et al., 2014).

Another question of interest is to what extent links overlap across layers. This can be quantified by edge overlap (or multiplexity (Gemmetto & Garlaschelli, 2015)). Results are reported in Fig. 5.1 (d). As expected, we find some overlap between connections in the feature, co-occurrence and association layers; these layers capture semantic aspects of the ML. Beyond the overlap between these layers, links tend not to overlap significantly more than one would expect in configuration model across the other layers. This suggests that different MLN layers tend to capture different relational aspects of the ML, again emphasising that the MLN is a more complete model than single network representations. We now explore the importance of individual layers and of the MLN as a whole in explaining acquisition trajectories.

Empirical Network	$\langle k \rangle$	CC	a	$Conn.$	$\langle d \rangle$	$CC\,(CM)$	$a\,(CM)$	$Conn.$ (CM)	$\langle d \rangle$ (CM)
Associations (Asso.)	9.3	0.20	-0.1	99.6%	3.2	0.03(1)	−0.01(6)	99.6%	3.00(1)
Feature Norms (Feat.)	9.0	0.63	-0.01	24.2%	1.8	0.38(2)	−0.07(6)	24.2%	1.72(1)
Co-occurrences (Co-occ.)	8.1	0.69	-0.44	62.4%	2.2	0.75(3)	−0.40(2)	62.4%	2.19(1)
Phonological Sim (Phon.)	1.31	0.37	0.48	33.1%	7.7	0.02(1)	0.03(2)	45.5%(8)	5.31(7)
Multiplex Aggregate	26.5	0.33	-0.07	100%	2.4	0.18(5)	−0.10(8)	100%	2.25(1)

Table 5.1: Metrics for the MLN layers with $N = 529$ nodes, listing mean degree $\langle k \rangle$, mean clustering coefficient CC, assortativity coefficient a, percentage of nodes in the largest connected component $Conn.$, and mean shortest path length of the largest connected component $\langle d \rangle$. The quantities are measured for the layers of the MLN and for randomised equivalents with the same degree sequences (denoted with CM). Error bars represent standard deviations and are reported in brackets behind the last significant digit.

5.2.2 Multiplex Orderings: Results and Discussion

Here we explore the process of word learning through the MLN assuming a preferential acquisition scenario (Hills et al., 2009), where words are learned earlier if they are central in the language environment. We assume that relevance of the language environment can be summarised by connectivity in a language network (Sigman & Cecchi, 2002; Motter et al., 2002; Baronchelli et al., 2013; Beckage & Colunga, 2015) rather than making assumptions on the statistics of relationships to infer relevance, as is more standard in a Bayesian approach (Xu & Tenenbaum, 2007; Frank et al., 2009).

We generate a word acquisition ordering τ by ranking words according to measures computed on the MLN. We assume that if one ranking is predictive, it would indicate that the feature generating it is relevant to early language learning. We compare MLN-based orderings to those based on non-relational, word-specific information such as word frequency and word length. For consistency with the literature, we investigated features previously considered in linguistic networks such as word degree (Steyvers & Tenenbaum, 2005; Beckage et al., 2011; Vitevitch et al., 2014), closeness (Motter et al., 2002; Sigman & Cecchi, 2002), betweenness (Borge-Holthoefer & Arenas, 2010; Motter et al., 2002; Sigman & Cecchi, 2002) and PageRank (Griffiths et al., 2007). For more details see Methods and SI Sect. 5.5.6.

We compare our orderings against an ensemble of empirical age of acquisition orderings, obtained from a probabilistic interpretation of the normative age of acquisition of words based on the MacArthur-Bates Communicative Development Inventory(MacWhinney, 2000) (CDI) and over a population of roughly 1000 toddlers (see Methods). The comparison is performed by computing the average word overlap $O(\tau, t)$ of ordering τ with normative orderings, namely how many words τ correctly predicts as learned until the inventory includes t words. Word *gains* are then obtained by subtracting $O(\tau, t)$ from the expected number of correctly predicted words by random guessing (see Methods and SI Sect. 5.5.6). Word gains are normalised by the inventory size t and their Z-scores relative to statistics from random guesses are estimated. Positive *vocabulary normalised word gain* indicates better than random performance of ordering τ and negative values indicate performance inferior to random guessing.

Fig. 5.2 reports normalised word gains (a) and their Z-scores (b) for the best performing orderings from our analysis. No ordering always outperforms the others: we notice shifts in which word features are the most relevant for different periods of word learning. This agrees with the idea that the progression of acquisition changes over the course of development: children learn some initial words and then tend to generalise those words based on network connectivity later, i.e. preferential acquisition (Hills et al., 2009). This cognitive interpretation of the observed shifts at different acquisition stages allow us to distinguish at least three distinct learning phases. In a first stage, comprising months 19 and 20, which we call the *very early learning stage* (VELS), we are trying to predict the first 40 learned words (grey overlays in Fig. 5.2). Direct sampling (Grimmett & Stirzaker, 2001) for this period indicates that our results are below the 2.5% threshold for statistical significance – the network appears to contain little information relevant for predicting acquisition. This phase is followed by an *early learning stage* (ELS) covering an age range between 20 and 23 months. This phase is characterised by *closeness centrality outmatching* every other ordering (see also SI Sect. 5.5.6). Last we discriminate a *late learning stage* (LLS) comprising ages between 23 and 28 months. During this stage the dominant hub words have already been learned and more localised structure is likely affecting learning. LLS is characterised by degrees in the association layer and frequency performing equally to or slightly better than closeness and other word features (cf. SI Sect. 5.5.6).

Interestingly, while previous work has found frequency to be a good predictor for normative age of acquisition (Kuperman et al., 2012), results in Fig. 5.2 suggest that word length and some topological network information are more predictive of early acquisition ordering. Good performance is achieved by word length in ELS, possibly because of a *least effort effect* (i Cancho & Solé, 2003) that suggests that short words are easier to memorise and learn. This hypothesis is also supported by the observation that word length loses its predictive power at later learning stages (word gain Z-scores are compatible with random fluctuations after 300 words have been learned). We interpret this happens when more sophisticated learning strategies are being used. One of them might be mediated by degree orderings. Among the four layers, ordering words according to their degree in the association layer gives the best performance and it can be up to 100% more predictive than random guessing (cf. SI Fig. S7). This confirms previous results suggesting that association norms are generally good predictive models for early language learning (Hills et al., 2009; Steyvers & Tenenbaum, 2005; Beckage et al., 2015). We conjecture associations perform well in our case because they are strongly related with semantic memory (Nelson et al., 2004; Zock, 2015; De Deyne & Storms, 2008).

Previous work has pointed out that in early learning, during VELS and ELS, it is very difficult to outperform random guessing because children may not have a clear strategy for learning new words early in language acquisition (Beckage et al., 2015). Instead here we find that better than random prediction is possible even in very early learning stages, at least for normative acquisition. Using the multiplex closeness of words it is possible to predict up to 160% words more than at random (cf. SI Fig. S7) and up to 25% more than by using degrees in the association layer (cf. SI Fig. S8). As reported in SI Sect. 5.5.6, we checked that multiplex closeness consistently outperforms every single-layer orderings during early learning stages. This is a promising result in that multiplex approaches may allow us to quantify learning strategies in toddlers. Also other multiplex measures such as versatile PageRank (De Domenico et al., 2015) perform up to 10% better than degrees in the association layer (cf. SI Fig. S8). The enhanced prediction ability of multiplex features indicates that the multiplex representation of the ML captures meaningful influences of multiple types of feature representations. Altogether, we can conclude that early learning is strongly related to the multiplex structure, specifically the closeness of words considering *all* layers that form the MLN.

Results for the LLS show a reversal of trends. Here we find that a single layer based on association degree ordering gives best predictability. Global and local multiplex measures, as well as word specific features such as word frequency, perform similarly or slightly worse until 400 words have been learned (see Fig 5.2 and SI Fig. S8). This indicates that preferential acquisition is very important, but at this point in development the dominant relevant relationships are those captured by the association layer. This is compatible with the observed emergence of semantic learning in children (Wiethan et al., 2014). However it has to be noticed that statistical significance deteriorates markedly during later stages of learning making comparisons of performance in the LLS difficult.

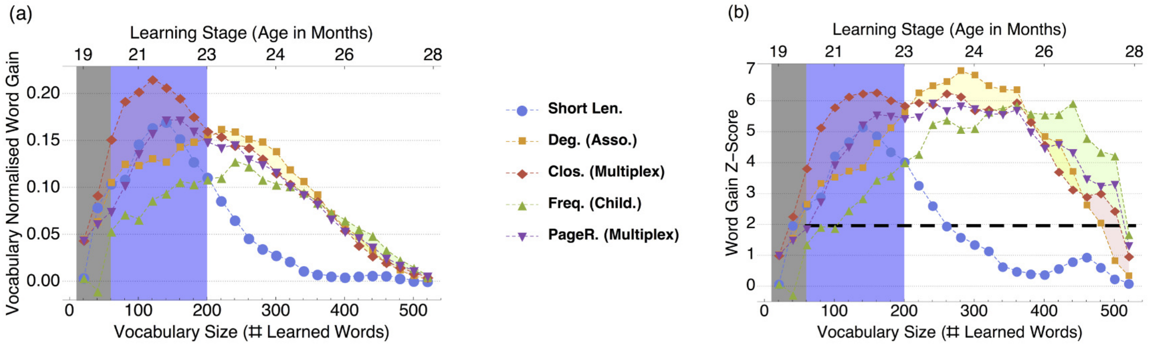


Figure 5.2: Left: Vocabulary normalised word gains for different orderings: shorter words first (Short Len.), words with higher degree in the association layer first (Deg. Asso.), closer words on the whole MLN first (Clos. Multiplex), more frequent words first (Freq. Child.) and higher multiplex Page Rank first (PageR. Multiplex). A normalised word gain of 0.20 means that 20% of words in the vocabulary have been predicted by the ordering strategy on top of expectations from random guessing. Higher word gains indicate better predictability. Right: Statistical significance indicated by word gain Z-score of the respective orderings. The distribution of the random overlaps can be approximated by a Gaussian after 60 words have been acquired (see SI Sect. 5.5.5). In this range, represented by the black dashed line, a score $Z \geq 1.96$ denotes word acquisition patterns that are different from random fluctuations with a 2.5% significance level. In both plots, error bars are the same size as the dots. Best performing orderings are highlighted with different overlays, which identify different learning stages: VELS (in grey), ELS (in light purple) and LLS (in white). Direct sampling was used in VELS for testing the statistical significance of results.

So far, we evaluated orderings based on single network layers and the combined multiplex, showing that prediction performance can benefit from taking account of information from multiple layers. In this analysis we assumed that each layer plays an equal role. In the next section we evaluate the changing influence of layers during the evolution of the lexicon.

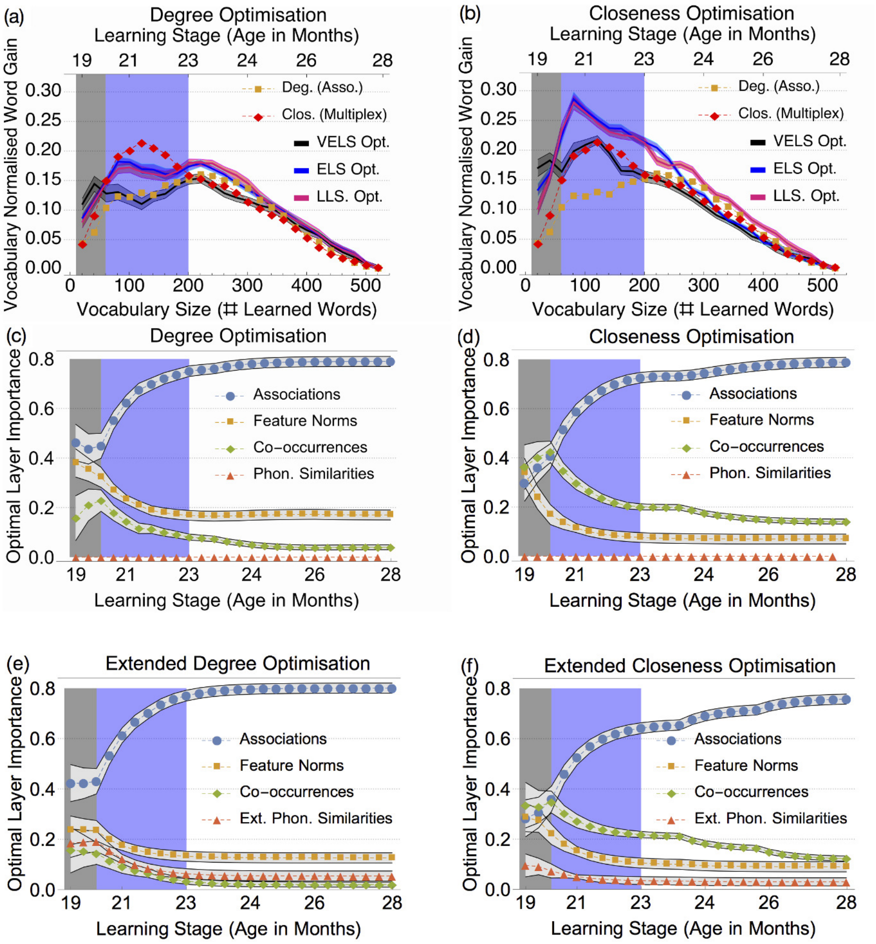


Figure 5.3: Top (a-d): Optimisation results for the original MLN. (a,b): Vocabulary normalised word gains corresponding to the optimal layer influences at the end of VELS (black), middle of ELS (blue), and middle of LLS (pink) for degree optimisation (a) and closeness optimisation (b). Error margins represent standard deviations over randomised ensembles. In (b) no optimisation of degree outperforms multiplex closeness during ELS. (c,d): Average optimal layer weights indicating the influence or importance of layers over different learning stages obtained from Monte Carlo experiments with degree (c) and closeness (d) optimisation. The phonological layer resulted in influences around 10−3. The same optimisation trends were obtained when the phonological layer was excluded from the optimisation. Bottom (e,f): For comparison, optimal layer weights for the MLN where rather than the phonological layer with 529 words an extended one with 30,000 words is used, instead. Considering the extended phonological layer leads to word gains at least 6% better than (a) and (b) (cf. SI Table 3).

5.2.3 Optimisation of layer influences

To explore the influence of different layers on word acquisition, we consider linear combinations of word features on each layer in order to obtain weighted multiplex measures (cf. Methods section). For a given network metric, we optimise the coefficients of the linear combinations to maximise vocabulary normalised word gains. The resulting layer weights, that maximise predictability, indicate the influence of the respective layers on word acquisition. We explored several types of optimisation, based on degree, closeness, betweenness and local clustering (cf. SI Sect. 5.5.7 and SI Tab. S3). We focus on the two best performing quantities from the ordering experiments, namely degree and closeness.

In order to avoid over-fitting, we perform a Monte Carlo robustness analysis: we optimise over subsets of word trajectories sampled uniformly at random, consisting of only 80% of the original words. Averages are computed over these trajectories and over different realisations of the age of acquisition trajectories from the normative age of acquisition ordering. We also confirmed that the improvement in performance of optimised multiplex parameters is strongly dependent on the structure and overlap of layers, as experiments on randomised multiplex networks result in far smaller word gains (see SI Sect. 5.5.8).

Optimisation results (cf. Fig. 5.3) indicated negligible contribution of the phonological layer, with weights ($\approx 10^{-3}$). Degree and closeness orderings based only on the phonological layer perform poorly as well, yielding word gains compatible with random fluctuations at a 5% significance level (see SI Sect. 5.5.6 and SI Fig. S5). We retrieved the same optimisation results for the other layers when the phonological layer was excluded. In addition, we considered phonological word scores coming from an "extended" phonological layer for adults, already analysed by Stella and Brede (Stella & Brede, 2015) and including 30000 words (see also SI Sect. 5.5.6.). Assessing the influence of adults' phonology over word acquisition in toddlers matches previous approaches in the relevant literature (Carlson et al., 2014) and we present results in our MLN framework in Fig. 5.3 but also in SI Sect. 5.5.7.

Optimal layer influences are displayed as a function of age for the MLN layers in Fig. 5.3 (c) and (d). Similar patterns of layer influence are found for rankings based on both degree and closeness. We see a clear distinction between different learning stages. When degree is considered, cf. Fig. 5.3 (c), in VELS the main layer contributions stem from the association and the feature norm layers. Instead, when optimising closeness, cf. Fig. 5.3 (d), in VELS all layers are found to contribute equally. In both cases the ELS, the next learning stage, marks a transition region between the VELS (before) and the LLS (after) in which only the association layer dominates. This transition is what motivated our choice of distinguishing between VELS and ELS. For degree-based optimisations, contributions from the co-occurrence layer become very small in the LLS, while associations contribute 80% of weight. An analogous transition is observed in the ELS phase when closeness is optimised and after that associations make up 80% of the total layer influence. The dominant influence of associations in later stages of learning is indicative of the emergence of preferential acquisition (Hills et al., 2009) that we already observed in ordering experiments.

When extended phonology is considered, in Fig. 5.3 (e) and (f), the influences of semantic layers remain consistent, however, phonology has a higher impact when phonological neighbourhood size/degree is considered rather than closeness. In VELS the extended phonology has importances similar to co-occurrences but the influence of both approaches 0 later in development.

In Fig. 5.3 (a) and (b) we compare vocabulary normalised word gains of trajectories based on optimal layer influences against our ordering experiments (cf. SI Fig. S9 for the respective Z-scores and previous section). On the one hand, when predicting only words learned in VELS, the optimal influence orderings outmatch both degree in the association layer and multiplex closeness during VELS but perform worse at later stages. Direct sampling indicates that these results are significant with a p-value of 0.01. Optimal influences up to the middle of ELS and LLS provide similar word gains, which are both up to 25% higher than the optimal word gains from

VELS. This suggests that the *same learning strategies* may be adopted *after* VELS, starting in ELS and lasting throughout LLS. On the other hand, results in Fig. 5.2 (a) highlight the importance of closeness centrality for predicting word acquisition at early stages. In fact, in ELS even the optimal linear combination of degrees results in *smaller word gains* compared to the unweighted multiplex closeness centrality. Optimal linear combinations of closeness perform even better than the unweighted multiplex closeness, see Fig. 5.2 (b). Notice that closeness is a *global* network feature, accounting for the position of a node relative to all other nodes. This is distinct from *local* features, such as degree, which only measure a node's relationship to its first neighbours. Our findings give support to the hypothesis that, particularly in the early learning stages, word learning is influenced by the global structure of the multiplex lexical network. This global multiplex structure is able to capture some important word patterns that influence early lexicon development.

5.3 Conclusions and future work

Introducing the framework of *multiplex lexical networks* (MLN), we have described and analysed normative word acquisition patterns of children between the ages of 19 and 28 months. Our approach considers semantic, phonological and syntactic contexts and captures more linguistic information than separate analysis of any single layer network previously could (Beckage et al., 2011, 2015; Hills et al., 2009; De Deyne & Storms, 2008). We go beyond a purely topological description by exploiting network properties to predict the order of normative word learning in children. Interestingly, we find that the best topological feature in predicting word learning changes through the course of development. This allows us to use network information to distinguish three stages of learning: (i) a very early learning stage where all but the phonological layer contribute substantially to prediction, (ii) an early learning stage which marks a transition period, and (iii) a late learning stage in which contribution from word associations dominates word learning. We believe the last point is due to free associations serving as a much better proxy for detecting superordinate words compared to other network features (cf. SI Sect. 5.5.6 for a quantitative discussion). Further assessment of the extent to which the taxonomic organisation serves as a mediator variable influencing word prediction is an interesting direction for future work.

Comparing the predictive power of various MLN features over time confirms: (i) the superiority of some multiplex network characteristics relative to single layer networks and word specific measures such as frequency or word length, and (ii) the special role closeness centrality might play in early word learning. This is emphasised in two ways: (i) strong performance of multiplex closeness, outperforming all combinations of local network features in the early learning phase, and (ii) the best overall predictions of the order words are learned is related to a weighted multiplex version of single-layer closeness centralities. We thus find strong indications that, particularly at early stages of learning, word acquisition in children is driven by minimising relational distances of learned words relative to other words in both the semantic and syntactic space of the mental lexicon.

With the MLN, we explored questions about normative lexical acquisition, uncovering developmental learning stages and quantifying the influence of certain types of linguistic information on early acquisition. Nevertheless, our MLN is limited in important ways. It is important to bear in mind that the MLN representation is only a projection of an individual child's full ML. Additionally, the results presented in this Chapter consider normative lexical acquisition. That is to say, a specific child may not learn words in the same order as the normative orderings. Instead normative order is obtained by averaging over roughly 1000 productive vocabulary reports. We attempt to address this shortcoming by sampling multiple orders probabilistically from these norms. Extending the model and testing the universality of the learning strategies in VELS and ELS on longitudinal data for individual children would be an interesting future research direction (Aitchison, 2012).

Another limitation is the use of a relatively small vocabulary. While the CDI is a commonly used check list of words a child produces, there are many words that a child may know which are not on the CDI. By expanding our model to the full vocabulary of a child, and for longer periods of development, we can further increase our understanding of word learning.

While the present study focuses on word learning in English, results should be corroborated by considering word learning in different languages. Future work can extend our analysis by considering correlations in word learning across other cognitive and linguistic domains to build an even richer picture of language organisation and learning than presented here. A stream of work in psycholinguistics has explored detailed mechanisms of word learning (Xu & Tenenbaum, 2007). Such work typically explores how words are memorised when clues are presented in different order and focuses on the underlying mechanisms of inferring meaning-object mappings, pointing out that learning is strongly dependent on context. Integrating context-dependent information in our MLN framework would surely be an exciting future research direction for investigating individual children's learning trajectories.

The marginal influence of the original phonological layer found in this study can be reconciled with previous findings.Dautriche et al. (2015) showed that 18 months old toddlers failed to learn an object label when it was a phonological neighbour of a noun they already knew. In our dataset 80% of words learned by toddlers within the first two months are nouns. This noun-richness is in opposition with orderings extracted from phonological degree and closeness, which identify mainly non-noun words, verbs in particular, as learned initially. This trend is not surprising, as having learned a verb earlier makes it easier for toddlers to learn its phonological neighbours (Dautriche et al., 2015). However, this contrast in noun-richness at early development stages reflects in the poor performance of the original phonological layer in predicting normative acquisition ordering. Nonetheless, we show that phonology influences word acquisition by considering a phonological network extended to the adults' lexicon, like in (Carlson et al., 2014). This extended layer identifies nouns as early candidates for learning and provides significant prediction results. This interplay between phonology and acquisition is in agreement with previous findings in the relevant literature (Carlson et al., 2014; Stamer & Vitevitch, 2012; Vitevitch et al., 2012, 2014).

We have shown strong evidence that multiplex lexical networks capture a richer picture of the mental lexicon than previous works using only single-layer networks, in spite of the above limitations. More importantly, this work provides novel formalised methods for exploring and explaining the influence of linguistic features on early acquisition.

5.4 Methods

5.4.1 Age of acquisition dataset

Age of acquisition orderings are constructed based on the MacArthur-Bates Communicative Development Inventory (CDI) norms (Dale & Fenson, 1996). The CDI, based on parent report of the productive vocabulary of children aged between 16 and 30 months, has been shown to be related to future language ability in young children (Dale & Fenson, 1996; Fenson et al., 1994). The norms are averaged over more than 1000 vocabulary reports and indicate the percentage of children, at a given age, that reportedly produce (and understand) a specific word. From these norms, we sample orderings of word learning. For example, "mommy" and "ball" are reported as produced by 93% and 64% of children by 16 months, respectively; these words are (usually) learned earlier than "chair" which is produced by only 14% of 16 month olds. We assume a word is known once 50% of children in a given month produce that word. We then sample an order of words learned within a month assuming that higher rates of production probabilistically indicate earlier learning. Starting from position one of the ranking, we: (i) sample a learned word with probability proportional to the percentage of children which are able to produce it, conditioned on the overall production rate being over 50%, (ii) erase it from the list of candidate words to be learned at later positions and (iii) proceed one position down in the ranking and start from (i). The sampling starts from month 16 and it stops when every word has been sampled. Note that a non-probabilistic version of this ordering has been used in other developmental modelling contexts (Hills et al., 2009; Beckage et al., 2015).

5.4.2 Construction of the multiplex network

The four layers of the MLN were selected because of their use in previous single-layer network studies investigating word learning of young children (Beckage et al., 2011; Beckage & Colunga, 2015; Hills et al., 2009; Borge-Holthoefer & Arenas, 2010). The four layers we consider here are (i) a semantic layer based on the Florida Free Association Norms (Nelson et al., 2004) where a link exists from word A to word B if word B is a free associate of the cue word A; (ii) a layer capturing feature similarity based on the McRae Feature Norms (McRae et al., 2005) where words A and B are connected if they share at least one semantic feature; (iii) a layer based on word co-occurrences (measured in child-directed speech (MacWhinney, 2000)) where words A and B are connected if they co-occur more than 45 times, where the threshold 45 was chosen approximately to match the link density of the other semantic layers; and a (iv) layer capturing phonological similarities (based on IPA transcription from WordNet 3.0 (Miller, 1995)) where words A and B are connected if they have IPA transcriptions with edit distance one. We treat the association layer as undirected and edge weights in layers (i) to (iii) are ignored. The resulting MLN is composed of 529 words which represent the intersection between the CDI data and the set of words which have at least one connection on any layer (see SI).

5.4.3 Cognitive interpretation of network metrics

We investigate the same local and global features of words/nodes that were investigated in previous works about network representations of language: degree, closeness, betweenness and PageRank (see SI). In a given network, the degree of a node is the number of its links (Newman, 2010). In our MLN, a word is characterised by one degree measure per layer. For instance, on the association layer a word with degree $k^{(Asso.)}$ has $k^{(Asso.)}$ associates. On the phonological layer the degree of a word coincides with its phonological neighbourhood size (Vitevitch et al., 2014): a word with degree $k^{(Phon.)}$ has $k^{(Phon.)}$ similar sounding words. Higher degree words in semantic networks were found to have lower age of acquisition norms (Steyvers & Tenenbaum, 2005), while degree on the phonological layer correlated positively with word confusability (Vitevitch et al., 2012). On the whole MLN, we can associate each word to its multidegree, namely the sum of its degrees on all the layers (De Domenico et al., 2013). Fundamentally, the usefulness of measures such as degree would suggest that words with multiple associates, multiple phonological neighbours and multiple shared features are those words that are learned earliest. The mechanisms behind this phenomenon could be that words with many associates and shared features are those words that are important in the language environment because they play a central role or because they are important to the parent or child. We remain agnostic as to the exact reason behind the increased 'importance' of a word, but suggest that it can be summarized by the degree. Closeness relates to how fast information spreads from a node to others (Newman, 2010). It is interesting for cognitive science since previous work showed that words closer on semantic topologies tend to be processed together in shorter time (Collins & Quillian, 1969; Collins & Loftus, 1975; Motter et al., 2002). Betweenness captures the extent to which a node falls on the shortest path between pairs of words (Newman, 2010). It captures centrality of words through a hypothetical navigation on the ML (Sigman & Cecchi, 2002; Motter et al., 2002; Borge-Holthoefer & Arenas, 2010) (cf. SI Sect. 5.5.6). PageRank identifies the likelihood of reaching a given node by a random walk in a network. High PageRank words on semantic networks were found easier to retrieve in fluency tasks (Griffiths et al., 2007). Closeness, betweenness and PageRank on multiplex networks exploit jumps through layers (cf. SI Sect. 5.5.6).

5.4.4 Overlap measures and word ranking

A word trajectory is an ordered list $\tau = (w_1, w_2, ..., w_N)$ indicating the exact order in which words are learned. We define τ_{aoa} to be the ensemble of rankings derived from the CDI norms. We sample each specific ordering τ_{aoa} based on the probability of production as reported in the age of acquisition norms.

Predicted word orderings are generated as follows. First, each word is ranked according to a word score, s_i, which is derived from network specific or extrinsic word features (e.g. frequency or word length). Words are then ordered according to word score, starting with the largest score. Words receive a position in the predicted trajectory according to their position in the ordering of word scores. Ties in word orderings s_i are taken into account by averaging over all resolutions. For example, one might wish to consider orderings based on degree in the association layer and would obtain the following word scores $s_{food} = 62$, $s_{water} = 45$, $s_{eat} = 20$. The corresponding predicted acquisition trajectory would be $\tau = (food, water, eat)$. Apart from the features mention in the main text of the Chapter, we also considered many others (cf. SI Sect. 5.5.6).

To evaluate predictive performance of a word score, we first measure the overlap $O(\tau, t)$ between the predicted learning trajectory τ and the empirically known trajectory τ_{aoa} at time t by counting the number of words that co-occur in τ and τ_{aoa} up to time t and averaging over different τ_{aoa}. We define the word gain $g(\tau, t)$ as the overlap minus random fluctuations, i.e.

$$g(\tau, t) = O(\tau, t) - \langle O(\tau_{ran}, t) \rangle. \tag{5.1}$$

The vocabulary size normalised word gains are computed as $G(\tau, t) = g(\tau, t)/t$ while the word gain Z-Scores consider deviations of overlaps from random guessing in terms of standard deviations σ, as $Z = g(\tau, t)/\sigma(O(\tau_{ran}, t))$.

5.4.5 Calculating optimal combinations of layers

To allow for varying influences of layers within the multiplex we construct word scores as convex combinations of individual layer influence, i.e. a word score s_w for word w is obtained as:

$$s_w = \alpha s_w^{(Asso)} + \beta s_w^{(Feat)} + \gamma s_w^{(Co-occ)} + (1 - \alpha - \beta - \gamma) s_w^{(Phon)}, \tag{5.2}$$

where $s_w^{(Asso)}$, $s_w^{(Feat)}$, $s_w^{(Co-occ)}$, and $s_w^{(Phon)}$ are word scores obtained from the respective single layer metrics and the coefficients α, β, γ give the influence of each layer on the overall word score s_w. This linear combination of single-layer features is similar in spirit to previous decompositions of multiplex centrality, see Solá et al. (2013) and discussions in (De Domenico et al., 2015). The optimisation finds the influences α, β, γ that maximise vocabulary normalised word gain (either over the entire time period or over specific learning phases) after removing 20% of the words at random. 20% was chosen in order not to remove too many words particularly in the very early word learning stages. Performance is evaluated on all words. Optimisation was performed using a differential evolution method (Price et al., 2006). Averages are calculated over 50 configurations of left out words for 30 normative age of acquisition orderings.

5.5 Supplementary Information

In the following text we provide additional details to the main Chapter. In Section 1 we give further details about the age of acquisition dataset and characterise the normative ordering. Further technical details on network definitions are given in Section 2. In Section 3 we briefly review the use of single layer network measures referenced in the main text. In the same section we also report and discuss a k-core analysis of the individual MLN layers, in order to explore core-periphery structure of the MLN. Section 4 contains an overview of the multiplex measures adopted in the main text. Section 5 provides some additional analytical results on the random word guessing baseline, which was used as a reference for the overlap word measures introduced in the Methods section of the main Chapter. Section 6 reports the word gains for many additional ordering experiments carried out and only cited in the main text. Orderings are either based on network features or on word characteristics. Section 7 discusses the optimisation results obtained for betweenness centrality and local clustering, where each layer of the MLN has a different influence. We conclude with Section 8, where we test the influence of individual word features on the predictability outcome of the optimisation procedure.

5.5.1 Vocabulary size of children over time

In order to establish an empirical relationship between the number of words learned (i.e. the vocabulary size) and the actual age of children, we considered CDI data collected in a research setting and publicly available (Frank et al., 2016). These CDI data exhibit the same trends as the original norming data used in the main text (Dale & Fenson, 1996) but allow for identification of a child's specific vocabulary size. A linear fitting $y = ax + b$ of the number of words x learned by children of age in months y up to month 30 allowed for us to attribute an average age to children of a given vocabulary size ($a = 0.0151 \pm 0.0006$, $b = 19.7 \pm 0.2$, adjusted R squared $R^2 = 0.53$). Notice that despite individual points in the dataset ranging between ages of 16 up to 30 months, the dependent variable of the linear fit only ranges between 19 and 28 months (cf. SI Fig. S1). This estimated information was used for the plots in Figures 2 and 3 of the main text.

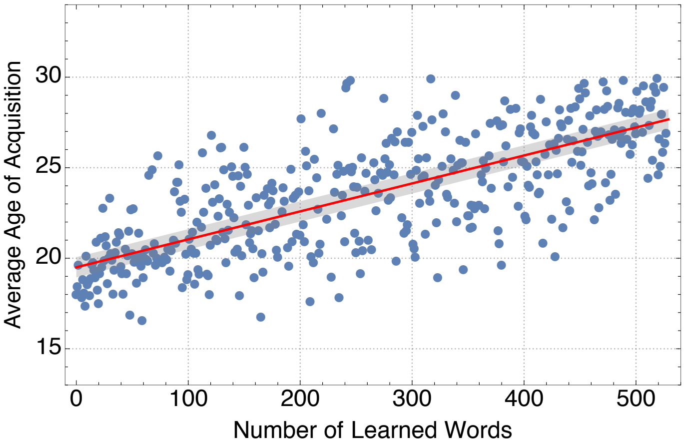


Figure 5.4: Linear fitting $y = ax + b$ of the number of words x learned by children of age in months y up to month 30 from the dataset. The overlay indicates standard errors estimated from the fitting procedure.

5.5.2 Edge definitions

Every node in the MLN represents a word and is replicated across layers. The association and feature norm layers are semantic, in that their relationships provide information about word meanings, while the phonological layer is indeed phonological, i.e. it is based on similarity patterns across word pronunciations. The layer of co-occurrences in child directed speech likely contains information related to semantic, syntactic, and phonological similarity. Notice that the order of layer is not important in our MLN framework.

The free association layer is based on the empirical University of South Florida Free Association Norms (Nelson et al., 2004). The dataset was built over almost 750,000 empirical free association pairs produced by 6,000 participants as responses to 5,019 stimulus words. Participants were asked to indicate the first target word that came to mind which was related to the presented cue word. The free association norms were obtained by considering only statistically significant associations. From these norms directed edges were retrieved from a cue word (e.g. "eat") to a target word (e.g. "food") with a given normalised frequency (e.g. 0.41). Therefore a weighted directed edge from word A to word B in the original dataset means that "A is freely associated with B". When restricted to the overlap with the CDI vocabulary check list, the network contained 529 words. This CDI association network displayed a reciprocity of 0.38 (i.e. 38% of the edges were bidirectional). However, in this preliminary investigation of multiplex lexical networks, we considered only non-weighted and undirected network layers for the sake of simplicity. We ignored edge weights and converted links from directed to undirected. Hence, in the MLN an edge between words A and B means that one of them is empirically associated with the other (e.g. “food" could be a response to “eat" or “eat" a response to “food"). This undirected network representation used by Steyvers & Tenenbaum (2005) and Hills et al. (2009) in their modelling of early language growth using network representations. Instead, a directed version was used in Beckage et al. (2015).

The feature norms layer is based on the McRae feature norms dataset (McRae et al., 2005), i.e. a set of feature norms collected from approximately 725 participants for 541 living (e.g. dog) and non-living (e.g. chair) basic (noun) concepts. Participants were asked to list semantic features of each concept, capturing the most salient and most relevant features. In the semantic features layer generated from this feature norming study, words A and B are connected if they share at least $X = 1$ semantic features (e.g. "balloon" and "ball" share the SHAPE feature).

The co-occurrence layer is based on the CHILDES dataset (MacWhinney, 2000) and it considers word co-occurrences based on child directed speech. In the co-occurrence layer an edge exists between word A and B if word A occurs within five words before or after word B. Because of the high rate of spurious connections, the co-occurrence between two words must have a frequency higher than a given threshold C to be considered as an edge in our graph. We set $C = 45$ in order for the co-occurrence layer to display a connectivity as close as possible to that of the association and feature norms layers. As an example words, "be" and "back" co-occur more than 45 times in the dataset and are therefore connected in the co-occurrence layer of the MLN. From a cognitive perspective, spurious connections are expected in children co-occurrence data because sentences from or directed to children tend to be short and because topics tend to change often (MacWhinney, 2000).

The phonological layer of the MLN is based on phonological similarity among words. We computed similarities based on the IPA phonological word transcriptions obtained from WordNet 3.0 (Miller, 1995). An edge in the phonological layer between words A and B means that the IPA phonological transcriptions of these words have edit distance one, in agreement with the definition of phonological word similarity adopted in other studies(Vitevitch, 2008). For instance, the words "bad" and "bat" are phonologically similar, as they differ in their last phoneme, and are therefore connected in the phonological layer.

5.5.3 Single-layer network metrics for linguistic networks

All single layers, as defined above have isolated nodes and small connected components, i.e. in each layer some pairs of nodes exists that are not connected by paths. This is evident also from the example visualisation in panel (a) of Fig. 1 from the main text. On the other hand, the whole multiplex lexical network is connected. While the multiplex literature admits several definitions of multiplex connectedness (Wasserman & Faust, 1994; Kivelä et al., 2014; Boccaletti et al., 2014), we here adopt the notion of connectedness from De Domenico et al. (2014), where a multiplex network is considered connected if its aggregate is connected. As reported in Table 1 of the main text, the largest connected component of the MLN's aggregate network includes all 529 words: the full multiplex lexical network is therefore considered connected.

Notice that the MLN framework considers edge-coloured graphs as multiplex networks following the definition originally proposed within the social sciences (Wasserman & Faust, 1994; Battiston et al., 2014), where no explicit inter-layer connections are considered. This is a specific case of multiplex networks, which can be generalised in order to consider inter-layer connections as well (Kivelä et al., 2014; Boccaletti et al., 2014; De Domenico et al., 2013). In the present MLN representation we ignore costs of jumping across layers, hence no inter-layer edges are considered. This modelling assumption leads to shortest paths over the multiplex structure coinciding with shortest paths over the aggregate network (De Domenico et al., 2014). Thus a shortest path is the network path connecting any two nodes within the smallest number of hops Newman (2010).

Shortest path lengths determine, together with clustering, the so-called small-world feature in single-layer networks (Watts & Strogatz, 1998). We follow established definitions of small-worldness from the literature about cognitive networks (Steyvers & Tenenbaum, 2005; Vitevitch, 2008; Beckage & Colunga, 2015) according to which a network is a small-world if, compared to random graphs of the same size, it exhibits a significantly higher mean clustering coefficient and a comparable mean shortest path length. In the field of linguistic networks, the small-world feature has been found in semantic networks (Motter et al., 2002; Sigman & Cecchi, 2002; i Cancho & Solé, 2001; Steyvers & Tenenbaum, 2005) and phonological networks (Vitevitch, 2008; Stella & Brede, 2015) as well. As evident from the comparison to configuration models in Table 1 from the main text all the layers of the MLN display the small-world feature. As also suggested in previous literature (Borge-Holthoefer & Arenas, 2010; Beckage et al., 2011; Goñi et al., 2011), we conjecture that small-worldness may be cognitively beneficial to language learning and use, as it might allow for efficient navigation within semantic memory (Goñi et al., 2011; Borge-Holthoefer & Arenas, 2010; Baronchelli et al., 2013). Furthermore, empirical evidence has shown that small-worldness is related to language learning in children (Beckage et al., 2011): semantic network lexicons of late talkers, who are likely to exhibit language processing difficulties, do show small-worldness to a much smaller degree compared to lexicons from children learning words at normative pace.

Small-worlds are defined through the average shortest path length and clustering coefficient. It is worth reminding that the mean clustering coefficient $CC \in [0, 1]$ measures how much neighbourhoods resemble complete graphs. The local clustering coefficient for node i with k_i neighbours and Γ_i edges between its neighbours is:

$$CC_i = \frac{2\Gamma_i}{k_i(k_i - 1)}. \tag{5.3}$$

As evident from the formula, disconnected nodes and nodes with only one neighbour have an ill-defined local clustering coefficient. In order to exclude them from the mean local clustering $CC = (\sum_i CC_i)/N$, we use the *deformed clustering coefficient* CC_c (Kaiser, 2008) (see Table 1 from the main text):

$$CC(def.) = \frac{\sum_i CC_i}{N(1-\theta)}, \tag{5.4}$$

where θ is the disconnectedness ratio (Kaiser, 2008), i.e. the fraction of nodes of degree 0 and 1 in the network which by definition cannot contribute to clustering. Previous work (Kaiser, 2008) highlighted the importance of considering the deformed clustering coefficient when evaluating small-worldness especially when considering small networks (i.e. networks composed of a few hundreds of nodes). This is the motivation for using the deformed clustering coefficient in this work.

The other network measures reported in Table 1 in the main text are: (i) the coefficient of assortative mixing by degree a, measuring degree correlations across network edges, (ii) the fraction of nodes within the largest connected component (*Conn.*) and (iii) the average shortest path length $\langle d \rangle$. Several independent studies identified single-layer linguistic networks as being significantly either more assortative (Vitevitch, 2008; Stella & Brede, 2015; Beckage et al., 2015) or more disassortative (Beckage et al., 2015) when compared to random graphs: the assortativity of mixing by degree thus quantifies aspects of structure in language networks that are not present in random graphs. From a cognitive perspective, assortative (disassortative) behaviour indicates the tendency for words having larger numbers of associates/meanings/phonological neighbours (not) to be connected with each other. In the phonological layer this pattern is the consequence of word length correlating with network degree (Stella & Brede, 2015, 2016a). We conjecture that assortative or disassortative mixing by degree on the semantic MLN layers might be related to a hypothetical navigability through words in the mental lexicon. A higher than random assortative coefficient would express the possibility of reaching a well connected word directly from another well connected word. Instead, disassortative degree mixing, as observed in the association layer, would indicate the possibility of reaching directly specific, less connected terms from broader, more well connected terms.

For a more detailed review of the above network metrics we refer the interested reader to Newman (2010).

Individual network layers are compared against configuration models, i.e. random graphs with the same degree sequence (and hence the same degree distribution) as the empirical networks. This choice of null model is appropriate, as it rules out effects of the degree sequence on network properties which is important since some of the empirical networks have heavy tailed degree distributions (see SI Fig. 5.5). Configuration models were sampled by randomising edges while maintaining degree sequences (Newman, 2010). The resulting network properties, such as clustering coefficient, were tested against analytical estimates (Newman, 2010). All comparisons of MLN layers in the main text are relative to these configuration models.

5.5.4 k-core analysis

The high deformed clustering coefficients (Kaiser, 2008), the relatively high assortativity coefficients, and the relatively flat cumulative degree distributions for degrees up to six (see SI

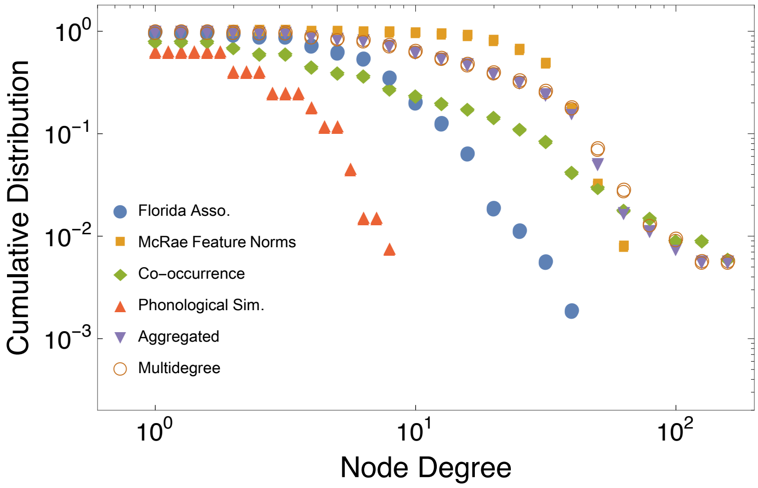


Figure 5.5: Cumulative degree distribution $P(K \geq k)$ of finding a node with individual layer degree or multidegree higher or equal to k within the MLN layers and the aggregate network.

Fig. 5.5) all suggest the presence of a core-periphery structure in the MLN layers. Notice that words in a densely connected core might have topological features significantly different from those of words in a poorly connected network periphery. Since we relate topological word features to word learning, it is useful to better assess the presence of strongly connected cores within the individual layers of the MLN. For this purpose we perform a k-core analysis (Newman, 2010). A k-core is a maximal subset of nodes such that each node in the set is connected to at least k others in the same subset. The "maximal" feature indicates that a group of nodes is a k-core if it is not a subset of any larger group that is a k-core. Hence, operatively k-cores can be obtained by repeatedly deleting all nodes of degree less than k in a given network.

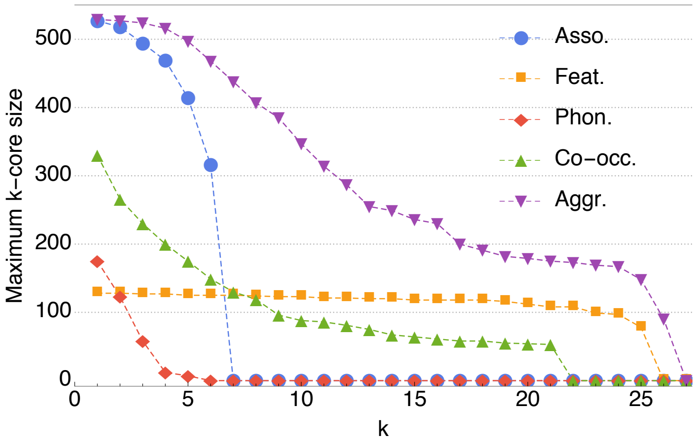


Figure 5.6: Size of the largest k-core for different values of k for the four layers and the aggregate network of the MLN.

In Supplementary Table S1 we list the number of k-cores, while the size of the largest k-core for each layer and the aggregated network are visualised in Supplementary Fig. S5.6. For instance, the phonological network has one largest connected component and 32 other connected components (i.e. k-cores with $k = 1$); these smaller components are also called *linguistic islands* (Vitevitch, 2008). A high number of linguistic islands, and no highly connected core seems

# k-cores \ k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Asso.	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Feat.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Co-occ.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Phon.	33	3	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aggr.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 5.2: Number of k-cores and maximum size of k-cores for k up to 20 in the 4 MLN layers and in the aggregate of the multiplex lexical network.

typical for phonological networks (Vitevitch, 2008; Stella & Brede, 2015). The association layer features a 6-core but no higher k-core. This is further indication that this layer is structurally different from the others, in that it does not display a marked core-periphery structure. In fact, from Table 1 of the main text, the association layer has a bigger largest connected component but comparable average connectivity when compared to the other MLN layers. On the other hand, the feature norms and the co-occurrence layers display k-cores up to $k = 20$, implying the presence of a very densely connected core (i.e. a marked core-periphery structure) in each one of these two layers. Notice that the core from the co-occurrence layer is smaller than the core of the feature norms layer and it also shrinks at a faster rate as k increases. The presence of a very densely connected core in the feature sharing layer is also indicated by the cumulative degree distribution reported in SI Fig. 5.5. In fact, for $k < 20$ the probability $P(X \geq k)$ of finding nodes with degree equal or higher to k stays nearly at the original value: this indicates the rarity of nodes having degrees $1 \leq k \leq 20$ in the network. Further analysis indicates that the feature norms layer displays up to a 25-core made up of 78 words. The presence of highly connected k-cores is evidence for a rich club effect (Newman, 2010), i.e. highly connected nodes tend to share links among themselves, particularly in the feature sharing layer. This result is in agreement with the high assortative mixing by degree exhibited by this layer and reported in Table 1 of the main text. Also the aggregated network displays k-cores but up to size $k = 26$ and considerably larger than those of the individual layers, again suggesting a rich club effect also when the whole edge-coloured multiplex structure is taken into account.

Interestingly words being part of densely connected k-cores are acquired earlier in the word trajectory based on the empirical age of acquisition. For instance, the 90 words in the 26-core of the aggregate network are ranked 30 ± 1 positions above the average position of ensembles of 90 randomly selected words in the empirical age of acquisition word ranking. Hence, when the normative age of acquisition ordering is considered, words in the densely connected core of the MLN structure are learned earlier than expected at random. This finding corroborates the idea of an interplay between topological features of words in the MLN and patterns in word acquisition. Further, since words in a densely connected core tend to be learned earlier but in general also display higher degree and higher closeness centrality (Newman, 2010), we focus on these two network features in the main text when exploring word acquisition patterns through ordering experiments.

5.5.5 Multiplex network metrics

Multiplex networks are convenient for considering multiple interactions within the same representation (Wasserman & Faust, 1994; De Domenico et al., 2015, 2016). Nonetheless, it is important to quantify whether considering all these interactions jointly is necessary or redundant in terms of topological patterns and interpretation of results. Structural reducibility (De Domenico et al., 2015) investigates whether the multiplex paradigm is a suitable model, identifying the presence and extent of redundant topological patterns. In order to assess the benefit of the multiplex representation, we adopt the greedy procedure suggested in (De Domenico et al., 2015) and implemented in *muxViz* (De Domenico et al., 2014). This structural reducibility analysis relies on: (i) identifying topologically similar layers, (ii) aggregating layers if appropriate, and (iii) comparing the richness in topological patterns of the aggregated multiplex layers against the aggregated network, obtained by projecting all the edges in the multi-layer structure on a single-layer network.

The procedure identifies similarity of layers and quantifies how distinguishable the multiplex is against aggregate versions of two or more layers. The whole procedure is based on the Von-Neumann entropy of each multiplex layer De Domenico et al. (2015). This measure is used for determining the richness of topological patterns of each multiplex network aggregation. Let us consider a multiplex network with M layers. In a given aggregation stage σ some of the layers might be aggregated, so that the multiplex network might count $X \leq M$ layers. A relative entropy function q of an aggregation stage σ where X multiplex layers are distinct is computed as:

$$q(\sigma) = 1 - \frac{\sum_{\alpha}^{X} h_\alpha}{X h_A}, \tag{5.5}$$

where h_α is the Von Neumann entropy of layer α while h_A is the Von Neumann entropy of the aggregated network. The higher the relative entropy q the more distinguishable the multiplex at the aggregation stage a is as compared to the aggregate network. The maximum value of q identifies the aggregation stage that is the most distinguishable compared to the aggregate network. In Fig. 1 (e) of the main text we show that for the MLN the maximum of q is reached when all the layer are kept as separate and no aggregation takes place. This means that the MLN is irreducible, i.e. aggregating any of its layers decreases the information on the topological patterns encapsulated in the multiplex structure. For the mathematical details behind the structural reducibility analysis we refer to De Domenico et al. (2015).

In the performed ranking experiments we adopt single-layer network measures such as degree, closeness and betweenness centralities and their multiplex counterparts. Here we give a brief technical overview over these measures. The *multidegree* (De Domenico et al., 2013) or *overlapping degree* (Bianconi, 2013; Boccaletti et al., 2014) m_i of node i is defined as the sum of all the degrees of the node replicas across the $M = 4$ layers of the multiplex:

$$m_i = \sum_{\alpha=1}^{M} k_i^{\alpha}, \tag{5.6}$$

where $k_i^{(\alpha)}$ is the degree of node i on layer α. The multidegree provides partial information about the connectivity of a node across the multiplex structure and it is one of the simplest and most frequently used measures in the relevant literature (Boccaletti et al., 2014; Battiston et al., 2014; De Domenico et al., 2013).

In single-layer networks, the *closeness centrality* c_i measures the distance of node i from all other nodes in the network:

$$c_i = \frac{1}{\sum_j d_{ij}}, \tag{5.7}$$

where d_{ij} is the shortest distance between nodes i and j. Disconnected nodes have distance infinity and thus do not contribute to the sum. Closeness relates to how fast information is expected to spread from a given node to others in the network. In the main text we consider the generalisation of closeness centrality where multiplex shortest paths are considered (i.e. we allow for "jumps" across layers in order to quantify how close a node is to another one). Notice that this multiplex version of closeness centrality corresponds to the closeness centrality on the aggregate network. In a more general setting one could consider costs for jumps between layers and might wish to introduce more general closeness centrality measures on multiplex networks (cf. De Domenico et al. (2014)). Our choice of identifying multiplex closeness with the closeness of the aggregate is motivated by simplicity of the modelling approach and difficulties of identifying costs of transitioning between layers. This weight of transitions between layers may be related to cognitive functioning but it is a methodologically open question in the literature (Aitchison, 2012; Zock, 2015). Notice that closeness centrality is a problematic measure on disconnected networks (Newman, 2010). In our case closeness centrality is defined on a fully connected multiplex network, so that biases due to disconnectedness are not present when all the layers are considered at once.

We also consider multiplex shortest paths in the multiplex version of the betweenness centrality b_i which quantifies the extent to which a node lies on the shortest paths between other nodes. For a mathematical definition we refer to Newman (2010). In SI Sect. 5.5.6 and 5.5.7 we report on the performance of word betweenness for predicting word learning.

Another measure tested in our ordering experiments is the PageRank centrality, which represents the likelihood that a random walker arrives at any particular node navigating a given network by hopping through links. Rather recently *versatile Page-Rank* (De Domenico et al., 2015) was introduced as a multiplex generalisation of PageRank, where random walkers can also transition across layers. For the mathematical details behind the formulation of the measure, we refer to De Domenico et al. (2013) and De Domenico et al. (2015). In SI Sect. 5.5.6 we report on the performance of PageRank for predicting word learning.

5.5.6 Random word guessing

To evaluate predictions of word acquisition we compare against the reference model of randomly guessing the words to be learned, i.e. the baseline of random permutations of the empirical data set for the word acquisition trajectory. As each word has the same probability of occurring at each position, the probability that a word is correctly guessed as having been learned by time t is $p_w = t/N$. Hence, the average number of correctly guessed words at time t is $\langle n_t \rangle = t^2/N$. Notice that the number of randomly guessed words $\langle n_t \rangle$ approaches the total number of words for larger t, i.e. the word gain measure (see Methods from the main text) becomes less sensitive at later learning stages. The standard deviation σ_t for random word guessing is:

$$\sigma_t = t p_w (1 - p_w) = \frac{t^2(N-t)}{N^2}. \tag{5.8}$$

σ_t is one of the three sources of variation affecting ordering experiments presented in the Results section in the main text. The second source is related to shuffling words tied in ranking positions and the third is related to the probabilistic reshuffling of words learned at each month according to normative CDI data. These sources of variation led to error margins on the vocabulary normalised word gains of the order of magnitude of $\approx 10^{-3}$, which are roughly the size of dots reported in Figure 2 from the main text and thus not displayed in that plot for clarity.

Notice that the number of randomly guessed words $\langle n_t \rangle$ represents the expected overlap of random orderings with the normative age of acquisition orderings. In the main text we use $\langle n_t \rangle$ as a reference value for defining both the vocabulary normalised word gain (see the Methods section) and the Z-scores of the overlaps observed through our ordering experiments. We approximate the binomial distribution relative to $\langle n_t \rangle$ and σ_t according to the commonly used constraint that $p_w t$ and $(1 - p_w)t$ have to be both greater than 5 (Grimmett & Stirzaker, 2001). This implies that we can approximate $\langle n_t \rangle$ and σ_t as coming from a Gaussian distribution when at least $t = 60$ words have been acquired. When valid, this approximation allows for us to use the Z-score statistical testing at a significance level of 5% when at least 60 words have been acquired. The black dashed line in Fig. 2 (b) from the main text marks this region and we adopt it also for indicating the end of the VELS.

5.5.7 Tested word orderings and word gains

In the main text we only reported detailed results for word rankings based on a subset of all measures explored. Additional details are given in Supplementary Fig. S5.7 for word gains for orderings based on degree, word frequency, closeness, word length, betweenness and PageRank. Average word gains are reported on the right end of the plot for each ordering. Notice that the highest word gain is relative to the ordering based on the association degree, which correctly predicts an average of about 19 words as learned on top of the expectation from random guessing. The maximum word gain obtained by the association degree is 42 ± 1 words.

As reported in the main text, adding degrees across different layers together does not guarantee better word gains. In fact the multidegree reaches a lower maximum word gain compared to the association degree only (34 ± 1 vs 42 ± 1). Error bounds are estimated over different normative age of acquisition orderings.

In the main text we tested the ordering based on word frequency computed from the child-directed speech data of the CHILDES dataset (MacWhinney, 2000). However we computed word frequency on an additional, independently obtained dataset, namely the Opensubtitle dataset (Barbaresi, 2014), which is based on TV subtitles. As reported in Supplementary Fig. S5.7, the Freq. (Adults) ordering based on Opensubtitle performs much worse (average word gain of 0 ± 2) compared to the frequency counting of the CHILDES dataset (average word gain of 17 ± 2). We conjecture this large gap in performances is due to the different nature of the two frequency datasets. While the CHILDES data set is relative to children speech, the Opensubtitle one is based mainly on TV-series written by and targeted to adults. Unsurprisingly, the lack of predictive power of the adult frequencies suggests that frequency counting from adults is a poor estimator of the word learning dynamics in children.

Word betweenness of the individual layers does perform worse than degree on all the MLN layers. This is not surprising as in all but the association layer several nodes are disconnected and hence have betweenness 0. However, when inter-layer paths are considered, the betweenness centrality computed on the multiplex structure performs noticeably better, with a maximum word gain of 39 ± 2. This finding is in agreement with the structural reducibility analysis from the main text: it supports the importance of considering the whole multiplex structure when investigating patterns among words in the modelled mental lexicon.

Single-layer PageRank for the association layer gives performances close to the degree. This is not a surprise, as PageRank and degree are correlated with each other (Newman, 2010). However, differently from what happens with the degree, the multiplex version of PageRank for the whole MLN provides predictive power similar to the best single-layer counterpart (i.e. the PageRank in the association layer). This indicates that random walks on the whole MLN are similar to random walks on the association layer, while hub nodes in the association layer might be very different from hub nodes in other layers, so that multidegree and degree in the association layer differ substantially. From a cognitive perspective, the relatively good performance of PageRank confirms previous results about the importance of this measure in exploring the mental lexicon (Griffiths et al., 2007).

The word gain was used in the main text for defining the vocabulary normalised word gains (see the Methods section from the main text). A vocabulary normalised word gain of X has to be interpreted as the percentage of words correctly guessed as learned by a given ordering in the whole vocabulary, in addition to those that could be learned at random. For instance, when the vocabulary is made of 100 words, an ordering identifies 24 but 4 would be guessed at random, then the vocabulary normalised word gain would be 0.2 or 20%. Supplementary Tab. S2 reports the average relative word gains for many of the tested orderings.

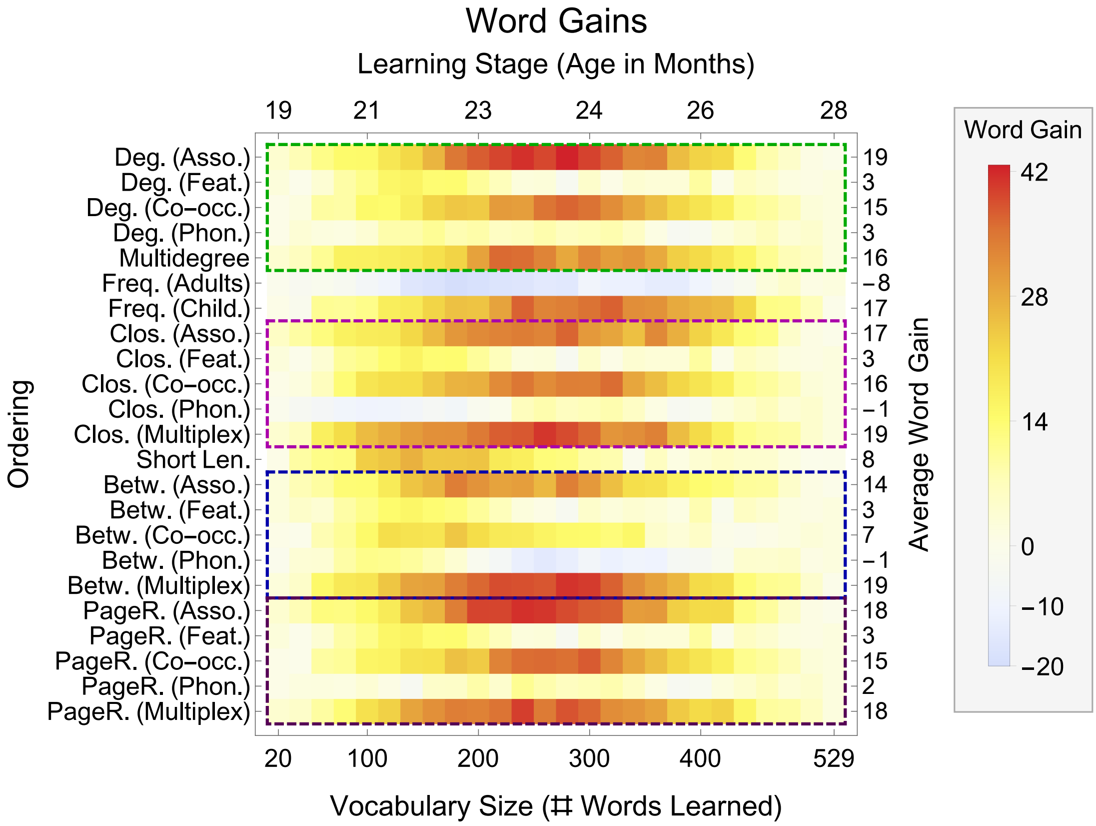


Figure 5.7: Word gains for different orderings at different learning stages. A word gain of 42 means that 42 words were correctly guessed as learned in addition to those that would have been guessed at random. Hence, the gain is considered over random guessing. Word gains averages over the whole learning trajectories are reported on the right. Word orderings are clustered according to their type (e.g. degree based, closeness based, betweenness based, PageRank based). Orderings based on exogenous features are presented as well: Freq. (Adults) comes from the Opensubtitle dataset, Freq. (Child.) comes from the CHILDES dataset, Short. Len. is based on word length.

Ordering	VELS	ELS	LLS
Deg. (Asso.)	0.077	0.13	0.089
Deg. (Feat.)	0.02	0.063	0.01
Deg. (Co–occ.)	0.008	0.089	0.069
Deg. (Phon.)	0.016	0.027	0.011
Multidegree	0.07	0.114	0.073
Freq. (Adults)	−0.006	−0.028	0.003
Freq. (Child.)	−0.006	0.091	0.078
Clos. (Asso.)	0.104	0.138	0.075
Clos. (Feat.)	0.02	0.064	0.01
Clos. (Co–occ.)	0.008	0.113	0.072
Clos. (Phon.)	−0.042	−0.063	0.005
Clos. (Multiplex)	0.07	0.191	0.083
Short Len.	0.045	0.143	0.027
Betw. (Asso.)	0.057	0.122	0.062
Betw. (Feat.)	0.042	0.065	0.009
Betw. (Co–occ.)	0.	0.095	0.03
Betw. (Phon.)	−0.005	0.046	−0.014
Betw. (Multiplex)	0.042	0.163	0.087
PageR. (Asso.)	0.077	0.129	0.087
PageR. (Feat.)	0.02	0.063	0.01
PageR. (Co–occ.)	0.	0.095	0.072
PageR. (Phon.)	0.008	0.009	0.011
PageR. (Multiplex)	0.062	0.144	0.082

Table 5.3: Average vocabulary normalised word gains for different orderings at different learning stages, i.e. the Very Early Learning Stage (VELS), the Early Learning Stage (ELS), the Late Learning Stage (LLS) and the Late Learning Stage (LLS). A vocabulary normalised word gain of 0.112 means that 11.2% words in the vocabulary were correctly guessed as learned by excluding effects due to random guessing. The gain is therefore over random guessing. Word orderings are clustered according to their type (e.g. degree based, closeness based, betweenness based, PageRank based). Orderings based on exogenous features are presented as well: Freq. (Adults) comes from the Opensubtitle dataset, Freq. (Child.) comes from the CDI data, Short. Len. is based on word length.

5.5.8 Ordering Experiments for the Phonological Layer

Supplementary Fig. S5.7 and Tab. S2 both demonstrate that the phonological layer does not perform well in terms of predicting the order in which words are learned. In Supplementary Fig. S5.8 we report the word gain Z-scores (see Methods from the main text) for the two estimators that work best in our analysis, namely word degree and closeness.

When we consider the phonological layer as isolated from the other MLN layers and rank the words according to their degree (closeness) in the phonological layer we obtain orderings whose word gains are compatible with random fluctuations in terms of 2 standard deviations over almost the whole learning trajectory. Hence, within a confidence interval of 95% we can say that degree (closeness) contributions coming from the phonological layer are compatible with random fluctuations. This is why the phonological layer does not perform well in the ordering experiments and in the optimisation experiments as well (see main text).

From a cognitive perspective our technical finding suggests that young toddlers do not use phonological similarities for boosting the likelihood of learning specific new words. This result is supported from recent empirical evidence with children of 18 months of age (Dautriche et al., 2015). As discussed also in the main text, different measures for capturing how similar sounding words are processed by children are required.

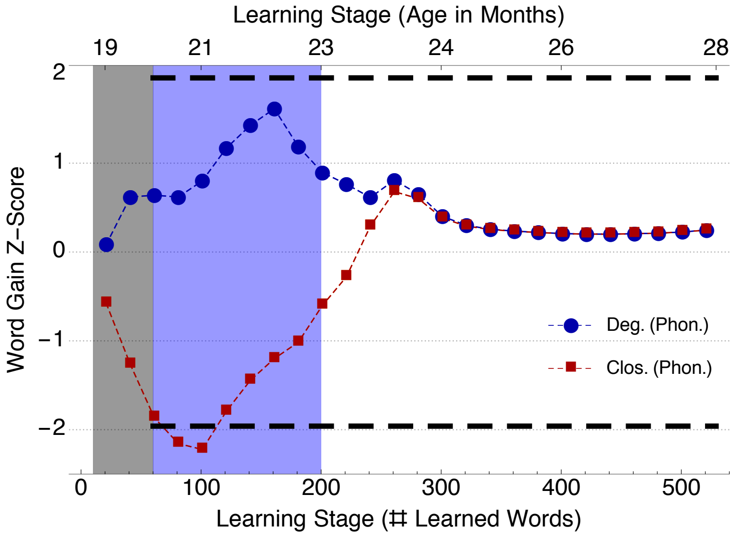


Figure 5.8: Word Gain Z-Score for orderings based on the phonological layer as separate from the other MLN layers. Words are ranked according to either their degree, Deg. (Phon.), or to their closeness, Clos. (Phon.). More connected and closer words are ranked higher. Almost all the retrieved Z Scores are compatible with overlaps provided by random guessing, within the domain of Z score $-1.96 \leq Z \leq 1.96$. Notice that after roughly 300 words are learned the two orderings provide the same average performances because the phonological layer features 268 disconnected words, i.e. words having degree 0 and closeness 1. When ranked, those words for which there is missing topological information become a tie and they are randomised, so that the ranking on that tie is identical to random guessing. We believe this is the technical reason why the phonological layer does not perform well in the optimisation procedure as well.

5.5.9 An Expanded Phonological Layer

Previous literature has found indications that phonological neighbourhood size (i.e. degree in a phonological network) has an impact on word acquisition in young toddles up to 30 months of age (Carlson et al., 2014). We refer the interested reader to (Carlson et al., 2014) also for a brief review of other works correlating phonology and lexical development in older children.

Based on the optimisation experiments reported in the main text, we cannot interpret the lack of predictive power of the particular phonological layer used in the MLN as a general lack of influence of phonological similarities on lexical acquisition. Instead, this is an indication that the phonological network induced from the 529 words considered in the study is too small (and thus too fragmented) for providing meaningful information about word acquisition. This finding is not in contrast with the irreducibility analysis: the phonological layer appears to encapsulate different edge patterns compared to other layers. Nonetheless, further analysis with ordering experiments (cf. the previous SI subsection) reveals that these patterns are not more predictive than random guessing for lexical acquisition. In order to test whether the poorly connected topology of the phonological layer is the source of this lack of predictive power we carried out additional experiments explained below.

We took inspiration from the work of Carlson et al. (2014), where the authors used a phonological network of 12000 words from adult caregivers for testing the influence of phonology on toddlers. Similarly, we considered the larger phonological network from adult native English speakers, already analysed in (Stella & Brede, 2015) and including almost 30000 English words. Detailed analysis of the network topology is provided in (Stella & Brede, 2015). The phonological layer of the multiplex analysed in the main Chapter is a subset of this network. This allows us to evaluate phonological degrees and closeness values of the same 529 words from the MLN on the "extended" topology of the adults' phonological similarities. Based on these two scores we performed additional ordering experiments, which are reported in SI Fig. 5.9.

We find that the word gains obtained from using either degree or closeness on the extended phonological layer are statistically significant within a 97.5% confidence level during ELS. This indicates that *phonology does indeed play a role on word acquisition rather early on*, compatibly with previous findings from the literature (Carlson et al., 2014). However, this effect reduces over time and word gains become compatible with random word guessing after 400 words have been learned. A similar result was found in (Carlson et al., 2014), where the influence of phonological word degree on word acquisition vanished around the 30th month of age.

The presence of a statistically significant influence of phonology over word learning motivated us to use the word centralities in the extended phonological network for our optimisation experiments. Results are reported both in the main text (cf. Figure 3) and in SI Sect. 5.5.7. Additional analysis relative to the original phonological layer is reported in SI Sect. 5.5.8.

5.5.10 Percentage Word Gains

In order to accompany the results from Fig. 2 of the main text, we introduce also the percentage word gain $P_O(\tau, \omega, t)$, as the difference of word overlaps $O_{aoa}(\tau, t)$ and $O_{aoa}(\omega, t)$ of orderings τ and ω, respectively, normalised by $O_{aoa}(\omega, t)$. We always consider overlaps with the normative

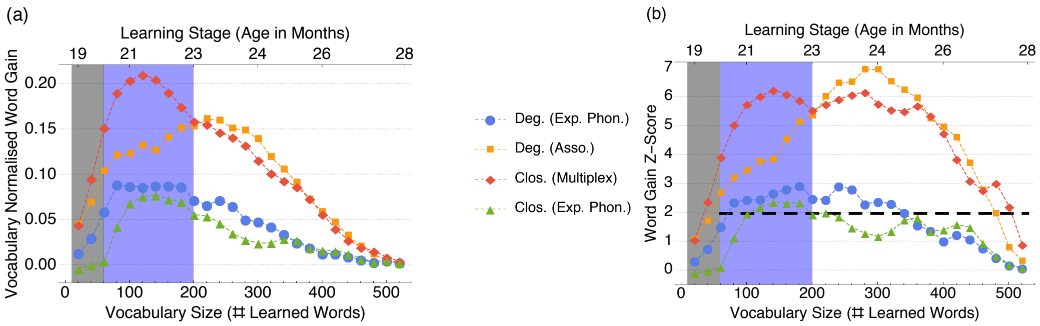


Figure 5.9: Vocabulary normalised word gains (a) and word gain Z-Score (b) for orderings based on the expanded phonological layer (Exp. Phon.) as separate from the other MLN layers. Orderings based on the original MLN association degree and multiplex closeness are reported as well for comparison. The dashed black line in (b) indicates Z-scores equal to 1.96. Both the extended phonological orderings provide predictions above this threshold and hence are statistically significant within a 97.5% confidence level.

age of acquisition orderings. Notice that when ω is a random ordering, then $O_{aoa}(\omega, t) = \langle n_t \rangle$ at time t. The choice of omega can be general. In formulas:

$$P_O(\tau, \omega, t) = \frac{O_{aoa}(\tau, t) - O_{aoa}(\omega, t)}{O_{aoa}(\omega, t)}. \tag{5.9}$$

The percentage word overlap quantifies the percentage of words that a given ordering provides in comparison to random guessing. For instance, a $P(\tau, \omega, 100) = 50\%$ indicates that when 100 words have been learned, the ordering τ correctly guesses as acquired 50% words more than ordering ω. Results for the case when ω is random ordering are reported in Supplementary Fig. S5.10. Error bars are estimated from standard deviations of the respective orderings. In VELS and ELS, except for the first point (i.e. when 20 words have been acquired) all our tested orderings display a percentage word gain that is several standard deviations away from the reference value of 0 (Sign Test p-value $< 10^{-5}$ for all the orderings). The highest percentage word gain is the one provided by multiplex closeness centrality, which predicts up to 160% more words compared to random guessing. A percentage word gain against random ordering compatible with zero would imply that the orderings would give improvements compatible with random fluctuations. This implies that all the word gains observed by ranking words according to network features or frequency are statistically not compatible with random fluctuations when at least 40 words are learned. This finding supports results coming from the Z-scores presented in the main text. Notice that while P_O allows for a statistical assessment of the significance of word gains in VELS (something that Z-scores cannot do), the percentage overlap P_O suffers from approaching the 0 reference level for all the orderings at the middle of LLS. In this phase, the Z-scores provide a clearer picture of the predictability of different orderings (see Fig. 2 in the main text). Notice that the Z-scores provide a more robust statistical testing of our findings.

Notice that we consider also another case when ω is the association degree ordering and results are reported in Supplementary Fig. S5.11. The multiplex closeness performs up to 25% better in ELS than association degree and the difference in performance between these two orderings is several standard deviations away from 0 (Sign Test p-value $< 10^{-5}$ for all the orderings). The unmatched predictive power of closeness centrality is what marks the whole early learning stage. At later stages, during LLS, the multiplex closeness and the association degree perform similarly. Interestingly, frequency performs worse than either multiplex closeness or association degree during early learning stages. Word length and multiplex PageRank perform only up to 10% better than association degree in ELS. The percentage word overlap confirms the claims from the main text that multiplex closeness outmatches all the other considered orderings during early learning stages.

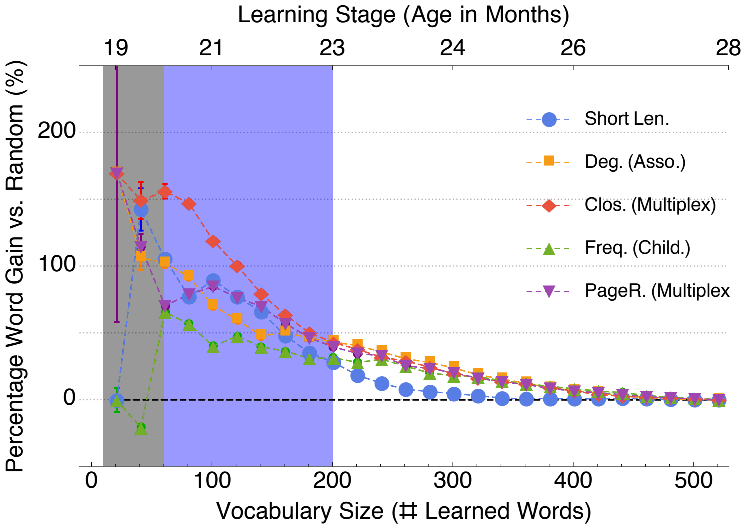


Figure 5.10: Percentage word gains against random guessing for the word orderings reproduced in Figure 2 of the main text. A percentage word gain of 100% at a given learning stage means that a word ordering leads to guessing 100% words more compared to random guessing. Error bars are based on standard deviations. All the orderings in VELS provide percentage word gains that are clearly away from 0 (Sign Test p-value $< 10^{-5}$ for all the orderings).

5.5.11 Influence of taxonomic relationships on word gains

Many English words refer to categories that are taxonomically organised, e.g. "horse" is a type of "animal". This taxonomic organisation results into basic, super-ordinate and sub-ordinate level object categories. Super-ordinate categories have broader semantic fields than sub-ordinate level categories. This taxonomic organisation can influence the semantic relationships giving rise to free associations, feature sharing norms and co-occurrences we represented in our multiplex lexical network.

Hence, explicitly quantifying how word features in individual multiplex layers correlate with a superordinate and subordinate categorisation of words in the children's vocabulary might be of relevance for understanding the mechanisms behind word acquisition in young children.

We focus here on quantifying how the taxonomic word organisation acts as a mediator variable correlating with the two most predictive word features in our framework: degree of words in the free associations and the closeness of words on the multiplex structure.

We retrieved from WordNet 3.0 (the version curated by Wolfram Research) a network of hyponymy relationships, where nodes represent words and the directed edge $A \to B$ means that A is a type of B. This hyponymy network included roughly 350000 English words. The

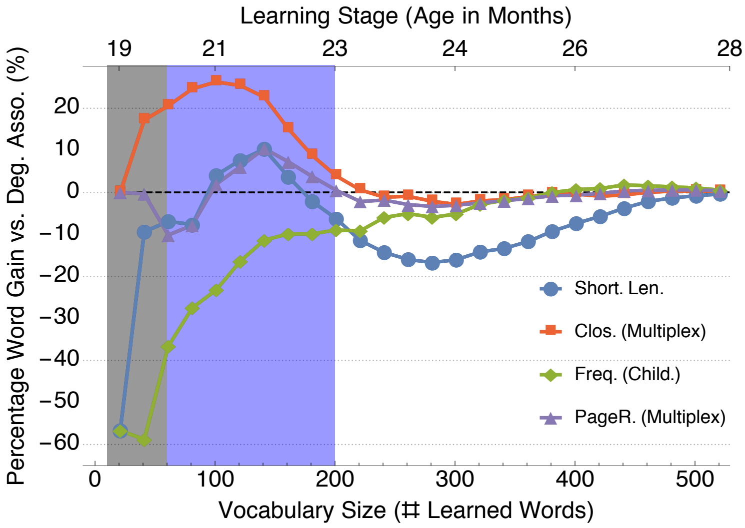


Figure 5.11: Percentage word gains against the association degree ordering for the word orderings reproduced in Figure 3 of the main text. A percentage word gain of 20% at a given learning stage means that a word ordering leads to guessing 20% words more compared to association degree. Error bars are based on standard deviations and are the same size of the dots. The peak in ELS for the multiplex closeness is clearly incompatible with a 0 difference (Sign Test p-value $< 10^{-5}$ for all the orderings).

in-degree of a node is defined as the number of directed edges pointing to that node. In this hyponym network words with higher in-degree represent broader words, as they have more words that fall into their semantic category. For instance, considering the edges "*pigeon → bird*" and "*dove → bird*", then bird would have indegree 2 while pigeon and dove would have indegree 0. We would then consider "bird" as being a broader term than either "pigeon" or "dove".

We assume that in-degree on the whole hyponym network is a good proxy of the mediator variable that distinguishes super-ordinate words (high indegree) from sub-ordinate (low indegree) in the English language.

We then correlated multiplex closeness and association degree of MLN words with their in-degree in the whole hyponymy network. We used the Kendall Tau to quantify correlations. Results indicate that the MLN association degree correlates with in-degree almost three times more compared to multiplex closeness (Kendall Tau of association degree ≈ 0.20, p-value $< 10^{-5}$ vs Kendall Tau of multiplex closeness ≈ 0.07, p-value < 0.01).

The different Kendall Taus indicate that the association degree is a better proxy for detecting superordinate words (i.e. words with higher indegree in our case) compared to closeness centrality.

As discussed in the main text, we conjecture that the association degree does not perform well in VELS and ELS because it would tend to guess more super-ordinate words than multiplex closeness would. Norming studies suggest (Hills et al., 2009) that children tend to learn general level categories before super- or subordinate classes (e.g. "dog" is learned before "animal", "chair" before furniture), a fact that is not captured well by the association degree ordering.

This is compatible with what we observe in our optimisation experiments (see the new Fig. 3, panels (c) and (d)), where the association layer contributes the most to optimised orderings using both degree and closeness but only outside of VELS. This network pattern suggests the emergence of a boosting effect in learning the generalisations of words previously acquired during VELS and ELS. These generalisations represent super-ordinate words that are captured predominantly by the association layer and not by the others (otherwise we would expect for all the layers to keep the same influence after VELS and ELS).

5.5.12 Optimisation Experiments

In Supplementary Tab. S3 we report the average vocabulary normalised word gains for the optimal orderings obtained by combining degrees and closeness centralities. Further, in Supplementary Fig. S5.12 we report the word gain Z-scores for the optimal orderings obtained by combining degrees and closeness centralities. These results complement the visualisation in panels (a) and (b) of Fig. 3 in the main text. In combination they demonstrate that not even the optimal combination of word degrees can perform better than multiplex closeness in terms of word predictability. Also, differently from the ordering experiments, for the optimised trajectories the Z-scores at the end of VELS are now clearly statistically significant. Ternary plots in panels (a) and (b) in SI Fig. S10 visualise the optimisation landscape of vocabulary normalised word gains over the remaining three layers at the end of VELS, middle of ELS, and middle of LLS. The region of optimal word gains tends to narrow and changes its location in the landscape over time, corroborating the idea that different linguistic information plays different roles at different stages of development.

Very similar results from SI Fig. S9 were obtained for the MLN with the extended phonological layer, for which the relative word gains are reported in SI Tab. S3 while the optimisation results are reported in SI Fig. S11.

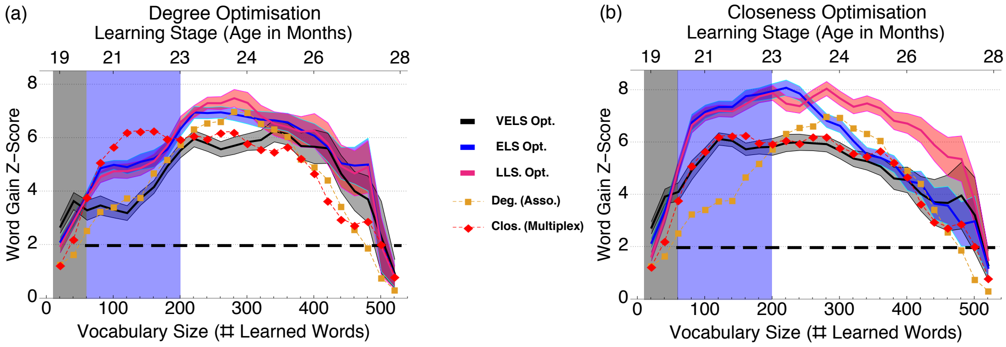


Figure 5.12: Word gain Z-scores for: (a) the optimal orderings of degree and (b) for closeness. The ordering experiments based on multiplex closeness and on association degree are reported as dots. The dashed black line indicates the range of words when the Z-Scores can be approximated to normal and it identifies the 95% confidence level (Z=1.96). Error bars are based on standard deviations from the optimised ensembles of word trajectories. In panel (a), the peak in ELS for the multiplex closeness is clearly larger than the peak for optimal combinations of word degrees (Sign Test p-value $< 10^{-5}$).

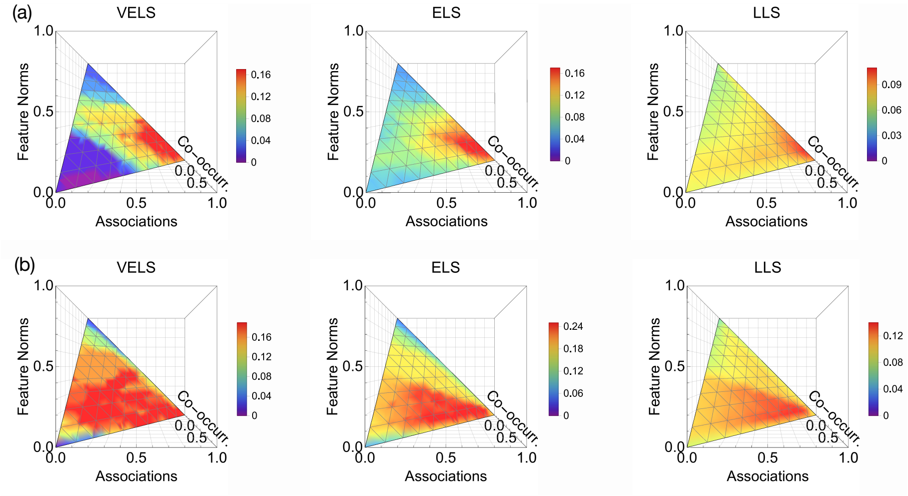


Figure 5.13: Ternary plots of the average relative word gains at the end of VELS, the middle of ELS and the middle of LLS for degree (a) and closeness (b) optimisation. Optimisation results relative to these ternary plots are reported in Figure 3 of the main text.

5.5.13 Linear optimisation of betweenness and local clustering

Since betweenness on the whole MLN performs slightly worse than association degree, as reported in SI Sect. 5.5.6, we also used the betweenness centralities of words within the MLN layers as a basis for scores to calculate optimal combinations of layers. We considered local deformed clustering as well, as defined in SI Sect. 5.5.3. We chose these two additional network features because they utilise the multiplex MLN structure at global and local levels.

When betweenness was used, the retrieved optimal layer importances did not change over the VELS, ELS or LLS stages. Therefore only one relative word gain curve is reported in Supplementary Tab. S3. Predictability results coming from the optimal linear combination of betweenness centralities are overall inferior to the ordering results with multiplex closeness and the single-layer association degree. Betweenness optimisation leads to similar results compared to the ordering experiments with multiplex betweenness (see Supplementary Tab. S2). The fact that betweenness provides lower predictability power than closeness on the whole MLN suggests that early learned words have higher closeness but lower betweenness. Hence they are words closer to others on the MLN structure but they are not necessarily part of many shortest paths as the MLN topology offers different short-cuts for navigating through words.

In the main text we reported the results of optimisation procedures based (i) on a local network statistics such as degree and (ii) on a global network statistics such as closeness centrality. Between the global and local extremes, we also investigated the optimisation of a second-order local network statistics such as the local clustering coefficient. Predictability results are reported in Supplementary Tab. S3. The optimisation over local clustering performs worse than the empirical multiplex closeness centrality and the linear optimisation results for both degree and closeness. This suggests further that the multiplex structure is supportive in capturing normative language learning trajectories of young children.

The inferior predictability performances of local clustering and betweenness, particularly in ELS, support the idea reported in the main text that closeness centrality, and therefore network distances, capture word correlations that are fundamental and relevant to normative word learning in young children. The idea that network distances are relevant for the cognitive processes regulating the ML is supported by previous experimental evidence (Collins & Quillian, 1969; Collins & Loftus, 1975; Motter et al., 2002).

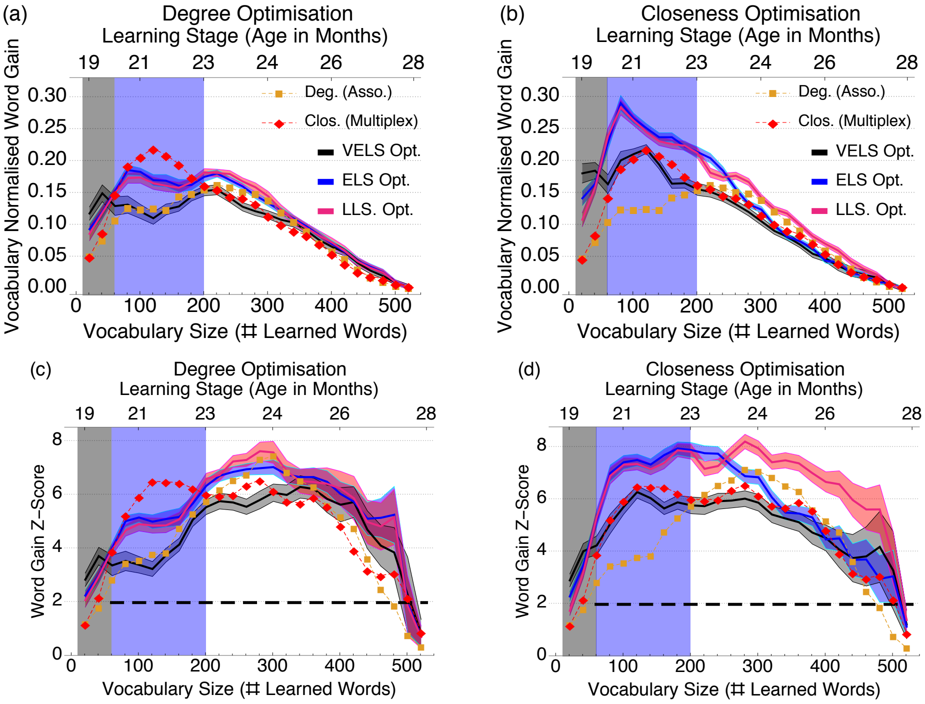


Figure 5.14: Optimisation results for degree (a) and closeness (b) in the networked lexicon with the extended phonological layer. Word gain Z-scores for (a) the optimal orderings of degree and (b) for closeness in the networked lexicon with the extended phonological layer. The ordering experiments based on multiplex closeness and on association degree are reported as dots. The dashed black line indicates the range of words when the Z-Scores can be approximated to normal and it identifies the 95% confidence level (Z=1.96). Error bars are based on standard deviations from the optimised ensembles of word trajectories.

5.5.14 Non-linear optimisations

Also non-linear combinations of word scores across layers were explored and investigated. In particular, we tested convex combinations of the type:

$$s_w = \alpha \frac{(s_w^{(Asso)})^a}{\sum_w (s_w^{(Asso)})^a} + \beta \frac{(s_w^{(Feat)})^b}{\sum_w (s_w^{(Feat)})^b} + \gamma \frac{(s_w^{(Co-occ)})^c}{\sum_w (s_w^{(Co-occ)})^c} + (1-\alpha-\beta-\gamma) \frac{(s_w^{(Phon)})^{3-a-b-c}}{\sum_w (s_w^{(Phon)})^{3-a-b-c}}, \tag{5.10}$$

where the coefficients $\alpha, ..., \gamma$ give layer influences, and the exponents $a, ..., c$ allow for non-linear influences of the various layers. Results are reported in Supplementary Tab. S3; we note that including non-linear influences only gives marginal corrections to the linear model explored in the main Chapter. Importantly, the main findings about the particular role of closeness centrality remains unchanged.

Ordering	VELS	ELS	LLS	Total
Deg. (Asso.)	0.08(1)	0.13(1)	0.09(1)	0.09(1)
Clos. (Multiplex)	0.07(1)	0.19(1)	0.08(1)	0.10(1)
Opt. Deg. (VELS)	0.16(4)	0.12(2)	0.08(1)	0.09(1)
Opt. Deg. (ELS)	0.14(2)	0.17(2)	0.09(1)	0.11(1)
Opt. Deg. (LLS)	0.14(2)	0.17(2)	0.08(1)	0.11(1)
Opt. Clos. (VELS)	0.19(3)	0.19(2)	0.07(1)	0.12(2)
Opt. Clos. (ELS)	0.16(2)	0.25(2)	0.09(1)	0.13(1)
Opt. Clos. (LLS)	0.14(2)	0.24(2)	0.10(1)	0.14(1)
Opt. Deg. Ext. (VELS)	0.17(4)	0.12(2)	0.09(1)	0.10(1)
Opt. Deg. Ext. (ELS)	0.15(2)	0.17(2)	0.10(1)	0.12(1)
Opt. Deg. Ext. (LLS)	0.15(2)	0.17(2)	0.09(1)	0.11(1)
Opt. Clos. Ext. (VELS)	0.19(3)	0.20(2)	0.08(1)	0.13(2)
Opt. Clos. Ext. (ELS)	0.16(2)	0.26(2)	0.09(1)	0.13(1)
Opt. Clos. Ext. (LLS)	0.14(2)	0.25(2)	0.10(1)	0.14(1)
Opt. Betweenness	0.11(2)	0.15(2)	0.08(1)	0.09(1)
Opt. Loc. Clus. (VELS)	0.10(2)	0.13(2)	0.06(2)	0.08(2)
Opt. Loc. Clus. (ELS)	0.10(2)	0.16(2)	0.08(2)	0.09(1)
Opt. Loc. Clus. (LLS)	0.10(2)	0.16(2)	0.07(2)	0.09(1)
Nonl. Opt. Deg. (VELS)	0.18(2)	0.13(2)	0.09(1)	0.10(1)
Nonl. Opt. Deg. (ELS)	0.15(2)	0.17(2)	0.09(1)	0.11(1)
Nonl. Opt. Deg. (LLS)	0.13(2)	0.17(2)	0.10(1)	0.11(1)
Nonl. Opt. Clos. (VELS)	0.19(3)	0.20(2)	0.08(1)	0.12(2)
Nonl. Opt. Clos. (ELS)	0.16(2)	0.25(2)	0.10(1)	0.13(1)
Nonl. Opt. Clos. (LLS)	0.15(2)	0.24(2)	0.11(1)	0.15(1)

Table 5.4: Optimised vocabulary normalised average word gains for the empirical degree in the association layer and multiplex closeness centrality (top), the optimal degrees and closeness for the whole MLN (middle - Opt.), the optimal degrees and closeness with the extended phonological layer (middle - Ext.), the optimal betweenness and local clustering linear combinations (middle) and the non-linear combinations of degree and closeness centralities, respectively (bottom). Entries are averaged while: 40 words have been learned in VELS, 160 words have been learned in ELS, when 369 have been learned in LLS and when all the 529 words have been learned (Total). Parentheses indicate error bounds as standard deviations: 0.14(2) indicates 0.14 ± 0.02.

5.5.15 Shuffling Models

5.5.15.1 Random shuffling models for the optimisation procedure

In this section we test how correlations of closeness and degree centralities among layers in the MLN influence results of optimisation procedure. In particular, we aim to test of prediction accuracies of the same quality as reported in the main text could be obtained from (multiplex) networks in which correlations have been destroyed. We consider null models for multiplex structure by shuffling the multiplex network, i.e.:

1. A multiplex network made of layers having an unchanged topology but randomly shuffled node labels. We call this the Global Label-Shuffled null model (GLS), and essentially for each network metric every word receives a word score randomly chosen from the word scores of the original empirical multiplex network. This gives word scores $s_w^{(GLS)}{=}s_{\Pi}(w)$ where Π indicates a random permutation of word labels.
2. A multiplex network preserving the topology of the empirical layers but with re-shuffled node labels in each layer. We call this the Independent Label-Shuffled null model (ILS). In this model associations of words between layers are randomised, i.e. replicas of a node on different layers can be associated with different topological identities of words in the original multiplex network. This results in word scores $s_w^{(ILS)} = \alpha s_{\Pi_1(w)}^{(asso)} + \beta s_{\Pi_2(w)}^{(feat)} + (1 - \alpha - \beta) s_{\Pi_3(w)}^{(co-occ)}$, where Π_1, Π_2, Π_3 indicate potentially different random permutations of the word list.

It is important to note the following:

- The Global Label-Shuffled (GLS) null model preserves the underlying distribution of global scores $\{S_i\}$ at a global network level, hence the name. However, it does not preserves the local multidegree or multi-closeness centralities of a given node (i.e. on the microscopic level). This model is a suitable null model because, in addition to the above global preservation of the scores, it also preserves inter-layer degree/closeness correlations (i.e. a hub in one layer might tend to be a poorly connected node in another layer).
- The Independent Label-Shuffled null model does not preserve the underlying distribution of global scores, since it reallocates node labels independently at random on each layer. This model preserves only the distributions of intra-layer scores $\{\{s_i^{(asso)}\}, \{s_i^{(feat)}\}, \{s_i^{(co-occ)}\}\}$. It disrupts inter- layer correlations.

Using these null multiplex models, we estimate the word gains. Results from Supplementary Tab. S4 show that:

1. Even when the global distributions of scores are preserved, the optimisation procedure does not lead to results close to the optimal results for the non shuffled multiplex network, both results are separated by at least a factor of six.
2. Results for the ILS are compatible with a random ordering when error bars are considered. Maintaining the distributions of scores in each individual layer is not enough to guarantee better results than random word guessing.

Ordering	VELS	ELS	LLS	Total
Opt. Deg. (VELS)	0.16(4)	0.12(2)	0.08(1)	0.09(1)
Opt. Deg. (ELS)	0.14(2)	0.17(2)	0.09(1)	0.11(1)
Opt. Deg. (LLS)	0.14(2)	0.17(2)	0.08(1)	0.11(1)
GLS Opt. Deg. (VELS)	0.03(2)	0.02(1)	0.01(1)	0.02(1)
GLS Opt. Deg. (ELS)	0.02(2)	0.03(1)	0.01(1)	0.02(1)
GLS Opt. Deg. (LLS)	0.02(2)	0.03(1)	0.01(1)	0.02(1)
ILS Opt. Deg. (VELS)	0.002(2)	0.002(2)	0.001(1)	0.002(2)
ILS Opt. Deg. (ELS)	0.001(1)	0.003(2)	0.002(2)	0.002(2)
ILS Opt. Deg. (LLS)	0.001(1)	0.003(2)	0.002(2)	0.002(2)
Opt. Clos. (VELS)	0.19(3)	0.19(2)	0.07(1)	0.12(2)
Opt. Clos. (ELS)	0.16(2)	0.25(2)	0.09(1)	0.13(1)
Opt. Clos. (LLS)	0.14(2)	0.24(2)	0.10(1)	0.14(1)
GLS Opt. Clos. (VELS)	0.03(3)	0.03(2)	0.01(1)	0.02(2)
GLS Opt. Clos. (ELS)	0.03(2)	0.04(2)	0.01(1)	0.03(2)
GLS Opt. Clos. (LLS)	0.02(2)	0.04(2)	0.01(1)	0.03(1)
ILS Opt. Clos. (VELS)	0.002(3)	0.002(2)	0.001(1)	0.002(2)
ILS Opt. Clos. (ELS)	0.001(2)	0.003(2)	0.001(1)	0.001(1)
ILS Opt. Clos. (LLS)	0.001(2)	0.003(2)	0.001(1)	0.001(1)

Table 5.5: Average vocabulary normalised word gain obtained when optimising degree (Deg.) or closeness centrality (Clos.) from the empirical dataset, the shuffled GLS model and the shuffled ILS model. The results were averaged over 500 shuffles. Error bars represent standard deviations and are denoted within parentheses, i.e. 0.19(3) means 0.19 ± 0.3.

3. Preserving the inter-layer multiplex correlations (as in the GLS null model) is fundamental in providing better performances of the optimisation procedure.

All in all, prediction on shuffled multiplex networks are substantially less accurate than predictions reported in the main Chapter. This gives further support for the main findings of the Chapter, i.e. correlations on the (multiplex) are an important determinant in word acquisition.

5.5.15.2 Importance testing for the phonological layer

Previous literature has reported correlations between degrees in the phonological layer and word learning in young children (cf. Carlson et al. (2014)). In contrast, the main Chapter does not find a significant influence of phonological information on word acquisition. The purpose of this section is to demonstrate that in principle our methodology can discover and appropriately quantify such correlations if present in the dataset. To demonstrate this, we re-allocate word labels on the phonological layer in such a way that phonological degrees correlate with acquisition orderings to a tuneable extent. This can be achieved by reshuffling randomly selected words in a directed manner until a desired correlation level has been achieved. Then we run optimisations for a multiplex composed of unaltered semantic layers and the reordered phonological layer. We randomise over 20 probabilistic age of acquisition orderings and perform 10 Monte Carlo robustness samplings for each age of acquisition ordering. We consider three hypothetical scenarios: a situation in which phonological degree and age of acquisition have a small correlation (Kendall Tau of 0.1), a situation with intermediate strength correlations (Kendall Tau 0.3) and a situation with very high correlations (Kendall Tau 0.8). Notice that the degree-based ranking of the original phonological layer correlated with the normative age of acquisition with a Kendall Tau of $\approx$ 0.01. Results of the reordering experiments are given in Figure SI Fig. 5.15.

In agreement with expectations, we observe that the optimisation assigns: (i) small but non-negligible weight to the phonological layer for the small correlation scenario, (ii) significant weight to the phonological layer for intermediate strength correlations and (iii) dominant weight to the phonological layer in the large correlation scenario. As shown in the main Chapter, in particular also the association layer correlates with intermediate strength with age of acquisition (Kendall Tau $\approx$ 0.24). Hence, in particular in the intermediate correlation scenario in which correlations of the phonological and of the association layer with age of acquisition are of similar magnitude, we observe interactions between these two layers, i.e. the phonological layer first dominates and then loses relative weight when the association layer gains influence. This effect is also detectable for the large correlation scenario, but here correlations in the phonological layer are always dominant.

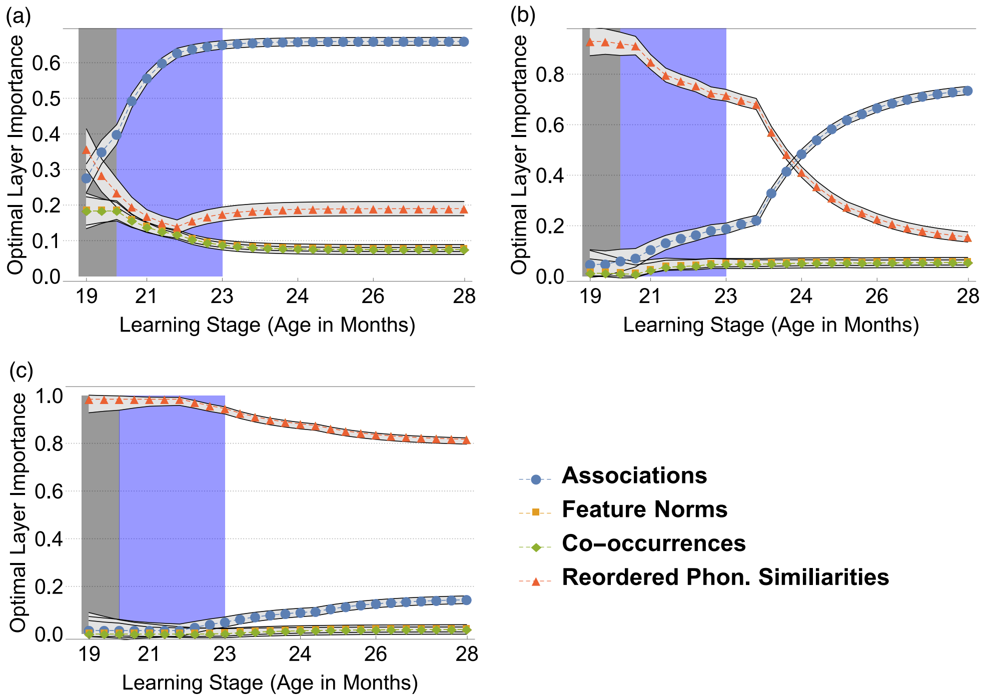


Figure 5.15: Optimal layer weights obtained from the degree optimisation where the labels on the phonological layer are reshuffled in order for word degrees to correlate with the age of acquisition ordering according to a Kendall Tau of 0.1 (top left), 0.3 (top right) and 0.8 (top left).

These experiments clearly demonstrate that our methodology is able to attribute layer weightings according to the predictive power of layers for word acquisition. In particular, the methodology is in principle capable of also quantifying correlations in the phonological layer, that is, the observation of phonological layer weights in the order of 10^{-3} is not to be attributed to the methodology but rather to the lack of predictive power encapsulated within network statistics of the layer. Compared to the extended phonological layer we used in SI Sect. 5.5.6, the original smaller layer is much less connected and more fragmented. Both the layers are based on the same phonological similarity measure but the extended phonological layer includes more than 29000 words, while the one originally used in the MLN includes only 529 words. We therefore consider the lack of performance of the MLN phonological layer to be a matter of quantity (rather than quality).

Chapter 6

Multiplex model of mental lexicon reveals explosive learning in humans

The multiplex structure proved powerful in predicting word learning in toddlers. But could it be used for further quantitative testing of psycholinguistic patterns during cognitive development up to adulthood? In adults, single-layer network studies confirmed the psycholinguistic conjecture about a kernel core in the mental lexicon, i.e. a cluster of tightly connected words that facilitates mental navigation through concepts and definition of words in a given language. These analyses focused on individual aspects of the lexicon, considering one type of word similarities at once. In this Chapter[1] we overcome this limitation by testing the above conjecture of lexicon coreness through extended multiplex lexical networks on large-scale datasets.

Abstract

Similarities among words affect language acquisition and processing in a multi-relational way barely accounted for in the literature. We propose a multiplex network representation of word similarities in a mental lexicon as a natural framework for investigating large-scale cognitive patterns. Our model accounts for semantic, taxonomic, and phonological interactions and identifies a cluster of words of higher frequency, easier to identify, memorise and learn and with more meanings than expected at random. This cluster emerges around age 7 yrs through an explosive transition not reproduced by null models. We relate this phenomenon to polysemy, i.e. redundancy in word meanings. We show that the word cluster acts as a core for the lexicon, increasing both its navigability and robustness to degradation in cognitive impairments. Our findings provide quantitative confirmation of existing psycholinguistic conjectures about core structure in the mental lexicon and the importance of integrating multi-relational word-word interactions in suitable frameworks.

---

[1] This Chapter follows the layout of the "Three-paper" PhD thesis in that it is an original, publishable research manuscript available as: *Stella, M., Beckage, N. M., Brede, M., & De Domenico, M. (2017). Multiplex model of mental lexicon reveals explosive learning in humans. arXiv preprint arXiv:1705.09731.* In order to better present the content, the Chapter/paper is briefly introduced by a short summary. The Chapter follows the same structure of the original paper (Stella et al., 2017), in adherence to the guidelines of the "Three-paper" PhD thesis format. For a better presentation within the thesis, the words "paper" and "manuscript" were changed to Chapter. The Supplementary Information of the original paper is reported at the end of the Chapter rather than in an appendix for increased readability and easier consultation with the main text.

6.1 Introduction

Investigating relationships between words offers insights into both the structure of language and the influence of cognition on linguistic tasks (Karuza et al., 2016; Beckage & Colunga, 2015). As a result, cognitive network science is rapidly emerging at the interface between network theory, statistical mechanics and cognitive science (De Deyne et al., 2016; Baronchelli et al., 2013; Beckage & Colunga, 2015; Karuza et al., 2016). The field is influenced by the seminal work from Collins & Quillian (1969), who assumed that concepts in the human mind are cognitive units, each representable as a node linked to associated elements. These connections represent a complex cognitive system known as the mental lexicon (Aitchison, 2012). Extensive empirical research has shown that relationships in the lexicon can be modelled as a network of mental pathways influencing both how linguistic information is acquired (Storkel, 2002; Beckage & Colunga, 2015; Casas et al., 2016; Carlson et al., 2014; Hills et al., 2009; Vitevitch et al., 2012), stored (Aitchison, 2012; Storkel, 2002; De Deyne et al., 2016; Vitevitch, 2008), and retrieved (Collins & Loftus, 1975; i Cancho & Solé, 2001; Vitevitch et al., 2012; De Deyne et al., 2016).

The cognitive role of quantification of lexical navigability as network distances finds empirical support in several experiments related to word identification and retrieval tasks (Collins & Loftus, 1975; Collins & Quillian, 1969; Dehaene et al., 1998; Meyer & Schvaneveldt, 1971). For instance, Collins and Loftus(Collins & Loftus, 1975) showed a correlation between network topology of semantic networks and word processing times: words farther apart on the network require longer identification times, thus indicating higher cognitive effort. More recently, the structural organisation of mental pathways among words was analysed in several large-scale investigations, considering similarity of words in terms of their semantic meaning (Sigman & Cecchi, 2002; De Deyne et al., 2016; Dorogovtsev & Mendes, 2001), their phonology (Vitevitch, 2008; Vitevitch et al., 2012; Siew, 2013; Stella & Brede, 2015, 2016a), or their taxonomy (i Cancho & Solé, 2001; Picard et al., 2009; Liu & Cong, 2014). Remarkably, all these networks, based on different definitions of relationships between words, were found to be highly navigable (sometimes called small-worlds (Watts & Strogatz, 1998)): words were found to be clustered with each other and separated by small network distances. This may suggest a universal structure of language organisation related to minimising cognitive load while maximising navigability of words (Beckage et al., 2011; Baronchelli et al., 2013; Solé & Seoane, 2015; Beckage & Colunga, 2015).

The above studies, however, have not yet attempted to use multi-relational information for characterising and quantifying the mental lexicon, often focusing on only one relationship at a time (Collins & Loftus, 1975; Hills et al., 2009; Carlson et al., 2014; Sigman & Cecchi, 2002; De Deyne et al., 2016; Dorogovtsev & Mendes, 2001; Beckage et al., 2011; Vitevitch, 2008). Some researchers have considered the aggregation of several of these relationships into single-layer networks (Sigman & Cecchi, 2002) and others have considered multi-relational models but only to capture the syntactic structure of language (Liu & Cong, 2014). However, the above approaches offer only limited insight into the cognitive complexity that allow individuals to use language (Aitchison, 2012) with diversity and ease.

More information about the lexical structure can indeed be obtained by accounting, simultaneously, for multiple types of word-word interactions. A natural and suitable framework for this purpose are multilayer networks (De Domenico et al., 2016, 2013; Kivelä et al., 2014;

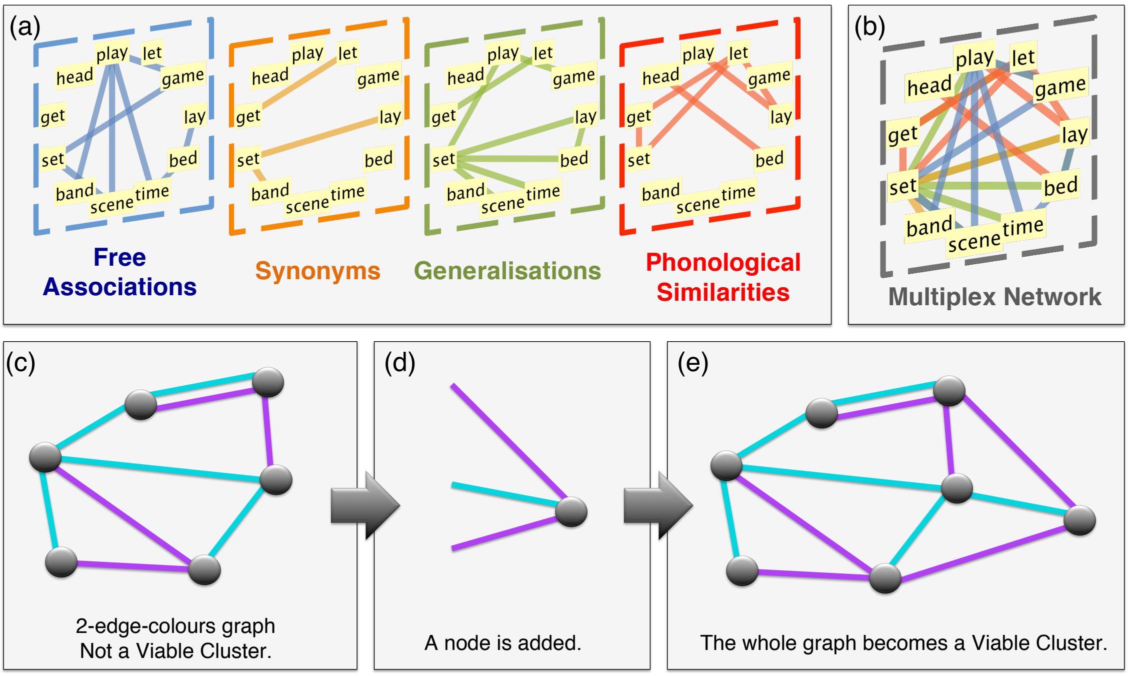
Boccaletti et al., 2014; Battiston et al., 2017). Multilayer networks simultaneously encode multiple types of interaction among units of a complex networked system. Therefore, they can be used to extract information about linguistic structures beyond information available from single-layer network analysis (Stella et al., 2017). The usefulness of multiplex representations has recently been shown for diverse applications including the human brain(Bassett & Sporns, 2017; De Domenico, 2017), social network analysis(Szell et al., 2010; Mucha et al., 2010; De Domenico et al., 2015), transportation(De Domenico et al., 2014; Cardillo et al., 2013) and ecology (Stella et al., 2016; Pilosof et al., 2017).

Here, on an unprecedented scale from a multi-relational perspective, we investigate the semantics, phonology, and taxonomy of the English lexicon as a model of distinct layers of a multiplex network (see Fig. 1). We study the evolution of multiplex connectivity over the developmental period from early childhood (2 years of age) to adulthood (21 years of age).

The proposed multiplex representation provides a powerful framework for the analysis of the mental lexicon, allowing for the capture of sudden structural changes that can not be identified by traditional methods. More specifically, when modelling lexical growth, we observe an explosive emergence of a cluster of words in the lexicon around the age of 7 years, which is not observed in single-layer network analyses. We show that this cluster is beneficial from a cognitive perspective, as its sudden appearance facilitates word processing across several linguistic pathways at once. This boost to cognitive processing also enhances the resilience of the lexicon network when individual words become progressively inaccessible, such as what may happen in cognitive disorders like anomia(Laine, 2013). These findings represent the first quantitative confirmation and interpretation of previous conjectures (Barsalou, 2008; Solonchak & Pesina, 2015; Picard et al., 2009; Aitchison, 2012) about the presence and cognitive impact of a core in the human mental lexicon.

6.2 Results

Our multilayer lexical representation (MLR) of words in the mind is a multiplex network(Wasserman & Faust, 1994; Baxter et al., 2016; De Domenico et al., 2016; Kivelä et al., 2014) made of $N = 8531$ words and four layers. Each layer encodes a distinct type of word-word interaction (cf. Fig. 6.1 (a)): (i) empirical free associations (Coltheart, 1981), (ii) synonyms (Miller, 1995), (iii) taxonomy/generalisations (Miller, 1995), and (iv) phonological similarities (Vitevitch, 2008). As shown in Fig. 6.1 (b), different relationships can connect words that would otherwise be disconnected in some single-layer representations. We considered these relationships with the aim of building a representation accounting for different types of semantic association, either from dictionaries (i.e. synonyms and generalisations) or from empirical experiments (i.e. free associations). We also include sound similarities (i.e. phonological similarities) to capture the main aspects involved in lexical retrieval from the mental lexicon (Collins & Loftus, 1975; Vitevitch, 2008; Vitevitch et al., 2012). This set of relationships represents a first approximation to the multi-relational structure of the mental lexicon. Compared to previous work on multiplex modelling of language development (Stella et al., 2017), the multiplex representation is enriched with node-level attributes related to cognition and language: (i) age of acquisition ratings (Kuperman et al., 2012), (ii) concreteness



2-edge-colours graph
Not a Viable Cluster.

A node is added.

The whole graph becomes a Viable Cluster.

Figure 6.1: (a): Visual representation of a subset of the multiplex lexical representation (MLR) for adults with $N = 8531$ words and four types of word relationships forming individual layers: free associations, synonyms, generalisations, and phonological similarities. (b) Multiplex visualisation as an edge-coloured network. (c) Using only purple links does not allow navigation of the whole network. Therefore the network is not a viable cluster. Notice, however, that the two nodes with overlapping links constitute the smallest possible viable cluster in a simple graph (which we refer to as "trivial" in the main text). (d-e) The appropriate addition of one node and three coloured links makes the resulting graph a viable cluster, with paths between all nodes using either only cyan or only purple colours.

ratings (Brysbaert et al., 2014), (iii) identification times in lexical decision tasks (Keuleers et al., 2012), (iv) frequency of word occurrence in Open Subtitles (Barbaresi, 2014) and (v) the number of unique definitions of a word in WordNet, used to quantify polysemy (Sigman & Cecchi, 2002) (see Methods). The analysis of structural reducibility of our multiplex model (see SI) quantifies the redundancy of the network representation (De Domenico et al., 2015). Results suggests that layers should not be aggregated with each other as each network layer contributes uniquely to the structure of the multiplex representation, confirming the suitability of the framework for further investigation.

As already discussed, investigating navigation on linguistic networks proved insightful (Collins & Quillian, 1969; Collins & Loftus, 1975; Sigman & Cecchi, 2002). Hence we focus on analysing the navigability of our multiplex network (De Domenico et al., 2014), identifying word clusters that are fully navigable on every layer, i.e. where any word can be reached from any other word on every layer when considered in isolation. An example is reported in Fig. 6.1 for a representative multiplex network with two layers. In network theory, these connected subgraphs are also called viable clusters (Baxter et al., 2016) (see Methods). Notice that the largest viable cluster of a single-layer network coincides with its largest connected component (Newman et al., 2011), i.e. the largest set of nodes all that can be reached from each other within one layer. In multiplex networks the two concepts are distinct, as viable clusters are required to be connected on *every* *layer when considered individually.* Removing this constraint of connectedness on every layer leads to the more general definition of multi-layer connected components(De Domenico et al., 2014), i.e. the largest set of nodes all connected to each other when jumps across layers are allowed. Fig. 6.1 (c-e) conveys the idea that the emergence of viable clusters can be due to the addition of particular links in the network.

Our multiplex model contains a single non-trivial (i.e. with more than two nodes) viable cluster composed of 1173 words, about 13.8% of the network size. In the following we refer to this cluster as the largest viable cluster (LVC). For easier reference, we indicate words in the empirical LVC as "LVC-in words" and words outside of the LVC as "LVC-out words". Reshuffling network links while preserving word degrees leads to configuration model-layers (Newman et al., 2011) that still display non-trivial LVCs (cf. LVC Rew. in Tab. 1). Further, on average $98.1 \pm 0.1\%$ of LVC-in words persist in the viable cluster after rewiring 5% of all the intra-layer links at random. We conclude that the LVC does not break but rather persists also in case of potentially missing or erroneous links in the network dataset (e.g. spurious free associations or mistakes in phonological transcriptions).

In order to further test correlations between network structure and word labels, we also consider a full reshuffling null model (see SI), in which word labels are reshuffled independently on every layer and thus word identification across layers is not preserved. Hence, full reshuffling destroys inter-layer correlations but preserves network topology. Fully reshuffled multiplex networks did not display any non-trivial viable clusters, emphasising the important role of inter-layer relationships for the presence of the LVC in the empirical data.

In the next section we analyse the evolution of the LVC during language learning over a time period of more than 15 years. We demonstrate the existence of an explosive phase transition (Baxter et al., 2016) in the emergence of the LVC and explore the significance of this transition from the perspective of cognitive development.

6.2.1 Emergence of the Largest Viable Cluster

To study the emergence of the LVC during cognitive development, we simulate probabilistic normative word orderings by smearing the age of acquisition dataset (Kuperman et al., 2012). We refer to these orderings as normative acquisition. Smearing allows us to account for the variance in age of acquisition across individuals by introducing a probabilistic interpretation of these orderings (see Methods). We compare the trajectories of normative acquisition against four null models: (i) random word learning (i.e. words are acquired at random), (ii) frequency word learning (i.e. higher frequency words are acquired earlier), (iii) polysemy word learning (i.e. words with a higher count of context-dependent meanings are learned earlier) and (iv) multidegree word learning (i.e. words with more connections –across all layers– are learned earlier). We investigate if modelling the development of the mental lexicon as growth of the empirical multiplex representation according to a given learning scheme matches the explosive transition observed in normative learning. Results are reported in Fig. 2 (a).

Normative acquisition indicates a sudden emergence of the LVC around age 7.7 ± 0.6 years, almost four years earlier than expected if learning words at random. Further analysis reveals two distinctive patterns. Firstly, this sudden appearance is robust to fluctuations in word rankings

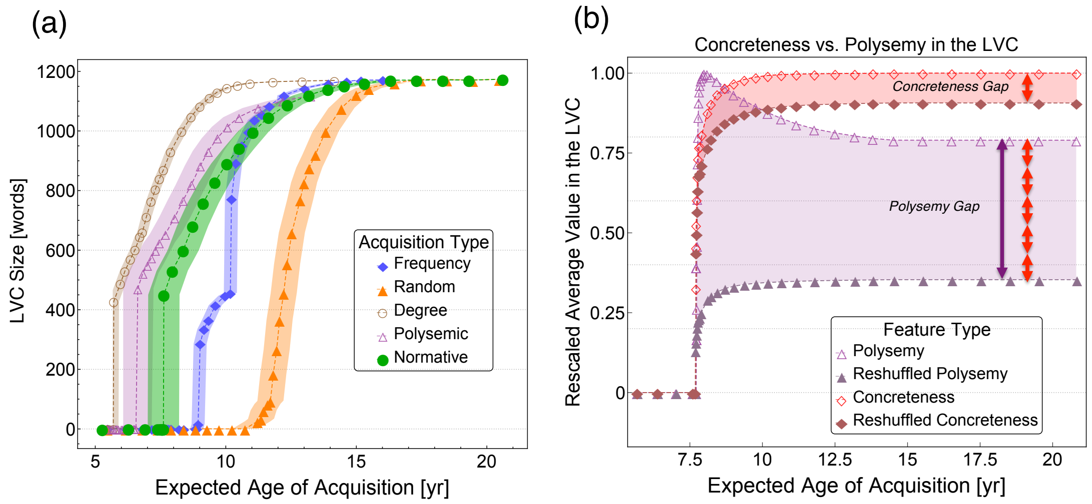


Figure 6.2: (a): Evolution of the size of the LVC when words are acquired in ascending order based on: age of acquisition (green dots), frequency (blue diamonds), polysemy (purple triangles), multidegree in the multiplex (brown circles) and at random (orange triangles). The LVC emerges with an explosive transition at 7.7 ± 0.6 years in normative acquisition. Areas represent standard deviations considering randomisations of smeared age of acquisition or ties in the rankings. (b): Comparison of average linguistic features for words in the LVC with normative acquisition in the empirical data and for a partial reshuffling null model with reshuffled node attributes. The curves are rescaled from 0 to 1 by their empirical maximum value and they represent averages over 200 iterations. Error margins are approximately the same size as the dots. Reshuffling node attributes results in a LVC with both reduced concreteness and polysemy values. We note significant gaps between the empirical and randomised data. The observed gap in polysemy values is almost 5 times larger than for concreteness values.

in the age of acquisition ratings (AoA): in all simulations based on AoA reports, after roughly 2500 words have been acquired, a LVC with at least 260 words suddenly appears after adding just a single word to the lexicon. Secondly, the average magnitude of this explosive change is $\Delta L_{AoA} = (420 \pm 50)$ words. These patterns suggest an explosive phase transition(D'Souza & Nagler, 2015; Grassberger, 2015; Baxter et al., 2016) in the structural development of the mental lexicon. To the best of our knowledge, this work is the first detection of such an explosive behaviour in cognitive language data.

Explosive behaviour in the emergence of the LVC is not observed in the random acquisition null model (see Methods), with only a few cases ($\chi_{Ran} = 32\%$) displaying a discontinuity of more than ten words. Further, the average magnitude of the LVC size change is only $\Delta L_{Ran} = (30 \pm 10)$ words, a full order of magnitude smaller than in the normative cases. Therefore explosiveness characterises normative acquisition as a genuine pattern of language learning.

Is the explosive appearance of the LVC due to the acquisition of specific links or rather to specific words? In order to test this, we focus on the set of "critical" words, i.e. the single words whose addition allows for the sudden emergence of the LVCs. We then compare features of these critical words with features of words already within the LVC at the time of its emergence. We test features like node-attributes (e.g. frequency, polysemy, etc.) and node degree. At a 95% confidence level, no difference was found for any feature (sign test, $p - value = 0.007$). This lack of difference suggests that the emergence of the LVC is indeed due to higher-order link correlations rather than local topological features (such as degree) or psycholinguistic attributes. Hence, it is the global layout of links that ultimately drive the explosive appearance of the LVC. As shown also in Fig. 6.1 (c-e), links crucial to the formation of the viable cluster might be acquired earlier (Fig. 6.1 (c)) but the LVC might appear only later (Fig. 6.1 (e)), after some key pathways completing the viable cluster are added to the network (Fig. 6.1 (d)).

The explosive emergence of the LVC has an interesting cognitive interpretation. Work in psycholinguistics suggests that frequency is the single most influential word feature affecting age of acquisition (Kuperman et al., 2012) (mean Kendall $\tau \approx -0.47$ between frequency and AoA). We thus test whether the LVC growth can be reproduced through early acquisition of highly frequent words, with frequency counts gathered from Open Subtitles (Barbaresi, 2014). All simulations on the frequency-based ordering display an explosive transition ($\chi_{fre} = 100\%$), however, the magnitude of the explosive transition is $\Delta L_{fre} = 280 \pm 30$ words, which is almost 2/3 of the normative one. At a confidence level of 95%, the distribution of frequency-based LVC magnitude changes differs from the normative one (sign test, p-value = 0.01). The distribution of ages at which the LVC emerges in the frequency null model overlaps (see Methods) in 21% of cases with the analogous normative one. We further observe that the frequency null model differs from the normative one not only quantitatively (i.e. magnitude and appearance of explosiveness) but also qualitatively: the frequency null model displays a second explosive phase transition in LVC size later in development, at around 10 ± 0.2 years of age. This second transition might be due to the merging of different viable clusters, since we focused only on the largest viable cluster, rather than on viable clusters of non-trivial size. Further analysis reveals that the multiplex network has only one viable cluster, which suddenly expands through a second explosive transition in the frequency-based vocabulary growth model (but not in the normative AoA model). The above differences provide strong evidence that explosiveness in the mental lexicon is not an artefact of correlation of word frequency with language learning patterns.

We next test preferentially learning words with high degree in the multiplex network to see if the LVC emerges earlier than in normative acquisition. Learning higher degree words first makes more links available in the multiplex network. As we said above, it is links that drive the LVC emergence, hence the expectation of earlier LVC appearance. The multidegree null model confirms this expectation and it displays a distribution of explosive transitions with average magnitude of 430 ± 30 but happening almost two years earlier than in normative acquisition, around age 5.8 ± 0.1, cf. Fig. 6.1. The distribution of critical ages overlaps with the normative one only for 2%. We conclude that the degree acquisition is significantly different from the normative case (mean Kendall $\tau \approx -0.31$ between multidegree and AoA).

Another feature that can influence language acquisition is polysemy (Casas et al., 2016; Solé & Seoane, 2015; Sigman & Cecchi, 2002), i.e. how many different context-dependent meanings a word has. As a quantification for polysemy, we consider the number of word definitions in the Wolfram dataset *WordData* (WolframResearch). When words with higher polysemy are acquired earlier, we find the appearance of the LVC at around age 6.6 ± 0.6 years, with an average magnitude of 470 ± 60 words, close to the normative one. The distribution of critical ages at which the LVC emerges in the polysemy null model displays the highest overlap (35%) with the analogous distribution from the normative case across all the null models we tested. Despite polysemy displaying a smaller correlation with the age of acquisition (mean Kendall

$\tau \approx -0.26$) compared to frequency or multidegree, it actually provides the highest overlap in terms of age at which the LVC emerges. This indicates that polysemy might play a role in driving the LVC emergence.

Another attribute that could impact language development is concreteness, i.e. how tangible a given concept is based on human judgements (Brysbaert et al., 2014; Hanley et al., 2013). Experimental research has shown that children tend to learn words earlier if a word is rated higher on concreteness (Kuperman et al., 2012; Brysbaert et al., 2014; Piaget, 2000; Aitchison, 2012). In order to test how concreteness can influence the LVC evolution, we develop a partial reshuffling null model (see Methods) where the topology of words is fixed but node attributes are reshuffled at random. Partial reshuffling destroys the correlations between word features and the network topology, such that we can quantify the role of the relational structure in the absence of correlation with word features. Partial reshuffling gives rise to LVCs of the same size but containing words that are less concrete and less polysemous than in normative acquisition, cf. Fig. 2 (b). Partial reshuffling of word frequency leads to a gap in frequency of similar size as we see for concreteness (see SI). The gap in polysemy between the empirical and the reshuffled LVCs is five times larger than the analogous concreteness gap, suggesting that polysemy has a greater influence than concreteness over the emergence of the LVC. We also notice a peak in polysemy: the "backbone" of the LVC (i.e. the LVC emerging around 8 yr) is composed of significantly more polysemous words compared to the LVC at age 20 (cf. Fig. 2 (b), sign test, p-value $= 0.001 < 0.05$). This early peak is absent in the partial reshuffling null model for polysemy. Furthermore, frequency (see SI) and concreteness do not display peaks early on after the LVC emergence. Such an early richness in high-polysemy words further indicates the idea that polysemy strongly influences the emergence of the LVC.

Polysemy is universal across languages (Aitchison, 2012; Solé & Seoane, 2015), even though it can represent an ambiguity in communication (Solé & Seoane, 2015): one word can have different meanings depending to the context. Usually in semantic networks(Steyvers & Tenenbaum, 2005; Sigman & Cecchi, 2002; Aitchison, 2012) one meaning of a word can be represented by its links in its neighbourhood (e.g. "character" and "nature" are linked in the synonyms network as they overlap in meaning). Hence one word meaning corresponds to a given neighbourhood layout for that word. Polysemic words have different meanings, therefore their neighbourhoods can aggregate (e.g. "character" is linked to "nature" in the context of complexion but also to "font" in the context of typography). Reshuffling at random the word labels constituting all the neighbourhoods in the network evidently disrupts meanings and their aggregation, thus destroying polysemy. We call this reshuffling "full" as it preserves the structure of connections in the layers but it fully destroys both intra-layer correlations at the endpoints of links and inter-layer correlations of words, thus fully disrupting word meanings. We use full reshuffling as a null model (see Methods and SI) for testing how important polysemy is in determining the presence of the LVC. We fully reshuffle 2025 high-polysemy words (i.e. the words making up the heavy tail of the polysemy distribution) and compute the LVC size in the resulting reshuffled multiplex networks. Results are compared against a reference case in which the same number of low-polysemy words are fully reshuffled. No viable cluster emerges on the multiplex networks with fully reshuffled high-polysemy words, while the LVC only shrinks by roughly 13% in case of fully reshuffling low-polysemy words. We conclude that correlations between word meanings and polysemy are indeed necessary in determining the presence of the LVC.

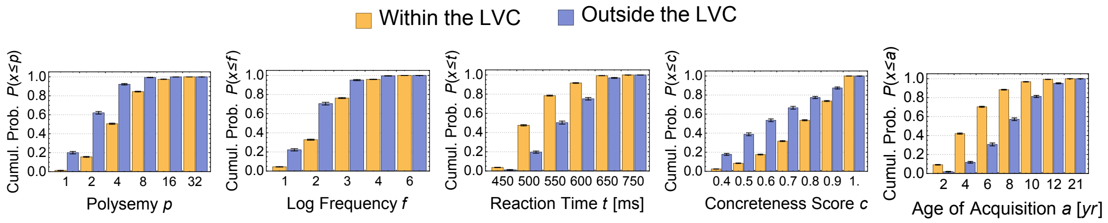
Node Attributes	LVC In	LVC Out	Asso. LCC In	Syno. LCC In	Hyp. LCC In	Phon. LCC In	LCC Int.	LVC Rew.
Age of Acquisition [ys]	6.43(2)	9.4(1)	8.5(1)	8.8(1)	9.0(1)	7.8(1)	7.4(1)	7.3(1)
Concreteness	3.93(3)	2.83(4)	3.63(4)	3.35(5)	3.45(5)	3.87(2)	3.72(2)	3.71(3)
Reaction Times [ms]	552(1)	600(3)	579(1)	581(2)	588(2)	581(2)	576(1)	569(1)
Log Frequency [Counts]	3.40(1)	2.57(1)	2.86(1)	2.85(1)	2.79(1)	2.95(1)	3.20(1)	3.30(1)
Polysemy [Meanings]	9.7(2)	3.6(2)	4.9(1)	5.6(1)	4.6(2)	5.8(1)	7.6(1)	8.2(1)

Degree Corrections	LVC In	LVC Out	Asso. LCC In	Syno. LCC In	Hyp. LCC In	Phon. LCC In	LCC Int.	LVC Rew.
Age of Acquisition [ys]	6.43(2)	7.62(1)	7.2(1)	8.1(1)	8.1(1)	7.5(1)	6.62(2)	6.61(2)
Concreteness	3.93(2)	3.67(3)	3.79(2)	3.42(4)	3.40(4)	3.89(2)	3.89(2)	3.90(2)
Reaction Times [ms]	552(1)	565(1)	559(1)	570(2)	566(2)	575(3)	556(1)	555(1)
Log Frequency [Counts]	3.40(2)	2.86(2)	3.26(1)	3.21(1)	3.21(1)	3.30(2)	3.32(1)	3.36(1)
Polysemy [Meanings]	9.7(2)	5.48(2)	6.8(1)	8.0(1)	7.7(1)	6.4(2)	8.5(1)	8.7(1)

Table 6.1: Average node attributes for words within the LVC and within the largest connected component (LCC) for each individual layer. All the values are medians, except for heavy-tail distributions such as the frequency and polysemy ones, where the arithmetic mean was used instead. All the values are sample-size corrected via Monte Carlo sampling. The last five rows refer to degree-corrected samplings, where the sampled LVC-out words have the same degree of the sampled LVC-in words. Error bars are reported in parentheses for brevity: 3.93(3) means 3.93 ± 0.03.

The above results indicate that polysemy does increase lexicon navigability by ultimately giving rise to the LVC, i.e. a relatively small cluster of words that is fully navigable under both semantic, taxonomic, and phonological relationships in the mental lexicon. Such view is in agreement with previous works (i Cancho & Solé, 2001; Sigman & Cecchi, 2002; Solé & Seoane, 2015), which point out how polysemy provides long-range connections in the lexicon which can increase navigability through different word clusters on semantic single-layer networks (Sigman & Cecchi, 2002).

6.2.2 Psycholinguistic characterisation of the Largest Viable Cluster (LVC)



(a) (b) (c) (d) (e)

Figure 6.3: Cumulative probabilities of finding a word with a given feature less than a threshold T for LVC-in (orange boxes) and LVC-out (blue boxes). For instance, the probability of finding a low frequency word (with frequency ≤ 10) at random is 0.05 within the LVC but almost five times larger outside of the LVC.

Next, we explore the impact of the presence of the LVC on cognitive aspects of language such as word processing. Our aim is to explore if words belonging to the LVC (LVC-in) are processed differently than those words not in the LVC (LVC-out), more from a language use perspective rather than a developmental one (which was analysed with the previous null models). Hence, we turn to large-scale datasets of node attributes (see Tab. 1 and Methods). We find (cf. Tab. 1) that words in the largest viable cluster (i) are more frequent in the Open Subtitles dataset (Barbaresi, 2014), (ii) acquired earlier according to AoA reports (Kuperman et al., 2012), (iii) quicker to identify as words in lexical decision tasks (Keuleers et al., 2012), (iv) are rated as more concrete concepts (Brysbaert et al., 2014) and thus more easily memorised (Hanley et al., 2013; Binder et al., 2005; Brysbaert et al., 2014) and (v) represent more meanings in different semantic areas(Casas et al., 2016) when compared to LVC-out words.

In Fig. 3 (a-e), we report the cumulative probabilities of finding a word with a given feature less than a certain value for a set of particular node-level attribute within and outside of the LVC. The difference between LVC-in and LVC-out further indicates how different the words in the LVC are compared to LVC-out words. For instance, let us consider reaction times, which indicate how quickly people classify stimuli as words or nonwords in lexical decision tasks (Keuleers et al., 2012). The probability of finding at random an LVC-in word correctly identified in less than $500\,ms$ is 0.48 while it is less than half, 0.2, for LVC-out words. Hence the LVC is rich in words identified more quickly. Analogous results hold for all the tested attributes.

Since LVC-in words have higher degree compared to LVC-out words (see SI) and degree correlates with many of the psycholinguistic attributes used in our study, it is interesting to quantify to which extent the difference between LVC-in and LVC-out is due to correlations with degree. Results shown below the thick line, in the lower part of Tab. 1, suggest that the degree effect does not fully explain the observed psycholinguistic features of the LVC: a sign test indicates that all the median node-attributes of LVC-in words are higher than those of LVC-out words, at 95% confidence level. Notice that the comparison that does not account for degree is still important since one could easily argue that degree itself can be interpreted as a cognitive component that affects word processing (Steyvers & Tenenbaum, 2005; Vitevitch et al., 2012).

Tab. 1 also compares the statistics of the LVC against its single-layer counterparts, i.e. the largest connected components (De Domenico et al., 2013) (LCC In). We also consider multiplex alternatives to the LVC such as: the intersection across all layers of words in the LCC of each layer (LCC Int, cf. SI) and the LVC in configuration models (LVC Rew.), which consist on average of 40% more words. The empirical LVC consists of words with the most distinct linguistic features compared to the other tested sets of words, in terms of all tested node attributes. Even rewiring all links does not completely disrupt such distinctness (cf. LVC Rew.). These differences in linguistic attributes suggest that the LVC is a better measure of "coreness" for words in the mental lexicon than either the LCCs or their intersection, an idea we further test in the next section.

6.2.3 Robustness of the multiplex lexicon and LVC to cognitive impairments

The LVC has been characterised as a set of higher degree words that differ in psycholinguistic features when compared to words located outside the LVC in our multiplex. This suggests that the higher degree, and cognitive correlations, of the LVC may be because the LVC is acting as a core for the mental lexicon. Let us denote the total number of links on a given layer as L and the link density as p. As shown in Fig. 6.4 (a), there are more links within the LVC ($Lp_{In/In}$) across all layers than outside of it ($Lp_{Out/Out}$) or at the interface of the LVC ($Lp_{In/Out}$). Further, across all individual layers the inequality $p_{In/In} > p_{In/Out} > p_{Out/Out}$ holds, denoting the presence of a core-periphery structure for the node partition $\{In, Out\}$ (Newman, 2012).

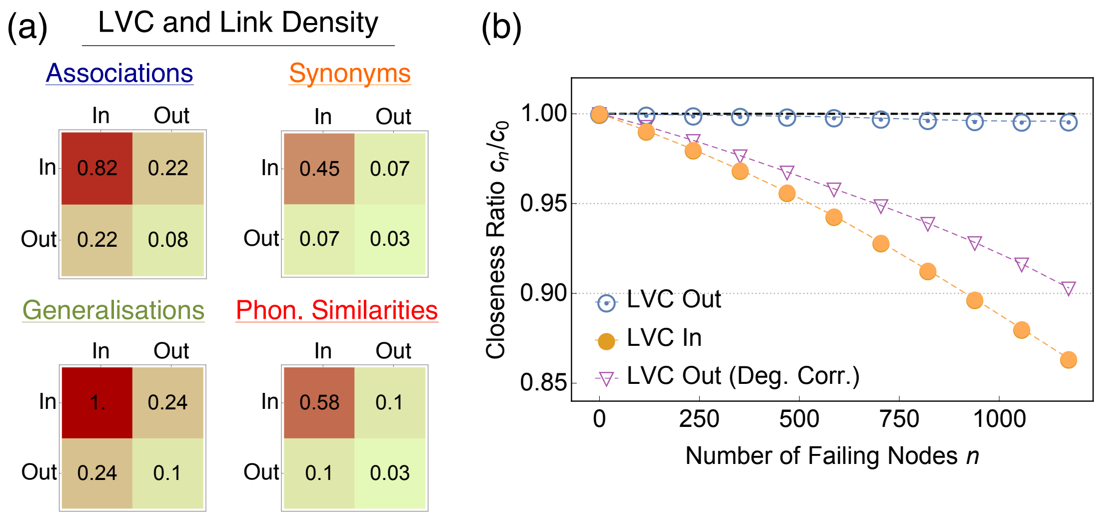


Figure 6.4: (a) Normalised link densities across layers for couples of nodes either in the LVC (In), out of the LVC or on the boundary (one node in, one node out). Densities are normalised by the maximum value ($Lp_{In/In}$ for generalisations) and colour coded (the higher the value, the more red the cell). (b) Resilience analysis with respect to random word failure, mimicking progressive aphasia in the mental lexicon. Words are targeted at random and then removed from the whole multiplex. In LVC Out (Deg. Corr.) we remove words from outside the LVC but with the same degree as the words removed inside the LVC, thus correcting for a degree effect seen in the LVC which will also effect efficiency. As a measure of efficiency we use the median closeness of words in the network, providing the inverse of the average number of network hops necessary for reaching any word from any other one through the multiplex topology. Error margins represent standard deviations and they are about the size of the dots.

In order to better interpret both the coreness and cognitive impact of the LVC, we perform a resilience analysis of the network by means of numerical experiments. Random word failure provides a plausible toy model for progressive anomia (Laine, 2013) driven by cognitive decline, where words become progressively non-accessible on all the lexicon levels without a clear trend (Laine, 2013).

To simulate progressive anomia, we randomly remove LVC-in and LVC-out words in separate experiments. The maximum number of removed words is 1173, corresponding to the size of the LVC. As a proxy for robustness, we consider the average multiplex closeness centrality, which correlates with the average cognitive effort for identifying and retrieving words within the lexicon (Collins & Quillian, 1969; Sigman & Cecchi, 2002) and plays a prominent role in early word acquisition as well (Stella et al., 2017). The results of this analysis are shown in Fig. 6.4 (b).

We find that the multiplex representation is robust to random LVC-out word removal: removing almost 1170 LVC-out words only reduces average closeness to a level that is still within a 95% confidence level of the original multiplex. Therefore failure of LVC-out words does not impact the cognitive effort in identifying and retrieving words within the lexicon. Instead, the multiplex lexicon is fragile to random LVC-in word removal: removing 50% of words from the LVC leads to a decrease in closeness 20 times larger than the drop observed for LVC-out words. While considering random removal in both cases, it is true that in general

LVC-in words have higher degree than LVC-out words, which might influence the robustness results from a technical perspective. The discrepancy in closeness degradation is only partly due to the higher degree of LVC-in words. Performing degree-corrected LVC-out word deletions still leads to less of a decrease in navigability as compared to LVC-in word deletion, as evident from Fig. 6.4 (b).

In summary, the multiplex lexicon is fragile to word failures of LVC-in words and robust to random failures of LVC-out words. This difference is a strong indicator that the LVC provides the necessary shortcuts for efficient navigation – with high closeness and thus low cognitive effort – of the mental lexical representation. It is worth remarking that the network's navigability is expected to increase in the presence of cores(Csermely et al., 2013; Newman, 2012), further supporting the interpretation that the LVC acts as a core of the multiplex structure. It has been conjectured that the mental lexicon has a core set of concepts (Barsalou, 2008; Solonchak & Pesina, 2015; Aitchison, 2012; Picard et al., 2009); we showed here how various cognitive metrics can be correlated with the LVC, suggesting that future work may benefit from considering the LVC as a quantification of lexical core structure.

6.3 Discussion

Previous literature from psycholinguistics has conjectured the existence of a core set of words in the lexicon (Barsalou, 2008; Solonchak & Pesina, 2015; Picard et al., 2009; Aitchison, 2012). Here, for the first time we give large-scale quantitative evidence to support these conjectures. In fact, we identify the largest viable cluster (LVC) of words which: (i) favours the emergence of connectivity which allow for navigation across all layers at once and (ii) acts as a core for the multiplex lexical representation. Words within the LVC display distinct cognitive features, being (i) more frequent in usage (Barbaresi, 2014), (ii) learned earlier (Kuperman et al., 2012), (iii) more concrete (Brysbaert et al., 2014) and thus easily memorised (Aitchison, 2012; Brysbaert et al., 2014) and activating perceptual regions of the brain (Binder et al., 2005), (iv) with more context-dependent meanings (Casas et al., 2016) and (iv) more easily identified in lexical decision tasks (Keuleers et al., 2012) than words outside the LVC. Remarkably, the explosive emergence of the LVC happens around age 7 years, which is also a crucial stage for cognitive development in children. According to Piaget's theory of cognitive development, age 7 years is the onset of the *concrete operational stage*, in which children develop more semantic and taxonomic relationships among concepts (e.g. recognising that their cat is a Siamese, that a Siamese is a type of cat and that a cat is an animal, thus drawing the conclusion that their cat is an animal among several). Experimental evidence (Ozcan, 2012) has also shown that, in this developmental stage, children display an increased ability of mental planning and usage of context-dependent words in a connected discourse such as narratives (Ozcan, 2012). Interestingly, both these findings can be interpreted in terms of an increased ability to navigate context-dependent meanings in the mental lexicon, which we quantitatively link to the explosive emergence of LVC core structure above. This indicates that the multiplex lexical network is a powerful representation of the mental lexicon: the network structure can indeed capture and translate well documented mental processes driving cognitive development into quantifiable information.

One limitation of our current approach is that we do not consider lexical restructuring over time, i.e. the adults' representation of word relationships could be different compared to children's or adolescents'. Previous work on the phonological level (Storkel, 2002) showed partial differences in phonological neighbourhoods between pre-schoolers and pre-adolescents. However, we show that the LVC persists even when all connections are randomly rewired and it stills identifies relevant words, e.g. more frequent, more concrete, etc. suggesting that the role of the LVC may still hold even with restructuring. Link rewiring also allows consideration of the variance in word learning due to individual differences. Individual difference modelling may be especially important for quantification, diagnosing, explaining, and correcting various language learning and usage issues (Beckage et al., 2011).

Another limitation is that the network representation might not be exact, e.g. there might be spurious links in the empirical free association layer or mistaken phonetic transcriptions in the phonological layer. In order to address this issue, we randomly reshuffle 10% of word labels, 2.5% on each layer separately, and find that the largest viable clusters are 10% smaller than the empirical LVC (t-test, $p-value = 0.009$). However, the LVC after reshuffling exhibits analogous performance in the features discussed in Tab 1 (sign test, $p-value = 0.96$). Together with the random rewiring experiments, this is an indication that the LVC structure is robust to small perturbations due to errors in the annotation of links or word labels.

Core/periphery network organisation is commonly found in many real-world systems (Csermely et al., 2013; Brede & de Vries, 2009), even though the definition of cores in multiplex networks remains an open challenge. We interpret the robustness experiments as quantitative indication that the LVC is acting as a core for the whole multiplex lexical network, increasing navigability in two ways. Within the LVC, words must be connected to each other, implying navigability from every word within the LVC across all individual layers. Outside of the LVC, connections to the viable cluster facilitate network navigation by making words closer to each other. Since closeness correlates with the cognitive effort in word processing (Collins & Loftus, 1975; Collins & Quillian, 1969; Sigman & Cecchi, 2002), the LVC can be considered as facilitating mental navigation through pathways of the mental lexicon. This quantitative result is in agreement with previous conjectures about polysemy facilitating mental navigation of words (i Cancho & Solé, 2001; Sigman & Cecchi, 2002; Solé & Seoane, 2015). Additionally, our results also indicate that the LVC acts as a multiplex core, with robustness to node failure due to densely entwined links and connections allowing navigation even in cases where words become inaccessible, as in cognitive disorders like progressive anomia (Laine, 2013) or due to individual variability early in development. It is worth remarking that we identify such a core with the largest LVC as no other non-trivial viable cluster exists in the MLR.

Indeed, identifying a core in the mental lexicon provides quantitative evidence supporting previous claims(Barsalou, 2008; Solonchak & Pesina, 2015) about the existence of a core of highly frequent and concrete words in the lexicon facilitating mental navigation and thus word retrieval in speech production experiments (Barsalou, 2008; Solonchak & Pesina, 2015; Hanley et al., 2013). Alongside the cognitive perspective, interpreting the LVC as a lexicon core provides support for further previous findings about the presence of a "kernel lexicon" in language (Dorogovtsev & Mendes, 2001; i Cancho & Solé, 2001; Picard et al., 2009), a set of a few thousand words which constitute almost 80% of all written text (Aitchison, 2012) and can define every other word in language (Picard et al., 2009). Previous works on semantic

(Dorogovtsev & Mendes, 2001; i Cancho & Solé, 2001), taxonomic (Picard et al., 2009) and phonological (Siew, 2013; Vitevitch et al., 2012) single-layer networks identified a kernel lexicon for the English language with roughly 5000 words which has not changed in size during the evolution of languages. This kernel lexicon was identified with the largest connected component of the English phonological network (Vitevitch et al., 2012; Siew, 2013). The LVC we present here is a subset of the largest connected components, even of the phonological one, and it also persists across semantic, taxonomic and phonological aspects of language. Hence, the LVC represents a further refinement of the kernel lexicon that: (i) it greatly rich in polysemous words, (ii) facilitates mental navigation and (iii) it is robust to rewiring or cognitive degradation. These three features suggest an interpretation of the LVC as a linguistic core of tightly interconnected concepts facilitating mental navigation through key words.

While the framework presented here has been applied only for the English language, comparison with other languages and linguistic representations to assess how universal the LVC core is remains an exciting challenge for future experimental and theoretical work.

6.4 Methods

6.4.1 Dataset and cognitive interpretation

The data used in this work come from different sources and thus the resulting multiplex network representation is based on independent studies. For our multiplex we construct four layers that model semantic, taxonomic and phonological relationships. We further distinguish semantic relationships in free associations and synonyms. For free associations, e.g. "A reminds one of B", we used the Edinburgh Associative Thesaurus (Coltheart, 1981). For both, generalisations (e.g. "A is a type of B") and synonyms (e.g. "A also means B") we used WordNet 3.0 (Miller, 1995). For phonological similarities we used the same dataset analysed in (Stella & Brede, 2016a) based on WordNet 3.0 (Miller, 1995). We treat every layer as undirected and unweighted. Words in the multiplex representation are required to be connected on at least one layer.

Free associations indicate similarities within semantic memory, i.e. when given a cue word "house", human participants respond with words that remind them of "house", for example "bed" or "home". Networks of free associations play a prominent role in capturing word acquisition in toddlers (Hills et al., 2009; Stella et al., 2017) and also word identification (De Deyne et al., 2016; Collins & Loftus, 1975). Also networks of synonyms are found to play a role in word identification (Aitchison, 2012). The hierarchy provided by taxonomic relationships deeply affects both word learning and word processing (Collins & Quillian, 1969; Aitchison, 2012). Phonological networks provide insights about the competition of similar sounding words for confusability in word identification tasks (Vitevitch, 2008; Vitevitch et al., 2012, 2014; Stella & Brede, 2015).

For the linguistic attributes we combined several different sources. We sourced word frequency from OpenSubtitles (Barbaresi, 2014), a dataset of movie subtitles whose word frequencies were found to be superior to frequencies from classical sources in explaining variance in the analysis of reaction times from lexical decision experiments (Barbaresi, 2014; Keuleers et al., 2012). Concreteness scores (Brysbaert et al., 2014) and age of acquisitions ratings (Kuperman et al.,

2012) were gathered from Amazon Turk experiments, allowing for large-scale data collection and confirmation of previous findings based on small-scale experiments (Brysbaert et al., 2014; Kuperman et al., 2012). The concreteness ratings indicate how individual concepts are rated (on a scale of 1 - "abstract" to 5 - "concrete") as abstract rather than perceptual (Brysbaert et al., 2014). Polysemy was quantified as the number of different definitions for a given word in WordData from Wolfram Research (WolframResearch, 2017). Reaction times were obtained from the British Lexicon Project (Keuleers et al., 2012) and indicate the response time in milliseconds for the identification of individual words were compared against non-words.

6.4.2 Smearing normative acquisition

Smearing is a technique used in statistics for improving the quality of data (Fayyad et al., 1996). We smear the age of acquisition data from Kuperman et al. (2012), where the average age of acquisition a_i and standard deviation $\sigma_a(i)$ around each word are provided, e.g. $a_{aim} = 6.72\,yrs, \sigma_a(aim) = 2.11\,yrs$. In our case, smearing consists of sampling possible age of acquisitions for word i from a Gaussian distribution $\mathcal{N}[a_i, \sigma_a(i)]$ rather than considering only the average value. Sampling independently an age of acquisition for each word in the dataset, we can build multiple artificial acquisition rankings r_{AoA} of words from the empirical data.

6.4.3 Lexicon growth experiments

We simulate lexicon growth over time $t(n)$ by considering subgraphs of the multiplex lexicon where only the first $n \leq 8531$ words in a given ranking r are considered. Rankings indicate the way words are acquired in the lexicon over time and can be based on word features or age of acquisition reports. The rankings we use are based on: (i) smeared age of acquisition (Kuperman et al., 2012), (ii) frequency(Barbaresi, 2014; Kuperman et al., 2012) (higher frequency words are learned earlier), (iii) multidegree (De Domenico et al., 2013) (words with more links across all layers are learned earlier), and (iv) polysemy (words with more meanings are learned earlier). As a randomised null model, we consider random word rankings. When the first n words in a ranking are considered, a subgraph of the multiplex lexicon with these words is built and its LVC is detected. By using the non-smeared age of acquisitions, we relate the number of learned words to the developmental stage in years $t(n)$, e.g. $n = 1000$ corresponds to $t = 5.5$ years.

The size of the LVC $L(t)$ is then obtained as a function of developmental stage $t(n)$ for every specific type of ranking. Results for the smeared age of acquisitions and the random null model are averaged over an ensemble of 200 iterations. Results for the frequency, degree and polysemy orderings are averaged over 200 iterations where words appearing in ties are reshuffled. Results are reported in Fig. 2.

Each iteration represents the evolution of the LVC size through the acquisition of an individual word. This acquisition trajectory may be related to different developmental stages. For every iteration, we detect the magnitude of the transition on the LVC size due to its appearance when adding words one by one to the network. We then compute the fraction χ of iterations presenting a discontinuity of more than 10 LVC-in words size. We also compute the average magnitude of the explosive transition ΔL.

Comparisons of the empirical distributions of ages at which the LVC emerges considers the overlapping coefficient (Fayyad et al., 1996), i.e. the overlap of two distributions normalised by the maximum overlap obtained when shifting the central moment of one of the distributions. An overlap of 100% means that one distribution is fully contained in the other one. An overlap of 0% means that the distributions have no overlap.

6.4.4 Robustness experiments

We carried out robustness testing via word/node removal: individual words are removed at random across all layers. Closeness centrality is then measured by considering shortest paths across the whole multiplex network structure, i.e. also including jumps between layers. We consider closeness centrality as a measure for the spreading of information and the mental navigability of the lexicon (Collins & Loftus, 1975; i Cancho & Solé, 2001; Siew, 2013). In our case closeness is well defined, since even the deletion of the whole LVC leaves the multiplex network connected (De Domenico et al., 2014). We consider a multiplex network as connected if it is possible to reach any pair of nodes by allowing for traversal along links on any layers.

With reference to Fig. 3, we perform random attacks of words within the LVC (LVC In) and outside of it (LVC Out). Since LVC-in words are more connected compared to words outside, we also perform degree corrected attacks: we perform attacks of random words within the LVC and words of equivalent degree outside the LVC. This degree correction (LVC Out - Deg. Corr.) allows for the attack of LVC-out words but reduces the number of links by the same amount as LVC-in attacks.

6.5 Supplementary Information

6.5.1 Topology of Individual Layers

The Supplementary Table S1 contains different network metrics for the individual layers of the **multiplex lexical representation (MLR)** described in the main text. The reported network features are: mean degree $\langle k \rangle$, mean local clustering coefficient CC. size of the largest connected component LCC, network diameter D and mean shortest path length $\langle d \rangle$.

In the MLR, degree counts the number of similarities a given word has on a given layer (Newman, 2010). For instance, on the association layer a word with degree 5 has exactly 5 words associated to it. On the phonological layer, the degree coincides with its phonological neighbourhood size (Vitevitch, 2008): a word with degree k has exactly k similar sounding words. Words with higher degree in semantic networks were found to have lower age of acquisition norms (Steyvers and Tenenbaum, 2005), while degree on the phonological layer correlated positively with word confusability (Vitevitch, 2014). The main cognitive intuition of degree as a network metric is that words with multiple semantic/taxonomic associates or multiple phonological neighbours are indeed important in the lexicon as they can be accessed through a variety of links by mental navigation of the lexical pathways. As reported in Supplementary Tab. S1, the layers of free associations and generalisations have higher mean degree compared to the layers of synonyms and phonological similarities.

Network	**⟨*k*⟩**	***CC***	**|*LCC*|**	***D***	**⟨*d*⟩**
Free Associations	8.7	0.093	6431	7	3.8
Free Associations CM	8.7	0.007(6)	6420(10)	8(1)	3.6(2)
Synonyms	3.1	0.28	5261	18	6.2
Synonyms CM	3.1	0.002(1)	6370(40)	13(1)	5.4(3)
Generalisations	9.9	0.048	7910	11	4.0
Generalisations CM	9.9	0.014(7)	7930(30)	8(1)	3.6(2)
Phonological Similarities	3.6	0.34	3668	22	6.7
Phonological Similarities CM	3.6	0.005(1)	4560(30)	11(1)	4.3(2)
Aggregate Network	23	0.084	8519	8	3.3
Aggregate Network Random	23	0.014(7)	8520(10)	7(1)	3.1(1)

Supplementary Tab S1: Network metrics for the individual layers in the Multiplex Lexical Representation and for randomised null models (configuration models): mean degree $\langle k \rangle$, mean local clustering coefficient CC. size of the largest connected component LCC, network diameter D and mean shortest path length $\langle d \rangle$. As null models we chose configuration models, i.e. networks with randomised links preserving the original degree of words in the layers of the MLR. Error margins indicate standard deviations based on 200 random realisations. For instance 8(1) means 8 ±1.

The mean local clustering coefficient CC measures how much, on average, neighbourhoods of nodes resemble complete graphs (Newman, 2010). The mean clustering coefficient CC is the mean local clustering computed across all nodes of a single-layer network. Local clustering on the phonological network was found to positively correlate with word confusability in lexical recognition tasks (Vitevitch, 2014): the more locally clustered a word was the easier it was for participants to fail in identifying it. Together with the shortest path length, clustering determines the so-called small-world property (Watts and Strogatz, 1998).

The shortest path length among any two nodes in a single-layer network is defined as the minimum number of network hops necessary for navigating the network from one node to the other. In the multiplex lexical representation we ignore any cost for jumping across layers, hence no inter-layer links are considered. This modelling assumption leads to shortest paths over the multiplex structure coinciding with shortest paths over the aggregate network.

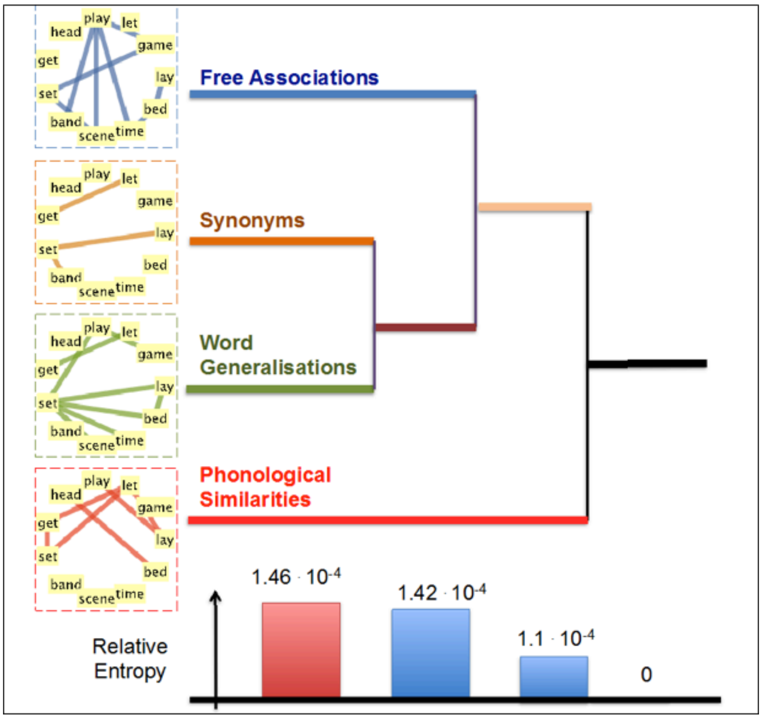
As reported in Supplementary Tab. S1, all of the MLR layers are small-world networks. We stick to the established definition of small-worldness from the literature of cognitive networks (Steyvers and Tenenbaum, 2005 and Beckage et al., 2011) according to which a single-layer network is a small-world if, compared to equivalently sized Erdős–Rényi random graphs, it exhibits a significantly higher mean clustering coefficient and a comparable mean shortest path length. Small-worldness has been found independently in different types of semantic networks (Steyvers and Tenenbaum, 2005) and in phonological networks as well (Vitevitch, 2008). As already suggested in previous works (Beckage et al. 2011, Vitevitch, 2014), small-worldness might be cognitively beneficial to language learning and use, as it might allow for efficient navigation within the mental lexicon. Furthermore, empirical evidence has shown that small-worldness is related to language learning in children (Beckage et al., 2011): semantic network lexicons of late talkers, who are likely to exhibit language learning deficits, do show small-worldness to a much smaller degree compared to lexicons from children learning words at normative pace.

6.5.2 Structural Reducibility Analysis

Multiplex networks are a natural framework for considering multi-relational interactions within a single representation (Battiston et al., 2017b). Nonetheless, it is important to quantify model parsimony: does including different layers or relations give additional information compared to an aggregated view? Aggregating here means considering an “aggregate” network where edges are present if they appeared in any of the individual network layers. Structural reducibility analysis (De Domenico et al., 2015) investigates whether a given multiplex representation could be aggregated or not, identifying the presence and extent of redundant topological patterns across layers. Structural reducibility analysis relies on: (i) identifying layers similar in their topologies, (ii) aggregating layers if appropriate, and (iii) comparing the richness in topological patterns of the aggregated multiplex layers against the aggregated network, obtained by projecting all the links in the multi-layer structure on a single-layer network. Analogously to a hierarchical clustering, the procedure identifies similarity of layers and quantifies how distinguishable the multiplex is against aggregate versions of two or more layers. The whole procedure is based on the Von Neumann entropy of each multiplex layer and we refer to (De Domenico et al., 2015) for further details.

In order to assess the benefit of the multiplex representation, we adopt the procedure suggested in (De Domenico et al. 2015). Results are reported in Supplementary Fig. S2. Interestingly, all the three semantic layers are clustered together and are only aggregated with the topological layer at the last stage of the hierarchical clustering. At every level of the hierarchy a relative entropy measure is reported, expressing how different is the average entropy of the considered layers compared to the aggregated multiplex network with all layers collapsed into one network. The higher the relative entropy, the more different the multiplex layers are, on average, compared to the aggregated single-layer network. Aggregating layers with different topological structure might cause a loss of information. In fact, after the aggregation, one would consider as indistinguishable (i.e. coming from the same layer) two classes of links that differed greatly in their assortment among nodes and were thus, in the multiplex representation, distinguishable (i.e. coming from different layers). Notice that the ideas of distinguishability and mixing among links coming from different layers relates to the quantum mechanics machinery behind irreducibility analysis (De Domenico et al., 2015).

Beyond the technical details, it is the information loss when aggregating multiplex layers that can indeed motivate pursuing a multiplex representation rather than an aggregated one. As reported in Supplementary Fig. S2, the highest information loss is relative to aggregating the original MLR with all its 4 layers kept as distinct. This result indicates that no aggregation



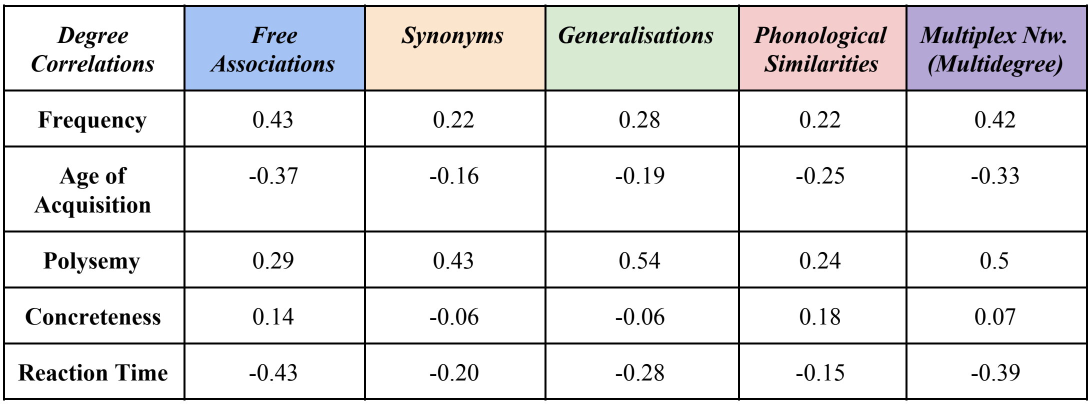
Supplementary Figure S2 : Structural reducibility analysis of the different layers in the multiplex lexical representation of the mental lexicon. Layers are hierarchically clustered according to the similarity in their topologies. At every level of the resulting dendrogram, an entropy measure quantifies how much more topological information a certain collection of layers has compared to the case in which all layers are projected into a single-layer network only (i.e. the aggregate network). The configuration having most topological information compared to the aggregate is the one in which all the layers are kept as separate: our MLR cannot be aggregated or clustered without losing topological information on the patterns among words.

can be performed without losing topological information on the pathways among words, thus motivating the importance of a multiplex representation of the mental lexicon over the 4 aspects we considered.

6.5.3 Degree Correlations and Corrections

As already discussed in SI Sect. 6.5.1, the degree of words in linguistic networks encapsulates some information about the cognitive and linguistic patterns recurring in word processing or acquisition. For instance, Stella et al., (2016) found that degree of words in the free association layer of young children was a powerful predictor of word learning trajectories in the early phases of development. On the phonological level, degree corresponds with the phonological neighbourhood size, which correlates with word confusability and retrieval time in lexical decision tasks (Vitevitch et al., 2014).

It is therefore expected for degree to correlate with other linguistic node attributes such as the ones tested in the main Chapter. We report such correlations in Supplementary Tab. S2. As a degree measure on the whole multiplex network we consider multidegree, i.e. the sum of the individual degrees across layers (cf. De Domenico et al., 2013). Positive (negative) correlation values of Kendall Tau indicate that higher degree implies higher (lower) node attribute.



Degree Correlations	*Free Associations*	*Synonyms*	*Generalisations*	*Phonological Similarities*	*Multiplex Ntw. (Multidegree)*
Frequency	0.43	0.22	0.28	0.22	0.42
Age of Acquisition	-0.37	-0.16	-0.19	-0.25	-0.33
Polysemy	0.29	0.43	0.54	0.24	0.5
Concreteness	0.14	-0.06	-0.06	0.18	0.07
Reaction Time	-0.43	-0.20	-0.28	-0.15	-0.39

Supplementary Table S2 : Kendall Tau correlations between degree on a given multiplex network layer and node-level attributes (frequency, age of acquisition, polysemy, concreteness and reaction time). The Kendall Tau correlation expresses the fraction of agreements minus disagreements in ordering when words are ranked according to the two tested features (e.g. if word i has degree higher than word j, how more likely is it to for word i to have a higher attribute than word j compared to word i having a smaller attribute than word j?). Kendall Tau significance tests indicate that all the reported correlations are statistically significant at a 0.05 significance level.

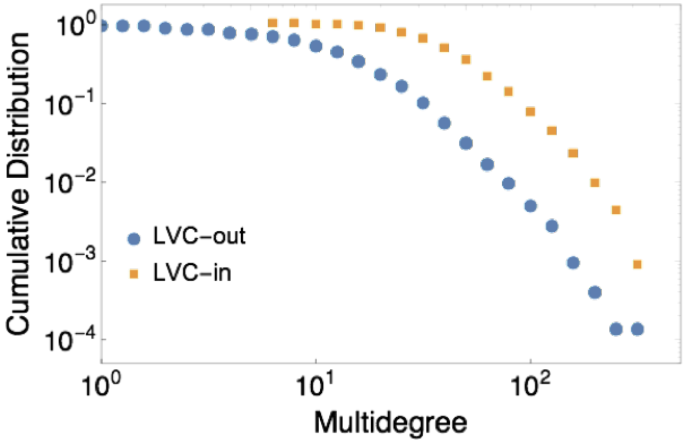
As reported in Supplementary Figure S2b, the LVC-out words are in general less connected compared to LVC-in words. One could then test if the differences in node level attributes between LVC-in and LVC-out words are merely an artefact of degree, provided that degree itself correlates with the node level attributes (cf. Tab. S2). Degree corrections allow to sample words with the same degree both in the LVC and outside of it, thus erasing first order differences in local connectedness (see Tab. 1 in the main Chapter). In the main text (cf. Tab. 1), degree corrections for the average node attributes of LVC-out words are performed by considering the degrees on the multiplex layer with the highest correlation to that specific node attribute. For instance, frequency correlates the most with word degrees in the association layer, therefore when computing the average frequency of LVC-out words with the same degree distribution as LVC-in words the degrees on the association layer are considered (see Methods of the main text for further details).

6.5.4 Reshuffling Null Models

Our reshuffling null models disrupt correlations between words within a given layer and across layers mainly in two ways (see also Methods in the main text):

1. *partial reshuffling* keeps words aligned with their topological position but it reshuffles individual node-attributes, i.e. it makes individual *rows* of Table S2 close to 0 when performed on all the layers;
2. *full reshuffling* performs a reshuffling of word labels, thus destroying any correlation between topology and node attributes, i.e. it makes the *whole* Table S2 close to 0 when performed on all the layers.

A visualisation of the different reshuffling models used in the main text is reported in SI Fig. S3. Notice how both the models preserve the link structure across layers but change the labels



Supplementary Figure S2b : Cumulative probability distribution of finding a node with a multidegree higher than a given value in the multiplex lexical representation for LVC-out words (blue) and LVC-in words. For instance, the probability of finding nodes with more than 100 connections across all layers is $2.5 \cdot 10^{-2}$ outside of the LVC and roughly $1 \cdot 10^{-1}$ within the LVC. Hence, the LVC is richer in higher degree nodes compared to the rest of the multiplex lexical representation.

or the node attributes at the endpoints. As evident from SI Fig. S3, full reshuffling does not change the node-level attributes a given word has, since the word label and all its features are reshuffled together (see the arrow in SI Fig. 3 (b)). Instead, partial reshuffling allows to consider reshuffled configurations where a given word is attributed another attribute (see the red arrow in SI Fig. 3 (c)).

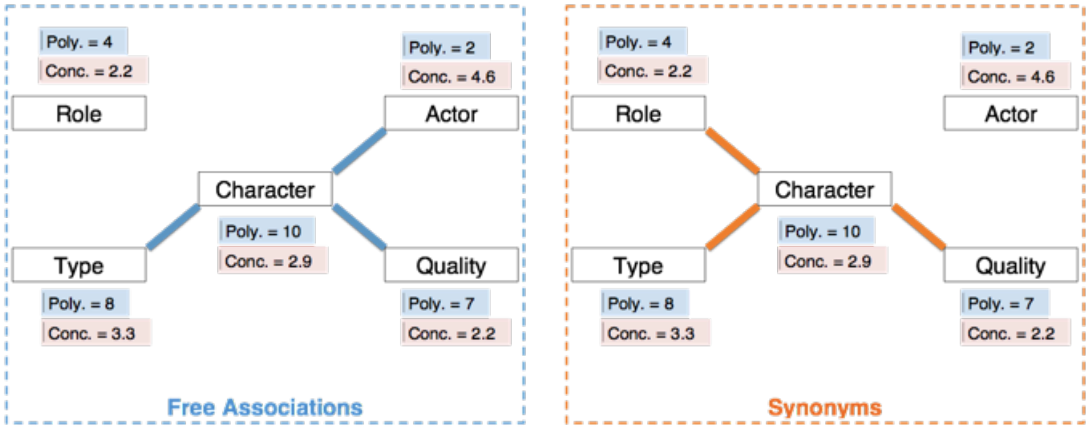
In the main text we performed full reshuffling of either high-polysemy words or of the same number of low-polysemy words. High-polysemy words were defined according to the distribution of polysemy counts in the MLR, which is reported in Supplementary Figure S3b. High-polysemy words are those in the heavy tail of the distribution, starting at polysemy count 5 (i.e. words with strictly more than 5 different context-dependent meanings in the dataset). Low-polysemy words were defined as the remaining words (i.e. words with less than 5 different context-dependent meanings in the dataset).

6.5.5 Attribute Correlations

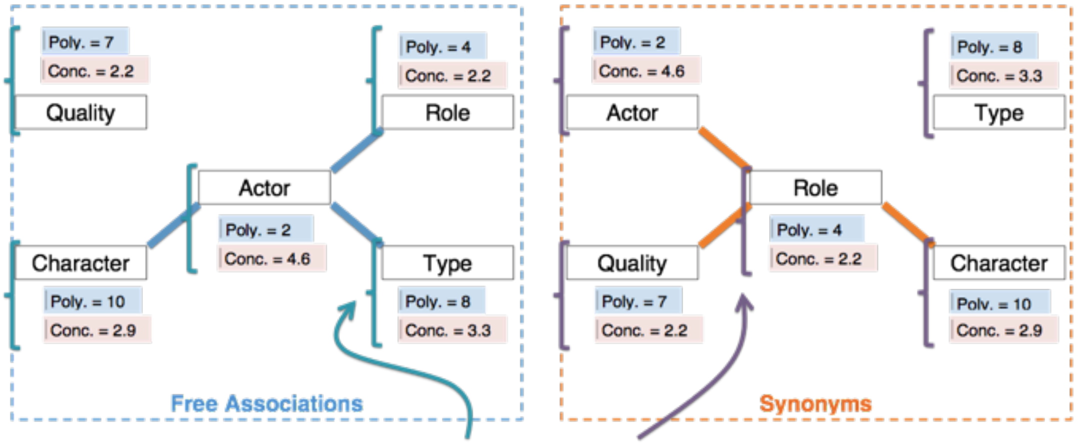
Partial reshuffling evidently destroys also correlations among node-level attributes of a given word. Supplementary Tab. S3 contains the Kendall Tau correlations among node-level attributes used in the main text.

6.5.6 Overlap Between Individual Largest Connected Components

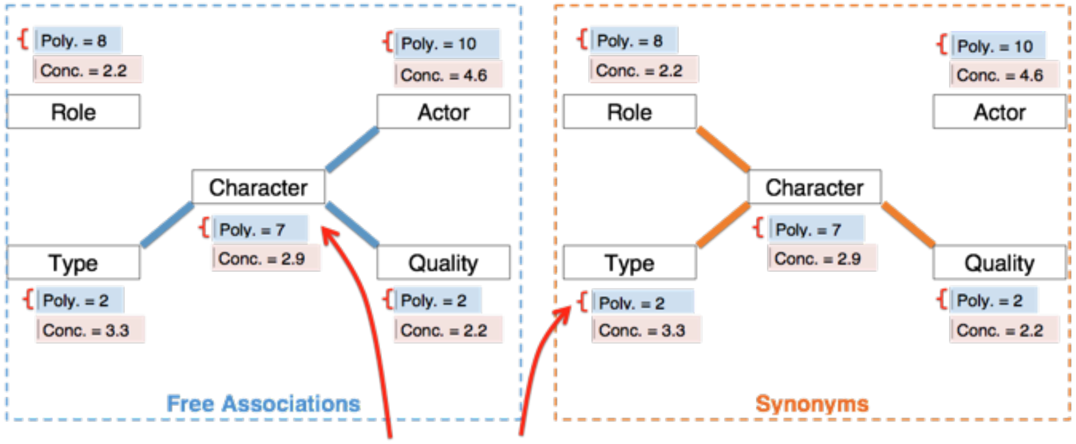
In the main text we focus on the largest viable cluster, whose definition for single-layer networks coincide with the one of largest connected component. The largest connected component (LCC) of a given network is defined as the largest set of nodes that can be reached from each other (Newman, 2010). Analogously to what we do in the main text for distributions of LVC emergence times, we define the *percentage intersection* as the intersection of word lists



(a) Empirical Multiplex Network

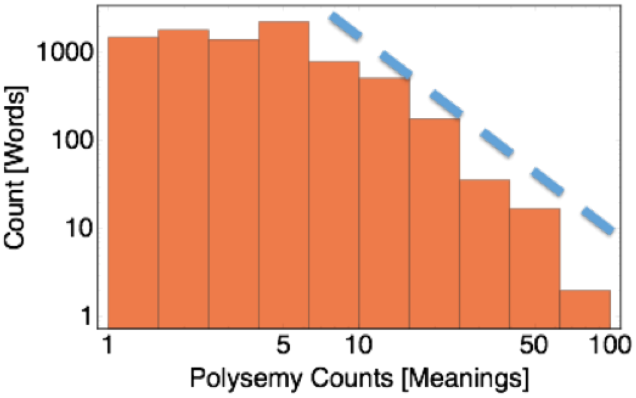


(b) Full reshuffling



(c) Partial Reshuffling

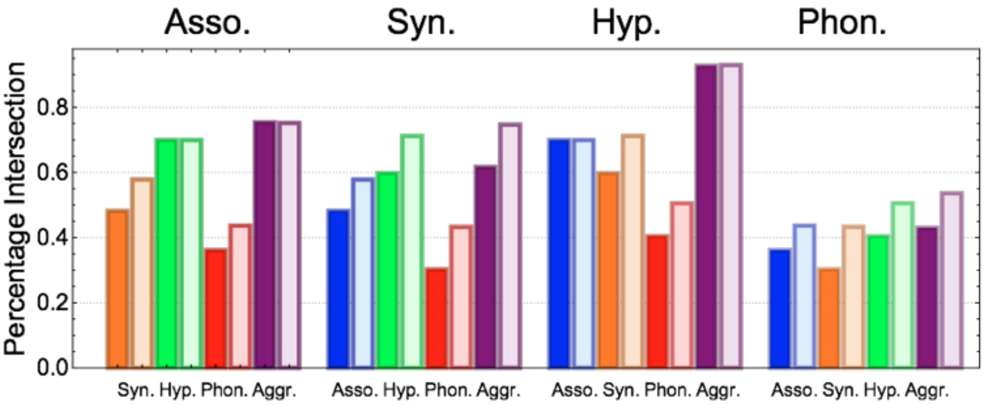
Supplementary Figure S3 : Examples of the reshuffling models for a subgraph of the multiplex lexical network limited to 5 words, 2 layers and 2 node-attributes (polysemy and concreteness). (a): Empirical network. (b) Full reshuffling of word labels, independently on each layer. Polysemic links are mostly disrupted. Arrows of different colour indicates that words are reshuffled independently on every layer (c) Partial reshuffling of node-attributes (e.g. polysemy counts), aligned across layers. Arrows of the same colour indicate that node attributes are reshuffled but kept consistent for a word across all layers. In both (b) and (c), the layout of links is untouched and degree distribution is preserved.



Supplementary Figure S3b: Log-log plot of the distribution of counts of words with a given polysemy count (i.e. how many words in the MLR have polysemy count 1, 2, 3, etc.). The dashed blue line is merely highlighting the heavy tail of the distribution, which includes words having more than 5 different meanings in the Wolfram's dataset WordData. No power-law fitting was performed on the data.

Attribute Correlations	***Frequency***	***Age of Acquisition***	***Polysemy***	***Concreteness***	***Reaction Time***
Frequency		-0.47	0.36	0.08	-0.54
Age of Acquisition	-0.47		-0.26	-0.28	0.44
Polysemy	0.36	-0.26		0	-0.35
Concreteness	0.08	-0.28	0		0
Reaction Time	-0.54	0.44	-0.35	0	

Supplementary Table S3 : Kendall Tau correlations between node attributes used in the main text. Correlations with p-value above a significance level of 0.05 were considered equivalent to 0 (i.e. statistically independent) within a Kendall Tau significance test and thus reported as “0” in the above table.



Supplementary Figure S4 : Percentage intersection between largest connected components of different layers ("Asso." for free associations, "Syn." for synonyms, "Hyp." for word generalisations and "Phon." for phonological similarities). The aggregated multiplex network is identified with "Aggr.". The plot can be read from top to bottom, e.g. at the top the Asso. layer has a LCC overlapping with the Phon. LCC (bottom) for 35% of the maximum possible value. Empty boxes are relative to configuration models, i.e. randomly rewired networks preserving word degrees. Error bars are of the same size of the box edge.

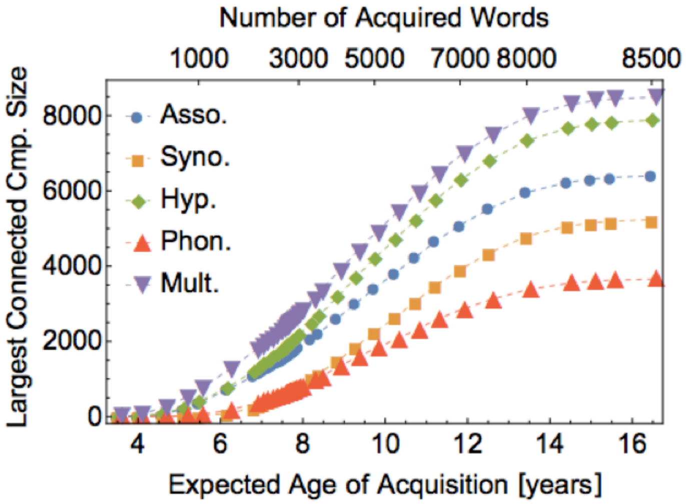
being in two largest connected components normalised by the maximum intersection possible (i.e. the largest connected component of the aggregated multiplex network). A percentage overlap of 50% between two layers means that half the words in the LCC of the aggregated multiplex network fall in the LCC of both the layers.

Notice that the LVC will always be either equal or smaller in size to the smallest pairwise percentage intersection. Pairwise intersections among largest connected components on individual layers are reported as percentage intersections in Supplementary Fig. S4. Full boxes report intersections among empirical layers while empty boxed indicate the reference value for configuration models.

The phonological layer is the layer overlapping the least among the four in the multiplex lexical representation. In all the possible cases, the phonological layer displays smaller intersections with other layers compared to random expectation. These findings agree with the results from the structural reducibility analysis in providing the idea that the phonological layer greatly differs in terms of pathways against all the other semantic layers in the MLR. Differently put, the above topological patterns confirm that phonological similarities among words do indeed look different from either semantic or taxonomic relationships. Our multiplex lexical representation naturally allows to keep phonological word-word interactions as separate from the others.

6.5.7 No Explosive Transitions in the Growth of Largest Connected Components

Contrary to what happens in the emergence of the LVC, the growth over time of largest connected components does not display any explosive behaviour. Considering 200 iterations where words are inserted in the multiplex lexical network according to their smeared age of acquisition, none of the simulated word trajectories exhibits an explosive discontinuity in the size of the largest connected component for any of the four individual layers.



Supplementary Figure S5 : Size of the largest connected component for the individual layers over normative word acquisition (“Asso.” for free associations, “Syn.” for synonyms, “Hyp.” for word generalisations and “Phon.” for phonological similarities). The aggregated multiplex network is identified with “Aggr.”. Error margins indicate standard deviations and are of the same size of the plot markers. Results are averaged over 200 iterations of smeared normative acquisition (in which words are acquired in the same order of their age of acquisition).

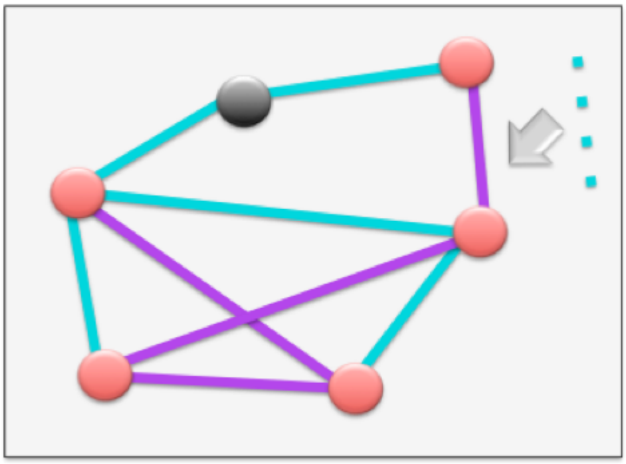
Also the connected component of the whole multiplex network (i.e. the connected component of its aggregate, De Domenico et al., 2013) does not display any explosive transition. Average behaviours are reported in Supplementary Fig. S5.

The above numerical results indicate that the explosive emergence of connected structures across all layers, such as the LVC, are indeed distinct topological patterns, different from those exhibited by other connected structures such as largest connected components.

6.5.8 No Explosive Transitions in the Growth of the Intersection of Largest Connected Components

We want to explore if the explosive pattern present in the LVC is replicated also in other reasonable definitions of cores. As another multiplex measure we explored also the intersection of largest connected components across all layers, which includes 2146 words. Notice that the intersection of LCCs does not require connectedness simultaneously across all layers, i.e. words in the intersection might not be connected with each other by paths using one of type of relationships only. The intersection of LCCs can therefore be considered as having some missing pathways in it, whose potential addition would make it a viable cluster, cf. Supplementary Fig. S6.

Differently from the LVC, the intersection of largest connected components does not display any explosive transition in its size when words are acquired according to normative acquisition (see Methods), neither when individual word trajectories are analysed. Supplementary Figure S7 reports the results for the average case, computed over 200 iterations of smeared normative acquisition.



Supplementary Figure S6 : Example network to illustrate differences in the notion of viable clusters and intersection between connected components, with 2 layers (represented by different colours). The intersection of LCCs is highlighted in red. Nodes in the intersection do not constitute a viable cluster as there is one cyan link missing connecting the red node to others by using only links of cyan colours. The addition of this pathway (the dashed link above in the picture) would make the whole intersection a viable cluster. Notice that the 4 nodes in the lower part of the figure do constitute a viable cluster. In general, the LVC is a subset of the intersection of largest connected components.

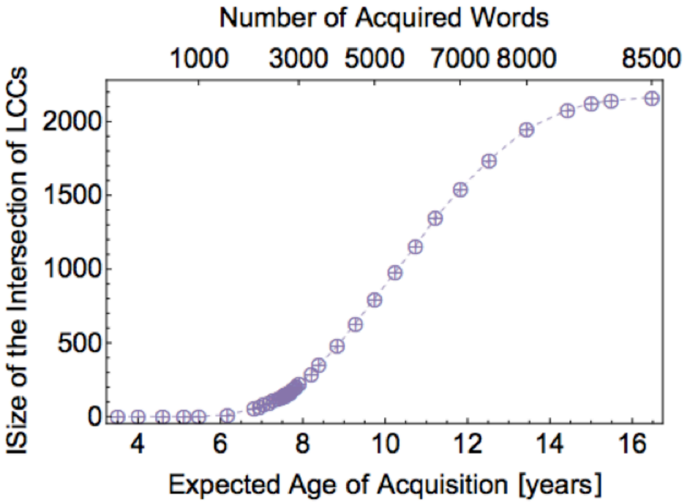
Even though the intersection of LCCs is indeed a multiplex measure, as it takes into account connectedness on different multiplex layers, it does not require simultaneous connectedness across all layers, as the LVC does instead. The LVC has therefore an additional requirement of viability on *every* layer, with “no holes” in connectedness of the individual layers. Such holes could instead present in the intersection of LCCs. This is the main different between the LVC and the intersection of LCCs: the former requires viability on every layer while the second does not. Therefore, comparison between the multiplex intersection of LCCs and the LVC points out that it is indeed viability the ultimate cause of explosive behaviour in the multiplex lexical representation..

Notice that, from a cognitive perspective, the stricter requirement of connectedness across all layers allows to navigate across all words of a viable cluster, without having to “jump” across different linguistic aspects. Hence, in terms of mental navigation, the LVC allows for a smaller number of words to be reached from each other within a given aspect of the mental lexicon while the intersection of LCCs might require jumps across different aspects of the lexicon (e.g. between semantic and phonological layers) in order to connect a larger number of concepts.

The absence of explosive behaviour in LCCs and in their intersection further supports the idea of the distinctness of the LVC emergence as presented in the main text.

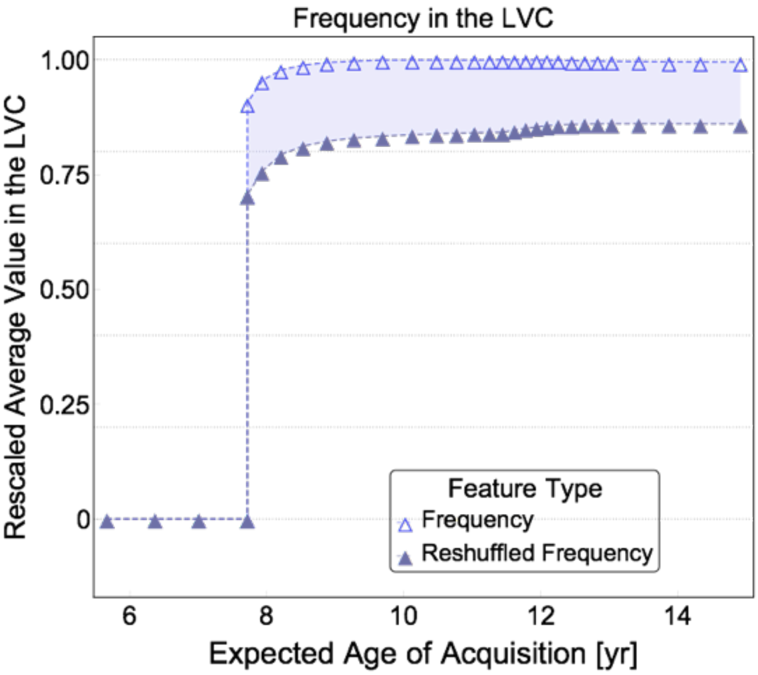
6.5.9 Frequency Gap in the Partial Reshuffling Null Model

In the main text we quantified the gap in median concreteness and polysemy of words in the LVC when partial reshuffling was applied to words in the MLR. In this section we report the results relative to partial reshuffling of word frequency, presented in Supplementary Fig. S8.



Supplementary Figure S7 : Size of the intersection of largest connected components over normative word acquisition. The aggregated multiplex network is identified with "Aggr.". Error margins indicate standard deviations and correspond to symbol sizes. Results are averaged over 200 iterations of smeared normative acquisition.

The observed gap between the frequencies in the empirical and in the partially reshuffled models is of similar size to the one observed for concreteness in the main text and it is therefore considerably smaller than the gap observed for polysemy. This suggests that polysemy has a higher influence on the evolution of the LVC compared to either concreteness or frequency of words.



Supplementary Figure S8 : Comparison of average frequency for words in the LVC with normative acquisition in the empirical data and for a partial reshuffling null model. Reshuffling node attributes results in a LVC with reduced frequency. The curves are rescaled by their maximum value and they represent averages over 200 iterations. Error margins are approximately the same size as the dots. We note significant gaps between the empirical and randomised data.

Part II

Investigating parasite spread through ecomultiplex networks

A Foreword to Part II

The next two Chapters composing Part II of this thesis originated from a scientific collaboration I started at the Santa Fe Institute Summer School on Complex Systems 2014 with the following collaborators:

- Cecilia S. Andreazzi (Fundação Oswaldo Cruz, Rio de Janeiro, RJ, 22713-375, Brazil);
- Sanja Selakovic (Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands);
- Alireza Goudarzi (RIKEN Brain Science Institute, Wako, Japan);
- Alberto Antonioni (Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain).

The next two Chapters represent papers I published with the above collaborators. In both the papers I am the main author, as I contributed to most of the technical work behind them. More in detail, I conceived, developed and simplified most of the ecomultiplex framework presented in both the papers, conceived the numerical experiments and analysed the results, conceived and developed the analytical parts, structured and wrote most the manuscripts reported here. A.A., C.S.A. and S.S. contributed to designing the ecomultiplex framework, C.S.A. and S.S. provided and cleaned the empirical data, A.A. and A.G. developed the simulation code and checked the analytical results. All the co-authors contributed to writing and reviewing the manuscripts. An older, unrevised version of the published, revised manuscript reported here in Chapter 8 was included by Sanja Selakovic at the end of her PhD thesis *Infection in ecosystems*, Utrecht University Repository (2016).

Chapter 7

A review of ecological networks

> It is interesting to contemplate a tangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent upon each other in so complex a manner, have all been produced by laws acting around us.
>
> *On the Origin of Species*
> Charles Darwin

An ecosystem is a set of animal populations living spatially close to each other and interacting with each other and with the environment such as sources of nutrients, energy and geographic landscape (Pielouet al. , 1969; May, 1973). In the last century, a great deal of effort has been put into the mathematical modelling of ecosystems (May, 1973; Pielouet al. , 1969; Hagen et al., 2012), in particular in terms of predator-prey models and robustness against animal extinction.

7.1 Single-layer ecological networks

Network science proved to be particularly powerful into investigating food webs in ecosystems (Patten, 1991; Dunne et al., 2002a,b). In a food web, nodes represent species, although individual nodes can sometime represent also collections of species, such as birds of a given family that are all predated by the same animal. Connections are directed and indicate “who-eats-who”, e.g. a predator feeding on a prey. It is common practice to have links directed from prey to predators, in order to mean the energy flowing from the prey to the feeding predator. Food webs are usually acyclic and can therefore represent a hierarchy in the energy flow within an ecosystem: on top of the hierarchy there are the so-called top predators (i.e. species that are not eaten by any other species) while on the bottom there are basic producers (i.e. species that do not feed on other animal populations), with different shades in between (i.e. species predating and being predated at the same time).

The robustness of food-webs has been extensively studied in order to address the effects of species extinction in a given ecosystem (May, 1973; Dunne et al., 2002a,b). In fact, neglecting other types of interactions, predator-prey relationships in a given ecosystem determine the evolution of population sizes $\mathbf{x}(t)$ close to the stable equilibrium of the community through the differential equation[1] $\dot{\mathbf{x}} = A\mathbf{x}$, where A is the adjacency matrix of the food-web.

Instead of empirical food-webs, May (1973) considered interaction matrices A as coming from ensembles of random networks where N species give rise to C interactions weighted with average strength s (which can be interpreted as the frequency of each interaction). The author showed that increasing either N, C or s leads to a critical point at which the probability of the ecosystem remaining stable in an equilibrium $\mathbf{x}^*$ under small perturbations sharply decreases. May's results challenged previous conjectures in ecology about increasing biodiversity and complexity in ecosystems always increasing robustness of ecosystems to fluctuations and extinctions.

However, a common criticism of May's approach is that species in food webs do not interact at random (Hagen et al., 2012; Dunne et al., 2002a; Daraganova et al., 2012; Dunne et al., 2013). Empirical food-webs tend to display a community structure (Dunne et al., 2002a), so that species interact mainly with other species in the same community rather that from outside. The presence of these clusters actually boosts robustness to perturbations.

Contrary to what happens in May's random topologies, in empirical food webs robustness increases with the density of interactions C (Dunne et al., 2002b), mainly because unstable clustered communities deteriorate more gradually than random uncorrelated links.

7.2 Precursors of multi-layer ecological network models

Predator-prey interactions are just one of the many types of ecological interactions present in ecosystems and impacting the evolution of populations, even in subtle ways. Let us consider the following example provided by Lafferty et al. (2008). Think of the African Savannah. It's not difficult to notice how lions eat zebras. Less obvious are the 51 species that can kill lions, which include a notable diversity of infectious agents (or parasites): two arthropods, two bacteria, 31 helminths, six protozoans and 10 viruses. Parasites transmit diseases within a population, thus affecting it over time. Further, multi-host parasites can get transmitted over populations of different species, as they evolved general contaminative mechanisms for different mammals at once (Pilosof et al., 2015). This example perfectly provides the idea that food-webs cannot fully address the problem of ecosystem robustness by themselves (Kéfi et al., 2015; Kéfi et al., 2016). It is only recently that ecologists and network scientists tried to couple food webs with networks of different ecological interactions (Lafferty et al., 2008; Melián et al.,

---

[1] The linear differential equation $\dot{\mathbf{x}} = A\mathbf{x}$ comes from the linearisation of nonlinear Lotka-Volterra differential equation close to the stable equilibrium point for population sizes.

2009; Pocock et al., 2012) provided their importance in spreading pathogens within ecosystems (Lafferty et al., 2006; Pelosse & Kribs-Zaleta, 2012). For instance, Pocock et al. (2012) built an ecological network including predator-prey, host-parasite, seed dispersal networks, and plant-pollinator interactions between 560 different species. No multiplex structure was considered: all the different interactions were aggregated into one network, which was used for testing robustness to species extinction via numerical simulations. In the aggregated network the authors identified keystone plants, i.e. plants whose population dynamics has substantial cascading effects across the whole network and whose preservation or restoration could have positive results for preserving biodiversity. While Pocock et al. (2012) did not distinguish between ecological interactions, Melián et al. (2009) considered separately mutualistic plant-pollinator and antagonistic plant-herbivore interactions among 390 species in the same network, representing the Donana Biological Reserve, Southern Spain. The authors did not characterise the multiplex representation rising from the two interactions but rather focused on a specific metric: The ratio of the number of mutualistic to antagonistic interactions per plant species. By means of numerical simulations for population dynamics, the authors related the high mutualistic to antagonistic ratios observed in a few plants to a higher richness of species and thus higher biodiversity in the modelled ecosystem. However, Melián et al. (2009) did not consider plant parasites in their analysis, which can heavily influence the population dynamics of plants (Pilosof et al., 2017).

Parasite spreading can be made even more subtle by the so-called vector species ("vectors" from now on). In biology a vector is an organism transmitting not the disease but rather its cause from one organism to another one. In ecology a vector can be an insect feeding on blood: the insect can get infected with the parasite via blood-exchange and transmit it to another animal, always by exchanging blood with it. Although this can greatly vary across species, parasites usually need to grow and develop for some time in their hosts before causing functional disorders or diseases. Vectors can help parasite diffusion not only by facilitating transmission but also by helping parasites to grow and develop (Roellig et al., 2009).

A first attempt in considering food webs with parasitic interactions is the work by Lafferty et al. (2006), further refined in Lafferty et al. (2008). The authors considered parasites and their host species as nodes in a simplex network originating from trophic interactions aggregated together, i.e. interactions involving energy exchanges. In fact, Lafferty et al. (2008) suggested to consider parasites as consumers, feeding on the resources made available by their hosts. The inclusion of parasitic interactions can dramatically increase the overall instability of the networked ecosystem against species extinction. However, the approach of Lafferty et al. (2006) of having parasites in a network representation only as consumers did not consider the existence of vectors in ecosystems, adding network loops among a given parasite, its vectors and its hosts.

The work by Fontaine et al. (2011) suggested to go beyond simplex networks by considering k-partite networks with nodes representing sub-networks of a given type (e.g. a food-web or a network of host-parasite interactions). In a k-partite network nodes can be categorised in k sets and links can be present only between nodes from different sets. Fontaine et al. (2011) argued on the ecological evidence motivating interactions among sub-networks of different types but limited their approach to a qualitative framework which was later fully developed into a multi-layer network model (Pilosof et al., 2017).

7.3 Multi-layer and multiplex ecological network models

To the k-partite structure of Fontaine et al. (2011), Pilosof et al. (2017) added also potential interactions among sub-networks of the same type, thus providing a general framework for modelling interactions within an ecosystem at different levels. To illustrate how multilayer networks can better account for the stability of ecosystems, the authors consider two coupled sub-networks or layers: plant/pollinators and plant/parasite interactions. Plants between the two layers are coupled via inter-layer connections. Nodes are removed progressively from the ecosystem at random. Either plants, pollinators or parasites go extinct when they get disconnected. Pilosof et al. (2017) show that parasite extinctions occur more slowly in the multilayer network rather than on its aggregate when pollinators start disappearing from the ecosystem. Thus considering multilayer network representations of an ecosystem can indeed change the qualitative conclusions about robustness of species to extinction (Pilosof et al., 2017).

To the best of our knowledge, the first framework with multiple layers of interactions for the investigation of parasites in ecosystems was the work from (Andreazzi et al., 2014). The authors investigated parasitic and trophic interactions within the Brazilian ecosystem of Serra de Canastra for the spread of a parasite causing the Chugs disease (Noireau et al., 2009). The authors found that having trophic and contaminative interactions at the same time was not always beneficial for accelerating the disease spreading across the networked ecosystem: the multiplex structure could drastically alter the infection spreading dynamics compared to using an aggregate network or the food-web only. These findings were tested more systematically within a more general model of multiplex ecological interactions discussed in Chapters 7 and 8.

Prominent contributions establishing the field of ecological multiplex networks were the works by Kéfi et al. (2015) and Kéfi et al. (2016). Kéfi et al. (2015) analysed more than 100 coexisting species in a marine ecosystem from central Chile coupling a food web structure with layers of non-trophic interactions such as competition for space and habitat provisioning. The authors showed how non-trophic interactions exhibited highly non-random structures, both alone and with respect to the topology of the food-web. In a principal component analysis, the observed trophic interactions explained about 65% of the non-trophic interactions, thus highlighting significant correlations between the food-web and other non-trophic mechanisms of interactions. Further analysis in (Kéfi et al., 2016) pointed out that these correlations guarantee higher species persistence and higher total biomass than is expected by chance alone, thus promoting a higher robustness to extinctions of species within the ecosystem.

Chapter 8

Parasite spreading in ecological multiplex networks

Empirical data collected from a few wild mammals suggested a synergy between parasitic and feeding interactions in Brazilian environments for the spreading of the Chagas disease (Rocha et al., 2013). Suitable large-scale, quantitative models for interpreting the data constituted a gap in the relevant ecological literate. In fact, although preliminary attempts at considering also parasitic interactions in food webs were performed (Lafferty et al., 2008; Fontaine et al., 2011), methodological frameworks for the theoretical and quantitative investigation of parasite spreading were missing in the literature. This gap motivated the introduction of the ecological multiplex framework introduced and tested in this Chapter[1].

---

[1] This Chapter follows the layout of the "Three-paper" PhD thesis in that it is an original, published research manuscript published as: *Stella, M., Andreazzi, C. S., Selakovic, S., Goudarzi, A., & Antonioni, A. (2016). Parasite spreading in spatial ecological multiplex networks. Journal of Complex Networks, 5(3), 486-511.* In order to better present the content, the Chapter/paper is briefly introduced by a short summary. An older, unrevised version of the published manuscript was included by Sanja Selakovic at the end of her PhD thesis *Infection in ecosystems*, Utrecht University Repository. This Chapter follows the same structure of the revised, published paper, in adherence to the guidelines of the "Three-paper" PhD thesis format. For a better presentation within the thesis, the words "paper" and "manuscript" were changed to Chapter. The Supplementary Information of the original paper is reported at the end of the Chapter rather than in an appendix for increased readability and easier consultation with the main text.

Abstract

Network ecology is a rising field of quantitative biology representing ecosystems as complex networks. A suitable example is parasite spreading: several parasites may be transmitted among their hosts through different mechanisms, each one giving rise to a network of interactions. Modelling these networked, ecological interactions at the same time is still an open challenge. We present a novel spatially embedded multiplex network framework for modelling multi-host infection spreading through multiple routes of transmission. Our model is inspired by Trypanosoma cruzi, a parasite transmitted by trophic and vectorial mechanisms. Our ecological network model is represented by a multiplex in which nodes represent species populations interacting through a food web and a parasite contaminative layer at the same time. We modelled Susceptible-Infected dynamics in two different scenarios: a simple theoretical food web and an empirical one. Our simulations in both scenarios show that the infection is more widespread when both the trophic and the contaminative interactions are considered with equal rates. This indicates that trophic and contaminative transmission may have additive effects in real ecosystems. We also find that the ratio of vectors-to-host in the community (i) crucially influences the infection spread, (ii) regulates a percolating phase transition in the rate of parasite transmission and (iii) increases the infection rate in hosts. By immunising the same fractions of predator and prey populations, we show that the multiplex topology is fundamental in outlining the role that each host species plays in parasite transmission in a given ecosystem. We also show that the multiplex models provide a richer phenomenology in terms of parasite spreading dynamics compared to more limited mono-layer models. Our work opens new challenges and provides new quantitative tools for modelling multi-channel spreading in networked systems.

8.1 Introduction

Pathogens and parasites ("parasites" hereafter) are one of the most widespread and diverse life forms (Poulin & Morand, 2014; Dobson et al., 2008). Several parasites infect multiple host species and many of these parasites may infect their host using different routes of transmission (Poulin, 2011). Multi-host parasites include many zoonoses with complex dynamics that challenge infection control and prevention efforts (Dobson, 2004). For instance, several multi-host protozoan parasites of public health concern exhibit more than one mode of transmission: *Toxoplasma gondii* can infect its hosts by fecal-oral transmission, the consumption of an infected prey, and through the placenta (Dubey, 2004); *Cryptosporidium* directly infects its hosts via sexual contact or via fecal-oral transmission (Fayer et al., 2000); *Trypanosoma cruzi* can be transmitted by insect vectors, the consumption of an infected prey, and also through the placenta (Noireau et al., 2009; Jansen et al., 2015). This complexity of host types and transmission modes challenges the development of models that account for the different sources of variation. The network approach is a promising alternative because it allows accounting for the individual, species-level and spatial sources of heterogeneity (Craft & Caillaud, 2011; Barter & T, 2016).

Contact networks can be explicitly used to understand the epidemiological consequences of complex host interaction patterns (Keeling, 2005; Meyers et al., 2005; Bansal et al., 2006; Ferrari et al., 2006; Craft et al., 2009; Dalziel et al., 2014). In a contact network, each individual is represented as a node and each contact that potentially results in transmission between two nodes is represented as a link. Interactions can also be embedded in space (Craft et al., 2009; Davis et al., 2008, 2015) where the probability of interaction between nodes may depend on the distance between them. The number of contacts of a node is called the degree of the node and is a fundamental quantity in network theory (Dalziel et al., 2014). All epidemiological models make assumptions about the underlying network of interactions, often without explicitly stating them. For example, classical mean-field models used in epidemiology assume that all the interactions have the same probability of leading to transmission (Anderson & May, 1992). Contact network models, however, mathematically formalise this intuitive concept so that epidemiological calculations can explicitly consider complex patterns of interactions (Bansal et al., 2007). A different approach consists in considering meta-population dynamics (Colizza & Vespignani, 2008), instead of individual contacts.

Recently, the recognition that real-world networks may include different types of interactions among entities prompted the development of methods that take into account the heterogeneity of interactions as well (Kivelä et al., 2014; Boccaletti et al., 2014). Examples include multi-modal transportation networks in metropolitan areas (Barthélemy, 2011; Morris & Barthelemy, 2012; Lima et al., 2015), or proteins that interact with each other according to different regulatory mechanism (Cardillo et al., 2013; Cozzo et al., 2013). Ecological systems are also characterised by multiple types of relationships among biological entities, organised and structured on different temporal and spatial scales (Kéfi et al., 2015; Kivelä et al., 2014). Different interaction types can be described as "multiplex networks" (Wasserman & Faust, 1994; Mucha et al., 2010; Battiston et al., 2014; De Domenico et al., 2013; Kéfi et al., 2015). Multiplex networks are a particular kind of multi-layer networks where the same nodes appear on all the layers but they can be connected differently on each layer. Each multiplex layer contains edges of a given type. In the context of parasites that can be transmitted over multiple transmission channels, multiplex networks can be used to include distinct mechanisms of parasite transmission (Kéfi et al., 2015). This approach encapsulates the heterogeneity in the transmission of real-world diseases and helps us understand how the interplay between different modes of transmission affects infection dynamics in an ecosystem (Buono et al., 2014; Lima et al., 2015; Salehi et al., 2015).

Descriptions of ecological multiplex networks (Kéfi et al., 2015; Kéfi et al., 2016) and studies of infection spreading over multiplex structures (Buono et al., 2014; Gomez et al., 2013; Salehi et al., 2015) have recently appeared in the literature. Previous approaches have already described the structural characteristic of food webs that include parasites (Lafferty et al., 2006) and tried to incorporate parasites in food webs using network framework (Lafferty et al., 2008). The effect of multiple hosts on parasite spreading dynamics have also been explored in the context of disease risk (Keesing et al., 2006), disease emergence in a target host (Fenton & Pedersen, 2005), parasite sharing and potential transmission pathways (Pilosof et al., 2015) and also in a multilayer network exploring cross-species transmission (within and between host species) (Pilosof et al., 2015). However, the consideration of real ecological scenarios in the analysis of parasite spreading through multiple transmission mechanisms is still an open problem. We propose a spatial multiplex-based framework to model multi-host parasite transmission through multiple transmission mechanisms. In this framework, each transmission mechanism can be represented in a different layer of the multiplex network structure. Our model is inspired by the complex ecology of *Trypanosoma cruzi* (Kinetoplastida: Trypanosomatidae) in its multiple host community. *T. cruzi* is a relevant example of a multi-host parasite and in humans it causes the Chagas disease, a serious infection affecting 6-9 million people (Jansen et al., 2015). The main infection route to humans involves the insect vectors (triatomine kissing bugs), but oral transmission is also recurrent (Shikanai-Yasuda & Carvalho, 2012). Vectors get infected when consuming blood meals from an infected host, while host infection occurs through the contact of vector's faeces and the biting wound or mucosa (stercorarian transmission). In sylvatic hosts the stercorarian transmission may occur when the animal scratches the bite and inadvertently rubs the parasite-contaminated matter into the lesion (Kribs-Zaleta, 2006). Infection by the oral route occurs when a mammal host ingests infected triatomine faeces, food contaminated with the parasite or by preying on infected vectors or mammals (Jansen et al., 2015).

Preliminary studies (Kribs-Zaleta, 2006, 2010b; Pelosse & Kribs-Zaleta, 2012) used mean-field methods to model *T. cruzi* transmission among its hosts and vectors. Their results indicate that in a fully connected network with no explicit spatial structure, vectorial and oral transmission effects are additive in maintaining and furthering the spread of the infection (Kribs-Zaleta, 2006). We use a Susceptible-Infected (SI) model to describe parasite transmission dynamics in spatially embedded multiplex networks. The multiplex framework can help us understand how infection spread is related to different ecological interactions and what is the epidemiological importance of vectors and hosts in different ecological scenarios. We first investigate the parasite spreading across aggregated parasite-host and trophic interactions. In order to measure the influence of the spatial embedding, we contrast the behaviour of a non-spatial model against one where nodes are embedded in space. We then study a reference spatial multiplex network in order to understand the interplay between the multiplex structure and epidemiological dynamics. In the vectorial transmission layer, vectors are contaminated after interacting with infected hosts and transmit the parasite when interacting with non-infected hosts. In the trophic transmission layer hosts acquire the parasite after feeding on infected vector or host. Finally, we use empirical data of a local *T. cruzi* host community, the Serra da Canastra ecosystem (Rocha et al., 2013), to model the dynamics of *T. cruzi* multiple transmission routes on its multiple hosts.

With the multiplex framework we aim to understand the effect of multiplex topology and the relative importance of vectorial and trophic transmission for parasite spreading dynamics. We use multiplex cartography (Battiston et al., 2014) to characterise species structural importance in the network and compare scenarios with different relative frequency of vectors. We then explore the speed of parasite spreading depending on the importance of vectorial and trophic transmission in scenarios with different frequency of vectors. Finally, we explore the effect of species structural importance on parasite spreading by simulating immunisation experiments.

8.2 Methods

We model a set of N populations interacting within an ecosystem via a network framework. Our aim is to model the diffusion of a multi-host parasite within the ecosystem. Nodes represent populations and they have identities, i.e. their species types (predator, prey, and vector). We denote with $S = \{s_k\}_{k=1}^{s}$ the set of all the s species types. Each node in the network is of a given species type s_k with frequency f_k, normalised such that $\sum_{k=1}^{s} f_k = 1$.

Given that we do not have enough information about the individual-level patterns of interactions, we will consider the food-webs in terms of interacting populations. We consider nodes as populations that follow the same formalism of individual-based dynamics. Our approach is based on the following assumptions: (i) we consider that the parasite transmission is fast and that all the individuals within a population instantaneously gets infected once transmission occurs (in other words, we do not consider meta-population dynamics such as considering parasite spreading within the population and dispersal among populations (Colizza & Vespignani, 2008)); (ii) we consider the parasite spreading happening at a much faster rate than any birth-death dynamics (which we do not consider).

We assume that individuals from populations can disperse across the system and potentially interact with other populations, according to a *dispersal layer*. The dispersal layer is an undirected graph with adjacency matrix D, so that $d_{ij} = d_{ji} = 1$ if population i can interact with j and vice-versa. In the following subsections, we define the topology of the dispersal layer as being either an Erdös-Rényi random graph or a random geometric graph. The main difference between the two is that the latter includes the notion that only spatially close enough populations can interact with each other (since on random geometric graphs nodes are embedded in space and linked if closer than a certain threshold distance ρ).

In our model, population interaction can potentially give rise to either (i) trophic interactions (a given species feeding on another one) or (ii) contaminative interactions (a given species of host getting in touch with vectors and transmitting the parasite through blood exchanges). Considering only trophic (or contaminative) interactions gives rise to the trophic (or vectorial) layer. Alternatively, considering both interactions together gives rise to an aggregated layer. A visualisation of the dispersal, trophic and contaminative layers is provided in Figure 8.1.

Transmission on a given network layer are allowed according to node identities $\{s_k\}$ and are defined according to the corresponding $s \times s$ *interaction matrices*, T for the trophic layer, V for the vectorial layer and $A = T \oplus V$ for the aggregated layer, where $\oplus$ indicates the Boolean OR function. There is no direct interaction between populations of the same species type because there is no cannibalism in the trophic layer and also no parasite transmission among vectors in the vectorial layer. This means the main diagonal of all interaction matrices are all 0s. The sifting of the dispersal layer through either T or V or A produces s-partite graphs, i.e. there are no edges between nodes of the same species types. We notice that sifted trophic interactions give rise to a directed network layer while we obtain an undirected vectorial layer from allowed contaminative interactions.

Providing the collection of species types S, the topology of the dispersal layer D, choosing if considering trophic and vectorial layers as separate or aggregated, and defining the corresponding interaction matrices fully determines the model. We explore the following models, enlisted in order of presentation:

- a random graph as dispersal layer, with 3 species types and aggregated interactions, called Random Aggregated Network (RAN);
- a random geometric graph as dispersal layer, with 3 species types and aggregated interactions, called Spatial Aggregated Network (SAN);
- a random geometric graph as dispersal layer, with 3 species types, interactions kept separate across a 2-layer multiplex structure, called Spatial Multiplex Network (SMN);

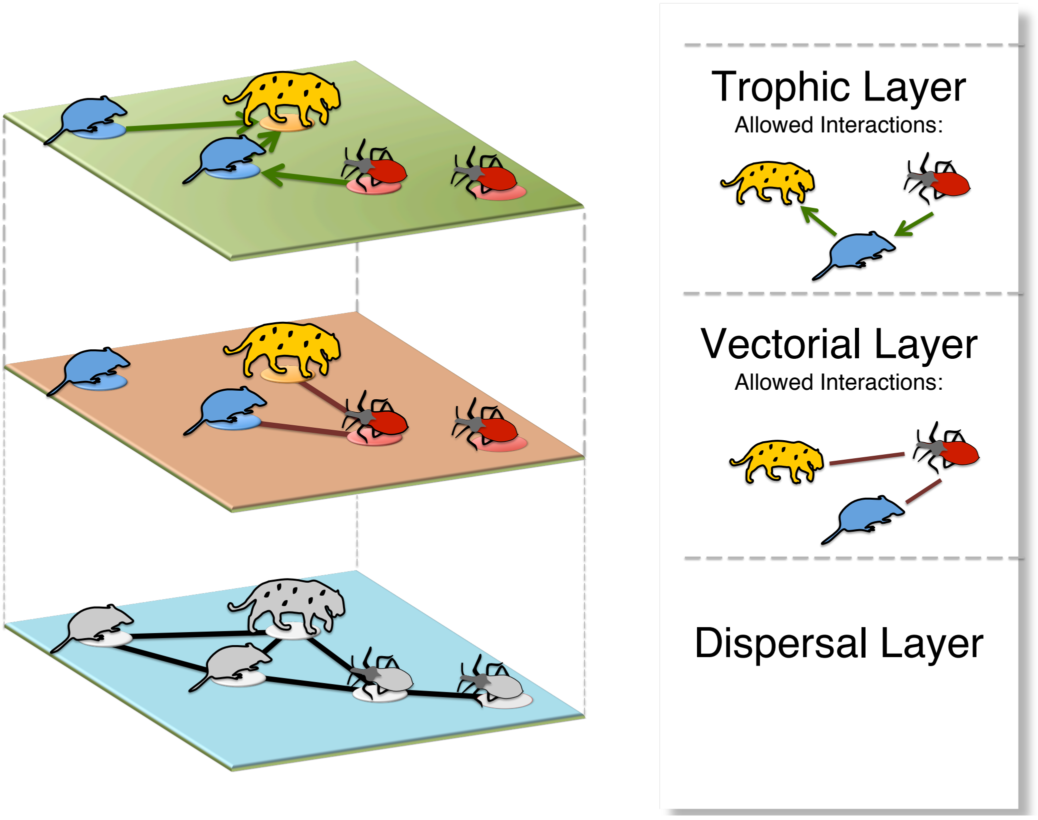


Figure 8.1: Visual representation of our model over the three layers: a trophic layer, a vectorial layer, and their underlying dispersal layer. Nodes are relative to the three-species example and they are drawn according to their species types, e.g. “predator", “prey" and “vector". Trophic and vectorial layers allow only for specific interactions to be present within the system, according to the species types involved in them. For instance, the allowed interactions in the three-species model are reported on the right. The parasite can spread on both such layers. When a node gets infected in one layer it gets infected on all the others as well. While the dispersal layer induces the other two, it is only the trophic and the vectorial layers that constitute our ecological multiplex networks.

- a random geometric graph model, with 20 species, interactions kept separate across a 2-layer multiplex structure according to ecological empirical interactions. This model is called Spatial Ecological Multiplex Network (SEMN).

We considered both smaller ($N = 1,000$ nodes) and larger networks ($N = 10,000$ nodes) with the same average degree. While the results obtained in both cases were robust to the network size change, the networks with $N = 10,000$ nodes displayed less finite-size effects. Therefore, in the following we present simulation and analytic results for networked ecosystems made of $N = 10,000$ nodes. The average degree of considered networks has been tuned in order to obtain connected dispersal layers, in which there is at least one path connecting each pair of nodes. This minimises statistical biases due to disconnectedness of a non-negligible fraction of populations.

8.2.1 Random aggregated network model

In the random aggregated network model (RAN) nodes have $s = 3$ possible identities, $S = s_1, s_2, s_3 =$ predator,prey,vector with species frequencies f_1, f_2, f_3 respectively. Herbivorous mammals are in general more abundant than carnivorous (Damuth, 1981) and for sake of simplicity we assume prey populations being double as frequent as predator populations, $f_2 = 2f_1$. Therefore, given that $f_1 + f_2 + f_3 = 1$, one obtains that $f_1 = (1 - f_3)/3$ and $f_2 = 2(1 - f_3)/3$, thus leaving the vector frequency $f_3 = f_v$ as a free parameter of the model. In this model the dispersal layer has the topology of an Erdös-Rényi with probability p_{ER}. Therefore, no space is included in the RAN model. In order to consider fully connected graphs in our simulations and to reduce the effects of degree heterogeneity we fixed a p_{ER} giving rise to networks with average degree $\langle k_{ER}\rangle = p_{ER} \cdot (N - 1) \approx 28.27$. The RAN model sifts interactions among predator, prey and vector populations from the dispersal layer according to the interaction matrix A defined as:

$$A = T \oplus V = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \tag{8.1}$$

For instance, $t_{21} = 1$ means that $s_2 =$ prey populations are eaten by $s_1 =$ predator populations. Notice that allowed interaction in T are directed (from the eater to the eaten, as usual in food-webs (Dunne et al., 2002a)) while they are undirected in V, since they represent ecological exchanges of infected fluids between the host and the vector species (Roque et al., 2013)). The above sifting creates the aggregated single layer of the model, where trophic and contaminative interactions are combined and where parasite diffusion occurs.

8.2.2 Spatial aggregated network model

In the spatial aggregated network model (SAN) the dispersal layer is a random geometric graph (RGG). Therefore, populations are embedded in a space. Nodes are scattered uniformly at random within the 2D space $\Omega = [0, 1]^2$ with periodic boundary conditions, i.e. a toroidal space. As known from previous works (Sattenspiel, 2009), the average degree of an RGG is

$\langle k_{RGG} \rangle = \pi N \rho^2$. For the sake of comparisons with the RAN model, we chose $\rho = 0.03$, thus having $\langle k_{RGG} \rangle = \langle k_{ER} \rangle = 28.27$. The interaction matrix A sifting the only aggregated network layer is the same as in the RAN model. Also species types are distributed as in the RAN model.

8.2.3 Spatial multiplex network model

In the spatial multiplex network model (SMN) the dispersal layer is a random geometric graph (RGG) with nodes spatially embedded and species types distributed as in the SAN model. However, we keep trophic and contaminative interactions as distinct on two separate layers. These structured interactions give rise to a multiplex network (Kivelä et al., 2014; Kéfi et al., 2015; Kéfi et al., 2016; Boccaletti et al., 2014), where populations are replicated across both layers and no explicit inter-layer edges are considered (De Domenico et al., 2013). The interaction matrices sifting the trophic and the vectorial layer are respectively T and V, as defined above in Equation 8.1. A multiplex network visualisation of the SMN model is provided in Figure 8.1.

8.2.4 Spatial ecological multiplex network model

In our last model, the spatial ecological multiplex network (SEMN), the dispersal layer is a random geometric graph (RGG), as in the SAN model. Also, trophic and contaminative interactions are kept separate analogously to the SMN model. In SEMN we used empirical ecological data within the model (Rocha et al., 2013). Specifically, we use data from an epidemiological study of *T. cruzi* infection in wild hosts in Southeast Brazil (Rocha et al., 2013) to estimate the trophic and vectorial interaction matrices T_{eco} and V_{eco} (see Supplementary Information), considering a total of 20 species. For the trophic interaction matrix T_{eco}, we build a qualitative potential food-web based on the animals diets (Carvalho & Santos, 2012; Cavalcanti, 2010; Amboni, 2007; Reis et al., 2006). As there was no species-level classification of the biological vectors present in the area, we considered the vectors as one single species type. We use species prevalence to estimate contaminative interactions in V_{eco} (Rocha et al., 2013). We assume that positive parasitological diagnostics for *T. cruzi* could be used as a proxy for vectorial transmission, since only individuals with positive parasitaemia (i.e. with high parasite loads in their blood) are able to transmit the parasite (Jansen et al., 2015). The vectorial layer was constructed based on the assumption that species with positive prevalence in hemocultive transmit the parasite to vectors and that species with positive prevalence in serology can be infected from vectors. The SEMN model has a total of 20 species types: $a = 7$ predators, $b = 12$ prey and 1 vector species. As in the previous models, we assumed that prey populations have double the frequency of predator populations (see RAN model). We considered all the predator and prey species populations having identical frequencies f_{pred} and f_{prey} respectively, such that:

$$a f_{pred} + b f_{prey} + f_v = 1 \rightarrow f_{prey} = 2\frac{1 - f_v}{a + 2b} = 2 f_{pred}. \tag{8.2}$$

Therefore, by tuning f_v we change also the frequency of predator and prey populations. The SEMN model is the most realistic one of this study since it takes into account spatial embedding, multiplex structure and empirical ecological data.

8.2.5 Parasite transmission dynamics

To simulate the parasite transmission dynamics a node, i.e. a population of a given species type can be either susceptible or infected. We start the simulation by infecting a fraction $\phi_0 = 0.28\%$ of all populations. In the RAN model we infect one node at random and let the infection spread along a random walk on the dispersal layer. We start measuring the infection dynamics after $N\phi_0$ nodes are infected. Similarly, in the other three spatial models, we infect all the nodes in a random circle of radius $r_0 = 0.03$, that is, $\pi N r_0^2 \approx 28.2$ populations become infected at the beginning, on average (a sensitivity analysis proves that the results presented in the following sections are robust up to 5% of initially infected populations). Subsequently, the parasite spreading evolves in SMN and SEMN models as follows:

1. A random node i is chosen together with one of its neighbours j on the dispersal layer.
2. The vectorial layer is chosen to be considered for the parasite transmission with probability p_v, which is a measure of the vectorial layer importance. Step 3 is then performed when the vectorial layer is chosen. Otherwise, step 4 takes place.
3. If node i is infected and the edge (i, j) exists in the vectorial layer, node j becomes infected as well (vectorial layer parasite transmission).
4. If node i is infected and the edge (i, j) exists in the trophic layer, node j becomes infected as well (trophic layer parasite transmission).
5. Steps 1-4 are repeated $N = 10^4$ times per each time step, i.e. an average of 1 update per node per time step, until T_{max} time steps are reached.

For RAN and SAN models parasite transmission occurs only on the aggregate layer without considering steps 2, 3 and 4. This is equivalent in treating contaminative and trophic interactions in an aggregate, unweighted way. Each population can be randomly chosen at each time step and at the end of the transmission process every node is chosen once, on average. This parasite transmission model is equivalent to an SI model with contact rate $\beta = 1$, where β is the probability for an individual to become infected when exposed to the disease (Sattenspiel, 2009). For the sake of simplicity, we assume $\beta = 1$ in both the trophic and vectorial layers and across all the species. Our assumption leads to the disease firstly spreading across the geodesic paths of the multiplex topology (Sattenspiel, 2009; Jeub et al., 2015) so that our infection process depends solely on the multiplex network structure. Notice that a more complicated model with two different β, one for each layer, would still be expected to reproduce a similar phenomenology to the one reported in the following (with one β only). This is because our infection dynamics is an SI model and because even the simpler model with one β only still potentially weights differently each layer through p_v.

8.2.6 Model parameter values

Let us summarise the main parameters of our models and relative values. In this study we consider networks of $N = 10,000$ populations (nodes) and average degree $\langle k \rangle = 28.27$ for the dispersal layer ($p_{ER} = \langle k \rangle / (N - 1)$ for random graphs, $\rho = 0.03$ for RGGs). We chose these parameter values in order to consider fully connected multiplex networks. Let us underline that we consider a multiplex connected component as the set of all nodes that can be reached from each other by considering all edge types of a node (De Domenico et al., 2014). Given that we have directed edges in the trophic layer, we have to consider the notion of *strongly connected component* on the multiplex topology, i.e. a set of nodes that can be reached from each other considering oriented paths along directed edges of any colour.

The maximum number of time steps $T_{max} = 10^4$ has been numerically tuned in order to let the system reach equilibrium. Each time step considers $N = 10^4$ updates for the parasite spreading dynamics, i.e. an average of 1 update per node per time step. The frequency of vector populations f_v is a free parameter of the model, together with the vectorial layer importance p_v, i.e. the probability for the parasite to spread along the vectorial layer, in the SMN and SEMN models.

8.2.7 Immunisation

In order to investigate the role played by predators and prey populations in spreading the parasite we focus on multiplex models (SMN and SEMN models). Using immunisation simulations we study the dynamics of parasite spreading when the same number of either predator or prey populations have been immunised. An immune node is not susceptible to the parasite. The number of immune nodes is determined per species by specifying the probability of immunisation π_k for each species $k \in S$. To perform the immunisation, populations of species s_k are randomly chosen with probability π_k and are set to be immune.

We consider two immunisation scenarios to investigate the relative role that predator or prey populations have in spreading the parasite. In the first scenario only prey populations are immunised while in the second scenario only predator populations are immunised. For simplicity, the π_k values for all prey and predator populations are set uniformly, however they are chosen in order to immunise the same total number of predators and the same total number of prey. From an ecological point of view, the immunisation simulations answer the following question: given the fictional possibility of vaccinating a limited number $\phi \ll N$ of populations against the parasite, is it more efficient to immunise predator populations or prey ones in order to hinder the parasite spreading?

8.2.8 Multiplex cartography

A multiplex cartography visually represents the role played by a given node across different layers according to its topological features (Guimerá & Amaral, 2005; Battiston et al., 2014). In this way, multiplex cartography becomes a rather simple yet powerful network metric providing information on the topological patterns of nodes across the multiplex structure. We chose it compared to other multiplex measures because of its simplicity, its powerfulness and its appealing analogy with maps. We build on previous literature (De Domenico et al., 2013; Battiston et al., 2014) by considering a cartography based on the following two measures: the multidegree or overlapping degree o_i and the participation coefficient P_i of node i. As in (Battiston et al., 2014; De Domenico et al., 2013), the multidegree o_i is defined as the sum of all the degrees of node i across the M multiplex layers:

$$o_i = \sum_{\alpha} k_i^{(\alpha)}. \tag{8.3}$$

where $k_i^{(\alpha)}$ is the degree of node i in the layer $\alpha \in \{1, ..., M\}$. The overlapping degree o_i represents a proxy of the overall local centrality that a node has within the multiplex network. Differently from Battiston et al. (2014), we consider o_i rather than its standardised counterpart $z_i = \frac{(o_i - \langle o_i \rangle)}{\sigma(o_i)}$ because our multiplex networks do not display Gaussian-like multidegree distributions. We consider hubs in our multiplex networks as those nodes being in the 95th percentile of the multidegree distribution.

The distribution of the connections over the different layers can be expressed via the participation coefficient P_i of node i:

$$P_i = \frac{M}{M-1} \left[1 - \sum_{\alpha=1}^{M} \left(\frac{k_i^{(\alpha)}}{o_i} \right)^2 \right]. \tag{8.4}$$

P_i ranges between 0 (for nodes that concentrate all their connections in one level only) and 1 (for nodes that distribute connections over all the M layers uniformly). In the following, we visualise our multiplex network cartography by clustering together individual points (each one referring to a given node) into 2D bins, thus obtaining a 2D histogram resembling a heat-map. The binned quantities are the overlapping degree on the y-axis and the participation coefficient on the x-axis.

8.2.9 Infection measures

On a macroscopic scale, we investigate parasite spreading by computing the *global infection time*, defined as the time step at which the largest (in node size) weakly connected component of the multiplex network is infected. Alternatively, the infection time indicates the time step t_{inf} at which the disease infects most of the nodes within the network. If $R(t) = N_{inf}(t)/N$ if the ratio of infected populations/nodes at time t, then $\mathrm{Max}_t(R(t)) = R(t_{inf})$.

Infection times represent a global, macroscopic statistics of the parasite spreading. To analyse the evolution of transmission in more detail we use the *parasite ratio increase* $\Delta R(t) := R(t+1) - R(t)$, i.e. the increase of the ratio of infected populations in one time step. The $\Delta R(t)$ is a measure for the rate at which the parasite is spreading within the multiplex network.

In order to capture the spatial features of our SMN and SEMN models we measure also $\langle \lambda \rangle$ defined as the average distance of the infected nodes from the centre of the embedding square $\Omega := [0,1]^2$ (where the infection originates). Given our assumption of uniform spreading of species populations within Ω, it is relatively straightforward to compute an upper bound $\langle \lambda \rangle^*$ for $\langle \lambda \rangle$ as:

$$\langle \lambda \rangle^* = \iint_0^1 \sqrt{(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2} dxdy \approx 0.3826. \tag{8.5}$$

$\langle\lambda\rangle^*$ represents the maximum average distance of infected populations from the centre of the embedding space (also the origin of the infection).

8.3 Results

Our results focus on: (i) highlighting the role of spatial correlations on the parasite spreading dynamics, (ii) assessing the differences between aggregated and multiplex models, (iii) highlighting the topological features of our models through cartography (Battiston et al., 2014) while relating them to parasite spreading at different values for the vector frequency f_v and importance of vectorial transmission p_v, and (iv) quantifying how different species promote or not parasite spreading by means of immunisation simulations. We first report the results concerning the aggregate models (RAN and SAN), then the three-species reference one (SMN) and the spatial ecological multiplex network (SEMN) as last. In particular, we show that: (i) the presence of spatial correlations slows down the parasite spreading in the SAN model compared to the RAN one, (ii) the multiplex structure deeply influences the parasite spreading dynamics in both SMN and SEMN models, (iii) the vector frequency determines a percolation threshold in the parasite spreading rate over the whole networked ecosystem in both SMN and SEMN models, (iv) a higher biodiversity in the SEMN model significantly modifies the infection times in similarly sized ecosystems from the SMN model and (v) prey and predator populations play different roles in promoting the parasite spreading in the empirical SEMN scenario.

8.3.1 Aggregate network models: the role of space

Comparing the results of the aggregate models RAN and SAN provides quantitative information about the role played by space. In Figure 8.3.1 (a) we compare the ratio of infected nodes over time for the RAN and SAN models by means of simulations and analytical results. Assuming a mean-field approximation, where every population can be potentially infected by any other one in the system, it is possible to write down the following equations for the infection dynamics:

$$\dot{n}_1 = f_1 N \left(\frac{f_1 N - n_1}{N}\right)\left(\frac{n_2}{N} + \frac{n_3}{N}\right) \tag{8.6}$$

$$\dot{n}_2 = f_2 N \left(\frac{f_2 N - n_2}{N}\right)\frac{n_3}{N} \tag{8.7}$$

$$\dot{n}_3 = f_3 N \left(\frac{f_3 N - n_3}{N}\right)\left(\frac{n_1}{N} + \frac{n_2}{N}\right) \tag{8.8}$$

where $n_k = n_k(t)$ is the number of infected nodes of species type $k \in 1, 2, 3$ at time t. Each equation considers how a given susceptible species population can be potentially infected in the model through its edges with other species population types. For instance, let us consider the infection dynamics of predator populations ($k = 1$). At time t, the probability of finding a susceptible predator population in the system is $(Nf_1 - n_1)/N$. However, in all models which

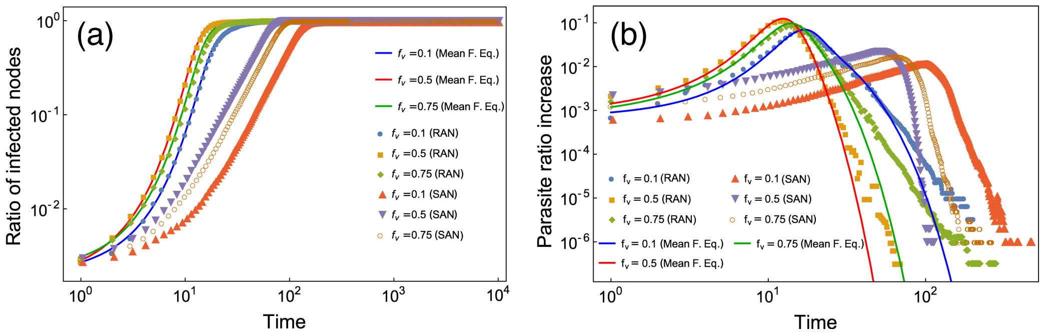


Figure 8.2: (a): Ratio of infected nodes over time for the random aggregate network (RAN) and the spatial aggregate network (SAN) models, at different frequencies f_v of vector populations in the system. (b): Parasite ratio increase of infected nodes over time for the random aggregate network (RAN) and the spatial aggregate network (SAN) models, at different frequencies f_v of vector populations in the system.

consider 3 species, a susceptible predator population can receive the parasite infection either from feeding on infected prey populations (the probability of sampling one is equal to n_2/N) or from being contaminated by an infected vector population (the probability of sampling one is equal to n_3/N). Analogous reasoning leads to the Equations 8.7 and 8.8. Notice that having directed edges leads to prey getting infected only through infected vectors in Equation 8.7.

Even though the mean field approximation does not consider the networked structure of the underlying dispersal layer, Figure 8.3.1 (a) shows that analytical results from the mean field equations reasonably approximate simulation results on ER random graph topologies (in RAN) at different vector frequencies f_v. Theory and simulations agree in indicating that the infection spreading dynamics reaches its maximum value around 20 time steps in the RAN model. Increasing the vector frequency does not always lead to the infection dynamics reaching its maximum value in less time steps. In fact, when we have $f_v = 0.1$ the ratio of infected nodes reaches its maximum value later than in the $f_v = 0.5$ case, i.e. the global infection time decreases. However, further increasing vector frequency from $f_v = 0.5$ to $f_v = 0.75$ leads to an increase rather than to a reduction in the global infection time.

For completeness, we also show in Figure 8.3.1 (b) the relative parasite ratio increases indicating the rate of parasite diffusion over time. We notice that the RAN model always displays a peak over time in the parasite ratio increases. This means that the parasite diffusion initially accelerates and it later slows down since susceptible populations become rarer in the system. Simulations and analytical results for the RAN model also agree in the appearing ordering of these peaks. Here, reaching earlier the maximum ratio of infected nodes means reaching earlier the peak in the parasite ratio increase. This is because we assume that populations of the same species type do not interact with each other (i.e. our networks are $k-$partite graphs). Since infection must always pass through a vector-host-vector path in order to infect other vectors, adding too many vector populations is detrimental for the global infection time.

In the SAN model, when the dispersal layer changes from an ER random graph to an RGG, the infection reaches its maximum spread at a much later stage (around 100 time steps). We observe that inserting spatial correlations makes the mean field approximation unreliable in describing the simulation results. This is due to the spatial embedding giving rise to non-negligible correlations among nodes.

Parasite ratio increases reveal that the RAN model displays also a faster infection spreading dynamics when compared to its spatial counterpart, the SAN model. Interestingly, both the aggregated models display a peak in the evolution of the parasite ratio increases. Overall, the addition of space increases the global infection time and it reduces the parasite spreading rate.

8.3.2 Spatial multiplex network model: the role of trophic and contaminative interactions

The 3-species reference model (SMN) consists of the simplest epidemiological scenario for the multiplex transmission. It is based on the simplest trophic chain in which vectors are consumed by prey populations and prey are consumed by predator populations. In the vectorial layer the vectors contaminate both prey and predator populations, see also Figure 8.1.

In Figures 8.3 (a)-(d), the multiplex cartographies highlight the degree centrality and participation coefficient of each species type at different vector frequencies f_v. Individual nodes are binned according to colour-coded two dimensional tiles so that the resulting plot resembles a heat-map.

When vector populations are rare in the system ($f_v = 0.01$, Figure 8.3 (a)), predators' participation coefficient is low. This means that predators interactions are concentrated mostly in the trophic layer and predator populations interact mostly with prey populations. Prey populations show a broader range of participation and this indicates that prey interact with predators and vectors on both layers. Vector populations have the highest participation coefficient and are hubs in the multiplex, since their links are uniformly distributed between both layers.

When f_v goes from 0.1, Figure 8.3 (b), to 0.25, Figure 8.3(c), vector populations show a broader range of participation coefficients indicating that their connections are distributed on both layers. Similar behaviour is reported when $f_v = 0.5$ (plot not presented). At vector frequency $f_v = 0.75$, vector populations are the most frequent in the system and each species type occupies a different region in the cartography (Figure 8.3 (d)). Thus, we have: (i) prey populations linked to vectors on both trophic and vectorial layers becoming almost truly multiplex hubs (participation coefficient value close to one and high multidegree), (ii) predator populations with a broad range of participation coefficients, (iii) vector populations with a broader range of participations coefficients but loosely connected to other populations because vectors do not interact with each other.

The multiplex structure in the SMN model allows for the infection to spread either on the vectorial layer (with probability p_v) or on the trophic layer (with probability $1 - p_v$) at each time step (see section 2.5). This interplay leads to the global infection time potentially being a function of the vectorial layer importance p_v. As reported in Figure 8.3.2 (b), when vector frequency is $f_v = 0.1$, the global infection time has its minimum for $0.4 < p_v < 0.8$. Hence, when the parasite spreads across both trophic and contaminative edges with roughly the same probability, its spreading on the whole multiplex networked ecosystem requires less time. Since the trophic layer in the SMN model is not fully connected and thus the infection cannot reach the entire network, we do not show infection times for $p_v = 0$. On the other hand, we do not

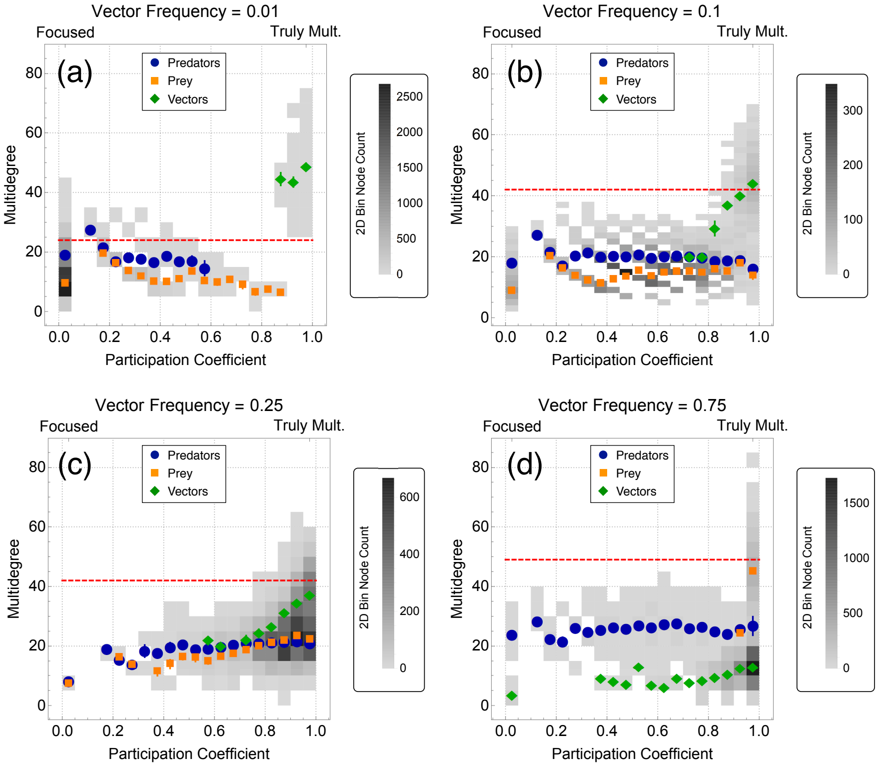


Figure 8.3: Cartographies as 2D histograms for the SMN model for vector frequency $f_v = 0.01$ (a), $f_v = 0.1$ (b), $f_v = 0.25$ (c), and $f_v = 0.75$ (d). The 10000 multiplex nodes are binned in 2D bins, according to their coordinates in the cartography. Bins are colour-coded according to the number of points falling within them: more coloured tiles have the most nodes in them. Coloured dots identify individual species: predators (blue), prey (orange) and vectors (green). Nodes falling above the horizontal red line have degrees above the 95th percentile in the multidegree distribution and they are therefore considered hubs. Error bars represent standard error of the mean.

consider the $p_v = 1$ case in order to always consider the food-web while focusing on the multiplex structure.

Increasing the frequency of vector populations does not accelerate parasite spreading in the multiplex network and the faster spreading occurs when $f_v = 0.5$ (8.3.2). The infection time decreases monotonically with the increase of vectorial layer importance p_v when $f_v = 0.25, 0.5$ or 0.75, but this pattern was not observed when $f_v = 0.1$. This is related to the topology of the allowed interactions in the SMN vectorial layer. In SMN the vectorial layer is undirected and vector populations are connected to both predator and prey populations. The trophic layer has directed interactions and parasite transmission requires at least two steps to spread from vector to predator populations. These topological features of the SMN model enables a faster parasite transmission on the vectorial layer rather than on the trophic layer. However, the frequency of different species types also influences parasite transmission in the model. Increasing the vector frequency from $f_v = 0.1$ to 0.25 or even up to 0.5 leads to an overall decrease of the infection times, depending on p_v. This trend changes when vectors are the most frequent species type in the system ($f_v = 0.75$). When the majority of nodes are vector populations the speed of parasite spreading increases in relation to $f_v = 0.5$ because vectors are not directly connected in neither of the layers. Therefore, a smaller number of predator and prey populations constraints

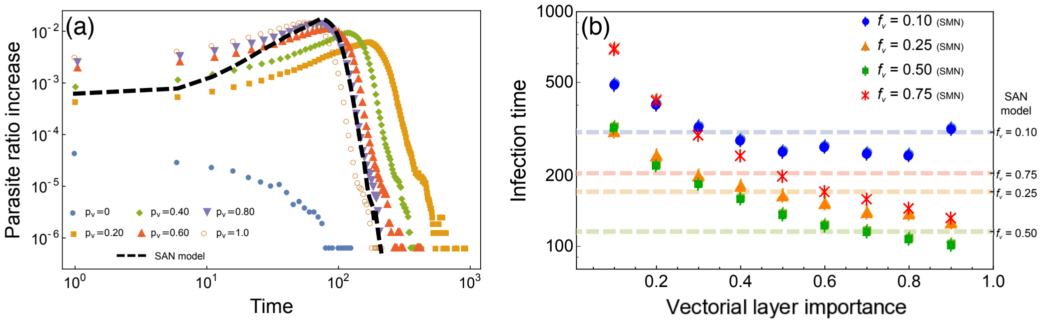


Figure 8.4: (a): global infection rate over time for $f_v = 0.75$ expressing the diffusion speed of the disease over time for SMN model. A qualitatively similar behaviour was observed also for other vector frequencies. (b): global infection time versus vectorial layer importance p_v for different vector frequencies in the SMN model. Dotted lines represent infection time in the SAN model for different vector frequencies. Results in both plots are averages of 100 repetitions.

parasite transmission to vectors. In Figure 8.3.2 (b) we also show the infection time for the SAN model represented as dotted lines for the different vector frequencies. We remember that in the SAN model there is only one aggregated layer where the infection spreads, thus all edges have the same importance. Comparing the infection time of the SAN and SMN models highlights the effect of multiplex structure for parasite spreading dynamics. Independently on the vector frequency, tuning the parasite spreading across trophic and contaminative interactions changes the infection time with respect to the aggregate case.

The speed of parasite spreading across the multiplex structure also reveals interesting patterns. As reported in Figure 8.3.2 (a) for $f_v = 0.75$, when $p_v > 0$ the parasite transmission initially accelerates within the system ($t < 100$). This behaviour is somehow similar to the one already observed in the SAN model (see Figure 8.3.1 and the black line in Figure 8.3.2 (a)). On the other hand, when the infection spreads only on the trophic layer ($p_v = 0$) a qualitatively different behaviour is observed, with no acceleration phase. This is because of the trophic layer topology (see T in the Methods section): the parasite can spread only from vectors to prey and from prey to predator populations. As the infection spreads, it becomes increasingly difficult to infect more populations over time. Vector populations which are susceptible at the beginning will never be infected. The aggregated model (SAN) does not capture this trend since it includes trophic and contaminative interactions mixed together. We observed a consistent behaviour for other vector frequencies $f_v \neq 0.75$. The only difference was in the order of the peaks of parasite spreading rate: the higher p_v the sooner the peak is reached when $f_v > 0.2$. We conjecture that this is because, in environments with many vector populations, the parasite spreads at a faster rate with respect to the trophic layer, so that increasing p_V accelerates the parasite spreading.

We also investigated the infection dynamics for very small values of vector frequencies (Figure 8.3.2). Simulations indicate that the SMN model displays a critical threshold in the emergence of pandemics around $f_v \approx 0.02$. Very small variations in the abundance of vector populations within the simulated ecosystem leads to dramatic changes in the ratio of infected populations after a suitably long relaxation time of $10,000$ time steps (Figure 8.3.2). By simulating larger ecosystems for $N = 25,000, 50,000, 100,000$ and $150,000$, we extrapolated the scaling behaviour of the critical threshold of vector frequency f_v. Simulation results suggest that the threshold does indeed not vanish in the thermodynamic limit (i.e. $N \to \infty$) but it is rather close to

$f_v \approx 0.02$ and lower bounded by the value $f_v = 0.019$. We conjecture that this critical transition is due to vector populations being fundamental in infecting prey populations. Considering the sifting matrices T and V, prey populations can be infected only by interacting with infected vector populations. When vectors are very rare in the system, prey populations (that are quite frequent in the system) get infected at a much slower rate. This bottle-neck translates into a phase transition in the infection rate. Our simulations show that the vectorial layer importance p_v slightly shifts the critical threshold of the phase transition, which occurs across all the different values of p_v (for $p_v = 0$ or $p_v = 1$ plots not reported for clarity). This phase transition marks the beginning of a distinct “phase" of the model ($f_v > 0.02$), for which the parasite percolates throughout the whole system at a faster rate, even when vector frequencies are low. Notice that when $0.02 < f_v < 0.1$, vector populations are multiplex hubs (see (a) and (b) in Figure 8.3), therefore they promote the parasite spreading on both the SMN layers.

As indicated by the grey area in Figure 8.3.2, the mean distance of infected nodes $\langle\lambda\rangle$ after 10,000 time steps also undergoes a phase transition around $f_v = 0.02$. However, $\langle\lambda\rangle$ converges to its upper bound $\langle\lambda\rangle^*$ at a faster rate compared to the ratio of infected population. Let us consider the case $f_v = 0.04$. The relative ratio of infected nodes is $\approx 70\%$ (see dotted lines in Figure 8.3.2), variations in the vectorial layer importance provide no evident fluctuations. However, always at $f_v = 0.04$, the mean distance of infected populations from the centre of infection is not 70% of the maximum value, but rather $\langle\lambda\rangle(f_v = 0.04) \approx \langle\lambda\rangle^* \approx 0.384$ (see the grey shape and the dashed black line in Figure 8.3.2). Therefore, in the same time steps, the infection spreads only across 70% of populations but it covers almost all the distances from the infection origin, in the embedding space. We interpret this as the parasite spreading at a faster rate uniformly over the whole embedding space rather than uniformly across all the considered populations. These different spatial and number diffusion rates are relative to our selected SI dynamics. When the infection probability $\beta = 1$ (as in our case) and only one neighbour node becomes infected at a time, the infection spreads firstly through geodesics in the network (Sattenspiel, 2009; Jeub et al., 2015). Having the parasite spreading on geodesics through our spatial multiplex network is compatible with our finding from Figure 8.3.2: the mean distance of infected nodes from the infection centre saturates faster than the ratio of infected nodes.

8.3.2.1 Immunisation scenarios in the SMN model

In order to relate the topological features of each species population in the multiplex to their roles in spreading the parasite across the networked ecosystem, we analyse immunisation scenarios. In the immunisation scenarios a fraction of populations of a given species type (e.g. predators) is immunised against the parasite (see Section 2.7). As reported in the previous section, we found different species having different degree and participation patterns within the SMN model (see the cartographies in Figure 8.3) at high vector frequencies ($f_V = 0.75$). In fact, when $f_v = 0.75$ prey, predator and vector populations occupy different regions in the multiplex cartography. In Figure 8.3.2.1 we report the global infection times when the same total number $\phi = 417$ of predator or prey populations is immunised. The chosen ϕ corresponds to immunising half the predator populations in the system. Our results show that immunising prey over predators leads to a greater increase in the system infection times for all values of vectorial layer importance p_v. The better performance of immunising prey over predators is also reflected in the increase of parasite ratio $\Delta R(t)$ (Figure 8.3.2.1): immunising prey not only delays a pandemic but it also

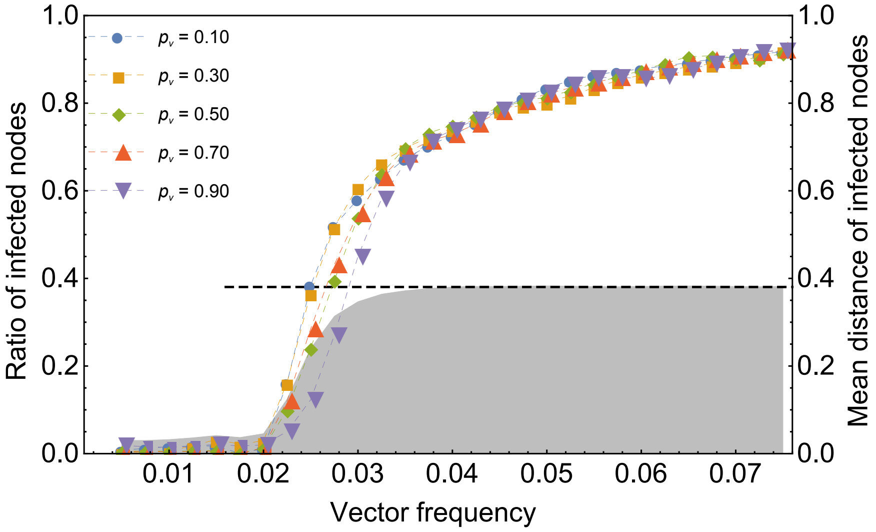


Figure 8.5: Ratio of infected populations after 10^4 steps, sampled at different values of p_v, against vector frequency f_v in the SMN model. When vectors are rare in the system, the system displays a phase transition in the rate of infection. The critical threshold is localized around $f_v \approx 0.02$, for all the values of p_v. The grey shape represents the mean distance of infected population from the origin of the parasite spreading and it is averaged over different p_v values. When $f_v > 0.02$ the infection radius saturates faster than the global percentage of infected populations. All curves are averages of 100 repetitions.

significantly slows down the parasite spreading in the initial accelerating phase (i.e., it lowers the $\Delta R(t)$ when $t < 140$). Even though slowing down the parasite transmission and reaching a pandemic at a later stage might sound equivalent, the parasite ratio increase reveals that in the predator immunisation scenario there is a higher diffusion speed in the decelerating infection phase, $t > 140$ (Figure 8.3.2.1). Because of this behaviour, we report on both patterns.

This difference could be attributed to the different topology of prey and predator populations in the trophic layer, i.e., the parasite spreads from vector to prey and then from prey to predator populations, so that prey have a higher betweenness in the sifted trophic interactions. Further numerical experiments indicate that this is not the case. Immunisation experiments performed with the same ϕ but with vector frequency $f_v = 0.25$ show that immunising either predators over prey gives statistically equivalent results in terms of both the parasite spreading times and the propagation rates. Therefore, at $f_v = 0.25$ immunising one species type over the other does not change parasite spreading. However, both the $f_v = 0.25$ and the $f_v = 0.75$ instances are relative to the same interaction matrices T and V and to the same number of immunised prey ϕ. Therefore, the relative difference in immunisation performances has to be attributed to the role played by each species within the global network topology. Immunising prey is different from immunising predator populations only when they have different topological patterns within the multiplex network, i.e. they occupy different areas of the multiplex cartography. This evidence points to the meaningfulness of the concept of network cartography for the parasite spreading dynamics: at $f_v = 0.75$ prey populations become truly multiplex hub nodes and assume an important role for parasite spreading, as demonstrated by our immunisation experiments.

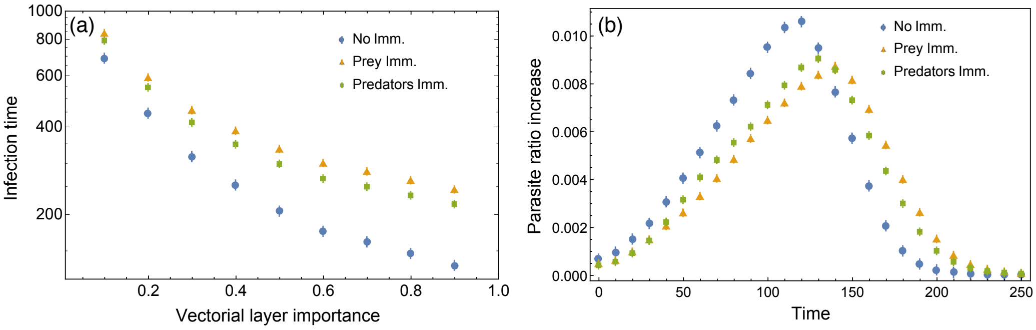


Figure 8.6: (a): global infection time versus vectorial layer importance p_v for different immunisation experiments with $f_v = 0.75$ in the SMN model. No immunisation means that no immunised populations are present in the system while two other dot types represent scenarios in which only prey or predators are immunised, respectively. For immunisation scenarios the same number of populations has been immunised. (b): parasite ratio increase of infected nodes over time for the SMN model for different immunisation scenarios with $f_v = 0.75$. Error bars are computed over 10 independent experiments. Immunising prey is the best choice in terms of both reducing the global infection time and slowing the infection spread over time.

8.3.3 Spatial ecological multiplex network model: the role of biodiversity

The SEMN model considers empirical interaction matrices T_{eco} and V_{eco} compared to SMN. Notice that the in V_{eco} the vector contaminates only 7 of the 20 species in the ecosystem, while in SMN it is allowed to contaminate all the other 2 species. In this section we relate the empirical ecological structure to the results for SEMN. The cartographies reported in Figure 8.7 (a-d) represent snapshots of the spatial ecological multiplex network with increasing frequencies of vectors. In all the cartographies there is one predator species that displays a wide variation in the participation coefficient, while the participation coefficients of the other predator species populations is zero. This is because, differently from SMN, the SEMN model has one predator species that can be contaminated by vectorial transmission (see V_{eco} in the Supporting Information), while the other predator species populations have links only on the trophic layer. When vector populations are rare ($f_v = 0.01$), predator and prey populations occupy the same regions of the cartography, as in the SMN model, see Figure 8.7 (a) and (b). A similar case occurs with prey populations, since only half of them have connections on the vectorial layer (see V_{eco} in the Supporting Information). Analogously to the SMN model, increasing the frequency of vectors leads to scenarios where some predator and prey populations display a wide range of participation coefficients. However, at both $f_v = 0.1$ and $f_v = 0.25$ predator populations have a higher multidegree than prey populations. This occurs because predators receive more connections than prey in the trophic layer. Therefore, for values as low as $f_v = 0.1$ the species types show varied and distinct patterns in the cartography. At $f_v = 0.25$, prey populations show an increased participation in the multiplex network as a sign of increased connectivity in the vectorial layer (Figure 8.7 (c)). When vector populations are highly frequent in the system, $f_v = 0.75$, the cartography reveals some extreme patterns: prey species populations that interact with vectors on the vectorial layer display participation coefficient close to 1 while the other prey species show focused interactions (Figure 8.7 (d)). This same pattern was observed between predator species populations that

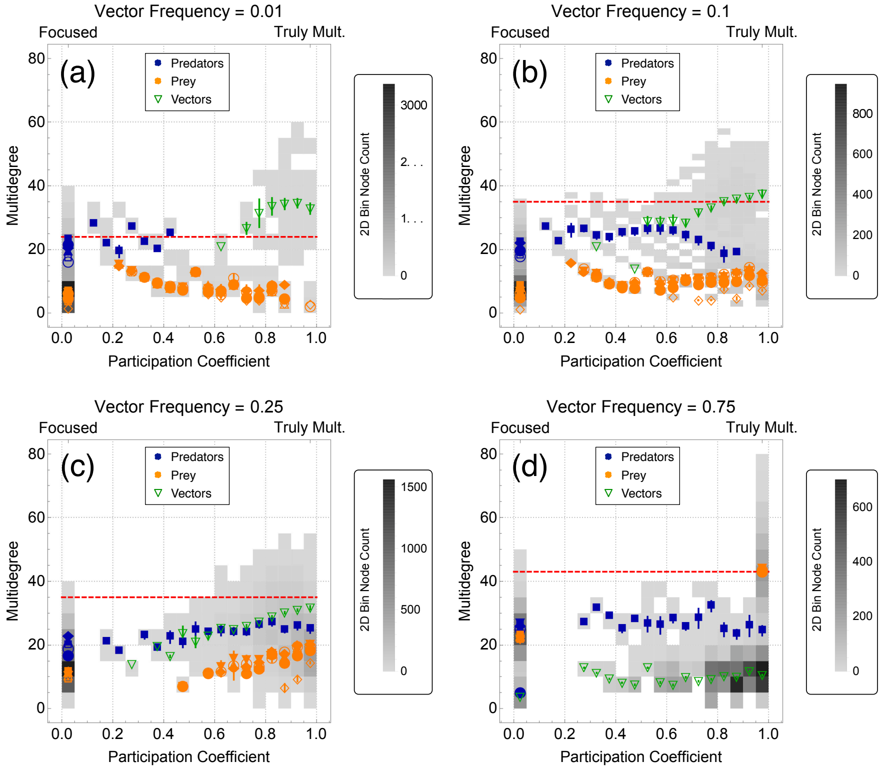


Figure 8.7: Cartographies as 2D histograms for the SEMN model for vector frequency $f_v = 0.01$ (a), $f_v = 0.1$ (b), $f_v = 0.25$ (c), and $f_v = 0.75$ (d). The 10,000 multiplex nodes are binned in 2D bins, according to their coordinates in the cartography. Bins are colour-coded according to the number of points falling within them: more coloured tiles indicate a higher number of nodes. Coloured dots identify individual species: predators (blue), prey (orange) and vectors (green). Nodes falling above the horizontal red line have degrees above the 95th percentile in the multidegree distribution and they are therefore considered hubs. Error bars represent standard error of the mean.

interact with vectors and the predator populations that do not when $f_v = 0.75$ (Figure 8.7 (d)). This was not observed in the SMN model.

As reported in Figure 8.8 (b), the time required to infect almost all the populations in SEMN is minimised when there is a high frequency of vectors in the environment and a high importance of vectorial layer for parasite transmission. Infection times decrease monotonically when $f_v = 0.5$ and 0.75. However, at vector frequencies $f_v = 0.1$ and 0.25 parasite spreading is optimised when the vectorial layer importance p_v is around 0.6 (8.8 (b)), that is, when vectorial and trophic transmission mechanism have similar importance. Therefore, vectorial and trophic transmission mechanism have an additive effect for parasite spreading only when $f_v < 0.5$. Comparing the results against a spatial aggregate network model using the Canastra matrices (Canastra SAN model) reveals how the multiplex structure can change dramatically the infection time. For instance, when $f_v = 0.1$, the infection time of the Canastra SAN model is halved compared to the SEMN one for $p_V = 0.1$, see also the dashed lines in Figure 8.8 (b). The multiplex structure not always increases the speed of parasite spreading and the multiple dynamics that resulted from the interplay of vectorial layer importance and community composition justifies the value of investigating different transmission routes via multiplexity. Despite the higher connectivity of the trophic layer in the SEMN model, parasite ratio increases behave similarly to the SMN

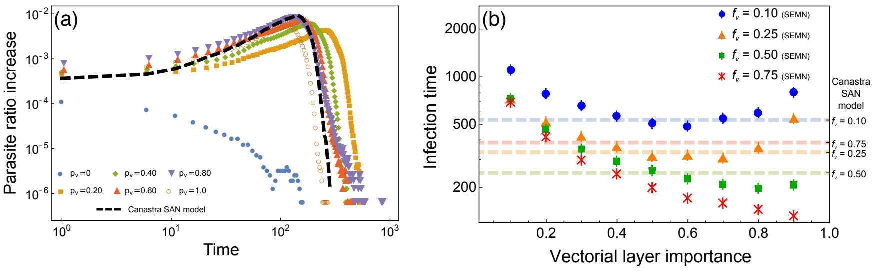


Figure 8.8: (a): global parasite ratio increase over time for $f_v = 0.75$ for SEMN model and different vectorial layer importance. A qualitatively similar behaviour was observed also for other vector frequencies. (b): global infection time versus vectorial layer importance p_v for different vector frequencies in the SEMN model. Dotted lines represent infection time in the SAN model applied to Canastra empirical data for different vector frequencies. Results in both plots are averages of 100 repetitions.

model (8.8 (a)). The parasite spreading propagates much slower on the trophic layer alone than on the full multiplex structure, see the $p_V = 0$ trajectory. Again, considering also contaminative interactions provides qualitatively different dynamics of parasite ratio increases than considering trophic interactions only (8.8 (a)). However, the dynamics of parasite ratio increases in time for the SEMN model are qualitatively similar to the SAN model relative to $p_V > 0$. Increasing the vectorial layer importance accelerates the parasite spreading even though no monotonous relationship is evident from the plots. For $p_v > 0$ the slow-down phase following the increase peaks does not behave independently of p_v (8.8 (a)). Therefore, these peaks cannot be considered good proxies of the infection times in the SEMN model. When the spreading deceleration occurs in different time windows, it sums up differently to the peak times, thus establishing global infection times that are not straightforwardly related to the peak times. For instance, the peak for $p_v = 0.8$ is reached sooner for the $p_v = 0.6$ but the deceleration phase takes longer for $p_v = 0.8$ than for the $p_v = 0.6$ and $p_v = 0.8$ has a higher global infection time compared to $p_v = 0.6$.

The SEMN model also displayed a phase transition in the emergence of a global epidemic, similarly to what happened for the SMN model. However, the different topology of trophic and vectorial layers brought to a slight increase in the critical vector frequency value, from $f_v = 0.02$ (SMN) to $f_v = 0.04$ (SEMN).

8.3.3.1 Immunisation scenarios in the SEMN model

Unlike the SMN model, the SEMN model has predator and prey populations exhibiting different cartography patterns only at low vector frequencies. Therefore, we investigated immunisation scenarios at $f_v = 0.1$ and $f_v = 0.25$. The results for $f_v = 0.1$ are reported in Figure 8.9 and are analogous to the $f_v = 0.25$ case (plots not shown for brevity).

Both the SMN and the SEMN models are spatially embedded, but SEMN has a higher number of species with interaction patterns based on empirical data. In SEMN, immunising prey over predator populations does not always hamper more the parasite spreading, as it happened in the SMN model. From the cartography in Figure 8.7 (a) one would expect predator populations to play a pivotal role in spreading the parasite, given their higher multidegree compared to prey populations, on average. However, in the same cartography 6 out of 12 prey species display a higher average participation coefficient compared to 6 out of 7 predator

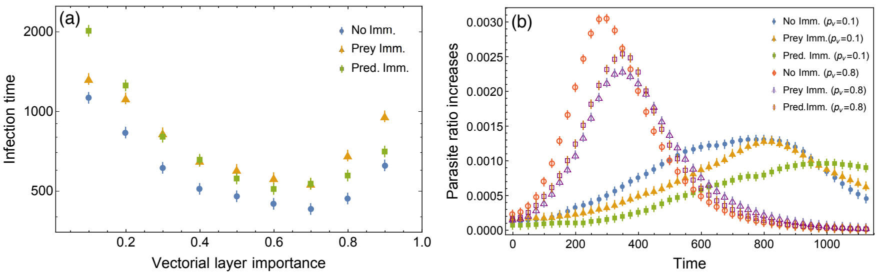


Figure 8.9: (a): global infection time versus vectorial layer importance p_v for different immunisation experiments with $f_v = 0.75$ in the SEMN model. The no immunisation scenario means that no immunised populations are present in the system while other dot types represent scenarios in which only prey or predators are immunised, respectively. For immunisation scenarios the same number of populations has been immunised. (b): parasite ratio increase of infected nodes over time for the SEMN model for different immunisation scenarios with $f_v = 0.75$. Error bars are computed over 10 independent experiments. Differently from the behaviour we observe in the SMN model, immunising prey is less effective than immunising predators in slowing down the disease spread for small p_v values. The opposite scenario happens when $p_v > 0.2$ where immunising prey is more effective than immunising predators, as shown in panel (b) comparing $p_v = 0.1$ and $p_v = 0.8$ immunising scenarios.

species (with participation coefficient equal to zero). Hence, from the cartography both predator and prey populations could play a central role in promoting the parasite spreading: predators are hubs while prey can spread the infection across both layers. In contrast to the SMN model, it is not possible to make predictions based on the cartography alone. Our immunisation simulations reveal the presence of two scenarios: when the parasite spreads mainly across the trophic layer ($p_v < 0.3$), then immunising the same number $\phi = 346$ of predator over prey populations significantly increases the infection times, (8.9 (a)), and slows down parasite diffusion (Figure 8.9 (b)). This finding relates to the SEMN cartography: predator populations have a high multidegree because they are hubs in the trophic layer (here $f_v = 0.1$) and hence promote the parasite spreading through trophic interactions. However, when the vectorial layer importance p_v increases above 0.3, then immunising predator or prey populations does not make noticeable difference. When $p_v > 0.7$ and the parasite spreads mainly through contaminative interactions the most effective immunisation strategy becomes immunising prey populations, since vectors contaminate mostly prey populations in the SEMN model (Figure 8.9). Again, this is compatible with the patterns in the multiplex cartography: when p_v is higher, the multiplex structure becomes predominant and the species populations that have higher participation coefficients, such as prey, can promote the infection spread.

8.4 Discussion

It is only recently that network scientists started addressing the multiplex structure of real-world systems such as ecological and epidemiological systems (Kivelä et al., 2014; De Domenico et al., 2013; Boccaletti et al., 2014). Multi-layer networks were used in ecological systems to approach different interaction types (Melián et al., 2009; Fontaine et al., 2011) and levels of organisation (Belgrano, 2005; Barter & T, 2016; Scotti et al., 2013). More in particular, multiplex networks were used for the first time by Kéfi et al. (2015), in order to consider trophic and non-trophic interactions together in a Chilean ecosystem. In epidemiological systems multilayer networks were used to describe parasite spreading with Susceptible-Infected-Susceptible dynamics (Saumell-Mendiola et al., 2012; Lima et al., 2015; Sahneh et al., 2013; Sanz et al., 2014), susceptible-infected-recovered dynamics (Dickison et al., 2012; Marceau et al., 2011; Buono et al., 2014), and multiple types of interactions between random layers (Zhao et al., 2014; Cozzo et al., 2013; Salehi et al., 2015). The modelling of multi-host parasites that are transmitted through multiple mechanisms in the ecosystem can be improved by applying the framework of multiplex networks. We used the multiplex approach to study both a simple predator-prey-vector system as a reference case, and an empirical data from host communities of *T. cruzi* in natural habitat (Canastra). Compared to their aggregate counterparts, both our multiplex network models displayed a richer phenomenology in terms of infection dynamics. Our three-species-system (SMN) as well as our empirical-based model (SEMN) showed that the epidemiological importance of vectors, hosts and parasites might be mapped on the multiplex cartography. Considering the node and link heterogeneity in a spatial context allowed for us to identify percolation thresholds for parasite spreading according to vector frequency. This is particularly interesting because the susceptible-infected dynamics in homogeneous hosts always leads to epidemic waves (in other words, when nodes are not spatially embedded there is no percolation threshold). In addition, we found that multiplex cartography had important implications in parasite spreading dynamics and that parasite transmission depends on: (i) the relative importance of the distinct transmission mechanisms, (ii) the role species play on the overall multiplex structure and (iii) the species relative frequencies in the system.

There is a strong debate in ecology on whether biodiversity reduces or not the risk of infection in host communities (Johnson et al., 2013; Keesing et al., 2006; Wood et al., 2014). In general, the effect of host diversity on parasite transmission depends on the ecological characteristics of hosts and on the mechanism of transmission (Wood et al., 2014). The spatial multiplex modelling framework that we propose in this study could be applied to address questions related to the role of multiple host community biodiversity on parasite transmission. In fact, we found that the spatial component has a significant impact on the speed of parasite spreading: spatial correlations slowed the speed of parasite spreading when compared to mean-field approximations. Therefore, considering the spatial structure of host communities in order to infer the importance of different host species for parasite transmission is a fundamental next step in future ecological disease studies (Craft & Caillaud, 2011; Kéfi et al., 2016). Percolation thresholds are spatially explicit tipping points that indicate the presence, in some regimes, of non-local correlations within a given system (Davis et al., 2008). For instance, if a network is not strongly connected, then the parasite will not be transmitted to the whole system. In our model the connectivity of the multiplex network was crucially affected by the frequency of different species. For very small frequency of vectors f_v, our model showed a percolation threshold in both the SMN and the Canastra SEMN model. The presence of such phase transition in the infection rate in an SI dynamics for a non-zero value of f_v is mainly related to (i) the spatial structure and to (ii) directed trophic interactions in the multiplex network. In the SMN model the parasite can percolate through the whole system only if $f_v > 0.02$, while in the Canastra SEMN model the critical vector frequency was found to be around $f_v = 0.04$. No phase transition for $f_v > 0$ was found in the RAN model, where nodes are not spatially embedded. We conjecture that the increase in the percolation threshold from the SMN to the SEMN models might be due to a higher diversity of potential hosts: with more species available there is an increased chance that vectors will interact with animals that do not become infected with the parasite. Interestingly, our theoretically computed frequencies agree with previous findings that even a small frequency of vectors in the ecosystem is sufficient to maintain Chagas disease in a human population (Reithinger et al., 2009).

Multiplex cartography (Battiston et al., 2014) considers both the relative frequency of each species and the interactions they have in both the trophic and the vectorial layers. Comparisons with aggregated networks revealed that considering trophic and vectorial transmission routes together can change dramatically the parasite spreading dynamics, depending on the relative frequency of vectors in the ecosystem. More in detail, the parasite spreading dynamics depends on the interplay between community species composition and the relative importance of the transmission mechanisms. In fact, when there is homogeneity in species composition (i.e. when the relative frequency of vectors $f_v \sim 0.5$), the lowest infection time is registered when the parasite spreads on both layers at the same time (i.e. for intermediate values of p_v) in both the SMN and the SEMN models. Therefore, our theoretical network models indicate that vectorial and trophic mechanisms of transmission can be additive in sustaining the spread of multi-host parasites such as *T. cruzi*, further agreeing with previous studies (Kribs-Zaleta, 2006). In random multiplex networks Saumell-Mendiola et al. (2012) the epidemic process also depends on the strength and nature of the coupling between the layers. In our case the vectorial layer importance p_v can be thought of as an implicit coupling between the layers, quantifying how much the vectorial layer is more important than the trophic layer in spreading the parasite. Previous investigation (Saumell-Mendiola et al., 2012; De Domenico et al., 2013; Kivelä et al., 2014; Boccaletti et al., 2014) showed that epidemic dynamics on a multiplex structure can be fundamentally different from the same dynamics on each multiplex layer considered as separate. Our results indicate that multiple mechanisms may speed up parasite spreading. The multi-layered transmission, which is observed in many parasites with complex life cycles and multiple mechanisms of infection, seems to be a very efficient strategy for spreading in communities of multiple hosts.

In vector-borne diseases, densities of hosts and vectors as well as the ratio of their densities, have strong implications for parasite transmission (Ross & Thomson, 1911; Pelosse & Kribs-Zaleta, 2012; Velascohernandez, 1994; Kribs-Zaleta, 2010a). The SMN model shows that higher vector frequencies make the vectorial layer faster in spreading the parasite from vectors to predator and prey populations. This relationship explains why infection times decrease monotonically with increased importance of the vectorial layer. On the other hand, if the vector frequency is low and the parasite spreads only on the trophic layer, it becomes increasingly difficult to infect more populations over time. In this situation the fastest global infection is achieved when both mechanisms of transmission are likewise selected for parasite spreading (there is a minimum in the infection time around $p_v = 0.6$). Moreover, in the Canastra SEMN model, we observe an analogous minimum even with higher vector frequencies. This suggests that global infection time is minimised when both mechanisms of transmission have similar importance in more complex ecological scenarios. Notice that considering both the transmission mechanisms but with one layer much more important than the other (e.g. $p_v = 0.1$) can lead to drastic increases in the infection time. The evolution and maintenance of mutually important multiple routes of transmission may be selected in parasites that infect a high number of host species.

Furthermore, using the multiplex cartography we predict that the relative importance of each mechanism for parasite spreading depends on the host community composition and relative frequency of species. We find that species structural patterns, encapsulated within the multiplex cartography, are a valuable measure to evaluate the importance of each species for parasite spreading. These findings are confirmed by the immunisation simulations. For instance, in the SMN model, a higher frequency of vectors ($f_v > 0.5$) increases prey populations connectivity and therefore their participation in the multiplex topology. We find different results when considering a more realistic ecological scenario. In the SEMN model, predator populations dominated the multiplex topology because of their higher connectivity and higher average multidegree. Immunising prey populations in the reference SMN model dramatically increases global infection time and the rate of disease spreading in the populations. However, in the SMN model immunising prey over predators results in different infection times only when these species occupy distinct regions in the multiplex cartography. This result points to the meaningfulness of the network cartography for understanding the parasite spreading dynamics. In fact, the multiplex cartography shows that prey participate more and have higher degree in the three-species multiplex network and thus could be a better target for immunisation. The immunisation simulations confirm this: immunising prey populations hampers the parasite spreading with respect to immunising the same number of predator populations. In the Canastra SEMN model, predators are the species type that attain most of their connections in the multiplex network and thus have a higher importance in the cartography. This pattern suggests that the predators are acting as a sink for the parasite and can thus reduce the overall parasite transmission in the SEMN model. This is mainly due to the fact that predators are hubs in the trophic layer and hence show a higher multidegree in the cartography. When the parasite spreads mainly in the trophic layer ($p_v < 0.3$) the immunisation experiments indicate that immunising predators hampers the disease more compared to immunising prey. This is in agreement with empirical studies pointing out the potential importance of predators as parasite bio-accumulators (Rocha et al., 2013; Jansen et al., 2015). However, prey also display a slightly higher average participation in the Canastra cartography and hence could also play a central role in spreading the parasite. In fact, when the vectorial layer importance p_v is above 0.7, immunising prey populations becomes the most effective immunisation strategy. This is because vectors contaminate mostly prey in the Canastra multiplex network. Again, the roles played by each species in the multiplex cartography depended on the frequency of vectors and is related to their importance for parasite spreading.

It has to be underlined that the main aim of our multiplex model is not to provide a realistic mechanism for the spreading dynamics of *T. cruzi* in wild hosts. Instead, our approach aims at providing a comprehensive framework for investigating the spreading of multi-host parasites across different transmission mechanisms. Additional information should be taken into account if one would want to study the dynamics of *T. cruzi* in wild hosts and Chagas disease epidemiology. For instance, it is known that the stercorarian transmission results in a much higher probability of parasite transmission from host to vector than from vector to host (Rabinovich et al., 1990). More realistic models should include these differences via different contact rates on different layers. In addition, host physiological and ecological characteristics influence their probability to transmit *T. cruzi*. A higher proportion of insects in host diets increase host probability of infection (Roellig et al., 2009; Rabinovich et al., 1990; Rocha et al., 2013). Finally, host species that share ecological habitat with vector species are more likely to be exposed to the infection (Jansen et al., 2015). Many zoonoses, which are infections naturally transmitted between vertebrate animals and humans, may have multiple hosts and mechanisms of transmission. Examples of zoonoses transmitted to humans by arthropod vectors include Malaria, Leishmaniasis, Chagas disease, West Nile virus, plague and Lyme disease (Schmidt & Ostfeld, 2001). The multiplex framework presented here could improve our understanding of the epidemiology and evolution of these parasites and help us elaborate more efficient control strategies for reducing disease incidence in humans. For instance, different or additional layers could be included within our multiplex framework to make the model more realistic, such as direct transmission mechanism or the network of human interactions with its socio-ecological characteristics. Outside of the ecological perspective, our spatial multiplex network model could be applied to modelling systems made of spatially embedded interacting agents where instead of parasite infection there is a given information spreading process.

Acknowledgements

We thank Hans Heesterbeek, Mason A. Porter, Darko Stefanovic, Jennifer Dunne, Markus Brede, the anonymous manuscript reviewers and the Santa Fe summer school scientific committee for their feedback on the early version of the manuscript.

Chapter 9

Community interactions determine the role of species in parasite spread amplification: the ecomultiplex network model

The previous Chapter ended by highlighting different roles played by predators and prey in facilitating parasite spread. In this Chapter[1] we further refine this finding by suggesting an improved version of the ecomultiplex model considering also metabolic theory for determining the abundance of animal groups in the simulated ecosystems. We use this framework for investigating how individual species, rather than groups of predators/prey, can indeed facilitate parasite spread through their ecological interactions (i.e. predator-prey or parasitic), confirming previous empirical findings and suggesting a potential mechanism for parasite transmission mediated by predators.

---

[1] This Chapter follows the layout of the "Three-paper" PhD thesis in that it is an original, publishable research manuscript available as: *Stella, M., Selakovic, S., Antonioni, A., & Andreazzi, C. S. (2017). Community interactions determine role of species in parasite spread amplification: the ecomultiplex network model. arXiv preprint arXiv:1706.05121.* In order to better present the content, the Chapter/paper is briefly introduced by a short summary. The Chapter follows the same structure of the original paper (Stella et al., 2017), in adherence to the guidelines of the "Three-paper" PhD thesis format. For a better presentation within the thesis, the words "paper" and "manuscript" were changed to Chapter. The Supplementary Information of the original paper is reported at the end of the Chapter rather than in an appendix for increased readability and easier consultation with the main text.

Abstract

Multiple routes of transmission for many diseases are investigated separately despite their potential interplay. As a unifying framework for modelling parasite spread through different transmission paths, we present the "ecomultiplex" model, where interacting animal groups are represented as a multiplex network. The model keeps into account: (i) multiple types of ecological interactions, (ii) spatial embedding and (iii) metabolic scaling. We adopt this framework for designing and testing potential control strategies for *T. cruzii* spread in two empirical host communities. Our investigation highlights two interesting results. Firstly, the ecomultiplex structure can be as efficient as more data-demanding epidemiological measures in identifying which species facilitate parasite spread. Secondly, the interplay between predator-prey and host-parasite interactions leads to a phenomenon of parasite amplification, in which top predators facilitate *T. cruzii* spread, offering theoretical interpretation of previous empirical findings. Our approach can provide novel insights in controlling pathogen spreading in real-world complex systems.

9.1 Introduction

Zoonoses are infections naturally transmitted between animals and humans, and are the most important cause of emerging and re-emerging diseases in humans (Perkins et al., 2005; Jones et al., 2008; Lloyd-Smith et al., 2009). The majority of the zoonotic agents are multi-host pathogens or parasites (Ostfeld & Holt, 2004; Alexander et al., 2012), whose various host species may differ in their contribution to parasite transmission and persistence over space and time (Jansen et al., 2015; Rushmore et al., 2014). This heterogeneity of host species contribution to parasite transmission is related to differences in host species' abundance, exposure and susceptibility to infection (Haydon et al., 2002; Altizer et al., 2003; Streicker et al., 2013). Further, many multi-host parasites have complex life cycles with multiple transmission modes, such as vertical, direct contact, sexual, aerosol, vector-borne and/or food-borne (Webster et al., 2017).

Among the zoonotic parasites with multiple hosts and transmission modes, *Trypanosoma cruzi* (Kinetoplastida: Trypanosomatidae), protozoan parasite which causes Chagas disease in humans, has a complex ecology that challenges transmission modelling and disease control (Noireau et al., 2009; Jansen et al., 2015). *T. cruzi* has already been found in more than 100 mammalian species and its transmission may be mediated by several interdependent mechanisms (Noireau et al., 2009; Jansen et al., 2015). For instance, *T. cruzi* has a contaminative route of transmission that is mediated by invertebrate vectors (Triatominae, eng. kissing bug); and a trophic route of transmission that cascades along the food-web when a susceptible predator feeds on infected prey (Noireau et al., 2009; Jansen et al., 2015).

Chemical insecticides and housing improvement have been the main strategies for controlling Chagas disease in rural and urban areas of Latin America (Dias & Schofield, 1999). However, these strategies are proving to be inefficient in reducing transmission (Roque et al., 2013). This is possibly related to the maintenance and transmission of parasites among local wild mammalian hosts and its association with sylvatic triatomine vectors (Roque et al., 2013, 2008). Therefore, modelling parasite transmission in a way that is explicitly considering the ecology of wildlife transmission, is fundamental to understanding and predicting outbreaks.

In this work we propose to address this challenge through the mathematical framework of multiplex networks (De Domenico et al., 2013; Kivelä et al., 2014; Boccaletti et al., 2014; De Domenico et al., 2016; Battiston et al., 2017), which have been successfully applied to epidemiology (Lima et al., 2015; De Domenico et al., 2016; Sanz et al., 2014) and ecology (Kéfi et al., 2015; Kéfi et al., 2016; Stella et al., 2016; Pilosof et al., 2017). Multiplex networks are multi-layer networks in which multi-relational interactions give rise to a collection of network layers so that the same node can engage in different interactions with different neighbours in each layer (De Domenico et al., 2013; Kivelä et al., 2014; Boccaletti et al., 2014).

We study the ecology of multi-host parasite spread by multiple routes of transmission and potential control strategies by developing the "ecomultiplex" framework (short for ecological multiplex framework). This framework is powerful in modelling complex systems such as infectious diseases transmission in wildlife. Firstly, it allows to account for multiple types of interactions giving rise to pathogen transmission with similar or different time scales. Secondly, the ecomultiplex framework uses metabolic theory (Jetz et al., 2004) for estimating species frequencies, which are known to influence parasite transmission (McCallum et al., 2001). Thirdly, by explicitly considering space, the model also allows to investigate the consequences of spatial structure on parasite transmission (Hudson et al., 2002).

The ecomultiplex framework is general in that it could include many ecological interactions among a diverse set of species in a realistic ecosystem. We apply this "ecomultiplex" formalism to investigating parasite spread in two vector and host communities in Brazil: Canastra (Rocha et al., 2013) and Pantanal (Herrera et al., 2011). We exploit the theoretical framework enriched with empirical data for designing and comparing different wild host immunisation strategies based on: (i) taxonomic/morphological features (e.g. immunising species of a same family); (ii) species interaction patterns (e.g. immunising species feeding on the vector); and (iii) species' epidemiological role (e.g. immunising species with higher parasite prevalence). Multiplex network structure proved to be an efficient measure in predicting species epidemiological role in both ecosystems. More importantly, considering together multiple transmission mechanisms allowed us to identify a parasite amplification role played by some species of top predators that would not be captured when considering the transmission mechanisms separately.

9.2 Material and Methods

9.2.1 Ecological multiplex network model

The "ecomultiplex" model describes an ecological community interacting in a spatially explicit ecosystem, see Fig. 9.1, using the novel framework of multiplex networks. Each layer of the ecomultiplex network represents a different type of interaction between species groups that can potentially lead to parasite transmission. We consider two alternative routes of transmission (i) food-web and (ii) contaminative. These infection routes give rise to a multiplex network of two layers where the nodes represent group of individuals of a given species, i.e. animal groups. Links on the food-web layer are directed to predator species and represent predator-prey interactions. Links on the vectorial layer are undirected and represent vector blood meals of parasitic insects acting as parasite vectors on their hosts. Distance among animal groups determines possible interactions: only groups sharing a spatial portion of their home range can interact with each other. We fixed the home range of all animal groups as a circle of radius $r = 0.03$ over a unitary squared space and studied a total of $N = 10000$ animal groups, (cf. the SI and Stella et al. (2016)). The small value of r has been tuned in order to keep the ecomultiplex network connected (De Domenico et al., 2013) so that the parasite can infect the whole ecomultiplex network.

9.2.2 Ecological data: Trophic interactions and body masses

Predator-prey and vector-host interactions in the ecomultiplex network are based on ecological data related to *T. cruzi* infection in wild hosts within two different areas: Canastra, a tropical savannah in Eastern Brazil (Rocha et al., 2013) and Pantanal, a vast floodplain in Southern Brazil (Herrera et al., 2011). Both biomes are highly diverse environments where epidemics of *T. cruzi* have been registered (Herrera et al., 2011). See SI Sections 9.5.1 and 9.5.2 for further details.

Trophic interactions in the food web are assigned according to literature data about animals' diets (Bueno et al., 2002; Cavalcanti, 2010; Amboni, 2007; Santos, 2012; Reis et al., 2006) (cf. SI Sect. 9.5.1). All vector species are grouped as one functional group due to missing species-level classification. Species prevalence is used to estimate the contaminative interactions in the vectorial layer (Herrera et al., 2011; Rocha et al., 2013). Positive parasitological diagnostics for *T. cruzi* (hemoculture) are used as a proxy for connections on the vectorial layer, since only individuals with positive parasitaemia (i.e. with high parasite loads in their blood) are able to transmit the parasite (Jansen et al., 2015). Body masses of host species represent averages over several available references (Herrera et al., 2011; Myers et al., 2008; Reis et al., 2006; Bonvicino et al., 2008; Schofieldet al. , 1994).

9.2.3 Mathematical formulation for group frequencies

Geographical proximity and ecological data regulate link creation in the ecomultiplex model. Ecological data, in particular body masses, regulate the frequency of animal groups (which are all mammals, except for vectors). Previous literature (Jetz et al., 2004) showed that the density n_i^{-1} of individuals of the same species having the species average body mass m_i follows the metabolic scaling:

$$n_i^{-1} = \beta^{-1} R_i^{-1} m_i^{3/4} \tag{9.1}$$

when R_i is the species-specific energy supply rate and β a constant expressing species metabolism. The above equation from metabolic theory can be used for determining the scaling relationship between body mass m_i and frequency f_i of animal groups (rather than animal individual) for species i, depending on the frequency of vector colonies f_v (cf. SI Sect.2):

$$f_i = (1 - f_v) \frac{m_i^{-1/4}}{\sum_{j=1} m_j^{-1/4}}. \tag{9.2}$$

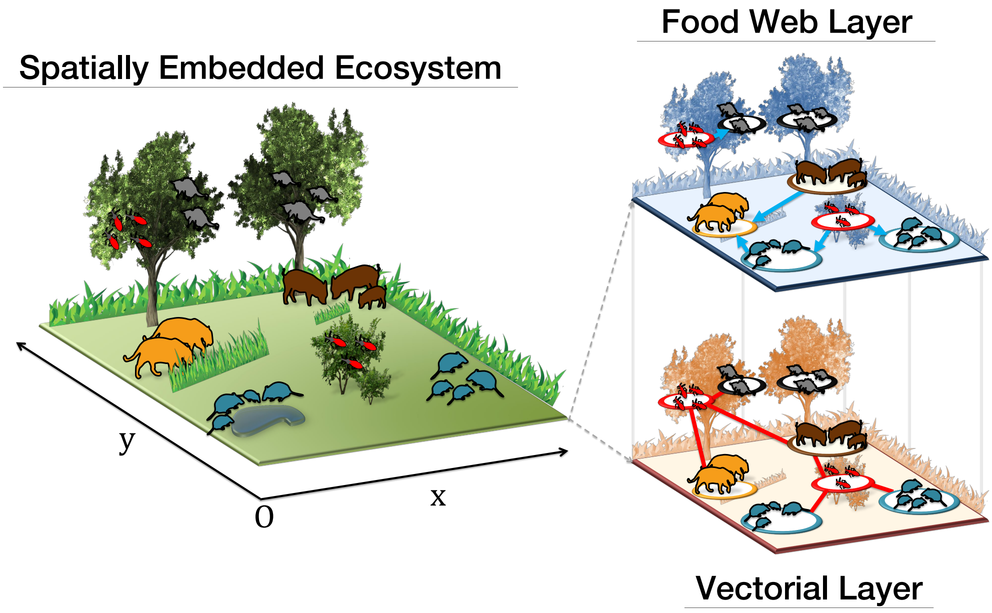


Figure 9.1: Visual representation of our ecological multiplex network model.

As a consequence of metabolic theory, the frequencies of animal groups in our ecomultiplex model scale as a power-law of body mass with exponent $-1/4$ rather than $-3/4$ (which is the scaling exponent for individuals rather than groups). We explicitly leave f_v as a free parameter of the model in order to investigate the influence of the frequency of vector colonies on parasite spreading.

9.2.4 Metrics for multiplex network analysis

We investigate the structure of a given ecomultiplex network through the concept of multiplex cartography (Battiston et al., 2014) (cf. SI Sect. 4). In our case, it describes how individual groups engage into trophic interactions on the ecomultiplex structure by considering: (i) the total number K of trophic interactions an animal group is involved in and (ii) the ratio U of uniform link distribution across layers, which ranges between 0 (all the links of a group are focused in one layer) and 1 (all links of a node are uniformly distributed across layers). The higher K the more an animal group interacts with other groups. The higher U the more an animal group will engage in feeding and contaminative interactions in equal measure.

9.2.5 SI model on the ecological multiplex network

Parasite spread is simulated as a Susceptible-Infected (SI) process: animal groups are susceptible or infectious. We assume that parasite transmission among animal groups happens considerably faster than both (i) group creation or extinction and (ii) parasite transmission within groups, so that meta-population and birth-death dynamics can be neglected. At each time step, the parasite can spread from an infected group to another one along a connection either in the vectorial (with probability p) or food-web (with probability $1-p$) layer. We consider p as a free parameter called vectorial layer importance, i.e. the rate at which transmission occurs through the consumption of blood by vectors rather than predator-prey interactions. We characterise globally the SI dynamics by defining the *infection time* t^* as the earliest time at which the parasite reaches its maximum spread within the networked ecosystem (Stella et al., 2016). The infection starts from a small circle of radius 0.03 in the middle of the unitary space infecting all animal groups within that area. Initial conditions are randomised over different simulations.

9.2.6 Immunisation Strategies

Immunisation strategies provide information on how species influence parasite spread: immunising species that facilitate parasite spreading should increase the global infection time t^* compared to immunising random species. We focus on immunising only[2] 10% of animal groups in ecomultiplex networks with 10000 nodes, in either high ($f_v = 0.25$) or low vector frequency scenarios ($f_v = 0.1$). Immunised groups are selected according to three categories of immunisation strategies (see also Table 1):

- **Biological features**: main taxonomic groups or body mass;
- **Ecomultiplex network features**: interaction patterns on the ecomultiplex structure;
- **Epidemiological features**: epidemiological measures of parasite prevalence in wildlife.

We define the infection time increase Δt_i as the normalised difference between the median infection time t_s when $\phi = 1000$ nodes are immunised according to the strategy s and the median infection time t_r when the same number of nodes is immunised uniformly at random among all animal groups, $\Delta t_i = \frac{t_i - t_r}{t_r}$. Infection times are averages sampled from 500 simulated replicates. Differences are always tested at 95% confidence level.

Immunisation Type	Strategy Name	Strategy Targets
Ecomultiplex Topological Features	Insectivores	Species feeding on the vector in a food-web
	Parasitised Didelphidae	Didelphidae contaminated by the vector on a vectorial layer
	Parasitised Mammals	All species contaminated by the vector on a vectorial layer
Taxonomical- morphological features	All Cricetidae	All Cricetidae
	All Didelphidae	All Didelphidae
	Large Mammals	All species with a body mass $>$ 1 kg
Epidemiological Features	Hemoculture N	The N species with the highest likelihood of being found infected with the parasite in field work (see SI).
	Serology N	The N species with the highest likelihood of having been infected with the parasite during their life time (see SI).

Table 9.1: Immunisation types, names and targets of the strategies we tested (cf. SI).

---

[2] By immunising groups at random in ecomultiplex networks with $N = 10000$ nodes, we identified $\phi = 1000$ as the minimum number of groups/nodes that have to be immunised in order to observe increases in t^* compared to random immunisation with a significance level of 5%.

Positive increases imply that the immunisation strategy slowed down the parasite in reaching its maximum spread over the whole ecosystem more than random immunisation. Negative increases imply that random immunisation performs better than the given immunisation strategy in hampering parasite diffusion.

9.3 Results

In this section we investigate how the ecomultiplex structure influences parasite spreading in both Canastra and Pantanal.

9.3.1 Network Analysis

Multiplex cartography for both Canastra and Pantanal (Fig. 2 (c-d)) shows that vectors are: (i) more connected and (ii) distribute their links more equally across the ecomultiplex layers than other species. Hence, vectors can get more easily infected in one layer and spread the parasite on another layer with equal likelihood. Hence, the cartography confirms that vectors facilitate parasite spread through their interactions. The local network structure around vectors in Canastra and Pantanal (cf. Fig. 2 (a-b)) shows that vector colonies are in the centre of star-like topologies on both layers. These results confirm that *Triatoma* species are pivotal for parasite spread, promoting it on both the food web and the vectorial layer. Although parasite diffusion can be hampered by removing vector colonies from the environment (Yamagata & Nakagawa, 2006), these immunisation strategies are not stable as vector reintroduction can happen shortly after elimination (Funk et al., 2013). Hence, we focus on immunisation strategies considering vectors' centrality in the ecomultiplex networks but immunising other species.

9.3.2 Immunisation Strategies

As expected, immunising species with the highest parasite prevalence (Hemoculture) is the best strategy for hampering parasite spread for both Canastra and Pantanal (cf. Fig. 9.3). This epidemiological strategy slows down parasite spread by almost 30% in Canastra and 26% in Pantanal when the parasite spreads mainly on the food-web layer ($p_v = 0.1$) Fig. 9.3. Immunising species interacting with vectors on the vectorial layer (an ecomultiplex strategy) also performs better than random. The difference between the epidemiological and the ecomultiplex strategies is present only at low vector frequencies ($f_v = 0.1$) in both Canastra (Fig. 4A) and Pantanal (Fig. 4C) but vanishes when $f_v = 0.25$ and $p_v > 0.2$ (Fig. 4B,D).

In Canastra, when 10% of the animal groups are vector colonies (Fig. 4A), biological immunisation strategies are equivalent to immunising species at random. The performance of biological immunisation changes dramatically when vector colonies become more frequent (Fig. 4B). Immunising large mammals decreases by 12% the global infection time when $p = 0.1$, suggesting that large mammals do not facilitate parasite transmission in the model. Immunising all the Didelphidae species leads to similar results (Fig. 4B). Modest increases in infection time are reported for immunising Cricetidae species when $p_v = 0.2$ (Fig. 4B).

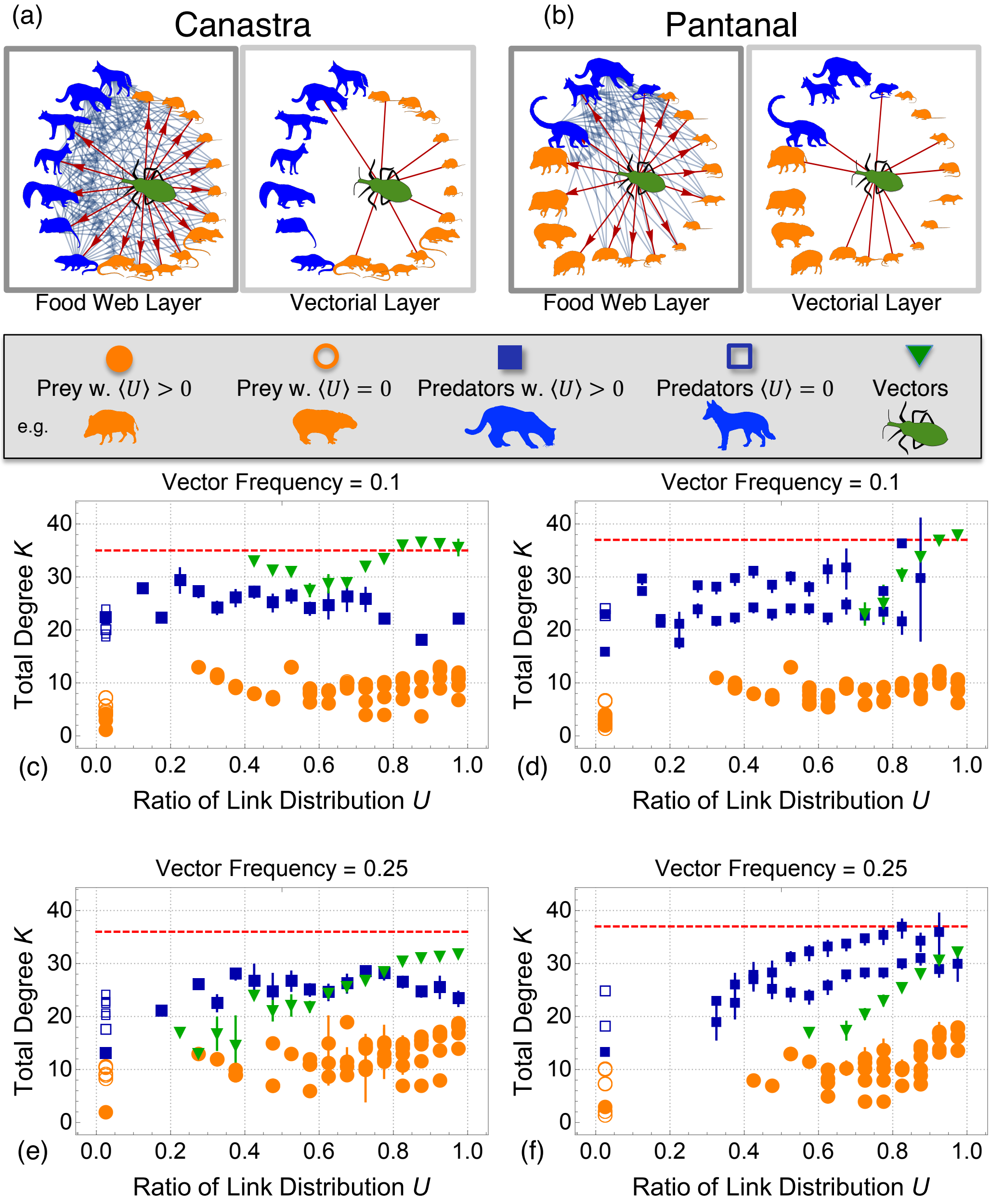


Figure 9.2: (a) and (b): Food web layer and vectorial layer in Canastra (left) and Pantanal (right) biomes. Predators are highlighted in blue, prey in orange and the vector in green. Interactions involving the insect are highlighted in red. Interactions involving other species are reported for completeness in blue. (c-e): Multiplex cartography of the Canastra ecomultiplex network with 10% (c) and 25% (e) of total groups as vectors. Multiplex cartography of the Pantanal ecomultiplex network with 10% (d) and 25% (e) of total groups as vectors. The red line separates hub nodes, i.e. the most connected nodes within the 95th percentile of the total degree distribution. The cartography highlights the average trends of species: blue for predators, orange for prey, and green for vectors. As evident from (a-e), vectors have higher total degree in the ecosystem and tend to distribute more equally their links across both the multiplex layers than all other species. Vectors are therefore pivotal in the ecosystem.

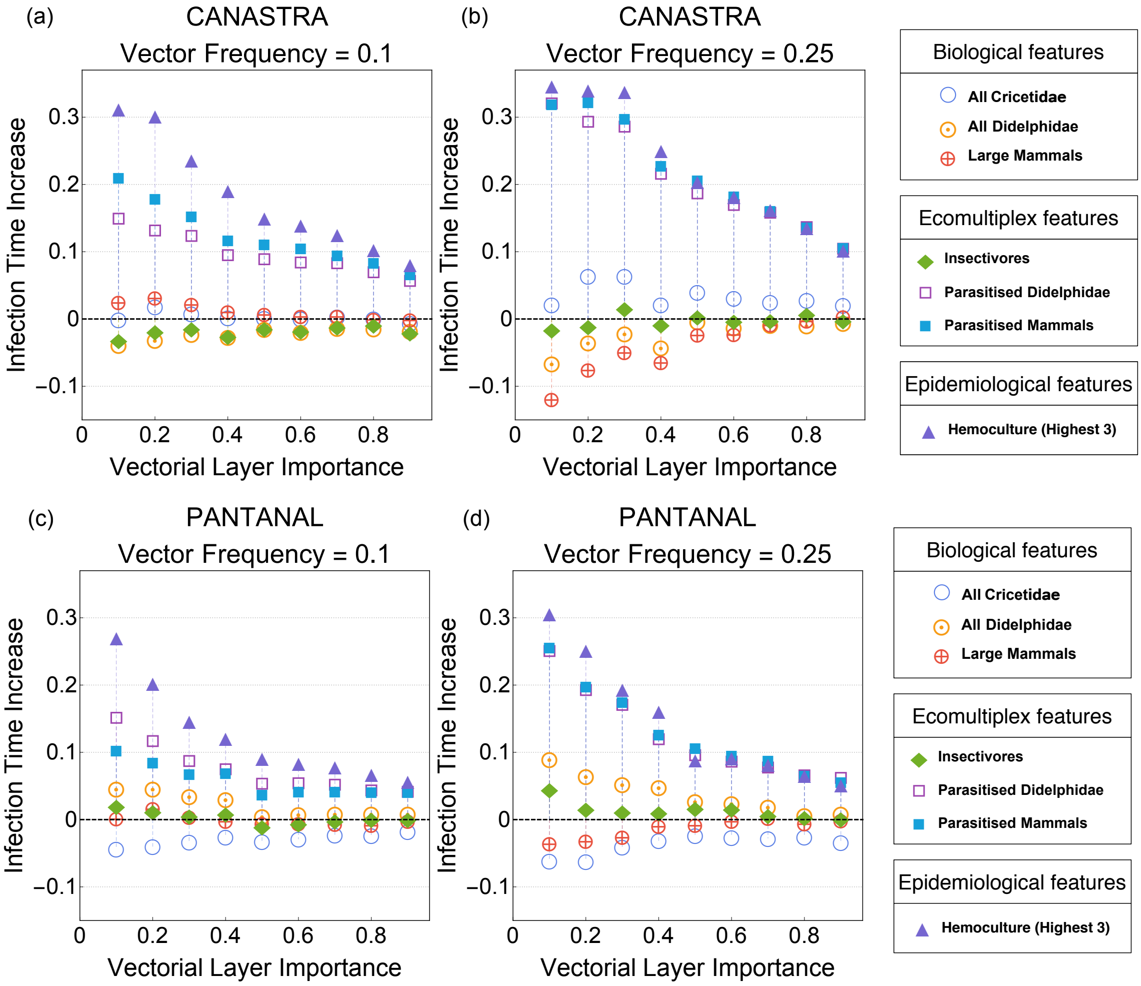


Figure 9.3: Immunisation strategies for the Canastra (top) and Pantanal (bottom) ecosystems when the vector frequency is 0.1 (left) and 0.25 (right).

Immunising species feeding on the vector (Insectivores) is equivalent to random immunisation (sign Test, p-values$>$ 0.1).

In Pantanal, immunising parasitised mammals, parasitised Didelphidae and species with the highest parasite prevalence (Hemoculture 3) are at least two times more effective in slowing down parasite spread compared to other strategies (Fig. 4C-4D). Contrary to what happens in Canastra, when $f_v = 0.1$ and the parasite spreads mainly on the food web ($p \leq 0.2$), immunising parasitised Didelphidae hampers parasite diffusion more than immunising all parasitised mammals (sign test, p-value$<$ 0.01) (Fig. 4C). Immunising insectivores or large mammals is equivalent to random immunisation (Fig. 4C). Immunising Cricetidae species always performs worse than random immunisation (Fig. 4C,D).

9.3.3 Top predators can lead to parasite amplification

In Canastra, the strategy Hemoculture 3 consists of immunising also one species of top predator, the *Leopardus pardalis* (ocelot) (see SI). We compare the performances of Hemoculture 3 against

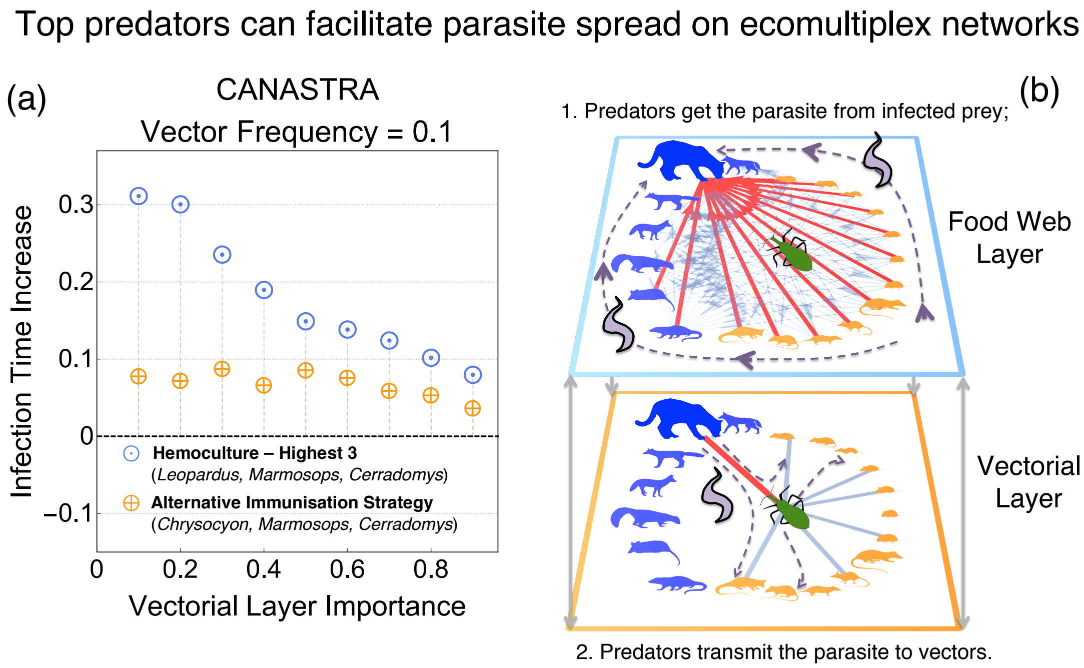


Figure 9.4: Difference in performances of the best immunisation strategy (hemoculture - Highest 3) when instead of the Leopardii the other top predator in the ecosystem (not parasitised by the vector) is immunised instead (hemoculture - H 3 No Leopardus). The other top predator is the maned wolf (Chrysocyon brachiurus).

another immunisation strategy where instead of the ocelot we immunise another top predator, the *Chrysocyon brachyurus* (maned wolf), which had negative parasite prevalence in this area (Rocha et al., 2013). In general, top predators are related to parasite transmission control in natural environments (Wobeser, 2013) so we do not expect differences.

Instead, results from Fig. 5A indicate a drastic increase of global infection time when a predator with positive parasite prevalence is immunised. This indicates that in Canastra the *Leopardus pardalis* (ocelot) has an amplification effect in spreading the parasite (Fig. 5B). This phenomenon crucially depends on the ecomultiplex structure, as discussed in the following section.

9.4 Discussion

We introduce ecomultiplex networks as a powerful theoretical framework for modelling transmission of multi-host parasites by multiple routes in species-rich communities. We identify three key points related to the model. Firstly, we show that network structure offers insights on which host species facilitate parasite spread. Secondly, we show that the structure of species-species interactions can be as useful epidemiological, taxonomic and morphological measures in controlling parasite transmission. Thirdly, we identify for the first time that the multiplex interactions of top predators affect their functional role in facilitating parasite transmission rather than hampering it.

Ecomultiplex strategies always outperform strategies based on species taxonomy, which neglect species' interactions. Further, network structure allows to design immunisation strategies performing as well as strategies considering parasite prevalence, with the advantage of requiring less fieldwork data. This quantitatively suggests the importance of jointly considering parasitoid-host and predator-prey interactions for understanding *T. cruzii* transmission in wildlife (Coura, 2005; Johnson et al., 2010; Penczykowski et al., 2016). Although Pantanal and Canastra differ in species and their ecological interactions, immunising species exposed to vectorial interactions proves to be efficient at all vectorial layer importances in both ecosystems. This underlines the importance of vectorial transmission for boosting parasite spread also in the food web. Importantly, the ecomultiplex model allows to quantitatively investigate the interplay between parasitoid-host and predator-prey interactions, an element conjectured being crucial for better understanding the ecology of wildlife diseases (Lafferty et al., 2008; Dunne et al., 2013; Funk et al., 2013).

The ecomultiplex model provides insights on how species influence parasite spread. In Pantanal, immunising only Didelphidae with positive parasitaemia slows down parasite spread more than immunising all mammals with positive parasitaemia. This finding agrees with previous studies that identify Didelphidae as important reservoirs for *T. cruzi* maintenance in natural ecosystems (Herrera & Urdaneta-Morales, 1992; Noireau et al., 2009). Reservoirs are epidemiologically connected populations in which the pathogen can be permanently maintained and from which infection is transmitted to a target population (Haydon et al., 2002). Being able to identify the functional role of species based on their topological interactions further highlights the powerfulness of the ecomultiplex framework in modelling parasite diffusion.

Within food webs, top predators are generally considered playing a regulating role in parasite spread by preying on infected individuals, thus eliminating infection sources for other animals (Packer et al., 2003; Hatcher et al., 2006; Wobeser, 2013). Our ecomultiplex framework shows that predators can also facilitate rather than just slow down parasite spread depending on their interactions with vectors. An example is the ocelot in Canastra, which is a generalist predator, i.e. it feeds on more prey than other species, and thus has an increased likelihood of becoming infected on the food web (Fig. 9.4 (b)). Once infected, ocelots can transmit the parasite to vectors through vectorial interactions. Since vectors facilitate parasite spread, then the ocelot can indeed amplify parasite diffusion. This is true for every generalist predator getting in contact with the vectors. This phenomenon of *parasite amplification* emerges only when both ecomultiplex layers are considered together. Therefore, this mechanism remarks the importance of unifying ecological and epidemiological approaches for better modelling parasite transmission. Importantly, this amplification mechanism provides a theoretical explanation by which the ocelot relates with the *T. cruzi* spread, as found in empirical studies (Rocha, 2006; Rocha et al., 2013).

Our theoretical model allows to design and test immunisation strategies in real-world ecosystems by relying on specific assumptions. For instance, since animal groups are embedded in space, home ranges need to be specified for them. For the sake of simplicity, in this ecological version of the model we considered only one average interaction radius for all species. Considering species-dependent empirical radii (home ranges) represents a challenging yet interesting generalisation for future work.

We consider the same parasite transmission probability across species in the SI dynamics. However, in our model the (i) structure of interactions and (ii) the different frequencies of animal groups are analogous to considering different transmission rates. In mean-field SI models (Stella et al., 2016), we showed that these two elements were sufficient for species to display different probabilities of catching the parasite. Immunisation strategies confirm this: immunising species that are more exposed to parasites leads to better immunisation performances compared to random immunisation. Considering species-dependent transmission rates as encapsulated in frequencies and links importantly reduces the number of model parameters.

We assume that parasite spread is happening at much faster rates compared to other meta-population dynamics (e.g extinction or migration), which are not currently considered in the model. However, including meta-population dynamics would allow to explore important research questions such as: (i) the interplay between predation and parasite amplification over top predators influencing parasite spread; (ii) the influence of migration on parasite diffusion; (iii) how extinction patterns influence parasite spread. Implementing he Markovian analytical approach from Gómez-Gardenes et al. (2015) in the ecomultiplex model would allow to reach even more realistic representations of real-world ecosystems.

9.5 Supplementary Information

9.5.1 Dataset

We parametrised the ecomultiplex models with data on species ecology (Tables 1 & 3) and interactions (Tables 2 & 4) in two localities: Canastra (Rocha et al., 2013) and Pantanal (Herrera et al., 2011). We obtained species body masses from the literature (Herrera et al., 2011; Myers et al., 2008; Reis et al., 2006; Bonvicino et al., 2008; Schofieldet al. , 1994) and calculated the average body mass when more then one reference was available for the same species. Interaction matrices contain data about existing trophic and contaminative interactions, respectively, blue and green squares. Predators are represented in columns and prey in rows. Trophic interactions in the food web layer were assigned according to literature information about animals' diets and ecology in Brazil (Bueno et al., 2002; Cavalcanti, 2010; Amboni, 2007; Santos, 2012; Reis et al., 2006; Rocha, 2006).

These studies identify categories of prey species such as "invertebrates" or "small mammals". We assigned a potential trophic interaction between a predator and a prey if the prey species belong to the prey category reported on species' diet (e.g. small mammal). However, we constrained these interactions based on information about species ecology such as use of habitat. For instance if a potential small mammal prey species is arboreal (e.g. *Caluromys philander*) we only considered predator species that are also able to use the arboreal strata (e.g. *Leopardus pardalis*). The study conducted in the Canastra area (herein Canastra ecomultiplex network) was located within the Serra da Canastra National Park and adjacent areas, in Minas Gerais state, South Eastern Brazil. It is an important remnant of the Cerrado biome, which is a vast tropical savana (Rocha et al., 2013). The Pantanal study (Pantanal ecomultiplex network) was conducted in the southern Mato Grosso do Sul state, midwest

Brazil. The Pantanal biome is a vast floodplain formed by a mosaic of seasonally inundated native grasslands, savannahs, scrub and semi-deciduous forests.

Positive parasitological diagnostics for *T. cruzi* (Hemoculture) (Herrera et al., 2011; Rocha et al., 2013) were used as a proxy for the interactions in the vectorial layer, since only individuals with positive parasitaemia (i.e. with high parasite loads in their blood) are able to transmit the parasite to vectors (Jansen et al., 2015).

Table 9.2: Taxonomic and ecological data of different animal species in Canastra area

ID	Species	Common name	Family	Diet type	Biomass [gr]	References
CHR	*Chrysocyon brachyurus*	Maned wolf	Canidae	omnivorous	25000	(Reis et al., 2006; Myers et al., 2008)
LEO	*Leopardus pardalis*	Ocelot	Felidae	carnivorous	9740	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
CER	*Cerdocyon thous*	Crab-eating fox	Canidae	omnivorous	6600	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
LYC	*Lycalopex vetulus*	Hoary fox	Canidae	omnivorous	3350	(Reis et al., 2006; Myers et al., 2008)
CON	*Conepatus semistriatus*	Striped˙hog-nosed skunk	Mustelidae	omnivorous	2567	(Reis et al., 2006; Myers et al., 2008)
DID	*Didelphis albiventris*	White-eared opossum	Didelphidae	omnivorous	1625	(Reis et al., 2006; Myers et al., 2008)
LUT	*Lutreolina crassicaudata*	Thick-tailed opossum	Didelphidae	omnivorous	600	(Reis et al., 2006; Myers et al., 2008)
CAP	*Caluromys philander*	Bare-tailed˙woolly opossum	Didelphidae	omnivorous	255	(Reis et al., 2006; Myers et al., 2008)
NEC	*Nectomys squamipes*	South American water rat	Cricetidae	omnivorous	270	(Reis et al., 2006; Myers et al., 2008)
MON	*Monodelphis* spp [a]	Short-tailed opossum	Didelphidae	omnivorous	72.8	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
MAR	*Marmosops incanus*	Gray slander opossum	Didelphidae	omnivorous	80	(Reis et al., 2006; Myers et al., 2008)
OXY	*Oxymycterus delator*	Spy hocicudo	Cricetidae	omnivorous	78.3	(Bonvicino et al., 2008)
CES	*Cerradomys subflavus*	Rice rat	Cricetidae	omnivorous	73	(Bonvicino et al., 2008)
NEC	*Necromys lasiurus*	Hairy-tailed bolo mouse	Cricetidae	omnivorous	52	(Bonvicino et al., 2008; Reis et al., 2006; Myers et al., 2008)
AKM	*Akodon montensis*	Montane grass mouse	Cricetidae	omnivorous	40.5	(Bonvicino et al., 2008; Reis et al., 2006; Myers et al., 2008)
AKO	*Akodon* spp [b]	Grass mouse	Cricetidae	omnivorous	33.25	(Bonvicino et al., 2008; Reis et al., 2006)
GRA	*Gracilinanus agilis*	Agile gracile opossum	Didelphidae	omnivorous	27.25	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
OLI	*Oligoryzomys* spp [c]	Pygmy rice rats	Cricetidae	omnivorous	21.23	(Bonvicino et al., 2008; Reis et al., 2006)
CAL	*Calomys* spp [d]	Vesper mouse	Cricetidae	herbivorous	18.65	(Bonvicino et al., 2008; Reis et al., 2006)
TRI	Triatominae	Kissing bug	Reduviidae	blood	0.2	(Schofieldet al. , 1994)

Rocha et al. (2013) report prevalence measurements for some genera, indicating the species that were collected during the study. Because those species are ecologically similar we calculated the average body size of the genus considering species that were collected. [a] *M. americana*, *M. domestica* and *M. sorex*. [b] *A. sp.*, *A. lindberghi*, *A. cursor*. [c] *O. sp*, *O. nigripes*, *O. rupestris*. [d] *C. sp*, *C. tener*

9.5.2 Mathematical formulation of scaling laws for the frequency of animal groups

In this section we report an extended version of the section "Mathematical formulation for group frequencies" providing additional details on the derivation of the animal frequencies we used for our model. Spatial embedding and ecological data regulate link creation in our ecomultiplex model. Ecological data, in particular body masses, regulate the frequency of animal groups.

Let us consider an ecosystem with S mammal species and a generic species of vector insect. We assume that animal groups of the same species $i \in 1, ..., S$ contain an average number of individuals n_i, each with average body mass m_i.. Therefore, the total body mass of a group of animal species i is $m_i n_i$. Let f_i be the frequency of the groups of species i. For instance, $f_i = 0.1$ means that 10% of the animal groups in the ecosystem are made of animals of species i. Then the number of animal groups N_i of species i is defined as $N_i = N f_i$. The sum of all the N_i provides the total number of animal groups N in the ecosystem minus the vector colonies (which have frequency f_v, instead).

	CHR	LEO	CER	LYC	CON	DID	LUT	CAP	NEC	MON	MAR	OXY	CES	NEC	AKM	AKO	GRA	OLI	CAL	TRI
CHR																				
LEO																				🟩
CER																				
LYC	🟦																			
CON	🟦																			
DID	🟦	🟦																		
LUT	🟦	🟦	🟦	🟦	🟦															
CAL		🟦																		🟩
NEC	🟦	🟦	🟦	🟦	🟦															
MON	🟦	🟦	🟦	🟦	🟦															
MAR	🟦	🟦	🟦	🟦	🟦															🟩
OXY	🟦	🟦	🟦	🟦	🟦	🟦	🟦													
CES	🟦	🟦	🟦	🟦	🟦	🟦	🟦													🟩
NEC	🟦	🟦	🟦	🟦	🟦	🟦	🟦													
AKM	🟦	🟦	🟦	🟦	🟦	🟦	🟦													🟩
AKO	🟦	🟦	🟦	🟦	🟦	🟦	🟦													🟩
GRA		🟦	🟦			🟦														
OLI	🟦	🟦	🟦	🟦	🟦	🟦	🟦													
CAL	🟦	🟦	🟦	🟦	🟦	🟦	🟦													🟩
TRI			🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	

Table 9.3: Interaction matrix includes trophic (predator-prey) and vectorial (vector-host) interactions in Canastra area

Table 9.4: Taxonomic and ecological data of different animal species in Pantanal area

ID	Species	Common name	Family	Diet type	Biomass [gr]	References
LEO	*Leopardus pardalis*	Ocelot	Felidae	carnivorous	9740	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
CET	*Cerdocyon thous*	Crab-eating fox	Canidae	omnivorous	6600	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
NAN	*Nasua nasua*	South American coati	Procyonidae	omnivorous	5140	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
SUS	*Sus scrofa*	Wild boar	Suidae	omnivorous	105750	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
PET	*Pecari tajacu*	Collared peccary	Tayassuidae	omnivorous	24000	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
HYH	*Hydrochaeris hydrochaeris*	Capybara	Caviidae	herbivorous	46200	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
TAP	*Tayassu pecari*	White-lipped peccary	Tayassuidae	omnivorous	30200	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
EUS	*Euphractus sexcinctus*	Six-banded armadillo	Chlamyphoridae	omnivorous	4450	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
PHF	*Philander frenatus*	Southeastern four-eyed opossum	Didelphidae	omnivorous	306.2	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
THP	*Thrichomys pachyurus*	Paraguayan punar	Echimyidae	herbivorous	291.25	(Reis et al., 2006; Herrera et al., 2011; Bonvicino et al., 2008)
CLL	*Clyomys laticeps*	Broad-headed spiny rat	Cricetidae	herbivorous	187.33	(Reis et al., 2006; Herrera et al., 2011; Bonvicino et al., 2008)
HOB	*Holochilus brasiliensis*	Web-footed marsh rat	Cricetidae	herbivorous	196	(Reis et al., 2006; Herrera et al., 2011; Bonvicino et al., 2008)
CES	*Cerradomys scotti*	Lindbergh's rice rat	Cricetidae	omnivorous	92.34	(Herrera et al., 2011; Bonvicino et al., 2008)
MOD	*Monodelphis domestica*	Gray short-tailed opossum	Didelphidae	omnivorous	111.2	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
OEM	*Oecomys mamorae*	Mamore arboreal rice rat	Cricetidae	herbivorous	83.75	(Reis et al., 2006; Herrera et al., 2011; Bonvicino et al., 2008)
THM	*Thylamys macrurus*	Long-tailed fat-tailed opossum	Didelphidae	omnivorous	46.6	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
CAC	*Calomys callosus*	Large vesper mouse	Cricetidae	herbivorous	37.45	(Reis et al., 2006; Herrera et al., 2011)
GRA	*Gracilinanus agilis*	Agile gracile opposum	Didelphidae	omnivorous	27.25	(Reis et al., 2006; Myers et al., 2008; Herrera et al., 2011)
TRI	Triatominae	Kissing bug	Reduviidae	blood	0.2	(Schofieldet al. , 1994)

We define the total mass M_i of animals from species i in the ecosystem as the sum over all animal groups N_i of species i:

$$M_i = (m_i n_i) N_i = m_i n_i f_i N, \tag{9.3}$$

where n_i represents the number of individuals within a given animal group exploiting a given home range. Previous literature Jetz et al. (2004) showed that the minimum density n_i^{-1} of individuals within a home range that is sufficient to sustain their metabolic requirements scales as:

$$n_i^{-1} = \beta^{-1} R_i^{-1} m_i^{3/4} \tag{9.4}$$

	LEO	CET	NAN	PHF	SUS	PET	HYH	TAP	EUS	THP	CLL	HOB	CES	MOD	OEM	THM	CAC	GRA	TRI
LEO																			
CET																			
NAN	🟦																		🟩
SUS																			🟩
PET																			
HYH	🟦																		
TAP																			
EUS	🟦	🟦	🟦																🟩
PHF	🟦	🟦	🟦																🟩
THP	🟦	🟦	🟦																🟩
CLL	🟦	🟦	🟦																🟩
HOB	🟦	🟦	🟦																
CES	🟦	🟦	🟦	🟦															
MOD	🟦	🟦	🟦	🟦															🟩
OEM	🟦	🟦	🟦	🟦															🟩
THM	🟦	🟦	🟦	🟦															
CAC	🟦	🟦	🟦	🟦															
GRA	🟦	🟦	🟦	🟦															🟩
TRI		🟦	🟦	🟦	🟦	🟦		🟦	🟦	🟦	🟦		🟦	🟦	🟦	🟦	🟦	🟦	

Table 9.5: Interaction matrix includes trophic (predator-prey) and vectorial (vector-host) interactions in Pantanal area

where R_i is the species-specific energy supply rate, i.e. the energy resources available to sustain the animal group in a given area and unit of time, expressed in W/km^2. β is a normalisation constant related to the species metabolism. Empirical work has shown that R_i is roughly independent on body mass Jetz et al. (2004). Inserting Eq.9.4 in Eq.9.3, we obtain:

$$M_i = m_i n_i f_i N = m_i(\beta R_i m_i^{-3/4}) f_i N. \tag{9.5}$$

Eq.9.5 relies on two assumptions: (i) the metabolic theory provides a good approximation for species densities and (ii) there is an average size n_i across animal groups of a given species. Eq.9.5 can be further used for determining how the frequency of animal groups f_i scales with body mass. To the best of our knowledge, there are no empirical data relating the frequency of animal groups with total body masses in ecosystems. However it is known from metabolic theory that the mass specific expenditure c of energy E_i per lifespan ΔT_i of an individual of mass m_i of species i is independent of body mass and it does not differ across different species Speakman (2005). In formulas we have: $c = E_i/(\Delta T_i m_i) \propto m_i^0$. Normalising the expenditure c by the home range area H_i over which energy E_i is gathered, then we have that:

$$c/H_i = E_i/(\Delta T_i H_i m_i) = R_i/m_i, \tag{9.6}$$

since R_i follows its definition (i.e. the energy available in the home range per unit of area and time). [3]

For the sake of simplicity, in our ecomultiplex model we approximate all home ranges as being equal (i.e. with one radius only). Therefore in our model the home range is constant across species, $H_i = H$, so that $m_i = R_i c/H = R_i c^* \propto m_i^0$. Summing over all individuals leads to $\sum m_i = \sum c^* R_i$. However while $\sum m_i = M_i$, the second sum is more delicate to deal with, as

---

[3] Eq. 9.6 is compatible with the findings from Jetz et al. (2004), where $R_i \propto m_i^0$ but $H \propto m_i$ across species.

home ranges can indeed overlap and hence cannot be considered as independent. In order to overcome this issue, we approximate $c^* \sum R_i \approx c'_i R_i$, where c'_i is an average energy expenditure rate at a global population level for species i. We approximate also c'_i as being constant across species, leading to the constraint:

$$M_i = c^{'} R(i), \tag{9.7}$$

Our assumption implies that a fraction of the energy acquired by individual animals, mediated by c', gets transformed into body mass M_i at the global population level, i.e. when all individuals of a species are considered. This implies that species with higher energy supply rates will also have higher total body mass. Inserting n_i into 9.7 and imposing the normalisation of frequencies, $\sum_{j=1}^{N_S} f_j = 1 - f_v$ leads to the following analytical expression for the frequency of animal groups of species i:

$$\gamma = \frac{1 - f_v}{\sum_{j=1} m_j^{-1/4}} \to f_i = (1 - f_v) m_i^{-1/4} \sum_{j=1} m_j^{-1/4} \to, \tag{9.8}$$

$$f_i = (1 - f_v) \frac{m_i^{-1/4}}{\sum_{j=1} m_j^{-1/4}}. \tag{9.9}$$

With our assumption for the total body mass, the frequencies of animal groups scale as a power-law with exponent $-1/4$. This scaling quantity is different from the scaling coefficient $-3/4$ that comes from metabolic theory and which refers to individuals rather than animal groups. Notice that our scaling for animal groups is indeed a consequence of metabolic theory for individuals. The above approximations allow to express frequencies of mammal groups as a function of the frequency of vector colonies f_v.

9.6 Model parameters and initial conditions

The ecomultiplex model adopts a few parameters that are the following:

- Number of total animal groups in the system N. We chose $N = 10000$ for obtaining numerically robust results. The same phenomenology was observed also at smaller scales ($N = 1000$ and $N = 500$).
- Average body mass m_i for individuals of species i, which are used for computing the frequencies f_i of animal groups of species i in the ecomultiplex network (see previous SI section);
- Frequency of vector colonies f_v. This quantity determines also the frequency of all the other animal groups in the ecosystem. In the main text we explored a low vector frequency scenario ($f_v = 0.1$) and a high vector frequency scenario ($f_v = 0.25$);
- Average interaction radius r: two animal groups are connected in any layer only if they are closer than r in the embedding space. This parameter was tuned numerically to $r = 0.03$ for getting fully connected multiplex networks where on average the parasite could potentially spread across all animal groups in the network;
- Probability β for an infected node of transmitting the infection to a susceptible node in the SI model (transmission rate). In order to be compatible with previous results (Stella et al.,

2016), we chose $\beta = 1$. This choice implies that the parasite spreads exactly accordingly to the topology of the multiplex network, as all contacts among infected and susceptible animal groups lead to parasite transmission with probability $\beta = 1$. Notice that the dependency of parasite spreading on the topology of the ecomultiplex network implies that even if $\beta = 1$ for all species, some species are still more (less) exposed to the parasite, as they can have more (less) connections or be more central along the patterns of infection and thus have more opportunities for getting infected.

- Vectorial layer importance p_v, determining the likelihood with which the parasite spreads along a link in the vectorial layer rather than using a link in the food web. When $p_v = 0$ the parasite spreads only across the food web layer. When $p_v = 1$ the parasite spreads only across the vectorial layer. When $p_v = 0.5$, the parasite spreads with equal likelihood across links in the vectorial layer and links in the food web layer. This is a free parameter in the model, ranging between 0 and 1.

Let us underline that all the results reported in the main text are relative to 500 random iterations of the SI model, each one with randomised initial conditions where only 30 adjacent animal groups are infected in the ecomultiplex network. Numerical analysis over the resulting simulations indicated the distributions of infection times being always unimodal, so that there was only a single most frequent value of infection time for every combination of vectorial layer importance and vector frequency. Unimodality justified the use of estimators robust to noise such as medians for computing average quantities for the infection times, as reported in the main text.

9.6.1 Network cartography for ecology

In the main text we used multiplex cartography for visualising topological patterns of different species in our ecomultiplex network model. The concept of network cartography was introduced by Guimerà and Amaral for distinguishing communities according to their connectivity (Guimerá & Amaral, 2005). This notion was later generalised by Battiston et al. (2014) for multiplex networks.

A multiplex cartography represents visually, like in a map, the role played by a given node across layers according to its topological features (Battiston et al., 2014). An example is reported in Supplementary Figure S9.5, where nodes of a fictional 2-layer multiplex network are mapped into regions of the cartography. We consider this measure to be of interest for ecological network science because it is a simple yet powerful measure for distinguishing different participation of nodes in a multiplex network with an appealing analogy for maps.

As in previous works (Battiston et al., 2014; De Domenico et al., 2013), we consider a cartography based on the following two measures: the multidegree or overlapping degree K_i and the participation coefficient or ratio of uniform link distribution U_i of node i. The multidegree K_i is defined as the sum of all the degrees of node i across the M multiplex layers:

$$K_i = \sum_{\alpha} k_i^{(\alpha)}. \tag{9.10}$$

where $k_i^{(\alpha)}$ is the degree of node i in layer $\alpha \in \{1, ..., M\}$. The multidegree K_i represents a proxy of the overall local centrality that a node has within the multiplex network. In our ecomultiplex network, the multidegree counts in how many trophic interactions is involved a given animal group and it is therefore a measure accounting for local information (i.e. the neighbourhood of an animal group) but across different interaction types (i.e. all trophic interactions, both eating and contaminative ones).

Differently from Battiston et al. (2014), we consider K_i rather than its standardised counterpart because our ecomultiplex networks do not display Gaussian-like multidegree distributions. We also use multidegree for defining hubs (i.e. animal groups interacting more than the average in the ecosystem) as those nodes being in the 95th percentile of the multidegree distribution.

However, the multidegree does not distinguish between interaction types. Two nodes could have the same multidegree, say $K = 10$, but one could be involved in 10 eating interactions while the other one in 5 eating interactions and in 5 contaminative ones, instead. In order to better assess the topology of individual nodes/animal groups in the ecomultiplex network we also consider the ratio U_i of uniform link distribution across layers for node i. In formulas, this is defined as:

$$U_i = \frac{M}{M-1}\left[1 - \sum_{\alpha=1}^{M}\left(\frac{k_i^{(\alpha)}}{K_i}\right)^2\right]. \tag{9.11}$$

U_i ranges between 0 (for nodes that concentrate all their connections in one level only) and 1 (for nodes that distribute connections over all the M layers uniformly). For instance, an animal group with degree 5 in the food-web and in the vectorial layer would have a ratio U of uniform link distribution equal to 1. Instead, another animal group with degree 10 in the food-web but 0 in the vectorial layer would have $U = 0$.

As reported also in Supplementary Figure S9.5, the couples (U_i, K_i) represent coordinates on a 2D "map" of a given node: nodes falling in the upper-right part of it are hubs that distribute uniformly their connections across layers while nodes falling in the lower-left part of the map are poorly connected nodes with links mainly in one layer only. Since in both our datasets vectors contaminate several animal species and are eaten also by several insectivores, we expect for vector groups to fall within the upper-right part of the multiplex cartography.

Since we simulate ecosystems with $N = 10000$ nodes, we do not visualise individual points but rather clusters of them, obtained by binning the original points in a 2D heat-map. We also plot the average trends of individual species in the ecomultiplex.

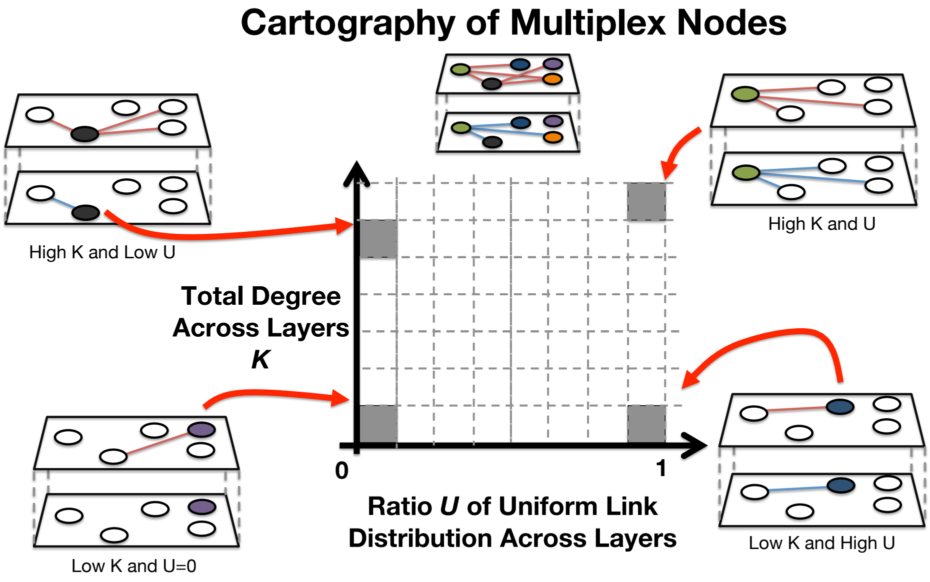


Figure 9.5: Scheme of how multiplex cartography works. In our ecomultiplex network, the total degree across layers counts how many trophic interactions given animal group participates in.

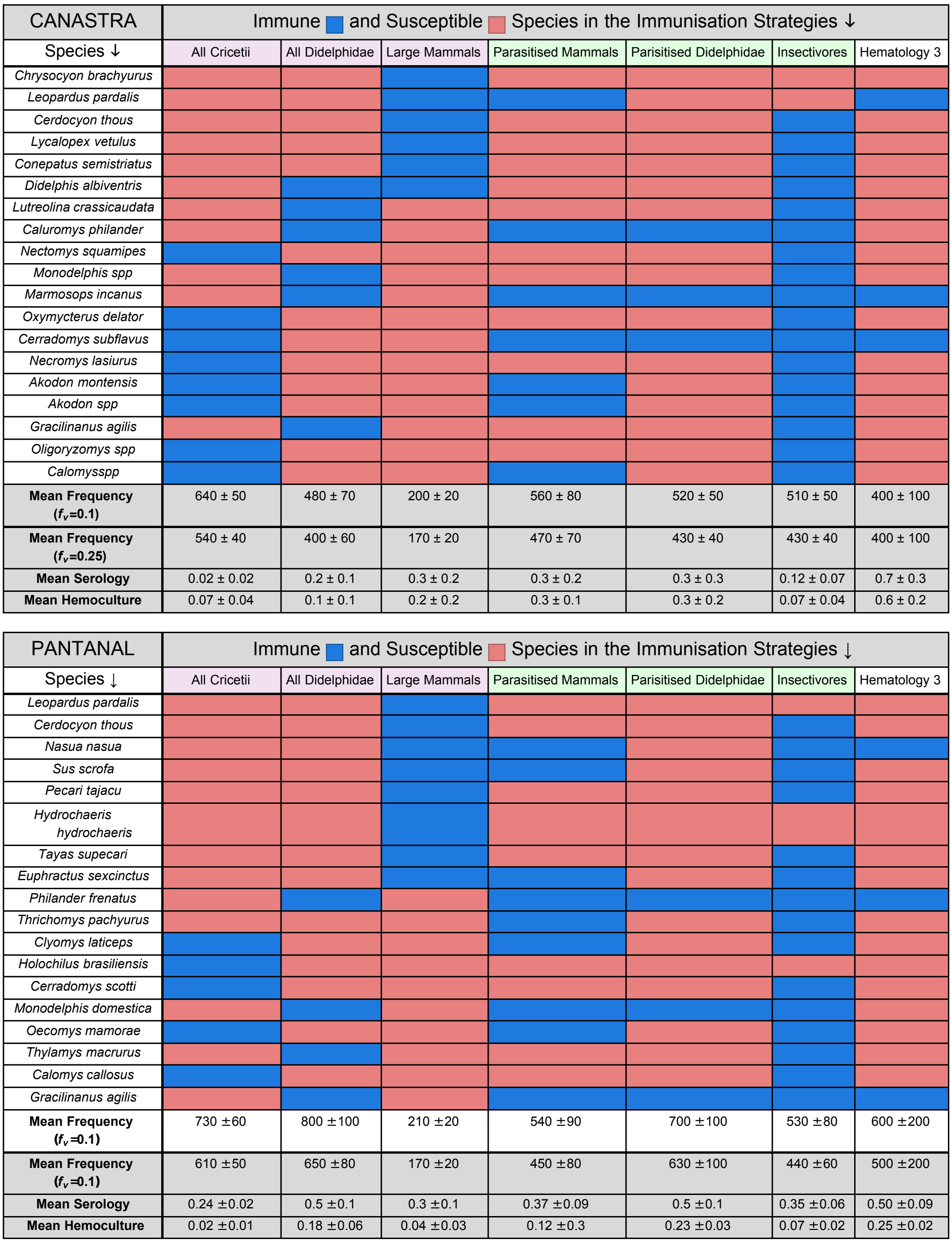
9.6.2 Immunisation Strategies

Supplementary Figure S9.6 summarises all the different species considered in each immunisation strategy presented in the main text. Remember from the main text that the best performing strategies are Hematology 3 and Parasitised Mammals. Notice that both in Canastra and Pantanal, Hematology 3 is not the strategy involving the most abundant species in the ecomultiplex network (cf. the average frequencies).

9.6.3 Additional Immunisation Strategies

Supplementary Figure S9.7 reports other immunisation strategies we explored, such as:

- Parasitised Cricetii: immunising the species from the Cricetus family that are linked to the vector on the vectorial layer;
- Parasitised Prey: immunising the prey species that are linked to the vector on the vectorial layer;
- Hematology (Highest 6): immunising the first 6 species with the highest likelihood of being found infected with the parasite in field work;
- Serology (Highest 3): immunising the first 3 species with the highest likelihood of having being found infected with the parasite during their lifetime in field work;
- Serology (Highest 6): immunising the first 3 species with the highest likelihood of having being found infected with the parasite during their lifetime in field work.



CANASTRA	Immune 🟦 and Susceptible 🟥 Species in the Immunisation Strategies ↓						
Species ↓	All Cricetii	All Didelphidae	Large Mammals	Parasitised Mammals	Parisitised Didelphidae	Insectivores	Hematology 3
Chrysocyon brachyurus	🟥	🟥	🟦	🟥	🟥	🟥	🟥
Leopardus pardalis	🟥	🟥	🟦	🟦	🟥	🟥	🟦
Cerdocyon thous	🟥	🟥	🟦	🟥	🟥	🟦	🟥
Lycalopex vetulus	🟥	🟥	🟦	🟥	🟥	🟦	🟥
Conepatus semistriatus	🟥	🟥	🟦	🟥	🟥	🟦	🟥
Didelphis albiventris	🟥	🟦	🟦	🟥	🟥	🟦	🟥
Lutreolina crassicaudata	🟥	🟦	🟥	🟥	🟥	🟦	🟥
Caluromys philander	🟥	🟦	🟥	🟦	🟦	🟦	🟥
Nectomys squamipes	🟦	🟥	🟥	🟥	🟥	🟦	🟥
Monodelphis spp	🟥	🟦	🟥	🟥	🟥	🟦	🟥
Marmosops incanus	🟥	🟦	🟥	🟦	🟦	🟦	🟦
Oxymycterus delator	🟦	🟥	🟥	🟥	🟥	🟦	🟥
Cerradomys subflavus	🟦	🟥	🟥	🟦	🟦	🟦	🟦
Necromys lasiurus	🟦	🟥	🟥	🟥	🟥	🟦	🟥
Akodon montensis	🟦	🟥	🟥	🟦	🟥	🟦	🟥
Akodon spp	🟦	🟥	🟥	🟦	🟥	🟦	🟥
Gracilinanus agilis	🟥	🟦	🟥	🟥	🟥	🟦	🟥
Oligoryzomys spp	🟦	🟥	🟥	🟥	🟥	🟦	🟥
Calomysspp	🟦	🟥	🟥	🟦	🟥	🟦	🟥
Mean Frequency (f_v=0.1)	640 ± 50	480 ± 70	200 ± 20	560 ± 80	520 ± 50	510 ± 50	400 ± 100
Mean Frequency (f_v=0.25)	540 ± 40	400 ± 60	170 ± 20	470 ± 70	430 ± 40	430 ± 40	400 ± 100
Mean Serology	0.02 ± 0.02	0.2 ± 0.1	0.3 ± 0.2	0.3 ± 0.2	0.3 ± 0.3	0.12 ± 0.07	0.7 ± 0.3
Mean Hemoculture	0.07 ± 0.04	0.1 ± 0.1	0.2 ± 0.2	0.3 ± 0.1	0.3 ± 0.2	0.07 ± 0.04	0.6 ± 0.2

PANTANAL	Immune 🟦 and Susceptible 🟥 Species in the Immunisation Strategies ↓						
Species ↓	All Cricetii	All Didelphidae	Large Mammals	Parasitised Mammals	Parisitised Didelphidae	Insectivores	Hematology 3
Leopardus pardalis	🟥	🟥	🟦	🟥	🟥	🟥	🟥
Cerdocyon thous	🟥	🟥	🟦	🟥	🟥	🟦	🟥
Nasua nasua	🟥	🟥	🟦	🟦	🟥	🟦	🟦
Sus scrofa	🟥	🟥	🟦	🟦	🟥	🟦	🟥
Pecari tajacu	🟥	🟥	🟦	🟥	🟥	🟦	🟥
Hydrochaeris hydrochaeris	🟥	🟥	🟦	🟥	🟥	🟥	🟥
Tayas supecari	🟥	🟥	🟦	🟥	🟥	🟦	🟥
Euphractus sexcinctus	🟥	🟥	🟦	🟦	🟥	🟦	🟥
Philander frenatus	🟥	🟦	🟥	🟦	🟦	🟦	🟦
Thrichomys pachyurus	🟥	🟥	🟥	🟦	🟥	🟦	🟥
Clyomys laticeps	🟦	🟥	🟥	🟦	🟥	🟦	🟥
Holochilus brasiliensis	🟦	🟥	🟥	🟥	🟥	🟥	🟥
Cerradomys scotti	🟦	🟥	🟥	🟥	🟥	🟦	🟥
Monodelphis domestica	🟥	🟦	🟥	🟦	🟦	🟦	🟥
Oecomys mamorae	🟦	🟥	🟥	🟦	🟥	🟦	🟥
Thylamys macrurus	🟥	🟦	🟥	🟥	🟥	🟦	🟥
Calomys callosus	🟦	🟥	🟥	🟥	🟥	🟦	🟥
Gracilinanus agilis	🟥	🟦	🟥	🟦	🟦	🟦	🟦
Mean Frequency (f_v=0.1)	730 ±60	800 ±100	210 ±20	540 ±90	700 ±100	530 ±80	600 ±200
Mean Frequency (f_v=0.1)	610 ±50	650 ±80	170 ±20	450 ±80	630 ±100	440 ±60	500 ±200
Mean Serology	0.24 ±0.02	0.5 ±0.1	0.3 ±0.1	0.37 ±0.09	0.5 ±0.1	0.35 ±0.06	0.50 ±0.09
Mean Hemoculture	0.02 ±0.01	0.18 ±0.06	0.04 ±0.03	0.12 ±0.3	0.23 ±0.03	0.07 ±0.02	0.25 ±0.02

Figure 9.6: Scheme on which species are immunised as animal groups in the ecomultiplex network in the different immunisation strategies presented in the main text. The average frequency, serology and hematology of the animal groups immunised in each strategy are presented as well. Error margins indicate standard deviations.

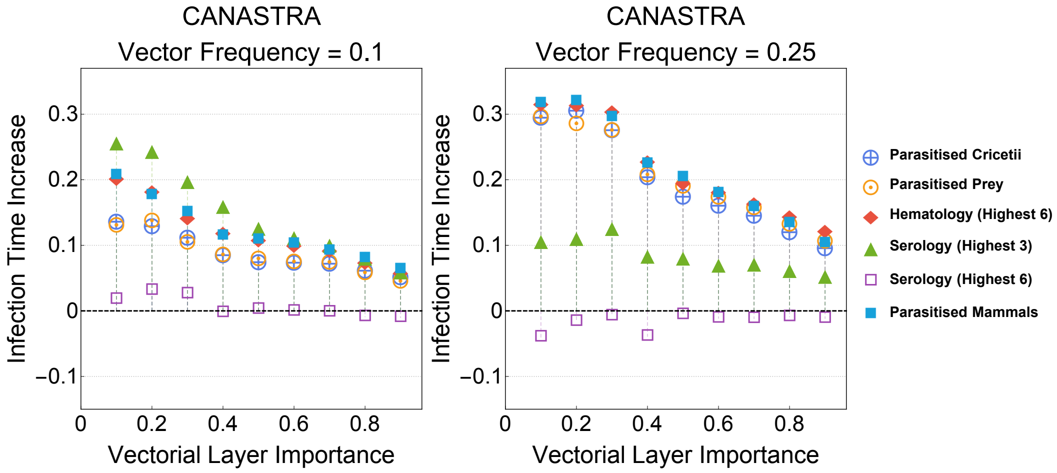


Figure 9.7: Infection time increases in Canastra for immunisation strategies that are not reported in the main text. Visual comparisons are made against the strategy Parasitised Mammals from the main text. For low vector frequency ($f_v = 0.1$) all the reported strategies behave worse than Hemoculture (Highest 3) and were therefore not discussed in the main text.

The immunisation strategies reported in Supplementary Figure S9.7 are compared against the best ecomultiplex strategy from the main text, i.e. Parasitised Mammals (in which the species linked to the vector on the vectorial layer are immunised). Interestingly, in terms of infection time increases, Serology (Highest 3) outperforms the ecomultiplex based strategy Parasitised Mammals when vector colonies compose 10% of the ecosystem ($f_v = 0.1$) and when the parasite spreads mainly through the food-web layer (sign test, p-value$< 10^{-}2$). In fact, for vectorial layer importance $p_v > 0.5$, Serology (Highest 3) and Parasitised Mammals give equivalent results. Notice, however, that Serology (Highest 3) performs consistently below the infection time increase 0.3 when $p_v \leq 0.2$ and $f_v = 0.1$, which is the infection time increase registered for the best epidemiological immunisation strategy, Hemoculture (Highest 3) (cf. the main text). Hence, serology performs worse than another epidemiological immunisation strategy and this is why it was not inserted for discussion in the main text. Notice also that Serology (Highest 6) performs almost equivalently to random immunisation.

Similarly, modifications of the Parasitised Mammals strategy by immunising only specific species according to their biology (e.g. Cricetii or prey) does not provide any evident improvement in terms of slowing down the parasite spread and increasing the infection time. As a result, Parasitised Cricetii and Parasitised Prey were not inserted in the main text. Similarly patterns were observed also in Pantanal.

Increasing the abundance of vector colonies in the ecomultiplex network leads to worse performances of the serology-based strategies. On the contrary, the measure based on Hematology provides equivalent result to the immunisation strategies based on the ecomultiplex network structure.

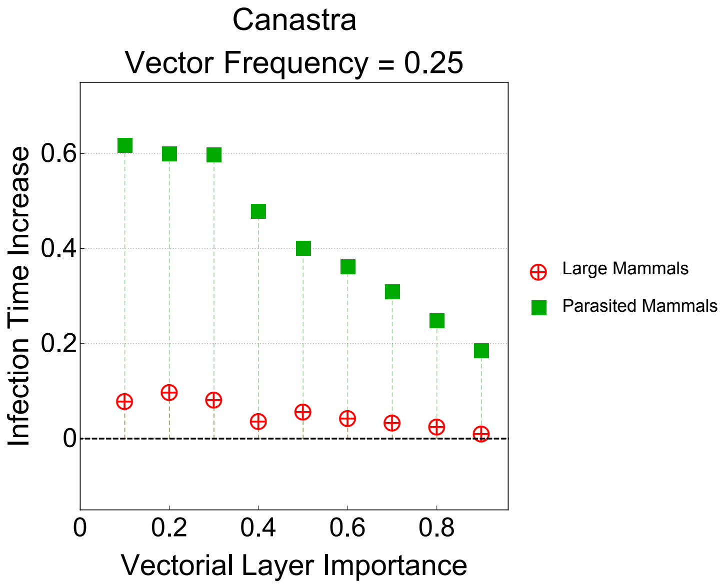


Figure 9.8: The best and the worst performing strategies presented in the main text for the Canastra ecosystem are here presented relatively to the null model with equal abundances. In this null model we consider an ecomultiplex network where all animal groups have equal abundance, i.e. occur with equal frequency. The error margins in the plot are the same size of the dots and are based on 500 iterations. Even providing equal abundances to different species does not remove the gap in global infection time that was observed in the main text.

9.6.4 A null model with equal abundances

In the ecomultiplex model animal groups of different species appear with a power-law frequency, so that animal groups of some species can be considerably more frequent than others. Previous findings indicated how abundance can indeed influence parasite spread among populations, so that a question can naturally arise: are the gaps in performances of immunisation strategies due just to heterogeneity of abundance distribution in the model?

In order to test this research question, we considered a null model equivalent to the scenarios described in the main text but where all species had the same body mass and hence the same group frequency. Results for the Canastra ecosystem are reported in Fig. 9.8. The results indicate that even providing equal abundances to different species does not remove the gap in global infection time observed in the main text. Therefore, we conclude that the inequality in abundances of different species cannot fully explain the gaps in immunisation strategies detected, which are rather considered being dependant on the topology of interactions in the ecomultiplex model. In fact, in the null model with equal abundances, animal groups do not differ for their frequency but rather for their topologies only.

Chapter 10

Discussion and Conclusions

> The road up and the road down are one and the same.
>
> *Fragment 60*
> Heraclitus of Ephesus

10.1 The need for multiplex frameworks in complexity science

The study of complex networks is crucial in investigating the structure and dynamics of complex systems through a multidisciplinary approach (Caldarelli, 2007; Newman, 2010). Despite the success of traditional network analysis, single-layer networks are limited to representing only one type of interactions occurring in a given complex system at a time. However, most real-world complex systems display interactions across different scales and levels. This additional complexity of multi-typed interactions or multiplexity can indeed have deep repercussions on the observed structural or dynamical patterns. Neglecting this multiplexity or considering only part of the available interactions in a system can severely reduce the understanding of complex systems (Kivelä et al., 2014; Boccaletti et al., 2014).

In order to overcome this limitation, the scientific community came up with more advanced network models such as multi-layer networks (De Domenico et al., 2013; Battiston et al., 2014, 2017; De Domenico et al., 2016), i.e. networks where different interactions are layered among different sets of agents. Multiplex networks preserve this multi-layer structure but they represent interactions among the same set of agents replicated across layers. The increasing availability of data and the multi-relational approach allow to use multiplex networks for modelling, detecting and understanding phenomena that are either absent or qualitatively different within single-layer networks. This is why multiplex networks rapidly became a powerful investigation tool for representing multiplex systems in the last few years s (De Domenico et al., 2013; Battiston et al., 2014; Kivelä et al., 2014; Boccaletti et al., 2014; Battiston et al., 2017; De Domenico et al., 2016).

Although the field of multiplex networks is a highly prolific and growing one within the physics community, there are still fields in which interdisciplinary multiplex modelling is either absent or quite uncommon. This is the main gap that motivated the suggestion of novel multiplex frameworks for the quantitative study of real-world systems within this thesis such as the mental lexicon and ecosystems.

10.2 The multiplex frameworks introduced in this thesis

Two main research areas were identified as fields where a traditional network approach had already been established but empirical work pointed out the need for a modelling approach accounting for multiplexity:

1. the investigation of the mental lexicon and cognitive development in psycholinguistics;
2. the investigation of parasite spread beyond food webs in network ecology.

In both the applications we focus on the representation of the interactions present in the system and enrich the resulting network representation with additional metadata, coming from the observation of the environment (e.g. word frequency or average body masses across populations). This perspective assumes that information on the dynamics of the whole system can be obtained when its interactions are observed and quantified, an approach that is also known also as ecological psychology in the psycholinguistic context (Clark, 2008) and which is viable for both the real-world systems analysed in this thesis.

10.3 Discussion of the key findings

10.3.1 Mental lexicon modelling through multiplex networks

The first part of research output of this thesis focuses on the above findings about the network representation of the mental lexicon. The work in this thesis focuses not only on the topological description of the structure of similarity among words in the lexicon but rather on the influence of this structure in cognitive dynamics such as word learning and mental navigation.

This part of thesis starts from phonology.

Phonological networks encode some specific constraints of language structure

In Chapter 4 of this thesis, we show that phonological networks are embedded in a combinatorial space, from which they inherit some network features as the heavy-tail degree distribution or the high assortative mixing by degree. We strongly believe this is a reasonable explanation for the partial results provided by Gruenenfelder & Pisoni (2009): phonological networks display some features that can be easily reproduced by random network models. In fact, what our investigation adds to this picture is that these random network models are embedded as well in the same combinatorial structure of empirical phonological networks: the features they reproduce are rather inherited from the underlying space and should not be considered as genuine linguistic patterns. By suggesting more refined assembly models, inspired by word derivation, we show how other features exhibited by phonological networks are exceedingly difficult to match even with network ensembles constrained on a core-periphery structure or in terms of phoneme correlations. Features that are difficult to match without external constraints are: (i) the cut-off on the degree and (ii) the distribution of local clustering coefficients.

Interestingly, both degree and local clustering relate to psycholinguistic empirical patterns. Degree in a phonological network indicates how many words sound similar to a given one while local clustering indicates how many sound similarities there are among words sounding similar to a given one. Experiments have shown that the more sound similarities in a neighbourhood the easier it is to confuse a word (Chan & Vitevitch, 2010), so that the detected patterns on degree and clustering could be the effect of a constraint limiting the confusability of words by limiting the number of connections locally available. Notice that, even if sound similarities on the one hand increase confusability and should not be desirable, they also seem to facilitate word recollection in long-term memory tasks (Vitevitch et al., 2012). Our quantitative network findings could therefore be interpreted as constraints in phoneme assembly resulting from the interplay of reducing word confusion while still facilitating word memorisation.

An important limitation of our phonological network models is that they are based on word similarities computed from the current English language. Our models cannot therefore account for the historical development of language over the centuries, although its analysis would be an interesting research direction. Another limitation of this network representation of the lexicon is that it relies on a rather restrictive measure of phonological similarity (Vitevitch, 1997), which captures “cat” and “can” as sounding similar but not “sight” and “sizes”. While phonological similarity has been extensively used in the psycholinguistic literature (Vitevitch, 1997; Luce & Pisoni, 1998; Vitevitch, 2008; Siew, 2013; Sadat et al., 2014), the scientific community is currently trying to overcome its restrictiveness by considering empirical definitions of sound similarities, i.e. designing experiments where participants provide subjective examples of sound similarities among words.

The interplay between phonology and semantics: multiplex lexical networks

Phonology is only one of several aspects of the mental lexicon. In Chapter 5 we go beyond a single-layer representation of the mental lexicon and use a multiplex network approach for investigating strategies of word acquisition in toddlers aged between 18 and 30 months. Within the resulting framework of multiplex lexical networks, we measure the prediction power that different network measures have in predicting the specific words learned by most of the children. We also measure the relative influence that different semantic and phonological aspects of word similarities have in word acquisition.

Multiplexity is a powerful predictor of early word learning

Results indicate that semantics and co-occurrence of words do have the same influence in determining the very first words learned by most of English toddlers, thus indicating that multiplexity is important early on for lexicon development. However, later in the development semantic free associations take the lead in predicting most of the words learned over time. This quantitative finding indicates that word learning through semantic associations of meanings starts at much earlier stages than the scholar age (between years 4 and 6) conjectured in empirical studies (Wiethan et al., 2014; Storkel, 2004).

Another quantitative indication of the importance of multiplexity is that the whole multiplex structure outperforms any individual layer in predicting word acquisition better than random guessing very early on during development (i.e. before month 23). The multiplex topology greatly outperforms even word frequency, which is considered to be the best predictor for age of acquisition of words in linear models (Kuperman et al., 2012). However not all multiplex measures are outstanding predictors: the multiplex closeness centrality outmatches even optimised combinations of intra-layer degrees (i.e. weighted versions of multidegree). This finding underlines the importance of considering a multiplex representation of the mental lexicon at a global rather than local level in quantifying cognitive development and its likely mechanisms. Our results confirm empirical findings about the importance of multi-relational information for word acquisition in toddlers (Dautriche et al., 2015).

Interestingly, phonology plays a different role compared to the other semantic network layers. When restricted to the same words in the other layers, the phonological one is too disconnected for producing any prediction different from random guessing. When extended to a larger corpus, similar in size to the lexicon of an adult speaker, phonology does influence word learning in the earliest phases of development, in agreement with previous findings (Carlson et al., 2014). This difference is another sign that possibly the measure of phonological similarity is too restrictive as it leads to networks noticeably worse in predicting word learning patterns than semantic networks made of the very same words.

It has to be underlined that our approach relies on speech production: children are considered to have acquired a word when they can pronounce it but it might be that cognitive acquisition happened at earlier stages. An important approximation made in this work is that the lexicon of the children approximates the learning environment they learn from. This modelling approach resulted being the most predictive one in previous work on modelling word acquisition in toddlers (Hills et al., 2009), hence we adopted it for the very first application of the framework of multiplex lexical networks. Other approaches are indeed possible and should be tested as future research directions. Another limitation of this work is that it predicts the average learning strategies of the majority of English toddlers, thus not addressing individual patterns and deviations from the average.

Multiplex lexical networks: from toddlers to adults

The work from Chapter 6 extends the analysis of the multiplex lexicon structure from early childhood to adulthood. Rather than focusing again on word acquisition, we rather focus on the emergence over time of structural patterns in the topology of multi-relational similarities among words and relate these patterns to mental navigation.

Multiplexity in adults confirms the presence of a lexicon core in the human mind

We corroborate previous conjectures about the presence of a network core within the mental lexicon by detecting a viable cluster (Baxter et al., 2016) in our multiplex network representation of the lexicon. This structure emerges through word acquisition via an explosive transition around age 7.7 years, an age considered as critical for cognitive development and logic reasoning in psycholinguistics (Piaget, 2000). Null models with random word acquisition or reshuffled word labels fail to reproduce this emergence, thus suggesting a cognitive importance for the cluster to appear within the lexicon. The analysis of psycholinguistic features of words in the cluster both indicate that it acts as a core for the network, i.e. a densely connected sub-region made of words being more frequent, more connected, more easily identifiable, with more semantic meanings and expressing more tangible concepts compared to words those outside of it. Hence, our work represents the first large-scale, quantitative confirmation of a core of words persisting across different aspects of the lexicon at once, whose presence was conjectured in previous psycholinguistic works (Barsalou, 2008; Solonchak & Pesina, 2015; Hanley et al., 2013) and in computational linguistics (Dorogovtsev & Mendes, 2001; Picard et al., 2009; i Cancho & Solé, 2001). According to such conjectures, a lexicon core would be advantageous in facilitating mental navigation among words and providing also semantic definitions of all words in a language starting from a few core meanings/words/concepts.

Our simulations with random attacks indeed indicate that the viable cluster facilitates mental navigation across words and layers in the networked representation of the lexicon and it makes it robust against progressive word failure, like in cognitive degradation due to impairments such as aphasia (Aitchison, 2012).

The approach in Chapter 6 is based on one approximation: we use the mental lexicon of adults and the age of acquisition of individual words for inferring the network structure of the mental lexicon up to 20 years earlier. This is reasonable as 3 out of 4 layers considered here are based on similarity rules that are defined independently of age: synonymy, generalisations and phonological similarities. Empirical evidence has shown that the layer of free associations might undergo a restructuring over time (De Deyne & Storms, 2008; Kenett et al., 2014). However this does not reduce the validity of our results, as we show that the viable cluster and its distinct patterns are robust against noise (e.g. random rewiring of small fractions of connections).

Additional limitations of multiplex lexical networks

It has to be noted that Chapters 4, 5 and 6 rely only on the English language but make general assertions about cognitive development. This is a limitation mainly due to the availability of corpora in languages different from English. This limitation has to be acknowledged and the increasing availability of linguistic data is definitely leading the research towards a multi-lingual approach. It has to be noted also that independent studies have confirmed some universal patterns in the structure of semantic (Youn et al., 2015) and phonological (Shoemark et al., 2016) networks of different languages, so that the current expectation is for the results pointed out in the above analysis to potentially hold across several languages.

Another limitation is that the multiplex network representations of Chapter 5 and 6 are based on a limited number of layers. Structural reducibility (De Domenico et al., 2015) indicates that the related multiplex networks are irreducible for both adults and children, so that considering fewer layers by aggregating any of them would have implied a loss of topological information about the structure of word similarities. However more layers could have been added, a critique that is in general true for every data-driven approach. In our case, the multiplex lexical networks we consider represent approximations of the mental lexicon encapsulating for the first time semantic and phonological information. It is on these types of similarities among words that we focused our first studies. Although one could have considered a generalisation layer also for children, we did not because of previous studies showing how toddlers are unable to understand generalisations of concepts (Piaget, 2000) before six or seven years of age. From a network perspective, considering generalisations in toddlers would have led to a highly disconnected layer over the same set of words present in the others. Rather than inserting generalisations in the multiplex network for toddlers we considered the whole generalisation layer of adults in Chapter 5. Further comparing the representations of mental lexicons of children and adults, in Chapter 6 we did not consider co-occurrences as it is highly debated in the relevant literature whether they provide more spurious relationships compared to syntactic dependencies in adults (Ninio, 2014; Zweig, 2015).

It has to be underlined that the framework of multiplex lexical networks remains general so that additional layers could be considered in future works, depending on data availability and overlap with the layers already considered here, in order to avoid structural reducibility. Testing the robustness of the above results to the addition of layers would be an interesting research direction. However, notice that, as shown in Chapter 5, each layer influences linguistic patterns with different magnitudes, with the association layer influencing word acquisition considerably more than feature norms and co-occurrences. In the same chapter we also showed that our results were robust against the addition of phonology to the multiplex model. Hence, we expect for our results to display a certain degree of robustness even in case of more layers being added to the multiplex representation.

The multiplex approach reported in the above Chapters has to be interpreted in the sense of the original definition of multiplexity in computational social science (Wasserman & Faust, 1994): although intra-layer edges of different types are allowed, inter-layer edges are not explicitly considered. This approximation relies on the difficulty of attributing or even interpreting a given cost for transitioning between different cognitive aspects. This problem is indeed common in multiplex modelling and although heuristics on how to weight inter-layer connections based on spectral features do exist (Boccaletti et al., 2014), their lack of cognitive interpretation motivated us in pursuing a simpler approach in the first multiplex studies suggested to the scientific community.

All in all, while the work reported in Chapter 3 presents theoretical models supporting a previously defined network topology, the multiplex framework from Chapters 5 and 6 represents a “first of its own” approach in cognitive network science: it fills a gap in the literature about the quantification of multi-relational correlations across semantics and phonology in the mental lexicon and their impact in terms of both cognitive development and word processing.

10.3.2 Ecosystems modelling through multiplex networks

The second part of the thesis deals with the ecological multiplex network modelling of parasite spreading in both fictional and real-world ecosystems. The multiplex structure allows to investigate the influence feeding interactions and parasitic contaminations have on spreading a parasite within a spatially embedded ecosystem.

Reconciling food webs and epidemiology in parasite spread: ecological multiplex networks

The work from Chapter 8 introduces a model of interacting communities on multiplex layers, where constituents have an identity and only interactions between certain identities are allowed on a given layer, provided that constituents are close enough in space. From an ecological perspective, constituents represent groups of animals dispersing over space, getting in touch with each other and interacting via two trophic channels: predator-prey relationships and blood exchange or parasitic interactions. Notice that parasitic interactions are technically “trophic” as well: they represent an energy exchange since the vector considered in the real-world ecosystem of Serra de Canastra feeds through blood (Velascohernandez, 1994; Poulin & Morand, 2014).

We use an SI model for the parasite spread mainly because if the infection is not treated within the first few hours, the parasitic infection basically lasts for the whole lifetime of the mammalian host (Jansen et al., 2015).

The quantitative results for both a fictional predator-prey-vector ecosystem and for the more complex trophic interactions in the Serra de Canastra environment indicate that parasite spread on a food web changes dramatically when parasitic interactions are considered. In addition, results point out the presence of an additive effect of predator-prey and vectorial interactions in accelerating parasite diffusion. Interestingly, this does not always happen but it rather emerges from a specific combination of frequency of vectors in the ecosystem and time scale (i.e. the likelihood of the disease spreading mainly through contaminative links rather than food-web ones). Such a quantification of the interplay between trophic interactions offers theoretical insights in previous empirical work suggesting a potential additive effect in the parasite spread in wild life (Rocha et al., 2013). In fact, our model indicates that considering more than one type of interactions does not trivially lead to quicker parasite spread, as this phenomenon crucially depends on the topology of interactions and the time scales of parasitic vs predator-prey interactions in the ecosystem.

Another interesting finding is evidence for a percolation phase transition, whose approximated critical threshold saturates around the critical value $f_c = f_v = 0.02$ when simulating ecosystems of size $N = 2.5 \cdot 10^4, 5 \cdot 10^4, 10^5, 1.5 \cdot 10^5$. We conjecture that this critical transition is due to vector populations being fundamental in infecting prey populations. If f_c vanished in the thermodynamical limit, then no vector populations would be necessary in the ecosystem for the parasite to spread across large portions of the network. But at $f_v = f_c = 0$ no vectors would be present in the system, thus the parasite would spread only from prey to predator nodes across directed, loop-less trees. Simulations indicate that these structures include only a very small fraction of nodes in the ecosystems. The addition of vectors to the ecosystem for $f_v = f_c > 0$

adds loops in the multiplex networks. We interpret the point f_c as the emergence in the multiplex network of a largest weakly connected component (De Domenico et al., 2014; Baxter et al., 2016) due to undirected connections provided by vectors and incorporating increasingly many nodes for $f_v > f_c$. It is the increasing frequency of vectors that increase the connectivity of the whole multiplex and also the number of nodes potentially reachable from the parasite diffusion. Hence, from an ecological perspective even a small but growing number of vectors in the ecosystem can lead to a phase transition on the percolation of the parasite across a large part of the ecosystem. This phenomenon represents a theoretical interpretation of what is called “incomplete treatment of vectors” in ecology (Patten, 1991; Noireau et al., 2009; Pelosse & Kribs-Zaleta, 2012): partial removal of vectors in an environment can still lead to pandemics even in case of small fractions of vectors left.

The ecomultiplex model assesses how individual species can facilitate parasite spread

Ecologists are usually interested in finding keystone species, i.e. species playing a pivotal role for the dynamics of an ecosystem.

Within the theoretical framework of ecomultiplex networks we also test the role that different species have in accelerating parasite spread in a given ecosystem. We started this analysis in Chapter 8 and better developed it in Chapter 9 by suggesting immunisation experiments.

In Chapter 8 we also better develop the idea of animal groups and relate their frequency to body masses, by using metabolic theory (Jetz et al., 2004). This modelling choice represents an ecological approach: similarly to the ecological psychology approach where words where learned according to their similarities with the learning environment, in here animals obtain energy from the surrounding environment in order to sustain their metabolism. The outcome of this interaction with the environment is the body mass, which regulates the density of animal species in the ecosystem (Jetz et al., 2004). We further develop this metabolic theory for obtaining the frequency of animal groups in the environment and implement it in the ecological multiplex network representation (or ecomultiplex representation, for brevity).

We also notice that the network topology is a powerful proxy for predicting which species are more exposed to parasitic infection and hence accelerate parasite diffusion, as it is mainly those with higher connectivity across both the trophic and vectorial layers. Vectors do indeed fulfil these requirements and their role of accelerating parasite spread is confirmed. However, our theoretical approach finds that other species accelerate parasite spreading as well, such as opossums (from the *Didelphidae* family) and cheetahs (from the *Leopardus* family), corroborating previous conjectures based on epidemiological measures of parasite prevalence in specific species (Herrera et al., 2011; Herrera & Urdaneta-Morales, 1992) and on the biology of the parasite development (Rocha et al., 2013; Roque et al., 2013).

Interestingly, we offer a theoretical mechanism potentially explaining why cheetahs have been recently found as prevalently infected with the parasite *T. cruzii* in wild life studies (Jansen et al., 2015). When considering only the food web for parasite spread, top predators such as the cheetah would be at the highest trophic level, predating on most of the other species and thus having a higher likelihood of catching the parasite by feeding on infected prey (cf. also the mean-field model from Chapter 8). On the food web cheetahs are top predators and cannot pass the parasite to other species. This is why in previous network studies based on food web only, top predators are considered as potentially hampering parasite diffusion (Wobeser, 2013): they eat infected animals and then die, removing infected substrate from the ecosystem in the process. However, on the ecomultiplex structure top predators are parasitised by the vector: cheetahs can catch the disease on the food web and then pass it to vectors on the vectorial layers. Since vectors do indeed facilitate parasite spread, top predators can do as well. We call this mechanism parasite amplification and it is a purely multiplex phenomenon: it cannot be observed when the food web and the vectorial layer are considered individually. Notice that in its current formulation the model does not include species removal due to predation, so that additional work is required before establishing if real-world top predators do indeed facilitate parasite transmission via parasite amplification or hamper parasite spread through predation.

Further limitations of the ecomultiplex framework

The main limitation of the ecomultiplex model is that it neglects meta-population dynamics: animal groups appearing/disappearing because of predation or lethal diseases originating from the parasite. This approximation crucially depends on the biology of the considered parasite. Although relatively little data is available about mortality rates in animal populations for the *T. cruzii* (Jansen et al., 2015), it is known that an infected animal host can spread the parasite already after 12 h since infection (Herrera & Urdaneta-Morales, 1992; Herrera et al., 2011) while in general extinction of animal groups due to predation is considered to happen with characteristic time scales of years (Wobeser, 2013). Furthermore, only 30% to 40% of mammalian hosts usually die out of complications due to the parasitic infection and this process still require years (Jansen et al., 2015). Hence, not considering a meta-population dynamics could be a reasonable approximation for modelling the spread of the *T. cruzii.*

We also show that those species that are found to accelerate parasite spread within the multiplex topology and the simulations are also those displaying a higher seropositivity to the parasite in empirically collected data (Herrera et al., 2011; Rocha et al., 2013). Seropositivity indicates that a given individual was infected with the parasite and developed antibodies for it during its lifetime. This agreement between data and simulations further corroborates the validity of a multiplex modelling approach for getting insights in the modelled ecosystem and possibly design efficient strategies on which species to vaccinate against the parasite for hampering its diffusion the most.

Another limitation is that the same mobility is considered across very different species, i.e. all the species have the same interaction radius when multiplex connections are built. This choice was motivated by the need of using random geometric graphs and metabolic scaling for reducing the number of parameters in the first instances of the model. As a future research direction, it would be interesting to add different mobility rates across species in the model.

Last but not least, the model is based on a limited number of ecological interactions obtained from field work. Therefore, more realistic predictions through the model do require higher resolution datasets of empirical ecological interactions among as many species as possible in a given ecosystem.

10.3.3 Similarities in multiplex approaches

Both the ecomultiplex models from Part II are similar in spirit to the linguistic models of Part I in that they offer a theoretical network representation for the quantification of real-world patterns independent of the details of the modelled system: nodes can represent either words or animal groups but their network topology can be studied with tools coming from network theory. Notice that our multiplex frameworks work in synergy with other disciplines for (i) getting inspiration from the modelling system, (ii) interpreting the modelling results and (iii) reconciling them with previous findings. This is also why the published works reported in the above Chapters featured international collaborations with other experts from cognitive theory and ecology.

10.4 Future research directions

This thesis presents new frameworks based on multiplex networks for quantifying patterns in real world systems. Being able to account for multiple relations at once in a network representation opens new, unexplored challenges both in cognitive network science and in network ecology.

10.4.1 Phonological networks

An interesting research direction for complementing the models introduced in Chapter 4 would be the investigation of networks of sound similarities based on measures more general than phonological similarity but with a given cognitive relevance. Candidates such as the fraction of common phonemes have been suggested in the past (Kapatsinski, 2006) but their cognitive interpretation remained largely unexplored by the cognitive scientific community.

As an alternative, Dautriche et al. (2016) suggested to consider orthographic similarity as a proxy for sound similarity, with words sounding similar if they are written the same. While this approach works reasonably well with languages having a close correspondence between orthography and phonology, such as Italian or German, this might not be the case for other languages such as English or Finnish. Despite this, using orthographic information is much more convenient for computing large-scale corpora of different languages, provided the scarcity of phonological transcriptions for languages other than English, French, Dutch and German. In their work, Dautriche et al. (2016) managed to analyse the approximated phonological structure of almost 100 languages by data mining Wikipedia webpages. Rather than assessing the network structure, the authors focused on the analysis of correlations between phonological and semantic distances among words, finding a consistent positive Pearson's correlation across all the investigated languages.

Instead, Shoemark et al. (2016) did focus on comparison of the network structure across 8 different languages. The authors confirmed the results in Chapter 4: those network features such as high assortativity and heavy tail degree distribution were consistently exhibited by phonological networks even across languages historically as different as Basque and Spanish and can therefore be interpreted as consequences of spatial embedding rather than linguistic patterns. Shoemark et al. (2016) also identified structural differences across languages against randomised null models equivalent to our Type 0 models from Chapter 4. In English, German, French, Dutch and Polish, less frequent words were found to be less likely of occurring within the largest connected component, in contrast with what happened with Spanish, Portuguese, and to a lesser extent Basque. Provided that both Dautriche et al. (2016) and Shoemark et al. (2016) partially filled the gap and took the novelty of cross-linguistic studies via phonological networks, the focus can be shifted to other future research directions.

10.4.2 Multiplex lexical networks

Word acquisition in Chapter 5 has been compared against the normative one, i.e. against the way most of the individual learners acquire novel word over time. This modelling approach was based on the data available in the study. However, with the advent of computer-driven experiments and online surveys such as the Amazon Mechanical Turk, longitudinal studies mapping the behaviour of individuals are becoming increasingly widespread in the literature.

Having access to how individual children learn words over time would allow to increase the resolution power of the multiplex lexicon analysis. The novelties of this this approach are mainly two. Firstly, if the learning strategies emerging from our multiplex framework hold also for individual children then they could be considered as universal. This would corroborate the conjecture that word learning reflects universal cognitive mechanisms that are encoded within the human mind and persist across different subjects (Aitchison, 2012; Beckage et al., 2011; Carlson et al., 2014).

In Chapter 5, we showed that the detected learning strategies are indeed resilient to noise on the acquisition order of words, which motivates analysis of further data for children in different learning environments. Secondly but more importantly, the presence or absence of word learning strategies (e.g. learning earlier words that are more central on multiplex lexicon structure) might be correlated to the presence of linguistic disorders for developing a novel methodology for early detection of linguistic impairments in toddlers younger than 3 years, something that is currently difficult to do with the current psycholinguistic procedures. For instance, late talkers might not present the observed high predictability of the multiplex features such as closeness, which performed quite well in the normative case. This approach would require simple interviews with individual children rather than EEG measurements as in alternative and recent approaches from neuroscience (Zare et al., 2016), where children's brain networks are built and their topology is correlated with the risk of developing language learning disorders via machine learning approaches. The cognitive and the neuroscience approaches could even work in synergy, for providing a higher accuracy of assessment. However, this research direction is strongly dependent on the availability of data.

Another similar data-driven question could be the testing of the observed patterns in Chapter 5 but based on different datasets for the normative age of acquisition. We tried using the Amazon Turk ages of acquisition from Chapter 6 but they revealed large inconsistencies with the normative ages of acquisition from the CHILDES dataset used in Chapter 5. We suspect this is because in the age-range between 18 and 30 months the ages of acquisitions from AmazonTurk suffer from a memory effect. In fact, in the Amazon Turk dataset (Kuperman et al., 2012), adults are asked to estimate when they did learn a specific words but memory cannot be considered to be chronologically reliable before age 3 years (Aitchison, 2012). This limitation is also what motivated our analysis on randomised ages of acquisition via smearing. It did not impair our study, as the emergence of the largest viable cluster started after the scholar age after age 7.7 years old and it was robust to smearing.

Another possible follow of Chapters 5 and 6 would be using again the same ecological psychology approach but by using the lexicon of adults as a representation of the learning environment rather than the children's one. This study would be analogous to what was done already with the extended phonological layer: the 529 words known by children would be compared and acquired according to their local topology on the adults' lexicon rather than on the toddlers' one. This methodology would not exploit a full multiplex representation but rather the local topology of it, at least partially. For instance, the closeness scores for the 529 examined words would still depend on the whole network structure and this would not only lead to potentially different results but also to a change of perspective, as it would give more weight to adults influencing directly word acquisition in children. Therefore, rather than a simple extension, this approach would become another approach of its own, exploiting a similar methodology but being based on a different perspective of word acquisition in toddlers known as preferential attachment. Previous models of preferential attachment on networks of free associations (Hills et al., 2009) suggested a minor role played by preferential attachment compared to the higher influence on word acquisition displayed by the toddlers' lexicon structure. The difficult interpretation for the "extended" network structure as a multiplex network and the previous findings in the literature motivated us in not pursuing this approach for the preliminary study but rather keep it a potential, separate follow up.

10.4.3 Ecomultiplex networks

In terms of the ecomultiplex models presented in Chapters 8 and 9, the most interesting future research direction is the presentation of the spatial model with identities among nodes in terms of a self-contained theoretical reference model for detecting correlations in real-world systems. By defining the available identities of nodes and the interaction matrices on the different layers, the model is general enough for describing also systems different from ecosystems. As an example, nodes could represent transportation stations have the tube, the bus station and the train station in geographical proximity. The transportation modes would then become multiplex layers and the identities could be larger or smaller stations in terms of crowd affluence. Analytical results about the degree distributions or the clustering coefficient on each layer can be retrieved, in order to have both analytical results and simulations within one reference model.

Another interesting research direction would be the addition of the meta-population dynamics in the modelled ecosystem. A viable approach would be the Markovian machinery recently suggested by Gómez-Gardenes et al. (2015) for epidemiology in social multiplex networks. Within this approach nodes would not represent animal groups any more but rather geographical patches, connected together if close enough in a given embedding space (e.g. the 2D Euclidean one as in the ecomultiplex models), as done also in other spatial models of food chains (Barter & T, 2016). Patches would have identities but now related to the species available in the ecosystem so that there would be one layer per species. Individual animals would have their own home patch on a given layer, according to their species, e.g. in one layer there would be only cheetahs occupying patches. Several animals could be in the same patch. At each time step, individuals would disperse in space and move through a random walk on the multiplex structure, interact with the other animals occupying the same patch and then going back home. Interactions could be either of predator-prey type or rather parasitic. As in the ecomultiplex model, by interacting with each other animals would transmit the parasite, so that an SI model could be simulated over time. The parasitic infection could start in one patch but then transmit across the whole ecosystem because of the mobility of the infected animals. In the SI, infected animals could then be removed, modelling a death dynamics and allowing to measure extinction trends in the model, which is a feature missing from the current ecomultiplex formulation. Although elaborated in its formulation, the model would be analytically treatable via a Markovian formulation. In Gómez-Gardenes et al. (2015), spectral analytical results matched very well agent-based models with significantly lower computational demands (e.g. computing the eigenvalues of the multiplex Laplacian). This approach represents a viable future research direction but it also feels extremely different from the original model formulation so that rather than a mere extension it should be developed as a work of its own: a generalised ecomultiplex model.

Bibliography

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

Alexander, K. A., Lewis, B. L., Marathe, M., Eubank, S., & Blackburn, J. K. (2012). Modeling of wildlife-associated zoonoses: Applications and caveats. *Vector-Borne and Zoonotic Diseases*, *12*(12), 1005–1018.

Altizer, S., Nunn, C. L., Thrall, P. H., Gittleman, J. L., Antonovics, J., Cunningham, A. A., Dobson, A. P., Ezenwa, V., Jones, K. E., Pedersen, A. B., et al. (2003). Social organization and parasite risk in mammals: Integrating theory and empirical studies. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1), 517–547.

Amancio, D. R., Oliveira Jr, O. N., & da F Costa, L. (2012). Using complex networks to quantify consistency in the use of words. *Journal of Statistical Mechanics: Theory and Experiment*, *2012*(01), P01004.

Amboni, M. P. d. M. (2007). *Dieta, disponibilidade alimentar e padrão de movimentação de lobo-guará, Chrysocyon brachyurus, no Parque Nacional da Serra da Canastra.* PhD thesis, Universidade Federal de Minas Gerais.

Anderson, R. M. & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control.* Oxford University Press.

Andreazzi, C. S., Antonioni, A., Goudarzi, A., Selakovic, S., & Stella, M. (2014). Disease spreading in spatial ecological multiplex networks. *Proceedings of the Santa Fe Institute Summer School 2014.*

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

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

Bansal, S., Grenfell, B. T., & Meyers, L. A. (2007). When individual behaviour matters: Homogeneous and network models in epidemiology. *Journal of the Royal Society Interface*, *4*(16), 879–891.

Bansal, S., Pourbohloul, B., & Meyers, L. A. (2006). A comparative analysis of influenza vaccination programs. *PLoS Medicine*, *3*(10), e387.

Barabási, A.-L. (2015). *Network Science*. Cambridge University Press.

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

Barbaresi, A. (2014). *Language-classified Open Subtitles (LACLOS): Download, extraction, and quality assessment*. PhD thesis, Berlin-Brandenburgische Akademie der Wissenschaften.

Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford University Press.

Baronchelli, A., Ferrer-i Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in Cognitive Sciences*, *17*(7), 348–360.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Barter, E. & T, G. (2016). Meta-food-chains as a many-layer epidemic process on networks. *Physical Review E*, *93*(2), 022303.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, *499*(1), 1–101.

Bassett, D. S. & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364.

Battiston, F., Cairoli, A., Nicosia, V., Baule, A., & Latora, V. (2016). Interplay between consensus and coherence in a model of interacting opinions. *Physica D: Nonlinear Phenomena*, *323*, 12–19.

Battiston, F., Nicosia, V., Chavez, M., & Latora, V. (2017). Multilayer motif analysis of brain networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(4), 047404.

Battiston, F., Nicosia, V., & Latora, V. (2014). Structural measures for multiplex networks. *Physical Review E*, *89*(3), 032804.

Battiston, F., Nicosia, V., & Latora, V. (2017). The new challenges of multiplex networks: Measures and models. *The European Physical Journal Special Topics*, *226*(3), 401–416.

Baxter, G. J., Cellai, D., Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. (2016). A unified approach to percolation processes on multiplex networks. In *Interconnected Networks*. Springer.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS One*, *6*(5), e19348.

Beckage, N. M., Aguilar, A., & Colunga, E. (2015). Modeling lexical acquisition through networks. *Proceedings of the 37th Conference of the Cognitive Science Society*.

Beckage, N. M. & Colunga, E. (2015). Language networks as models of cognition: Understanding cognition through language. In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. Springer.

Belgrano, A. (2005). *Aquatic Food Webs: An Ecosystem Approach*. Oxford University Press.

Bianconi, G. (2013). Statistical mechanics of multiplex networks: Entropy and overlap. *Physical Review E*, *87*(6), 062806.

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, *17*(6), 905–917.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendina-Nadal, I., Wang, Z., & Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, *544*(1), 1–122.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4), 175–308.

Bonvicino, C. R., Oliveira, J. d., & D'Andrea, P. S. (2008). *Guia dos roedores do Brasil, com chaves para gêneros baseadas em caracteres externos.* Rio de Janeiro: Centro Pan-Americano de Febre Aftosa-OPAS/OMS.

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

Brede, M. & de Vries, B. J. (2009). Networks that optimize a trade-off between efficiency and dynamical resilience. *Physics Letters A*, *373*(43), 3910–3914.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Bueno, A. d. A., Belentani, S. C. d. S., & Motta-Junior, J. C. (2002). Feeding ecology of the maned wolf, Chrysocyon brachyurus (Illiger, 1815)(Mammalia: Canidae), in the ecological station of Itirapina, São Paulo State, Brazil. *Biota Neotropica*, *2*(2), 1–9.

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., & Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, *464*(7291), 1025–1029.

Bullmore, E. & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198.

Buono, C., Alvarez-Zuzek, L. G., Macri, P. A., & Braunstein, L. A. (2014). Epidemics in partially overlapped multiplex networks. *PLoS One*, *9*(3), e92200.

Caldarelli, G. (2007). *Scale-free networks: Complex webs in nature and technology.* Oxford University Press.

Callaway, D. S., Hopcroft, J. E., Kleinberg, J. M., Newman, M. E., & Strogatz, S. H. (2001). Are randomly grown graphs really random? *Physical Review E*, *64*(4), 041902.

Cardillo, A., Gómez-Gardenes, J., Zanin, M., Romance, M., Papo, D., del Pozo, F., & Boccaletti, S. (2013). Emergence of network features from multiplexity. *Scientific Reports*, *3*, 1344.

Carlson, M. T., Bane, M., & Sonderegger, M. (2011). Global properties of the phonological networks in child and child-directed speech. In *Proceedings of the 35th Boston University Conference on Language Development.*

Carlson, M. T., Sonderegger, M., & Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory and Language*, *75*, 159–180.

Carvalho, F. G. N. & Santos, E. (2012). Predacisaio do roedor Calomys sp. (Cricetidae) pelo marsupial Monodelphis domestica (Didelphidae) em Buiquea, Brazil. *Biotemas*, *25*(3), 317–320.

Casas, B., Català, N., Ferrer-i Cancho, R., Hernández-Fernández, A., & Baixeries, J. (2016). The polysemy of the words that children learn over time. *arXiv preprint arXiv:1611.08807*.

Cavalcanti, G. N. (2010). *Biologia comportamental de Conepatus semistriatus (Carnivora, Mephitidae) em Cerrado do Brasil Central*. PhD thesis, PhD thesis, Universidade Federal de Minas Gerais.

Chan, K. Y. & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, *34*(4), 685–697.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press.

Colizza, V. & Vespignani, A. (2008). Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology*, *251*(3), 450–467.

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

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

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33*(4), 497–505.

Cong, J. & Liu, H. (2014). Approaching human language with complex networks. *Physics of Life Reviews*, *11*(4), 598–618.

Coura, J. (2005). Transmission of Chagasic infection by oral route in the natural history of Chagas disease. *Revista da Sociedade Brasileira de Medicina Tropical*, *39*, 113–117.

Cozzo, E., Banos, R., Meloni, S., & Moreno, Y. (2013). Contact-based social contagion in multiplex networks. *Physical Review E*, *88*(5), 050801.

Craft, M. E. & Caillaud, D. (2011). Network models: An underutilized tool in wildlife epidemiology? *Interdisciplinary Perspectives on Infectious Diseases*, *2011*.

Craft, M. E., Volz, E., Packer, C., & Meyers, L. A. (2009). Distinguishing epidemic waves from disease spillover in a wildlife population. *Proceedings of the Royal Society of London B*, *276*(1663), 1777–1785.

Csermely, P., London, A., Wu, L.-Y., & Uzzi, B. (2013). Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, *1*(2), 93–123.

Dale, P. S. & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 125–127.

Dalziel, B. D., Huang, K., Geoghegan, J. L., Arinaminpathy, N., Dubovi, E. J., Grenfell, B. T., Ellner, S. P., Holmes, E. C., & Parrish, C. R. (2014). Contact heterogeneity, rather than transmission efficiency, limits the emergence and spread of canine influenza virus. *PLoS Pathogens*, *10*(10), e1004455.

Damuth, J. (1981). Population density and body size in mammals. *Nature*, 699–700.

Daraganova, G., Pattison, P., Koskinen, J., Mitchell, B., Bill, A., Watts, M., & Baum, S. (2012). Networks and geography: Modelling community network structures as the outcome of both spatial and network processes. *Social Networks*, *34*(1), 6–17.

Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2016). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science*.

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, *143*, 77–86.

Davis, S., Abbasi, B., Shah, S., Telfer, S., & Begon, M. (2015). Spatial analyses of wildlife contact networks. *Journal of the Royal Society Interface*, *12*(102), 20141004.

Davis, S., Trapman, P., Leirs, H., Begon, M., & Heesterbeek, J. (2008). The abundance threshold for plague as a critical percolation phenomenon. *Nature*, 634–637.

De Deyne, S., Kenett, Y. N., Anaki, D., Faust, M., & Navarro, D. J. (2016). Large-scale network representations of semantics in the mental lexicon. In *Big data in cognitive science: From methods to insights*. Psychology Press.

De Deyne, S. & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213–231.

De Domenico, M. (2017). Multilayer modeling and analysis of human brain networks. *Giga Science*, *6*(5), 1–8.

De Domenico, M. & Biamonte, J. (2016). Spectral entropies as information-theoretic tools for complex network comparison. *Physical Review X*, *6*, 041062.

De Domenico, M., Granell, C., Porter, M. A., & Arenas, A. (2016). The physics of spreading processed in multilayer networks. *Nature Physics*, *12*, 901–906.

De Domenico, M., Lancichinetti, A., Arenas, A., & Rosvall, M. (2015). Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X*, *5*(1), 011027.

De Domenico, M., Nicosia, V., Arenas, A., & Latora, V. (2015). Structural reducibility of multilayer networks. *Nature Communications*, *6*, 6864.

De Domenico, M., Porter, M. A., & Arenas, A. (2014). MuxViz: A tool for multilayer analysis and visualization of networks. *Journal of Complex Networks*, *3*(2), 159–176.

De Domenico, M., Sasai, S., & Arenas, A. (2016). Mapping multiplex hubs in human functional brain networks. *Frontiers in Neuroscience*, *10*.

De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., Gomez, S., & Arenas, A. (2013). Mathematical formulation of multilayer networks. *Physical Review X*, *3*, 041022.

De Domenico, M., Solé-Ribalta, A., Gómez, S., & Arenas, A. (2014). Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences*, *111*(23), 8351–8356.

De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S., & Arenas, A. (2015). Ranking in interconnected multilayer networks reveals versatile nodes. *Nature Communications*, *6*, 6868.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.-F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600.

Dias, J. & Schofield, C. J. (1999). The evolution of Chagas disease (American trypanosomiasis) control after 90 years since Carlos Chagas discovery. *Memórias do Instituto Oswaldo Cruz*, *94*, 103–121.

Dickison, M., Havlin, S., & Stanley, H. E. (2012). Epidemics on interconnected networks. *Physical Review E*, *85*(6), 066109.

Dobson, A. (2004). Population dynamics of pathogens with multiple host species. *The American Naturalist*, *164*(5), 64–78.

Dobson, A., Lafferty, K. D., Kuris, A. M., Hechinger, R. F., & Jetz, W. (2008). Homage to Linnaeus: How many parasites? How many hosts? *Proceedings of the National Academy of Sciences*, *105*(19), 1482–1148.

Dorogovtsev, S. N. & Mendes, J. F. F. (2001). Language as an evolving word web. *Proceedings of the Royal Society of London B*, *268*(1485), 2603–2606.

D'Souza, R. M. & Nagler, J. (2015). Anomalous critical and supercritical phenomena in explosive percolation. *Nature Physics*, *11*(7), 531–538.

Dubey, J. (2004). Toxoplasmosis–a waterborne zoonosis. *Veterinary Parasitology*, *126*(1), 57–72.

Dunne, J. A., Lafferty, K. D., Dobson, A. P., Hechinger, R. F., Kuris, A. M., Martinez, N. D., McLaughlin, J. P., Mouritsen, K. N., Poulin, R., Reise, K., et al. (2013). Parasites affect food web structure primarily through increased diversity and complexity. *PLoS Biology*, *11*(6), e1001579.

Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002a). Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences*, *99*, 12917–12922.

Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002b). Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecology Letters*, *5*, 558–567.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*(7), 301–306.

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

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Petropolitanae*, *8*, 128–140.

Fay, D. & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, 505–520.

Fayer, R., Morgan, U., & Upton, S. J. (2000). Epidemiology of Cryptosporidium: Transmission, detection and identification. *International Journal for Parasitology*, *30*(12), 1305–1322.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*. AAAI Press.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), 1–185.

Fenton, A. & Pedersen, A. B. (2005). Community epidemiology framework for classifying disease threats. *Emerging Infectious Diseases*, *11*(12), 1815–1821.

Ferrari, M. J., Bansal, S., Meyers, L. A., & Bjørnstad, O. N. (2006). Network frailty and the geometry of herd immunity. *Proceedings of the Royal Society of London B*, *273*(1602), 2743–2748.

Ferrer-i Cancho, R. (2014). Beyond description: Comment on “Approaching human language with complex networks”. *Physics of Life Reviews*, *11*(4), 621–623.

Fontaine, C., Guimarães, P. R., Kéfi, S., Loeuille, N., Memmott, J., Van Der Putten, W. H., Van Veen, F. J., & Thébault, E. (2011). The ecological and evolutionary implications of merging different types of networks. *Ecology Letters*, *14*(11), 1170–1181.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers’ referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585.

Fronczak, A. (2014). Exponential random graph models. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of Social Network Analysis and Mining*. Springer.

Funk, S., Nishiura, H., Heesterbeek, H., Edmunds, W. J., & Checchi, F. (2013). Identifying transmission cycles at the human-animal interface: The role of animal reservoirs in maintaining gambiense human African trypanosomiasis. *PLoS Computational Biology*, *9*(1), e1002855.

Gallotti, R. & Barthelemy, M. (2014). Anatomy and efficiency of urban multimodal mobility. *Scientific Reports*, *4*, 6911.

Gemmetto, V. & Garlaschelli, D. (2015). Multiplexity versus correlation: The role of local constraints in real multiplexes. *Scientific Reports*, *5*, 9120.

Gomez, S., Diaz-Guilera, A., Gomez-Gardeñes, J., Perez-Vicente, C. J., Moreno, Y., & Arenas, A. (2013). Diffusion dynamics on multiplex networks. *Physical Review Letters*, *110*(2), 028701.

Gómez-Gardenes, J., de Barros, A. S., Pinho, S. T., & Andrade, R. F. (2015). Abrupt transitions from reinfections in social contagions. *Europhysics Letters*, *110*(5), 58006.

Gómez-Gardenes, J., Reinares, I., Arenas, A., & Floría, L. M. (2012). Evolution of cooperation in multiplex networks. *Scientific Reports*, *2*, 620.

Goñi, J., Arrondo, G., Sepulcre, J., Martincorena, I., de Mendizábal, N. V., Corominas-Murtra, B., Bejarano, B., Ardanza-Trevijano, S., Peraita, H., Wall, D. P., et al. (2011). The semantic organization of the animal category: Evidence from semantic verbal fluency and network theory. *Cognitive Processing*, *12*(2), 183–196.

Grassberger, P. (2015). Percolation transitions in the survival of interdependent agents on multiplex networks, catastrophic cascades, and solid-on-solid surface growth. *Physical Review E*, *91*(6), 062806.

Gravino, P., Servedio, V. D., Barrat, A., & Loreto, V. (2012). Complex structures and semantics in free word association. *Advances in Complex Systems*, *15*(03n04), 1250054.

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

Grimmett, G. & Stirzaker, D. (2001). *Probability and random processes.* Oxford University Press.

Gruenenfelder, T. M. & Pisoni, D. B. (2009). The lexical restructuring hypothesis and graph theoretic analyses of networks based on random lexicons. *Journal of Speech, Language, and Hearing Research*, *52*(3), 596–609.

Guillon, J., Attal, Y., Colliot, O., La Corte, V., Dubois, B., Schwartz, D., Chavez, M., & Fallani, F. D. V. (2017). Loss of inter-frequency brain hubs in Alzheimer's disease. *Scientific Reports*, *7*, 10879.

Guimerá, R. & Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *Nature*, *433*, 895–900.

Hagen, M., Kissling, W. D., Rasmussen, C., Carstensen, D., Dupont, Y., Kaiser-Bunbury, C., O'Gorman, E., Olesen, J., De Aguiar, M., Brown, L., et al. (2012). Biodiversity, species interactions and ecological networks in a fragmented world. *Advances in Ecological Research*, *46*, 89–120.

Halu, A., Mondragón, R. J., Panzarasa, P., & Bianconi, G. (2013). Multiplex PageRank. *PLoS One*, *8*(10), e78293.

Hanley, J. R., Hunt, R. P., Steed, D. A., & Jackman, S. (2013). Concreteness and word production. *Memory & Cognition*, *41*(3), 365–377.

Hatcher, M. J., Dick, J. T., & Dunn, A. M. (2006). How parasites affect interactions between competitors and predators. *Ecology Letters*, *9*(11), 1253–1271.

Haydon, D. T., Cleaveland, S., Taylor, L. H., Laurenson, M. K., et al. (2002). Identifying reservoirs of infection: A conceptual and practical challenge. *Emerging Infectious Diseases*, *8*(12), 1468–1473.

Herrera, H. M., Rocha, F. L., Lisboa, C., Rademaker, V., Mourão, G., & Jansen, A. (2011). Food web connections and the transmission cycles of Trypanosoma cruzi and Trypanosoma Evansi (Kinetoplastida, Trypanosomatidae) in the Pantanal region, Brazil. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, *105*(7), 380–387.

Herrera, L. & Urdaneta-Morales, S. (1992). Didelphis marsupialis: A primary reservoir of Trypanosoma cruzi in urban areas of Caracas, Venezuela. *Annals of Tropical Medicine & Parasitology*, *86*(6), 607–612.

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

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

Horn, A., Ostwald, D., Reisert, M., & Blankenburg, F. (2014). The structural–functional connectome and the default mode network of the human brain. *Neuroimage*, *102*, 142–151.

Hudson, P. J., Rizzoli, A., Grenfell, B. T., Heesterbeek, H., & Dobson, A. P. (2002). *The Ecology of Wildlife Diseases*. Oxford University Press.

i Cancho, R. F. & Solé, R. V. (2001). The small world of human language. *Proceedings of The Royal Society of London B*, *268*(1482), 2261–2265.

i Cancho, R. F. & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, *100*(3), 788–791.

Jansen, A. M., Xavier, S. C., & Roque, A. L. R. (2015). The multiple and complex and changeable scenarios of the Trypanosoma cruzi transmission cycle in the sylvatic environment. *Acta Tropica*, *151*, 1–15.

Jetz, W., Carbone, C., Fulford, J., & Brown, J. H. (2004). The scaling of animal space use. *Science*, *306*(5694), 266–268.

Jeub, L. G. S., Balachandran, P., Porter, M. A., Mucha, P. J., & Mahoney, M. W. (2015). Think locally, act locally: Detection of small, medium-sized, and large communities in large networks. *Physical Review E*, *91*, 012821.

Johnson, P. T., Dobson, A., Lafferty, K. D., Marcogliese, D. J., Memmott, J., Orlofske, S. A., Poulin, R., & Thieltges, D. W. (2010). When parasites become prey: Ecological and epidemiological significance of eating parasites. *Trends in Ecology & Evolution*, *25*(6), 362–371.

Johnson, P. T., Preston, D. L., Hoverman, J. T., & LaFonte, B. E. (2013). Host and parasite diversity jointly control disease risk in complex communities. *Proceedings of the National Academy of Sciences*, *110*(42), 16916–16921.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*(7181), 990–993.

Kaiser, M. (2008). Mean clustering coefficients: The role of isolated nodes and leafs on clustering measures for small-world networks. *New Journal of Physics*, *10*(8), 083042.

Kapatsinski, V. (2006). Sound similarity relations in the mental lexicon: Modeling the lexicon as a complex network. *Speech research Lab Progress Report*, *27*, 133–152.

Karuza, E. A., Thompson-Schill, S. L., & Bassett, D. S. (2016). Local patterns to global architectures: Influences of network topology on human learning. *Trends in Cognitive Sciences*, *20*(8), 629–640.

Keeling, M. (2005). The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, *67*(1), 1–8.

Keeling, M. J. & Eames, K. T. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, *2*(4), 295–307.

Keeling, M. J. & Rohani, P. (2008). *Modeling infectious diseases in humans and animals.* Princeton University Press.

Keesing, F., Holt, R. D., & Ostfeld, R. S. (2006). Effects of species diversity on disease risk. *Ecology Letters*, *9*(4), 485–498.

Kéfi, S., Berlow, E. L., Wieters, E. A., Joppa, L. N., Wood, S. A., Brose, U., & A, N. S. (2015). Network structure beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology*, *506*, 506.

Kéfi, S., Miele, V., Wieters, E. A., Navarrete, S. A., & Berlow, E. L. (2016). How structured is the entangled bank? The surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience. *PLoS Biology*, *14*(8), e1002527.

Kenett, Y. N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience*, *8*.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304.

Kiss, G., Armstrong, C. A., & Milroy, R. (1972). *An associative thesaurus of English.* Medical Research Council, Speech and Communication Unit, University of Edinburgh, Scotland.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, *2*(3), 203–271.

Kribs-Zaleta, C. (2006). Vector consumption and contact process saturation in sylvatic transmission of T. cruzi. *Mathematical Population Studies*, *13*(3), 135–152.

Kribs-Zaleta, C. (2010a). Alternative transmission modes for Trypanosoma cruzi. *Mathematical Biosciences and Engineering*, *7*(3), 657–673.

Kribs-Zaleta, C. (2010b). Estimating contact process saturation in sylvatic transmission of Trypanosoma cruzi in the United States. *PLoS Neglected Tropical Diseases*, *4*(4), e656.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990.

Lafferty, K. D., Allesina, S., Arim, M., Briggs, C. J., De Leo, G., Dobson, A. P., Dunne, J. A., Johnson, P. T. J., Kuris, A. M., Marcogliese, D. J., Martinez, N. D., Memmott, J., Marquet, P. A., McLaughlin, J. P., Mordecai, E. A., Pascual, M., Poulin, R., & Thieltges, D. W. (2008). Parasites in food webs: The ultimate missing links, *Ecology Letters*. *11*(6), 533–546.

Lafferty, K. D., Dobson, A. P., & Kuris, A. M. (2006). Parasites dominate food web links. *Proceedings of the National Academy of Sciences*, *103*(30), 1121–1211.

Laine, M. (2013). *Anomia: Theoretical and clinical aspects*. Psychology Press.

Lee, K.-M., Kim, J. Y., Cho, W.-k., Goh, K., & Kim, I. (2012). Correlated multiplexity and connectivity of multiplex random networks. *New Journal of Physics*, *14*(3), 033027.

Leech, G. (1992). 100 million words of English: The British National Corpus (BNC). *Language Research*, *28*(1), 1–13.

Lima, A., De Domenico, M., Pejovic, V., & Musolesi, M. (2015). Disease containment strategies based on mobility and information dissemination. *Scientific Reports*, *5*, 10650.

Liu, H. & Cong, J. (2014). Empirical characterization of modern Chinese as a multi-level system from the complex network approach. *Journal of Chinese Linguistics*, *42*(1), 1–38.

Lloyd-Smith, J. O., George, D., Pepin, K. M., Pitzer, V. E., Pulliam, J. R., Dobson, A. P., Hudson, P. J., & Grenfell, B. T. (2009). Epidemic dynamics at the human-animal interface. *Science*, *326*(5958), 1362–1367.

Luce, P. A. & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1).

MacWhinney, B. (2000). *The CHILDES project: The database*, volume 2. Psychology Press.

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication Theory*, *84*.

Marceau, V., Noël, P.-A., Hébert-Dufresne, L., Allard, A., & Dubé, L. J. (2011). Modeling the dynamical interaction between epidemics on overlay networks. *Physical Review E*, *84*(2), 026105.

Martinčić-Ipšić, S., Margan, D., & Meštrović, A. (2016). Multilayer network of language: A unified framework for structural analysis of linguistic subsystems. *Physica A: Statistical Mechanics and its Applications*, *457*, 117–128.

May, R. M. (1973). *Stability and complexity in model ecosystems*. Princeton University Press.

McCallum, H., Barlow, N., & Hone, J. (2001). How should pathogen transmission be modelled? *Trends in Ecology & Evolution*, *16*(6), 295–300.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559.

Melián, C. J., Bascompte, J., Jordano, P., & Krivan, V. (2009). Diversity in a complex ecological network with two interaction types. *Oikos*, *118*(1), 122–130.

Meyer, D. E. & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227.

Meyers, L. A., Pourbohloul, B., Newman, M. E., Skowronski, D. M., & Brunham, R. C. (2005). Network theory and SARS: Predicting outbreak diversity. *Journal of Theoretical Biology*, *232*(1), 71–81.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*(11), 39–41.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*(5594), 824–827.

Min, B., Do Yi, S., Lee, K.-M., & Goh, K.-I. (2014). Network robustness of multiplex networks with interlayer degree correlations. *Physical Review E*, *89*(4), 042811.

Mitchell, M. (2009). *Complexity: A guided tour.* Oxford University Press.

Morris, R. G. & Barthelemy, M. (2012). Transport on coupled spatial networks. *Physical Review Letters*, *109*(12), 128703.

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

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, *328*(5980), 876–878.

Myers, P., Espinosa, R., Parr, C., Jones, T., Hammond, G., & Dewey, T. (2008). The animal diversity web. *Last Accessed May 16, 2017.* http://animaldiversity. org.

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

Newman, M. (2010). *Networks: An introduction.* Oxford University Press.

Newman, M., Barabasi, A.-L., & Watts, D. J. (2011). *The structure and dynamics of networks.* Princeton University Press.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167–256.

Newman, M. E. (2012). Communities, modules and large-scale structure in networks. *Nature Physics*, *8*(1), 25–31.

Nicosia, V., Bianconi, G., Latora, V., & Barthelemy, M. (2013). Growing multiplex networks. *Physical Review Letters*, *111*(5), 058701.

Nicosia, V. & Latora, V. (2015). Measuring and modeling correlations in multiplex networks. *Physical Review E*, *92*(3), 032805.

Ninio, A. (2014). Syntactic networks, do they contribute valid information on syntactic development in children? *Physics of Life Reviews*, *11*, 632–634.

Noireau, F., Diosque, P., & Jansen, A. M. (2009). Trypanosoma cruzi: Adaptation to its vectors and its hosts. *Veterinary Research*, *40*(2), 1–23.

Omodei, E., De Domenico, M., & Arenas, A. (2017). Evaluating the impact of interdisciplinary research: A multilayer network approach. *Network Science*, *5*(2), 235–246.

Ostfeld, R. S. & Holt, R. D. (2004). Are predators good for your health? Evaluating evidence for top-down regulation of zoonotic disease reservoirs. *Frontiers in Ecology and the Environment*, *2*(1), 13–20.

Ozcan, M. (2012). Developmental differences in the naming of contextually non-categorical objects. *Journal of Psycholinguistic Research*, *41*(1), 51–69.

Packer, C., Holt, R. D., Hudson, P. J., Lafferty, K. D., & Dobson, A. P. (2003). Keeping the herds healthy and alert: Implications of predator control for infectious disease. *Ecology Letters*, *6*(9), 797–802.

Padgett, J. F. & Ansell, C. K. (1993). Robust action and the rise of the Medici, 1400-1434. *American Journal of Sociology*, *98*(6), 1259–1319.

Patten, B. C. (1991). Network ecology: Indirect determination of the life–environment relationship in ecosystems. *Theoretical Studies of Ecosystems: The Network Perspective*, 288–351.

Pelosse, P. & Kribs-Zaleta, C. (2012). The role of the ratio of vector and host densities in the evolution of transmission modes in vector-borne diseases: The example of sylvatic Trypanosoma cruzi. *Journal of Theoretical Biology*, *312*, 133–142.

Penczykowski, R. M., Laine, A.-L., & Koskella, B. (2016). Understanding the ecology and evolution of host–parasite interactions across scales. *Evolutionary Applications*, *9*(1), 37–52.

Perkins, S. E., Cattadori, I., & Hudson, P. J. (2005). The role of mammals in emerging zoonoses. *Mammal Study*, *30*(sp1), S67–S71.

Piaget, J. (2000). Piaget's theory of cognitive development. *Childhood Cognitive Development: The Essential Readings*, 33–47.

Picard, O., Blondin-Massé, A., Harnad, S., Marcotte, O., Chicoisne, G., & Gargouri, Y. (2009). Hierarchies in dictionary definition space. *arXiv preprint arXiv:0911.5703*.

Pielou, E. C. et al. (1969). *An Introduction to Mathematical Ecology.* Wiley-Interscience.

Pilosof, S., Greenbaum, G., Böhm, M., Krasnov, B. R., & Zelnik, Y. R. (2015). Asymmetric disease dynamics in multihost interconnected networks. *arXiv Preprint, arXiv:1512.09178.*

Pilosof, S., Morand, S., Krasnov, B. R., & Nunn, C. L. (2015). Potential parasite transmission in multi-host networks based on parasite sharing. *PLoS One*, *10*(3), e0117909.

Pilosof, S., Porter, M. A., Pascual, M., & Kéfi, S. (2017). The multilayer nature of ecological networks. *Nature Ecology & Evolution*, *1*, 0101.

Pimm, S. L. (1982). Food webs. In *Food webs*. Springer.

Pocock, M. J., Evans, D. M., & Memmott, J. (2012). The robustness and restoration of a network of ecological networks. *Science*, *335*(6071), 973–977.

Poulin, R. (2011). *Evolutionary ecology of parasites*. Princeton University Press.

Poulin, R. & Morand, S. (2014). Parasite biodiversity. *Smithsonian Institution*.

Price, K., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution: A practical approach to global optimization*. Springer Science & Business Media.

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

Rabinovich, J. E., Wisnivesky-Colli, C., Solarz, N. D., & Gürtler, R. E. (1990). Probability of transmission of Chagas disease by Triatoma infestans (Hemiptera: Reduviidae) in an endemic area of Santiago del Estero, Argentina. *Bulletin of the World Health Organization*, *68*(6), 737.

Reis, N. R., Peracchi, A. L., Pedro, W. A., & Lima, I. P. (2006). *Mamíferos do Brasil*. Universidade Estadual de Londrina.

Reithinger, R., Tarleton, R. L., Urbina, J. A., Kitron, U., Gürtler, & *et al*, R. E. (2009). Eliminating Chagas disease: Challenges and a roadmap. *British Medical Journal*, *338*.

Ristad, E. S. & Yianilos, P. N. (1998). Learning string-edit distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *20*(5), 522–532.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 173–191.

Rocha, F. (2006). *Área de uso e seleção de habitats de três espécies de carnívoros de médio porte na fazenda Nhumirin, e arredores, Pantanal da Nhecolândia, MS*. PhD thesis, Dissertação (Mestrado em Ecologia e Conservação)–Curso de Pós-graduação em Ecologia, Universidade Federal de Matogrosso do Sul, Campo Grande.

Rocha, F. L., Roque, A., de Lima, J. S., Cheida, C. C., Lemos, F. G., de Azevedo, F. C., Arrais, R. C., Bilac, D., Herrera, H. M., & Mourao, G. (2013). Trypanosoma cruzi infection in neotropical wild carnivores (Mammalia: Carnivora): At the top of the T. cruzi transmission chain. *PLoS One*, *8*(7), e67463.

Rocha, F. L., Roque, A. L. R., Arrais, R. C., Santos, J. P., Lima, V. D. S., Xavier, S. C. d. C., Cordeiro-Estrela, P., D'Andrea, P. S., & Jansen, A. M. (2013). Trypanosoma cruzi TcI and TcII transmission among wild carnivores, small mammals and dogs in a conservation unit and surrounding areas, Brazil. *Parasitology*, *140*(02), 160–170.

Roellig, D. M., Ellis, A. E., & Yabsley, M. J. (2009). Genetically different isolates of Trypanosoma cruzi elicit different infection dynamics in raccoons (Procyon lotor) and Virginia Opossums (Didelphis Virginiana). *International Journal for Parasitology*, *39*(14), 1603–1610.

Roque, A. L. R., Xavier, S. C., da Rocha, M. G., Duarte, A. C. M., D'Andrea, P. S., & Jansen, A. M. (2008). Trypanosoma cruzi transmission cycle among wild and domestic mammals in three areas of orally transmitted Chagas disease outbreaks. *The American Journal of Tropical Medicine and Hygiene*, *79*(5), 742–749.

Roque, A. L. R., Xavier, S. C., Gerhardt, M., Silva, M. F., Lima, V. S., D'Andrea, P. S., & Jansen, A. M. (2013). Trypanosoma cruzi among wild and domestic mammals in different areas of the Abaetetuba municipality (Pará State, Brazil), an endemic Chagas disease transmission area. *Veterinary Parasitology*, *193*(1), 71–77.

Ross, R. & Thomson, D. (1911). A case of sleeping sickness studied by precise enumerative methods: Further observations. *Proceedings of the Royal Society of London B*, 187–205.

Rushmore, J., Caillaud, D., Hall, R. J., Stumpf, R. M., Meyers, L. A., & Altizer, S. (2014). Network-based vaccination improves prospects for disease control in wild chimpanzees. *Journal of the Royal Society Interface*, *11*(97), 20140349.

Sadat, J., Martin, C. D., Costa, A., Alario, F., et al. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive psychology*, *68*, 33–58.

Sahneh, F. D., Scoglio, C., & Chowdhury, F. N. (2013). Effect of coupling on the epidemic threshold in interconnected complex networks: A spectral analysis. *American Control Conference (ACC)*, IEEE.

Salehi, M., Marzolla, M., Magnani, M., Sharma, R., Siyari, P., & Montesi, D. (2015). Diffusion processes on multilayer networks. *Network Science and Engineering*, *2*.

Santos, E. M. D. (2012). Predação do Roedor calomys sp.(Cricetidae) pelo marsupial Monodelphis domestica (Didelphidae) em Buíque–pe, Brasil. *Biotemas*, *25*(3), 317–320.

Santos, F. C. & Pacheco, J. M. (2005). Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters*, *95*(9), 098104.

Sanz, J., Xia, C.-Y., Meloni, S., & Moreno, Y. (2014). Dynamics of interacting diseases. *Physical Review X*, *4*(4), 041005.

Sattenspiel, L. (2009). *The Geographic Spread of Infectious Diseases: Models and Applications.* Princeton University Press.

Saumell-Mendiola, A., Serrano, M. Á., & Bogu ná, M. (2012). Epidemic spreading on interconnected networks. *Physical Review E*, *86*(2), 026106.

Schmidt, K. A. & Ostfeld, R. S. (2001). Biodiversity and the dilution effect in disease ecology. *Ecology*, *82*(3), 609–619.

Schofield, C. J. et al. (1994). *Triatominae: Biology & control.* Eurocommunica Publications.

Scott, J. (2012). *Social network analysis.* Sage.

Scotti, M., Ciocchetta, F., & Jordán, F. (2013). Social and landscape effects on food webs: A multi-level network simulation model. *Journal of Complex Networks*, *1*(2), 160–182.

Shikanai-Yasuda, M. A. & Carvalho, N. B. (2012). Oral transmission of Chagas diseas. *Clinical Infectious Diseases*, *54*(6), 845–852.

Shoemark, P., Goldwater, S., Kirby, J., & Sarkar, R. (2016). Towards robust cross-linguistic comparisons of phonological networks. *ACL 2016*, 110.

Siew, C. S. (2013). Community structure in the phonological network. *Frontiers in Psychology*, *4*, 553.

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

Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, *3*(2), 1–40.

Solá, L., Romance, M., Criado, R., Flores, J., del Amo, A. G., & Boccaletti, S. (2013). Eigenvector centrality of nodes in multiplex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *23*(3), 033131.

Solé, R. V. & Seoane, L. F. (2015). Ambiguity in language networks. *The Linguistic Review*, *32*(1), 5–35.

Solé-Ribalta, A. (2013). Spectral properties of the Laplacian of multiplex networks. *Physical Review E*, *88*(3).

Solonchak, T. & Pesina, S. (2015). Lexicon core and its functioning. *Procedia - Social and Behavioral Sciences*, *192*, 481–485.

Speakman, J. R. (2005). Body size, energy metabolism and lifespan. *Journal of Experimental Biology*, *208*(9), 1717–1730.

Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, *8*(9), 418–425.

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology*, *1*(4), e42.

Stamer, M. K. & Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning spanish as a foreign language. *Bilingualism: Language and Cognition*, *15*(03), 490–502.

Stauffer, D. & Aharony, A. (1991). *Introduction to percolation theory*. Taylor and Francis.

Stella, M., Andreazzi, C. S., Selakovic, S., Goudarzi, A., & Antonioni, A. (2016). Parasite spreading in spatial ecological multiplex networks. *Journal of Complex Networks*, *5*(3), 486–511.

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, *7*.

Stella, M., Beckage, N. M., Brede, M., & De Domenico, M. (2017). Multiplex model of mental lexicon reveals explosive learning in humans. *arXiv preprint arXiv:1705.09731*.

Stella, M. & Brede, M. (2015). Patterns in the English language: Phonological networks, percolation and assembly models. *Journal of Statistical Mechanics*, *2015*, P05006.

Stella, M. & Brede, M. (2016a). Investigating the phonetic organisation of the English language via phonological networks, percolation and Markov models. In *Proceedings of ECCS 2014*. Springer.

Stella, M. & Brede, M. (2016b). Mental lexicon growth modelling reveals the multiplexity of the English language. In *Complex Networks VII*. Springer.

Stella, M., Selakovic, S., Antonioni, A., & Andreazzi, C. S. (2017). Community interactions determine role of species in parasite spread amplification: the ecomultiplex network model. *arXiv preprint arXiv:1706.05121*.

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

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, *29*(02), 251–274.

Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, *25*(02), 201–221.

Streicker, D. G., Fenton, A., & Pedersen, A. B. (2013). Differential sources of host species heterogeneity influence the transmission and control of multihost parasites. *Ecology Letters*, *16*(8), 975–984.

Szell, M., Lambiotte, R., & Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, *107*(31), 13636–13641.

Towlson, E. K., Vértes, P. E., Ahnert, S. E., Schafer, W. R., & Bullmore, E. T. (2013). The rich club of the C. elegans neuronal connectome. *The Journal of Neuroscience*, *33*(15), 6380–6387.

Velascohernandez, J. X. (1994). A model for Chagas disease involving transmission by vectors and blood transfusion. *Theoretical Population Biology*, *46*(1), 1–31.

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech*, *40*(3), 211–228.

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

Vitevitch, M. S., Chan, K. Y., & Goldstein, R. (2014). Insights into failed lexical retrieval from network science. *Cognitive Psychology*, *68*, 1–32.

Vitevitch, M. S., Chan, K. Y., & Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language*, *67*(1), 30–44.

Wang, Z., Wang, L., & Perc, M. (2014). Degree mixing in multilayer networks impedes the evolution of cooperation. *Physical Review E*, *89*(5), 052813.

Wasserman, S. & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

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

Webster, J. P., Borlase, A., & Rudge, J. W. (2017). Who acquires infection from whom and how? Disentangling multi-host and multi-mode transmission dynamics in the elimination era. *Philosophical Transactions of the Royal Society B*, *372*(1719), 20160091.

Weibull, J. W. (1997). *Evolutionary game theory.* MIT press.

Wiethan, F. M., Nóro, L. A., & Mota, H. B. (2014). Early lexical and phonological acquisition and its relationships. In *CoDAS, Sociedade Brasileira de Fonoaudiologia.* SciELO Brasil.

Wobeser, G. A. (2013). *Essentials of disease in wild animals.* John Wiley & Sons.

WolframResearch (2017). WordData source information. *Last Accessed: 1 Dec. 2017.* http://reference.wolfram.com/language/note/WordDataSourceInformation.html.

Wood, C. L., Lafferty, K. D., DeLeo, G., Young, H. S., Hudson, P. J., & Kuris, A. M. (2014). Does biodiversity protect humans against infectious disease? *Ecology*, *95*(4), 817–832.

Xu, F. & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245.

Yamagata, Y. & Nakagawa, J. (2006). Control of Chagas disease. *Advances in Parasitology*, *61*, 129–165.

Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., Croft, W., & Bhattacharya, T. (2015). On the universal structure of human lexical semantics. *arXiv preprint arXiv:1504.07843.*

Zare, M., Rezvani, Z., & Benasich, A. (2016). Automatic classification of 6-month-old infants at familial risk for language-based learning disorder using a support vector machine clinical neurophysiology. *Clinical Neurophysiology*, *127*(7), 2695–2703.

Zhao, D., Li, L., Peng, H., Luo, Q., & Yang, Y. (2014). Multiple routes transmitted epidemics on multiplex networks. *Physics Letters A*, *378*(10), 770–776.

Zock, M. (2015). Words in books, computers and the human mind. *Journal of Cognitive Science*, *16*(4), 355–378.

Zweig, K. A. (2015). Are word-adjacency networks networks? In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks.* Springer.