Universidad de Granada

Facultad de Ciencias

On the Architectural Features of Ecological and Biological Networks

Tesis presentada por Virginia Domínguez García
para obtener el grado de Doctora en Física y Ciencias del Espacio

2014

Departamento de Electromagnetismo y Física de la Materia
e Instituto Carlos I de Física Teórica y Computacional

Editor: Universidad de Granada. Tesis Doctorales
Autor: Virginia Domínguez García
ISBN: 978-84-9125-031-9
URI: http://hdl.handle.net/10481/39822

ugr | Universidad de Granada

La doctoranda Virginia Domínguez García y el director de la tesis D. Miguel Ángel Muñoz Martínez, catedrático de Universidad,

GARANTIZAMOS: al firmar esta tesis doctoral, *On the Architectural Features of Ecological and Biological Networks* ("Sobre las características arquitectónicas de redes ecológicas y biológicas"), que el trabajo ha sido realizado por el doctorando bajo la dirección del director de la tesis. Y hasta donde nuestro conocimiento alcanza, en la realización del trabajo se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones, así como que el doctorando ha disfrutado de una estancia en el extranjero, durante un periodo de tres meses, en el *Department of Engineering Mathematics* de la Universidad de Bristol (Reino Unido).

Granada, a 20 de Noviembre de 2014.

Director de la tesis: Doctorando:

Fdo: Miguel Ángel Muñoz Martínez Fdo: Virginia Domíguez García

Contents

1 Introduction **6**

Introducción **14**

2 Topological features of ecological and biological networks **22**

2.1 Unimodal networks 22

2.1.1 Community structure 24

2.1.2 Hierarchy 26

2.1.3 Coherence 27

2.1.4 Feedback loops and motifs 29

2.1.5 Intervality 30

2.2 Bipartite networks 32

2.2.1 Degree distribution 33

2.2.2 Community structure 34

2.2.3 Nestedness 34

3 Food webs coherence determines stability **37**

3.1 Trophic coherence and stability 38

3.1.1 Measuring stability 41

3.2 Preferential Preying:
foodwebs with tunable coherence 49

3.2.1 The model 50

3.2.2 Preferential preying performance 51

3.3 The origins of stability 55

3.3.1 Effect of biomass distribution 59

3.3.2 Efficiency . . . 59
3.3.3 Weighted networks . . . 62
3.4 May's paradox . . . 64
3.4.1 Effect of biomass distribution . . . 66
3.5 Chapter summary . . . 68
3.6 Data supplement to chapter 2 . . . 69

4 Inherent directionality in biological and ecological networks 74
4.1 A model of network directionality . . . 76
4.1.1 Number of ascents and Eulerian cyclic numbers . . . 80
4.1.2 Devising an asymptotic result . . . 81
4.2 Infering directionality:
Counting loops in empirical networks . . . 82
4.2.1 A novel algorithm for measuring loops in networks . . . 87
4.2.2 Network randomizations . . . 88
4.3 Measuring directionality in empirical networks . . . 90
4.3.1 Algorithm details . . . 92
4.4 Chapter summary . . . 94
4.5 Data supplement to chapter 3 . . . 96

5 Factors determining nestedness in complex networks 99
5.1 Analytical quantification of nestedness . . . 101
5.1.1 introducing a refined measure . . . 101
5.1.2 Nestedness in the configuration model . . . 103
5.2 Nestedness in finite-size random networks . . . 106
5.2.1 Emergence of effective correlations in finite-size networks . . . 106
5.2.2 Effective correlations imply nestedness in finite networks . . . 107
5.3 Nestedness in empiric networks . . . 109
5.3.1 Degree correlations in real vs randomized networks . . . 109
5.3.2 Nestedness in real vs randomized networks . . . 111
5.3.3 Nestedness vs degree correlations in empirical networks . . . 112
5.3.4 A more refined null model . . . 113
5.4 Chapter summary . . . 116

5.5 Data supplement to chapter 4 117

6 Ranking species in mutualistic networks 122

6.1 MusRank: non-linear ranking algorithm for mutualistic networks 125

6.1.1 Other algorithms used in the study 126

6.2 Assessing the quality of a given ranking 128

6.3 MusRank performance 129

6.3.1 Optimally packed matrices 132

6.4 Chapter summary 134

6.5 Data supplement to chapter 5 135

7 Conclusions 138

Conclusiones 145

A Devising an interaction matrix from an adjacency matrix 151

B Mimicking reality: foodweb modeling 155

C Possible amendments to the PPM 162

D Analytical theory for maximally coherent networks 164

E Degree-degree correlations and nestedness in heterogeneous networks 167

Bibliography 170

Chapter 1

Introduction

Many interesting phenomena occurring in living systems emerge as the result of the interaction among many constitutive elements. Memory is an attribute of the whole brain network; it cannot be simply inferred from the properties of independent neurons, even assuming a perfect understanding of their one-to-one interactions. Response of cells to changing environments is regulated by their whole metabolic network, composed by many different interdependent metabolic pathways, which means that under different external conditions the cell reaction can be dramatically different. The coherent behaviour of large moving flocks of birds or fish is not easily understood when one just investigates isolated individuals. These three examples show a common trend: the global emergent behaviour of these systems cannot be trivially explained in terms of the features of their individual components. Even more remarkably, details of individual interactions do not play a determinant role on the emergent outcome; regardless of whether one studies synchronisation of fireflies, circadian rhythms, neural activity, or even economic cycles, it is possible to neglect specific details of the interactions between animals, cells, neurons or companies and treat all these elements as oscillators of some kind, and in many cases, similar phenomenologies emerge. Thus, these examples are what we call "complex systems", and as such, they should be tackled from a general "systemic approach". In other words, they seem natural candidates to be studied within the general framework of statistical mechanics, which studies noting but the "macroscopic" or "collective" properties using knowledge from the "microscopic" or 'individual" elementary components. In particular, one expects that the traditional methods

and tools of statistical mechanics –complemented with novel and specifically devised ones– may constitute the most appropriate framework to study, within a quantitative perspective, the fascinating collective properties of living systems and communities of them. In this way statistical mechanics has become, in the last decade or so, a highly multidisciplinary discipline.

Within this system approach one of the concepts that pervades all fields and has proven extremely valuable is that of "complex networks". For example, in the case of molecular biology, recent years have witnessed a transformation from studies limited to the specific traits of a few particular genes into the development of extensive global genetic maps that describe how such genes regulate (activate/repress) each other [63].

Networks are useful descriptors of complex systems that evidence the interactions or links among their individual components, thus defining characteristic architectures. Complex networks observed in nature have highly non-trivial structural properties, such as compartmentalisation, clustering, scale-free connectivity patterns, etc. Scrutinising how these traits influence the behaviour of the underlying dynamical systems and understanding how such features come spontaneously about are mayor key goals in modern complex systems theory.

A systemic (or "holistic") framework like this is conceptually useful for at least two reasons:

First, a similar network architecture, shared by two completely different types of systems, may arise from common requirements for proper functioning, such as overall stability and robustness, or similar assembling processes. This allows us to compile theoretical knowledge, to gain insight, as well as to make specific predictions.

Second, network structure greatly influences its response to perturbations. For example, the robustness of a given ecosystem against species extinctions is highly dependent on its connectivity pattern, given that this underlies the sequence of possible secondary extinctions that the first can potentially induce [6]. Similarly, in gene networks, it is believed that the network architecture is such that its outcome is robust to small mutations, individual gene knock-outs or to environmental fluctuations. More in general, different dynamical models operating on top of underlying networked systems have been reported to show qualitatively different behaviour depending on the network degree distribution [192].

The network or systemic approach can be straightforwardly applied to theoretical ecology, were ecosystems consist of a large number of species interacting –in different ways, such as, predation, mutualism, parasitism, etc– with each other, following quite specific patterns. Ecosystems fall into the category of complex systems since the behaviour of a structured community as a whole cannot be straightforwardly inferred from the properties of the individual species. To gain insight on how species-rich communities operate and are sustained one has to look first at the structure of the complex network they form. It is the study of the whole system –and not that of the individual pairwise interactions– what provides crucial knowledge about its global functioning.

All these ideas and concepts, are very useful to rationalise ecological systems [9, 10, 27, 167]. Indeed, the main goal of this thesis, is to analyse ecological complex communities from a network viewpoint. We will scrutinise their very distinctive architectural traits and make an attempt to correlate them with prominent features such as stability, resilience, robustness, and, in general, biodiversity. Our studies focus on ecological networks, but some aspects of biological and technological systems are also analysed up to a much lesser extent.

The idea of a complex network of interactions among species is as old as Darwin's contemplation of the tangled bank [62]. Lindenmayer, Odum, Margalef, and many others described ecological communities as graphs of energy transfer. Feeding interactions within a community can also be described in that form. Indeed, foodwebs depict trophic relationships between species, that is, who eats whom. These are directed networks were the species are represented by the nodes, and feeding interactions by directed links, encoding the aforesaid energy transference from the prey species to the predator.

Difficult to sample and difficult to model, foodwebs are nevertheless of central practical and theoretical importance. The interactions involving species on different trophic levels mediate species' responses to natural or intentional perturbations such as habitat loss or species extinction. Understanding the ecology and mathematics of foodwebs, and more broadly, ecological networks, is central to understanding the ecosystems response to perturbations, and may serve as a first step toward a more predictive ecology.

For several decades up to the 1970*s* a dominant ecological paradigm was that complex communities were more stable than simple ones [79, 141]. The argument in favour

of complexity fostering stability in ecological communities was presented in a general way by MacArthur [141], who stated that "a large number of paths through each species is necessary to reduce the effects of overpopulation of one species". He concluded that "stability increases as the number of links increases" and that stability is easier to achieve in more diverse communities of species, thus linking community stability with both high number of trophic links and increased numbers of species. This convention was challenged by May in a seminal paper [150] were, employing local stability analyses of randomly assembled community matrices, he demonstrated that network stability decreases with complexity. May found that simple, abstract communities of interacting species will tend to change sharply from stable to unstable behaviour as the complexity of the system increases. This apparent contradiction between the existence of large and complex ecosystems and the instability associated to those two features started was is known as the "diversity-stability debate". Many relevant questions arise around this controversy. Where does the exceptional stability of the ecosystems come from? Are there any special feature in foodwebs architecture that could account for it? Taken together recent ecological advances indicate that diversity can be expected, on average, to give rise to ecosystem stability [156], however this is not a settled issue and the debate remains open. Unravelling which features of foodwebs structure are involved in the enhanced stability of the ecosystems, if any, is of central importance in ecology, and one of the main issues we assess along this thesis.

A feature long related to system stability is that of the remarkable absence of loops (closed path of nodes) in empirical networks. Feedback loops are well-known to have a profound impact on dynamical stability in food webs [7, 21, 151, 164, 172, 173] as well as in biological and generic networks [4, 12, 15, 35, 60, 124, 136, 143, 180, 198, 201, 216, 248]. Therefore their absence has been frequently associated with the enhanced stability of these systems [7, 35, 121, 140, 164, 172]. However, is this the only explanation or could some other feature be behind the absence of feedback loops in directed networks? This is another of the points to be covered along this thesis.

Apart from the feeding relations described before, there are many other kind of interactions between species in ecosystems. Of particular relevance are mutualistic interactions. Ubiquitous in nature, up to 90% of tree species in the tropics depend on interactions with animals to complete their life cycles, either through pollination of

flowers or dispersal of seeds [119]. These interactions are mutually beneficial: plants obtain the dispersing services of the animals, and the animals, in turn, obtain food or other benefits in exchange for their services. Historically, the first studies on mutualism focused on highly specialised interactions. The examples of an almost perfect matching between the morphology of a flower and that of an insect that pollinates it seemed to be accepted as a common feature of coevolution. However, these examples of extreme, pairwise specialisation may more often be the exception than the rule. Mutualistic interactions can involve dozens or even hundreds of species interacting in complex ways, with different levels of specialisation [118, 259]

Mutualistic communities, such as plant-pollinator or seed-disperser communities, can be described as a network of interactions between mutualistic partners (e.g. plants and animals). However there is a marked difference between foodwebs and mutualistic networks. As said before foodwebs are represented as directed networks, links depicting who eats whom, with only a single type of node. In principle, all species could be connected to any other. Mutualistic networks are described by means of bipartite graphs, which encode the relationships between (but not within) two distinct sets: plants and animals.

Plant-animal mutualistic networks exhibit common features among different communities. Interestingly enough, the same architecture appears across systems regardless of the mutualism type (pollination or seed dispersal), geographic location, or species composition [29, 120]. Many of these topological properties found in mutualistic networks are emergent properties that result from the complex interaction between plants and animals across time and space [264]. Inspection of plant-frugivore and plant-pollinator networks evidenced that these networks are neither randomly assembled nor organised in compartments arising from tight, reciprocal specialisation [29]. Plant-animal mutualistic networks are highly *nested*, that is, species interaction are arranged in such a way that specialists (species that have few mutualistic partners) interact only with generalists (species with many mutualistic partners).

This particular feature has been associated with enhanced stability and coexistence of mutualistic communities [33, 181], suggesting that nestedness has been, to some extent, selected for by the forces of nature. However this effect is challenged in a recent work [234] were the opposite is found: nested networks tend to promote instability.

Other evidence indicate that neutral processes also have an important influence on the architecture of mutualistic communities [129, 263], meaning that species abundance distributions and random interactions between individuals may be the determinants of mutualistic network properties. Many questions arise regarding this unexpected property. Which are the factors that determine the nestedness of these important networks sustaining much part of earth's biodiversity? Is it necessary to call upon optimisation processes to obtain nested networks [242] or can neutral theories explain the nestedness present in empirical networks [129, 263]? How does this connectivity patter affect other network features? If we were to decide on a protection policy, what species should go first? the most connected? the most central? the better connected? Those are the some of the issues we consider along this thesis regarding mutualistic communities.

In this work we use the system approach of the statistical mechanics of complex networks to answer some relevant questions regarding how does the architecture of biological and ecological networks affects their performance. Could there be some properties of foodweb structure that explains its stability? Is stability the solely responsible for the "tree-like" appearance of many biological and ecological networks, or could any other feature explain this? What are the main factors determining the nestedness of a mutualistic community? Is it possible to identify the importance of the species in a mutualistic network with the only information provided by its connectivity pattern? These are the main questions we will investigate trough this thesis.

Summary

We reserve **chapter two** to briefly introduce the most relevant topological features of ecological and biological networks that have been exposed by close inspection and statistical analyses of empirical data. What are the main features of foodwebs and other ecological networks? Are some of these properties shared by other kind of biological networks? Do these features play some special role in the network performance? That and similar questions is what we will cover in the first chapter.

Chapter three presents an analysis of the stability-complexity debate. Why are large, complex ecosystems stable? Here we show that trophic coherence – a hitherto ignored feature of food webs which current structural models fail to reproduce – is

a better statistical predictor of linear stability than size or complexity. Furthermore, we prove that a maximally coherent network with constant interaction strengths will always be linearly stable. We also propose a simple model which, by correctly capturing the trophic coherence of food webs, accurately reproduces their stability and other basic structural features. Most remarkably, our model shows that stability can increase with size and complexity. This suggests a key to May' s Paradox, and a range of opportunities and concerns for biodiversity conservation.

In **chapter four** we explore the hypothesis that the presence of feedback loops in many empirical complex networks is severely reduced owing to the presence of an inherent global directionality. Aimed at quantifying this idea, we propose a simple probabilistic model in which a free parameter γ controls the degree of inherent directionality. Upon strengthening such directionality, the model predicts a drastic reduction in the fraction of loops which are also feedback loops. To test this prediction, we extensively enumerated loops and feedback loops in many empirical biological, ecological and socio-technological directed networks. We show that, in almost all cases, empirical networks have a much smaller fraction of feedback loops than network randomisations. Quite remarkably, this empirical finding is quantitatively reproduced, for all loop lengths, by our model by fitting its only parameter γ. Moreover, the fitted value of γ correlates quite well with another direct measurement of network directionality, performed by means of a novel algorithm. We conclude that the existence of an inherent network directionality provides a parsimonious quantitative explanation for the observed lack of feedback loops in empirical networks.

In **chapter five** we move on to mutualistic communities. Along this chapter we suggest a slightly refined version of the measure of nestedness and study how it is influenced by the most basic structural properties of networks, such as degree distribution and degree-degree correlations (i.e. assortativity). We find that most of the empirically found nestedness stems from heterogeneity in the degree distribution. Once such an influence has been discounted – as a second factor – we find that nestedness is strongly correlated with disassortativity and hence – as random networks have been recently found to be naturally disassortative – they also tend to be naturally nested just as the result of chance.

Assessing the importance of any given species in mutualistic networks is a key task

when evaluating extinction risks and possible cascade effects. In **chapter six** we develop a technique to measure the importance of species in mutualistic networks. Inspired in a recently introduced algorithm –similar in spirit to Google' s PageRank but with a built-in non-linearity– here we propose a method which allows us to derive a sound ranking of species importance in mutualistic networks. This method clearly outperforms other existing ranking schemes and can become very useful for ecosystem management and biodiversity preservation, where decisions on what aspects of ecosystems to explicitly protect need to be made.

Network research has become a fundamental framework for the study of complex systems across many fields. Here we deal with its application to biological systems and ecological communities. The fact that common features appear across so many levels of biological and ecological organisation thrust forwards the idea that there really are general principles shaping the function of these complex systems. The impact the system approach has in modern science can be summed up in Strogatz words: "*I hope I 've given you a sense of how thrilling it is to be a scientist right now. It feels like the dawn of a new era. After centuries of studying nature by teasing it into smaller and smaller pieces, we're starting to ask how to put the pieces back together again*". We trust that this approach will provide a path towards a better understanding in the search for principles governing living systems.

Introducción

Muchos de los fenómenos más interesantes que presentan los sistemas vivos son fruto de una interacción entre sus constituyentes. La memoria, por ejemplo, es un atributo del cerebro entero, y no se puede entender simplemente considerando las propiedades de las neuronas independientes (incluso aunque asumiésemos que conocemos perfectamente como funcionan, que no es el caso). El comportamiento de las células depende de toda una red metabólica, compuesta por muchas rutas metabólicas relacionadas entre si, y por tanto puede presentar respuestas totalmente diferentes en función de las condiciones externas. El movimiento coherente que presentan las grandes bandadas de pájaros o los bancos de peces no se entiende fácilmente si se estudian solamente unos cuantos de esos individuos de manera aislada. Todos estos ejemplos tienen algo en común: el comportamiento colectivo emergente no se puede explicar de una manera directa teniendo sólo en cuenta las características de los componentes individuales. Algo incluso más llamativo es que los detalles de las interacciones individuales no parecen jugar un papel muy relevante en ese comportamiento emergente. Así, independientemente de si estudiamos la sincronización de las luciérnagas, ritmos circadianos, actividad neuronal o ciclos económicos incluso, podemos no tener en cuenta los detalles de las interacciones que tienen lugar entre los insectos, células, neuronas o compañías y tratarlos a todos como osciladores de algún tipo, que presentarán unas fenomenologías similares. Todos estos son ejemplos de lo que se denominan como "sistemas complejos", y como tal, deberían entenderse desde un enfoque sistémico. Así visto, parecen candidatos ideales para ser estudiados dentro del marco de la mecánica estadística, que estudia las propiedades "macroscópicas" (o colectivas) de un sistema partiendo de la información "microscópica" (o individual) de sus componentes. Es de esperar que los métodos tradicionales y las herramientas de la mecánica estadística -junto otros métodos nuevos y

especialmente diseñados- constituya el mejor marco para estudiar, desde un punto de vista cuantitativo, las propiedades fascinantes de los sistemas vivos y las comunidades que los forman. En este sentido, la mecánica estadística ha pasado a transformarse en una disciplina, valga la redundancia, multidisciplinar.

Dentro de este enfoque sistémico uno de los conceptos que encontramos en todas las ramas y que ha demostrado ser de gran utilidad es el de las "redes complejas". En el caso de la biología molecular, hemos visto como el uso de estas herramientas ha transformado los estudios limitaos a unos cuantos genes partículares en la construcción de mapas genéticos globales que describen como los genes se regulan entre si [63]. Las redes son una forma muy útil de describir este tipo de sistemas, puesto que ponen de manifiesto las interacciones existentes entre los elementos del sistema, así como si se encuentra una estructura particular en ellas. Las redes complejas que aparecen en la naturaleza, de hecho, suelen tener unas características estructurales altamente no triviales, como por ejemplo la compartimentación, clustering, distribuciones de grado muy heterogéneas, etc. Entender cómo estas estructuras determinan el comportamiento colectivo y cuales son los fenómenos que llevan a su aparición son los objetivos clave en la teoría de sistemas complejos actual.

Un enfoque sistémico (u holístico) como este es interesante, al menos, por dos razones:

En primer lugar, redes que tengan una estructura similar aun siendo de sistemas muy diferentes pueden ser el resultado de la existencia de necesidades similares en ambos sistemas (como por ejemplo que sean estables o robustos) o de que ambas redes se formen de una manera similar.

En segundo, la estructura de la red tiene una gran influencia en cómo responde el sistema a las perturbaciones. La robustez de un ecosistema frente a la extinción de especies está determinada, en gran parte, por su patrón de conectividad, puesto que esto decidirá las extinciones secundarias que pueden producirse como resultado [6]. De manera parecida, en las redes genéticas se cree que la arquitectura es tal que las hace muy robustas frente a pequeñas mutaciones o al bloqueo de algunos genes o a pequeños cambios del ambiente. En general muchas dinámicas sobre redes complejas han mostrado tener un comportamiento diferente dependiendo de la distribución de grado[192]. Este enfoque de "redes" (o sistémico) puede utilizarse también en ecología teórica, donde los

ecosistemas están compuestos por un gran numero de especies interaccionando entre si de diferentes formas (predación, mutualismo, parasitismo, etc) pero con patrones específicos. Podemos considerar a los ecosistemas como sistemas complejos porque el comportamiento de estas comunidades estructuradas en su conjunto no puede ser inferido de manera directa de las propiedades de las especies que forman parte de él. Para entender realmente cómo se forman y funcionan estas comunidades es necesario considerar la estructura completa de la red que forman. Es el estudio del sistema en su conjunto, y no el de las interacciones individuales lo que nos dará información crucial sobre su funcionamiento a nivel global.

Todas estas ideas y conceptos son muy útiles a la idea de racionalizar los sistemas ecológicos [9, 10, 27, 167]. De hecho, el objetivo principal de esta tesis es analizar comunidades ecológicas complejas desde un enfoque sistémico, usando redes. Estudiaremos sus características arquitectónicas distintivas e intentaremos correlacionarlas con comportamientos generales tales como estabilidad, resiliencia, robustez y en general biodiversidad. Nuestros estudios se centran en redes ecológicas, pero algunos aspectos de redes biológicas, y en menos medida tecnológicas, se analizarán también.

La idea de una red compleja de interacciones entre especies es, de hecho, muy antigua, tanto como Darwin y su madeja enmarañada [62]. Lindermayer, Odum, Margalef y muchos otros han descrito las comunidades ecológicas como grafos en los que se transporta energía. Las interacciones de predación/alimentación en una comunidad pueden ser vistas también de esta forma. De hecho, las redes tróficas representan este tipo de interacciones entre especies, es decir, quién se come a quien. Son redes dirigidas donde las especies se representan mediante nodos, y las interacciones de predación mediante links, que codifican el transporte de masa desde la presa hasta el predador.

Aunque estas redes son muy difíciles de obtener y de modelar son también de una importancia crucial. Estas interacciones entre especies de diferentes niveles tróficos determinan su respuesta a perturbaciones en los ecosistemas, ya sean intencionadas o naturales. Entender la ecología y las matemáticas de estas redes, y en general, de las redes ecológicas es central para entender la respuesta de los ecosistemas a las perturbaciones, y podría ser un paso hacia una ecología más predictiva.

Durante varias décadas, hasta los 70, el paradigma dominante era que las comunidades complejas son mas estables que las simples [79, 141]. El argumento a favor de

que la complejidad es beneficiosa para la estabilidad de las comunidades ecológicas lo presentó de una manera general McArthur [141]: "*un gran numero de caminos entre especies es necesario para reducir los efectos de superpoblación en una especie*". El concluyó que "*la estabilidad aumenta como el numero de links aumenta*" y que la estabilidad se alcanza mas fácilmente en comunidades mas diversas, conectando, por lo tanto, la estabilidad tanto con un número grande de especies como con gran número de links entre ellas. Esto fue puesto en cuestión por May en un artículo donde, aplicando análisis de estabilidad lineal sobre matrices de comunidades ensambladas de forma aleatoria, demostró que la estabilidad de esas redes disminuía con la complejidad [150]. Estas comunidades sufrían una transición de estable a inestables que era mas acentuada cuento mas aumentaba la complejidad y el tamaño de los sistemas. Esta aparente contradicción entre la existencia de ecosistemas grandes y complejos y la inestabilidad asociada con esas dos características inició el debate de "diversidad vs estabilidad". Muchas preguntas relevantes surgen relacionadas con esto ¿De dónde viene esta estabilidad?¿Existe alguna estructura en las redes tróficas que pueda dar cuenta de ello? En conjunto, los avances ecológicos recientes indican que la diversidad, en media, aumenta la estabilidad de los ecosistemas [156], aunque este no es un tema cerrado. Descubrir cuales son las características topológicas implicadas en la gran estabilidad de los ecosistemas (si la hubiere) es un tema de una importancia central en ecología, y uno de los temas más importantes de esta tesis.

Una propiedad que también se ha relacionado con la estabilidad de los sistemas es la ausencia de loops (caminos cerrados de nodos) en las redes reales. Se sabe que los feedback loops tienen un impacto muy relevante en la estabilidad de redes tróficas[7, 21, 151, 164, 172, 173] , así como en las biológicas, o incluso en general [4, 12, 15, 35, 60, 124, 136, 143, 180, 198, 201, 216, 248]. Su ausencia ha sido frecuentemente asociada a la estabilidad que estos sistemas deberían tener [7, 35, 121, 140, 164, 172]. Aun así, ¿es esta la única explicación para la ausencia de loops? Este es otro de los puntos que trataremos a lo largo de la tesis.

Aparte de las relaciones de predación de la que hemos hablado ya, existen muchos otros tipos de interacciones en los ecosistemas. Una de gran relevancia es la interacción mutualista. Estas interacción son ubicuas en la naturaleza. De hecho, casi el 90% de las plantas tropicales depende de su interacción con animales para completar su ciclo

vital, ya sea por polinización o por dispersión de semillas [119]. Originalmente los primeros estudios de mutualismos se centraron en interacciones muy especiales. Los ejemplos de una coincidencia perfecta entre la morfología de una flor y la del animal que la poliniza eran aceptados como algo normal. Resulta que estos ejemplos de extrema especialización entre dos especies pueden ser más la excepción que la norma. Las interacciones mutualistas pueden involucrar decenas o incluso cientos de especies interaccionando entre si, con diferentes niveles de especialización [118, 259] . Las comunidades mutualistas, como las formadas por plantas y sus polinizadores o por semillas y sus dispersores, pueden describirse como una red de interacciones entre compañeros mutualistas (plantas y animales en este caso). Estas, sin embargo, muestran una diferencia fundamental con las redes tróficas . Mientras que en las redes tróficas todos los nodos son del mismo tipo y podrían, en un momento dado conectarse todos con todos, este no es el caso de las interacciones mutualistas. Estas comunidades son descritas por una red bipartita, que contiene interacciones entre dos grupos diferenciados: plantas y animales, pero no dentro de ellos. Las redes mutualistas de plantas y animales tienen unos rasgos muy característicos que aparecen repetidos a través de sistemas muy diferentes, independientemente del tipo de mutualismo, localización geográfica, o composición de especies [29, 120]. Muchas de estas características son propiedades emergentes que resultan de la interacción entre plantas y animales [264]. La inspección de estas redes mutualistas evidenció que estas redes no son ni aleatorias, ni forman compartimentos con mucha especialización recíproca [29]. Las redes mutualistas de plantas y animales son *anidadas*, lo que significa que las interacciones están dispuestas de tal manera que los especialistas (especies que tienen muy pocos compañeros mutualistas) interaccionan solo con generalistas (especies con muchos compañeros). Esta característica particular se ha asociado a una aumento de la estabilidad y coexistencia de especie en estas comunidades [33, 181], lo que sugiere que el anidamiento ha podido ser, de alguna manera, seleccionado por las fuerzas de la naturaleza. Esto es disputado en un trabajo reciente [234] donde se encuentra justo lo contrario:las redes anidadas tienden a promover la inestabilidad. Otras evidencias indican que los procesos neutrales tienen también una gran importancia en la arquitectura de estas comunidades [129, 263], lo que significaría que es la distribución de abundancia de las especies y las interacciones aleatorias entre individuos lo que determina las propiedades de estas redes. Muchas

preguntas surgen también alrededor de este tema ¿Cuáles son los factores que determinan el anidamiento de estas redes tan importantes?¿es necesario recurrir a procesos de optimización para obtener redes anidadas [242] o pueden las teorías neutrales explicar el anidamiento presente en muchas de estas redes [129, 263]?¿Como afecta a todo esto el patrón de conectividad?. Si quisiésemos diseñar una política de protección de especies ¿cuales deberían protegerse primero? ¿las más conectadas? ¿las más centrales? ¿las mejor conectadas?. Esta son algunas de los problemas que trataremos en esta tesis relacionados con redes mutualistas.

A lo largo de este trabajo usamos el enfoque sistémico de la mecánica estadística de redes complejas para responder algunas preguntas relevantes sobre cómo la arquitectura de las redes ecológicas y biológicas afecta a su funcionamiento. ¿Existe alguna propiedad responsable de la estabilidad? ¿Es la estabilidad la única explicación para la estructura con pocos loops presente en redes ecológicas y biológicas o existen otros factores que podrían explicarlo? ¿Cuáles son los principales factores que determinan el anidamiento de una comunidad mutualista? ¿Es posible identificar la importancia de las especies en las comunidades mutualistas solamente con la información del patrón de conectividad? Estas son las principales cuestiones que consideraremos a lo largo de esta tesis.

Resumen

Nos reservamos el **capitulo dos** para introducir de forma breve las características mas relevantes de las redes ecológicas y biológicas que han sido expuestas gracias al análisis estadístico de los datos. ¿Cuáles son sus características principales? ¿Comparten otras redes las mismas propiedades? ¿Qué papel juegan en el funcionamiento de la red? Esto es lo que cubriremos en el segundo capitulo.

En el **capitulo tres** presentamos un análisis del debate complejidad vs estabilidad. ¿Por qué son tan estables los ecosistemas a pesar de ser grandes y complelos? Aquí mostramos que la coherencia trófica -una característica de estas redes que no había sido tenida en cuenta hasta ahora- es un mejor predictor estadístico de la estabilidad lineal que el tamaño o la complejidad. Demostramos también que una red totalmente coherente con interacciones constantes será siempre estable. Proponemos un modelo muy simple que, capturando correctamente la coherencia de las redes tróficas reproduce su estabilidad y otras propiedades básicas. El modelo sugiere que la estabilidad puede

aumentar con la complejidad y el tamaño en algunas circunstancias. Esto podría ser algo a considerar en la paradoja de May, y presenta una gama de oportunidades para la conservación de la biodiversidad.

En el **capitulo cuatro** exploramos la hipótesis de que la presencia de loops en muchas redes complejas se reduce mucho si existe una dirección global inherente. Para cuantificar esta idea proponemos un modelo muy simple para generar redes con direccionalidad, usando un solo parámetro. En este modelo un aumento de la direccionalidad genera una disminución del numero de loops. Para probar esto cuantificamos los loops existentes en mucas redes empíricas de diferente tipo. Ejn general las redes muestran una ausencia de loops cuando se comparan con las randomizaciones, y el modelo es capaz de predecir esta ausencia con bastante precisión ajustando el parámetro de direccionalidad. Esta medida de la direccionalidad se correlaciona con una medida directa. Concluimos que la existencia de una direccionalidad inherente es capaz de explicar la ausencia de loops de una manera sencilla pero efectiva.

En el **capitulo cinco** pasamos a trabajar con redes mutualistas. A lo largo de este capitulo introducimos una nueva medida del anidamiento y estudiamos como influyen en ella las propiedades estructurales mas básicas (distribución de grado y correlaciones). Encontramos que la mayoría del anidamiento que encontramos en las redes empíricas procede de la gran heterogeneidad de su distribución de grado. Una vez descontamos este efecto, encontramos que la nestedness esta relacionada con la disasortatividad, y puesto que las redes finitas muestran un nivel residual de anticorrelación, también tienden a estar anidadas de forma natural.

En el **capitulo seis** desarrollamos una técnica para medir la importancia de las diferentes especies en una red mutualista. El algoritmo, con un espíritu similar al del Google-Rank, pero con un planteamiento no-lineal, nos permite obtener rankings robustos en redes mutualistas. Este método supera a todos los demás que hemos usado como marco comparativo, y puede ser útil para conservación de la biodiversidad o administración de ecosistemas, donde haya que tomar decisiones explicitas sobre que especies proteger.

Las redes se han convertido en un marco fundamental para el estudio de los sistemas complejos en diversos campos. Aquí tratamos con sus aplicaciones a sistemas biológicos y ecológicos. El hecho de que patrones comunes aparezcan a través de campos muy

diversos y a través de muchos niveles de organización da alas a la idea de que existen realmente unos principios generales que definen la forma de funcionar de estos sistemas complejos. El impacto que el enfoque sistémico ha tenido en la ciencia moderna puede ser resumido en estas palabras de Strogatz :"*Espero haberos transmitido lo emocionante que es ser científico ahora mismo. Parece que estamos en el amanecer de una nueva era. Después de siglos en los que hemos dividido la naturaleza en piezas pequeñas para estudiarla, estamos empezando a preguntarnos cómo montar las piezas de nuevo*". Confiamos en que este enfoque puede proveer un camino hacia un mejor entendimiento en la búsqueda de los principios que gobiernan los sistemas vivos.

Chapter 2

Topological features of ecological and biological networks

2.1 Unimodal networks

The last 15 years have witnessed an explosion in the research field of complex networks [5, 14, 24, 176, 178, 192, 217, 239]. Several systems ranging from genetic networks, metabolic pathways and ecosystems to societies or the WWW could be described with a common framework, in which elements (genes, chemical compounds, ecological species, persons or web pages) are nodes connected by links. These links, that encode the interplay between them, can take the form of gene activation/deactivation, reactions, species interactions, relationships or hiperlinks). A major consequence of this multidisciplinary approach was the recognition that several networks, notwithstanding differences in the nature of their nodes, exhibit similar statistical properties. Despite the fact that all these empirical networks are dynamical in their nature and thus may change its structure over time (i.e predators can switch prey if a particular species is scarce, some metabolic reactions only take place when certain compounds are present, etc), the description in terms of static networks has proven valuable to identify structural features responsible for emerging functions [5, 16, 25, 29, 178]. Some structural features, including clustering, degree assortativity[177], and the relative abundance of specific motifs [12, 13], characterize the topology (i.e the structure) at the local scale. Other traits, such as nestedness [29], community structure [87], and the existence of

a hierarchy [56, 207] are related to the large-scale organization. From here on we are going to focus on some characteristic patterns of biological and ecological networks, and how these structural features affects their functioning.

Inspection of a wide collection of empirical networks evidenced that the *connectivity pattern* they exhibit is far from what could be naively, and randomly, expected. For instance, the degree distribution in both metabolic and protein networks is highly heterogeneous. It displays a scale-free connectivity pattern [110, 112], characterised by the presence of a bulk of nodes with few connections and few nodes with many connections. These kind of highly heterogeneous degree-distributions make networks highly tolerant to random attacks/errors [6] and may be behind the alleged robustness of these systems to small mutations or genes knock-outs. In addition to the number of connections of each node, it is also possible to examine how is it connected to the other nodes in the network. In this sense, we can speak of *degree-degree correlation* as the tendency of the nodes to be connected to other nodes as a function of their neighbour's degree. Several biological and ecological networks exhibit a dissasortative pattern [177] in which nodes with many connections are preferentially linked to nodes with few connections (social network, on the contrary, tend to have nodes with similar number of connections link together, i.e, they are assortative).

Akin to this idea but with different implications is the *rich-club phenomenon*, which refers to the tendency of nodes with high centrality (the dominant elements of the system) to form tightly interconnected communities. It is one of the crucial properties accounting for the formation of dominant communities in both computer and social sciences [54]. Interestingly enough this kind of organisation is also present in biological systems. In some cases such as the human and macaque brains these "rich-clubs" are identified as regions that play a key role in global information integration between different parts of the neural network [98, 258]. In others, such as protein-protein networks, the presence of the phenomenon may indicate that key proteins act in concert, suggesting a certain degree of stability in the activities for which they are responsible [155]. This relevant effect seems, however, to be absent in foodwebs, maybe because in these networks there is not any structure scting as a "control-centre". On the other hand, inspection these type of networks evidenced a high connectance, with species interacting with many others, forming rather dense webs. Study of he *"K-subwebs"* (or *K-cores*)

of foodwebs, defined as a subset of nodes that are connected to at least k nodes from the same subset, revealed a highly cohesive organisation, with a high number of small subwebs highly connected among themselves through the most dense subweb [160].

All these properties We have just seen that many different ecological and biological systems can be visualized as networks made up of units linked whenever there is some sort of interaction or "flow" between them. When these interactions have a determined direction, i.e links have an origin and a target node, they constitute the so called "directed networks". Genetic regulatory circuits, metabolic pathways, neural networks and foodwebs are examples of this type of webs. In general, the direction defined by the interaction in these systems can be thought of in terms of flows, such as the energy transfer in food webs or the flow of biological information in genetic or neural networks. Directed networks also show features that appear repeatedly across different systems. Following we cover some important structural characteristics of both ecological and biological directed networks that will be of use along this study, paying special attention to the structural features of ecological systems, such as foodwebs and plant-animal mutualistic networks.

2.1.1 Community structure

Much attention has been paid in recent years to the *community structure* of complex networks: how the nodes can be classified in groups such that a high proportion of links fall within groups and few connections occur between groups [87]. One can, thus, identify clusters, or communities, of densely linked nodes in the network structure, connected to each other.

In biological networks this particular feature is frequently associated with task differentiation: different functions being accomplished by different modules. For example, the brain is known to exhibit a hierarchical-modular structure [168], where each module specialises in carrying out a determined task [67]. Modules in genetic regulatory networks and metabolic networks are associated with diverse kinds of subcircuits, each performing a specific kind of function [64, 101]. Some subcircuits are used in many diverse biological contexts, others are more complex and are dedicated to similar biological functions wherever they appear.

The idea of compartmentalisation in foodwebs has been present almost since the

beginning. At the end of his seminal paper, May [150] noticed that model communities tended to be more stable if organized in blocks, suggesting both that compartmentalisation could enhance stability and that real ecological communities may be modular. The high number of connections in foodwebs, compared with other types of networks, jeopardices the search of such groups. However, application of powerful methods for community structure detection to foodwebs resulted in revealing a compartmentalised structure related to species habitat [127] or other features such as body size or phylogeny [209]. The dynamical consequence of this kind of organization seems to be an enhance of the network robustness against the propagation of perturbations. Indeed, the existence of these compartments favours species extinctions to be confined within a module rather than to spread over the whole community[238].

Regardless of the nature of the network and the reason behind the community structure it is possible to quantify the degree of compartmentalisation. For a system with S nodes and mean degree $K = L/S$, the *configuration model* [165] holds that the probability of there being a link from j to i is $k_i^{in}k_j^{out}/(KS)$ (where k_i^{in} and k_i^{out} are the numbers of i's prey and predators, respectively) [178]. Using this, and given a particular partition (i.e., a classification of nodes into groups) of the network, one can define the "modularity" of that particular partition of the network as:

$$Q = \frac{1}{KS}\sum_{ij}\left(a_{ij} - \frac{k_i^{in}k_j^{out}}{KS}\right)\delta(\mu_i, \mu_j),$$

where μ_i is a label corresponding to the partition that node i finds itself in, and $\delta(x, y)$ is the Kronecker delta [178]. The modularity of the network is taken to be the maximum value of Q obtainable with any partition. Since searching exhaustively is prohibitive for all but very small and sparse networks, a stochastic optimization method is usually called for. Throughout this work we use the algorithm of Arenas *et al.* [17], although there are many in the literature and the most appropriate can depend on the kind of network at hand [61, 68].

2.1.2 Hierarchy

Hierarchy is a polysemous word, involving order, levels, inclusion, or control as possible descriptors [56]. However all of them stem from the idea of some nodes being "domi-

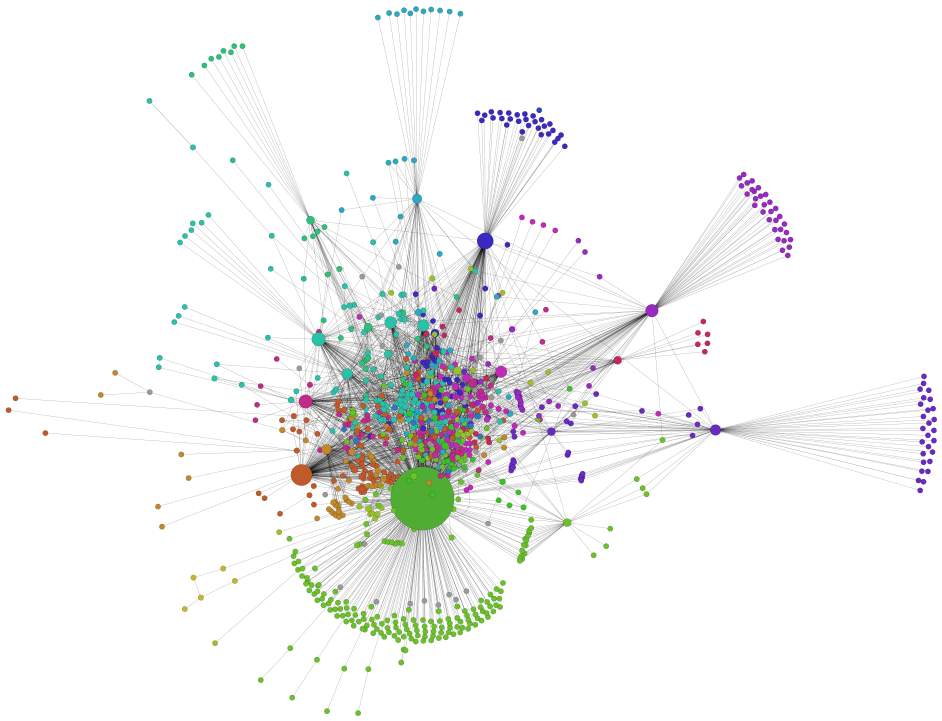

Figure 2.1: Transcriptional regulatory network of the E. Coli [220], with different modules in different colours. The modules have been defined by application of community detection software (Gephi [31]) and evidence the existence of different groups interlinked to compose the whole regulatory network.

nant" over others. Different kind of biological and ecological directed networks present such type of organization, generally associated with control or transport of some class. For example in neuronal networks the flow of information propagates from sensory neurons at the bottom of the hierarchy of control, to neurons in the central system, and from there to the level of motor neurons[**?**]. Hierarchical structure appears associated to control also in genetic and transcriptional regulatory networks, were controller nodes act upon controlled ones [80, 131, 213, 244, 276]. In networks were there is a transfer of matter, as in foodwebs or metabolic networks, one can identify a hierarchy of "trophic" levels were links tend to point from lower levels to higher ones. In this way, a hierarchy comes up, in which the *trophic level* of a species indicates how high up the "food pyramid" it is (see Fig. 2.2).

To extract this hierarchical ordering it is often useful to assign to each species a number representing its "trophic" level as follows: the *basal species*, those with no prey (no incoming connectivity), are assigned level one. The ones that only prey on basal species are set to level two, and, in general, a given species takes the average level of its prey plus one [135]

$$l_i = \frac{1}{k_i^{in}} \sum_j a_{ij} l_j + 1 \tag{2.1}$$

where $k_i^{in} = \sum_j a_{ij}$ is the number of prey of species i (or i's *in-degree*), and a_{ij} are elements of the predation (or adjacency) matrix A. Along this study we will only deal with unweighted networks in which the elements of A are either 1's or 0's, and hence we omit the link strength term usually included in Eq. (2.1) [135] [1]. Therefore, the trophic level of each species can take non-integer values and is a purely structural property that can be determined by solving a system of linear equations.

Although we have illustrated this algorithm to extract the hierarchical ordering with foodwebs, it can be applyed to any directed network. As a matter of fact similar methods have been proposed to infer the hierarchical ordering in biological networks [131, 213, 276]. These methods are able to extract a hierarchical organization from a given network and classify the nodes into a few discrete levels. In contrast, the method we propose here produces more refined orderings since it is able to resolve possible degeneracies between the coarser levels produced by previous methods.

2.1.3 Coherence

Once species are asigned a trophic level (section. 2.1.2), it is possible to define the "trophic distance" spanned by each link in a foodweb ($a_{ij} = 1$) as $x_{ij} = l_i - l_j$ (which is not a distance in the mathematical sense since it can take negative values). The distribution of trophic distances over the network, $p(x)$, will have mean $\langle x \rangle = 1$ (since for any node i the average over its incoming links is $\sum_j a_{ij}(l_i - l_j)/k_i^{in} = 1$ by definition). We define the "trophic coherence" of the network as the homogeneity of $p(x)$: the more similar the trophic distances of all the links, the more coherent. To say it

[1] However it is customary, when working with weighted networks, to weight each prey species contribution with the strenght of its link

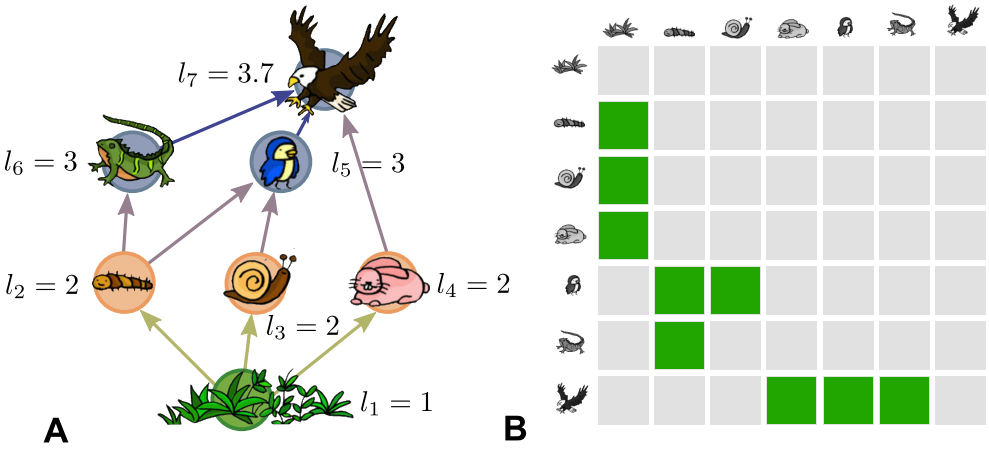

Figure 2.2: Example of directed ecological network encoding feeding relations in a simple community. Left panel (A) shows the network schematic representation were the nodes represent trophic species and the links the feeding interactions. The links are directed from the prey to the predator, reflecting the actual flow of biomass. Right panel (B) presents the interaction (or adjacency) matrix where the green squares represent a existing interaction ($a_{ij} = 1$) meaning species i consumes species j, and the grey squares stand for non-existent interactions. In this network a hierarchy appears related with the transport of biomass, with some species being dominant over others. The trophic level of the species (l_i) gives an idea of how far is from the source of biomass.

simply, coherent networks have a well defined "stratified" structure, with highly defined "trophic layers" while incoherent ones have not (see Fig. 2.3).

As a quantification of coherence, we use the standard deviation of the distribution of trophic distances, which we will refer to, from now on, as an incoherence parameter:

$$q = \sqrt{\langle x^2 \rangle - 1} \tag{2.2}$$

where $\langle \cdot \rangle = L^{-1} \sum_{ij} (\cdot) a_{ij}$, and L is the total number of links, $L = \sum_{ij} a_{ij}$.

Trophic coherence bears a close resemblance to Levine's measures of "trophic specialization" [135]. However, our average is computed over links instead of species, with the consequence that we need not consider the distinction between resource and consumer specializations. It is also related to measures of omnivory: in general, the more omnivores one finds in a community, the less coherent the foodweb.

We have introduced this concept in the context of foodwebs, however many kinds of directed networks transport energy, matter, information or other entities in a similar way to how food webs carry biomass from producers to apex predators. The adaptation of this measure to any type of directed network is straightforward.

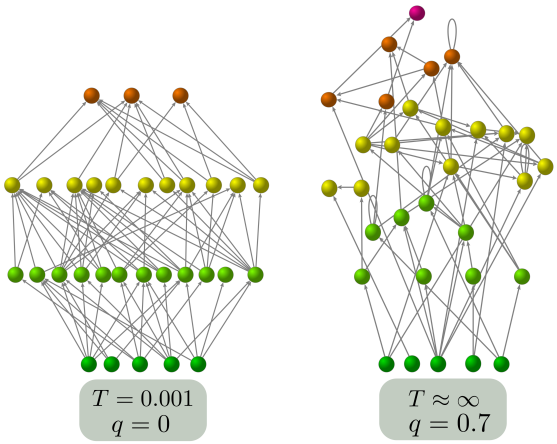

Figure 2.3: Two networks illustrating the extremes of trophic coherence: the network on the left is completly coherent, $q = 0$, (all links are between species that are only one trophic level apart) and hence have perfectly defined layers. On the other hand, the network one on the right shown a highly incoherent structure, $q = 0.7$, without a clear stratification.

2.1.4 Feedback loops and motifs

In a directed network a "feedback loop" of length k is defined as a closed sequence of k different nodes in which a walker following the directions of the arrows returns to the starting point after visiting once, and only once, all k nodes (fig. 2.4). Although not many studies have dealt with loops statistics, inspection of empirical biological networks evidence a under-representation of long feedback loops in the E. coli gene regulatory network and the over-representation of short feedback loops in the S. cerevisae's [245]. Ref.[85] studies the statistics of the total number of feedback loops in different complex networks, and reports a general absence of directed loops in different foodwebs. We also study loops statistics in different directed complex networks and obtain similar results (see chapter 4 and ref.[P2]): biological and ecological networks

show an under-representation of long (i.e involving more than two nodes) loops. These kind of network patterns are well-known to have a profound impact on dynamical stability. For instance, feedback loops are reported to be ubiquitous basic elements of biological systems such as transcriptional regulatory networks, signaling transduction pathways and cell cycle regulatory networks, where they are responsible for bistability (i.e. it may creates discontinuous output response from continuous input) and oscilations, [124, 143, 180, 198, 248]. However, as oberved in [140], biological networks display a kind of antiferromagnetic ordering – were contiguous links have a statistical tendency to point in opposite directions– causing a depletion of feedback loops which, the authors claim, lead to an enhancement of network stability.

In foodwebs the impact of loops in stability is also wide acknowledged [7, 21, 151, 156, 164, 172, 173, 241], and studies mainly focus on the stabilising effects that weak interactions taking place in long trophic loops confers to the community.

Overall, motif identification (i.e Patterns that occur in the real network significantly more often than in randomized networks) in both ecological and biological networks exhibit a characteristic pattern [163], that in biological networks represents the existence of repeated circuits with similar functions [223, 273] and in foodwebs different ecological relations such as apparent competition or intraguild predation [28], or different species roles in the community [237].

2.1.5 Intervality

When Cohen first compiled a database of empirical foodwebs and studied them quantitatively [51], in addition to observing that certain ratios (known as scaling laws) between kinds of species are to some degree universal, he discovered that many of the networks were what he called *interval*: the columns of the adjacency matrix could be ordered in such a way that all the prey of any predator were contiguous [52](see fig. 2.5). In other words, if all the species were placed appropriately on an axis – now labelled as the niche dimension – then the set of prey of any given predator would lie on an unbroken interval. More recently the degree of intervality of foodwebs has been measured with several estimatorss that are close to unity if most of the prey lie on unbroken intervals, or approach zero when very few do [236]. This streaking feature of foodwebs has been related to the number of trophic dimensions characterizing the niches in a community,

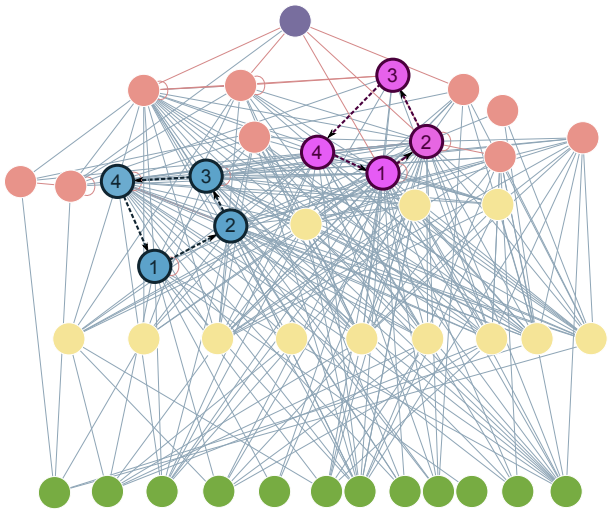

Figure 2.4: Empirical trophic web of the Mondego Stuary [194], showing the only two loops of length 4 present in the network. Species trophic level is represented by the species' height, and the color is just a help to the eye.

and it is the core idea behind many customarily employed foodweb models such as the Cascade [53] or the Niche model[270].

A simple way to quantify it is to measure, for each predator, the size of its largest unbroken interval as a fraction of its total number of prey, and use the average value over predators as a measure of the intervality – also sometimes called contiguity – of the web. This is the magnitude we shall use in this work, represented as ξ.

To illustrate the concept of intervality the top left panel of fig.2.5 shows the adjacency matrix for the foodweb of Skipwith pond (U.K), with the species ordered so as to maximize ξ. Since there are very many possible orderings of the columns (the factorial of the number of species) it is not possible in general to exhaustively search for the best – i.e., the most interval – one. Instead we do as Stouffer et al. [236] and use a Simulated Annealing (SA) algorithm. For comparison, in the right panel of Fig.2.5 we show the best ordering (also obtained with the SA method) for a random graph generated with the same numbers of species, S, of basal species (producers), B, and of links, L, as the empirical web. Within these constraints, the links are placed at random, as in an Erdos Renyi graph [178]. The most immediate observation is that the empirical foodweb is indeed significantly more interval than the random graph, indicating (as has often been

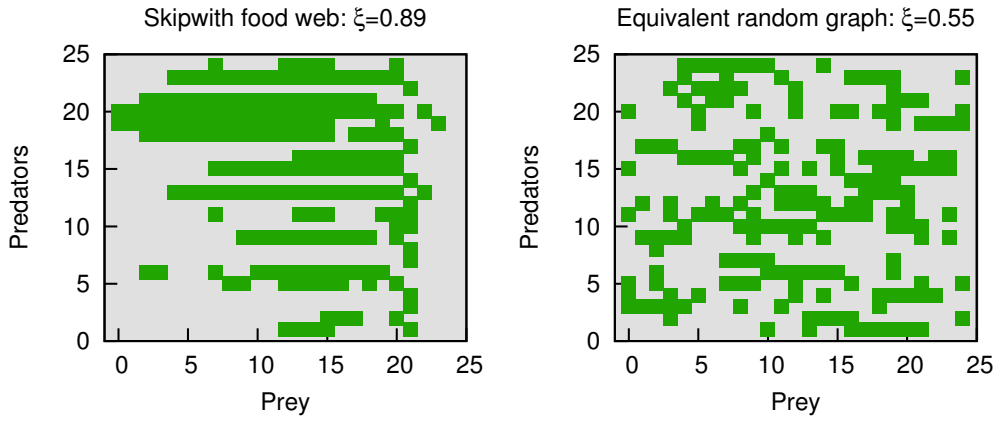

Figure 2.5: Left panel: Adjacency matrix corresponding to the foodweb of Skipwith pond [267]. Green square means that the predator (vertical axis) eats the prey species (horizontal axis) in question. An ordering of the species has been sought which maximizes prey-intervality, ξ: the prey of any predator tend to be contiguous. Right panel: The same for a random null-model corresponding to the Skipwith pond web, in which links have been allocated randomly.

stated) that intervality is a highly non-trivial feature of natural foodwebs.

2.2 Bipartite networks

Whenever a biological or ecological system presents two different type of elements interacting between each other they are susceptible to be described in terms of bipartite networks. For instance, transcriptional regulatory bipartite networks depict the interaction between two different set of nodes: genes and the proteins they encode. The links, that represent which genes codify each of the proteins, will link only elements in one set with elements on the other. Metabolic reactions can also be described in terms of bipartite networks, were reactions and metabolites are linked whenever the metabolite takes part in a reaction.

There are different kinds of ecological bipartite networks, encoding different type of interactions. Examples of them are host-parasitic networks, which represent interactions between two species in which one of them takes advantage on the other, facilitation networks, which encode the interactions in which an organism profits from the pres-

ence of another (as happens with nurse plants that provide shade for new seedlings or saplings) or mutualistic networks, representing mutually beneficial interactions. Here we will only focus on this last kind.

Mutualistic communities, thus, can be described as a bipartite network of interactions between two mutualistic partners, in our case plants and animals. The links represent mutualistic interactions such as pollination of plants by different animals (bees, butterflies, hummingbirds...) or dispersion of seeds by animals which feed on the plants producing them (birds, monkeys...). Fig.2.6 represents an example of such a network. Alternatively, mutualistic networks can also be described as a matrix, with animals as rows and plants as columns. Each element of such interaction matrix is 1 if that particular plant and animal interact, or 0 otherwise [2]

As in the case of directed networks, inspection of diverse mutualistic communities revealed the existence of common features. In fact, mutualistic networks have well-defined structures and several patterns have been found in a vast array of datasets encompassing highly diverse ecosystems and varying degrees of complexity.

2.2.1 Degree distribution

One important feature of complex networks is that their degree distribution is heavily heterogeneous, characterized by having many nodes with few links and very few nodes with many links [230]. Regardless of the differences in high, latitude, or species composition, mutualistic networks display a common and well-defined connectivity pattern [120]. These networks are much more heterogeneous than expected by chance, although not as heterogeneous as scale-free networks, they show an exponential cut-off. This implies that as the number of interactions reaches a critical connectivity value k_c, the probability of finding species with more connections drops faster than expected for a power-law. Mutualistic communities, as other networks with these type of connectivity patterns, exhibit a high tolerance to random extinctions [42].

[2]These element can also be a positive number describing the strength of the interaction (i.e., the relative frequency of visits, or relative frequency of pollen or seeds dispersed). In this case this would be a weighted bipartite network.

2.2.2 Community structure

As we have seen before, many biological and ecological systems are organized into modules, where different subsets of units have a specific functionality. In mutualistic networks modularity can emerge from spatio-temporal structure (as habitat separation) and/or evolutionary processes leading to non-random patterns of interactions. Modules in mutualistic networks have been suggested to be units of coevolution, and some phylogenetic studies support this idea [210]. Plant-animal mutualistic networks exhibit a modular estucture with generalist species connecting peripheral species together into modules, but also connecting modules between them [183]. These generalist species act as modular hubs, spanning connections among modules, and so they are crucial to maintain the cohesiveness of the network. The extinction of a module hub can lead to fragmentation of the network, that is why they are considered very important for the conservation of mutualistic communities.

2.2.3 Nestedness

Nestedness is a concept borrowed from island biogeography to illustrate how a pool of animals is redistributed among a set of islands [19]. In the mutualistic network literature, nestedness describes a defined pattern of species interactions. Inspection of plant-frugivore and plant-pollinator networks evidenced that these networks are neither randomly assembled nor organized in compartments arising from tight, reciprocal specialization. Plant-animal mutualistic networks are highly *nested*, that is, specialists (those with only a few mutualistic partners) interact with species that form well-defined subsets of the species with which generalists (those with many mutualistic partners) interact [29]. In other words, if we rank plants from the most specialized to the least specialized, we find that the set of animals a plant interacts with are contained in a larger set, which in turn is contained in a larger set, and so on, as in nested Chinese boxes, see fig. 2.6.

Nested networks have two important features:

- First, most part of generalist species interacts among them, which means that these networks are very cohesive. In other words: generalist species generate a dense core of interactions to which the rest of the community is attached.

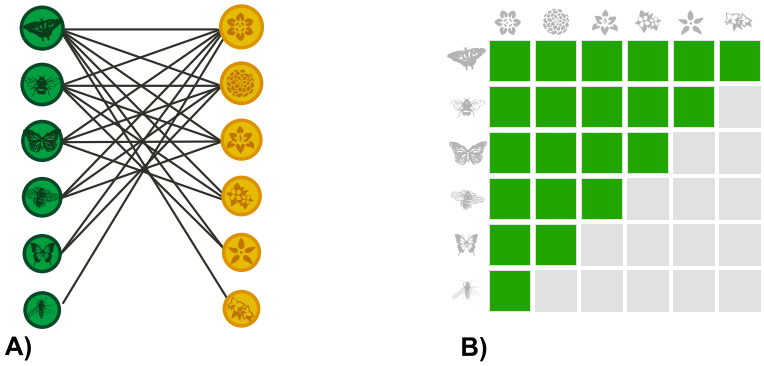

Figure 2.6: Example of mutualistic community composed by plants and its pollinators. It can be represented by a network of interactions (A) or by a plant-animal interaction matrix (B). Left panel shows an explicit representation of the network, where green nodes represent animal species and yellow nodes represent plant species. The lines connecting them are the observed interactions in the community. Right panel shows the interaction matrix, with animal species in rows and plant species in columns. Green squares depict observed interactions between a plant and an animal while light gray squares are non-observed interactions. This community is perfectly nested because specialist species form perfect subsets of more generalized species interacting with their mutualistic partners. The box outlined in B represents the central core of the network.

- Second, specialists tend to interact with the most generalist species, wich creates a high asymmetry in terms of specialization levels.

The role this feature plays in mutualistic networks is still open to debate. Several studies suggest that nestedness may increase the stability and foster coexistence in mutualistic communities [33, 181]. The general idea behind this connection is that since generalists tend to be more abundant and less-fluctuating species, when compared with specialists (because generalists rely on so many other partners), this asymmetrical structure provides pathways for the persistence of rare species [118]. Also, the cohesive pattern can provide alternative routes for system responses to perturbation, since a species is more unlikely to become isolated after the elimination of other species when embedded on a highly cohesive network [29].

This positive effect is, however, challenged in a recent study were the opposite is found: nested networks tend to promote instability. This work also suggest that in order to correctly evaluate the effect of the nested structure one should consider the

weight of the interactions and not only their presence [234].

The network properties mentioned above share a common feature: they are often asymmetrical. Asymmetry is present at all levels in mutualistic networks, from species number of interactions (degree-heterogeneity), to their pattern (nestedness) and strengths. In fact, the interaction strength (or species dependence) is highly heterogeneous and asymmetrical, which means that some plant species are highly dependent on the service (e.g. dispersal) provided by an animal species, but this animal species might depend much less on the resources (e.g. fruits) provided by the plant. Theoretical studies in foodwebs suggest that this distribution of interaction strengths may promote community persistence and stability [37]. In mutualisitc networks the reasoning goes as follows: if both plant and animal depend strongly on each other, a decrease in plant abundance will be followed by a similar decrease in the animal abundance, which in turn will turn back on its partner. If their relation is assymetrical this kind of negative feedback is less pronounced.

Chapter 3

Food webs coherence determines stability

In the early seventies, Robert May addressed the question of whether a generic system of coupled dynamical elements randomly connected to each other would be stable. He found that the larger and more interconnected the system, the more difficult it would be to stabilise [150, 152]. His deduction followed from the behaviour of the leading eigenvalue of the interaction matrix, which, in a randomly wired system, grows with the square root of the mean number of links per element.

This result clashed with the received wisdom in ecology – that large, complex ecosystems were particularly stable – and initiated the "diversity-stability debate" [73, 141, 156, 190]. Indeed, Charles Elton had expressed the prevailing view in 1958: "the balance of relatively simple communities of plants and animals is more easily upset than that of richer ones; that is, more subject to destructive oscillations in populations, especially of animals, and more vulnerable to invasions" [79]. Even if this description were not accurate, the mere existence of rainforests and barrier reefs seems incongruous with a general mathematical principle that "complexity begets instability", and has become known as May's Paradox.

One solution might be that the linear stability analysis used by May and many subsequent studies does not capture essential characteristics of ecosystem dynamics, and much work has gone into exploring how more accurate dynamical descriptions might enhance stability [41, 65, 156]. But as ever better ecological data is gathered,

it is becoming apparent that the leading eigenvalues of matrices related to foodwebs do not exhibit the expected dependence on size or link density [109]. Food webs must, therefore, have some unknown structural feature which accounts for this deviation from randomness – irrespectively of other stabilising factors.

We show here that a network feature we call *trophic coherence* accounts for much of the variance in linear stability observed in a dataset of 46 food webs, and we prove that a perfectly coherent network with constant link strengths will always be stable. Furthermore, the simple model that we proposed to capture this property suggests that networks can become more stable with size and complexity if they are sufficiently coherent.

3.1 Trophic coherence and stability

Each species in an ecosystem is generally influenced by others, via processes such as predation, parasitism, mutualism or competition for various resources [78, 200, 228, 241]. A food web is a network of species which represents the first kind of influence with directed links (arrows) from each prey node to its predators [70, 74, 199, 214]. Such representations can therefore be seen as transport networks, where biomass originates in the basal species (the sources) and flows through the ecosystem, some of it reaching the apex predators (the sinks).

The trophic level of a species (l_i) can be defined as the average trophic level of its prey, plus one [135, 195]. Thus, plants and other basal species are assigned level one, pure herbivores have level two, but many species will have fractional values [1] (see sec. 2.1.2). A species' trophic level provides a useful measure of how far it is from the sources of biomass in its ecosystem. We can characterise each link in a network with a *trophic distance*, defined as the difference between the trophic levels of the predator and prey species involved as $x_{ij} = l_i - l_j$ (it is not a true "distance" in the mathematical sense, since it can be negative). We then look at the distribution of trophic distances over all links in a given network. The mean of this distribution will always be equal to one,

[1] In computing the mean trophic level, it is customary to weight the contribution of each prey species by the fraction of the predator's diet that it makes up. Since we are here only considering binary networks, we do not perform this weighting. We also use the words predator and prey as synonyms of consumer and resource, respectively, even in referring, say, to plants and herbivores.

while we refer to its degree of homogeneity as the network's *trophic coherence*. We shall measure this degree of order with the standard deviation of the distribution of trophic distances, q (we avoid using the symbol σ since it is often assigned to the standard deviation in link strengths). A perfectly coherent network, in which all distances are equal to one (implying that each species occupies an integer trophic level), has $q = 0$, while less coherent networks have $q > 0$. We therefore refer to this q as an "incoherence parameter". (For a technical description of these measures, see secc.2.1.3).

A fundamental property of ecosystems is their ability to endure over time [214, 228]. "Stability" is often used as a generic term for any measure of this characteristic, including for concepts such as robustness and resilience [92]. When the analysis regards the possibility that a small perturbation in population densities could amplify into runaway fluctuations, stability is usually understood in the sense of Lyapunov stability – which in practice tends to mean linear stability [103]. This is the sense we shall be interested in here, and henceforth "stability" will mean "linear stability". Given the equations for the dynamics of the system, a fixed (or equilibrium) point will be linearly stable if all the eigenvalues of the Jacobian matrix evaluated at this point have negative real part. Even without precise knowledge of the dynamics, one can still apply this reasoning to learn about the stability of a system just from the network structure of interactions between elements (in this case, species whose trophic interactions are described by a food web) [9, 103, 152, 153]. Indeed, in Appendix A we describe how an interaction matrix W can be derived from the adjacency (or predation) matrix, A, representing a food web, such that the real part of W's leading eigenvalue, $R = Re(\lambda_1)$, is a measure of the degree of self-regulation each species would require in order for the system to be linearly stable. In other words, the larger R, the more unstable the food web. For the simple yet ecologically unrealistic case in which the extent to which a predator consumes a prey species is proportional to the sum of their (biomass) densities, the Jacobian coincides with W and R describes the stability for any configuration of densities (global stability). For more realistic dynamics – such as Lotka-Volterra, type II or type III – the Jacobian must be evaluated at a given point, but we show that the general form can still be related to W (see Appendix A). Furthermore, by making assumptions about the biomass distribution, it is possible to check our results for such dynamics (see subsection 3.1.1 below). However, in gneral we shall focus simply on

the matrix W without making any further assumptions about dynamics or biomass distributions, and hence estimate the community stability with R (leading eigenvalue of W).

May considered a generic Jacobian in which link strengths were drawn from a random distribution, representing all kinds of ecological interactions [150, 152]. Because, in this setting, the expected value of the real part of the leading eigenvalue (R) should grow with $\sqrt{SC}$, where S is the number of species and C the probability that a pair of them be connected, larger and more interconnected ecosystems should be less stable than small, sparse ones [229]. (Allesina and Tang have recently obtained stability criteria for random networks with specific kinds of interactions: although predator-prey relationships are more conducive to stability than competition or mutualism, even a network consisting only of predator-prey interactions should become more unstable with increasing size and link density [9]).

Throughout this chapter we analyse the stability of a set of 46 empirical food webs from several kinds of ecosystem (the details and references for these can be found in Section 3.6 at the end of this chapter). We start by scrutinizing how do the complexity and size of ecosystems relate to their stability. In Fig. 3.1.A we plot the R of each web against $\sqrt{S}$, observing no significant correlation. Figure 3.1.B shows R against $\sqrt{K}$, where $K = SC$ is a network's *mean degree* (often referred to as "complexity"). In contrast to a recent study by Jacquet *et al.* [109], who in their set of food webs found no significant complexity-stability relationship, we observe a positive correlation between R and $\sqrt{K}$. However, less than half the variance in stability can be accounted for in this way. In Section 3.1.1 below we also compare the empirical R values to the estimate derived by Allesina and Tang for random networks in which all links are predator-prey. Surprisingly, the correlation is lower than for $\sqrt{K}$ ($r^2 = 0.230$). The conclusion of Jacquet and colleagues – namely, that food webs must have some non-trivial structural feature which explains their departure from predictions for random graphs – therefore seems robust.

Might this feature be trophic coherence? In Fig. 3.1.C we plot R for the same food webs against the incoherence parameter q. The correlation is significantly stronger than with complexity – stability increases with coherence. However, there are still outliers, such as the food web of Coachella Valley. We note that although most forms of intra-

species competition, such as the struggle for water, habitat or mating, are not described by the interaction matrix, there is one form which is: cannibalism. This fairly common practice is a well-known kind of self-regulation which contributes to the stability of a food web (mathematically, negative elements in the diagonal of the interaction matrix shift its eigenvalues leftwards along the real axis). In Fig. 3.1.D we therefore plot the R and q we obtain after removing all self-links. Now Pearson's correlation coefficient is $r^2 = 0.804$. In other words, cannibalism and trophic coherence together account for over 80% of the variation in stability observed in this dataset. In contrast, when we compare stability without self-links to the other measures, we find that for $\sqrt{S}$ the correlation becomes negative (though insignificant), for $\sqrt{K}$ it rises very slightly to $r^2 = 0.508$, and for Allesina and Tang's estimate it drops below significance (see Section 3.1.1). In Section 3.1.1, we measure stability in a slightly different way, according to Lotka Volterra, type II and type III dynamics, and show that in every case trophic coherence is still the best predictor of stability.

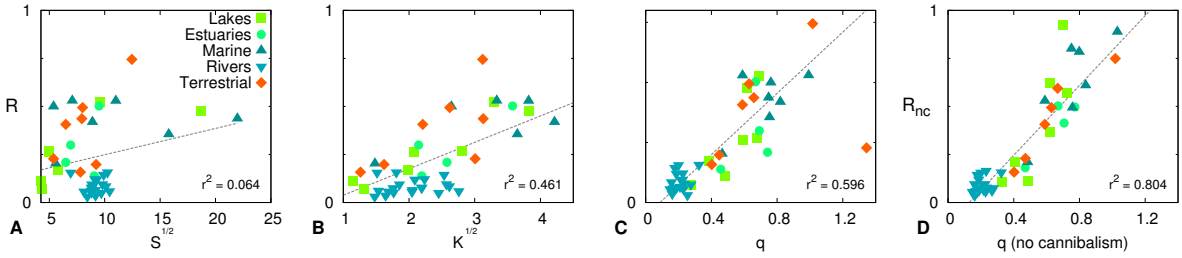

Figure 3.1: Scatter plots of stability (as measured by R, the real part of the leading eigenvalue of the interaction matrix) against several network properties in a dataset of 46 food webs; Pearson's correlation coefficient is shown in each case. **A**: Stability against $\sqrt{S}$, where S is the number of species ($r^2 = 0.064$). **B** Stability against $\sqrt{K}$, where K is the mean degree ($r^2 = 0.461$). **C** Stability against trophic incoherence, q ($r^2 = 0.596$). **B** Stability after all self-links (representing cannibalism) have been removed (R_{nc}) against trophic coherence, q ($r^2 = 0.804$).

3.1.1 Measuring stability

Let us assume that the populations of species making up an ecosystem (each characterised by its total biomass) change through time according to some set of nonlinear

differential equations, the interactions determined by the predation matrix, A (whose elements a_{ij} take the value one if species i preys on species j, and zero otherwise). If the system persists without suffering large changes it must, one assumes, find itself in the neighbourhood of a fixed point of the dynamics. We can study how the system would react to a small perturbation by expanding the equations of motion around this fixed point and keeping only linear terms. The subsequent effect of the perturbation is then determined by the corresponding Jacobian matrix, and the system will tend to return to the fixed point only if the real parts of all its eigenvalues are negative [103].

Even without knowledge of the details of the dynamics, it is possible to draw some conclusions about the stability of a food web solely from its predation matrix [153]. Independently of the exact interaction strengths, we know that not all the biomass lost by a prey species when consumed goes to form part of the predator – in fact, this efficiency is relatively low [137]. It is therefore natural to assume that the effect of species j on species i will be mediated by $w_{ij} = \eta a_{ij} - a_{ji}$, where η is an efficiency parameter which, without further information, we can consider equal for all pairs of species. We can thus treat the interaction matrix, given by

$$W = \eta A - A^T \tag{3.1}$$

as the Jacobian of some unspecified dynamics. However, we have ignored the stabilising effect of intra-species competition – the fact that individuals within a species compete with each other for resources which are not specified by the predation matrix, such as habitat or water. These would correspond to real values to be subtracted from the diagonal elements of W, thereby shifting its set of eigenvalues (or spectrum) leftwards along the real axis. Therefore, the eigenvalue with largest real part of W, as defined above, can be seen as a measure of the minimum intra-species competition required for the system to be stable. Thus, the lower this value, $R = Re(\lambda_1)$, the higher the stability.

In Appendix A we describe this analysis in more detail. Beginning with a general consumer-resource differential equation for the biomass of each species, we obtain the Jacobian in terms of the function $F(x_i, x_j)$ which describes the extent to which species i consumes species j. For the simple (and unrealistic) case $F = x_i + x_j$, the Jacobian reduces to the matrix W as given above, independently of the fixed point. For more

realistic dynamics, the Jacobian depends on the fixed point. For instance, for the Lotka-Volterra function $F = x_i x_j$, the off-diagonal elements of the Jacobian are $J_{ij} = w_{ij} x_i$. If we set $F = x_i H(x_j)$ (with $H(x) = x^h/(x^h + x_0^h)$, x_0 the half-saturation density and h the Hill coefficient), we have either type II ($h = 1$) or type III ($h = 2$) dynamics [208]. Then the off-diagonal elements are $J_{ij} = [\tilde{\eta}(x_i, x_j)a_{ij} - a_{ji}]H(x_i)$, where the effective efficiency is $\tilde{\eta}(x_i, x_j) = \eta h x_0^h x_i x_j^{-(h+1)} H(x_j)^2/H(x_i)$.

The Jacobians for Lotka-Volterra, type II and type III dynamics are all similar in form to the matrix W, although for an exact solution we require the fixed point. Throughout this chapter we therefore use the leading eigenvalue of W as a generic measure of stability. However, in the next subsection (3.1.1) we consider the effects that different kinds of biomass distribution have on each of these more realistic dynamics. The results are qualitatively the same as those for the matrix W, although we find that both the squatness of a biomass pyramid and the level of noise in this structure affect the strength of the diversity-stability relationship described in the main text. After that we go on and scrutinize how do other features such as the proportion of herbivores (sec. 3.1.1), stability criteria (sec. 3.1.1) and sampling errors or inconsistencies in the resolution of the data (sec. 3.1.1) influence the correlation between trophic coherence and stability.

It is evident that this measure of stability depends on the parameter η. In sec. 3.3.2 below we show that the results reported here remain qualitatively unchanged for any $\eta \in (0, 1)$, and discuss how stability is affected when we consider $\eta > 1$ or $\eta < 0$. Further on we also look into the effects of including a noise term so that η does not have the same value for each pair of species, and find that our results are robust to this change too. For the main results in the chapter, however, we use the fixed value $\eta = 0.2$.

Biomass distribution

As discussed at the beginning of section 3.1.1, the Jacobian corresponding to most kinds of biologically plausible dynamics will depend on details of the fixed point. In other words, we need to know the biomass of each species in order to evaluate the Jacobian. Since only a fraction of the energy produced by a species can be used by its consumers, ecosystems can often be regarded as pyramids in which biomass is a

decreasing function of trophic level [3]. More specifically, if we assume that the biomass of a species is a constant fraction of the combined biomass of its resources, biomass will be exponentially related to trophic level. We can thus write

$$x_i = e^{a(l_i - 1)} \tag{3.2}$$

with a a parameter determining the difference in biomass between predator and prey species (for $a = 0$ there is no dependence of biomass on trophic level), and set the basal species to unity biomass. A negative value of a then corresponds to a pyramid in which biomass decreases with trophic level (note that a graphical representation of this situation will look like a pyramid if the size of each echelon corresponds to the logarithm of its biomass). Although terrestrial food webs have this distribution, in certain aquatic environments inverted pyramids can arise, corresponding to a positive a. This is due to the effect of increasing longevity with trophic level, which can compensate to some extent for the inefficiency of predation [3].

In order to examine the robustness of results to fluctuations in this exponential law, we can consider instead a biomass given by

$$x_i = (1 + \xi_i) e^{a(l_i - 1)} \tag{3.3}$$

where the variables ξ_i are randomly drawn from a normal distribution with mean zero and standard deviation σ_x. We can then use these values of $\boldsymbol{x}$ to evaluate the Jacobian for each kind of dynamics and study the behaviour of its leading eigenvalue, R.

Table 3.1 shows the correlations between stability and the various network measures shown in Fig. 3.1 above, over the 46 food webs in the dataset. The first row displays the values for the simple case where the Jacobian is considered equal to the interaction matrix W. The second, third and fourth rows are for the cases of Lotka-Volterra, type II and type III dynamics, with biomass distributed according to Eq. (3.2) and $a = -0.2$. The general pattern shown in Fig. 3.1 is conserved for these more realistic dynamics.

Jacobian	$\sqrt{S}$	$\sqrt{K}$	q	q (no self-links)
W	0.064	0.461	0.596	0.804
W_I	0.045	0.219	0.431	0.730

W_{II}	0.088	0.359	0.456	0.658
W_{III}	0.107	0.426	0.608	0.582

Table 3.1: First column: Jacobian used to compute stability of the empirical food webs of Table 3.4 in sec. 3.6. W is simply the interaction matrix, as used throughout the main text; W_I, W_{II} and W_{III} correspond to types I, II and III, respectively (where Lotka-Volterra is type I). For these cases, we assume an uncorrupted biomass pyramid, as given by $a = -0.2$ in Eq. (3.2). Second, third and fourth column, respectively: value of the correlation coefficient r^2 obtained for R (stability) against $\sqrt{S}$ (where S is the number of species), $\sqrt{K}$ (where K is the mean degree), and q (incoherence parameter). Fourth column: as the third column, after removing all self-links. Compare with Fig. 3.1 above.

Herbivory

Links from basal species (producers) to species which only consume basal species (herbivores) will necessarily have a trophic distance equal to one (see sec. 2.1.3). Since the proportion of basal species, B/S, varies considerably among food webs, we can expect this measure to have a strong bearing on trophic coherence. On the other hand, a large number of basal species may provide a more stable configuration than a network in which many species depend on just a few producers. Might this be the underlying reason for the relation between trophic coherence and stability?

Figure 3.2.A is a scatter plot of q against B/S for the food webs listed in Table 3.4 (70 pg.). There is indeed a significant negative correlation ($r^2 = 0.559$). Figures 3.2.B and 3.2.C show how stability, as measured both before and after removing self-links, varies with the proportion of basal species in the same dataset. The correlations are also significant ($r^2 = 0.475$ for R and $r^2 = 0.505$ for R_{nc}), but slightly lower than we observe in Fig. 3.2.A. In any case, they are much weaker than the correlations shown in Fig. 3.1 between trophic coherence and stability. We can therefore conclude that trophic coherence is the most powerful explanatory variable of stability, while the effect of the proportion of basal species is either less important, or simply an artefact of its

correlation with trophic coherence.

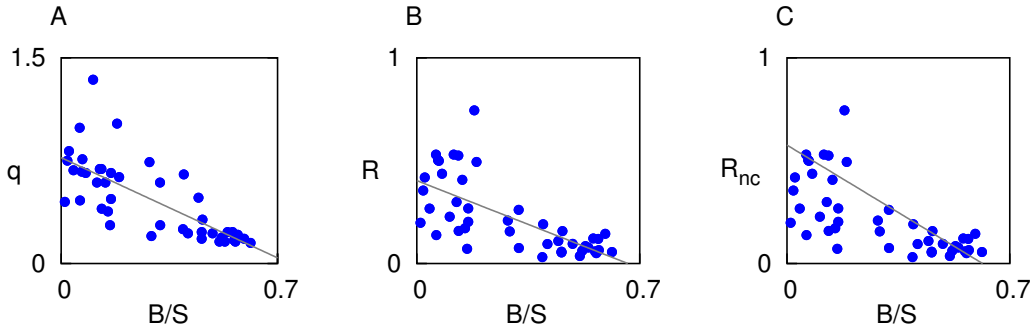

Figure 3.2: Scatter plots, for the food webs listed in Table 3.4, of three network measures against the proportion of basal species, B/S. **A**: Trophic coherence, q, against B/S ($r^2 = 0.559$). **B**: R (real part of the leading eigenvalue of W) against B/S ($r^2 = 0.475$). **C**: R_{nc} (real part of the leading eigenvalue of W after self-links have been removed) against B/S ($r^2 = 0.505$).

Stability criteria

Throughout this chapter we discuss May's result for random networks, according to which the real part of the leading eigenvalue should scale as $R \sim \sqrt{SC} = \sqrt{K}$. We also show that R does not exhibit a significant correlation with $\sqrt{S}$, although we do observe a modest positive correlation ($r^2 = 0.480$) with $\sqrt{K}$. In Figs. 3.3.A and 3.3.B we show scatter plots, for the food webs listed in Table 3.4 in the Data supplement at the end of the chapter, of the leading eigenvalue after self-links have been removed, R_{nc}, against $\sqrt{S}$ and $\sqrt{K}$. In the former case the correlation is now negative but still insignificant, while in the latter the correlation increases slightly to $r^2 = 0.508$. However, food webs are network in which all the links stand for predation (as opposed to other ecological relationships, such as competition or mutualism). Allesina and Tang have recently derived stability criteria for specific kinds of interactions [9]. In particular, when the links stand for predation but are randomly placed among the species, they find that

the real part of the leading eigenvalue should scale as

$$R \sim (1+\rho)\sqrt{SV}, \tag{3.4}$$

where V is the variance of the off-diagonal elements of the interaction matrix W, and ρ is Pearson's correlation coefficient between the elements W_{ij} and W_{ji}. Figure 3.3.C is a scatter plot of R_{nc} against the prediction of Eq. (3.4). Somewhat surprisingly, the correlation is very weak ($r^2 = 0.083$). In Fig. 3.3.D we swap R_{nc} for R (the leading eigenvalue when cannibalism is included) and now the correlation becomes significant ($r^2 = 0.230$), although still relatively low. These results provide further evidence that the structure of food web is non-random in a way which is particularly relevant for their stability.

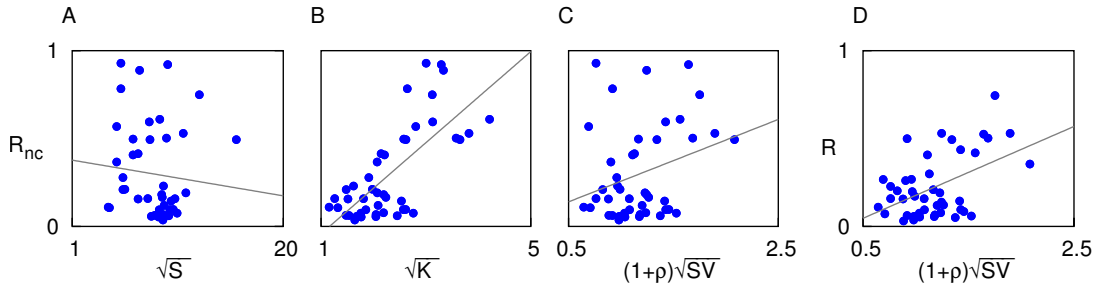

Figure 3.3: Scatter plots, for the food webs listed in Table 3.4, of stability measures against various network values. **A**: R_{nc} (real part of the leading eigenvalue after self-links have been removed) against $\sqrt{S}$ ($r^2 = 0.008$). **B**: R_{nc} against $\sqrt{K}$ ($r^2 = 0.508$). **C**: R_{nc} against Allesina and Tang's prediction, given by Eq. (3.4) ($r^2 = 0.083$). **D**: R (real part of the leading eigenvalue without removing self-links) against Allesina and Tang's prediction ($r^2 = 0.230$).

Missing links and trophic species

Despite important recent developments in food-web inference techniques, it is often hard to ascertain from observation whether a given species consumes another (and even more difficult to quantify the extent of predation). Furthermore, the food webs we have used here for our analysis (described in Section 3.6) were obtained with a variety of different

techniques. To assess whether the patterns we have observed in this dataset, shown in Fig. 3.1 , are robust to possible experimental errors, we remove from each food web a percentage of links, chosen randomly, and recompute each of the magnitudes of interest. After averaging over 100 such tests for each food web, we then recalculate each of the correlation coefficients shown in Fig. 3.1. These are shown in Table 3.2 for different percentages of links removed. As we can see, the dependency of stability on the other magnitudes is barely affected by the random deletion of links: the correlation of R with size is never significant, while the correlation with both complexity and coherence actually increases slightly with the percentage of deleted links.

Missing links	$\sqrt{S}$	$\sqrt{K}$	q	q (no self-links)
0%	0.064	0.461	0.596	0.804
1%	0.061	0.484	0.598	0.814
5%	0.064	0.497	0.635	0.831
10%	0.014	0.545	0.752	0.857
20%	0.002	0.582	0.783	0.845

Table 3.2: First column: percentage of links randomly deleted from the empirical food webs of Table 3.4. Second, third and fourth column, respectively: value of the correlation coefficient r^2 obtained for R (stability) against $\sqrt{S}$ (where S is the number of species), $\sqrt{K}$ (where K is the mean degree), and q (incoherence parameter). Fourth column: as the third column, after removing all self-links. Compare with Fig. 3.1 in page 41.

The nodes in the food webs found in the literature often represent "trophic species". This means that if two or more species in the community share their full sets of prey and predators, they are coalesced into a single node, even if they are in fact taxonomically distinct. However, with recent advances in empirical techniques of food-web inference, larger networks are now being obtained in which nodes represent taxonomic, rather than trophic, species. To find out whether our empirical findings are affected by the degree of taxonomic resolution, we perform a similar test to that of link deletion: for each food

web, we randomly choose a percentage of species to be duplicated – that is, we introduce a new species with the same sets of predators and prey. As before, we average over 100 such tests and recalculate the correlation coefficient for each pair of magnitudes of interest. In Table 3.3, below, we show these results for various percentages of duplicated species. As with the deleted links, we find that the correlations are fairly robust to these modifications, implying that they are not severely affected by the taxonomic resolution of the food webs.

Species duplicated	$\sqrt{S}$	$\sqrt{K}$	q	q (no self-links)
0%	0.064	0.461	0.596	0.804
20%	0.002	0.582	0.783	0.845
50%	0.122	0.406	0.713	0.797

Table 3.3: First column: percentage of species duplicated (as described in Section 3.1.1) in the empirical food webs of Table 3.4. Second, third and fourth column, respectively: value of the correlation coefficient r^2 obtained for R (stability) against $\sqrt{S}$ (where S is the number of species), $\sqrt{K}$ (where K is the mean degree), and q (incoherence parameter). Fourth column: as the third column, after removing all self-links. Compare with Fig. 3.1 in page 41.

3.2 Preferential Preying: foodwebs with tunable coherence

We have seen in the previous section that trophic coherence is a relevant factor in the stability of foodwebs. However, as we show in Figure 3.13 (at the end of this chapter) and Figure 3.7.A modern models of foodweb assembly do not seem to be able to reproduce this feature of empirical networks. In order to be able to study this characteristic we put forward a model we call "Preferential Preying". The basic idea behind this model is that the coherence of the network can be defined by the "trophic preferences" of the species composing the community.

3.2.1 The model

We begin with B nodes (basal species with no incoming connectivity) and no links. New nodes (the consumer species) are sequentially added to the system until we have a total of S species. Their prey is assigned in the following way: a new node i is awarded a **first prey** j chosen randomly from the ones already existent in the community. Then another κ_i nodes m are chosen with a probability P_{im} that decays with the trophic distance between j and m. Specifically, we use the exponential form

$$P_{im} \propto \exp\left(-\frac{|l_j - l_m|}{T}\right),$$

where j is the first node chosen by i, and T is a parameter that sets the degree of trophic specialization of consumers (see Fig. 3.4). In this way, at high T all the prey are chosen randomly, whereas for T close to zero a given node's prey will all have similar trophic levels and the network will be highly coherent ($q = 0$).

The number of prey of each species is drawn from a Beta distribution with a mean value proportional to the number of available species, since this has been shown to provide the best approximation to the in-degree distributions of food webs [235]. The number of extra prey is determined in a similar manner to the Niche Model prescription: $\kappa_i = x_i n_i$, where n_i is the number of nodes already in the network when i arrives, and x_i is a random variable drawn from a Beta distribution with parameters

$$\beta = \frac{S^2 - B^2}{2L} - 1,$$

where L is the expected number of links. In this work, we only consider networks with a number of links within an error margin of 5% of the desired L; thus, for all the results reported, we have imposed this filter on the PPM networks and those generated with the other models.

To allow for cannibalism, the new node i is initially considered to have a trophic level $l_i = l_j + 1$ according to which it might then choose itself as prey. Once i has been assigned all its prey, l_i is updated to its correct value.

The PPM is reminiscent of Barabási and Albert's model of evolving networks [23], but it is also akin to a highly simplified version of an "assembly model" in which species

enter via immigration [32, 158]. It assumes that if a given species has adapted to prey off species A, it is more likely to be able to consume species B as well if A and B have similar trophic levels than if not. It may seem that this scheme is similar in essence to the Niche Model, with the role of niche-axis being played by the trophic levels. However, whereas the niche values given to species in niche-based models are hidden variables, meant to represent some kind of biological magnitude, the trophic level of a node is defined by the emerging network architecture itself. We shall see that this difference has a crucial effect on the networks generated by each model.

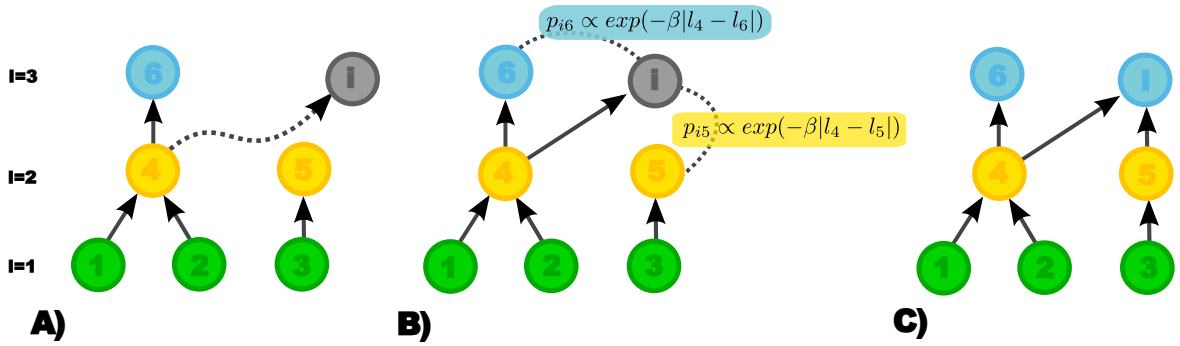

Figure 3.4: Diagram showing how networks are assembled in the Preferential Preying Model (PPM). In Panel **A** a new node, labelled i, is introduced into the network, and is randomly assigned node 4 as its first prey species. In Panel **B**, the probabilities of next choosing node 5 or node 6 are calculated, as functions of their trophic distance to node 4 ($\beta = 1/T$). Node 5 is the closest, and in this case is taken as the second prey species, as shown in Panel **C**.

3.2.2 Preferential preying performance

The model we put forward here is aimed to generate stratified networks (or networks with tunable trophic coherence) in a simple way, and is not initially designed to capture other features of foodweb structure. For the set of natural foodwebs considered in Table 3.4 we have adjusted the free parameter T so as to fit the empirical value of the coherence. This allows us to generate an "imitation" of the original community, in a similar way to how other customarily used models do. However so as to asses to what extent the topologies generated with our model are realistic we have compared

the performance of the networks generated using the PPM against the ones generated with other static models of wide use (The complete description of the models we use as a workbench are in Appendix B). We have considered a set of important measures usually studied in foodwebs, such as stability (as measured in sec. 3.1.1), mean chain lenght, modularity (sec. 2.1.1), mean trophic level and number of apex predators and cannibalistic species. All of these measures have an imporant impact on the network behaviour. We already know that a compartmentalized structure is associated with enhanced resilience of the ecosystems to be perturbed by species extinctions (sec. 2.1.1), and that coherence and cannibalism are strongly correlated with its stability (sec. 3.1). However other features also have an important impact in foodweb performance. For instance, a low *mean chain length* (MCL) has been associated with a high stability [200]. It is defined as the average length (number of links) of all the food chains, i.e. a directed path beginning at at basal species and ending at an apex predator [78], composing a foodweb. The *mean trophic level*, is an average over all the species in a food web of their trophic levels (i.e., $\bar{l} = S^{-1}\sum_i l_i$). It constitutes other importan measure, since it has come to be regarded as an indicator of an ecosystem's health, to the extent that the Convention on Biological Diversity has mandated that signatory states report changes in this measure [195].

We go on and study the overall performance of all the models regarding 8 important measures in foodwebs. For each of the foodweb models and each network measure, we compute the Mean Average Deviation (MAD) of the theoretical prediction, X_{theo} from the empirical value, X_{empi}, simply as $MAD = \langle |X_{theo} - X_{empi}| \rangle$, where $\langle \cdot \rangle$ stands for an average over the 46 food web listed in Table 3.4. The results for each of the eight network measures are shown in the panels of Fig. 3.5. The first panel (summing up the results in Fig. 3.13) clearly show that the niche-based models tend to overestimate the value of q significantly. The fact that none of these models differs substantially as regards q from the predictions of the Cascade Model implies that the various features which they are designed to capture – such as intervality, multiple niche dimensions or phylogenetic constraints – have very little bearing on trophic coherence. The Preferential Preying Model, on the other hand, can reproduce the correct value of q in 45 out of 46 food webs by adjusting its parameter T. The odd web out is that of Coachella Valley, which is slightly more incoherent even than the PPM achieves with low, negative T. This

food web is also the only one in our dataset in which more than half the species indulge in cannibalism, which allows the Coachella Valley food web to exhibit a relatively low R, which it loses when we remove self-links.

The second and third panels show how the models fare as regards stability, both with and without self-links. The PPM achieves significantly better results than the other models in both cases, something we attribute to its reproducing the correct level of trophic coherence. Furthermore, the niche-based models tend to predict less stability than the food webs exhibit, especially in the case without cannibals. This is in keeping with the observation by Allesina and Tang [9] that "realistic" food web structure (i.e., that generated with current structural models) is not conducive to stability.

Next we look at mean chain length and modularity, two measures which have been associated with ecosystem robustness. In particular, a low mean chain length is thought to increase stability [200], while a high modularity might contain cascades of extinctions [95]. In keeping with the first observation, the niche-based models tend to predict longer chains than found in nature; however, they also somewhat overestimate modularity. In any case, the PPM also outperforms the other models on these two measures.

The numbers of cannibals and of apex predators are not very well predicted by any of the models. All but the Nested Hierarchy Model tend to overestimate the cannibals and underestimate the apex predators. Finally, we look at the mean trophic level – a measure which, as mentioned above, is used nowadays to assess the health of marine ecosystems and to monitor the effects of overfishing [195]. As we might expect from this measure's relationship to trophic structure, the PPM does significantly better than the other models at predicting the mean trophic level of food webs. In general, the niche-based models tend to overestimate the mean trophic level.

The comparison we have made here is not as rigorous as one might wish to establish the best food-web model, and this was not our intention. For instance, we have not controlled for the number of parameters, nor attempted to derive likelihoods for each model, as Allesina *et al.* have done [10]. However, we believe it is sufficient to show that a) the failure of current structural models to capture trophic coherence is an important shortcoming; and b) the Preferential Preying Model, which overcomes this problem, generates networks at least as realistic as any of the other structural models. In fact, the PPM significantly outperforms the others on six out of the eight measures

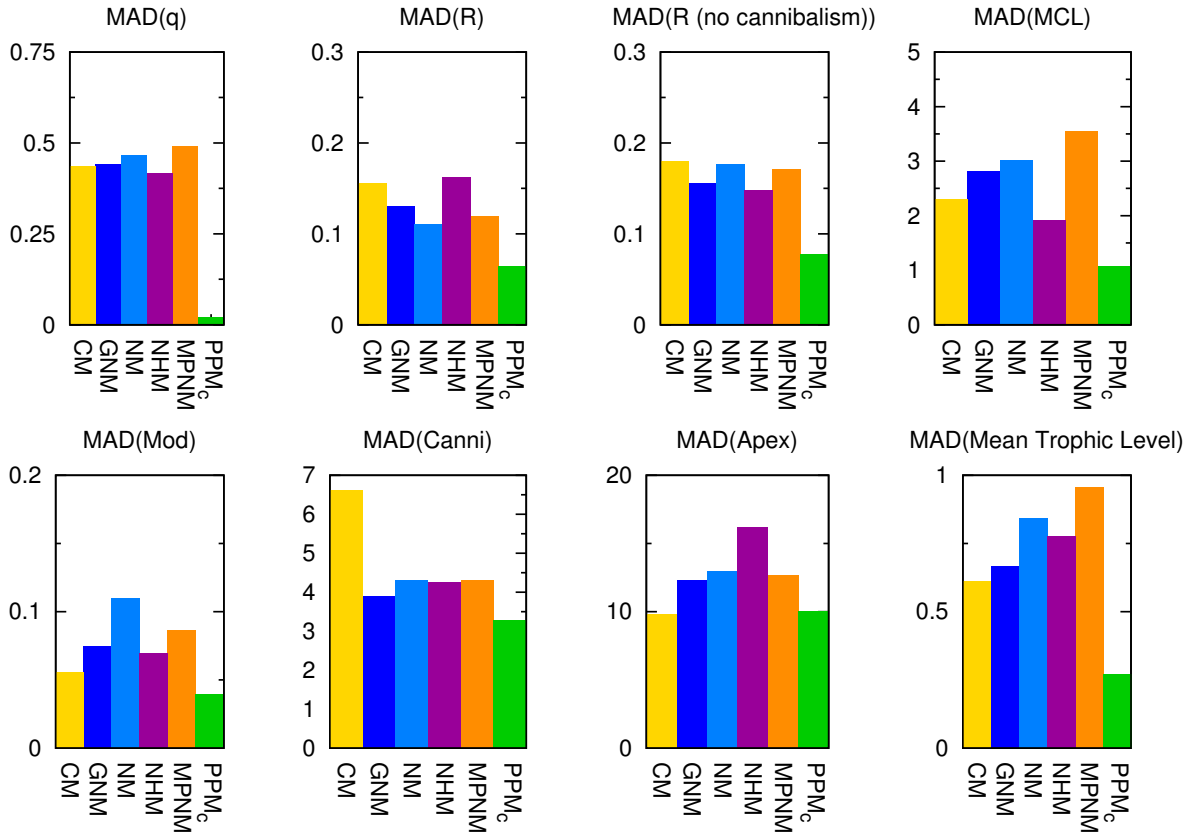

Figure S 3.5: Mean Average Deviation (MAD) form the empirical values returned by each of the food web models discussed in Section Appendix B– Cascade, Niche, Nested Hierarchy, Generalized Niche, Minimum Potential Niche and Preferential Preying – for the network measures described before: trophic coherence q, stability R, stability after removing self-links R_{nc}, mean chain length, modularity, and numbers of cannibals and of apex predators.

we have analysed, and fares no worse on the remaining two. However, the PPM does not capture some of the features known to be relevant in food webs, in particular a phylogenetic signal [171]. The high degree of intervality exhibited by many food webs [236] might be a spurious effect of phylogeny and trophic coherence (both of which we know, from preliminary simulations, to contribute to intervality) or may need to be modelled explicitly, as in the Niche Model. In any case, we hope to have shown that any attempt to build a model which generates networks as similar as possible to real

food webs must take account of trophic coherence.

3.3 The origins of stability

Figure 3.6.A shows three networks with varying degrees of trophic coherence. The one on the left was generated with the PPM and $T = 0.01$, and since it falls into perfectly ordered, integer trophic levels, it is maximally coherent, with $q = 0$. For the one on the right we have used $T = 10$, yielding a highly incoherent structure, with $q = 0.5$. Between these two extremes we show the empirical food web of a stream in Troy, Maine [247], which has the same number of basal species, consumers and links as the two artificial networks, and an intermediate trophic coherence of $q = 0.18$. Figure 3.6.B shows how trophic coherence varies with T in PPM networks. At about $T = 0.25$ we obtain the empirical trophic coherence of the Troy food web (indicated with a dashed line). We also plot q for networks generated with the Generalized Niche Model against "diet contiguity", c, its only free parameter [236]. At $c = 0$ and $c = 1$ we recover the Cascade and Niche Models, respectively (see Appendix B for more information about the models). However, diet contiguity has little effect on trophic coherence.

Figure 3.6.C shows the stability – as measured by R, the leading eigenvalue of the interaction matrix – for the same network. For the PPM networks, stability closely mirrors trophic coherence: as T decreases, the networks become more stable (smaller R) as well as more coherent (smaller q). The empirical value of R is obtained at about the same T which best approximates the empirical q. The Generalized Niche Model also generates more stable networks as diet contiguity is increased, but this effect cannot be due to trophic coherence, which remains nearly constant. The origin of increasing stability in this model is revealed when we measure R_{nc} (R after removing all self-links from the networks): the Generalized Niche Model now displays only a very small dependence of stability on diet contiguity. In contrast, the behaviour of R_{nc} with T in the PPM networks remains qualitatively the same as in the previous case, and the empirical stability continues to be obtained at $T \simeq 0.25$ (in this case, the empirical stabilities R and R_{nc} coincide, since the Troy food web has no cannibals).

We perform this analysis for each of the 46 food webs in our dataset, obtaining the value of T which best captures the empirical trophic coherence according to the

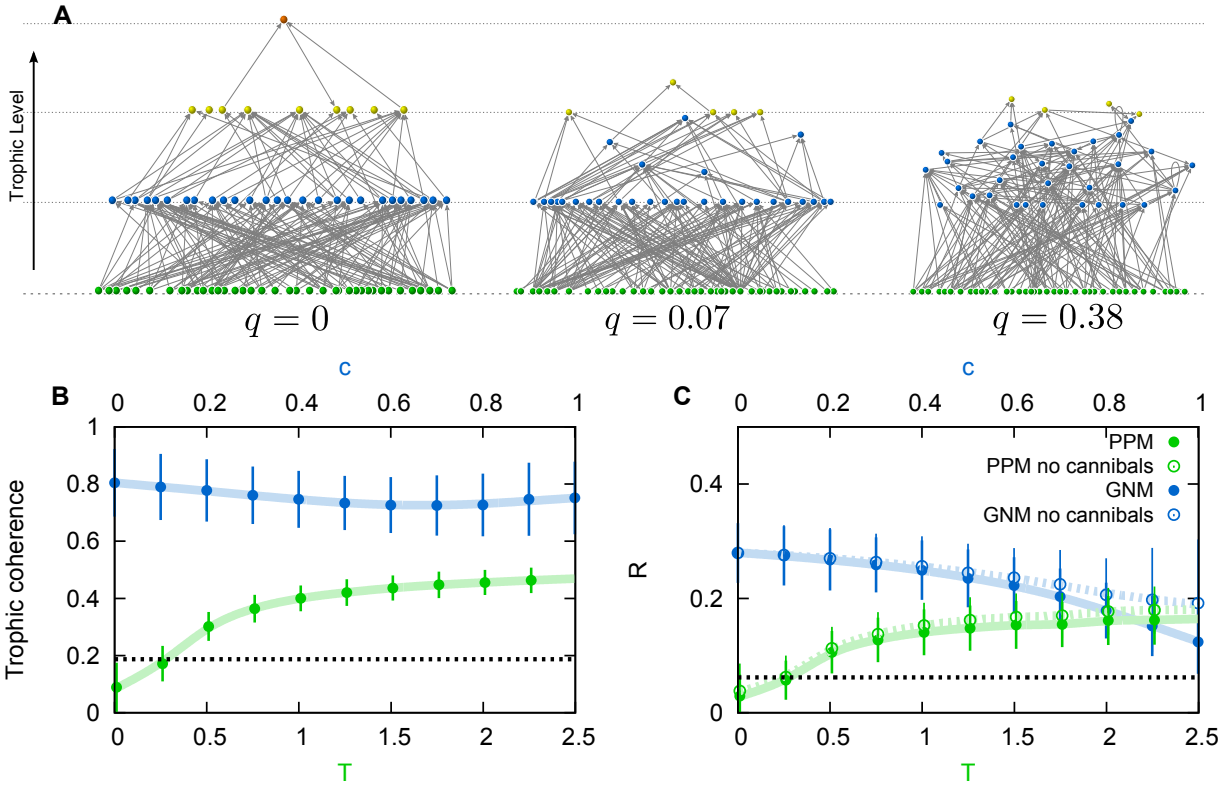

Figure 3.6: **A**: Three networks with differing trophic coherence, the height of each node representing its trophic level. The networks on the left and right were generated with the Preferential Preying Model (PPM), with $T = 0.01$ and $T = 10$, respectively, yielding a maximally coherent structure ($q = 0$) and a highly incoherent one ($q = 0.5$). The network in the middle is the food web of a stream in Troy, Maine, which has $q = 0.18$ [247]. All three have the same numbers of species, basal species and links. **B**: Incoherence parameter, q, against T for PPM networks with the parameters of the Troy food web (green); and against c for Generalized Niche Model networks with the same parameters (blue). The dashed line indicates the empirical value of q. **C**: Stability (as given by R, the real part of the leading eigenvalue of the interaction matrix) for the networks of panel **B**. Also shown is the stability of networks generated with the same models and parameters, but after removing self-links (empty circles). In panels **B** and **C**, the dashed line represents the empirical value of R, while bars on the symbols are for one standard deviation.

PPM. We then compute the ensemble averages of R and R_{nc} generated at this T, for comparison with the empirical values. Similarly, we compute the average values of these measures predicted by each of the niche-based models described above – the Cascade, Niche, Nested Hierarchy, Generalized Niche and Minimum Potential Niche Models (see B for details). The last two models have free parameters, but as these do not have a significant effect on trophic coherence, we use the values published as optimal in Refs. [235] and [10], respectively (or the mean optimal values for those food webs which were not analysed in these papers). Figures 3.7.A-C show the average absolute deviations from the empirical values for trophic coherence and stability, before and after removing self-links, for each model. In Fig. 3.7.A we observe that, as mentioned above, the niche-based models fail to capture the trophic coherence of these food webs. Stability, whether with or without considering self-links, is predicted by the PPM significantly better than by any of the other models, as shown in Fig. 3.7.B and Fig. 3.7.C. This is in keeping with Alessina and Tang's observation that current structural models cannot account for food-web stability [9]. In the previous section (3.2.2) we show the results of similar model comparisons for several other network measures: modularity, mean chain length, mean trophic level, and numbers of cannibals and of apex predators. The PPM does as well as any of the other models as regards numbers of cannibals and apex predators, and is significantly better at predicting the other measures.[2]

Why does the trophic coherence of networks determine their stability? The case of a maximally coherent structure, with $q = 0$ (such as the one on the left in Fig. 3.6.A), is amenable to mathematical analysis. In Appendix D we consider the undirected network that results from replacing each directed link of the predation matrix with a symmetric link, the non-zero eigenvalues of which always come in pairs of real numbers $\pm\mu_j$. We use this to prove that the eigenvalues of the interaction matrix we are actually interested in, if $q = 0$, will in turn come in pairs $\lambda_j = \pm\sqrt{-\eta}\mu_j$, where η is a parameter related to the efficiency of predation (considered, for the proof, constant for all pairs of species). All the eigenvalues will therefore be real if $\eta < 0$, zero if $\eta = 0$, and imaginary if $\eta > 0$. A positive η is the situation which corresponds to a food web – or any system in which

[2]Allesina and co-workers have developed a likelihood-based approach for comparing food-web models [10]. We have not yet been able to obtain the corresponding likelihoods for the PPM, but if this is done in the future it would provide a firmer basis from which to gauge the models' relative merits, and perhaps to build a more realistic model drawing on each one's strengths.

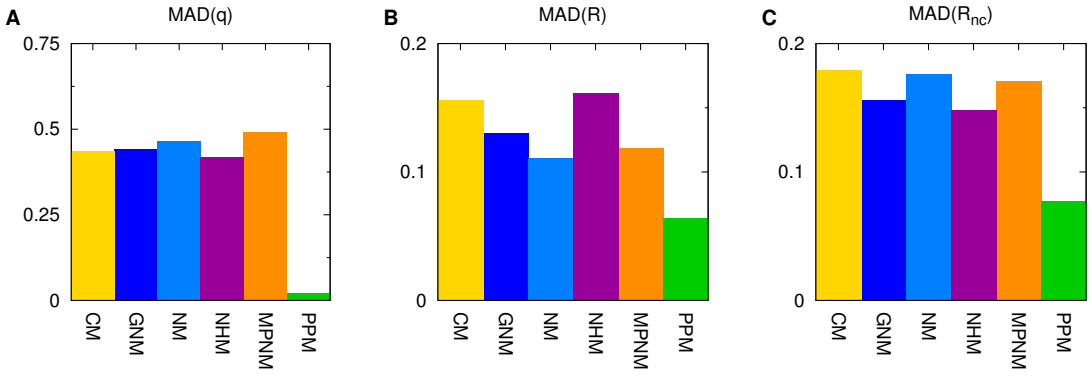

Figure 3.7: **A**: Mean Absolute Deviations (MAD) from empirical values of trophic coherence, q, for each food-web model – Cascade (CM), Generalized Niche (GNM), Niche (NM), Nested Hierarchy (NHM), Minimum Potential Niche (MPNM) and Preferential Preying (PPM) – as compared to a dataset of 46 food webs. **B**: MAD from empirical values of stability, R, for the same models and food webs as in panel **A**. **C**: MAD from empirical values of stability, R, after removing self-links, for the same models and food webs as in panels **A** and **B**.

the gain in a "predator" is accompanied by some degree of loss in its "prey". Therefore, a perfectly coherent network is a limiting case which can be stabilised by an infinitesimal degree of self-regulation (such as cannibalism or other intra-species competition) at any of the nodes. Any realistic situation would involve some degree of self-regulation, so we can conclude that a maximally coherent food web with constant link strengths would be stable.

Although a general, analytical relationship between trophic coherence and stability remains elusive, it is intuitive to expect that a deviation from maximal coherence will drive the real part of the leading eigenvalue towards the positive values established for random structures, as is indeed observed in our simulations.

Before we go on, in the rest of this section further inspect how do different considerations of the stability criteria affect the results of coherence in networks generated with the PPM and Generalized Niche models. In particular we will study how does the biomass distribution (sec. 3.3.1), efficiency parameter η (sec. 3.3.2), and weight of links (sec. 3.3.3) affect the overall stability of these networks.

3.3.1 Effect of biomass distribution

So as to scrutinize how these results depend on the particular choice of stability criteria, we go on and repeat the previous study but using Lotka-Volterra dynamics. We know from section 3.1.1 that in order to evaluate the jacobian we need to impose some biomass distribution. In Fig.3.8 we show the values of R obtained from the Lotka-Volterra Jacobian given by Eq. (3.2) with different values of a, corresponding to pyramid, flat and inverted pyramid distributions of biomass. The empirical values found for the Chesapeake Bay food web [1, 253] with each distribution are compared to the predictions of the Preferential Preying Model against T (left panel), and the Generalized Niche Model against contiguity c (right panel). The effect of the parameter T on stability in the PPM networks remains qualitatively the same as the results reported in the main text for the matrix W given by Eq. (3.1). The more squat the biomass pyramid (the more negative the parameter a), the more stable are both the empirical and PPM networks. This is in keeping with observations of ecosystems [3]. In the Generalized Niche Model networks, however, the effect is opposite: it is the inverted pyramid (positive a), which is most stable. We do not have an explanation for such an effect, but note that it marks a qualitative difference between the networks generated with this model and real food webs.

3.3.2 Efficiency

According to the definition of R above, we must give a value to the parameter η in order to measure stability. The value of this parameter affects the kind of interaction we intend to model with the interaction matrix, $W = \eta A - A^T$, and has a strong bearing on the values of R measured. The definition of W captures the fact that the effect of a prey species on one of its predators is a proportion η of the effect of the predator on the prey. If we are considering the flow of biomass from prey to predator, this should be a relatively small fraction – for instance, the "ten percent law" is often used as a rough estimate of the efficiency of predation [137]. On the other hand, our definition of stability is only strictly independent of the fixed point for a dynamics such as the one described above. For a more realistic dynamics, we might expect a multiplicative factor to appear relating the fixed-point biomass of a prey species to that of one of its

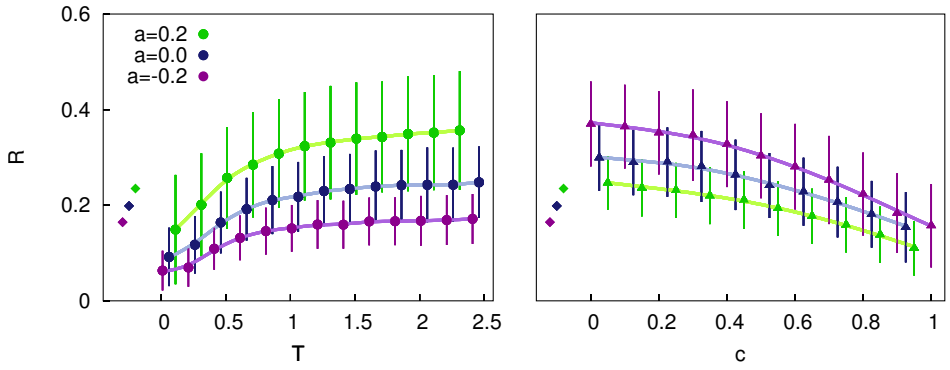

Figure 3.8: Value of R obtained for the Lotka-Volterra Jacobian given by Eq. (A.3), with biomass distributed according to Eq. (3.2) for $a = -0.2$ (pyramid), 0 (flat), and 0.2 (inverted pyramid). In each panel, the diamonds represent the values for the empirical food web of Chesapeake Bay [1, 253]. Circles in the panel on the left show the corresponding results for PPM networks against T using the same parameters; triangles in the panel on the right are for networks generated with the Generalized Niche Model against contiguity, c.

predators. The parameter η might therefore be increased (or decreased) by this effect.

As mentioned above, throughout this chapter we use the value $\eta = 0.2$. However, simulations of the PPM show that using the value of the parameter T which best approximates the empirical degree of trophic coherence is enough to predict the empirical R for a wide range of η. In Fig. 3.9 we show R against T for PPM networks constructed with the parameters of the Chesapeake Bay food web [1, 253] for four cases. We also plot, with an asterisk, the empirical value of R observed in each case, always at the value $T = 0.67$ found to adjust the empirical trophic coherence, $q = 0.47$ (see Table 3.4 at the end of this chapter). The top left panel is for the case of $\eta = 0$, which represents a situation in which the biomass of prey species is completely unaffected by the biomass of their predators. We show in the proof we include in Appendix D that a perfectly coherent network with $\eta = 0$ would have only zero eigenvalues. As incoherence increases, R grows somewhat, though it remains small compared to most cases in which the parameter η simulates a measure of feedback from predators to prey. The top right panel is for $\eta = 0.7$, implying a relatively high efficiency and a strong negative feedback acting on prey species. At $\eta = 1$, all the eigenvalues of W would have

zero real part because it would be an antisymmetric matrix (intuitively, any increase in one node's biomass will be compensated by a decrease in another, so perturbations will be maintained and neither dampened nor amplified). At $\eta > 1$ we simulate a situation such that a predator extracts more biomass form its prey than the latter loses. As we would expect intuitively, this scenario of runaway growth is significantly more unstable than the ones described above. However, the behaviour of R with T is qualitatively similar to that observed for $0 < \eta < 1$. Finally, the bottom right panel corresponds to the case $\eta = -1$, implying that predation reduces the biomass of a predator as well as that of its prey. We know from the proof described in Methods that at $q = 0$ all the eigenvalues of W are purely real for any $\eta < 0$. Similarly, the behaviour of R with T is now inverted: the most coherent networks are now the most unstable.

In the panels corresponding to $\eta = 0$, 0.7 and -1, the value of T which adjusts the empirical trophic coherence also predicts the empirical R very accurately (as we have found for all the food webs in our dataset when using $\eta = 0.2$; see main text). The case of $\eta = 2$ is slightly out: the PPM predicts a slightly higher value of R at $T = 0.67$, although it is not out by much more than a standard deviation. This case of $\eta > 1$ is unlikely to be relevant for ecology; but the small discrepancy serves to remind us that the PPM does not capture all the structural features of real food webs.

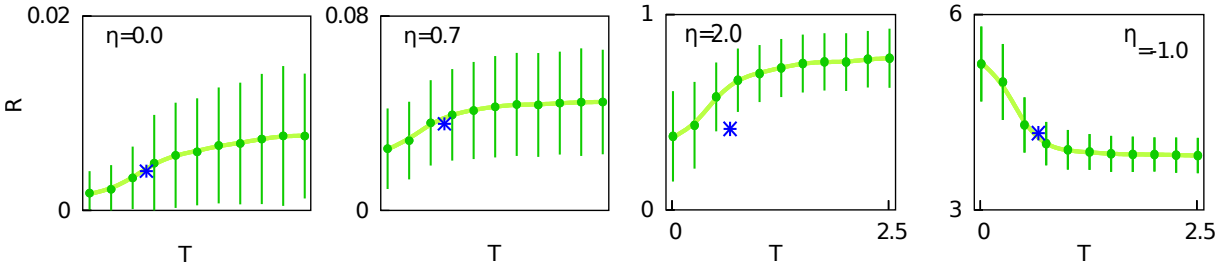

Figure 3.9: Real part of the leading eigenvalue, R, of the interaction matrix $W = \eta A - A^T$ against the parameter T, from averages over networks generated with the PPM for the parameters of the Chesapeake Bay food web [1, 253]. In each panel a different value of the parameter η is used, and the corresponding empirical value of R is represented with a blue asterisk at the value $T = 0.67$, found to predict the empirical trophic coherence $q = 0.47$ (as shown in Table 3.4 in sec. 3.6).

3.3.3 Weighted networks

Although we have been considering the food webs as unweighted networks (the elements in A are either zero or one), in reality certain interactions will be more important than others, and the efficiency η need not be the same for all links. A simple way to look into how these considerations might affect our results is as follows. We make the change $W_{ij} \to (1 + \xi_{ij})W_{ij}$, with ξ_{ij} drawn from a Gaussian distribution of mean zero, standard deviation σ and no correlation between ξ_{ij} and ξ_{ji}. For a given network we then obtain the value of R for many different realizations of the noise $\{\xi\}$. In the left panel of Fig. 3.10 we show the average and standard deviations of R thus defined for three different levels of noise – $\sigma = 0.0$, 0.2 and 0.4 – for PPM networks with the parameters of the Chesapeake Bay food web [1, 253]. We also show (with diamonds) the corresponding averages and standard deviations obtained by performing the same test on the empirical food web. As is to be expected, increased noise leads to a higher average R (lower stability) and a wider standard deviation. However, the behaviour of the average R against the parameter T remains similar with increasing noise, and the value $T = 0.67$ which best adjusts the empirical trophic coherence (as given by Table 3.4 at the end of this chapter) continues to predict the empirical average R at each σ. This is not, however, the case for the Generalized Niche Model. We show the mean and standard deviation of R generated with this model against its contiguity parameter c for the same food web. Whereas the empirical and simulated average values of R correspond at $c \lesssim 1$ when there is little noise, as σ increases the model average R grows faster than the empirical value. This suggests that trophically coherent networks, such as the Chesapeake Bay food web or those generated by the PPM, are more robust to fluctuations in interaction strengths than those generated with niche-based models.

The allometric relationship according to which metabolic rates decline with increasing body size has been shown to reduce predation strength per unit biomass, thereby contributing to stability [41]. Since body size tends to augment (exponentially) with trophic level, this would mean that a more coherent structure would also involve a more homogeneous distribution of link strengths (for a given predator). Therefore, in a more realistic setting in which body sizes and link strengths are considered, we expect the stabilising effect of trophic coherence to be greater than we have shown here for binary networks.

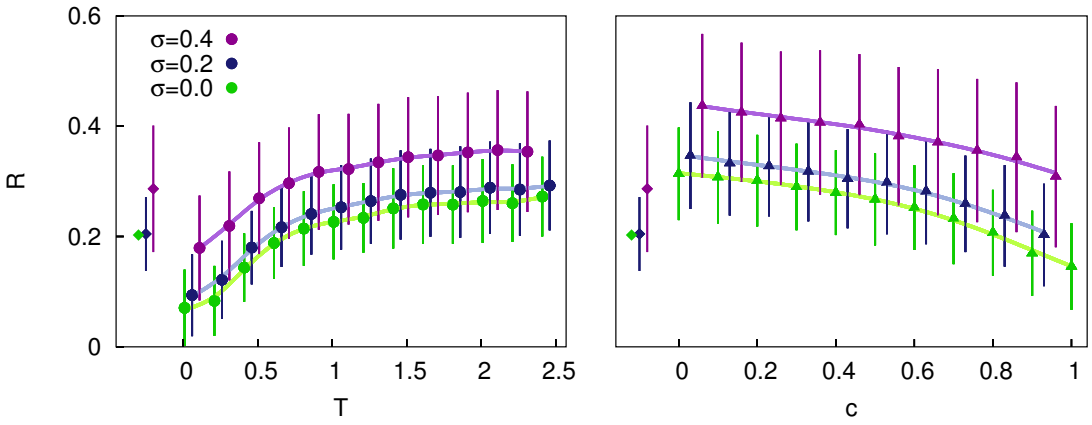

Figure 3.10: Value of R obtained after defining the modified interaction matrix $\tilde{W}_{ij} = (1 + \xi_{ij})W_{ij}$, where ξ_{ij} is drawn from a Gaussian distribution of mean zero and standard deviation σ, and averaging over realizations of the noise $\{\xi\}$. In each panel, the diamonds represent the average values for the empirical food web of Chesapeake Bay [1, 253], with standard deviations as error bars, for noise levels $\sigma = 0$, 0.2 and 0.4. The panel on the left shows the corresponding results for PPM networks against T using the same parameters, while the panel on the right is for those generated by the Generalized Niche Model against contiguity, c.

Feasibility

We have been discussing the potential stability of fixed points of ecosystem dynamics, but for this to be relevant such a fixed point has to be *feasible*. That is, there must exist a fixed point such that every species has a positive biomass. To determine a potential fixed point one must, in general, know the details of the dynamics (as mentioned above). However, even with these specifications, given an unweighted network is is highly unlikely that the fixed point will involve only positive biomasses. However, nature does not have this problem, among other reasons because species' biomasses co-evolve with the interaction weights. If we are granted a certain freedom to set these weights, even if other details of dynamics are set, the problem of finding a fixed point becomes under-specified, and configurations allowing for feasible fixed points might be located. We saw above that the stability of real food webs and those generated by the PPM seem to be more robust to random changes in interaction strengths than their niche-based model counterparts. This suggests that, given a prescription to modify

interaction weights, trophic coherence might enhance the feasibility of fixed points as well as their stability. Such an exercise lies beyond the scope of this paper, but we believe it is a promising avenue of research to be undertaken in the future.

3.4 May's paradox

As we have seen, the PPM can predict the stability of a food web quite accurately just with information regarding numbers of species, basal species and links, and trophic coherence. But what does this tell us about May's Paradox – the fact that large, complex ecosystems seem to be particularly stable despite theoretical predictions to the contrary? To ascertain how stability scales with size, S, and complexity, K, in networks generated by different models, we must first determine how K scales with S – i.e. if $K \sim S^{\alpha}$, what value should we use for α? Data in the real world are noisy in this regard, and both the "link-species law" ($\alpha = 0$) and the "constant connectance hypothesis" ($\alpha = 1$) have been defended in the past, although the most common view seems to be that α lies somewhere between zero and 1/2 [200, 215, 229]. The most recent empirical estimate we are aware of is close to $\alpha \simeq 0.5$, depending slightly on whether predation weights are considered [22]. In our dataset, the best fit is achieved with a slightly lower exponent, $\alpha = 0.41$.

In Fig. 3.11.A we show how stability scales with S in each of the niche-based models when complexity increases with size according to $\alpha = 0.5$. The dashed line shows the slope that May predicted for random networks ($R \sim \sqrt{K} = S^{0.25}$) [150].

We also plot the curve recently shown by Allesina and Tang to correspond to random networks in which all interactions are predator-prey [9], which has a similar slope to May's at large S. This scaling is indeed closely matched by the Cascade Model. The behaviour of the other models is similar (except for the Nested Hierarchy Model, in which R increases more rapidly at high S), and, as expected, networks always become less stable with increasing size and complexity. In Fig. 3.11.B we show how the stability of PPM networks scales in the same scenario. For high T, their behaviour is similar to that of the Cascade Model: $R \sim S^{\gamma}$, with $\gamma \simeq 0.25$. However, the exponent γ decreases as T is lowered, until, for sufficiently large and coherent networks, it becomes negative – in other words, *stability increases with size and complexity.* The inset in Fig. 3.11.B

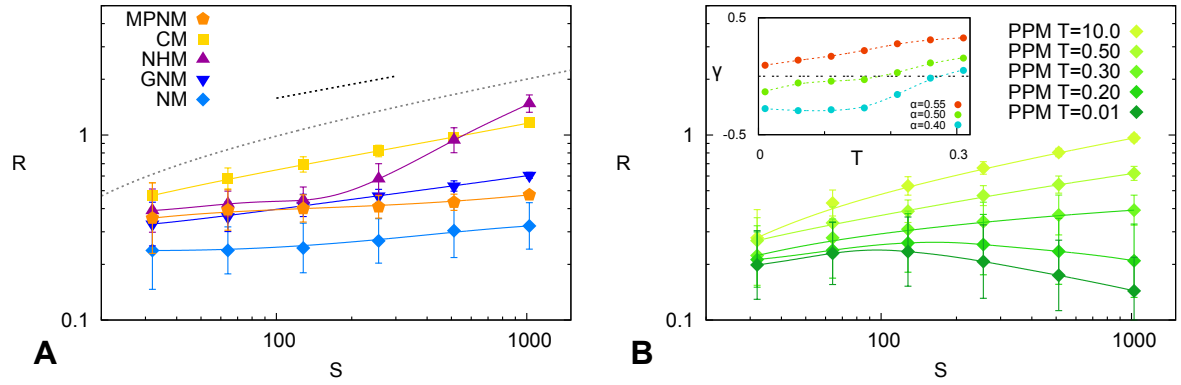

Figure 3.11: **A**: Scaling of stability, R, with size, S, in networks generated with each of the models of previous panels except for the PPM. Mean degree is $K = \sqrt{S}$. The dashed line indicates the slope predicted for random matrices by May [150], while the dotted curve is from Allesina and Tang [9]. **B**: Scaling of stability, R, with size, S, in PPM networks generated with different values of T. In descending order, $T = 10$, 0.5, 0.3, 0.2 and 0.01. $B = 0.25S$. Inset: Slope, γ, of the stability-size line against T for $\alpha = 0.55$, 0.5 and 0.4, where the mean degree is $K = S^{\alpha}$. Bars on the symbols are for one standard deviation.

shows the exponent γ obtained against T, for different values of α. The smaller α, the larger the range of T which yields a positive complexity-stability relationship.[3]

In Section 3.4.1 we extended this analysis to specific dynamics – Lotka-Volterra, type II and type III – by assuming an exponential relationship between biomass and trophic level which can be described as a pyramid. The positive complexity-stability relationship does not appear to depend on the details of dynamics. However, the slope of the $R-S$ curve varies with both the squatness of the biomass pyramid and the extent to which the pyramid is corrupted by noise. A squat pyramid (more biomass at low trophic levels than at high ones) has the strongest relationship, while for an inverted pyramid (more biomass at high trophic levels than at low ones) the slope can flatten out or change sign. Noise in the biomass pyramid tends always to weaken the positive

[3]Plitzko and colleagues recently showed that there exists a range of parameters (in a Generalized Modeling framework [93]) for which Niche Model networks can increase in stability with complexity [202]. However, for this study networks were rejected unless they were stable and had exactly four trophic levels. This selection may have screened for trophic coherence, cannibalism or other structural features.

complexity-stability relationship, and can also change its sign.

3.4.1 Effect of biomass distribution

In Fig. 3.12 we look into how the biomass distribution (as initially defined in sec. 3.1.1) affects the diversity-stability relationship. All networks are generated with the Preferential Preying Model and $T = 0.01$. The first row of panels is for the case where biomass decays with trophic level as an uncorrupted exponential ($\sigma_x = 0$), for Lotka-Volterra, type II and type III dynamics (top panels from left to right). As compared with the constant biomass case ($a = 0$), a decaying distribution is seen to increase the slope whereby R falls with S. In other words, placing more biomass at the bottom of the food web than at the top not only increases stability, but also strengthens the positive diversity-stability relationship exhibited by trophically coherent networks. This occurs for all three kinds of dynamics, although the effect is strongest for type III and weakest for type II. For an inverted pyramid (positive a), R is approximately constant with S.

We go on to analyse the effect of corrupting the exponential distribution of biomass with a noise of standard deviation σ_x. The second row of panels is for $\sigma_x = 0.1$. Although the slope is now less pronounced in all cases, this degree of noise does not undermine the positive diversity-stability relationship for any of the dynamics considered. Finally, in the bottom row we apply a higher noise, $\sigma_x = 0.4$. Now the relationship is inverted and diversity decreases stability. It is not, perhaps, surprising that noise in the distribution of biomass (large σ_x) should have a similar effect on scaling as incoherence in the trophic structure (large T). However, it is interesting that the noise level at which the transition from a positive to a negative diversity-stability relationship occurs does not seem to depend on a or on the kind of dynamics.

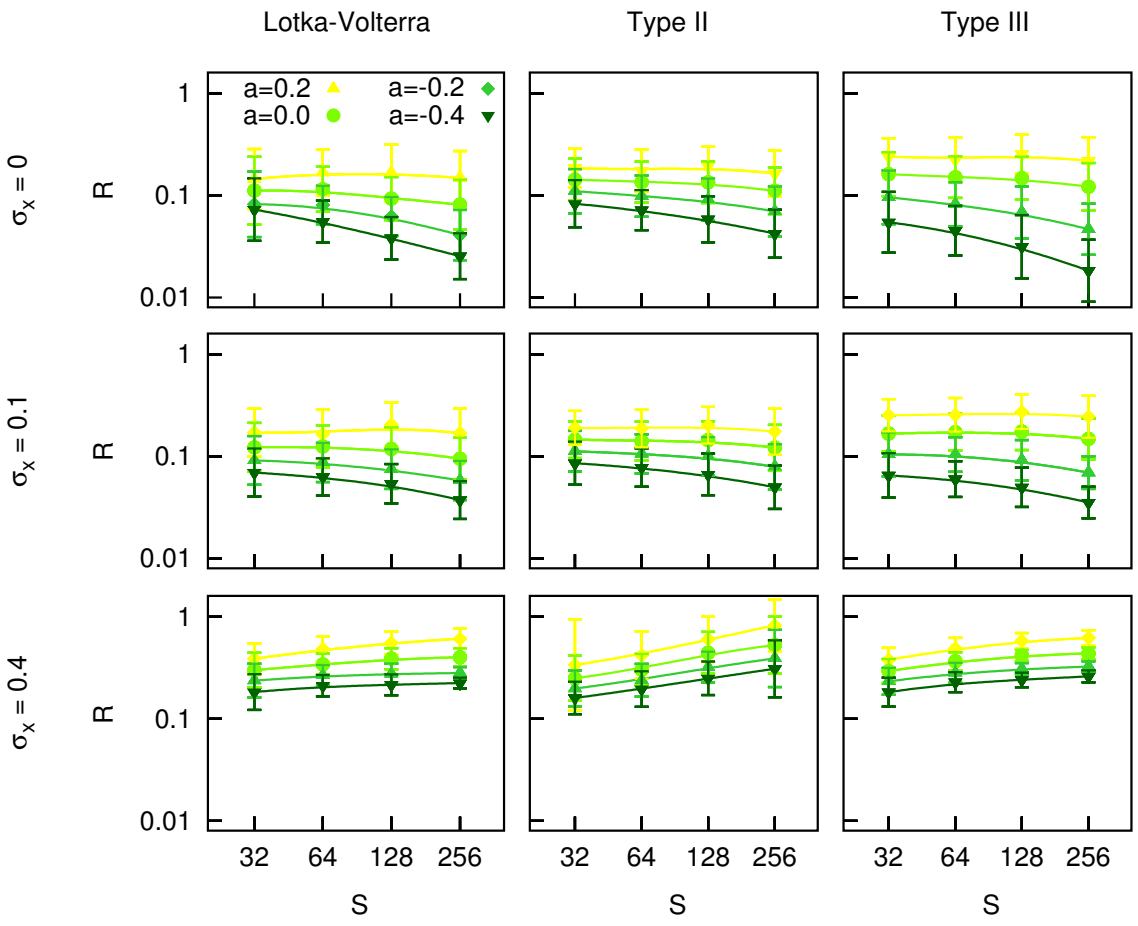

Figure 3.12: Scaling of R with S in PPM networks generated with $T = 0.01$, $K = S^{0.4}$, and $B = 0.25S$. In each panel, from top to bottom, lines are for $a = 0.2$, 0, -0.2 and -0.4. From left to right, columns of panels are for Lotka-Volterra, type II and type III dynamics, as given by Eqs. (A.3) and (A.4). From top to bottom, rows of panels are for levels of biomass noise $\sigma_x = 0$, 0.1 and 0.4 in Eq. (3.3). In types II and III, the half-saturation is set at $x_0 = 1/2$.

3.5 Chapter summary

- Inspection of the predation matrices of natural ecosystems shows they are clearly peculiar in some way since their largest eigenvalues do not depend solely on their size or complexity (as we would expect both from random graph theory and structural foodweb models). We show that the network feature we call **trophic coherence** is strongly correlated with linear stability of food webs. In fact, trophic coherence and cannibalism account for an 80% of the variation in stability in our dataset (Fig. 3.1). This strong correlation of trophic coherence with stability constitues the main finding of this chapter.

- In Appendix A we detail a method to obtain the interaction matrix from the adjacency in a general way, which allows us to computationally study many different dynamics.

- We prove that a perfectly coherent network with constant link strengths will always be stable (Appendix D), establishing an analytical connection between those two features in these particular choice of interaction matrix.

- We put forward a model of foodweb assembly with tunable trophic coherence. Most of the foodweb models to date fails to reproduce the actual cohernece of trophic webs and generate less stratified networks. Although this simple model does not attempt to replicate other characteristic features of food webs – such as a phylogenetic signal or body-size effects – it reproduces the empirical stability of the analysed empirical webs quite accurately once its only free parameter has been adjusted to the empirical degree of trophic coherence (Fig. 3.11). Comparation with other models shows that PPM performs at least as well as them in many other relevant foodweb features (Fig. 3.5).

- Most remarkably, the model also predicts that networks should become more stable with increasing size and complexity, as long as they are sufficiently coherent and the number of links does not grow too fast with size (Fig. 3.11). Although this result should be followed up with further analytical and empirical research, it suggests that we need no longer be surprised at the high stability of large, complex ecosystems

3.6 Data supplement to chapter 2

Food-web data

We have compiled a dataset of 46 food webs available in the literature, pertaining to several ecosystem types. The methods used by the researchers to establish the links between species vary from gut content analysis to inferences about the behaviour of similar creatures. In Table 3.4 we list the food webs used along with references to the relevant work. We also list, for each case, the number of species S, of basal species B, the mean degree K, the ecosystem type, the trophic coherence q, the value of the parameter T found to yield (on average) the empirical q with the Preferential Preying Model, and the numerical label used to represent the food web in several figures below.

Food web	S	B	K	Type	q	T	Reference	Label
Akatore Stream	84	43	2.70	River	0.16	0.26	[246, 247, 249]	18
Benguela Current	29	2	7.00	Marine	0.76	0.87	[275]	11
Berwick Stream	77	35	3.12	River	0.18	0.25	[246, 247, 249]	34
Blackrock Stream	86	49	4.36	River	0.19	0.25	[246, 247, 249]	27
Bridge Brook Lake	25	8	4.28	Lake	0.59	1.15	[99]	14
Broad Stream	94	53	6.01	River	0.16	0.16	[246, 247, 249]	35
Canton Creek	102	54	6.83	River	0.16	0.18	[249]	2
Caribbean (2005)	249	5	13.31	Marine	0.75	0.70	[30]	17
Caribbean Reef	50	3	11.12	Marine	0.99	-0.24	[186]	13
Carpinteria Salt Marsh Reserve	126	50	4.29	Marine	0.65	-8.27	[130]	33
Catlins Stream	48	14	2.29	River	0.20	0.27	[246, 247, 249]	19
Chesapeake Bay	31	5	2.19	Marine	0.47	0.67	[1, 253]	5
Coachella Valley	29	3	9.03	Terrestrial	1.34	-0.02	[203]	12
Crystal Lake (Delta)	19	3	1.74	Lake	0.28	0.33	[252]	37
Cypress (Wet Season)	64	12	6.86	Terrestrial	0.63	0.73	[254]	42
Dempsters Stream (Autumn)	83	46	5.00	River	0.23	0.30	[246, 247, 249]	36
El Verde Rainforest	155	28	9.74	Terrestrial	1.02	-0.82	[265]	15
Everglades Graminoid Marshes	63	5	9.79	Terrestrial	0.66	0.47	[256]	44
Florida Bay	121	14	14.60	Marine	0.59	0.48	[254]	26

German Stream	84	48	4.20	River	0.21	0.29	[246, 247, 249]	28
Grassland (U.K)	61	8	1.59	River	0.40	0.72	[145]	4
Healy Stream	96	47	6.60	River	0.22	0.24	[246, 247, 249]	29
Kyeburn Stream	98	58	6.42	River	0.18	0.18	[246, 247, 249]	30
LilKyeburn Stream	78	42	4.81	River	0.23	0.29	[246, 247, 249]	31
Little Rock Lake	92	12	10.84	Lake	0.69	0.75	[144]	8
Lough Hyne	349	49	14.66	Lake	0.62	0.66	[76, 211]	46
Mangrove Estuary (Wet Season)	90	6	12.79	Marine	0.67	0.47	[254]	43
Martins Stream	105	48	3.27	River	0.32	0.49	[246, 247, 249]	20
Maspalomas pond	18	8	1.33	Lake	0.48	-9.22	[11]	39
Michigan Lake	33	5	3.91	Lake	0.38	0.21	[147]	40
Mondego Estuary	42	12	6.64	Marine	0.74	10.07	[**?**]	41
Narragansett Bay	31	5	3.65	Marine	0.66	1.18	[166]	38
Narrowdale Stream	71	28	2.18	River	0.25	0.38	[246, 247, 249]	21
N.E. Shelf	79	2	17.76	Marine	0.82	0.67	[138]	10
North Col Stream	78	25	3.09	River	0.28	0.34	[246, 247, 249]	22
Powder Stream	78	32	3.44	River	0.22	0.28	[246, 247, 249]	23
Scotch Broom	85	1	2.62	Terrestrial	0.45	0.49	[162]	16
Skipwith Pond	25	1	7.88	Lake	0.68	0.23	[267]	6
St. Marks Estuary	48	6	4.60	Marine	0.69	1.02	[49]	9
St. Martin Island	42	6	4.88	Terrestrial	0.59	0.60	[89]	7
Stony Stream	109	61	7.61	River	0.17	0.18	[249]	3
Sutton Stream (Autum)	80	49	4.19	River	0.15	0.19	[246, 247, 249]	32
Troy Stream	77	40	2.35	River	0.18	0.30	[246, 247, 249]	24
Venlaw Stream	66	30	2.83	River	0.23	0.33	[246, 247, 249]	25
Weddell Sea	483	61	31.81	Marine	0.75	1.01	[108]	45
Ythan Estuary	82	5	4.82	Marine	0.46	0.38	[104]	1

Table 3.4: Details of the 46 foodwebs used throughout this chapter. From left to right, the columns are for: name, number of species S, number of basal species B, mean degree K, ecosystem type, trophic coherence q, value of the parameter T found to yield (on average) the empirical q with the Preferential Preying Model, references to original work, and the numerical label.

Trophic coherence Dataset

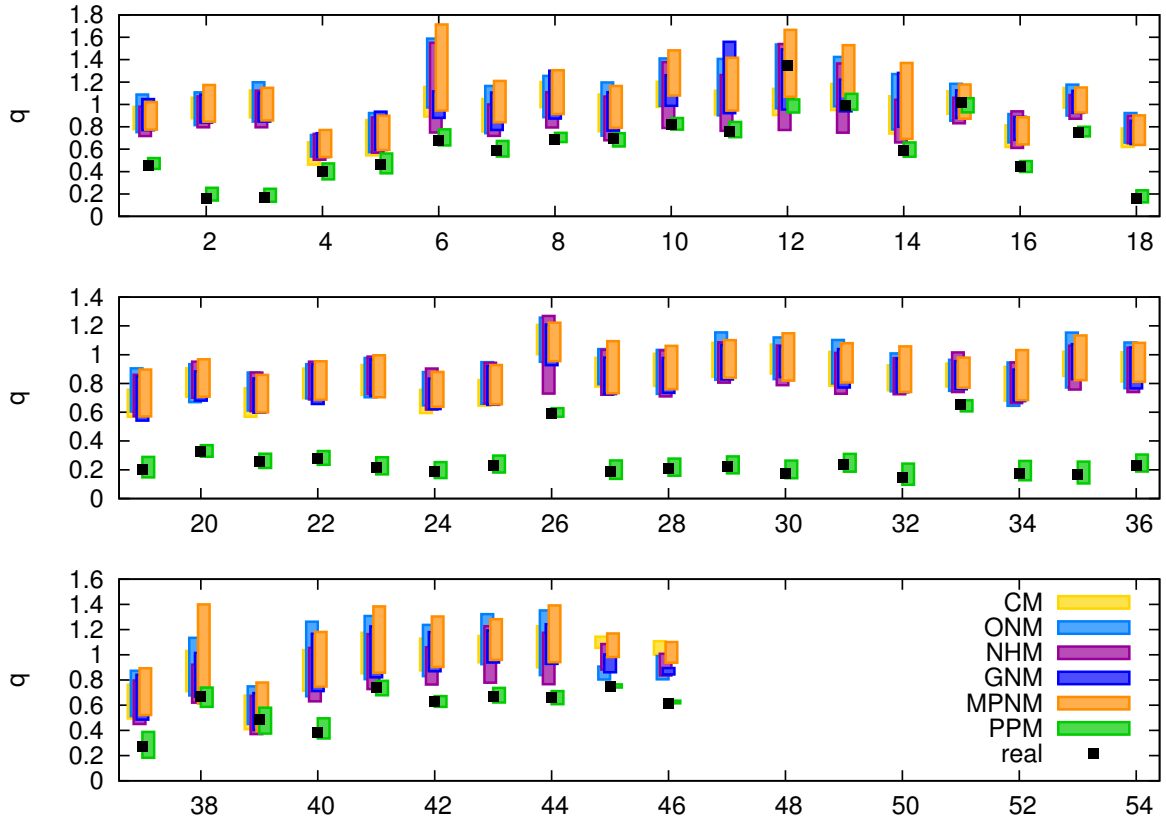

Figure S 3.13: Trophic coherence, as measured by q, for each of the food webs listed in Table 3.4. The corresponding predictions of each food-web model discussed in Appendix B – Cascade, Niche, Nested Hierarchy, Generalized Niche, Minimum Potential Niche and Preferential Preying – are displayed with bars representing one standard deviation about the mean. Empirical values are black squares. The labelling of the food webs is indicated in the rightmost column of Table 3.4.

Stability Dataset

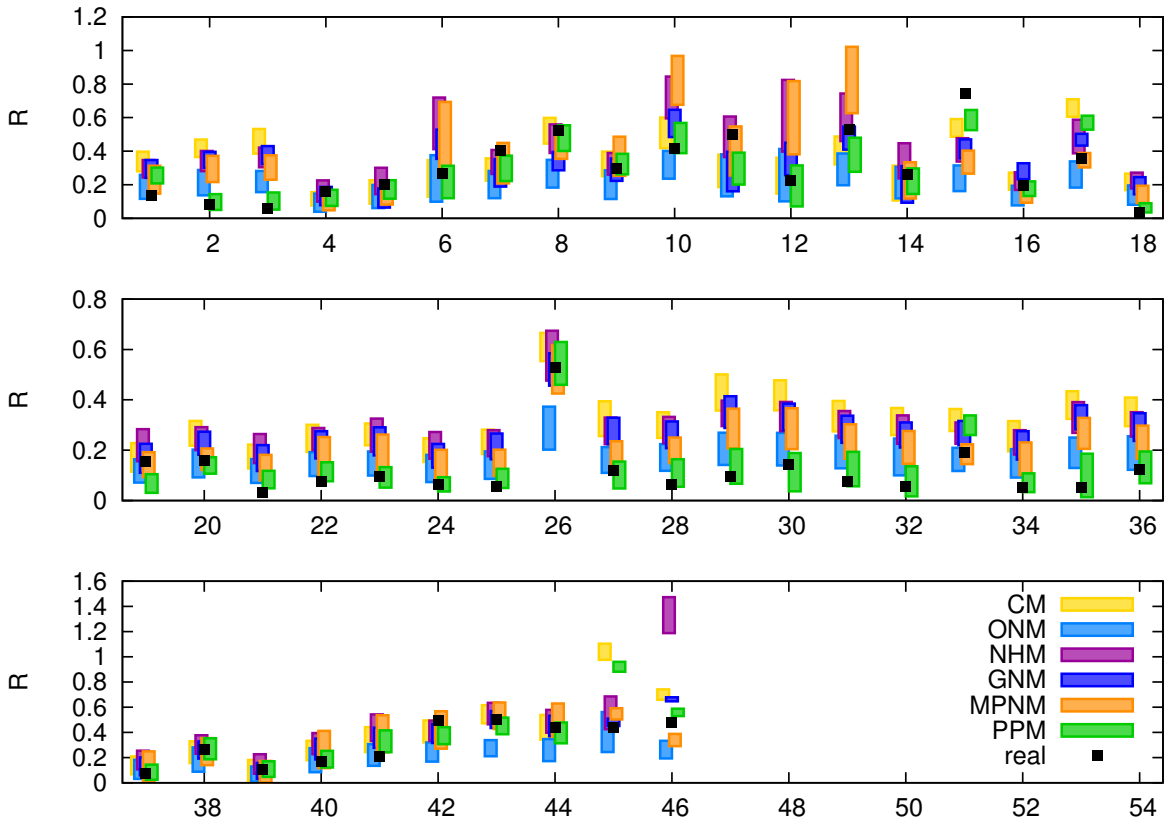

Figure S 3.14: Stability, as measured by R, for each of the food webs listed in Table 3.4. The corresponding predictions of each food-web model discussed in Appendix B – Cascade, Niche, Nested Hierarchy, Generalized Niche, Minimum Potential Niche and Preferential Preying – are displayed with bars representing one standard deviation about the mean. Empirical values are black squares. The labelling of the food webs is indicated in the rightmost column of Table 3.4.

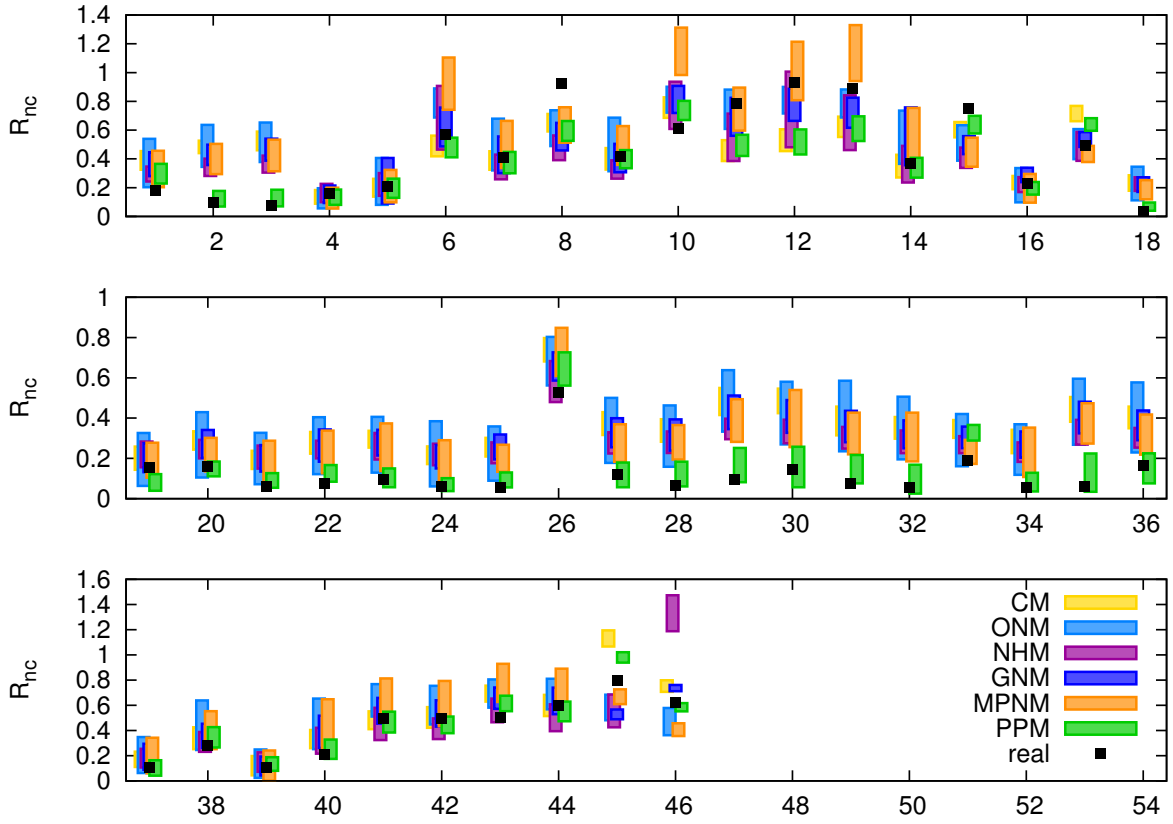

Figure S 3.15: Stability after removal of all self-links, R_{nc}, as measured by R, for each of the food webs listed in Table 3.4. The corresponding predictions of each food-web model discussed in Appendix B – Cascade, Niche, Nested Hierarchy, Generalized Niche, Minimum Potential Niche and Preferential Preying – are displayed with bars representing one standard deviation about the mean. Empirical values are black squares. The labelling of the food webs is indicated in the rightmost column of Table 3.4.

Chapter 4

Inherent directionality in biological and ecological networks

Genetic regulatory circuits, metabolic pathways, foodwebs, and many different biological or even socio-technological systems can be visualized as networks made up of units linked pairwise whenever there is some sort of "interaction" or "flow" between them. In many cases, real networks are dynamical, time-changing entities, and most of the existing compiled datasets represent static snapshots or time-averages over some observation interval of these more complex processes. Nevertheless, the description in terms of static networks has proven to be useful to identify structural features which are responsible for emerging functions [5, 16, 25, 178]. Some structural features, including clustering, degree assortativity [174], and the relative abundance of specific motifs [12, 13], characterise the topology at the local scale. Other traits, such as nestedness [33, P3], community structure [87, 238], and the existence of a hierarchy [56, 207] are related to the large-scale organization. Clearly, these features are not necessarily independent.

In many real networks, interactions are directed, i.e. links have an origin and a target node. This direction can be generally thought of in terms of flows, such as the energy transfer in food webs [74] and the flow of biological information in genetic or neural networks. Often, this flow identifies a global inherent directionality. By "inherent directionality" we mean that all nodes can be ordered on a one-dimensional axis, in such a way that links point preferentially from low to high values of their coordinates in such

an axis. In this sense, the existence of an inherent directionality is deeply related to the existence of a hierarchical structure [56, 269]. For example, (*i*) in networks where there is a transfer of matter, such as foodwebs or metabolic networks, one can identify a hierarchy of "trophic" levels (links tend to point from lower levels to higher ones), (*ii*) in gene regulatory networks there is a hierarchy of control (controller nodes act upon controlled ones), and (*iii*) in neural networks, the flow of information propagates from sensory neurons at the bottom of the hierarchy, to neurons in the central system at intermediate levels, and from there to the level of motor neurons.

The existence of an inherent directionality can have a deep impact on the network small-scale structure, in particular on the statistics of motifs, such as feedback loops. In a directed network, a "feedback loop" of length k is defined as a closed sequence of k different nodes in which a walker following the directions of the arrows returns to the starting point after visiting once and only once all k nodes. Feedback loops are well-known to have a profound impact on dynamical stability in foodwebs [7, 21, 151, 156, 164, 172, 173, 241] as well as in biological and generic networks [4, 12, 15, 35, 60, 124, 136, 143, 180, 198, 201, 216, 248]. "Structural loops" or simply "loops", defined as closed sequences of pairwise connected nodes, independently of the direction of links are also of interest. Clearly, the set of feedback loops is a subset of that of structural loops.

The relationship between the existence of a inherent directionality and feedback loops can be intuitively understood by considering the case of perfect directionality –or feedforwardness– in which all links are aligned with the inherent directionality. In such perfectly directional networks, feedback loops are completely absent, as at least one link against the directionality is required to close a feedback loop. The impact of directionality on the statistics of feedback loops is less trivial to assess in cases of incomplete feedforwardness, where directionality only partially determines the direction of links.

In this chapter we present a simple model relating an assumed degree of inherent directionality with the statistics of feedback loops in networks. Our model depends on a single parameter, γ, defined as the probability of any link in the network to point along the inherent direction (see Fig. 4.1). An analytical calculation allows us to predict the fraction $F(k)$ of loops of length k which are feedback loops. We show that, as long as

there exist a inherent directionality, i.e. as long as $\gamma \neq 1/2$, the fraction of feedback loops $F(k)$ of any loop length k –for which we provide analytical estimations– is much smaller than it would be in network randomizations.

To test the model predictions against empirical data, we scrutinize a number of real biological, ecological, and also socio-technological directed networks. For each of these real networks, we perform an extensive computational study of the number of structural and feedback loops it includes. In nearly all the networks we analyzed, we find that $F(k)$ is dramatically smaller than in randomizations of the same networks. Remarkably, the model reproduces the curves $F(k)$ with good precision for all the real networks we studied, just by fitting its only free parameter, quantifying the degree of inherent directionality.

Furthermore, we introduce a method to directly estimate the degree of directionality in any given network by employing topological information only. The resulting measurement for each specific network correlates quite well with the directionality parameter employed to obtain the fit for the statistics of feedback loops. We also verify that our results are robust against network subsampling or lack of knowledge of existing connections. Therefore, we conclude that the lack of feedback loops stems from the existence of a inherent directionality in real-world networks.

4.1 A model of network directionality

Let us consider a network consisting of N nodes and L directed links and imagine that the fraction of loops which are also feedback loops, $F(k)$, is known. We now aim at constructing a probabilistic model able to predict the empirically-measured function $F(k)$. The model consists in taking the real network under consideration and randomizing the direction of each single link with the constraint that some degree of inherent directionality exists. We therefore assume that nodes can be characterized by an index or coordinate $i = 1 \dots N$ representing their position along the directionality axis. As a convention, we choose higher nodes in the hierarchy to have larger labels, as shown in Fig. 4.1.A. A direction to each existing link is (re-)assigned as follows (see Fig. 4.1.B): a link is set to point from a lower label to the higher one, with probability γ, where the "directionality parameter" γ satisfies $0 \leq \gamma \leq 1$. With the complementary

probability $1-\gamma$ the link points against the inherent directionality. In particular, $\gamma = 1$ (or $\gamma = 0$) stands for perfect inherent directionality, while for $\gamma = 1/2$, the inherent directionality does not affect the direction of the links.

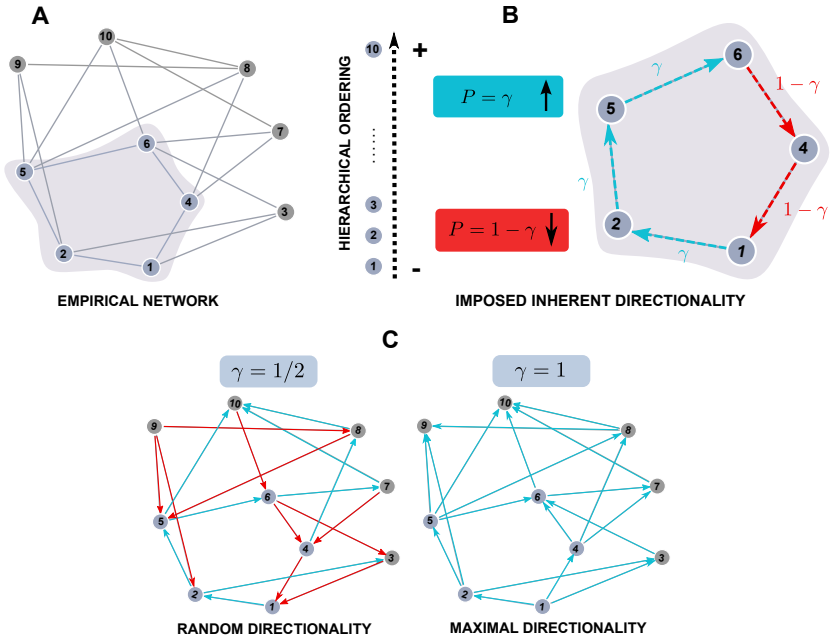

Figure 4.1: **Schematic representation of the directionality model**. (A) A network in which nodes are labeled according to some existing inherent ordering or hierarchy, which identifies an inherent directionality. (B) In any given feedback loop, arrows point in the direction of increasing labels, i.e. along the inherent directionality, with probability γ (blue arrows) or against it with probability $1 - \gamma$ (red arrows). (C) Example of networks with $\gamma = 1/2$ (random directionality) and with $\gamma = 1$ (perfect directionality).

Our goal is to analytically estimate the expected value of $F(k)$ for any given loop length k as a function of the only parameter. To make progress, we consider loops independently, i.e. we neglect possible correlations between for example loops having common links in a same network. We also neglect the impact of possible heterogeneities in the distribution of loops across hierarchical levels. In the case of real networks we are interested in, we shall assume these as working hypotheses, whose validity will be tested *a posteriori* by comparing our results against data.

Under these assumptions, we focus on a specific loop of arbitrary length k (see Fig.

4.1). Without loss of generality, we re-label the node indexes onto the integer numbers $1 \dots k$ by preserving the ordering, i.e. we label the node having the lowest index in the loop with 1, the second lowest with 2 and so on. In this way, the loop is associated with a permutation $\{n\} = n_1, n_2 \dots n_k$, where n_i is the label of the $i-th$ node in the loop. Formally, we define $n_{k+1} = n_1$ to ensure that the loop is closed.

Under the assumptions above, we consider that all the $k!$ possible loop permutations are equally likely to be found. In this way, the maximum number of feedback loops is expected to occur for $\gamma = 1/2$, for which the two directions are equi-probable. In this case, $F(k) = 2^{1-k}$ as only 2 out of the possible 2^k loops of length k are feedback loops. In a more general case, the probability of a given loop to be a feedback loop depends on the distribution of *the number of ascents*, i.e. the number $A(l,k)$ counting how many permutations of the basic sequence of length k are such that $n_i < n_{i+1}$ holds for exactly l distinct values of i. For a non-periodic sequence, i.e. without establishing any relation between n_k with n_1, the solution to this problem is given by the so-called Eulerian numbers (see e.g. [55] chapter 6 or [91]). Since loops are closed, we need to generalize the concept of Eulerian numbers to the periodic or cyclic case, i.e. we need to count the number of ascents in a generic closed loop, which we call "cyclic Eulerian numbers", $A(l,k)$. In sec. 4.1.1 we prove a recursion relation

$$(k-1)A(l,k) = k[(k-l)A(l-1,k-1) + lA(l,k-1)] \tag{4.1}$$

which generalizes a similar relation for standard Eulerian numbers (see e.g. [55]) and which allows us to recursively find all cyclic Eulerian numbers. Notice, in particular, that $A(0,k) = A(k,k) = 0 \ \forall k$ as it is clearly impossible to have all ascents/descent in a closed loop. Examples of cyclic Eulerian numbers for values of k up to 9 are also presented later in section 4.1.1.

The expected fraction $F(k,\gamma)$ of loops of length k which are feedback loops can be expressed as

$$F(k,\gamma) = \sum_{l=0}^{k} \frac{A(l,k)}{k!} \left[\gamma^l (1-\gamma)^{k-l} + \gamma^{k-l}(1-\gamma)^l\right], \tag{4.2}$$

where the two terms in square brackets account for the two different possible orientations of a feedback loop. The function $F(k,\gamma)$ is plotted in Fig. (4.2) as a function of γ for different values of k. $F(k,\gamma)$ is symmetric by exchanging γ by $1-\gamma$, corre-

sponding to reversing the direction of the inherent directionality. Note that imposing the normalization condition $\sum_l A(l,k) = k!$, one can easily retrieve from Eq.(4.2) the probability $F(k,1/2) = 2^{1-k}$ in the limiting case $\gamma = 1/2$.

The exact expression of Eq. (4.2) can be approximated in the asymptotic limit of large k and γ not too small by the following expression (see sec. 4.1.2)

$$F(k,\gamma) \approx 2\exp\left\{\frac{k}{2}\log[\gamma(1-\gamma)] + \frac{k}{24}\log^2\left(\frac{\gamma}{1-\gamma}\right)\right\}. \tag{4.3}$$

Eq.(4.3) predicts that the fraction of feedback loops decays exponentially with the loop length k with an amplitude factor 2 and with an exponential constant which depends on γ.

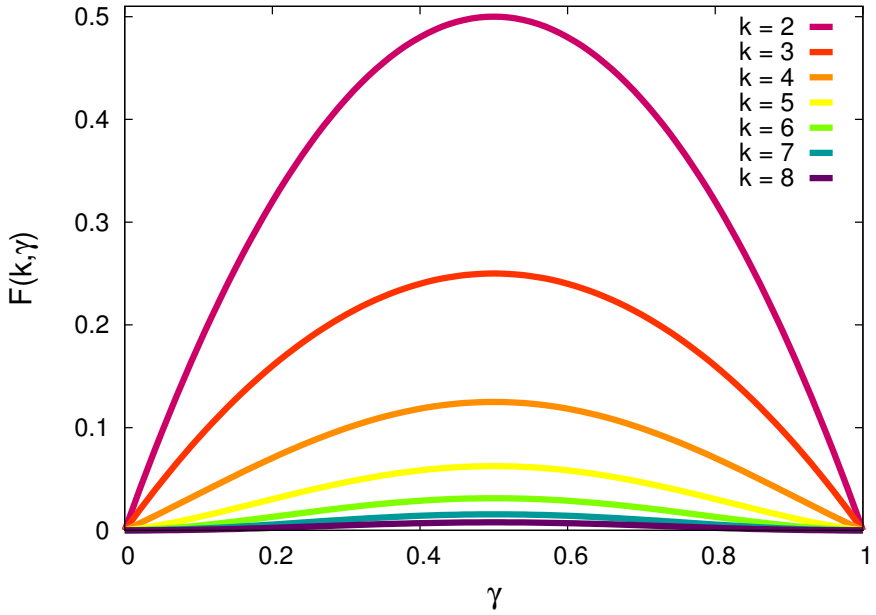

Figure 4.2: **Fraction of feedback loops, $F(k)$, versus the directionality parameter** γ. $F(k)$ has a maximum at $\gamma = 1/2$, for which all link directions are randomly set, giving rise to the largest possible fraction of directed loops. On the other side, $F(k)$ vanishes for $\gamma = 0$ and for $\gamma = 1$ as expected. Notice also that the curves are symmetric around $\gamma = 1/2$ and that for values of γ different from $1/2$ one has a directionality-induced lack of feedback loops.

4.1.1 Number of ascents and Eulerian cyclic numbers

Let us consider a loop of length k, formed by a closed chain of k nodes and k edges, and let us label the nodes with numbers from 1 to k. We consider all the $k!$ possible permutations of labels and aim at computing the number $A(l,k)$ of such permutations including l ascents, i.e. permutations in which exactly l labels in the sequence are immediately followed by a larger one. The first goal is to verify that the $A(l,k)$'s satisfy a simple recurrence relation, similar to that obeyed by standard Eulerian numbers (see e.g. [55] chapter 6 and [91]). To establish such a relation, let us first observe that the number of ascents does not depend on the specific ordering/permutation within a cycle. For instance the permutations 123(1), 231(2) and 312(3), which correspond to three different ways of labeling the cycle $A \to B \to C \to A$, have the same number of ascents (2, in this example). Therefore $A(l,k) = kC(l,k)$ where $C(l,k)$ corresponds to the number of ascents in the case in which the symmetry has been broken and one specific label has been chosen to be at the opening and closing extremes of the representation above. Now we look for a recurrence relation for $C(l,k)$, for which we need to express $C(l,k)$ as a function of $C(j,k-1)$, where $j = l$ or $j = l-1$. These correspond to two different cases that can occur when a new node is inserted in a loop to create a one-step larger sequence. If the node is inserted where there was an ascent, it simply replaces the previous one, so that the number of ascents remains unaltered. If it is inserted where there was a descent, a new ascent is created, so that l is increased by one. These two possibilities can be summarized in the recursive equation

$$C(l,k) = C(l,k-1)l + C(l-1,k-1)(k-1-(l-1)), \tag{4.4}$$

where the two cases above have been weighted with the number of ascents and descents, respectively. Eq. (4.1) follows straightforwardly from Eq. (4.4) and $A(l,k) = kC(l,k)$. Specific values for $k \leq 9$ obtained by iterating the recursive formula are shown in Table 4.1 below.

k l	0	1	2	3	4	5	6	7	8
1	1								
2		2							
3		3	3						
4		4	16	4					
5		5	55	55	5				
6		6	156	396	156	6			
7		7	399	2114	2114	399	7		
8		8	960	9528	19328	9528	960	8	
9		9	2223	38637	140571	140571	38637	2223	9

Table 4.1: Cyclic Eulerian numbers $A(l,k)$,where k is the size of the loop and l the number of ascents.

4.1.2 Devising an asymptotic result

The fraction $A(l,k)/k!$ can be interpreted as the normalised probability of a loop of length k to have l ascents. Using the recurrence relation (Eq. 4.1) it is possible to compute the moments of the distribution

$$\langle l^n \rangle_k \equiv \sum_l l^n \frac{A(l,k)}{k!}. \tag{4.5}$$

It is easy to see using symmetry arguments that its average value is $k/2$, while for large enough values of k, the distribution can be well approximated by a Gaussian with some variance σ_k^2. It is easy to recursively show that the variance of this distribution is exactly given by

$$\sigma_k^2 = \langle l^2 \rangle_k - \langle l \rangle_k^2 = k/12. \tag{4.6}$$

Indeed, first, we can verify it explicitly for the smallest non-trivial case, $k = 3$ (in this case one has $l = 1$ with probability $1/2$ and $l = 3$ with probability $1/2$). Multiplying

Eq. (4.1) by $l^2/k!$ and summing over l yields

$$(k-1)\langle l^2\rangle_k = \langle (k-l-1)(l+1)^2\rangle_{k-1} + \langle l^3\rangle_{k-1}. \tag{4.7}$$

Upon simplifying some terms and using the explicit expression for the first moment, $\langle l\rangle_k = k/2$, it becomes

$$(k-1)\langle l^2\rangle_k = (k-3)\langle l^2\rangle_{k-1} + k^2 - \frac{3k}{2} + \frac{1}{2}. \tag{4.8}$$

The latter expression involves the unknowns $\langle l^2\rangle_k$ and $\langle l^2\rangle_{k-1}$ only. Assuming $\langle l^2\rangle_{k-1} = \sigma^2_{k-1} + < l >^2_{k-1} = (k-1)/12 + (k-1)^2/4$, it is straightforward to show that Eq. (4.8) yields $\langle l^2\rangle_k = k/12 + k^2/4$, therefore proving that $\sigma^2_k = k/12$ for any value of k. Now, approximating the sum in eq.(4.2) by an integral gives finally

$$F(k,\gamma) \approx \int_{-\infty}^{\infty} dl \frac{e^{-\frac{(l-k/2)^2}{k/6}}}{\sqrt{\pi k/6}} \left[\gamma^l (1-\gamma)^{k-l} + \gamma^{k-l}(1-\gamma)^l\right]. \tag{4.9}$$

Evaluating the Gaussian integral explicitly leads to Eq.(4.3) . Let us caution that in principle the integral in Eq.(4.9) should be evaluated between 0 and k, yielding a more complicated expression than Eq.(4.3), involving error functions. Extending the integral over all the real axis is legitimate only when $1-\gamma$ is not too small, so that most of the weight in the integrand is concentrated between 0 and k. When $1-\gamma$ becomes very small ($1-\gamma \approx 0.0025$), this assumption breaks down, and the simple expression of Eq.(4.3) does not hold.

4.2 Infering directionality: Counting loops in empirical networks

Now we go on to analyze a large set of empirical biological, ecological and socio-technological directed networks taken from the literature (for the complete list see Appendix 4.5). We excluded from our analyses un-directed networks and tree-like networks with no single loop of any size. Self-loops –being unrelated to inherent directionality– have not been taken into account. For each network and each loop-length k, we ex-

haustively count the number of structural loops and the fraction of them which are also feedback loops, $F(k)$. We remark that knowledge of the hierarchical level of each node (if any) is not necessary for this computation. From a computational perspective, counting loops is a non-polynomial (NP) hard problem, thus becoming an unfeasible task for large network sizes. For this reason, previous studies often used less computationally-expensive proxies –such as the Estrada index [81] or analytical estimations for large network sizes [38]– to estimate the amount of loops in real networks. Despite the non-polynomial nature of the problem, present computer power allows us to count loops up to reasonably-large sizes by using an efficient breadth first algorithm (see sec. 4.2.1 for more information on the algorithm).

We compare the measured fraction of feedback loops $F(k)$ with two different randomizations of the same network. The first one –that we term directionality randomization (DR)– preserves the existing links, but fully randomizes their directions. The second one – configuration randomization (CR)– randomizes both links and directions, but preserving the in and out connectivity of each single node [165] (see sec.4.2.2 for a more detailed description of the randomization procedures).

Our results, shown in Fig. 4.4, exhibit a clear trend: the fraction of feedback loops of any length k is much smaller in biological and ecological networks than would be expected for any of the two different randomisations. Let us caution that randomly wired networks of finite size can exhibit small statistical deviations from the large-size limit $\gamma = 1/2$.

We remark that also the total number of feedback loops –not just its fraction– is severely reduced with respect to network randomizations in all the considered biological and ecological networks, as firstly noted in [85], and is shown in Figure 4.3. This conclusion holds both for the directionality and the configuration randomizations. Notice that, in most networks, this effect becomes exponentially more pronounced for larger loop sizes.

These trends are not so evident for socio-technological networks; while all of the considered ecological and biological networks have a smaller fraction of feedback loops than their directionality randomizations, some of the social ones (e.g. “twitter followings” and “political blogosphere”) have a larger $F(k)$ than configurational randomizations (see Fig.4.4).

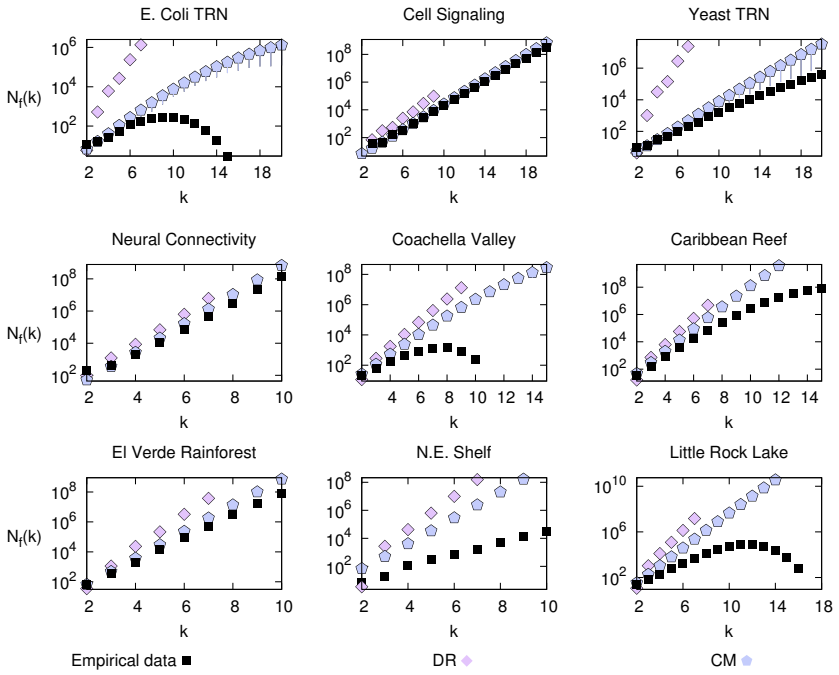

Figure 4.3: **Total number of feedback loops, $N(k)$, in real ecological and biological networks** (black squares) as compared with the directional randomization (magenta diamonds) and the configuration randomization (blue pentagons). Note the logarithmic scale in the vertical axis. In all cases, the number of long feedback loops in real networks is significantly smaller than the random ensemble averages. Errorbars, in most cases smaller than symbol size, correspond to one standard deviation in the randomized ensemble.

We now test the predictions of our probabilistic model against the empirically measured values of $F(k)$ in all real networks. For each of the analyzed empirical networks we consider loop lengths ranging from $k = 3$ to maximum values up to $k = 12$, determined by computational capabilities and depending crucially on network size and connectivity. For each network, we estimate the value of the directionality parameter γ which best describes the observed fraction of feedback loops via an unweighted least-square fit of $\log F(\gamma, k)$ as a function of k.

Results are summarized in Fig.4.4. The model reproduces remarkably well empirical data for all loop lengths by fitting the only free parameter. In some cases, such as for the neural connectivity (C. elegans) network, the agreement between empirical data and model predictions is quite impressive, while significant deviations are observed in some

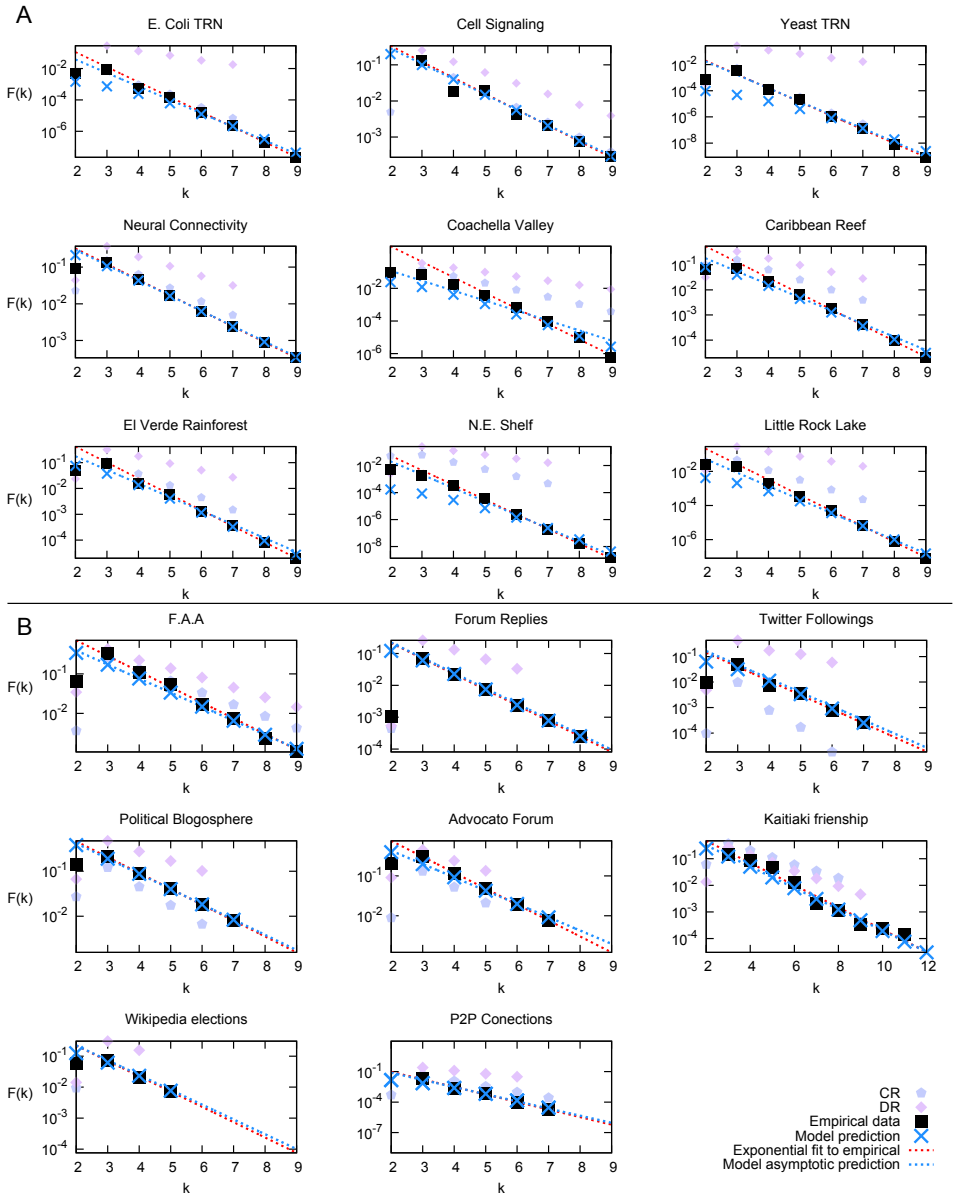

Figure 4.4: **Fraction of feedback loops, $F(k)$, as a function of the loop length, k, in empirical networks**. Black squares correspond to empirical data and red dashed lines stand for fits of the empirical data to an asymptotic exponential curve (fit done using data for $k > 4$). Pale blue pentagons stand for configurational randomizations and pale pink diamonds for directionality randomizations. Blue crosses mark the best fit of our probabilistic model (the parameter γ has been fitted using a least-squares method to $\log F(k)$ versus k). The resulting optimal γ values for the different networks are compiled in Table 4.2 in sec.4.5, at the end of this chapter. Blue dashed lines correspond to the asymptotic analytical estimate of Eq.(4.3) for the corresponding γ. Notice the closeness between the exponential fit to real data and the analytical prediction.

other cases for small loop lengths, $k \leq 4$. In particular, the worst agreement is obtained for the Coachella valley foodweb. However, this network, with only 29 nodes, is the smallest in the dataset, so it can deviate significantly from statistical predictions and it has been previously reported to be anomalous from other viewpoints [114]. In some cases, such as the N.E. Shelf foodweb and the two considered transcription regulatory networks (*E. coli* TRN and Yeast TRN), $\gamma > 0.999$ indicating a rather extreme level of inherent directionality (see Table 4.2). We obtained similar results for other empirical networks with very few loops (listed in Table 4.2 as well), providing additional support to our conclusion.

As the model predicts an asymptotic exponential decay of $F(k)$ as the loop-length k increases, we have performed –for each particular network– a fit of the empirical data (for $k > 4$) to an exponential function (see dashed red lines in Fig. 4.4). In this case, the quality of the fit of $\log F(k)$ versus k can be assessed via a linear regression coefficient, r. Obtained values of r^2 (Table 4.2) are larger than 0.99 in all cases except one –the Mammalian cell signaling network, for which $r^2 = 0.973$– indicating that even for relatively small loop-lengths the predicted asymptotic exponential decay holds. Furthermore, each of these exponential fits is very close to its corresponding analytically-obtained asymptotic result, Eq. (4.3) (blue discontinuous lines in Fig. 4.4). In the few cases in which the analytical asymptotic prediction breaks down (see sec. 4.1.2) the blue dashed lines correspond to a fit of the model data for $k \leq 4$. This shows that the asymptotic expression is reasonably accurate even for rather short loops.

We conclude this section with a remark on the possible impact of unknown links. Our knowledge of biological and technological networks is often incomplete and it is important to assess how this fact may affects our analyses. To test the robustness of our framework, we mimicked the effect of undersampling of empirical networks by eliminating a fraction of the links at random, and repeated the analysis above. While this operation clearly affects the number of links, the conclusions of our model (in particular the fitted value of γ) are very weakly modified even when a relatively large fraction of nodes is removed $20\% \sim 50\%$ (see figure 4.5 below).

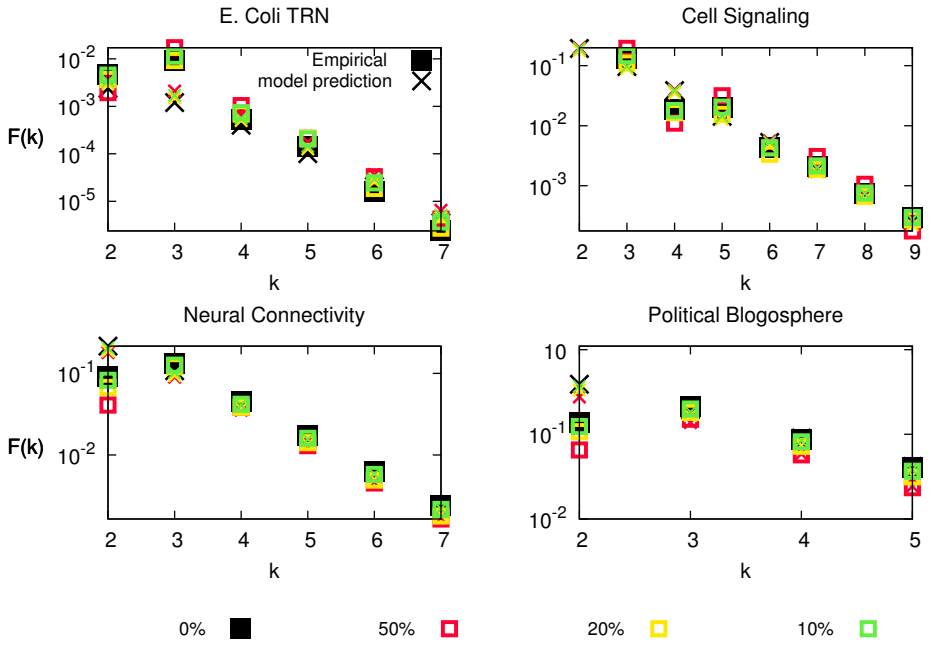

Figure 4.5: **Fraction of feedback loops at removing a percentage of links**. Black squares correspond to empirical data. Other curves are obtained by randomly deleting a fraction of the links (the fraction is shown in the legend). Fitted values of γ for a fraction $(0, 10\%, 20\%, 50\%)$ of removed links are: Ecoli $(0.9988, 0.0.9983, 0.9985, 0.9979)$ Cell Signaling $(0.89, 0.89, 0.90, 0.88)$ Neural Connectivity $(0.88, 0.88, 0.90, 0.90)$ Political Blogosphere $(0.73, 0.76, 0.79, 083)$. Observe the very mild variations induced by random deletion of links.

4.2.1 A novel algorithm for measuring loops in networks

We consider a generic network with N nodes and L directed edges, codified by its adjacency or connectivity matrix A, such that $a_{ij} = 1$ if there is a directed link from node i to node j and $a_{ij} = 0$ otherwise. We consider the simplest possible case of unweighted and unsigned directed networks and keep generalizations to those cases for a future study. Our aim is to count the number of directed and structural loops of any possible length k. A feedback loop is a self-avoiding directed walk starting and finishing at the same network node, i.e. the walk follows the direction of the links and any node appears at most once in any particular loop.

We have devised a breadth first algorithm [126] that counts closed paths starting from all possible root nodes, $i \in [1, N]$, in a sequential way. From any given node,

the tree of possible paths branches out following the adjacency matrix. Within a given search, a Boolean variable σ is assigned to each node to distinguish visited from unvisited nodes. The process is iterated until either: (i) the starting node is reached, (ii) a predetermined maximum allowed loop-size has been reached, or (iii) every accessible node has been visited. Note that searches with arbitrarily long loops would be prohibitively expensive for medium-sized and large networks and, therefore, we need to limit our search to paths up to a maximum length (typically, from 7 to 20 depending on network size and connectivity). Similarly, to determine the total number of structural loops we symmetrise the network (i.e. we construct a new adjacency matrix $B = A + A^T$) and run the same algorithm as above. Given the proliferation of links and paths, the search becomes much slower in this case. The algorithm has been tested using as a benchmark Erdős-Rényi networks [178] for which analytical expressions for the number of (directed and undirected) loops are known.

4.2.2 Network randomizations

Our measurements of the fraction of feedback loops $F(k)$ for real networks have been compared along the main text with two different network randomizations.

Configuration randomization (CR) consists in randomizing the original network with the constraints that the total number of nodes and links, as well as the (in and out) connectivities of each node are preserved. Algorithmically, this is achieved by randomly selecting two different edges, say $i \to j$ and $k \to l$, switching the ending points: $i \to l$, $k \to j$, and repeating the operation many times. The ensemble generated iterating this process is usually call the "configuration model" [165].

Directionality randomization (DR) consists in maintaining the original topology of the network, but performing randomizations of the direction of all the links. For each link, each of the two directions is chosen with probability 1/2. This randomization preserves the number of nodes and links in the network as well as the degree sequence [178]. A random network ensemble can be build up by repeating the randomization as many times as needed. In such random ensemble, a loop of length k is a feedback loop with probability 2^{1-k} and it is non-directed with complementary probability $1 - 2^{1-k}$.

For the sake of completeness, we have verified that our conclusions remain unchanged when considering two further randomization ensembles[149, 197, 278]:

(*i*) **mean directionality randomization (MDR)**, in which directions are randomized but the total connectivity of each node is preserved only on average, i.e. links join randomly selected origin and target nodes, which are chosen with probability proportional to their total connectivity (including in and out links) in the corresponding real network, and
(*ii*) **mean configuration randomization (MCR)** similar to the CR above, but where the in and out connectivities of each node are preserved only on average (in particular, two nodes are connected with a probability proportional to the out-connectivity of the first times the in-connectivity of the second).

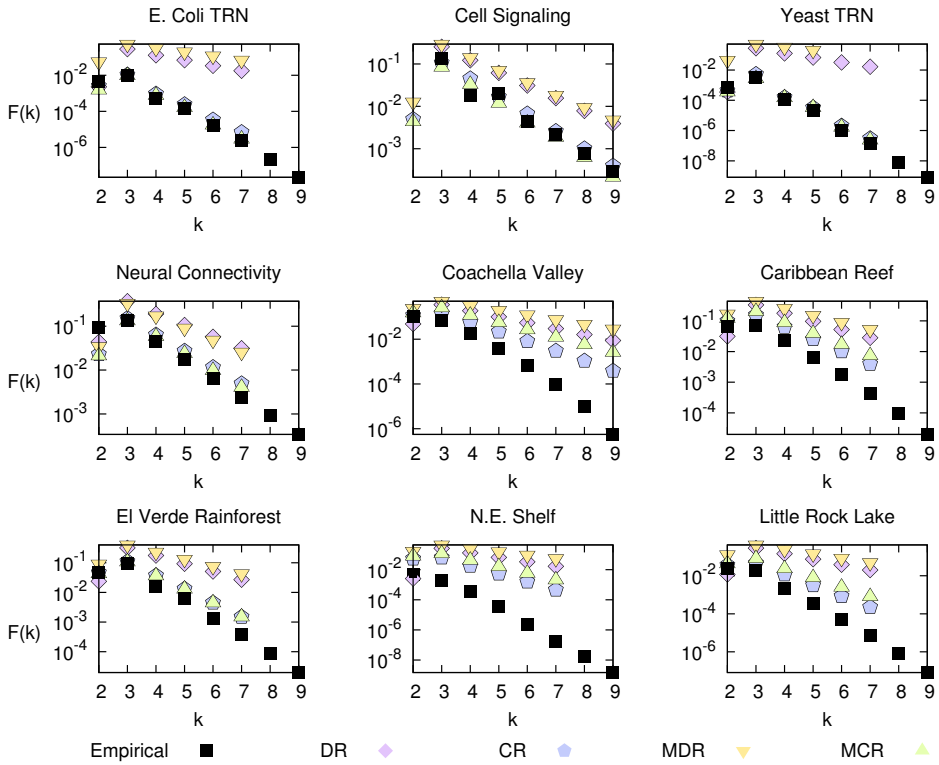

Figure 4.6: **Fraction of feedback loops, $F(k)$, as a function of the loop length, k, in empirical networks compared with 4 different randomisations**. All the attempted randomisations have a significantly larger fraction of feedback loops than empirical ones (Black squares correspond to empirical data; pale pink diamonds correspond to directionality randomisations (**DR**), pale blue pentagons to configurational randomisations (**CR**), yellow downward triangles to mean directionality randomisations (**MDR**), and, finally, green upward triangles to configuration randomisations (**MCR**).

Fig. 4.6 shows results analogous to those of Fig. 4.4, in which $F(k)$ is plotted as a function of k for real networks as well as for the four different randomizations discussed above. As it can be seen, the conclusions are robust for all these randomizations.

4.3 Measuring directionality in empirical networks

The directionality parameter γ in the probabilistic model represents how strongly the hypothesised hierarchical ordering affects the direction of the links in the network; $\gamma = 1$ (and also $\gamma = 0$) reflect perfect directionality while $\gamma = 1/2$ corresponds to random directionality. In the previous section, γ has been inferred from the statistics of feedback loops.

We now propose an algorithm to directly measure the degree of directionality of a network from its topology. Similar methods have been proposed for this purpose [131, 213, 276]. All of them are able to extract a hierarchical ordering from a given network and classify nodes into a few discrete levels. Instead, the method we propose produces more refined orderings, being able to resolve possible degeneracies between the coarser levels produced by previous methods.

Our method is inspired by algorithms for determining trophic levels in food-webs, but is applicable to any directed network; it can be also seen as a way to infer a "hidden variable" from network topology [222]. The general idea is that *basal* nodes (those with no incoming connections) are assigned the lowest value in the hierarchy ($l = 1$) and the rest of the nodes are assigned the mean level of their incoming nodes plus one. The set of equations defined by this method constitute a system of linear equations that can be solved with standard methods. For turther deatils see sec. 4.3.1. Notice that, while with existing methods [131, 213, 276] hierarchical levels associated to nodes are integer numbers, here they are in general real numbers.

Using the hierarchical ordering resulting from applying the algorithm above, it is straightforward to compute the fraction of links pointing from lower to higher hierarchical levels, i.e. aligned with the inherent directionality. We call this fraction "current parameter", and refer to it as χ (see sec. 4.2.1). In the limit of perfect feedforwardness one expects $\chi = 1$, while in the absence of a well-defined directionality $\chi \approx 1/2$.

Our results are summarized in Fig. 4.7, and they clearly show that all the consid-

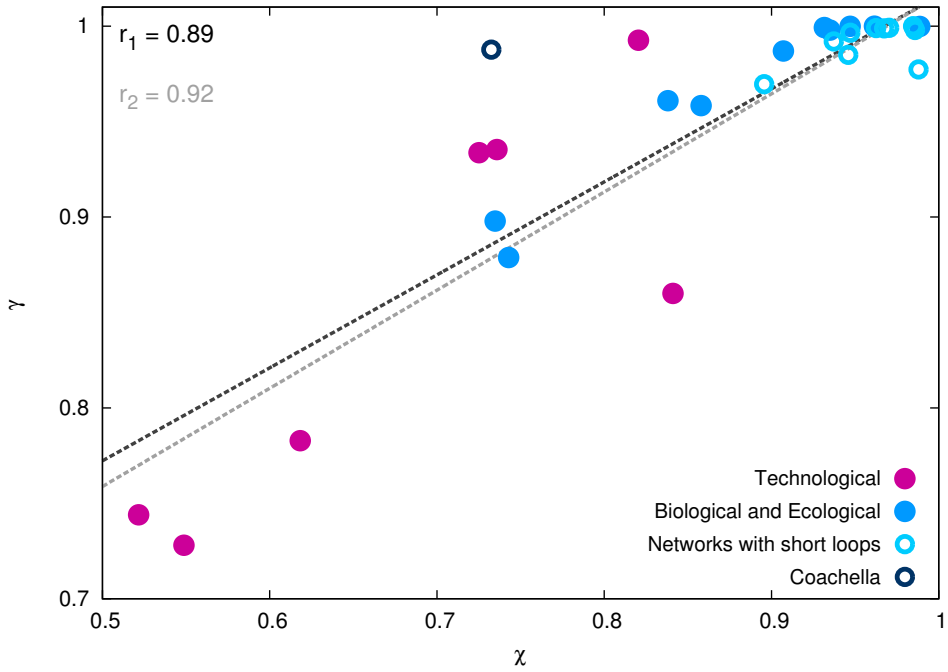

Figure 4.7: **Correlation between inferred and explicitly measured levels of directionality**. Scatter plot of the optimal values of the directionality parameter γ plotted against the current parameter χ. Values of either γ or χ close to 1 reflect a high degree of directionality while smaller values close to $1/2$ imply that link directions are nearly uncorrelated with directionality. The value of the linear correlation coefficient is $r = 0.89$ or $r = 0.92$ depending on whether the outlier "Coachella Valley" small network is included or not.The corresponding best fits are $\gamma = 0.487\chi + 0.529$ and $\gamma = 0.514\chi + 0.502$, respectively.

ered biological, ecologial, and also –to much lesser extent– socio-technological networks exhibit some degree of hierarchy, represented by a current parameter $\chi > 1/2$. More remarkably, the explicitly measured values of χ correlate quite well with the fitted value of the directionality parameter γ in the set of networks under study. This correlation implies that the free parameter we use to fit the directional model is consistent with a direct measure of directionality (current) in the same networks.

4.3.1 Algorithm details

Along this chapter we have proposed an algorithm to directly measure the degree of directionality in empirical networks. It is inspired in previous algorithms for determining the trophic hierarchy in foodwebs, but can be applied to any kind of directed network..

Extract hierarchical ordering: As customarily done with food-webs, one identifies "basal nodes" as those having zero in-connectivity, i.e. with no link pointing to them. In the conceivable case in which no basal node exists, we progressively identify sets made out of two, three... nodes which –taken as a unique coarse-grained node– are basal, i.e. no external node points to any node in the set. Basal nodes obtained in this way are placed at the lowest level of the hierarchical ordering, $l = 1$ (see fig.4.8). Then, the level of the remaining nodes is defined as the average of the trophic level of all nodes pointing to it (its preys in food-webs) plus 1:

$$l_j = \frac{1}{k_j} \sum_i a_{ij} l_i + 1 \tag{4.10}$$

where k_j is the in-connectivity of node j, a_{ij} is the connectivity or adjacency matrix and l_j is the hierarchical level of node j. The conditions (4.10) define a set of linear equations in the unknown l_j's, that can be written in terms of a modified laplacian matrix

$$\mathbf{\Lambda}\vec{l} = \mathbf{1}, \tag{4.11}$$

where $\vec{l}$ is the vector of trophic levels. The modified laplacian matrix has off-diagonal entries $\Lambda_{ij} = -1/k_i^{in}$ if i eats j or 0 otherwise, and diagonal entries $\Lambda_{ii} = 1$.

The only necessary condition for the system of linear equations defined in 4.11 to be solvable is that the matrix Λ has to be invertible. This requires at least one basal species (else zero would be an eigenvalue of Λ). However, note that cycles are not, in general, a problem, despite the apparent recursivity of Eq 4.10.

Quantifiying the current parameter χ: Once the trophic level of the nodes have been assigned the current parameters is determined as the portion of links that are aligned with the inherent directionality of the network (direction of increasing hierarchy), as in fig.4.8. In a case of perfect directionality all links should be aligned with the direction defined by the hierarchy, and hence $\chi = 1$. In the absence of a

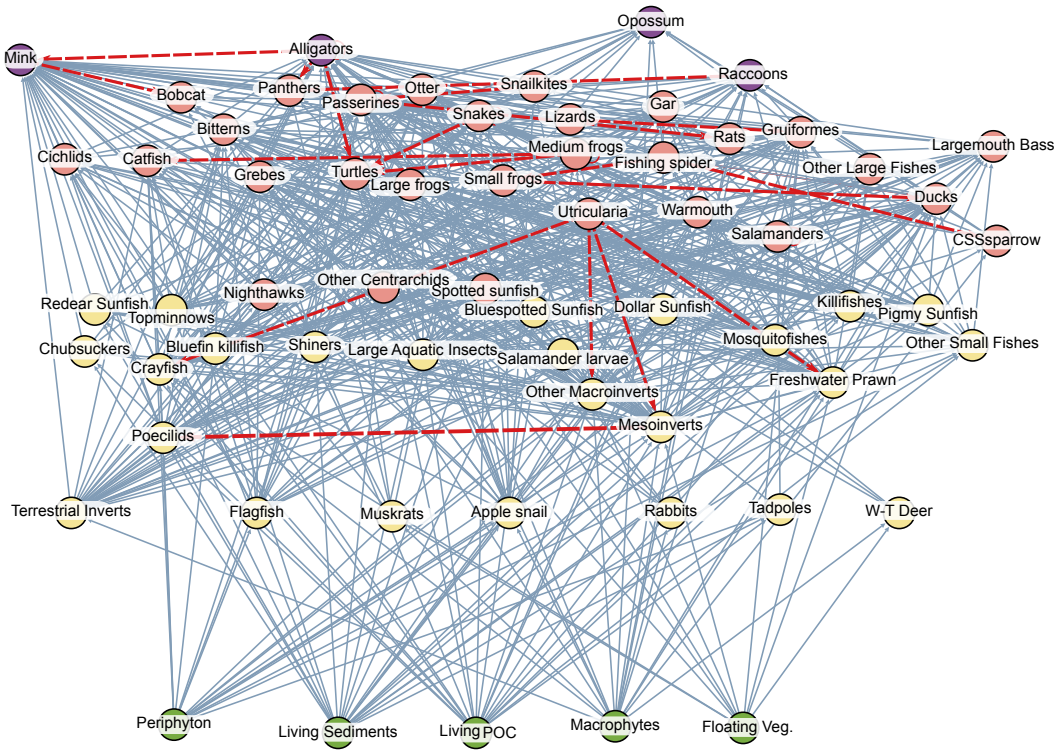

Figure 4.8: Empirical representation of a trophic web of a Everglades Graminoid Marshes ecosystem [256] in which the trophic hierarchy as been made apparent by application of our algorithm. The height of the nodes represents the trophic level of each species (labeled with its name), and the different colours are only a guide to the eye. Links spanning negative trophic distances (pointing from higher to lower hierarchies) are represented in red, while links aligned with the inherent directionality are in gray. Determination of the current parameter is straightforward once the levels have been assigned. In this case from the 617 links in the community only 20 (hihglighted in red) oppose the general directionality, and hence the current parameter of this network is $\chi = 0.97$. Curiously enough one of the species involved in many of the "against-directional" interactions is the *Utricularia*, a carnivorous plant.

well-defined directionality, the direction of the links would be random and one expects $\chi \approx 1/2$ (apart from small deviations due to finite-size effects). In general ecological and biological networks exhibit a rather large directionality, with values of $\chi > 0.8$ in most of cases (see table 4.2 at the end of this chapter).

4.4 Chapter summary

- In this chapter we have tackled the problem of exhaustively counting the number of structural and feedback loops in a variety of biological, ecological, and socio-technological networks. We then compared these numbers with those in randomized versions of the same graphs, where other basic structural features (such as total number of nodes, number of links, connectivity of each link, etc.) were preserved. In general, ecological and biological networks show a similar trend, an under-representation of long loops compared to the random expectations (Fig.4.3 and 4.6). This effect is much milder in socio-technological networks.

- We hypothesize that the (empirically observed) lack of feedback loops stems from the existence of an inherent directionality. To investigate this conjecture, we have constructed a simple computational model in which an inherent network directionality –quantified by a directionality parameter γ– is built in. For this model we are able to analytically compute the fraction of feedback loops of any given length as a function of γ. We find that this intrinsically directional model can reproduce quite well empirical curves of the fraction of feedback loops of any length by just tuning its only parameter γ (Fig.4.4).

- Using the proposed model it is possible to infer a directionality in empirical networks. The inferred directionality is the value γ which gives a better fit of the curves of fraction of feedback loops.

- We put forward a method to directly measure the directionality in empiric networks. Indeed, by employing a method inspired on how trophic levels are identified in foodwebs, we have been able to identify –just by looking at the network structure– the current parameter χ as the fraction of links in the direction defined by the hierarchy (Fig.4.8).

- The optimal value of the directionality parameter γ –derived from the statistics of loops– correlates quite well with the current parameter, χ, computed by quantifying the network degree of directionality. These two measures of network inherent directionality are quantitatively different but they are strongly correlated (Fig.4.7). In this way the existence of an inherent directionality constitutes

a simple yet satisfactory parsimonious explanation for the empirically observed lack of feedback loops in biological and ecological networks.

4.5 Data supplement to chapter 3

Networks data

We have considered a number of different directed networks from the literature, excluding loop-less ones. Let us remark that real networks are more flexible and dynamical than what is captured in existing databases. Current data can be considered as a snapshot or time-average of such dynamical networks. Nevertheless, this should not constitute a particular limitation for the analyses performed here. The list of employed networks is:

- *Foodweb:* N.E. Shelf [138] ($N = 79, L = 1,403$), Caribbean Reef [186] ($N = 50, L = 556$), Coachella Valley [203] ($N = 29, L = 262$), El Verde Rainforest [265] ($N = 155, L = 1509$), Little Rock Lake [144] ($N = 92, L = 997$), Lough Hyne [76, 211] ($N = 349, L = 5117$), and Weddell Sea [108] ($N = 483, L = 15362$).

- *Biological networks*: *C. elegans* Neural Connectivity Network ($N = 297, L = 2345$) [268], *E. coli* Transcriptional Regulatory Network (TRN) ($N = 1037, L = 2687$) [220], Mammalian Cell Signaling Network ($N = 599, L = 1399$) [140], and Yeast Transcriptional Regulatory Network (TRN) ($N = 4440, L = 12873$) [276]

- *Socio-technological networks*: F.A.A. Flights ($N = 1226, L = 2616$) [140], Digg News Forum Replies ($N = 30398, L = 87627$) [48], Twitter Followings ($N = 81306, L = 1768149$)[154], Political Blogs in the U.S.A. ($N = 1490, L = 19025$) [2], Advogato Forum Replies ($N = 6551, L = 51332$) [148], Kaitiaki Friendship ($N = 62, L = 119$) [250], Wikipedia Elections ($N = 7115, L = 103689$) [134], P2P Connections ($N = 6301, L = 20777$) [212], and

- *Other networks (with few loops)*:

  - *Foodwebs*: Everglades Graminoid Marshes ($N = 69, L = 916$) [256], Mangrove Estuary ($N = 97, L = 1,491$) [255], Mondego Estuary - Zostrea site ($N = 46, L = 400$) [**?**], and Skipwith Pond [267] ($N = 25, L = 197$).
  - *Biological networks*: Human TRN ($N = 3088, L = 6886$) [276] and Mouse TRN ($N = 72, L = 117$) [276].

– *Social networks*: Ownership ($N = 7252, L = 6726$) [179].

Other biological networks such as, for example, protein-protein networks –in which two proteins are associated if they co-participate in some metabolic process– are undirected and, thus, not covered by our study.

Correlation between γ and χ

Network	r^2	γ	χ
E. coli TRN	0.995	0.999	0.9316
Cell Signaling	0.973	0.888	0.7348
Yeast TRN	0.997	1.000	0.9887
Neural Connectivity	1.000	0.879	0.7429
Coachella Valley	0.984	0.988	0.7325
Caribbean Reef	0.997	0.958	0.8579
El Verde Rainforest	0.997	0.961	0.8381
N.E. Shelf	0.997	1.000	0.9470
Little Rock Lake	0.999	0.998	0.9350
Lough Hyne	0.980	0.999	0.9616
Weddell Sea	0.992	0.987	0.9072
F.A.A.	0.998	0.783	0.6183
Forum Replies	1.000	0.935	0.7359
Twitter Followings	0.984	0.967	-
Political Blogosphere	1.000	0.744	0.5217
Advocato Forum	0.999	0.714	0.5487
Kaitiaki friendship	0.975	0.860	0.8411
Wikipedia elections	0.998	0.934	0.7252
P2P Connections	0.991	0.993	0.8205
Everglades	0.979	0.999	0.9673
Mangrove Estuary	0.998	0.999	0.9704
Mondego Estuary	1.000	0.992	0.9373
Skipwith	0.000	0.997	0.9471
Human TRN	0.984	0.999	0.9626
Mouse TRN	1.000	0.970	0.8957

Ownership	1.000	0.977	0.9880
Tuberculosis TRN	1.000	0.998	0.9858
B. subtilis TRN	1.000	0.985	0.9459

Table 4.2: **Quantification of network directionality.** First and second columns: values of the linear correlation coefficient r^2 and of the fitted parameter γ, respectively, for the linear fit of $\log F(k)$ versus k with Eq. (4.2) for the considered networks. Third column: measures of the current parameter χ from the network structure (large values of χ indicate high levels of hierarchy and thus of directionality). Networks below the central double line are those with only a small number of short loops, i.e. not having any loop larger than $k = 6$. In the case of the Skipwith network, the value of r^2 is absent as we could not compute long enough loops to observe the exponential decay. In the Twitter followings network the value of χ could not be computed due to computational limitations.

Chapter 5

Factors determining nestedness in complex networks

Networks have become a paradigm for understanding systems of interacting objects, providing us with a unifying framework for the study of diverse phenomena and fields, from molecular biology to social sciences [24]. Most real networks are not assembled randomly but present a number of non-trivial structural traits such as the small-world property, scale freeness, hierarchical organization, etc [5, 178]. Network topological features are essential to determine properties of complex systems such as their robustness, resilience to attacks, dynamical behavior, spreading of information, etc. [25, 40, 178]. A paradigmatic case is that of ecosystems, in which species can be visualized as nodes of a network and their mutual interactions (predation, mutualism, facilitation, etc) encoded in the edges or links. One such feature of ecological networks, which has been studied for some time by ecologists, is called *nestedness* [19]. Loosely speaking, a bipartite network [178] –say, for argument's sake, of species and islands, linked whenever the former inhabits the latter– is said to be nested if the species that exist on a few islands tend always to be found also on those islands inhabited by many different species. This can be most easily seen by graphically representing a matrix such that species are columns and islands are rows, with elements equal to one whenever two nodes are linked and zero if not. If, after ordering all nodes by degree (number of neighbours), most of them can be quite neatly packed into one corner, the network is considered highly nested [19, 272]. Even if initially introduced for bipartite networks, the concept of nestedness

can be readily generalized for generic networks. The idea of nestedness is illustrated in Fig.5.1 where we plot different connectivity matrices with different levels of maximal "compactability" and, thus, with different levels of nestedness.

Nestedness is usually measured with purposely-designed software. The most popular *nestedness calculator* is the "temperature" of Atmar and Patterson (used to extract a temperature from the matrices in Fig.5.1) [19]. It estimates a curve of equal density of ones and zeros, calculates how many ones and zeros are on the "wrong" side and by how much, and returns a number between 0 and 100 called "temperature" by analogy with some system such as a subliming solid. A low temperature indicates high nestedness. It is important to caution that nestedness indices should not be used as black-boxes, as this can lead to false conclusions [83, 257]. The main drawback of these calculators is that they are defined by complicated algorithms, hindering further analytical developments.

In a seminal work, Bascompte and collaborators [29] showed that real *mutualistic networks* (i.e. bipartite networks of symbiotic interactions), such as the bipartite network of plants and the insects that pollinate them, are significantly nested. They also defined a measure to quantify the average number of shared partners in these mutualistic networks, and called it "nestedness" because of its close relation with the concept described above. They go on to show evidence of how the so-defined nestedness of empirical mutualistic networks is correlated with the biodiversity of the corresponding ecosystems [34]: the global species competition is significantly reduced by developing a nested network architecture and this entails a larger biodiversity. The principle behind this is simple. Say nodes A and B are in competition with each other. An increase in A will be to B's detriment and vice-versa; but if both A and B engage in a symbiotic relationship with node C, then A's thriving will stimulate C, which in turn will be helpful to B. Thus, the effective competition between A and B is reduced, and the whole system becomes more stable and capable of sustaining more nodes and more individuals. The beneficial effect that "competing" nodes (i.e. those in the same side of a bipartite network) can gain from sharing "friendly" partners (nodes in the other side) is not confined to ecosystems. It is expected also to play a role, for instance, in financial networks or other economic systems [240]. To what extent the measure introduced by Bascompte et al. is related to the traditional concept of nestedness has not, to the best of our knowledge, been rigorously explored. Irrespectively of this relation, however, the

insight that mutual neighbours can reduce effective competition in a variety of settings is clearly interesting in its own right, and it is for this reason that we analyze this feature here. On a different front, Staniczenko et al.[234] have made some promising analytical progress regarding the traditional concept of nestedness.

Here, we take up this idea of shared neighbours, though characterized, owing to reasons we shall explain in the Methods, with a slightly different measure (see 5.1.1) and study analytically and computationally how it is influenced by the most relevant topological properties, such as the degree distribution and degree-degree correlations. Our aim is to understand to what extent nestedness is a property inherited from imposing a given degree distribution or a certain type of degree-degree correlations.

5.1 Analytical quantification of nestedness

In this section we propose a new nestedness index that is inspired by the one introduced in ref.[34] but with some additional advances. Our main aim is to be able to easily discern the contribution of the degree-distribution to this measure, and hence obtain a estimation of the nestedness of the networks beyond its heterogeneity. In order to do so we will use the ensemble of the randomized networks provided by the (uncorrelated) configuration ensemble.

5.1.1 introducing a refined measure

Consider an arbitrary network with N nodes defined by the adjacency matrix A: the element a_{ij} is equal to the number of links (or edges) from node j to node i (typically considered to be either 1 or 0 though extensions to weighted networks have also been considered in the literature [234]). If A is symmetric, then the network is undirected and each node i can be characterized by a degree $k_i = \sum_j a_{ij}$. If it is directed, i has both an *in* degree, $k_i^{in} = \sum_j a_{ij}$, and an *out* degree, $k_i^{out} = \sum_j a_{ji}$; we shall focus here on undirected networks, although most of the results could be easily extended to directed ones.

Bastolla *et al.* [34] have shown that the effective competition between two species can be reduced if they have common neighbours with which they are in symbiosis. Therefore, in mutualistic networks it is beneficial for the species at two nodes i and

j if the number of shared symbiotic partners, $\hat{n}_{ij} = \sum_l a_{il}a_{lj} = (a^2)_{ij}$, is as large as possible. Going on this, and assuming the network is undirected, the authors propose to use the following measure:

$$\eta_B = \frac{\sum_{i<j} \hat{n}_{ij}}{\sum_{i<j} \min(k_i, k_j)}, \tag{5.1}$$

which they call *nestedness* because it would seem to be highly correlated with the measures returned by nestedness software. Note that, although the authors consider only bipartite graphs, such a feature is not imposed in the above definition.

Here, we take up the idea of the importance of having an analytical expression for the nestedness but, for several reasons, we use a definition slightly different from the one in [34]. Actually, η_B suffers from a serious shortcoming; if one commutes the sums in the numerator of Eq.(5.1), it is found that the result only depends on the heterogeneity of the degree distribution: $\sum_{ij} \hat{n}_{ij} = \sum_l \sum_i a_{il} \sum_j a_{lj} = N\langle k^2 \rangle$ (in an undirected network, $\sum_{i<j} = \frac{1}{2}\sum_{ij}$; we shall always sum over all i and j, since it is easier to generalize to directed networks and often avoids writing factors 2). Therefore, this index essentially provides a measurement of network heterogeneity. Also, although the maximum value $\hat{n}_{ij}$ can take is $\min(k_i, k_j)$, this is not necessarily the best normalization factor, since (as we show explicitly in the next Section) the randomly expected number of paths of length 2 connecting nodes i and j depends on both k_i and k_j. Furthermore, it can sometimes be convenient to have a local measure of nestedness (i.e. nestedness of any given node) which cannot be inferred from the expresion above. For all these reasons, we propose to use

$$\tilde{\eta}_{ij} \equiv \frac{\hat{n}_{ij}}{k_i k_j} = \frac{(a^2)_{ij}}{k_i k_j}, \tag{5.2}$$

which is defined for every pair of nodes (i, j). This allows for the consideration of a nestedness per node, $\tilde{\eta}_i = N^{-1} \sum_j \tilde{\eta_{ij}}$, or of the global measure

$$\tilde{\eta} = \frac{1}{N^2} \sum_{ij} \tilde{\eta_{ij}} \tag{5.3}$$

which is very similar in spirit to the measure introduced by Bastolla *et al.* in [34] but, as argued above, has a number of additional advantages.

This new index can be easily applied to bipartite networks. In this case the communities are represented by a bipartite graph, with two different set of nodes. The ones considered in Ref. [34], for instance, are composed of animals and plants which interact in symbiotic relations of feeding-pollination; these interactions only take place between animals and plants. Let us therefore consider a bipartite network and call the sets Γ_1 and Γ_2, with n_1 and n_2 nodes, respectively ($n_1 + n_2 = N$). Using the notation $\langle\cdot\rangle_i$ for averages over set Γ_i, the total number of edges is $\langle k\rangle_1 n_2 = \langle k\rangle_2 n_1 = \frac{1}{2}\langle k\rangle N$. Assuming that the network is defined by the configuration ensemble, though with the additional constraint of being bipartite, the probability of node l being connected to node i is

$$\hat{\epsilon}_{il} = 2\frac{k_i k_l}{\langle k\rangle N}$$

if they belong to different sets, and zero if they are in the same one. Proceeding as before, we find that the expected value of the nestedness for a bipartite network is

$$\eta_{bip} = \frac{1}{N^2}\left[\sum_{i,j\in\Gamma_1}\frac{1}{k_i k_j}\sum_{l\in\Gamma_2}\frac{k_i k_l}{\langle k\rangle_1 n_2}\frac{k_l k_j}{\langle k\rangle_2 n_1} + \sum_{i,j\in\Gamma_2}\frac{1}{k_i k_j}\sum_{l\in\Gamma_1}\frac{k_i k_l}{\langle k\rangle_1 n_2}\frac{k_l k_j}{\langle k\rangle_2 n_1}\right] = \frac{n_1\langle k^2\rangle_2 + n_2\langle k^2\rangle_1}{\langle k\rangle_1\langle k\rangle_2(n_1+n_2)^2}. \tag{5.4}$$

Interestingly, if $n_1 = n_2$, the fact that the network is bipartite has no effect on the nestedness: $\eta_{bip} = \eta_{conf}$.

Having an analytical definition of nestedness, it becomes feasible to scrutinize how it is influenced by the most basic structural features, such as the degree distribution and degree-degree correlations. The standard procedure to determine how significantly nested a given network is, is to generate randomizations of it (while keeping fixed some properties such as the total number of nodes, links, or degree distribution) and compare the nestedness of the initial network with the ensemble-averaged one. The set of features kept fixed in randomizations determine the *null-model* used as reference.

5.1.2 Nestedness in the configuration model

Many networks exhibit quite broad degree distributions $P(k)$, in many cases close to the fairly ubiquitous scale-free pattern $P(k) \sim k^{-\gamma}$ [5]. Since heterogeneity tends to

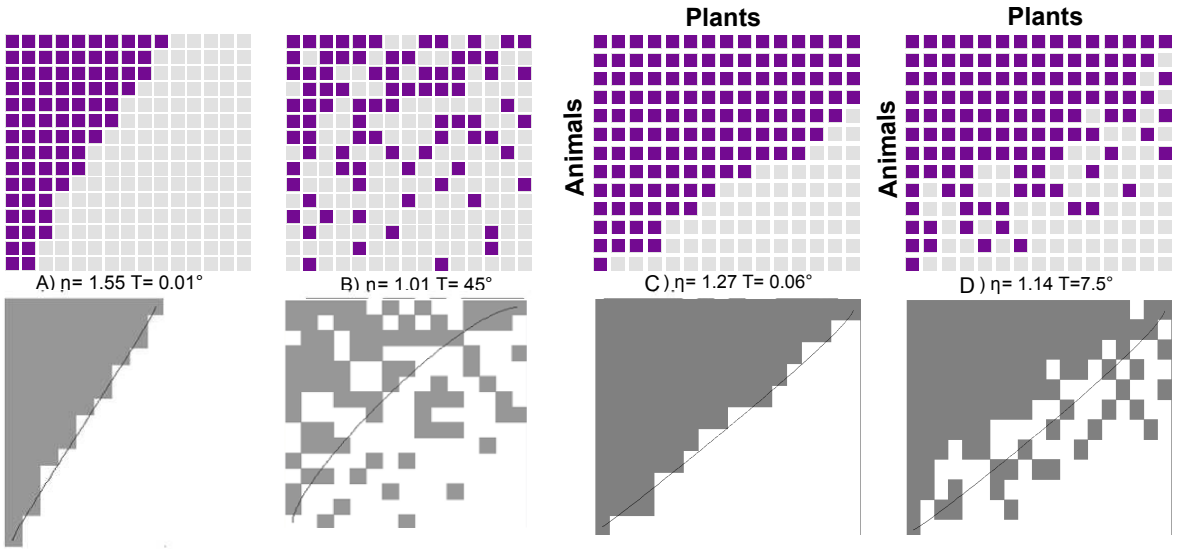

Figure 5.1: Measures of nestedness in networks. The figure shows different connectivity matrices with different levels of nestedness as measured by (i) our new nestedness index [Eq.(5.7)] and (ii) the standard nestedness "temperature' calculator". Panels A and B shows two different unimodal networks. As can be readily seen, the most packed matrix corresponds to a very low temperature and to a high nestedness index ($\eta > 1$) and, reciprocally, the least packed one exhibits a high temperature and an index close to its expected value for a random network ($\eta \simeq 1$). Panels C and D shows two bipartite networks, and again the most packed matrix has the higher value of nestedness ($\eta > 1$) while the least packed has higher temperature and a η closer to 1.

have a significant influence on any network measure, it is important to analytically quantify the influence of degree-distributions on nestedness. For any particular degree sequence, the most natural choice is to use the *configuration model* [165, 178] –defined as the ensemble of random networks wired according to the constraints that a given degree sequence $(k_1, ..., k_N)$ is respected– as a *null model*. In such an ensemble, the averaged value of any element of the adjacency matrix is

$$\overline{a_{ij}} \equiv \hat{\epsilon}^c_{ij} = \frac{k_i k_j}{\langle k \rangle N}. \tag{5.5}$$

We use an overline, $\overline{(\cdot)}$, to represent ensemble averages and angles, $\langle \cdot \rangle$, for averages over nodes of a given network.

Plugging Eq. (5.5) into Eq. (5.2), we obtain the expected value of $\tilde{\eta}$ in the configuration ensemble, which is our basic null model

$$\overline{\tilde{\eta}_{ij}} = \frac{\langle k^2 \rangle}{\langle k \rangle^2 N} \equiv \tilde{\eta}_{conf}. \tag{5.6}$$

It is important to underline that $\overline{\tilde{\eta}_{i,j}}$ is independent of i and j; hence, it coincides with the expected value for the global measure, $\overline{\tilde{\eta}} = \overline{\tilde{\eta}_{i,j}}$ (which justifies the normalization chosen in Eq. (5.2)). Also, it is noteworthy that for degree distributions with finite first and second moments, $\tilde{\eta}_{conf}$ goes to zero as the large-N limit is approached.

It is obvious from Eq. (5.6) that degree heterogeneity has an important effect on $\tilde{\eta}$; for instance, scale-free networks (with a large degree variance) are much more nested than homogeneous ones. Therefore, if we are to capture aspects of network structure other than those directly induced by the degree distribution it will be useful to consider the nestedness index normalized to this expected value,

$$\eta \equiv \frac{\tilde{\eta}}{\tilde{\eta}_{conf}} = \frac{\langle k \rangle^2}{\langle k^2 \rangle N} \sum_{ij} \frac{(a^2)_{ij}}{k_i k_j}. \tag{5.7}$$

Although η is unbounded, it has the advantage that it is equal to unity for any uncorrelated random network, independently of its degree heterogeneity, thereby making it possible to detect additional non-trivial structure in a given empirical network.

Degree-degree correlations in the configuration model

In the configuration ensemble, the expected value of the mean degree of the nearest neighbours (nn) of a given node is $\overline{k_{nn,i}} = k_i^{-1} \sum_j \hat{\epsilon}^c_{ij} k_j = \langle k^2 \rangle / \langle k \rangle$, which is independent of k_i. Still, specific finite-size networks constructed with the configuration model can deviate from the ensemble average results (which hold exactly only in the $N \to \infty$ limit). Real networks are finite, and they often display degree-degree correlations, which result in $\overline{k_{nn,i}} = \overline{k_{nn}}(k_i)$. If $\overline{k_{nn}}(k)$ increases (decreases) with k, the network is said to be assortative (disassortative), i.e. nodes with large degree tend to be connected with other nodes of large (small) degree.

The measure usually employed to measure this phenomenon is Pearson's coefficient applied to the edges [40, 177, 178]: $r = ([k_l k'_l] - [k_l]^2)/([k_l^2] - [k_l]^2)$, where k_l and k'_l are

the degrees of each of the two nodes belonging to edge l, and $[\cdot] \equiv (\langle k \rangle N)^{-1} \sum_l (\cdot)$ is an average over edges. Writing $\sum_l (\cdot) = \sum_{ij} a_{ij}(\cdot)$, r can be expressed as [177]

$$r = \frac{\langle k \rangle \langle k^2 \overline{k_{nn}}(k) \rangle - \langle k^2 \rangle^2}{\langle k \rangle \langle k^3 \rangle - \langle k^2 \rangle^2}. \tag{5.8}$$

In the infinite network-size limit we expect $r = 0$ in the configuration model (null model) as there are no built in correlations. However even if the index r is widely used to measure network correlations, some drawbacks of it have been put forward [69, 274]

5.2 Nestedness in finite-size random networks

The use of the configuration model allows us to determine to what extent some features of networks are influenced by the degree-distribution. In the previous section we study the effect the heterogeneity in the degree-distribution has in nestedness measures, and devise a measure of nestedness that takes it into account. However, the predictions of the configuration model ($r = 0$ and $\eta = 1$) only hold exactly in the infinite size ($N \to \infty$) limit. Now, we go on an further study the nestedness and degree correlations in uncorrelated networks with finite-size. In orther to do so we will use again the configuration model to generate finite networks, and study its correlations and nestedness, to determine whether the predictions of the infinite case hold or, on the other hand, the finite size effect have a significant contribution.

5.2.1 Emergence of effective correlations in finite-size networks

We have computationally constructed finite random networks with different degree distributions; in particular, Poissonian, Gaussian, and scale-free distributions, assembled using the configuration model as explained above (for the scale-free case see Ref.[46]) and measured their Pearson's correlation coefficient. Results are illustrated in Fig.5.2; the probability of obtaining negative (disassortative) values of r is larger than the one for positive (assortative) values (observe the shift between $r = 0$ and the curve averaged value). This means that the null-model expectation value of r is negative! i.e. finite random networks are more likely to be disassortative than assortative. This result is highly counterintuitive because the ensemble is constructed without assuming any type

of correlations and is, clearly, a finite-size effect. Indeed, for larger network sizes the averaged value of r converges to 0 as we have analytically proved and computationally verified. For instance, for scale-free networks with $P(k) \propto k^{-\gamma}$ with $2 < \gamma < 3$, r can be easily shown to converge to 0 as $r \propto N^{-1/3}$ in the large-N limit, as we detail below.

In this particular case the maximum expected degree, K_{max} in a network of size N is of the order $N^{\frac{1}{\gamma-1}}$ and this cut-off controls the scaling of the moments as

$$< k^m > \sim \int_1^{K_{max}} k^m k^{-\gamma} dk \sim k^{m-\gamma+1}|_1^{N^{\frac{1}{\gamma-1}}} \sim N^{\frac{m-\gamma+1}{\gamma-1}} - 1.$$

Combining the expressions for the first three moments appearing in the definition of r in Eq.E(5.8), one readily obtains;

$$r = \frac{aN^{\frac{3-\gamma}{\gamma-1}} - bN^{2\frac{3-\gamma}{\gamma-1}}}{cN^{\frac{4-\gamma}{\gamma-1}} - dN^{2\frac{3-\gamma}{\gamma-1}}} \sim -eN^{\frac{2-\gamma}{\gamma-1}} \tag{5.9}$$

where a, b, c, d, and e are un-specified positive constants. In the case of a scale-free network with $\gamma = 2.5$ this reduces to $r \sim N^{-\frac{1}{3}}$ in agreement with numerical results shown in Fig.5.2.B (observe that, as we use a logarithmic scale, the absolute value of r rather than r itself is employed).

A well-known effect leading to effective disassortativity, is that simple algorithms, which are supposed to generate uncorrelated networks, can instead lead to degree-degree anti-correlations when the desired degree distribution has a heavy tail and no more than one link is allowed between any two vertices (as hubs are not as connected among themselves as they should be without such a constraint) [146, 191]. Also, our observation is in agreement with the recent claim that, owing to entropic effects, real scale-free networks are typically disassortative: simply, there are many more ways to wire networks with disassortative correlations than with assortative ones [115]

5.2.2 Effective correlations imply nestedness in finite networks

A straightforward consequence of the natural tendency of finite networks to be disassortative is that they thereby also become naturally nested. Indeed, the nestedness index η was defined assuming there were no built-in correlations, but if degree-degree

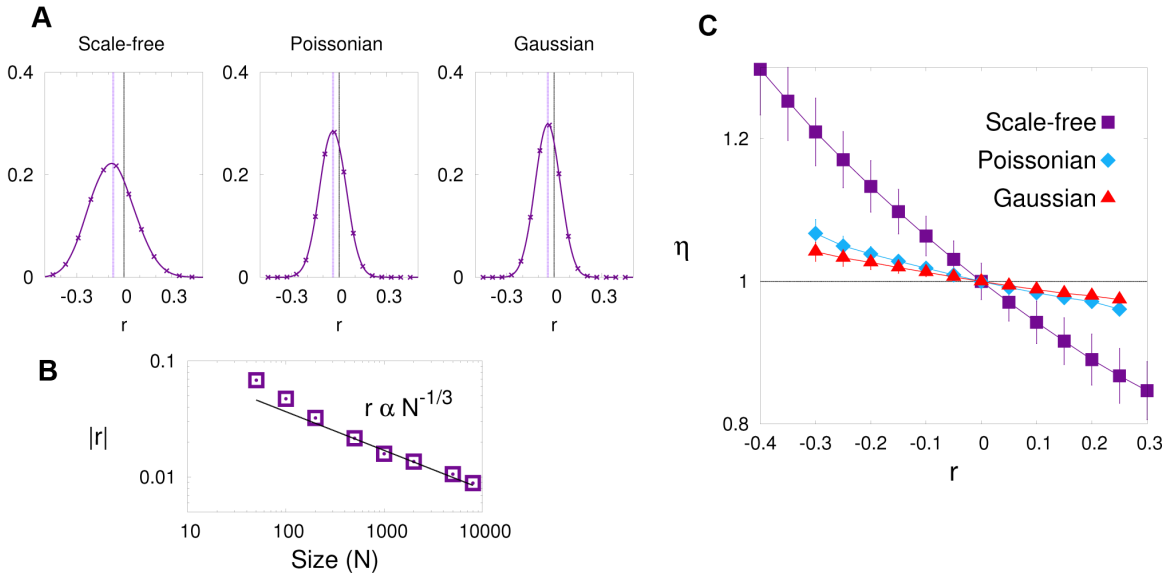

Figure 5.2: Correlation coefficient and nestedness in random networks. (Panel A): Correlation coefficient, r for 10^6 networks generated independently using the configuration model with $N = 50$ nodes, $< k >= 5$ and (from left to right) scale-free (with exponent $\gamma = 2.25$), Poissonian, and Gaussian ($\sigma^2 = 10$) degree distributions. (Panel B): Pearson's correlation coefficient as a function of network size for scale free networks with $\gamma = 2.25$. (Panel C): Averaged nestedness (with error bars corresponding to one standard deviation) as a function of Pearson's correlation index r in random (scale-free, Poissonian, and Gaussian) networks (as in the left panel). These curves are obtained employing the Wang-Landau algorithm as described in sec.5.3.4. All three curves show a positive (almost linear) correlation between disassortativity and nestedness: more disassortative networks are more nested. By restricting the corresponding configuration ensembles to their corresponding subsets in which r is kept fixed it is possible to define a more constraint null model as discussed in the main text.

correlations effectively emerge in finite-size random networks, then deviations from the neutral value $\eta = 1$ are to be expected.

Indeed, in Fig.5.2.C we have considered networks constructed with the configuration model, employing the same probability distributions (Gaussian, Poissonian and scale free) as above. For each so-constructed random network we compute both r and η and plot the average of the second as a function of the first (technical details on how to sample networks with extreme values of r –using the Wang-Landau algorithm [266]– are

given in sec.5.3.4). The resulting three curves exhibit a neat (almost linear) dependence of the expected value of η on r: disassortative networks are nested while assortative ones are anti-nested. As disassortative ones are more likely to appear, a certain degree of nestedness is to be expected in finite random networks. Observe, however, that for truly uncorrelated random networks, i.e. with $r = 0$, the expectation value of η is 1.

Finally, in Appendix E, we provide an analytical connection between disassortativity and nestedness in infinite random networks with explicitly built-in degree-degree correlations. Also in this case a clear relation between nestedness and disassortativity emerges (as shown in figure E.1 in the aforesaid apprendix) for scale-free networks.

5.3 Nestedness in empiric networks

In the previous section we have seen that finite-size random networks do exhibit some degree of nestedness, as expected from the existent effective degree-degree correlations. The next logical step is to look at empirical networks and wonder whether their values of nestedness and degree-degree correlations can be explained by a null model or not.

To study this we have considered 60 different empirical networks, both bipartite and unimodal, from the literature. The set includes foodwebs, metabolic, neuronal, ecological, social, and technological networks (for the whole dataset see sec. 5.5 at the end of this chapter). We have performed randomizations preserving the corresponding degree sequences (configuration ensemble) and avoiding multiple links between any pair of nodes. Results for a subset of 16 networks are illustrated in Figure 5.3, which shows the distribution of r and η values (see figure caption) compared with the actual values of r_{emp} and η_{emp} in the empirical networks.

5.3.1 Degree correlations in real vs randomized networks

The actual value of r in empirical networks coincides with the ensemble average within an error of the order of 1, 2, or 3 standard deviations in about two thirds of the cases (53%, 67%, and 76% respectively (see fig.5.4)). Similarly, the corresponding p-values are larger than the significance threshold (0.05) in 60% of the cases. Particularizing for bipartite networks, the z-scores rise to: 60%, 76%, and 89%, respectively, and the significant P-values go up to 68% (data are collected in Appendix 5.5).

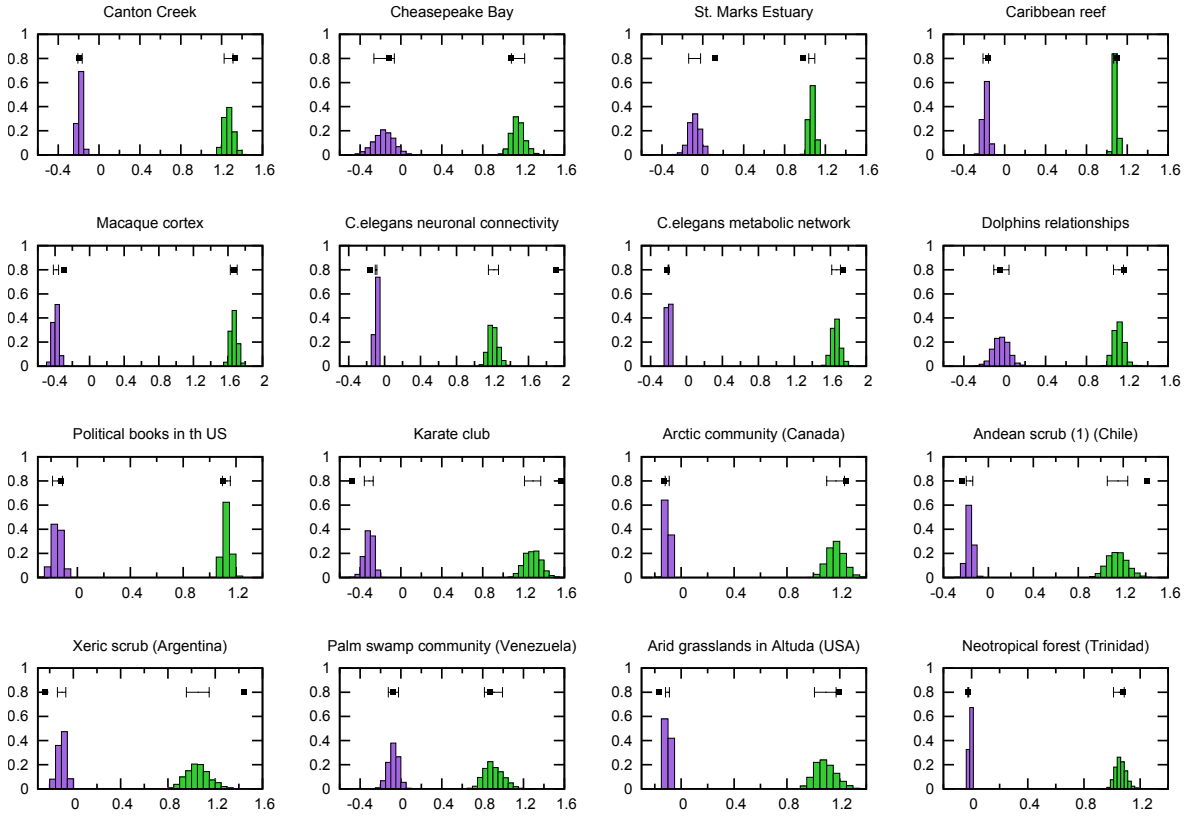

Figure 5.3: Correlation coefficient and nestedness in degree-preserving randomiaztions. Probability distribution of Pearson's coefficient r (red) and of the nestedness coefficient η (green) as measured in degree-preserving randomizations of a subset of 16 (out of a total of 60) real empirical networks (as described and referenced in Appendix 5.5). The actual empirical values in the real network are marked with a black box and compared (also in black) with a segment centered at the mean value of the random ensemble (configuration model) with width equal to one standard deviation. In most cases but not all, the empirical values lie in or near the corresponding interval, suggesting that typically empirical networks are not significantly more assortative/nested than randomly expected.

Therefore, roughly speaking, the null model –in which networks are randomly wired according to a specified degree sequence– explains well the correlations of about two-thirds (or more) of the networks we have analyzed and, more remarkably, it explains even better the correlations of bipartite networks. Thus, once it has been realized that random networks have a slight natural tendency to be disassortative, in many

cases, there does not seem to be a clear generic statistical tendency for real networks to be more correlated (either assortatively or disassortatively) than expected in the null model. For instance in almost all foodwebs we have analyzed the empirical value of r is well explained by randomizations, while in some other social and biological networks there are some residual positive correlations (assortativity).

5.3.2 Nestedness in real vs randomized networks

We have conducted a similar analysis for the nestedness index η and compare its value in real networks with the expected value in randomizations (see Fig.5.3). In this case, the actual value of η in empirical networks coincides with the ensemble average with an error of the order of 1, 2, or 3 standard deviations also in about two thirds of the cases (43%, 73%, and 83% respectively). As for the p-value, it is above threshold in 63% of the cases (which goes up to 76% for bipartite networks). Thus, in most of the analysed examples, empirically observed values of nestedness are in agreement with null-model expectations once the degree-distribution has been taken into consideration (data shown in Appendix 5.5).

All these results can be summarized in fig. 5.4 below, where the number of empirical networks with a z-score of 1,2,3 or more are represented.

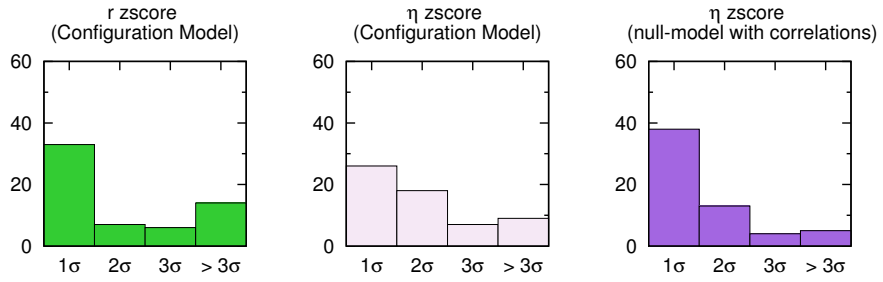

Figure 5.4: A: Number of empirical networks which value of r lies within $1, 2, 3$, or more standard deviations from the expected value in the randomized ensemble. B: Number of networks which value of η as measured by Eq.(5.7) lies within $1, 2, 3$, or more standard deviations from the expected value in the randomized ensemble. C: Same analysis but for the randomizations in the second null model, as defined in secc. 5.3.4. The use of a more refined null-model

5.3.3 Nestedness vs degree correlations in empirical networks

As said above, both Fig.5.2.C and Fig.5.3 reveal a global tendency: exceedingly disassortative networks tend to be nested while assortative ones are anti-nested. To further explore this relation, Fig.5.5 shows a plot of nestedness against assortativity for the selection of empirical networks listed in Appendix 5.5. Although these networks are highly disparate as regards size, density, degree distribution, etc., it is apparent that the main contribution to η comes indeed from degree-degree correlations. *The observation of such a strong generic correlation between the nestedness and disassortativity constitutes one of the main messages of this chapter.*

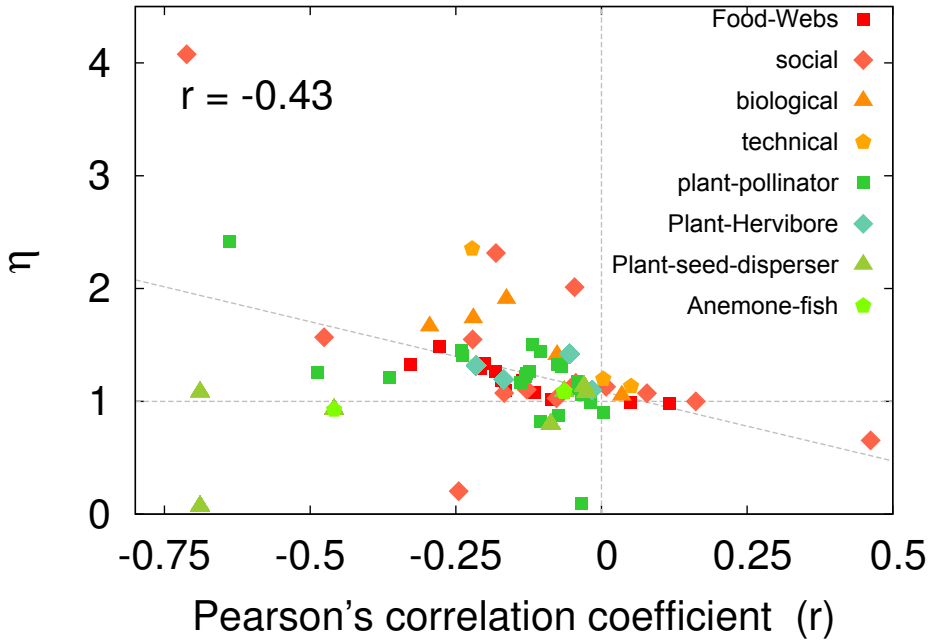

Figure 5.5: Nestedness against assortativity (as measured by Pearson's correlation coefficient) for data on a variety of networks. Warm-coloured items correspond to unimodal networks and green ones to bipartite networks of different kinds (see Appendix 5.5).

5.3.4 A more refined null model

A unique criterion for choosing a proper null model does not exist [90]. For instance, it is possible to go beyond the null model studied so far by preserving not just the degree sequence but also empirical correlations. Indeed, from the full set of networks generated with the configuration model for a given degree sequence, one could consider the subset of networks with a fixed value of r, as done in Fig.5.2.C (and as explained below). In particular, one could take the sub-ensemble with the same r as empirically observed. This constitutes a more refined null model in which the number of nodes, degree sequence, and degree-degree correlations are preserved. This more refined null model reproduces slightly better than the configuration model the empirical values of nestedness: 63% of networks fall within one standard deviation (as opposed to 42% in the first null model), 85% within two (versus 72%), and 92% for three standard deviations (as opposed to 82%). Interestingly, allowing for three standard deviations bipartite networks are explained in a 100% of the cases (see Table 5.2 at the end of this chapter). Thus, the null model preserving degree-degree correlations explains quite well the observed levels of nestedness.

Sampling networks with a given value of r

Given that networks with very large r (in absolute value) are rare, and thus they seldom appear in the randomization process used to built the configurational ensemble (or null model) we have implemented the Wang-Landau (multi-canonical) algorithm to enrich the sampling with such rare networks [266]. The gist of this technique is to perform a "random" walk in the r-space, in such a way that jumps toward frequently visited r-values are penalized and, instead, rarely visited r's are favoured, which requires storing the statistics of the number of times every value of r has been previously "extracted". Starting from an initial network (with r_1), a small change in its topology is tentatively made, and the resulting new network (with r_2) is accepted with probability

$$P(r_1 \to r_2) = min\left[\frac{g(r_1)}{g(r_2)}, 1\right] \tag{5.10}$$

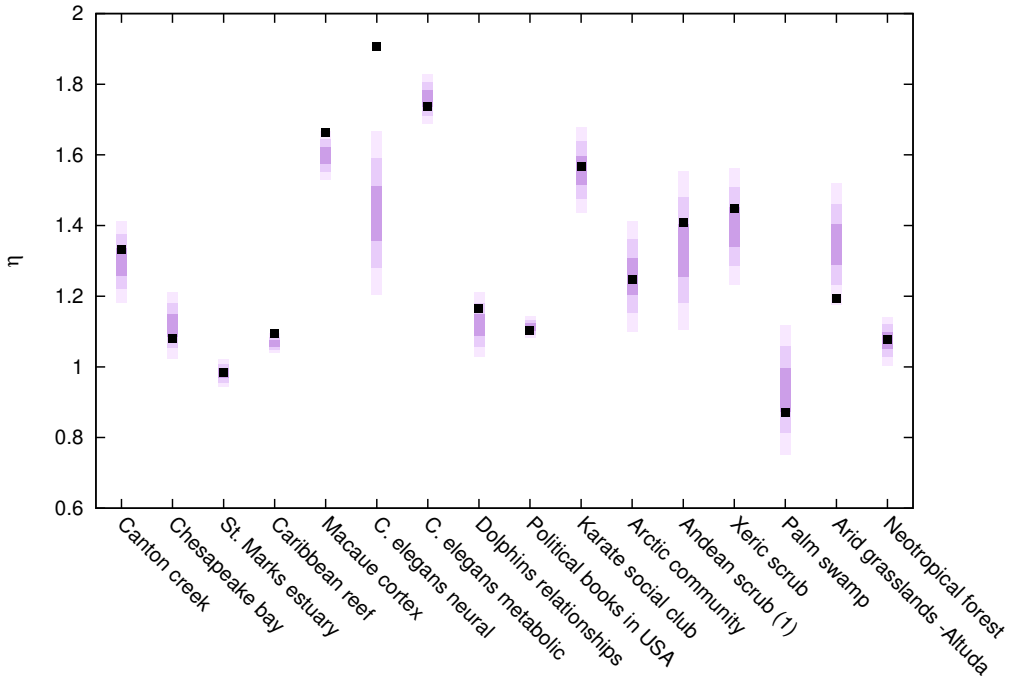

Figure 5.6: Averaged nestedness index, η, as measured in in the second null model in which the value of r is preserved. The actual value of η in the real network is marked with a black square, while the coloured intervals corresponds to one, two and three standard deviations respectively. In most cases the empirical value lies in or near the corresponding interval. Allowing for two or three standard deviations essentially all empirical points yield within the corresponding interval.

where $g(r)$ stands for the (previously observed) frequency. This algorithm allows for uniform searches in "r-space".

Obtaining an average η as a function of r The use of this algorithm allows us to uniformly explore the r-space, and as such, to measure the nestedness of these rare configurations that will be otherwise out of range. This has been the technique used to obtain the data shown inf Fig.5.2.C. In this case, for the three different types of degree-distributions considered in section 5.2, we have generated uncorrelated networks using the configuration model, and progresively change their structure so as to sweep different values of r. While running the search in the r-space, we also measure and store the

nestedness values of Eq.(5.7) for every (binned) value of r; after properly normalizing, we obtain the averaged value of nestedness as a function of r as shown in Fig.5.2.C. In all three cases (scale-free, Poissonian, and Gaussian distributions), we obtain a very clear (almost linear) dependence between r and η: disassortative networks are distinctly nested (on average) while assortative ones are anti-nested (on average). Let us caution that this conclusion holds "on average", i.e. our results do not necessarily imply that any particular disassortative network is actually nested.

Null model with correlations The configuration model generates uncorrelated networks with a given degree sequence. However, within this frame it is not possible to obtain random networks with a determined value of its correlations. Since we are interested in taking the correlations into account, we use the Wang-Landau algorithm to sample rare configurations. Analogously to what we do in random finite networks, we compute the average nestedness in this second ensemble for a fixed value of r. That is, we only consider random networks with correlations close to the empirical values ($r \approx r_{emp}$). In particular, we have considered the same 16 workbench networks as above, and produced Figure 5.6, where we show for each network, the averaged nestedness (with its corresponding standard deviation) as a function of r. The empirical values of η are marked with black boxes.

In this new, more constrained, ensemble the null model performs only slightly better than the configurational one, with in general, is to be expected given the fact that we are reducing the ensemble. However, empirical values are still reasonably well explained by randomized values, in almost all cases.

5.4 Chapter summary

- The first contribution of this chapter is that a new analytical nestedness index has been introduced. It is a variant of the one introduced in Ref.[34], allowing for analytical developments, which are not feasible with standard computational estimators (or calculators) of nestedness. Besides that, the new index exhibits number of additional advantages: (i) it allows us to identify the amount of nestedness associated with each single node in a network, making it possible to define a "local nestedness"; (ii) the new index is properly normalized and provides an output equal to unity in uncorrelated random networks, allowing us in this way to discriminate contributions to nestedness beyond network heterogeneity.

- There are more disassortative (negatively degree-degree correlated) networks than assortative ones even among randomly assembled networks. Different reasons for this have already being pointed out in the literature [115, 146, 191] and we have confirmed that indeed this is the case for finite networks built with the configuration model. Therefore ,the neutral expectation for finite random networks is to have some non-vanishing level of disassortativity ($r < 0$). Analogously, there is a very similar tendency for finite random networks to be naturally nested (Fig. 5.2).

- There is a clean-cut correspondence between nestedness and disassortativity: disassortative networks are typically nested and nested networks are typically disassortative (Fig. 5.2 and 5.5)

- Heterogeneity of natural networks can account for much of their nestedness. Analyses of 60 empirical networks (both bipartite and non-bipartite) taken from the literature reveal that in many cases the measured nestedness is in good correspondence with that of the degree-preserving null model (Fig. 5.4). In particular, almost 90% of the studied bipartite networks are well described by the null model and this figure rises up to 100% when a more refined null model is considered (Table 5.2). Degree heterogeneity together with the finite size of real networks suffice to justify most of the empirically observed levels of nestedness in ecological bipartite network.

5.5 Data supplement to chapter 4

Empirical unimodal networks on Fig 5.5							
Network	r	r z-score	r P-value	η	η z-score	η P-value	η z-score with correlations
Biological - Food Webs (13)							
Little Rock lake	-0.327	0.514 ()	0.300 ()	1.326	0.026 ()	0.199 ()	0.350 ()
Ythan Estuary	-0.278	2.000 ()	0.039 (*)	1.482	-2.459 ()	0.021 (*)	-1.339 ()
Stony Stream	-0.208	0.368 ()	0.401 ()	1.295	-0.500 ()	0.300 ()	-0.594 ()
Canton creek	-0.201	0.667 ()	0.266 ()	1.331	-1.454 ()	0.096 ()	-0.880 ()
Skipwith Pond	-0.086	-3.406 (*)	0.001 (*)	1.014	2.142 ()	0.017 (*)	-4.600 (*)
El Verde rainforest	-0.171	0.133 ()	0.306 ()	1.187	1.520 ()	0.072 ()	0.561 ()
Caribbean Reef	-0.164	-0.778 ()	0.245 ()	1.095	-0.687 ()	0.297 ()	-3.140 (*)
St. Martin Island	-0.135	0.723 ()	0.226 ()	1.181	-3.000 ()	0.005 (*)	-5.488 (*)
UK grassland	-0.182	1.093 ()	0.142 ()	1.266	-1.412 ()	0.069 ()	-0.985 ()
Chesapeake bay	-0.115	-0.510 ()	0.299 ()	1.080	1.015 ()	0.162 ()	1.191 ()
Northwest Shelf	-0.082	-5.125 (*)	0.000 (*)	1.024	5.000 (*)	0.001 (*)	-0.495 ()
Coachella Valley	0.050	-4.548 (*)	0.000 (*)	0.990	4.272 (*)	0.000 (*)	0.031 ()
St. Marks Forest	0.116	-3.379 (*)	0.000 (*)	0.984	2.733 ()	0.002 (*)	-0.205 ()
Biological - other (4)							
Macaque cortex [97]	-0.295	-3.167 (*)	0.001 (*)	1.664	0.076 ()	0.484 ()	-2.909 ()
C.elegans metabolic network [71]	-0.220	4.434 (*)	0.001 (*)	1.736	-1.500 ()	0.166 ()	0.5 ()
C.elegans neuronal connectivity [268]	-0.163	7.443 (*)	0.000 (*)	1.908	-12.446 (*)	0.000 (*)	-3.875 (*)
Yeast metabolic [111]	-0.076	-	-	1.411	-	-	-
Social (12)							

Facebook friendship [154]	-0.712	-	-	4.077	-	-	-
Karate club [277]	-0.476	3.744 (*)	0.000 (*)	1.568	-3.487 (*)	0.000 (*)	-0.303 ()
Kaitiaki web relations [250]	-0.245	1.698 (*)	0.042 (*)	0.203	-0.904 ()	0.165 ()	0.287 ()
Political blogs [2]	-0.221	-	-	1.549	-	-	
Ownership network [179]	-0.181	-	-	2.315	-	-	
Consulting firm [59]	-0.167	-	-	1.073	-	-	-
Political books in th US [128]	-0.128	-0.622 ()	0.229 ()	1.103	0.800 ()	0.189 ()	0.919 ()
Dolphins relationships [139]	-0.044	0.149 ()	0.460 ()	1.164	-1.020 ()	0.157 ()	-1.480 ()
Jazz musicians [88]	0.020	-7.000 (*)	0.000 (*)	0.997	4.631 (*)	0.000 (*)	1.631 ()
Football games [87]	0.162	-4.526 (*)	0.000 (*)	0.999	2.000 ()	0.002 (*)	-3.832 (*)
Net-science arXiv [175]	0.462	-	-	0.653	-	-	-
Technical (3)							
Airports network [187]	-0.221	-	-	2.352	-	-	-
Power grid [268]	0.003	-	-	1.198	-	-	-
World flights [188]	0.051	-	-	1.134	-	-	-

Table 5.1: Pearson's correlation coefficient and nestedness η for different empirical (unimodal) networks from different backgrounds appearing in Fig.4 . The z-score and P-values are calculated with respect to the random null model, and the "z-score with correlations" with the null model constraint to keep to the actual value of the Pearson's coefficient r. The asterisks $*$ stand for z-values larger than 3 (in absolute value) and P-values smaller than the threshold 0.05.

Empirical bipartite networks on Fig 5.5

Network	r	r z-score	r P-value	η	η z-score	η P-value	η z-score with correlations
Plant - Pollinator (28)							
Andean scrub (1) (Chile) [18]	-0.239	2.448 ()	0.009 (*)	1.409	-2.857 ()	0.005 (*)	-1.067 ()
Andean scrub (2) (Chile) [18]	-0.363	2.840 ()	0.003 (*)	1.208	-3.246 (*)	0.001 (*)	-1.595 ()
Montane forest and grassland (USA) [50]	-0.131	-2.364 ()	0.011 (*)	1.218	1.067 ()	0.131 ()	-0.442 ()
High-altitude desert (Canary Islands) [75]	-0.018	0.625 ()	0.253 ()	0.989	-0.340 ()	0.334 ()	0.397 ()
Arctic community (Canada) [102]	-0.128	1.467 ()	0.079 ()	1.246	-1.164 ()	0.116 ()	0.195 ()
Multiple Communities (Galápagos Islands) [157]	-0.638	5.444 (*)	0.000 (*)	2.416	-5.033 (*)	0.003 (*)	-1.957 ()
Xeric scrub (Argentina) [159]	-0.241	4.086 (*)	0.000 (*)	1.447	-4.104 (*)	0.000 (*)	-0.919 ()
Woody riverine veg. and xeric scrub (Argentina) [159]	-0.124	0.722 ()	0.224 ()	1.268	-1.550 ()	0.071 ()	-1.745 ()
Palm swamp community (Venezuela) [206]	-0.074	0.018 ()	0.492 ()	0.872	0.370 ()	0.382 ()	1.033 ()
Boreal forest (Canada) [26]	-0.069	1.815 ()	0.030 (*)	1.309	-0.926 ()	0.138 ()	-0.151 ()
Alpine subarctic community (sweden) [77]	-0.034	0.500 ()	0.298 ()	1.061	-0.500 ()	0.282 ()	-0.625 ()
High Arctic (Canada) [125]	-0.075	0.875 ()	0.171 ()	1.322	-2.866 ()	0.030 (*)	-2.977 ()
Medow (U.K) [161]	-0.104	0.049 ()	0.434 ()	1.442	-1.049 ()	0.156 ()	-0.658 ()
Arctic community (Canada) [169]	-0.104	-1.720 ()	0.031 (*)	0.821	1.472 ()	0.069 ()	-0.485 ()
Deciduous forest (USA) [170]	-0.040	-0.571 ()	0.307 ()	1.130	0.544 ()	0.297 ()	0.130 ()
Coastal forest (Mauritius Island) [182]	-0.488	2.278 ()	0.020 (*)	1.257	-3.378 (*)	0.002 (*)	-2.143 ()
Upland grassland (South Africa) [184]	-0.041	-1.000 ()	0.180 ()	1.180	0.885 ()	0.232 ()	0.017 ()
Maple-oak woodland (USA) [221]	-0.037	0.500 ()	0.320 ()	1.114	-0.166 ()	0.426 ()	0.578 ()
Snowy Mountains of Australia [106]	-0.139	0.778 ()	0.236 ()	1.165	-1.432 ()	0.076 ()	-0.804 ()
Peat bog (Canada) [225]	0.003	-2.857 ()	0.002 (*)	0.905	1.722 ()	0.035 (*)	-0.367 ()
Arroyo Goye (Argentina) [260–262]	-0.034	-0.091 ()	0.497 ()	0.114	-0.125 ()	0.446 ()	-0.507 ()
Cerro Lopez (Argentina) [260–262]	-0.030	0.071 ()	0.412 ()	0.133	-0.250 ()	0.352 ()	-0.078 ()

Llao Llao (Argentina) [260–262]	-0.059	0.182 ()	0.375 ()	0.124	-0.375 ()	0.317 ()	-0.504 ()
Mascardi (c) (Argentina) [260–262]	-0.034	0.500 ()	0.307 ()	0.083	0.000 ()	0.417 ()	1.300 ()
Mascardi (nc) (Argentina) [260–262]	-0.023	-1.400 ()	0.045 (*)	0.151	1.555 ()	0.040 (*)	0.937 ()
Quetrihue (c) (Argentina) [260–262]	-0.051	0.900 ()	0.144 ()	0.104	-1.428 ()	0.081 ()	-0.806 ()
Quetrihue (nc) (Argentina) [260–262]	-0.043	1.636 ()	0.073 ()	0.079	-1.750 ()	0.053 ()	-0.627 ()
Safariland (Argentina) [260–262]	-0.035	0.031 ()	0.432 ()	0.097	-0.142 ()	0.406 ()	-0.182 ()
Plant - Hervibore (4)							
Arid grasslands in Marathon (USA) [113]	-0.168	4.571 (*)	0.000 (*)	1.193	-1.256 ()	0.112 ()	2.696 ()
Arid grasslands in Altuda (USA) [113]	-0.215	3.200 (*)	0.001 (*)	1.316	-1.489 ()	0.048 (*)	1.226 ()
Whole country (Britain) [133]	-0.016	-0.500 ()	0.198 ()	1.100	0.875 ()	0.177 ()	0.693 ()
Whole country (Finland) [133]	-0.055	0.651 ()	0.275 ()	1.420	0.000 ()	0.240 ()	-0.320 ()
Plant - Seed Disperser (5)							
Forest (Papa New Guinea) [36]	-0.030	0.075 ()	0.427 ()	1.131	-0.588 ()	0.246 ()	-0.537 ()
Semideciduous tropical forest (Panama) [204]	-0.689	0.276 ()	0.322 ()	1.081	-0.14 ()	0.449 ()	-0.950 ()
Neotropical forest (Trinidad) [227]	-0.026	0.858 ()	0.192 ()	1.077	-0.736 ()	0.231 ()	-0.178 ()
Birds and Berries. Calton (England) [227]	-0.087	0.500 ()	0.346 ()	0.796	0.333 ()	0.358 ()	0.965 ()
Temperate woodland (Britain) [231]	-0.459	0.814 ()	0.266 ()	0.926	-0.561 ()	0.470 ()	1.559 ()
Anemone-Fish (1)							
Coral reefs [185]	-0.064	0.063 ()	0.483 ()	1.085	-0.633 ()	0.322 ()	-1.025 ()

Table 5.2: Pearson's correlation and nestedness η for empirical bimodal networks from different backgrounds appearing in Fig.4 . The z-score and P-values are calculated with respect to the random null model, and the "z-score with correlations" with the null model restricted to the actual correlation coefficient r of the network. As above, the asterisks $*$ stand for z-values larger than 3 (in absolute value) and P-values smaller than the threshold 0.05.

Chapter 6

Ranking species in mutualistic networks

Assessing the stability and robustness of complex ecosystems is a fundamental problem in conservation ecology [72, 74, 167, 228, 232]. The loss of an individual "keystone" species can induce cascade effects –i.e. a series of secondary extinctions triggered by the primary one– propagating the damage through the network. Thus, the relative "importance" of a given species within a ecological network could be gauged as a function of the eventual size of the cascade of extinctions its loss would potentially cause. A successful ranking of species importance should rank first those species that trigger larger extinction cascades.

In the context of food webs, species rankings have been long sought (see e.g. [82, 116]). For example, Allesina and Pascual [8] successfully applied the Google's PageRank algoritm [189] to order species within food webs, much as Google ranks webpages.

Mutualistic ecological communities such as those formed by plants and their pollinators, plant seeds and their dispersers, or anemone and the fishes that inhabit them, etc. constitute another broadly studied set of ecological networks. These comprise two different sets of living beings that benefit from each other and as such can be represented in terms of bipartite networks [178]. Mutualistic networks turn out to have a very particular "nested" architecture [29, 34, 233] in which specialist species –interacting with only a few mutualistic partners– tend to be connected with generalists (Figure 6.1). Such a nested design is believed to confer robustness against species loss and

other systemic damages, thus fostering biodiversity [107, 122]. Determining a ranking

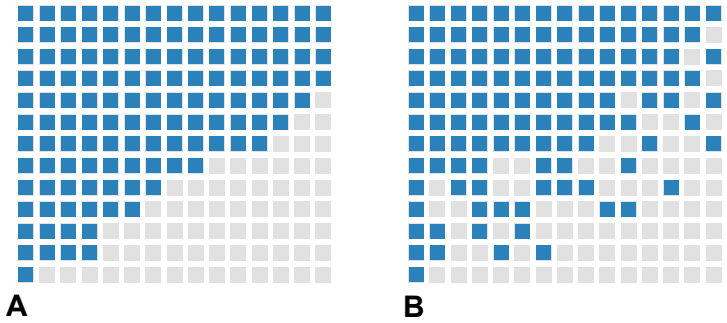

Figure 6.1: Example of two different bipartite networks with different levels of nestedness. For simplicity, we focus on binary networks: blue squares correspond to existing interactions while empty ones describe absent links. A perfectly nested network (A) shows a characteristic interaction matrix in which *specialist* species –with low connectivity– interact only with *generalist* ones. The matrix in (B) has a lesser degree of nestedness (see [233, P3] and [19] for quantifications of nestedness).

of species importance in mutualistic communities poses an important practical challenge, as it would be highly desirable to know which species are more crucial for the long-term stability of the community. The goal would be to establish a proper ordering of species, ranking them in order of decreasing importance for the community. This would facilitate the design of sound conservation policies protecting the most important species.

Following the experience from food webs we could, in principle, employ the PageRank algorithm to rank mutualistic species in bipartite networks. PageRank [8, 189] is a linear-algebra iterative algorithm which, in a nutshell, computes the "importance" of a given node as the linear superposition of the importance of the nodes connecting to it, in a recursive and self-consistent way [1]. However, in this work, taking inspiration from a recent breakthrough on economics/econometrics [57, 243], we propose to employ a novel **non-linear** algorithm specially designed for bipartite networks.

Tacchella *et al.* [57, 243] analyzed economic data from the world trade network (i.e.

[1] PageRank measures the steady-state probability of finding a random walker –which moves following the links of the network but also can perform random jumps to arbitrary nodes with small probability– at any given node of the network.

the bipartite network of countries and the products they export). The goal was to infer an objective ranking of countries in terms of their "fitness" and a classification of the products in terms of increasing "complexity". Inspection of such economic data reveals that rich (high fitness) countries are not specialized into producing complex products (such as high-tech devices) exclusively. Rather, they export a highly diversified variety of goods, including less-complex ones (e.g. cereals). On the other hand, poor (low fitness) countries only produce low-complexity merchandises. These facts are reflected in the nested structure of the corresponding bipartite network [57, 243], with a shape similar to that in Figure 6.1. The main idea behind the novel algorithm of Tacchella *et. al.* is that while the fitness of a country can be safely defined as the linear average of the complexity of the products it exports, the reverse does not make sense. Indeed, the complexity of a given product cannot be meaningfully estimated as the average fitness of the countries producing it, but is much better characterised by the minimal fitness required to produce it [243]. To implement this idea Tacchella *et al.* proposed an iterative non-linear algorithm (see below) and were able to compute the fitness of all countries and the complexity of all products in a self-consistent way, using solely information contained in the matrix of economic transactions. The novel algorithm clearly outperforms PageRank and leads to striking implications for understanding the global trade market [57, 243].

Here, we consider a set of 63 real mutualistic networks taken from the literature (45 pollination networks, 16 frugivore seed-dispersal, and 2 other networks) and rank the species accordingly to different criteria (such as node-connectivity, betweeness centrality, PageRank, etc.) including the novel non-linear algorithm. Each of the employed criteria leads to a different ranking of species. We analyze the quality of any of these orderings by monitoring how fast the network collapses if species are sequentially removed in order of decreasing ranking. The best ranking is the one for which the network breaks down more rapidly. Our conclusion is that the non-linear algorithm clearly outperforms all others, thus providing us with an efficient and powerful scheme to gauge the relative importance of species in mutualistic communities.

6.1 MusRank: non-linear ranking algorithm for mutualistic networks

Inspired by the work of Tacchella *et. al.*, [57, 243] we propose a novel ranking algorithm for mutualistic networks of ecological relevance. We shall refer to it as mutualistic species rank (**MusRank**). To establish a common terminology for plant-pollinator, seed-disperser, and anemone-fish networks, we refer to plants, seeds and anemones as "passive" (P) elements, while pollinators, dispersers, and fishes are their "active" (A) partners; rather than fitness and complexity now we use the terms *importance* and *vulnerability*, for the two emerging species rankings, respectively. It is natural to identify products with *passive* components and countries with *active* ones (but the opposite identification can also be made; see below). We assume that the importance of an active species, is determined by the number of its mutualistic passive partners, each one weighted with its own vulnerability: the more partners and the more vulnerable they are, the more important an active element is.

On the other hand, the vulnerability of a passive element will be bounded by the less important species it interacts with. The rationale behind this is that, given that mutualistic networks are nested, specialized species tend to interact with generalists. If a passive element interacts only with generalists it is most certainly a specialist and therefore highly vulnerable as it can disappear if a few generalists go extinct.

The non-linear algorithm, encoding these ideas, is summarized in Eq.(6.1). The importance of active elements, $I_{A=1,\ldots,A_{max}}$, and the vulnerability of passive ones, $V_{P=1,\ldots,P_{max}}$, are computed at iteration n as a function of their values in iteration $n-1$ using the interaction (or adjacency) matrix M_{AP} as the only input:

$$\begin{aligned} \tilde{I}_A^{(n)} = \sum_{P=1}^{P_{max}} M_{AP} V_P^{(n-1)} \quad &\longrightarrow \quad I_A{}^{(n)} = \frac{\tilde{I}_A^{(n)}}{\langle \tilde{I}_A^{(n)} \rangle_A} \\ \tilde{V}_P^{(n)} = \frac{1}{\sum_{A=1}^{A_{max}} M_{AP} \frac{1}{I_A{}^{(n-1)}}} \quad &\longrightarrow \quad V_P{}^{(n)} = \frac{\tilde{V}_P^{(n)}}{\langle \tilde{V}_P^{(n)} \rangle_P}. \end{aligned} \tag{6.1}$$

In a first step (left), intermediate values of the importance and vulnerability are cal-

culated for each species: the first as the average of vulnerabilities of its partners and the second as the inverse of the average of its partners inverse importances [243]. In a second step (right), both values are normalised to their mean values. In this way, starting from arbitrary initial conditions (e.g. ${I_A}^{(0)} = 1$ for all A and ${V_P}^{(0)} = 1$, for all P) the two-step transformation above is iterated until a fixed point is reached. Such a fixed point –which does not depend on initial conditions– defines the output of the algorithm: a ranking of importances and vulnerabilities for active and passive species, respectively.

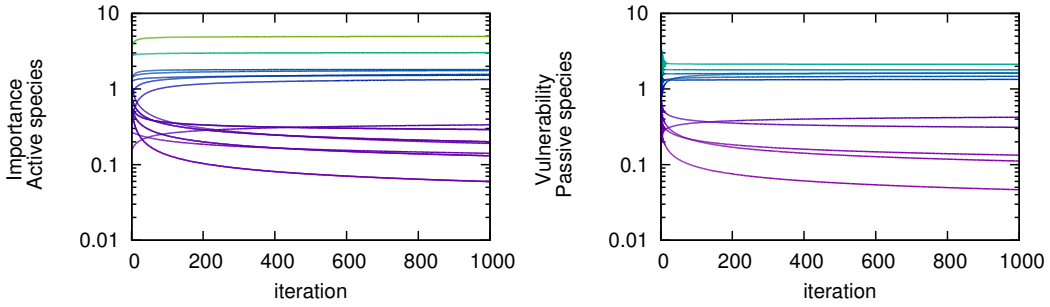

Figure 6.2: Left and right panels show the iterative process of the calculation of active species *Importance* and passive species *Vulnerability* defined by Eq.(6.1) in an empirical mutualistic community composed by 18 active species (pollinators) and 11 passive species (plants) [169]. Iteration of the non-linear algorithm determines a fixed point, independent of initial conditions (in this case ${I_A}i^{(0)}$ and ${V_P}i^{(0)}$ are randomly distributed), where two different rankings emerge: a classification of active species by its importance and a classification of passive species by its vulnerability.

6.1.1 Other algorithms used in the study

Along this study we have used different ranking techniques as a workbench. Each of them prioritizes a different quality of the nodes and provides a different ranking of species. The details of each algorithm are as follows:

- **CLOS**: Nodes are sorted in order of decreasing closeness centrality. The closeness centrality of a node is measured as the inverse of the average shortest distance to all other nodes in the network. The higher the closeness of a node, the bet-

ter it will spread information trough the network. We compute it using the *closeness_centrality* function of the bipartite package in Python *NetworkX*.

- **EIG**: Nodes are sorted in decreasing order of their overlap with the highest eigenvalue. To calculate the eigenvector centrality of the bipartite network we use the *gsl* functions for solving non-symmetric matrices.

- **BTW**: Nodes are sorted in order of decreasing betweeness centrality. The betweeness centrality of a node measures the fraction of shortest paths between all possible node pairs in the network, in which it appears. Nodes with a high betwenness acts as bridges amidst different "clusters" of the network, hence the higer the betwenness of a node, the more control it has over the flow of information. We use the *betweeness_centrality* function of the *bipartite* section of algorithms in the Python package *NetworkX*.

- **DEG**: Nodes are sorted in order of decreasing number of connections.

- **NES**: Nodes are sorted in order of the inverse contribution to network nestedness. We calculate the total nestedness of a given bipartite matrix, and the contribution of each species to the total as described in [P3]. Species that contribute most to the community nestedness are the most vulnerable ones [218]. In order to look for the fastest community collapse we target them in order of increasing contribution to nestedness .

- **PAGE**: Nodes are sorted in decreasing order of Google's PageRank.The ranking is given by the projection over each node of the leading eigenvalue of the matrix H, whose elements are defined as

$$h_{ij} = d \cdot a_{ij} / \sum_j a_{ij} + (1-d)/N.$$

The constant d is a "damping factor" needed to warrant that the matrix is irreducible, and a_{ij} are the elements of the adjacency matrix. The value of d has been set to 0.999, but results are not very sensitive to this choice.

Genetic algorithm

The genetic algorithm is designed to seek for those sequences of extinction that maximize the extinction area. We start with 10^4 different random orderings of the A_{max} active species. At each iteration-step two of these orderings are randomly selected. Each one beats the other with a probability proportional to its associated extinction area (normalized to the sum of both extinction areas). The loser sequence is erased from the set and a copy of the winner will occupy its place. With a small probability, $\mu = 0.005$, this copy suffers a mutation, meaning that two random nodes exchange their positions in the ordering. The algorithm is iterated until no better solutions are found in a sufficiently large time window, that is, until no appreciable changes are seen in the extinction area with increasing time. If the network is too large, this algorithm might not be able to find a stationary optimal solution within a reasonable computation time.

6.2 Assessing the quality of a given ranking

In order to evaluate the quality of any possible ranking of species for a given mutualistic network we proceed by computationally implementing the following protocol (see Figure 6.3.A). Active species are removed progressively following the ordering prescribed by any specified ranking algorithm. Secondary extinctions are monitored (a species is declared extinct when it no longer has any mutualistic partners to interact with). The process is iterated until all the species in the network have gone extinct. The total fraction of extinct species as a function of the number of deleted species defines a *extinction curve* [8]. For each possible sequence of eradications the extinction area is obtained as the integral of the extinction curve (see Figure 6.3.B). This procedure allows for a quantitative discrimination of species rankings: the best possible ordering of species would be the one for which the largest extinction area is obtained upon progressively removing active species in order of decreasing rank.

An exhaustive search of the optimal ranking (in the space of all possible orderings) can be performed for relatively small networks but becomes an unfeasible task for larger ones. To have an estimation of the optimal ranking we implemented a genetic algorithm (**GA**) (see sec. 6.1.1) devised to obtain the maximal possible extinction area by searching in the space of all possible orderings. For some of the largest networks we

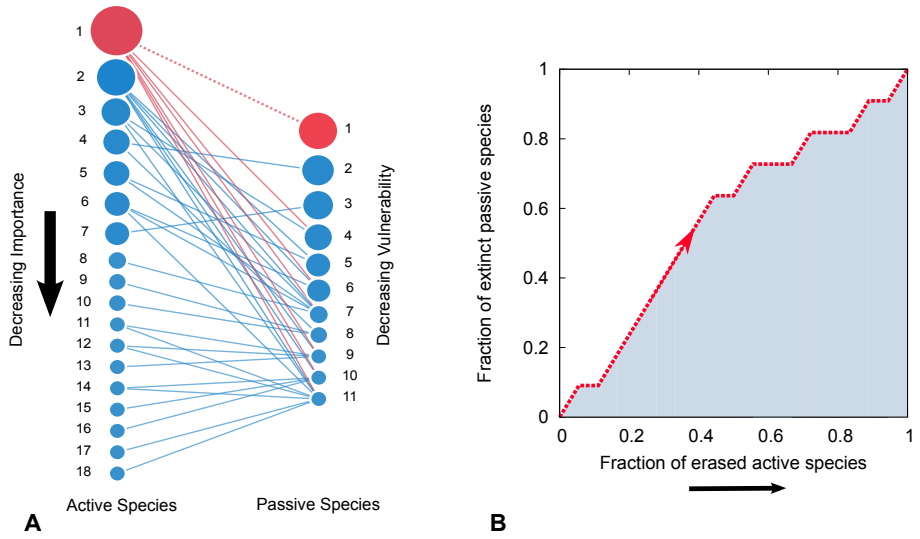

Figure 6.3: Left: schematic representation of the extinction protocol for an empirical mutualistic network: Arctic community [169]) with 18 active (pollinators) and 11 passive (plants) species. Both active (left) and passive (right) species are ordered following some prescribed ranking; from the highest ranked species (top) to the lower-ranked ones (bottom). The (blue and red) lines represent mutualistic interactions as encoded in the interaction (or adjacency) matrix. Active species are progressively removed from the community, their corresponding (red) links are erased, and passive species are declared extinct whenever they lose al their connections. Right: extinction curve, showing the fraction of extinct passive species as a function of the number of sequentially removed active ones for a given specified ranking. The shaded region is the extinction area for the ranking under consideration. Different rankings lead to different extinction areas. The larger the area the better the ranking.

studied (in particular, for Montane forest and grassland, Beech forest and Phryganic ecosystem with $275,678$, and 666 active nodes respectively; see table 6.5 at the end of this chapter) the computational time required for the genetic algorithm to converge is exceedingly large and satisfactory results were not found.

6.3 MusRank performance

We compared different rankings based on (see secc. 6.1.1): a) decreasing closeness centrality (**CLOS**), b) decreasing eigenvector centrality (**EIG**), c) decreasing betweeness

centrality (**BTW**), d) decreasing degree centrality (**DEG**), e) increasing contribution to nestedness (**NES**) as described in [P3], f) decreasing PageRank (**PAGE**), and g) decreasing importance as measured by MusRank (**MUS**).

The average extinction area of the different algorithms was obtained for all networks in the dataset. In the frequent case in which the order is degenerate (more than one node were rated with the same value), we considered 10^3 different randomizations and computed the averaged extinction area.

For the sake of completeness we have also repeated all the protocol above, but exchanging in Eq.(6.1) the roles of active and passive species, i.e. assigning importances to passive species and vulnerabilities to active ones. We refer to this as "reversed" algorithm. We have also studied extinction areas by progressively removing passive species (rather than active ones) and monitoring secondary extinctions of active species.

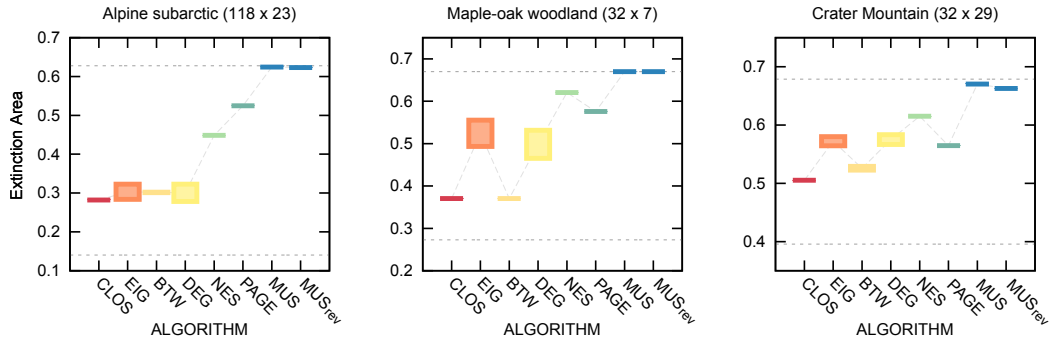

Figure 6.4: Extinction areas for three different mutualistic networks (names and sizes, specified above) as obtained employing the different ranking schemes described in the text. The upper dashed line shows the optimal performance corresponding to the ranking found by the genetic algorithm (**GA**) search, and the lower one the null-expectation, that is the averaged area obtained when targeting nodes in a random order. The different algorithms used to rank the nodes are: closeness centrality (**CLOS**), eigenvector centrality (**EIG**), betweeness centrality (**BTW**), degree centrality (**DEG**), nestedness centrality (**NES**), PageRank (**PAGE**), and importance as measured by the MusRank (**MUS**). $\mathbf{MUS}_{rev}$ corresponds to the reversed version of the algorithm in which the roles of active and passive species are exchanged. The height of the boxes corresponds to the standard deviation of the results when averaging over 10^3 random ways to break degeneracies in the orderings.

Figure 6.4 illustrates the performance of the different rankings/algorithms for three different instances of mutualistic networks. Extinction areas are plotted for each of the considered ranking algorithms. In the three cases MusRank gives results closest to the corresponding optimal solutions as derived from the genetic algorithm. In almost all of the 63 studied cases, results are much better for the novel ranking than for any of the other ones. PageRank gives similar results to MusRank in a few cases (including a relative large network with 102 nodes). Apart from this, only for very small networks (with less than 17 active species) some other method different from PageRank gives extinction areas similar to the ones of the novel algorithm. In about one third of the networks, the ranking provided by MusRank is as good as the one found by the **GA** and in some cases (networks for which the **GA** could not converge in a reasonable time) extinctions areas are larger for MusRank than for the **GA**.

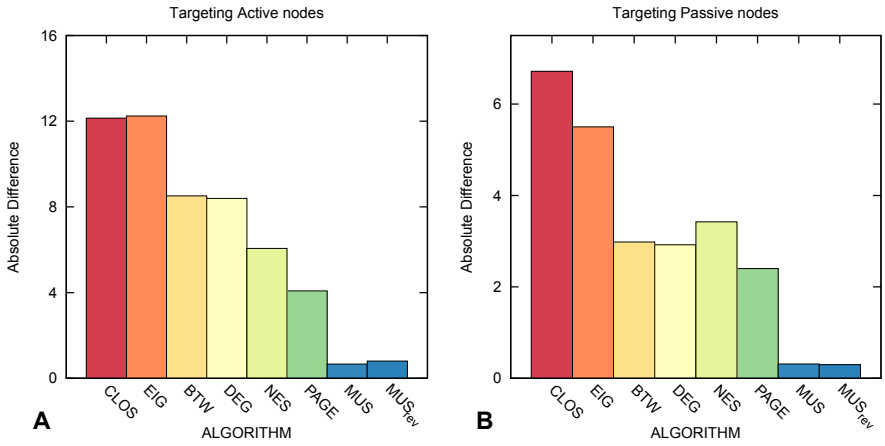

Figure 6.5: Averaged deviation of the extinction area obtained for each of the employed rankings (or algorithms) from the maximal possible value as determined using the genetic algorithm (average over 60 networks in the database). The left A (right B) panel shows results when active (passive) species are targeted and passive (active) species undergo secondary extinctions. Results are consistently much better for the MusRank, in either the direct or the reversed version, than for any other ranking scheme.

Figure 6.5 gives a global picture of the performance of the different rankings. It shows the difference, averaged over 60 mutualistic networks, between the optimal solution as found by the **GA** and that of each specific ranking (the 3 networks for which

the **GA** does not converge are excluded from this analysis). Figure 6.5.A illustrates that the ranking provided by the MusRank –either in the direct or the reversed form– greatly outperforms all others.

The same conclusion can be reached when progressively removing passive rather than active species, ordered in a sequence of increasing vulnerability rather than decreasing importance (we target first the strongest species), see Figure 6.5.B. Therefore, both targeting strategies and both the direct and the reversed versions of the algorithm provide results of similar quality.

6.3.1 Optimally packed matrices

The ranking provided by MusRank, in which nodes are arranged by their level of importance or vulnerability, permits us to obtain a highly packed matrix as illustrated in Figure 6.6. By "packed" we mean that a neat curve separates densely occupied and empty parts of the matrix. It could be thought that this ordering might be somewhat similar to the one that allegedly packs the matrix in the most efficient way (as defined by existing algorithms usually employed in the literature to measure nestedness [19]). However, as Figure 6.6 vividly illustrates, the ordering provided by MusRank gives a more packed matrix than that obtained by the standard method employed by nestedness calculators [19]. This suggests that MusRank should be used (rather than existing ones) to measure nestedness in bipartite matrices.

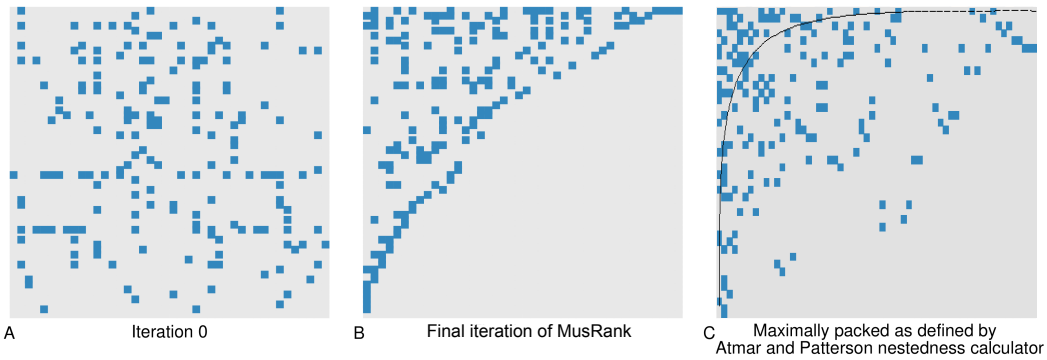

Figure 6.6: Interaction matrix of a mutualistic community in the Andes [18] composed of 42 pollinators and 61 plants ordered by decreasing importance and increasing vulnerability respectively, as measured by MusRank. Panels A and B show two different shots of the iteration process: the initial random condition and the final (fixed-point) ranking obtained after iteration. Panel C shows the same matrix but with nodes labelled in an order which gives the maximally packed matrix according to the nestedness calculator of Atmar and Patterson [19]. The novel algorithm provides a much more "packed" matrix than this frequently employed method.

6.4 Chapter summary

- In this chapter we present a new method to asses species importance in mutualistic communities: Musrank. It is similar in spirit to Google's PageRank, but of non-linear character. The algorithm provides two complementary rankings: one for active species (such as insects, birds, fish,...) in terms of their importance and one for passive species (plants and their seeds, anemone, etc) in terms of their vulnerability. This ranking outperforms all the others we study (fig. 6.5)

- We also put forward a general framework to asses the quality of importance rankings in mutualistic communities: good rankings lead to a fast break-down of the corresponding mutualistic network when species are progressively removed in decreasing ranking order (fig. 6.3).

- As a byproduct, the ranking with Musrank also provides a maximally packed matrix. When active species are ordered by decreasing importance and passive species by increasing vulnerability, a neat curve separates densely occupied and empty parts of the matrix (fig. 6.6). We suggest this kind of packing should be used in nestedness calculators.

6.5 Data supplement to chapter 5

Networks dataset

Network Name	A_{max}	P_{max}	Label
Plant-Pollinator communities			
Andean scrub (elevation 1), Cordon del Crepo (Chile) [18]	99	87	1
Andean scrub (elevation 2), Cordon del Crepo (Chile [18]	61	42	2
Andean scrub (elevation 3), Cordon del Crepo (Chile) [18]	28	41	3
Boreal forest (Canada) [26]	102	12	4
Montane forest and grassland (U.S.A.) [50]	275	96	5
grassland communities in Norfolk, Hickling (U.K.) [66]	61	17	6
grassland communities in Norfolk, Shelfanger (U.K.) [66]	36	16	7
High-altitude desert, Canary Islands (Spain) [75]	38	11	8
Alpine subarctic community (Sweden) [77]	118	23	9
Mauritius Island (un-published)	13	14	10
Mediterranean shrubland, Doñana (Spain) [100]	179	26	11
Arctic community (Canada) [102]	86	29	12
Snowy Mountains (Australia) [105]	91	42	13
Heathland -heavily invaded- (Mauritius Island) [122]	135	73	14
Heathland -no invaded- (Mauritius Island) [122]	100	58	15
Beech forest (Japan) [123]	678	89	16
Lake Hazen (Canada) [125]	110	27	17
Multiple Communities (Galápagos Islands) [157]	54	105	18
Woody riverine vegetation and xeric scrub (Argentina) [159]	72	23	19
Xeric scrub (Argentina) [159]	45	21	20
Meadow (U.K.) [161]	79	25	21
Arctic community (Canada) [169]	18	11	22
Deciduous forest (U.S.A.) [170]	44	13	23
Coastal forest, Azores Island (Portugal) [182]	12	10	24
Coastal forest, Mauritius Island (Mauritius) [182]	13	14	25
Coastal forest, Gomera Island (Spain) (un-published)	55	29	26
Upland grassland (South Africa) [184]	56	9	27
Coastal scrub (Jamaica) [196]	36	61	28

Phryganic ecosystem (Greece) (un-published)	666	131	29
Mountain, Arthur's Pass (New zealand) [205]	60	18	30
Mountain, Cass (New zealand) [205]	139	41	31
Mountain, Craigieburn (New zealand) [205]	118	49	32
Palm swamp community (Venezuela) [206]	53	28	33
Caatinga (N.E. Brazil) [219]	25	51	34
Maple-oak woodland (U.S.A.) [221]	32	7	35
Peat bog (Canada) [225]	34	13	36
Temperate rain forests, Chiloe (Chile) [226]	33	7	37
Evergreen montane forest, Arroyo Goye (Argentina) [261]	29	10	38
Evergreen montane forest, Cerro Lopez (Argentina) [261]	33	9	39
Evergreen montane forest, Llao Llao (Argentina) [261]	29	10	40
Evergreen montane forest, Mascardi (c) (Argentina) [261]	26	8	41
Evergreen montane forest, Mascardi (nc) (Argentina) [261]	35	8	42
Evergreen montane forest, Quetrihue (c) (Argentina) [261]	27	8	43
Evergreen montane forest, Quetrihue (nc) (Argentina) [261]	24	7	44
Evergreen montane forest, Safariland (Argentina) [261]	27	9	45
Seed-Disperser communities			
Eastern forest, New Jersey (USA) [20]	21	7	46
Forest (Papua New Guinea) [36]	9	31	47
Forested landscape, Caguana (Puerto Rico) [45]	16	25	48
Forested landscape, Cialitos (Puerto Rico) [45]	20	34	49
Forested landscape, Cordillera (Puerto Rico) [45]	13	25	50
Forested landscape, Frontón (Puerto Rico) [45]	15	21	51
Tropical rainforest, Queensland (Australia) [58]	7	72	52
Coastal dune forest, Mtunzini (South Africa) [84]	10	16	53
Forest, Santa Genebra Reserve T1.(Brazil) [86]	18	7	54
Forest, Santa Genebra Reserve T2.(Brazil) [86]	29	35	55
Submontane rainforest (Central Philippine Islands) [96]	19	36	56
Mediterranean shrubland, Hato Ratón (Spain) [117]	17	16	57
Rainforest, Krau Game Reserve (Malaysia) [132]	61	25	58
Crater Mountain Research Station (Papua New Guinea) [142]	32	29	59
Atlantic forest (SE. Brazil) [224]	110	207	60
Montane forest (Costa Rica) [251]	40	170	61

Other communities			
Anemone-fish interactions in coral reefs [185]	26	10	62
Ant-plant interaction in rainforest (Australia) [39]	41	51	63

Table 6.1: Dataset of different mutualistic networks used throughout the study, with A_{max} active and P_{max} passive species.

Chapter 7

Conclusions

Along this thesis we have studied the way in which some particular architectural features of ecological and biological networks determine their overall performance. The system approach used here, based on the application of statistical mechanics of complex networks to different biological and ecological systems, makes possible the study of the emergent or collective behaviour of those living systems, such as linear stability, global directionality, hierarchical structure, robustness against extinctions or the presence of a nested structure, among others. In this way, the study of static networks has proven extraordinarily useful. Our approach can be divided in three parts:

-This highly interdisciplinary approach let us contrast different systems using the same methods, leading us to identify general patterns in their architecture, (as for example the existence of a inherent directionality in ecological and biological networks).

-The design of simple models, subjected to diverse in-silico experiments, constitutes our main "laboratory". It let us contrast the natural network structures with the ones generated by these simple models. Here we have tried to stick to a minimal approach, introducing the less number of parameters and assumptions needed to model some phenomena.

-And last, but not least, the use of null-models is vital to analyse the statistical relevance of natural (or synthetic) structured networks, allowing us to discern whether a particular architectural feature is relevant or on the contrary, is present in random networks. It is worth noting that one should be very careful on the desing of these null-models, since this will determine the outcome. Within this minimal-system approach, we have

covered different properties regarding both directed (foodwebs, transcriptional regulatory genetic networks, signaling networks, neural networks) and undirected networks (plant-pollinator and seed-disperser bipartite mutualistic networks).

In the first part of this thesis we focus our attention in directed networks, mainly in the unaccounted stability of foodwebs and the inherent directionality present in many ecological and biological networks. Regarding foodwebs the most relevant finding is the detection of a strong correlation between lineal stability and the network feature we call trophic coherence. Indeed, cannibalism and trophic coherence together account for most of the variance in stability observed in our dataset [P1].

Along this line we have suggested the Preferential Preying Model as a simple algorithm for generating networks with tunable trophic coherence. Most remarkably, it is able to create "syntethic" networks with similar stability properties to the natural foodwebs. The model also predicts that networks should become more stable with increasing size and complexity, as long as they are sufficiently coherent and the number of links does not grow too fast with size. Although this result should be followed up with further analytical and empirical research, it suggests that we need no longer be surprised at the high stability of large, complex ecosystems.

We must caution that these findings do not imply that trophic coherence was somehow selected for by the forces of nature in order to improve foodweb function. It seems unlikely that there should be any selective pressure on the individuals of a species driving them to do what is best for their ecosystem. Rather, many biological features of a species are associated with its trophic level. Therefore, adaptations which allow a given predator to prey on species A are likely to be useful also in preying on species B if A and B have similar trophic levels. This kind of organisation leads to trophic coherence, which results in high stability. However, If real ecosystems are coherent enough that they become more stable with size and complexity, as our model predicts, then the reverse might be true. Ascertaining whether the loss of a few species would stabilise or destabilise a given community could be important for conservation efforts, particularly for averting "tipping points" [241].

On a different note, but still related to systems stability, we turn our attention to the empirically observed absence of feedback loops in directed biological and ecological networks [85, 94, 245, P2]. This feature have been oftentimes associated with the

required stability these systems should present. However, while the important role of feedback loops in determining dynamical properties of complex networks has been widely recognised in the literature, their statistics remained scarcely studied. Further broadening this line we hypothesize that the (empirically observed) lack of feedback loops stems from the existence of an inherent directionality. To check the hypothesis we devise a simple model with a built-in directionality, which reproduces quite well the empirical fraction of feedback loops of any length by just tuning its only parameter γ. It is worth noting that the qualities of the results seem remarkable considering that our models assumes a number of simplifications that are not trivial. In general loops are not independent (as we consider in the model) since they can share some nodes. In particular hubs are more likely to take part in loops than other nodes with less connectivity. What is more, node degree and hierarchical position could be correlated in empiric networks. Such effects are neglected by our simple model and could be responsible for the deviations of the empirical data from our model predictions.

This finding is similar in spirit to the remarkable observation by Mayaa'n *et al.* that biological networks display a kind of antiferromagnetic ordering. However, while they claim that this organization is behind a enhancement of the stability, our hypothesis is that the existence of an inherent directionality constitutes a simple yet satisfactory parsimonious explanation for the empirically observed lack of feedback loops [P2]. In fact, under the light of Fig. 4.4 the presence of an inherent directionality seems to pervade a vast majority of biological and ecological networks.

Clearly trophic coherence and inherent directionality are not uncorrelated. The fact that there is a trend on the links to align along a direction eases the formation of more "satratified" structures, and as such, is likely to foster the coherence of the networks. This relation poses some interesting questions ¿Is also a high directionality correlated with satbility? ¿Could trophic coherence be present in other biological networks? It should be interesting to gain analytical insight on how are those two features related to each other. We know that trophic coherence plays a relevant role on foodwebs stability. Unraveling whether the absence of loops could be related with stability trough trophic coherence seems a stimulating line of study.

Even if our findings of trophic coherence were reached by working with foodwebs, many directed networks of different type transport energy, matter, information, capital

or other entities in a similar way to how foodwebs carry biomass from producers to apex predators. In this way the existence of a more or less stratified structure could be latent in other networks. It seems likely that the relation between a network's trophic coherence and its leading eigenvalue will be of relevance to other disciplines, and perhaps the preferential Preying model, though overly simplistic for many scenarios, may serve as a first approximation for exploring these effects in a variety of systems.

Moving on to bipartite networks, our work is aimed both to determine which are the topological features behind the nested architecture present in mutualistic plant-animal networks and to bring forward a solid procedure to rank the importance of different species in these valuable communities.

Our first contribution in this topic is the introduction of a new analytical nestedness index. It is normalized so as to provide an output equal to unity in uncorrelated random networks. In this way, having removed the direct effects of the degree distribution –which has a dominant contribution to other measures of nestedness– it is possible to move one step forward and ask how degree-degree correlations (as quantified by Pearson's coefficient) influence nestedness measurements.

Curiously enough, there are more disassortative (negatively degree-degree correlated) networks than assortative ones even among randomly assembled networks. Therefore, the neutral expectation for finite random networks is to have some non-vanishing level of disassortativity ($r < 0$). Accordingly, there is a very similar tendency for finite random networks to be naturally nested. Moreover, there is a clean-cut correspondence between nestedness and disassortativity: disassortative networks are typically nested and nested networks are typically disassortative [P3] (as vividly illustrated in Figure 5.2.C and Fig.5.5). Along this line, analyses of a wide dataset reveals that in many cases the empirical measured nestedness is in good correspondence with that of the degree-preserving null model. As a result, degree heterogeneity together with the finite size of real networks suffice to justify most of the empirically observed levels of nestedness in ecological bipartite network [P3].

It is worth regarding that throughout all this work the use of random models have been of crucial importance in order to determine the statistical relevance of different features. It is the contrast of empirical structured systems with their randomized counterparts which makes some pattern relevant or not. We have taken on a computational

approach, that could be seen as a "microcanonical ensemble" where the network basic properties (N, L, degree-sequence) are kept completely constant. There are other interesting approaches were this restriction is loosened, and a "grandcanonical" approach is rather used (?), allowing for analytical developments. Although both approaches should give the same results in infinite size limit, maybe contrasting their results may give information on finite-size induced effects.

It should be noted that the fact that empirical nestedness measures are in many cases compatible with the random expectation does not mean that natural mutualistic communities are not nested. The pattern of connections where specialist species preferably interact with generalists is indeed present. In fact what our finding essentially highlights it that the main factor contributing to create a nested structure is the degree-heterogeneity (and degree-degree correlations to a much lesser extent). However we do not go further and give any prediction about its origin. Devising a model that render a similar connectivity pattern is crucial for fathom in the origin of the nestedness, and in which way (if any) it contributes to enhance the biodiversity in these communities.

With the idea in mind of contributing to the preservation of these communities we have put forward a novel framework to asses the relative importance of species in mutualistic networks. Inspired by a recent work on economics/econometrics we employ an algorithm, similar in spirit to Google's PageRank but of non-linear character, that we have named MusRank. We also propose a criterion to assess the quality of any given ranking of species: good rankings lead to a fast break-down of the corresponding mutualistic network when species are progressively removed in decreasing ranking order.

In most of the empirical mutualistic networks we analysed the use of our novel algorithm rendered a ranking which clearly outperforms all the alternative ones used as workbench. The emerging ordering allows for assessing the importance of individual species within the whole system in a meaningful, efficient and robust way [P4]. This novel approach –introduced in this work for the first time in the context of mutualistic ecological networks– may prove of practical use for ecosystem management and biodiversity preservation, where decisions on what aspects of ecosystems to explicitly protect need to be made. Furthermore, as a by-product, the excellent packing of nested matrices provided by this non-linear approach (see Figure 6.6) calls for a redefinition

of the way in which nestedness is measured. In particular we suggest that nestedness calculators should use the ranking provided by the present algorithm, which clearly outperforms others in making the nested architecture evident. It is worth noting that the existence of a nested structure (with specialist species preferentially interacting with generalists) is in the heart of the algorithm. Although we show that most of the contribution to network nestedness is contained in the degree-sequence, this particular choice of ranking (by degree) is clearly inferior to the one we propose. It should also be interesting to test whether the technique is also of use in bipartite networks without this particular feature.

Even if the system approach have render successful results in diverse fields, biological and ecological systems still pose a great challenge. There are two main reasons for it. On one hand the interactions that take place in these models are, in general, not well known, and while it is true that the emergent behaviours are, to some extent, indifferent to those "microscopic" details, a better understanding of those one-to-one interactions is crucial to advance towards more predictive results. On the other hand these systems are highly "interactive", that is, they involve many agents that communicate in diverse ways. Ecosystems, for instance, are build up of a vast number of components, interacting in diverse forms: predation, mututalism, parasitism..... Here we have considered two of these interactions independently, however in real ecosystems all these interaction are interlinked, forming a "even-more-complex" system. In this sense, the use of multi-layered (or multiplex) networks could prove of use in this context.

Besides that, we should remember that the systems underlying these webs are all of a dynamical nature, with changing interactions, and hence, studying its "fixed picture" is like studying their time-average. Useful though it have been, one of the main drawbacks of our "satic" approach is that many interesting phenomenology may be dependent on this dynamical behaviour, and hence, will be lost in the fixed view. What is more, these systems have emerged as a result of evolutionary processes. Our ecosystems are the result of millions of years of evolution. Biological networks such as signalling pathways and genetic regulatory webs are the results of different and parallel evolutionary processes in diverse organisms trough time. It is logical to think that in order to fathom how these systems come about and take shape an evolutionary, co-adaptative approach should be attempted, since some features may not be captured without this

perspective.

Answering how the topology of different biological and ecological systems emerge and how does it affects their dynamics is a subject of the utmost importance in our rapid-changing environment. With this work we tried to make further developments along this line, as also a increasing number of scientist do. Fortunately it seems we are immerse in a kind of golden-era of interdisciplinarity, were the paradigm of enlightenment through reductionism and division has been left behind. Now the transmission of knowledge between disciplines and the use of an integrative approach seems to be gaining momentum in the scientific community. The times they are a changing, let's hope the funding goes along with it.

Conclusiones

A lo largo de esta tesis hemos estudiado cómo algunas características arquitectónicas de las redes ecológicas y biológicas determinan su funcionamiento. Para este estudio hemos usado un enfoque "sistémico", considerando todo el sistema en su conjunto, basado en la aplicación de la mecánica estadística de redes complejas a estos sistemas vivos. Este enfoque hace posible el estudio de comportamientos emergentes de interés, como estabilidad, direccionalidad, estructura jerárquica, robustez frente a extinciones o la existencia de una estructura anidada. El estudio de la estructura "estática" de estas redes ha resultado ser de mucha ayuda, y podemos distinguir tres cuestiones fundamentales de nuestro enfoque:

- Al ser muy interdisciplinar (puede aplicarse a sistemas muy diversos) permite contrastar redes de tipo muy diferente, y por tanto permite identificar patrones muy generales.

-Nos basamos en el diseño de modelos y simulaciones computacionales. Esto es nuestro "laboratorio" básico. En este estudio intentamos mantener los modelos lo mas simple posibles.

-Y por ultimo, pero no menos importante, el uso de modelos nulos es fundamental en nuestro estudio. Comparar la estructura de las redes naturales (o sintéticas) con modelos sin estructura (random) permite establecer la relevancia estadística de diferentes patrones. Hay que prestar mucha atención al diseño de estos modelos, pues de eso puede depender el resultado. Dentro de este enfoque "minimal" hemos estudiado diferentes propiedades tanto de redes dirigidas (redes tróficas, genéticas, neuronales) como no dirigidas (redes mutualistas de polinizadores y plantas y de semillas y dispersores).

En la primera parte de la tesis nos centramos principalmente en la gran estabilidad que presentan las redes tróficas y en la direccionalidad inherente a muchas redes

ecológicas y biológicas. En cuanto a redes tróficas, el resultado mas relevante es la fuerte correlación entre estabilidad lineal y una característica de estas redes que hemos denominado "coherencia trófica" [P1]. De hecho la coherencia trófica, junto con el canibalismo, pueden explicar la mayor parte de la varianza en la estabilidad de las redes que usamos. En esta línea proponemos un modelo (Preferential Preying model) como algoritmo sencillo para generar redes sintéticas con una estructura similar a las redes tróficas. Este modelo es capaz de generar redes con un estabilidad similar a las naturales, y sugiere que si las redes están suficientemente estratificadas podría esperarse que la estabilidad del sistema aumentase con el tamaño y la complejidad. Aunque es necesario estudiar esto mas a fondo, esto sugiere que no tendríamos que estar tan sorprendidos por la gran estabilidad de los ecosistemas. Debemos mencionar que estos resultados no implican que la coherencia trófica haya sido, de alguna forma, seleccionada por las fueras de la naturaleza para mejorar el funcionamiento de estas redes. Parece poco probable que exista una presión selectiva actuando sobre los individuos de una especie que les lleve a hacer lo que es mejor para su ecosistema. Mas bien, podríamos pensar que muchas características biológicas de las especies están asociadas con su nivel trófico. De esta manera las adaptaciones que hacen que un predador pueda cazar a A es posible que le sirvan también para cazar a B si A y B tienen niveles tróficos similares. Este tipo de organización genera coherencia trófica, que a su vez genera una alta estabilidad. Aunque, si los ecosistemas reales son suficientemente coherentes como para volverse mas estables con el tamaño y la complejidad, como predice nuestro modelo, entonces lo contrario también podría ser verdad. Poder determinar si la pérdida de unas cuantas especies desestabilizaría o estabilizaría una comunidad es muy importante para la conservación, particularmente evitando puntos de no retorno [241].

Cambiando ligeramente de tema, pero todavía relacionado con la estabilidad, dirigimos nuestra atención a la ausencia de loops (observada experimentalmente) en redes ecológicas y biológicas[85, 94, 245, P2]. Esta característica ha sido asociada muchas veces a la alta estabilidad de este tipo de sistemas[]. Aunque esta relación ha sido ampliamente estudiada, la estadística de loops no lo ha sido tanto. Siguiendo esta linea hacemos la hipótesis de que la ausencia de loops proviene de la existencia de una direccionalidad intrínseca en estos sistemas. Para contrastar esta hipótesis diseñamos un modelo para generar redes con una direccionalidad determinada, dependiente de un

parámetro, γ. Este modelo reproduce bien las curvas experimentales. Merece la pena comentar que la calidad del resultado es sorprendente teniendo en cuenta que nuestro modelo asume muchas simplificaciones que no son triviales, puesto que en general los loops no son independientes, sino que hay nodos que tienen mas tendencia a aparecer en ellos (los hubs, con mucha conectividad). Además la conectividad y la jerarquía podrían estar correlacionados, cosa que tampoco consideramos en nuestro modelo simple. Este resultado es similar a la observación de Mayaa'n et al. de que las redes empíricas muestran un tipo de organización "anrtiferromagnética" , y lo relacionan con la estabilidad en estos sistemas. Nuestra hipótesis es que la existencia de una direccionalidad inherente constituye una explicación simple pero efectiva de esta ausencia de loops [P2].

Claramente la coherencia trófica y la direccionalidad inherente no son fenómenos independientes. El hecho de que exista una tendencia a que los links se alineen en una dirección y no en la otra facilita la formación de estructuras más estratificadas, y por tanto es posible que favorezca la coherencia de las redes. Esta relación nos hace preguntarnos varias cosas ¿Se correlaciona también una alta direccionalidad con una alta estabilidad?¿Podría la coherencia trófica estar presente en otro tipo de redes? Sería interesante avanzar en el conocimiento analítico de la relación entre estas dos magnitudes. Sabemos que la coherencia trófica juega un papel determinante en la estabilidad de las redes tróficas, descubrir si la ausencia de loops podría estar relacionada con la estabilidad a trabes de la coherencia trófica parece una interesante línea de estudio.

Hemos de decir que auqnue los resultados sobre la coherencia trófica los hemos obtenido trabajando con redes tróficas, muchas redes de diferente tipo transportan energía, materia u otras cosas de una manera similar a como las redes tróficas transportan biomasa. En este sentido podría existir una estructura estratificada en otro tipo de redes. Parece probable que la relación entre la coherencia trófica y el autovalor deominante podría ser relevante en otras disciplinar, y quizás el modelo de predación preferencial, aunque muy simple para muchos escenarios, podría servir como una primera aproximación para explorar estos efectos en diferentes sistemas.

Pasando a redes mutualistas, nuestra primera aportación es la introducción de un nuevo índice analítico para cuantificar el anidamiento. Este índice esta normalizado para generar un valor igual a la unidad en redes random no correlacionadas. Esto

nos permite no considerar el efecto que tiene la distribución de grado en la estructura anidada (que es una contribución dominante en otras medidas) y estudiar el efecto de otras características, como las correlaciones de grado(medidas con el coeficiente de correlación de Pearson). Curiosamente, existen mas redes disasortativas que asortativas incluso en redes random. Por lo tanto es esperable que las redes random de tamaño finito exhiban cierto nivel de correlación ($r < 0$). A la vez que esto ocurre, existe una tendencia de las redes de tamaño finito a estar anidadas también de manera natural. Lo que es más, existe una correlación clara entre nestedness y disasortatividad [P3] (como puede verse en la figura 5.2.C y 5.5). En esta línea,el análisis de nuestra base de datos de redes muestra que en muchos casos la medida empírica de la nesteedness es compatible con la del modelo nulo que conserva la distribución de grado. Como resultado, la heterogeneidad de grado, junto con el tamaño finito de las redes reales es suficiente para justificar la mayor parte de la estructura anidada presente en las redes mutualistas.

Es interesante tener en cuenta que a lo largo de todo nuestro trabajo el uso de modelos nulos es crucial para determinar la importancia estadística de algunas estructuras de las redes naturales. Es el contraste entre estos sistemas estructurados y sus versiones randomizadas (sin estructura) lo que nos da información sobre si un patrón es relevante o no. Aquí nosotros tomamos un enfoque computacional, que podría verse como una especie de "colectivo microcanónico" donde las características básicas de la red se mantienen constantes. Existen ostas maneras de acercarse a los modelos nulos dónde esta restricción se elimina, y se usa un enfoque más parecido a un "colectivo macrocanónico", lo que permite hacer más desarrollos analíticos. Aunque ambas formas deben dar resultados similares es posible que estudiarlo de ambas maneras nos de información sobre efectos inducidos por tamaño finito, por ejemplo. Queremos hacer notar que el hecho de que las medidas empíricas del anidamiento sean en muchos casos compatibles con lo que cabría esperar en un modelo nulo no significa que las comunidades mutualistas no tengan una estructura anidada. El patrón de conexiones donde las especies especialistas interaccionan preferiblemente con las generalistas sigue presente. De hecho lo que nuestro resultado resalta es que el factor que más contribuye a la existencia de este tipo de estructura es la existencia de una distribución de grado muy heterogénea (y la presencia de correlaciones a una escala mucho menor). En cualquier

caso, nosotros no hemos ido más allá y no damos ninguna hipótesis sobre el origen de esta distribución. Diseñar un modelo que genere unos patrones de conectividad similares sería crucial para ahondar en cual es realmente el origen del anidamiento, y la manera en la que contribuye a aumentar la biodiversidad en estas comunidades (si lo hace).

Con la idea de contribuir a la preservación de estas comunidades proponemos un marco novedoso para determinar la importancia que tiene cada una de las especies en la comunidad mutualista. Nuestro algoritmo está inspirado en el Page-Rank de Google, pero tienen un carácter no-lineal, y le hemos dado el nombre de MusRank. Además proponemos también un criterio para determinar la calidad de los rankings en estas redes: los buenos rankings llevan a un rápido colapso de la red cuando las especies son eliminadas en ese orden. En la amplia mayoría de las redes que analizamos el uso de nuestro algoritmo genera un ranking que supera a todos los otros que hemos usado como comparación. El ordenamiento emergente permite identificar la importancia de cada especie de una manera eficiente, significativa y robusta [P4]. Este nuevo enfoque , que introducimos aquí por vez primera en el contexto de redes mutualistas, podría ser útil para preservar la biodiversidad y gestionar los ecosistemas, en el caso de que haya que decidir que especies deben protegerse más. Además, si representamos las matrices con las especies ordenadas según el ranking del algoritmo las matrices muestran una forma de alto empaquetamiento, indicando que quizás habría que considerar otras maneras de medir el anidamiento. En particular sugerimos que las calculadoras de anidamiento deberían usar este algoritmo para hacer evidente la estructura anidada de estas redes. Hay que tener en cuenta que la existencia de este tipo de estructuras está en el corazón de nuestro algoritmo. Aunque demostramos que la mayor parte de la nestedness es una contribución de la heterogeneidad, el ordenamiento por grado es claramente inferior al que genera MusRank. Estaría también interesante probar si esta técnica funciona también en redes sin este tipo de estructura.

Aún considerando que el enfoque sistémico ha resultado muy exitoso en diversos campos, los sistemas ecológicos y biológicos aún suponen un reto muy grande. Esto tiene dos razones fundamentales. Por una parte las interacciones en estos modelos no son, en general, bien conocidas, y aunque sea verdad que las propiedades emergentes son, hasta cierto punto, independientes de los detalles, un mejor entendimiento de las

interacciones individuales es crucial para avanzar hacia resultados más predictivos. Por otra parte estos sistemas son muy "interactivos", lo que significa que implican a muchos agentes y de formas diversas. Los ecosistemas, por ejemplo, estan compuestos por un gran número de especies, interaccionando de diferentes formas: predación mutualismo, parasitismo,etc. Aquí hemos considerado dos de esas interacciones por separado, aunque en los ecosistemas reales todas ellas están interrelacionadas, formando un sistema "más-complejo", si cabe. En este aspecto, el uso de redes multi-capa (multiplex) prodía ser de gran utilidad. Aparte de todo esto, tenemos que recordar que los sistemas con los que estamos tratando son todos, en realidad, dinámicos, y que por tanto estudiar su "foto fija" es como estudiar su promedio temporal. Aunque ha probado ser muy útil, uno de los mayores problemas del enfoque con redes "estáticas" es que buena parte de la fenomenología interesante puede depender de un comportamiento dinámico, y por tanto se perderá con esta visión. Aparte, estos sistemas son el resultado de un proceso evolutivo. Los ecosistemas son el resultado de millones de años de cambio. Las redes biológicas son resultado de diferentes procesos evolutivos. Es lógico pensar, por tanto, que para comprender realmente como estos sistemas aparecen y toman forma deberían intentarse enfoques co-adaptativos y que tengan en cuenta la evolución que pueden sufrir las estructuras, puesto que es probable que sin ello haya efectos que no podamos explicar. Saber cómo emerge la topológica de los diferentes sistemas biológicos y ecológicos y como afecta a sus dinámicas es algo de la mayor importancia. En este trabajo hemos intentado aportar algo a este respecto, al igual que muchos otros componentes de la comunidad científica. Afortunadamente parece que estamos en una era dorada de la interdisciplinariedad, y que el reduccionismo y la compartimentación del conocimien ha sido dejado atrás . Ahora la transmisión de conocimiento entre disciplinas y los enfoques mas integradores parecen estar ganando. Los tiempos están cambiando, esperemos que la financiación acompañe.

Appendix A

Devising an interaction matrix from an adjacency matrix

Let us assume that we have a set of ordinary differential equations governing the evolution of the population of each species in an ecosystem, as measured, for instance, by its total biomass x_i. In vector form, we can write this as

$$\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x}).$$

The dynamics will have a fixed point at any configuration $\mathbf{x}^*$ such that $\mathbf{f}(\mathbf{x}^*) = \mathbf{0}$. Let us suppose that the system is placed at this fixed point but suffers a small perturbation $\boldsymbol{\zeta}(t)$:

$$\mathbf{x}(t) = \mathbf{x}^* + \boldsymbol{\zeta}(t).$$

For small enough $|\boldsymbol{\zeta}(t)|$, its dynamics will be given by the linearised equation:

$$\frac{d}{dt}\boldsymbol{\zeta}(t) = J(\mathbf{x}^*)\boldsymbol{\zeta}(t),$$

where $J(\boldsymbol{x}^*)$ is the Jacobian matrix $[\partial f_i/\partial x_j]$ evaluated at $\boldsymbol{x}^*$. The fixed point will be locally stable if all the eigenvalues of $J(\boldsymbol{x}^*)$ have negative real part [103].

Let us consider a fairly general dynamics for $\boldsymbol{x}^*$ given by a consumer-resource model:

$$\frac{d}{dt}x_i = \eta_{ij}\sum_j a_{ij}F(x_i, x_j) - \sum_j a_{ji}F(x_j, x_i) + G(x_i). \tag{A.1}$$

The first term on the right accounts for the increment in species i's biomass through consumption of its resources, the second term is the biomass lost to its consumers, and the function G represents any factors which are not due to interaction with other species. Since we are interested here in effects of interactions between species, we shall simply assume $G(x) = \gamma x$ with γ a constant. The function F describes how the interaction between a consumer and a resource species depends on their respective biomasses. The parameter η is the efficiency of predation – the proportion of biomass lost by a resource which goes on to form part of the consumer. We shall in general consider this parameter to be constant for all pairs of species ($\eta_{ij} = \eta$, $\forall i, j$), but in Sections 3.3.2 and 3.3.3 we look into the effects of varying its value. In the rest of this work we set this parameter to $\eta = 0.2$.

The Jacobian, J, will be obtained by taking the partial derivatives of Eq. (A.1), for each i, with respect to each x_j.

In the simple case where the interaction between species is given by a sum,

$$F(x_i, x_j) = x_i + x_j,$$

we have

$$J_{ij} = (\eta a_{ij} - a_{ji})(1 + \delta_{ij}) + \gamma\delta_{ij},$$

where δ_{ij} is the Kronecker delta (equal to one when $i = j$, or else zero). Positive terms added to or subtracted from the main diagonal of J simply shift its spectrum of eigenvalues to the right or left, respectively. Therefore, we concentrate on the matrix

$$W = \eta A - A^T, \tag{A.2}$$

where A^T is the transpose of A, and consider λ_1, the eigenvalue of W with the largest real part. Then, $R = Re(\lambda_1)$ can be regarded as a measure of the minimum degree of self-regulation at each node which this dynamics would require in order for the system to be stable. In other words, the smaller R, the more stable we shall say the system is.

In this simple case defined by $F(x_i, x_j) = x_i + x_j$ the Jacobian is independent of the point $\boldsymbol{x}^*$ where it is evaluated. However, this will not, in general, be the case and for other dynamics we would need to specify this point in order to characterise the stability of the system. For instance, in a generalised Lotka-Volterra dynamics, the interaction

is proportional to the biomass of both consumer and resource,

$$F(x_i, x_j) = x_i x_j,$$

and the Jacobian becomes

$$J_{ij} = (1 + \delta_{ij}) w_{ij} x_i + \gamma \delta_{ij}, \tag{A.3}$$

where w_{ij} are the elements of the matrix W as given by Eq. (3.1). Note that this expression depends on the biomass of species i (though not on j's) at the point of interest.

To capture the nonlinearities expected in a prey species' functional response, consumer-resource models often describe the interaction as

$$F(x_i, x_j) = x_i H(x_j),$$

where H is the Hill equation,

$$H(x) = \frac{x^h}{x_0^h + x^h},$$

with x_0 the half-saturation density. The Hill coefficient h determines whether the functional response is of type II ($h = 1$) or type III ($h = 2$) [208]. Now we find that the Jacobian is

$$J_{ij} = [\tilde{\eta}(x_i, x_j) a_{ij} - a_{ji}] H(x_i) \tag{A.4}$$

if $i \neq j$, where the effective efficiency of predation is

$$\tilde{\eta}(x_i, x_j) = \frac{x_i}{H(x_i)} \frac{\partial H(x_j)}{\partial x_j} \eta = \frac{h x_0^h x_i}{x_j^{h+1}} \frac{H(x_j)^2}{H(x_i)} \eta,$$

and, for the main diagonal elements,

$$J_{ii} = \{h[1 - H(x_i)] + 1\} H(x_i) w_{ii} + \gamma.$$

In each of these kinds of dynamics it is necessary to evaluate the Jacobian at a particular point: Equations (A.3) (Lotka-Volterra) and (A.4) (types II and III) are similar in

form to the matrix W of Eq. (3.1), but their terms are modified by the biomass of the predator, or the biomasses of both prey and predator, respectively. One might suggest that we only need identify a fixed point and evaluate the equations there. But, in general, a feasible fixed point (in which $x_i > 0$ for all i) will not exist. Feasible fixed points could be defined by attributing weights to the elements of the interaction matrix A, but this would involve decisions on how to do this in a realistic way which might render the results somewhat arbitrary. (For a discussion on the feasibility of fixed points, see Section 3.3.3.)

Throughout most of chapter two we focus simply on the matrix W as given by Eq. (3.1), for although the dynamics it describes exactly is not very realistic (corresponding to the interaction term $F(x_i, x_j) = x_i + x_j$ in Eq. (A.1)), it captures the essential behaviour of better motivated dynamics without requiring any assumptions about the fixed point. In fact, if all species had the same biomass at the fixed-point, then Eqs. (A.3) (Lotka-Volterra) and (A.4) (types II and III) would also reduce to the matrix W as given by Eq. (3.1), for an appropriate choice of the parameter η. However, so as to test the robustness of our results to details of the dynamics, in Section 3.1.1, 3.3.1, and 3.4.1 we look into the effects of different distributions of biomass according together with Lotka-Volterra, type II or type III dynamics. We find that the relationship between trophic coherence and stability reported in chapter two is robust to these considerations, although the dependence of biomass on trophic level introduces interesting effects, in particular for the complexity-stability scaling.

In chapter two text we describe how stability in directed networks (and food webs in particular) is determined to a large extent by their trophic coherence. In Fig. 3.14 we compare the predictions of each of the food-web models described in Appendix B for each of the food webs listed in Table 3.4. Another network feature which influences stability, as mentioned above, is the existence of self-links (representing cannibalism, in the case of food webs), since this is a form of self-regulation. We disentangle this effect from that of trophic coherence, we remove all self-links from the food webs and again measure the real part of the leading eigenvalue, R_{nc}. The predictions of each model are shown in Fig. 3.15.

Appendix B

Mimicking reality: foodweb modeling

The Cascade model

This was the first attempt to show how networks with a structure, in some sense, similar to real food webs could come about via simple rules. Cohen and Newman studied the proportion of basal, intermediate and top species composing the networks and concluded that it was necessary to suppose the existence of an ordering (or cascade) of species that constrains the possible predators and prey of each species. Is under this assumption that they put forward the cascade model [53] to explain "the phenomenology of observed food web structure, using a minimum of hypotheses". This model is based on two parameters: number of species S and link density L/S, and it distributes species and feeding links stochastically, with two simple constraints: species are randomly placed in a one-dimensional feeding hierarchy axis, and species can only feed on species that are lower in the hierarchy than they are, as shown in Figure B.1. As is straightforward feeding cycles or cannibalism are therefore not covered by this model.

Stouffer and co-workers later modified this model so that the number of prey would be drawn from the Beta distribution used by the Niche Model (see below), and called the new version the Generalized Cascade Model [235]. Since this amendment improves the model's predictions as regards distributions of prey and predators (without, to the best of our knowledge, involving any drawbacks), throughout this work we use the

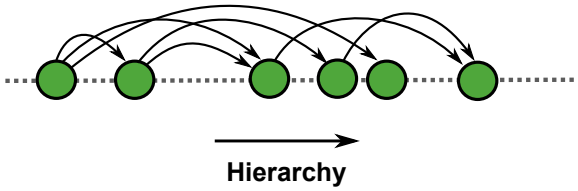

Figure B.1: Schematic representation of the Cascade model. Nodes are randomly placed along the hierarchical axis, and they feed *only* in those below them, which renders completely linear food webs.

Generalized Cascade Model.

Implementation details: In the Cascade model each species i is assigned a random number n_i drawn from a uniform distribution between 0 and 1. For any pair (i, j), we set i to be a consumer of j with a constant probability p if $n_i > n_j$, and with probability zero if $n_i \leq n_j$. With S species, we obtain an expected number of links L if we set

$$p = \frac{2L}{S(S-1)}.$$

The Niche model

Based in the idea of an underlying order and a one-dimensional niche proposed by Cohen [51] a new food-web model was proposed by Williams and Martinez [270], that solved some problems of the previous one (as the assumption of link-species scaling, the exclusion of looping, and the lack of trophic overlap among species).

This is an elegant way of generating non-trivial network topologies by randomly assigning each species a position on a "niche axis", together with a range of axis centered at some lower niche value. Each species then consumes all other species lying within its range, and none without. The rationale behind this model is that food webs were thought to be *interval* – i.e., the species could be arranged in an ordering such that the prey of any given predator were contiguous [52] (see sec. 2.1.5). The Niche Model achieves this by construction. More recent analysis have shown that food webs are

not, in general, perfectly interval, although they do usually exhibit a certain degree of intervality [43, 236]. Nevertheless, the Niche Model has been tremendously successful, since it outperforms the Cascade Model in approximating measurable features of food webs, and even compares well to more elaborate models which take the Niche Model as a basis [271]. It is still the most commonly used model whenever synthetic networks similar to food webs are required, and in any case, the niche model showed that food-web structure was far from random.

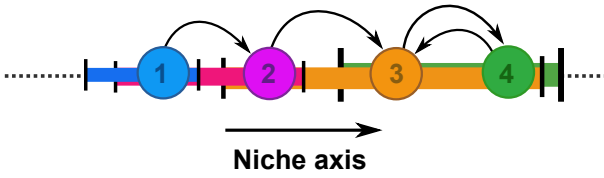

Figure B.2: Schematic representation of a construction of a foodweb with the Niche model: species are placed randomly along the niche axis. Each species is assigned a range r_i, centered at c_i and represented by a strip with the same color of the node, in which it will feed. The species will prey upon all the nodes which n_i falls within its predation range, and none without. Cannibalism and feeding cycles are now allowed.

simulation details: in the Niche Model, each species i is awarded a niche value n_i as in the Cascade Model [270]. However, instead of choosing species with lower niche values randomly for prey, i is constrained to consume the subset of species j such that $c_i - r_i/2 \leq n_j < c_i + r_i/2$ – i.e., all those lying on an interval of the niche axis of size r_i and centred at c_i, and none without. The range is defined as $r_i = x_i n_i$, where x_i is drawn from a Beta distribution with parameters $(1, \beta)$. For S species and a desired number of links L, we must set

$$\beta = \frac{S(S-1)}{2L} - 1.$$

The centre of the interval c_i is drawn from a uniform distribution between $r_i/2$ and $min(n_i, 1 - r_i/2)$.

The Nested Hierarchy Model

The Nested Hierarchy Model provides a way to take into account that phylogenetically similar species should have prey in common. The model, put forward by Cattin et. al [47], is based on the hypothesis that any species' diet is the consequence of both phylogenetic constraints and adaptation. In this way, the two-step assignation of links (see below) seeks to model, on one part, the phylogenetical constraints (two consumers that share prey are assumed to be phylogenetically related) and on the other, the effects of independent adaptation (the links assigned at random). We find this is particularly interesting model because phylogenetic constraints should indeed be taken into account. One problem we find with the Nested Hierarchy Model, however, is that a given species i is assumed to be related to a certain set A of species which share common prey with i; but i will also belong to the set B of common prey of a different set of consumers, and nothing constrains A and B to overlap. In other words, the species related to i due to its prey are not the ones related to i due to its predators, whereas in nature it is to be expected that phylogenetically similar species should have both prey and predators in common. In fact, it has recently been reported that common predators are statistically more significant than common prey [171].

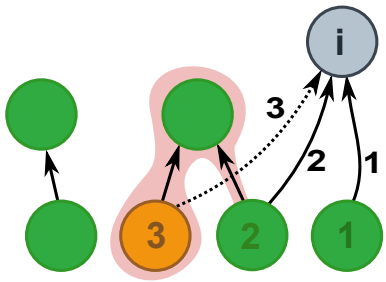

Figure B.3: Illustration of food web assembly using the Nested Hierarchy model. In this case we have a species with 3 potential preys. The first prey is randomly selected from the network, establishing the first link -labeled as 1-. To establish the second link, since the first prey did not have any other predator, we have to randomly select a prey again, stablishing the second link -labeled as 2. In order to asses the third link, now we look into the species that are being preyed upon by the predators of species 2, highlighted in a coloured background, and choose the third prey within the set. This aims to mimic the phylogenetic signal known to exist among predators who share prey.

simulation details: This modeling framework gives each species a niche value and a range, exactly as in the Niche Model. However, instead of establishing links directly to species within the range, first the number of prey to be consumed by each species is determined, in proportion to the range, $k_i^{in} \propto r_i$, so as to generate an expected number of links L. These links are then attributed in the following way. The species with lowest niche value has no prey, while the one with the highest has no predators (so there is always at least one basal species and one apex predator). Starting from the species with second smallest niche value and going up in order of n, we take each species i and apply the following rules to determine its k_i^{in} prey:
1. We choose a random species j already in the network (so $n_j \leq n_i$) and set it as the first prey species of i.
2. If j has no predators other than i, we repeat 1 until either the chosen prey does have other predators, or we reach k_i^{in}. Else we go to 3.
3. We determine the set of species which are prey to the predators of j. We select, randomly, species from this set to become also prey of i until we either complete k_i^{in}, or we go to 4.
4. We continue choosing prey species randomly from among those with lower niche values. If we still have not reached k_i^{in} when these run out, we continue choosing them randomly from those with higher niche values.

The Generalized Niche Model

The Generalized Niche Model was proposed to account for the fact that empirical food webs turned out not to be maximally interval, as predicted by the Niche Model [236]. This model allows for tunable prey contiguity, so only a proportion $1 - c$ of the prey are chosen from the fixed interval, while the remaining fraction c are randomly chosen from among all the species further down the axis. It is therefore a combination of the Niche Model and the Cascade Model, with the contiguity parameter, c, determining the relative importance of each mechanism. The Generalized Niche Model has been shown to emulate real food webs very successfully, at least as regards certain features, such as community structure [95]. It is also often used as a convenient model for generating synthetic networks with a view to studying foodwebs *in silico* [93].

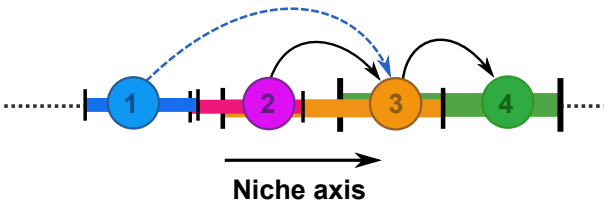

Figure B.4: Schematic representation of a food web constructed with the Generalized Niche model. In this case the *reduced* range, r_i, is represented by a strip of the same color of the node. All species lying in the predation range of any other will be preyed upon. However now some will be randomly selected among the ones with lower niche values (blue dashed line).

simulation details: the Generalised Niche Model is implemented as the Niche model but with reduced ranges $r_i = cx_i n_i$. Then, for each species, the number of extra prey $k_i^{cascade} = (1-c)x_i n_i S$ is drawn randomly from among the available species with niche values lower than n_i, as in the Generalized Cascade Model. For $c = 1$ we have the Niche Model, while $c = 0$ results in the Generalized Cascade Model.

The Minimum Potential Niche Model

The Minimum Potential Niche Model is similar to the Generalized Niche Model in that it is a modification of the Niche Model which breaks up complete intervality by means of a parameter, f [10]. However, the motivation is slightly different. The idea is that in reality there is more than one niche dimension constraining possible predation links (hence the lack of complete, one-dimensional intervality), which implies that some of the links determined by the Niche Model are actually "forbidden links". The species are all allocated niche values n_i and ranges $r_i = x_i n_i$ as in the Niche Model. The species at the extremes of this range are always consumed. However, the rest is considered a potential range and the β parameter used in the Beta distribution from which x_i is drawn is now

$$\beta = \frac{S(S-1)}{2(L+F)} - 1,$$

where $F = fP$, P being the total number of potential links given the ranges, minus the species at the extremes. Once all the species have their ranges, each species within will

be consumed with a probability $1 - f$. Therefore, $f = 0$ results in the original Niche Model, but $f > 0$ produces a proportion of forbidden links.

Allesina *et al.* suggested a framework for comparing niche-based models [10]; they computed the likelihood that the Cascade, Niche and Nested Hierarchy models have of generating the links in a set of ten real food webs, and found theirs (the Minimum Potential Niche Model) to be superior – and, in fact, the only one capable of generating all the observed links.

Appendix C

Possible amendments to the PPM

In chapter two, where we propose the Preferential Preying model as a method to generate synthetic networks with tunable trophic coherence, we were only interested in the effect this feature has in foodweb stability, and hence used the simplest version of the model. However, many of the details are somewhat arbitrary, and several possible amendments and generalisations spring to mind:

- **Basal species.** All the niche-based models discussed allow the number of producers, B, to emerge freely (although they are not, generally, particularly successful in predicting B [271]). We chose here to begin with a set number of basal species, as in the Preferential Attachment Model [23]. We imagine that for most applications where synthetic networks are required it would be useful to have control over this parameter (which is itself related to trophic coherence, as we show in Section 3.1.1). However, if a freely emerging B were preferred – for instance, for a rigorous comparison against models which do not allow this value to be set easily – it is straightforward to take the minimum κ_i equal to zero for incoming species, thereby allowing a proportion of them to become producers.

- **Numbers of prey.** We have drawn the number of prey for each incoming species from a Beta distribution, as in all the niche-based models, because Stouffer *et al.* [235] have shown that this method yields a particularly good fit to food-web data (we have also verified that this holds true for our 46 food-web dataset). However, were the model to be applied to systems other than food webs, it may be preferable

to use, for instance, a Poisson or a Pareto distribution, depending on the in-degree distributions of the networks to be emulated.

- **Boltzmann factor.** The functional form we have used to determine the second and subsequent prey of an incoming species (an exponential in the trophic distance divided by the parameter T) is arbitrary; careful fitting to data may suggest a better function. There is also no reason other than simplicity to use the same value of T for each incoming species: one could also draw a different value T_i for each incoming species form some distribution, perhaps dependent on the trophic level of its first prey.

- **Cycles**. Directed loops in food webs are relatively rare, yet often present. The PPM as described does not generate any of these cycles, but it could easily be amended to do so by assigning each incoming species a small number of predators as well as prey from amongst the species already in the network. However, directed loops require some predators to consume prey at higher trophic levels than theirs, so the more coherent a network, the fewer directed loops are to be expected.

- **Phylogeny and body size.** In this simple incarnation, the PPM ignores the main effects that most of the other models are based on, but these could be taken into account in a "Generalized Preferential Preying Model". Something akin to a phylogenetic signal could be induced by introducing a bias in the Boltzmann factor such that an incoming node tended to copy the prey and predators of a randomly chosen species already in the network – perhaps limiting in the Nested Hierarchy Model in the case where only prey are copied. The Niche, Generalized Niche and Minimum Potential Niche models assume that the niche ordering (usually thought to represent body size, possibly in combination with other biological features) to some extent constrains species to find prey within closed intervals thereof. A bias could likewise be introduced in the Boltzmann factor of the PPM such that intervals of the sequence of entry were preferred, if this constraint in empirical networks turned out to be more than a spurious effect of trophic coherence.

Appendix D

Analytical theory for maximally coherent networks

Let us consider a maximally coherent network, with $q = 0$. The S species will thus fall into M discrete trophic levels, with m_i species in each level i, so that the number of basal species is $B = m_1$, and $S = \sum_{i=1}^{M} m_i$. Each link of the predation (or *adjacency*) matrix A will lead from a prey node at some level i to a predator node a level $i + 1$. The interaction matrix $W = \eta A - A^T$ (where the efficiency η is assumed equal for all pairs of species) will therefore be an $S \times S$ block matrix where the only nonzero blocks are those above and below the main diagonal:

$$W = \begin{pmatrix} 0 & \eta A_1 & 0 & \dots & 0 & 0 \\ -A_1^t & 0 & \eta A_2 & \dots & 0 & 0 \\ 0 & -A_2^t & 0 & \dots & 0 & 0 \\ & & \dots\dots\dots\dots\dots & & & \\ 0 & 0 & 0 & \dots & 0 & \eta A_{S-1} \\ 0 & 0 & 0 & \dots & -A_{S-1}^t & 0 \end{pmatrix}. \tag{D.1}$$

Blocks A_i are $m_i \times m_{i+1}$ matrices representing the links between the species at level i and those at level $i + 1$.

Let us now consider the adjacency matrix $\tilde{A}$ of the undirected network we obtain

by replacing each directed link (or arrow) in A with an undirected (symmetric) one:

$$\tilde{A} = \begin{pmatrix} 0 & A_1 & 0 & \dots & 0 & 0 \\ A_1^t & 0 & A_2 & \dots & 0 & 0 \\ 0 & A_2^t & 0 & \dots & 0 & 0 \\ & & \dots\dots\dots\dots & & & \\ 0 & 0 & 0 & \dots & 0 & A_{S-1} \\ 0 & 0 & 0 & \dots & A_{S-1}^t & 0 \end{pmatrix}. \tag{D.2}$$

The eigenvalues $\{\mu_i\}$ of $\tilde{A}$ are all real since the matrix is symmetric. Furthermore, for every non-negative eigenvalue $\mu_j \geq 0$ there is another eigenvalue $\mu_l = -\mu_j$ since the network is bipartite (species can be partitioned into two groups with no links within each of them: species in even trophic levels and species in odd levels). Therefore, the eigenvalues of $\tilde{A}^2$ are either positive and doubly degenerate or zero. Moreover, the matrix $\tilde{A}^2$ can be written as:

$$\tilde{A}^2 = \begin{pmatrix} D_1 & 0 & B_1 & 0 & \dots \\ 0 & D_2 & 0 & B_2 & \dots \\ B_1^t & 0 & D_3 & 0 & \dots \\ 0 & B_2^t & 0 & D_4 & \dots \\ & & \dots\dots\dots & & \end{pmatrix}. \tag{D.3}$$

where

$$D_i = \begin{cases} A_1 A_1^t & \text{for } i = 1 \\ A_{i-1}^t A_{i-1} + A_i A_i^t & \text{for } 1 < i < M \\ A_{M-1}^t A_{M-1} & \text{for } i = M, \end{cases} \tag{D.4}$$

$$B_i = A_i A_{i+1}.$$

Now, the square of matrix W reads:

$$W^2 = \begin{pmatrix} -\eta D_1 & 0 & \eta^2 B_1 & 0 & \dots \\ 0 & -\eta D_2 & 0 & \eta^2 B_2 & \dots \\ B_1^t & 0 & -\eta D_3 & 0 & \dots \\ 0 & B_2^t & 0 & -\eta D_4 & \dots \\ & & \dots\dots\dots\dots & & \end{pmatrix}. \tag{D.5}$$

We introduce a diagonal matrix U with diagonal blocks

$$U_{ii} = (-\eta)^{\left\lfloor \frac{i-1}{2} \right\rfloor} I_i, \tag{D.6}$$

where I_i is the identity matrix of size m_i, and $\lfloor x \rfloor$ denotes the floor function of x:

$$U = \begin{pmatrix} I_1 & 0 & 0 & 0 & 0 & \dots \\ 0 & I_2 & 0 & 0 & 0 & \dots \\ 0 & 0 & -\eta I_3 & 0 & 0 & \dots \\ 0 & 0 & 0 & -\eta I_4 & 0 & \dots \\ 0 & 0 & 0 & 0 & \eta^2 I_5 & \dots \\ & & \dots\dots\dots\dots & & & \end{pmatrix}. \tag{D.7}$$

We can write

$$W^2 = -\eta U^{-1} \tilde{A}^2 U. \tag{D.8}$$

Therefore, the eigenvalues of W^2 can be obtained by multiplying those of $\tilde{A}^2$ by $-\eta$: they are either negative and doubly degenerate or zero. Denoting by λ_j the eigenvalues of W, we can write

$$\lambda_j^2 = -\eta \mu_j^2. \tag{D.9}$$

This means that for every $\mu_j = 0$ we have $\lambda_j = 0$, and for every pair of real eigenvalues $\pm\mu_j$ of $\tilde{A}$ there is a pair of imaginary eigenvalues $\lambda_j = \pm \mathrm{i}\sqrt{\eta}\mu_j$ of W. In any case, for $\eta > 0$, all the eigenvalues of the interaction matrix W have zero real part. If $\eta = 0$ all its eigenvalues would be zero, while for $\eta < 0$, the imaginary parts would vanish and all the eigenvalues would be real, all the nonzero ones coming in pairs $\lambda_j = \pm\sqrt{|\eta|}\mu_j$.

Appendix E

Degree-degree correlations and nestedness in heterogeneous networks

It is possible to provide an analytical connection between disassirtativity and nestedness in random networks with explicitly built-in degree-degree correlations. It has been recently shown [115] that there is a mapping between any mean-nearest-neighbour function $\overline{k_{nn}}(k)$ –accounting for degree-degree correlations– and its corresponding mean-adjacency-matrix $\hat{\epsilon}$, which is as follows:

$$\overline{k_{nn}}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} + \int d\nu f(\nu) \sigma_{\nu+1} \left[\frac{k^{\nu-1}}{\langle k^\nu \rangle} - \frac{1}{k} \right]. \tag{E.1}$$

This can be seen as an expansion of $\overline{k_{nn}}(k)$ in powers of k with some weight function f and $\sigma_{\nu+1} \equiv \langle k^{\nu+1} \rangle - \langle k \rangle \langle k^\nu \rangle$ (which can always be done [115]), the corresponding matrix $\hat{\epsilon}$ takes the form

$$\hat{\epsilon}_{ij} = \frac{k_i k_j}{\langle k \rangle N} + \int d\nu \frac{f(\nu)}{N} \left[\frac{(k_i k_j)^\nu}{\langle k^\nu \rangle} - k_i^\nu - k_j^\nu + \langle k^\nu \rangle \right]. \tag{E.2}$$

Without entering here the details of this decomposition (for which we refer the reader to Ref. [115]) let us just remark that the first term in Eq.(E.2) coincides with the expected value for the standard configuration model, while the second one accounts

for correlations. Hence, Eq.(E.2) can be seen as an extension of the configuration model including correlations, i.e. a *correlated configuration model*. In particular, Eq.(E.2) encodes the way a network should be wired (i.e. the probabilities with which any pair of nodes should be connected) to have the desired degree sequence and degree-degree correlations.

In many empirical scale-free networks, $\overline{k_{nn}}(k)$ can be fitted by $\overline{k_{nn}}(k) = A + Bk^{\beta}$, with $A, B > 0$ [40, 44, 193] – the mixing being assortative (disassortative) if β is positive (negative). Such a case is described by Eq. (E.1) with $f(\nu) = C[\delta(\nu - \beta - 1)\sigma_2/\sigma_{\beta+2} - \delta(\nu - 1)]$, with C a positive constant, which simplifies significantly the expressions above. This choice yields

$$\overline{k_{nn}}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} + C\sigma_2 \left[\frac{k^{\beta}}{\langle k^{\beta+1} \rangle} - \frac{1}{\langle k \rangle} \right] \tag{E.3}$$

After plugging Eq. (E.3) into Eq.(5.8) one obtains:

$$r = \frac{C\sigma_2}{\langle k^{\beta+1} \rangle} \left(\frac{\langle k \rangle \langle k^{\beta+2} \rangle - \langle k^2 \rangle \langle k^{\beta+1} \rangle}{\langle k \rangle \langle k^3 \rangle - \langle k^2 \rangle^2} \right). \tag{E.4}$$

It turns out that the configurations most likely to arise naturally (i.e those with maximal entropy) usually have $C \simeq 1$ [115]. Therefore, and for the sake of analytical simplicity, we shall consider this particular case (note that $C = 1$ corresponds to removing the linear term, proportional to $k_i k_j$, in Eq. (E.2), and leaving the leading non-linearity, $(k_i k_j)^{\beta+1}$, as the dominant one); that is, we shall use

$$\hat{\epsilon}_{ij} = \frac{1}{N} \left\{ \frac{\sigma_2}{\sigma_{\beta+2}} \left[\frac{(k_i k_i)^{\beta+1}}{\langle k^{\beta+1} \rangle} - k_i^{\beta+1} - k_j^{\beta+1} + \langle k^{\beta+1} \rangle \right] + k_i + k_j - \langle k \rangle \right\}. \tag{E.5}$$

Substituting the adjacency matrix for this expression in the definition of η in Eq.(5.7) , we obtain its expected value as a function of the remaining parameter β:

$$\overline{\eta}(\beta) = \frac{\langle k \rangle^2}{\langle k^2 \rangle} \left[1 + (\sigma_2 - \alpha_\beta^2 \rho_\beta) \left(2 \frac{\langle k^{\beta} \rangle \langle k^{-1} \rangle}{\langle k^{\beta+1} \rangle} - \langle k^{-1} \rangle^2 \right) + \alpha_\beta^2 \rho_\beta \left(\frac{\langle k^{\beta} \rangle}{\langle k^{\beta+1} \rangle} \right)^2 \right], \tag{E.6}$$

where $\alpha_\beta \equiv \sigma_2/\sigma_{\beta+2}$ and $\rho_\beta \equiv \langle k^{2(\beta+1)} \rangle - \langle k \rangle^{2(\beta+1)}$. Note that $\overline{\eta}_0 = 1$, as corresponds to uncorrelated networks. As r can be inferred from β using Eq.(E.4), then we can plot

the resulting η as a function of r for different networks. In particular, for scale-free networks with $P(k) \sim k^{-gamma}$ we obtain the curves shown in Fig. E.1; they exhibit a clear tendency (at least for $\gamma > 2$): disassortative networks tend to be nested and the other way around.

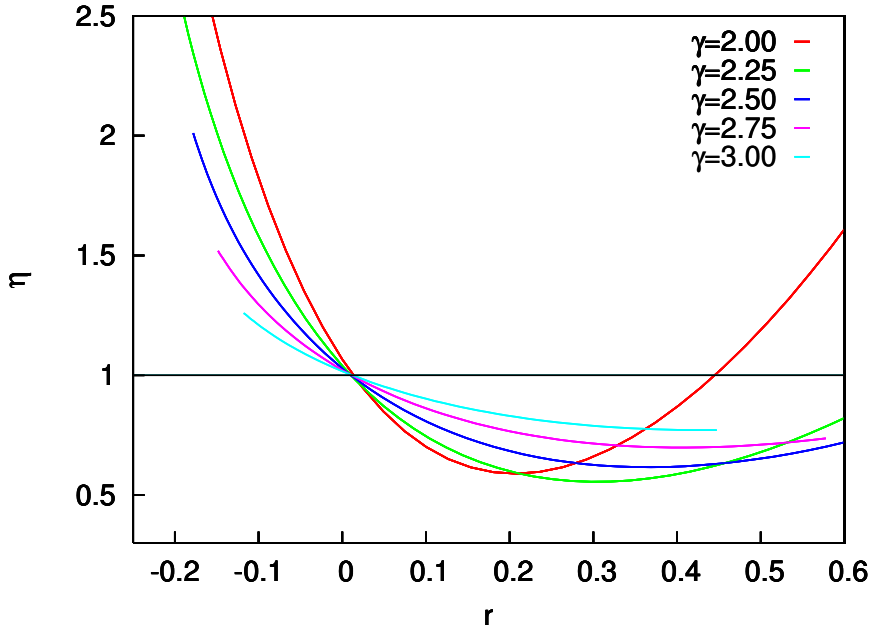

Figure E.1: Nestedness against assortativity (as measured by Pearson's correlation coefficient, r) for scale-free networks with different values of the degree-distribution exponent, γ. $\langle k \rangle = 10$, $N = 1000$.

Bibliography

[1] L. G Abarca-Arenas and R. E. Ulanowicz. The effects of taxonomic aggregation on network analysis. *Ecological Modelling*, 149(3):285 – 296, 2002.

[2] L. A. Adamic and N. Glance. The political blogosphere and the 2004 us election. *Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem*, 2005.

[3] A. Agrawal and K. Gopal. *Biomonitoring of Water and Waste Water*. Springer India, 2013.

[4] M. Aigner. *Lattice Paths and Determinants. in Computational Discrete Mathematics*, volume 2122 of *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 2001.

[5] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Rev Mod Phys*, 74:47–97, 2002.

[6] R. Albert, H. Jeong, and A. L. Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[7] S. Allesina and M. Pascual. Network structure, predator–prey modules, and stability in large food webs. *Theoretical Ecology*, 1:55–64, 2008.

[8] S. Allesina and M. Pascual. Googling food webs: Can an eigenvector measure species' importance for coextinctions? *PLoS Comp Biol*, 5(9):e1000494, 2009.

[9] S. Allesina and S. Tang. Stability criteria for complex ecosystems. *Nature*, 483 (7388):205–208, 2012.

[10] S. Allesina, D. Alonso, and M. Pascual. A general model for food web structure. *Science*, 320(5876):658–661, 2008.

[11] J. Almunia, G. Basterretxea, J. Arısteguia, and R. E. Ulanowicz. Benthic-pelagic switching in a coastal subtropical lagoon. *Estuarine, Coastal and Shelf Science*, 49(3):363 – 384, 1999.

[12] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, London, 1 edition, 2006.

[13] U. Alon. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, 8(6):450–461, 2007.

[14] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences USA*, 97(21): 11149–11152, 2000.

[15] D. Angeli, J. E. Ferrell Jr., and E. D. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. USA*, 101:1822–1827, 2004.

[16] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):3 – 153, 2008.

[17] A. Arenas, A. Fernández, and M. Pascual. Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics*, 10:053039, 2008.

[18] M. T. K. Arroyo, R. B. Primack, , and J. J. Armesto. Community studies in pollination ecology in the high temperate andes of central chile. i. pollination mechanisms and altitudinal variation. *American Journal of Botany*, 69:82–97, 1982.

[19] W. Atmar and B. D. Paterson. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, 96:373–382, 1993.

[20] J. W. Baird. The selection and use of fruit by birds in an eastern forest. *Wilson Bulletin*, (92):63–73, 1980.

[21] C. Banasek-Richter, L. Bersier, M. Cattin, R. Baltensperger, J. Gabriel, Y. Merz, R. Ulanowicz, A. Tavares, D. Williams, P. Ruiter, K. Winemiller, and R. Naisbit. Complexity in quantitative food webs. *Ecology*, 90(6):1470–1477, 2009.

[22] C. Banašek-Richter, L. Bersier, M. Cattin, R. Baltensperger, J. Gabriel, and J. Merz, *et al.* Complexity in quantitative food webs. *Ecology*, 90:1470–7, 2009.

[23] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[24] A.L. Barabási. *Linked: The New Science of Networks*. Perseus Books Group, 2002.

[25] A. Barrat, M. Barthelemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, Cambridge, 2008.

[26] S. C. H. Barrett and K. Helenurm. The reproductive-biology of boreal forest herbs.1. breeding systems and pollination. *Canadian Journal of Botany*, 65:2036–2046, 1987.

[27] J. Bascompte and P. Jordano. Plant-animal mutualistic networks: the architecture of biodiversity. *Annual Review of Ecology Evolution and Systematics.*, 38: 567–593, 2007.

[28] J. Bascompte and C. J. Melián. Simple trophic modules for complex food webs. *Ecology*, 86:2868–2873, 2005.

[29] J. Bascompte, P. Jordano, C. J. Melián, and J. M. Olesen. The nested assembly of plant animal mutualistic networks. *Proc. Nat. Acad. Sci.*, 100:9383–9387, 2003.

[30] J. Bascompte, C. Melián, and E. Sala. Interaction strength combinations and the overfishing of a marine food web. 102(15):5443–5447, 2005.

[31] M. Bastian, S.n Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, 2009.

[32] U. Bastolla, M. Laessig, S. Manrubia, and A. Valleriani. Diversity patterns from ecological models at dynamical equilibrium. *J. Theor. Biol.*, 212:11–34, 2001.

[33] U. Bastolla, M. A. Fortuna, A. Pascual-Garcia, A. Ferrera, B. Luque, and J. Bascompte. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458(7241):1018–1020, 2009.

[34] U. Bastolla, M.A. Fortuna, A. Pascual-García, A. Ferrera, B. Luque, and J. Bascompte. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458:1018–21, 2009.

[35] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405:590–593, 2000.

[36] B. Beehler. Frugivory and polygamy in birds of paradise. *The Auk*, 100:1–12, 1983.

[37] E. L. Berlow, A. M. Neutel, J. E. Cohen, P. C. De Ruiter, B. Ebenman, M. Emmerson, J. W. Fox, V. A. A. Jansen, J. Iwan Jones, G. D. Kokkoris, D. O. Logofet, A. J. McKane, J. M. Montoya, and O. Petchey. Interaction strengths in food webs: issues and opportunities. *Journal of Animal Ecology*, 73(3):585–598, 2004.

[38] G. Bianconi and N. Gulbahce. Algorithm for counting large directed loops. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224003, 2008.

[39] N. Blüthgen, N. E. Stork, and K. Fiedler. Bottom-up control and co-occurrence in complex communities: honeydew and nectar determine a rainforest ant mosaic. *Oikos*, 106:344–358, 2004.

[40] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.U. Hwang. Complex networks: Structure and dynamics. *Phys. Rep.*, 424:175, 2006.

[41] U. Brose, R. J. Williams, and N. D. Martinez. Allometric scaling enhances stability in complex food webs. *Ecology Letters*, 9:1228–1236, 2006.

[42] E. Burgos, H. Ceva, R.P.J. Perazzo, M. Devoto, D. Medan, M. Zimmermannd, and A. M. Delbue. Why nestedness in mutualistic networks? *Journal of Theoretical Biology*, 249:307–313, 2007.

[43] J.A. Capitán, A. Arenas, and R. Guimerà. Degree of intervality of food webs: From body-size data to models. *J. Theor. Bio.*, 334:35–44, 2013.

[44] A. Capocci, G. Caldarelli, and P. de los Rios. Quantitative description and modeling of real networks. *Phys. Rev. E.*, 68:047101, 2003.

[45] T. A. Carlo, J. A. Collazo, and M. J. Groom. Avian fruit preferences across a puerto rican forested landscape: pattern consistency and implications for seed removal. *Oecologia*, 134:119–131, 2003.

[46] M. Catanzaro, M. Boguña, and R. Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Phys. Rev. E*, 71:027103, Feb 2005. doi: 10.1103/PhysRevE.71.027103.

[47] M. F. Cattin, L. F. Bersier, C. Banasek-Richter, R. Baltensperger, and J. P. Gabriel. Phylogenetic constraints and adaptation explain food-web structure. *Nature*, 427:835–9, 2004.

[48] M. De Choudhury, H. Sundaram, A. John, and D. D. Seligmann. Social synchrony: Predicting mimicry of user actions in online social media. *Proc. Int. Conf. on Computational Science and Engineering*, pages 151–158, 2009.

[49] R. R. Christian and J. J. Luczkovich. Organizing and understanding a winter's Seagrass foodweb network through effective trophic levels. *Ecol. Model.*, 117: 99–124, 1999.

[50] R. E. Clements and F. L. Long. *Experimental pollination. An outline of the ecology of flowers and insects.* Carnegie Institute of Washington, Washington D.C., USA, 1923.

[51] J. E. Cohen. Food webs and the dimensionality of trophic niche space. *Proc. Natl. Acad. Sci. USA*, 74:4533–4563, 1977.

[52] J. E. Cohen. *Food Webs and Niche Space.* Princeton Univ. Press, Princeton, New Jersey, 1978.

[53] J. E. Cohen and C. M. Newman. A stochastic theory of community food webs I. models and aggregated data. *Proc. R. Soc. London Ser. B.*, 224:421–448, 1985.

[54] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks. *Nature Physics*, 2:110 – 115, 2006.

[55] L. Comtet. *Advanced Combinatorics: The Art of Finite and Infinite Expansions*. D. Reidel Publishing Company, Dodrecht-Holland/Boston U.S.A., 1974.

[56] B. Corominas-Murtra, J. Goñi, R. V Solé, and C. Rodríguez-Caso. On the origins of hierarchy in complex networks. *Proc. Natl. Acad. Sci. USA*, 103:13316–13321, 2013.

[57] M. Cristelli, A. Gabrielli, A. Tacchella, G. Caldarelli, and L. Pietronero. Measuring the intangibles: A metrics for the economic complexity of countries and products. *Plos ONE*, 8, 2013.

[58] F.H.J. Crome. The ecology of fruit pigeons in tropical northern queensland. *Australian Journal of Wildlife Research*, 2:155–185, 1975.

[59] R. Cross and A. Parker. *The Hidden Power of Social Networks*. Harvard Business School Press, Boston, 2004.

[60] J. M. Dambacher, H. K. Luh, H. W. Li, and P. A. Rossignol. Qualitative stability and ambiguity in model ecosystems. *American Naturalist*, 161:876–888, 2003.

[61] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas. Comparing community structure identification. *J. Stat. Mech.*, page P09008, 2005.

[62] C. Darwin. *On the Origin of Species*. John Murray, London,UK, 1859.

[63] E. H. Davidson, J. P. Rast, P. Oliveri, A. Ransick, C. Calestani, C. H. Yuh, T. Minokawa, G.e Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Y. Lee, R. Revilla, A. G. Rust, Z. J. Pan, M. J. Schilstra, P. J. Clarke, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood, and H. Bolouri. A genomic regulatory network for development. *Science*, 295(5560): 1669–1678, March 2002.

[64] Eric Davidson and Michael Levin. Gene regulatory networks. *Proc.s Nat. Acad. Sci. USA*.

[65] D. L. DeAngelis and J. C. Waterhouse. Equilibrium and nonequilibrium concepts in ecological models. *Ecol. Monogr.*, 57:1–21, 1987.

[66] L. V. Dicks, S. A. Corbet, and R. F. Pywell. Compartmentalization in plant–insect flower visitor webs. *J. Anim. Ecol.*, 71:32–43, 2002.

[67] I. Diez, P. Bonifazi, I. Escudero, B. Mateos, MA. Muñoz, S. Stramaglia, and J. M. Cortes. A novel brain partition highlights the modular skeleton shared by structure and function. *ArXiv*, 2014.

[68] L. Donetti and MA. Munoz. Detecting network communities: a new systematic and efficient algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, 10:P10012, 2004.

[69] S. N. Dorogovtsev, A. L. Ferreira, A. V. Goltsev, and J. F. F. Mendes. Zero pearson coefficient for strongly correlated growing treesex network. *Physical Review E*, 81:031135, 2002.

[70] B. Drossel and A. J. McKane. *"Modelling Food Webs", in A Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley-VCH, Berlin, 2003.

[71] J. Duch and A. Arenas. Community identification using extremal optimization. *Physical Review E*, 72:027104, 2005.

[72] J. Dunne, R. Williams, and N. Martinez. Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecol Lett*, 5:558–567, 2002.

[73] J. A. Dunne, U. Brose, R. J. Williams, and N. D. Martinez. Modeling food-web dynamics: complexity-stability implications. *Aquatic Food Webs: An Ecosystem Approach*, pages 117–129, 2005.

[74] J.A. Dunne, R.J. Williams, and N.D. Martinez. Network structure and robustness of marine food webs. *Mar. Ecol. Prog. Ser.*, 273:291–302, 2004.

[75] Y.L. Dupont, DM Hansen, and J.M. Olesen. Structure of a plant-flower-visitor network in the high-altitude sub-alpine desert of tenerife, canary islands. *Ecography*, 26:301–310, 2003.

[76] A. Eklöf, U. Jacob, J. Kopp, J. Bosch, R. Castro-Urgal, B. Dalsgaard, N. Chacoff, C. deSassi, M. Galetti, P. Guimaraes, S. Lomáscolo, A. Martín González, M.A. Pizo, R. Rader, A. Rodrigo, J. Tylianakis, D. Vazquez, and S. Allesina. The dimensionality of ecological networks. *Ecology Letters*, 16:577–583, 2013.

[77] H. Elberling and J. M. Olesen. The structure of a high latitude plant-flower visitor system: the dominance of flies. *Ecography*, 22:314–323, 1999.

[78] C. S. Elton. *Animal Ecology*. Sidgwick and Jackson, London, 1927.

[79] C. S. Elton. *Ecology of Invasions by Animals and Plants*. Chapman and Hall, London, 1958.

[80] D. H. Erwin and E. H. Davidson. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet*, 10:141–148, 2009.

[81] E. Estrada. Characterization of 3d molecular structure. *Chemical Physics Letters*, 319(5-6):713–718, 2000.

[82] E. Estrada. Food webs robustness to biodiversity loss: The roles of connectance, expansibility and degree distribution. *J Theor Biol*, 244:296–307, 2007.

[83] J. Fischer and D.B Lindenmayer. Treating the nestedness temperature calculator as a "black box" can lead to false conclusions. *Oikos*, 99:193–199, 2002.

[84] P.G.H. Frost. Fruit-frugivore interactions in a south african coastal dune forest. *R. Noring (ed.) Acta XVII Congresus Internationalis Ornithologici, Deutsches Ornithologische Gessenshaft, Berlin*, pages 1179–1184, 1980.

[85] N. Gulbahce G. Bianconi and A. E. Motter. Local structure of directed networks. *Physical Review Letters*, 100:118701, 2008.

[86] M. Galetti and M.A. Pizo. Fruit eating birds in a forest fragment in southeastern brazil. *Ararajuba, Revista Brasileira de Ornit*, 1996.

[87] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99:7821–7826, 2002.

[88] P. Gleiser and L. Danon. Community structure in jazz. *Adv. Complex Syst*, 6: 565, 2003.

[89] L. Goldwasser and J. A. Roughgarden. Construction of a large Caribbean food web. *Ecology*, 74:1216–1233, 1993.

[90] N.J. Gotelli. Research frontiers in null model analysis. *Global Ecology and Biogeography*, 10:337–343, 2001.

[91] R. Graham, D. Knuth, and O. Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, Boston, 1994.

[92] V. Grimm and C. Wissel. Babel, or the ecological stability discussions: An inventory and analysis of terminology and a guide for avoiding confusion. *Oecologia*, 109:323–334, 1997.

[93] T. Gross, L. Rudolf, S.A. Levin, and U. Dieckmann. Generalized models reveal stabilizing factors in food webs. *Science*, 325:747–50, 2009.

[94] N. Guelzim, S. Bottani, P. Bourgine, and F. Képès. Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet*, 31:60, 2002.

[95] R. Guimerà, D. B. Stouffer, M. Sales-Pardo, E. A. Leicht, M. E. J. Newman, and L. A. N. Amaral. Origin of compartmentalization in food webs. *Ecology*, 91(10): 2941–51, 2010.

[96] A. Hammann and B. Curio. Interactions among frugivores and fleshy fruit trees in a philippine submontane rainforest. 1999.

[97] L. Harriger, M. P. van den Heuvel, and O. Sporns. Rich club organization of macaque cerebral cortex and its role in network communication. *Plos One*, 7: e46497, 2012.

[98] L. Harriger, M. P. van den Heuvel, and O. Sporns. Rich club organization of macaque cerebral cortex and its role in network communication. *PLoS ONE*, 7: e46497, 2012.

[99] Karl Havens. Scale and structure in natural food webs. *Science*, 257(5073):1107–1109, 1992.

[100] J. Herrera. Pollination relatioships in southern spanish mediterranean shrublands. *Journal of Ecology*, 76:274–287, 1988.

[101] A. Hintze and C. Adami. Evolution of complex modular biological networks. *PLoS Comput Biol*, (2):e23, 2008. doi: 10.1371/journal.pcbi.0040023.

[102] B. Hocking. Insect-flower associations in the high arctic with special reference to nectar. *Oikos*, 19:359–388, 1968.

[103] P. Holmes and E. T. Shea-Brown. Stability. *Scholarpedia*, 1:1838, 2006.

[104] M. Huxham, S. Beaney, and D. Raffaelli. Do parasites reduce the chances of triangulation in a real food web? *Oikos*, 76:284–300, 1996.

[105] D. W. Inouye and G. H. Pyke. Pollination biology in the snowy mountains of australia: comparisons with montane colorado, usa. *Australian Journal of Ecology*, 13(191-210), 1988.

[106] D. W. Inouye and G. H. Pyke. Pollination biology in the snowy mountains of australia: comparisons with montane colorado, usa. *Australian Journal of Ecology*, 13(191-210), 1988.

[107] J. Memmott J, N. Waser, and M. Price. Tolerance of pollination networks to species extinctions. *Proc Roy Soc Ser B*, 271:2605–2611, 2004.

[108] U. Jacob, A. Thierry, U. Brose, W.E. Arntz, S. Berg, T. Brey, I. Fetzer, T. Jonsson, K. Mintenbeck, C. Möllmann, O.L. Petchey, J.O. Riede, and J.A. Dunne. The role of body size in complex food webs. *Advances in Ecological Research*, 45: 181–223, 2011.

[109] C. Jacquet, C. Moritz, L. Morissette, P. Legagneux, F. Massol, P. Archambault, and D. Gravel. No complexity-stability relationship in natural communities. *arXiv:1307.5364*, 2013.

[110] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, October 2000.

[111] H. Jeong, S. Mason, A.L. Barabási, and Z. N. Oltvai. Centrality and lethality of protein networks. *Nature*, 411:41, 2001.

[112] H. Jeong, S. P. Mason, Z. N. Oltvai, and A. L. Barabasi. Lethality and centrality in protein networks. *Nature*, 411:41–42, 2001.

[113] A. Joern. Feeding patterns in grasshoppers (orthoptera: Acrididae): factors influencing diet specialization. *Oecologia*, 38:325–347, 1979.

[114] J. C. Johnson, S. P. Borgatti, J. J. Luczkovich, and M. G. Everett. Network role analysis in the study of food webs: An application of regular role coloration. *Journal of Social Structure*, 2, 2001.

[115] S. Johnson, J.J. Torres, J. Marro, and M A. Muñoz. Entropic origin of disassortativity in complex networks. *Phys. Rev. Lett.*, 104:108702, 2010.

[116] F. Jordan, Z. Benedek, and J. Podani. Quantifying positional importance in food webs: A comparison of centrality indices. *Ecol Model*, 205:270–275, 2007.

[117] P. Jordano. El ciclo anual de los paseriformes frugívoros en el matorral mediterráneo del sur de españa: importancia de su invernada y variaciones interanuales. *Ardeola*, 32:69–94, 1985.

[118] P. Jordano. Patterns of mutualistic interactions in pollination and seed dispersal: Connectance, dependence asymmetries, and coevolution. *American Naturalist*, 129:657–677, 1987.

[119] P. Jordano. *Fruits and frugivory.* Seeds: the ecology of regeneration in natural plant communities. Commonwealth Agricultural Bureau International, Wallingford, UK, 2000.

[120] P. Jordano, J. Bascompte, and J. M. Olesen. Invariant properties in coevolutionary networks of plant animal interactions. *Ecology Letters*, 6:69–81, 2003.

[121] C. Kadelka, D. Murrugarra, and R. Laubenbacher. Stabilizing gene regulatory networks through feedforward loops. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(2):025107, 2013.

[122] C. N. Kaiser-Bunbury, S. Muff, J. Memmot, C. B. Muller, and A. Caflisch. The robustness of pollination networks to the loss of species and interactions: a quantitative approach incorporating pollinator behaviour. *Ecology Letters*, 13:442–452, 2010.

[123] M. Kato, T. Makutani, T. Inoue, and T. Itino. Insect-flower relationship in the primary beech forest of ashu, kyoto: an overview of the flowering phenology and seasonal pattern of insect visits. *Contr. Biol. Lab. Kyoto Univ*, 27:309–375, 1990.

[124] M. Kaufman, C. Soulé, and R. Thomas. A new necessary condition on interaction graphs for multistationarity. *Journal of Theoretical Biology*, 248(4):675 – 685, 2007.

[125] P. G. Kevan. *High arctic insect-flower visitor relations: the inter-relationships of arthropods and flowers at Lake Hazen, Ellesmere Island, Northwest Territories, Canada*. Ph.D. thesis thesis, University of Alberta, 1970.

[126] D. E. Knuth. *The Art Of Computer Programming Vol 1. 3rd ed.* Addison-Wesley, Boston, 1997.

[127] A. E. Krause1, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor. Compartments revealed in food-web structure. *Nature*, 426:282, 2003.

[128] V. Krebs. unpublished. *unpublished, http://www.orgnet.com/.*

[129] A. Krishna, P. Guimaraes, P. Jordano, and J. Bascompte. A neutral-niche theory of nestedness in mutualistic networks. *Oikos*, 117:1609–1618.

[130] K. D. Lafferty, R. F. Hechinger, J. C. Shaw, K. L. Whitney, and A. M. Kuris. Food webs and parasites in a salt marsh ecosystem. *Disease ecology: community structure and pathogen dynamics (ed. S. Collinge & C. Ray)*, pages 119–134, 2006.

[131] M. Cosentino Lagomarsino, P. Jona, B. Bassetti, and H. Isambert. Hierarchy and feedback in the evolution of the escherichia coli transcription network. *Proc. Natl. Acad. Sci. USA*.

[132] F. Lambert. Fig-eating by birds in a malaysian lowland rain fores. *J. Trop. Ecol.*, 32(5):401–412, 1989.

[133] S.R Leather. Feeding specialisation and host distribution of british and finnish prunus feeding macrolepidoptera. *OIKOS*, 60:40–48, 1991.

[134] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Governance in social media: A case study of the wikipedia promotion process. *Proc. Int. Conf. on Weblogs and Social Media*, pages 98–105, 2010.

[135] S. Levine. Several measures of trophic structure applicable to complex food webs. *J. Theor. Biol.*, 83:195–207, 1980.

[136] R. Levins. Discussion paper: The qualitative analysis of partially specified systems. *Annals of the New York Academy of Sciences*, 231(1):123–138, 1974.

[137] R. L. Lindeman. The trophic-dynamic aspect of ecology. *Ecology*, 23:399–418, 1942.

[138] J. Link. Does food web theory work for marine ecosystems? *Mar. Ecol. Prog. Ser.*, 230:1–9, 2002.

[139] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54:396–405, 2003.

[140] A. Ma'ayan, G. A. Cecchi, J. Wagner, A. R. Raob, R. Iyengara, and G. Stolovitzky. Ordered cyclic motifs contribute to dynamic stability in biological and engineered networks. *PNAS*, 105:19235–19240, 2008.

[141] R. MacArthur. Fluctuations of animal populations, and a measure of community stability. *Ecology*, 36:533–5, 1955.

[142] A.L. Mack and D. D. Wright. Notes on occurrence and feeding of birds at crater mountain biological research station, papua new guinea. *Emu*, 96:89–101, 1996.

[143] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA*, 100(21):11980–11985, 2003.

[144] N. D. Martinez. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.*, 61:367–392, 1991.

[145] N. D. Martinez, B. A. Hawkins, H. A. Dawah, and B. P. Feifarek. Effects of sampling effort on characterization of food-web structure. *Ecology*, 80:1044–1055, 1999.

[146] S. Maslov, K. Sneppen, and A. Zaliznyak. Detection of topological patterns in complex networks: Correlation profile of the internet. *Physica A*, 333:529–540, 2004.

[147] D.M Mason. Quantifying the impact of exotic invertebrate invaders on food web structure and function in the great lakes: A network analysis approach. *Interim Progress Report to the Great Lakes Fisheries Commission- yr 1*, 2003.

[148] P. Massa, M. Salvetti, and D. Tomasoni. Bowling alone and trust decline in social network sites. *Proc. Int. Conf. Dependable, Autonomic and Secure Computing*, pages 658 –663, 2009. doi: 10.1109/DASC.2009.130.

[149] C. Matias, S. Schbath, E. Birmelé, J. j. Daudin, and S. Robin. Assessing the exceptionality of network motifs. *REVSTAT – Statistical Journal*, 4:31–51, 2006.

[150] R. M. May. Will a large complex system be stable? *Nature*, 238:413, 1972.

[151] R. M. May. Food-web assembly and collapse: mathematical models and implications for conservation. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 364:1643–1646, 2009.

[152] Robert M. May. *Stability and complexity in model ecosystems*. Princeton University Press, Princeton, USA, 1973.

[153] Robert M. May. Qualitative stability in model ecosystems. *Ecology*, 54:638–41, 1973.

[154] J. McAuley and J. Leskovec. Learning to discover social circles in ego networks. *Advances in Neural Information Processing Systems*, pages 548–556, 2012.

[155] J. J. McAuley, L. F. Costa, and T. S. Caetano. The rich-club phenomenon across complex network hierarchies. *Applied Physics Letters*, 91:084103, 2007.

[156] K. S. McCann. The diversity-stability debate. *Nature*, 405:228–33, 2000.

[157] C. K. McCullen. Flower-visiting insects of the galapagos islands. *Pan-Pacific Entomologist*, 69:95–106, 1993.

[158] A. J. McKane and B. Drossel. *Models of food web evolution, in Ecological Networks: Linking Structure to Dynamics in Food Webs. M. Pascual and J.A. Dunne, eds.* Oxford University Press, Oxford, UK, 2006.

[159] D. Medan, N. H. Montaldo, M. Devoto, A. Mantese, V. Vasellati, and N. H. Bartoloni. Plant-pollinator relationships at two altitudes in the andes of mendoza, argentina. *Arctic Antarctic and Alpine Research*, 34:233–24, 2002.

[160] C. J. Melian and J. Bascompte. Food web cohesion. *Ecology*, 85:352–358, 2004.

[161] J. Memmott. The structure of a plant-pollinator food web. *Ecology Letters*, 2: 276–280, 1999.

[162] J. Memmott, N. D. Martinez, and J. E. Cohen. Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural food web. *J. Anim. Ecol.*, 69:1–15, 2000.

[163] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594): 824–827, 2002.

[164] E. G. Mitchell and A-M Neutel. Feedback spectra of soil food webs across a complexity gradient, and the importance of three-species loops to stability. *Theoretical Ecology*, 5(2):153–159, 2012.

[165] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6(2-3):161–180, 1995.

[166] M. E. Monaco and R. E. Ulanowicz. Comparative ecosystem trophic structure of three u.s mid-atlantic estuaries. *Marine Ecology Progress Series*, 161:239–254, 1997.

[167] J. Montoya, S. L.Pimm, and R. Sole. Ecological networks and their fragility. *Nature*, 442:259–264, 2006.

[168] P. Moretti and MA. Muñoz. Griffiths phases and the stretching of criticality in brain networks. *Nature Communications*, 4:2521, 2013.

[169] T. Mosquin and J. E. H. Martin. Observations on the pollination biology of plants on melville island, n.w.t., canada. *Canadian Field Naturalist*, 81:201–205, 1967.

[170] A. F. Motten. *Pollination Ecology of the Spring Wildflower Community in the Deciduous Forests of Piedmont North Carolina.* Doctoral Dissertation thesis, Duke University, Duhram, North Carolina, USA, 1982.

[171] R. E. Naisbit, R. P. Rohr, A. G. Rossberg, P. Kehrli, and L.-F. Bersier. Phylogeny versus body size as determinants of food web structure. *Proc. R. Soc. B*, 279: 3291–7, 2012.

[172] A-M. Neutel, J. A. P. Heesterbeek, and P. C. de Ruiter. Stability in real food webs:weak links in long loops. *Science*, 296:1120–1123, 2002.

[173] A-M. Neutel, J. A. P. Heesterbeek, J. van de Koppel, G. Hoenderboom, A. Vos, C. Kaldeway, F. Berendse, and P. C. de Ruiter. Reconciling complexity with stability in naturally assembling food webs. *Nature*, 449:599–602, 2007.

[174] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, 2002.

[175] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74:036104, 2006.

[176] Mark. E. J. Newman, , S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64:026118, 2001.

[177] M.E.J. Newman. Mixing patterns in networks. *Phys. Rev. Lett.*, 89:208701, 2002.

[178] M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[179] K. Norlen, G. Lucas, M. Gebbie, and J. Chuang. Eva: Extraction, visualization and analysis of the telecommunications and media ownership network. *Proceedings of International Telecommunications Society 14th Biennial Conference, Seoul Korea*, 2002.

[180] B. Novak and J. J. Tyson. Design principles of biochemical oscillators. *Nature Reviews Molecular Cell Biology*, 9(12):981–991, 2008.

[181] T. Okuyama and J. N. Holland. Network structural properties mediate the stability of mutualistic communities. *Ecol. Lett.*, 11:208–216, 2008.

[182] J. M. Olesen, L. I. Eskildsen, and S. Venkatasamy. Invasion of pollination networks on oceanic islands: importance of invader complexes and endemic super generalists. *Diversity and Distributions*, 8:181–192, 2002.

[183] J. M. Olesen, J. Bascompte, Y. L. Dupont, and P. Jordano. The modularity of pollination networks. *Proceedings of the National Academy of Sciences*, 104(50): 19891–19896, 2007.

[184] J. Ollerton, S. D. Johnson, L. Cranmer, and S. Kellie. The pollination ecology of an assemblage of grassland asclepiads in south africa. *Annals of Botany*, 92: 807–834, 2003.

[185] J. Ollerton, D. McCollin, D.G. Fautin, and G.R. Allen. Finding nemo: nestedness engendered by mutualistic organisation in anemonefish and their hosts. *Proceedings of the Royal Society B*, page doi:10.1098, 2006.

[186] S. Opitz. Trophic interactions in Caribbean coral reefs. *ICLARM Tech. Rep.*, 43: 341, 1996.

[187] T. Opsahl. Why anchorage is not (that) important: Binary ties and sample selection. *Available at http://wp.me/poFcY-Vw. Accesed 2013 Aug 10*, 2011.

[188] T. Opsahl, F. Agneessens, and J. Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 3(32):245–251, 2010.

[189] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.

[190] R.T. Paine. Food web complexity and species diversity. *Am. Nat.*, 100:65–75, 1966.

[191] J. Park and M.E.J. Newman. The origin of degree correlations in the internet and other networks. *Phys. Rev. E*, 66:026112, 2003.

[192] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86:3200, 2001.

[193] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87:258701, 2001.

[194] J. Patricio. Network analysis of trophic dynamics in south florida ecosystems, fy 99: The graminoid ecosystem. *Master's Thesis. University of Coimbra, Coimbra, Portugal*, 2000.

[195] D. Pauly, V. Christensen, J. Dalsgaard, R. Froese, and F. Torres. Fishing down marine food webs. *Science*, 279:860–3, 1998.

[196] M. Percival. Floral ecology of coastal scrub in southeast jamaica. *Biotropica*, 6: 104–129, 1974.

[197] F. Picard, J. J. Daudin, M. Koskas, S. Schbath, and S. Robin. Assessing the exceptionality of network motifs. *J. of Comp. Biol.*, 15:1–20, 2008.

[198] S. Pigolotti, S. Krishna, and M. H. Jensen. Oscillation patterns in negative feedback loops. *Proc Natl Acad Sci U S A*, 104(16):6533–6537, 2007.

[199] S. L. Pimm. *Food Webs*. Chapman and Hall, London, 1982.

[200] S. L. Pimm. *The Balance of Nature? Ecological Issues in the Conservation of Species and Communities*. The University of Chicago Press, Chicago, 1991.

[201] E. Plahte, T. Mestl, and S. W. Omholt. Feedback loops, stability and multistationarity in dynamical systems. *Journal of Biological Systems*, 03(02):409–413, 1995.

[202] S.J. Plitzko, B. Drossel, and C. Guill. Complexity–stability relations in generalized food-web models with realistic parameters. *J. Theor. Biol.*, 306:7–14, 2012.

[203] G. Polis. Complex trophic interactions in deserts: an empirical critique of food-web theory. *Am. Nat.*, 138:123–125, 1991.

[204] B. Poulin, S. J. Wright, G. Lefebvre, and O. Calderon. Interspecific synchrony and asynchrony in the fruiting phenologies of congeneric bird-dispersed plants in panama. *Journal of Tropical Ecology*, 15:213–227, 1999.

[205] R. B. Primack. Insect pollination in the new zealand mountain flora. *New Zealand J. Bot.*, 21:317–333, 1983.

[206] N. Ramirez and Y. Brito. Pollination biology in a palm swamp community in the venezuelan central plains. *Botanical Journal of the Linnean Society*, 110:277–302, 1992.

[207] E. Ravasz and A. L. Barabási. Hierarchical organization in complex networks. *Phys. Rev. E*, 67:026112, 2003.

[208] L.A. Real. Kinetics of functional response. *Am. Nat.*, 111:289–300, 1977.

[209] E. L. Rezende, E. M. Albert, M. A. Fortuna, and J. Bascompte. Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecology letters*, 12, 2009.

[210] Enrico L. Rezende, Jessica E. Lavabre, Paulo R. Guimaraes, Pedro Jordano, and Jordi Bascompte. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*, 448:925–928, 2007.

[211] J.O. Riede, U. Brose, B. Ebenman, U. Jacob, R. Thompson, C. Townsend, and T. Jonsson. Stepping in Elton's footprints: a general scaling model for body masses and trophic levels across ecosystems. *Ecology Letters*, 14:169–178, 2011.

[212] M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *Systems and Implications for System Design in IEEE Internet Computing Journal*, 2002.

[213] C. Rodríguez-Caso, B. Corominas-Murtra, and R. V. Solé. On the basic computational structure of gene regulatory networks. *Mol. BioSyst.*, 5:1617–1629, 2009.

[214] A. G. Rossberg. *Food Webs and Biodiversity: Foundations, Models, Data*. Wiley, 2013.

[215] A. G. Rossberg, K. D. Farnsworth, K. Satoh, and J. K. Pinnegar. Universal power-law diet partitioning by marine fish and squid with surprising stability-diversity implications. *Proc. R. Soc. B*, 278:1617–25, 2011.

[216] G. Rote. *Path Problems in Graphs*, volume 7 of *Computing Supplementum*. Springer Vienna, Austria, 1990.

[217] J. F. F. Mendes S. N. Dorogovtsev. Evolution of networks. *Advances in physics.*, 51:1079–1187, 2002.

[218] S. Saavedra, D. B. Stouffer, B. Uzzi, and J. Bascompte. Strong contributors to network persistence are the most vulnerable to extinction. *Nature*, 478(7368): 233–235, 2011.

[219] G. M M Santos, C. M. L. Aguiar, and M. A. R. Mello. Flower-visiting guild associated with the caatinga flora: trophic interaction networks formed by social bees and social wasps with plants. *Apidologie*, 41:466–475, 2010.

[220] J. Sanz, J. Navarro, A. Arbués, C. Martín, P. C. Marijuán, and Y. Moreno. The transcriptional regulatory network of mycobacterium tuberculosis. *PLoS ONE*, 6:e22178, 07 2011.

[221] D. W. Schemske, M. F. Willson, M. N. Melampy, L. J. Miller, L. Verner, K. M. Schemske, and L. B. Best. Flowering ecology of some spring woodland herbs. *Ecology*, 59:351–366, 1978.

[222] M. A. Serrano, D. Krioukov, and M. Boguñá. Self-similarity of complex networks and hidden metric spaces. *Phys. Rev. Lett.*, 100:078701, Feb 2008.

[223] R. Sharan, S. Suthram, R. M. Kelley, T. Kuhn, S. McCuine, P. Uetz, T. Sittler, R. M. Karp, and T. Ideker. Conserved patterns of protein interaction in multiple species. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6):1974–1979, 2005.

[224] W. R. Silva, P. de Marco, E. Hasui, and V.S.M. Gomes. Patterns of fruit-frugivores interactions in two atlantic forest bird communities of south-eastern brazil: implications for conservation. *In: D.J. Levey, W.R. Silva and M. Galetti (eds.) Seed dispersal and frugivory: ecology, evolution and conservation*, (Wallinford: CAB International):423–435, 2002.

[225] E. Small. Insect pollinators of the mer bleue peat bog of ottawa. *Canadian Field Naturalist*, 90:22–28, 1976.

[226] C. Smith-Ramírez, P. Martinez, M. Nuñez, C. González, and J. J. Armesto. Diversity, flower visitation frequency and generalism of pollinators in temperate rain forests of chiloé island,chile. *Botanical Journal of the Linnean Society*, 147: 399–416, 2005.

[227] B. K. Snow and D. W. Snow. The feeding ecology of tanagers and honeycreeperes in trinidad. *The Auk*, 88:291–322, 1971.

[228] R. V. Solé and J. Bascompte. *Self-Organization in Complex Ecosystems*. Princeton University Press, Princeton, USA, 2006.

[229] R. V. Solé, D. Alonso, and A. McKane. Scaling in a network model of a multi-species ecosystem. *Physica A*, 286:337–44, 2000.

[230] R. V. Solé and M. Montoya. Complexity and fragility in ecological networks. *Proc. R. Soc. Lond. B*, 268:2039–204, 2001.

[231] A.E. Sorensen. Interactions between birds and fruit in a temperate woodland. *Oecologia*, 50:242–249, 1981.

[232] U. Srinivasan, J. Dunne, J. Harte, and N. Martinez. Response of complex food webs to realistic extinction sequences. *Ecology*, 88:671–682, 2007.

[233] P. P. A. Staniczenko, Jason C. Kopp, and Stefano Allesina. The ghost of nestedness in ecological networks. *Nature Communications*, page 1391.

[234] P. P. A. Staniczenko, J.C. Kopp, and S. Allesina. The ghost of nestedness in ecological networks. *Nature Communications*, 4:139, 2013.

[235] D. B. Stouffer, J. Camacho, R. Guimerà, C. A. Ng, and L. A. N. Amaral. Quantitative patterns in the structure of model and empirical food webs. *Ecology*, 86: 1301–1311, 2005.

[236] D. B. Stouffer, J. Camacho, and L. A. N. Amaral. A robust measure of food web intervality. *Proc. Natl. Acad. Sci. USA*, 103:19015–19020, 2006.

[237] D. B. Stouffer, M. Sales-Pardo, M. I. Sirer, and J. Bascompte. Evolutionary conservation of species' roles in food webs. *Science*, 335:1489–1492, 2012.

[238] D.B. Stouffer and J. Bascompte. Compartmentalization increases food-web persistence. *Proc. Natl. Acad. Sci. USA*, 108, 2011.

[239] Steven H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

[240] G. Sugihara and H. Ye. Cooperative network dynamics. *Nature*, 458:979, 2009.

[241] William J. Sutherland, Robert P. Freckleton, H. Charles J. Godfray, Steven R. Beissinger, Tim Benton, Duncan D. Cameron, Yohay Carmel, David A. Coomes, Tim Coulson, Mark C. Emmerson, Rosemary S. Hails, Graeme C. Hays, Dave J. Hodgson, Michael J. Hutchings, David Johnson, Julia P. G. Jones, Matt J. Keeling, Hanna Kokko, William E. Kunin, Xavier Lambin, Owen T. Lewis, Yadvinder Malhi, Nova Mieszkowska, E. J. Milner-Gulland, Ken Norris, Albert B. Phillimore, Drew W. Purves, Jane M. Reid, Daniel C. Reuman, Ken Thompson,

Justin M. J. Travis, Lindsay A. Turnbull, David A. Wardle, and Thorsten Wiegand. Identification of 100 fundamental ecological questions. *Journal of Ecology*, 101(1):58–67, 2013.

[242] S. Suweis, F. Simini, J. R. Banavar, and A. Maritan. Emergence of structural and dynamical properties of ecological mutualistic networks. *nature*, 500:449–452, 2013.

[243] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero. A new metrics for countries fitness and products complexity. *Scientific Reports*, 2:723, 2012.

[244] B. Tang, H.K. Hsu, P.Y. Hsu, R. Bonneville, S.S. Chen, T.H. Huang, and V.X. Jin. Hierarchical modularity in era transcriptional network is associated with distinct functions and implicates clinical outcomes.

[245] D. Thieffry, A. M. Huerta, E. Pérez-Rueda, and J. Collado-Vides. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in escherichia coli. *Bioessays*, 20(5):433–440, 1998.

[246] R. M. Thompson and C. R. Townsend. Energy availability, spatial heterogeneity and ecosystem size predict food-web structure in stream. *Oikos*, 108:137–148, 2005.

[247] R. M. Thompson and C.R. Townsend. Impacts on stream food webs of native and exotic forest: an intercontinental comparison. *Ecology*, 84:145–61, 2003.

[248] G Tiana, S Krishna, S Pigolotti, M H Jensen, and K Sneppen. Oscillations and temporal signalling in cells. *Physical Biology*, 4(2):R1, 2007.

[249] C. R. Townsend, R. M. Thompson, A. R. McIntosh, C. Kilroy, E. Edwards, and M. R. Scarsbrook. Disturbance, resource supply, and food-web architecture in streams. *Ecol. Let.*, 1:200–209, 1998.

[250] Trustlet. http://www.trustlet. org/datasets/kaitiaki/kaitiaki-graph-2008-09-01.dot online; accessed 14-march-2014. *Trust network datasets project.*

[251] N. T.Wheelwright, W. A.Haber, K. G.Murray, and C. Guindon. Tropical fruit-eating birds and their food plants: a survey of a costa rican lower montane forest. *Biotropica*, 16:173–192, 1984.

[252] R. E. Ulanowicz. Growth and development: Ecosystems phenomenology. springer, new york. pp 69-79. *Network Analysis of Trophic Dynamics in South Florida Ecosystem, FY 97: The Florida Bay Ecosystem.*, 1986.

[253] R. E. Ulanowicz and D. Baird. Nutrient controls on ecosystem dynamics: the chesapeake mesohaline community. *Journal of Marine Systems*, 19(1–3):159 – 172, 1999.

[254] R. E. Ulanowicz, C. Bondavalli, and M.S. Egnotovich. Spatial and temporal variation in the structure of a freshwater food web. *Network Analysis of Trophic Dynamics in South Florida Ecosystem, FY 97: The Florida Bay Ecosystem.*

[255] R. E. Ulanowicz, C. Bondavalli, and M.S. Egnotovich. Spatial and temporal variation in the structure of a freshwater food web. *Network Analysis of Trophic Dynamics in South Florida Ecosystem, FY 97: The Florida Bay Ecosystem.*, 1998.

[256] R. E. Ulanowicz, J. J. Heymans, and M. S. Egnotovich. Network analysis of trophic dynamics in south florida ecosystems. *Network Analysis of Trophic Dynamics in South Florida Ecosystems FY 99: The Graminoid Ecosystem.*, 2000.

[257] W. Ulrich, M. Almeida-Neto, and N. J. Gotelli. A consumer's guide to nestedness analysis. *Oikos*, 118(1):3–17, 2009.

[258] M. P. van den Heuvel and O. Sporns. Rich-club organization of the human connectome. *The Journal of Neuroscience*, 31:15775–15786, 2011.

[259] D. P. Vázquez and M. A. Aizen. Asymmetric specialization: a pervasive feature of plant pollinator interactions. *Ecology*, 85:1251–1257, 2004.

[260] D. P. Vázquez. *Interactions among Introduced Ungulates, Plants, and Pollinators: A Field Study in the Temperate Forest of the Southern Andes. Doctoral*

Dissertation thesis. University of Tennessee, USA, Knoxville, Tennessee, USA, 2002.

[261] D. P. Vázquez and D. Simberloff. Ecological specialization and susceptibility to disturbance: conjectures and refutations. *American Naturalist*, 159:606–623, 2002.

[262] D. P. Vázquez and D. Simberloff. Changes in interaction biodiversity induced by an introduced ungulate. *Ecology Letters*, 6:1077–1083, 2003.

[263] D. P. Vázquez, C. Melian, N. Williams, N. Bluthgen, B. Krasnov, and R. Poulin. Species abundance and asymmetric interaction strength in ecological networks. *Oikos*, 116:1120–1127, 2007.

[264] D. P. Vázquez, N. Blüthgen, L. Cagnolo, and N. P. Chacoff. Uniting pattern and process in plant–animal mutualistic networks: a review. *Annals of Botany*, 103 (9):1445–1457, 2009.

[265] R. B. Waide and W. B. Reagan. *The Food Web of a Tropical Rainforest*. University of Chicago Press, Chicago, 1996.

[266] F. Wang and D.P. Landau. Efficient multiple-range random walk algorithm to calculate the density of states. *Physical Review Letters*, 86:2050–2053, 2001.

[267] P. H. Warren. patial and temporal variation in the structure of a freshwater food web. *Oikos*, 55, 1989.

[268] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[269] J. Wickens and R. Ulanowicz. On quantifying hierarchical connections in ecology. *J. Social Biol. Struct.*, 11:369–378, 1988.

[270] R. J. Williams and N. D. Martinez. Simple rules yield complex food webs. *Nature*, 404:180–183, 2000.

[271] R. J. Williams and N. D. Martinez. Success and its limits among structural models of complex food webs. *Journal of Animal Ecology*, 77:512–519, 2008.

[272] David H. Wright and Jaxk H. Reeves. On the meaning and measurement of nestedness of species assemblages. *OECOLOGIA*, 92(3):414–428, 1992.

[273] S. Wuchty, Z. N. Oltvai, and A. L. Barabási. Network motifs: theory and experimental approaches. *Nature Genetics.*, 35:176 – 179, 2003.

[274] X. Xu, J. Zhang, J. Sun, and M. Small. Revising the simple measures of assortativity in complex networks. *Physical Review E*, 80:056106, 2009.

[275] P. Yodzis. Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. *J. Anim. Ecol.*, 67:635–658, 1998.

[276] H. Yu and M. Gerstein. Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl. Acad. Sci. USA*, 103(40):14724–14731.

[277] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33:452–473, 1977.

[278] Z. Zhai, S. Y. Ku, Y. Luan, G. Reinert, M. S. Waterman, and F. Sun. The power of detecting enriched patterns: An hmm approach. *J. of Comp. Biol.*, 17:581–592, 2010.

Publications derived from this thesis

[P1] S. Johnson, V. Domínguez-García, and MA. Muñoz. Trophic coherence determines foodweb stability. *Proc. Natl. Acad. Sci. USA*, accepted, 2014.

[P2] V. Domínguez-García, Simone Pigolotti, and MA. Muñoz. Inherent directionality explains the lack of feedback loops in empirical networks. *submitted*, 2014.

[P3] S. Johnson, V. Domínguez-garcía, and MA. Muñoz. Factors determining nestedness in complex networks. *PloS ONE*, page e74025, 2013.

[P4] V. Domínguez-García and MA. Muñoz. Ranking species in mutualistic networks. *submitted*, 2014.