SPRINGER BRIEFS IN COMPLEXITY

Tiziano Squartini
Diego Garlaschelli

Maximum-Entropy Networks

Pattern Detection, Network Reconstruction and Graph Combinatorics

Springer

SpringerBriefs in Complexity

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems—cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The three major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "SpringerBriefs in Complexity" which are concise and topical working reports, case-studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

More information about this series at http://www.springer.com/series/8907

Tiziano Squartini · Diego Garlaschelli

Maximum-Entropy Networks

Pattern Detection, Network Reconstruction and Graph Combinatorics

Springer

Tiziano Squartini
IMT School for Advanced Studies
Lucca
Italy

Diego Garlaschelli
Lorentz Institute for Theoretical Physics
University of Leiden
Leiden
The Netherlands

ISSN 2191-5326 ISSN 2191-5334 (electronic)
SpringerBriefs in Complexity
ISBN 978-3-319-69436-8 ISBN 978-3-319-69438-2 (eBook)
https://doi.org/10.1007/978-3-319-69438-2

Library of Congress Control Number: 2017955644

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Angela and Costanza
our welcoming horizons

A Leonardo e Valeria
due Stelle luminose
che camminano sulla Terra

Preface

«… i' vegno per menarvi all'altra riva
ne le tenebre etterne, in caldo e 'n gelo.
E tu che se' costì, anima viva,
pàrtiti da cotesti che son morti.
Ma poi che vide ch'io non mi partiva,
disse: «Per altra via, per altri porti
verrai a piaggia, non qui, per passare:
più lieve legno convien che ti porti».
E 'l duca lui: «Caròn, non ti crucciare:
vuolsi così colà dove si puote
ciò che si vuole, e più non dimandare».

—Dante Alighieri, *Divina Commedia*

This book is intended as a complete, self-consistent introduction to a general methodology to study complex networks. This methodology combines concepts of information theory, statistical physics, and graph theory and provides a way to build *maximum-entropy models of networks*. These models have a rigorous theoretical origin and a range of practical applications. In this book, we emphasize the applications to pattern detection, network reconstruction, and graph combinatorics.

Most of the results in this book have been developed by ourselves and our direct collaborators. Our approach has been rigorously investigated both analytically and numerically and used in several applications, ranging from physics to economics and biology. Given the robustness of the method, its wide range of applicability, and the growing interest it has been attracting from our colleagues, we felt the need of writing a compact text unifying our results fragmented across many publications.

Pattern detection, network reconstruction, and graph combinatorics are by themselves three distinct and active fields of research. Several specialized and generally unrelated techniques have been proposed within each of these fields. Our

aim in this present book is *not* that of discussing the existing field-specific approaches, but that of emphasizing the *connections* between these seemingly independent problems.

Moving from common, general, first principles, we present a unified methodology which provides various explicit solutions to the aforementioned problems. Thus, while providing the references necessary to contextualize all the results discussed here, this book is deliberately not in a review-like structure. Rather, it focuses on an original, unifying approach to these three selected domains of network theory.

This book is intended for a broad audience, ranging from Ph.D. students in physics and mathematics looking for an unconventional introduction to network theory, to researchers in other disciplines (e.g., economics, social sciences, biology) interested in the application of network analysis to their topic(s) of interest. No prior knowledge of network theory or any of the specific topics discussed here is required.

Lucca, Italy — Tiziano Squartini
Leiden, The Netherlands — Diego Garlaschelli
May 2017

Acknowledgements

Frodo: But I am going to Mordor.
Sam: I know that well enough, Mr. Frodo.
Of course you are. And I'm coming with you.

—John Ronald Reuel Tolkien, *The Fellowship of the Ring*

We are indebted to a number of friends and colleagues who have collaborated with us in the development and application of the methodology presented in this book. Large parts of these results would have not been obtained without their help and dedication.

We thank in particular the co-authors of the main publications constituting the bulk of this book: Giulio Cimini, Giorgio Fagiolo, Andrea Gabrielli, Frank den Hollander, Iman van Lelyveld, Rossana Mastrandrea, Francesco Picciolo, Franco Ruzzenenti.

We also thank the people in our groups in Leiden, Lucca and Rome who have surrounded us with a great sense of teamwork: Assaf Almog, Alexandru Babeanu, Riccardo Di Clemente, Eli van Es, Elena Garuccio, Valerio Gemmetto, Janusz Meylahn, Joey de Mol, Vasyl Palchikov, Andrea Roccaverde, Fabio Saracco, Qi Zhang.

We also thank our former mentors and supervisors for all the fruitful discussions we have had with them: Guido Caldarelli, Tiziana Di Matteo, Maria I. Loffredo, Luciano Pietronero, Felix Reed-Tsochas.

This list is by far incomplete, and our gratitude equally goes to all the people that, even if not mentioned here, have contributed directly or indirectly to our adventure.

We warmly thank our parents for their continued affection and support: Tiziana and Bruno, Rosa and Danilo.

A final lovely thank goes to those who have encouraged us and been patient with us during the months spent writing this book: Angela, Costanza, Leonardo, Valeria.

Grazie!

Contents

Chapter 1
Introduction

Forthwith this frame of mine was wrenched
With a woful agony, Which forced me to begin my tale;
And then it left me free.
Since then, at an uncertain hour,
That agony returns:
And till my ghastly tale is told,
This heart within me burns.
I pass, like night, from land to land;
I have strange power of speech;
That moment that his face I see,
I know the man that must hear me:
To him my tale I teach.

– Samuel Taylor Coleridge, The Rime of the Ancient Mariner

Whoever has played *'Guess Who?'* knows the importance of gaining maximally useful information in absence of complete knowledge. To win the game, a player must be able to identify an unknown character among many possible ones. Each player should aim at discarding the highest number of 'wrong' characters by asking the smallest number of questions about the traits of the face of the unknown character. The key to winning the game is a careful choice of the traits about which questions are asked: at each round, the best choice is the one that selects the most informative traits, i.e. those that allow the player to identify most quickly the correct character.

A similar problem is encountered in most areas of science. Researchers often deal with complicated real-world systems about which there is only limited empirical information available, for instance because it is impossible to characterize all the components and/or all the interactions among them. The scientific challenge is that of *identifying the most informative properties* of the system, i.e. a relatively small set of properties which, once empirically measured, shed as much light as possible about the entirety of the system itself. The same problem underlies the practice

T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-69438-2_1

of *theoretical modelling* as well. A good theoretical model of a system or process is one that uses the smallest number of parameters and/or explanatory variables to replicate the observed phenomenology. So, like in *'Guess Who?'*, the identification of a few informative properties is the key to good scientific practice. At an underlying mathematical level, the identification of the most informative properties strongly depends on an *enumeration problem*: once a new set of properties is measured, how many configurations of the system are still compatible with the measured values? A set of properties for which the number of compatible configurations is small is a 'good' choice. In the jargon of information theory, such properties encode a lot of *information*. In that of of statistical physics, their measurement lowers the *entropy*, which is a measure of the uncertainty of the outcome of a random process.

* * *

The **focus** of this book is on *networks* (or *graphs* in the mathematical jargon), i.e. systems whose components are *nodes* (or *vertices*) connected by *links* (or *edges*).

Over the last two decades, the study of networks has been steadily flourishing into a very active and popular discipline, often referred to as 'Network Science', 'Network Theory', or simply 'Complex Networks'. The reason for this tremendous growth of interest is the fact that networks are encountered virtually ubiquitously across many problems of scientific and societal relevance. How infectious diseases, biological signals, financial crises, computer viruses, electrical breakdowns, mechanical fractures, rumours and opinions spread throughout organisms, materials, societies, economies and ecologies ultimately depends on the structure of some underlying network. Such a network may for instance describe how different cells, people, companies, computers, power plants, physical systems or biological species are connected together. As a matter of fact, different disciplines, including biology, social science, epidemiology, mathematics, physics, neuroscience, finance, economics, ecology, computer science and engineering, often focus their analyses on systems that admit a common abstract description in terms of graphs.

Quite surprisingly, the empirical topological properties of the vast majority of networks encountered across these disciplines are remarkably similar. For instance, most networks feature a large number of nodes with a few connections and a few nodes (or 'hubs') with a very large number of connections. Indeed, the observed frequency of nodes with a given number of connections is in most cases found to decrease slowly with, and typically as a negative power of, this number. Networks with this property are called *scale-free* networks. Similarly, most real-world networks turn out to display *community structure*, i.e. to be partitioned into groups of nodes that are much more densely connected internally than with each other. Other widespread properties include a high level of *clustering*, i.e. the formation of many connected triples of nodes, and a short average path length among most pairs of nodes, a property that goes under the name of *small-world* effect. The observation of these common phenomenological properties in the structure of many different networks has stimulated the development of a unifying mathematical language to charaterize the empirical properties of complex networks and a common framework to study generic processes for the spread of information, shocks and instabilities throughout

them. Statistical physics, graph theory and computer science have played a major role in developing methods, models and tools to characterize and understand complex networks.

Many excellent books and review articles introducing network science from the above interdisciplinary viewpoint already exist in the literature. This book does *not* want to replicate this material and focuses on networks from a different perspective.

* * *

The **aim** of this book is that of illustrating how seemingly unrelated challenges encountered in network science are highly intertwined and, ultimately, have a common theoretical underpinning. For the sake of illustration, we have selected three key problems: pattern detection, network reconstruction and graph combinatorics.

Pattern detection is the identification of empirical properties that systematically deviate from some simple benchmark or reference model. In the context of networks, these properties might be a surprising preference of certain nodes to connect to each other with respect to the rest of the network, or an unexpectedly large number of triangles forming around some particular nodes. These 'surprising' or 'unexpected' properties are likely to represent important structural patterns singling out nontrivial information about the (unknown) network formation process. Detecting such patterns requires the specification of a null model (or even multiple null models) of the network as a benchmark. In our *'Big Data'* era dominated by the need to identify meaningful information in huge streams of noisy data that are continuously produced, pattern detection is becoming more and more important as a core component of virtually any data-mining technique.

Network reconstruction is the problem of inferring the unknown structure of a network, given only partial knowledge about some of its structural properties. In many practical situations, network data are protected by privacy or confidentiality issues. For instance, when the nodes of the network represent banks and the links represents their credit relationships, generally only very limited information is publicly known. Banks only publicly disclose their total exposures towards the aggregate of all other banks and 'who is connected to whom' is not known. A similar condition applies to networks of social contacts, for which the available data typically only report the aggregate number of encounters of each node. In order to carry out any network-dependent analysis, e.g. the estimation of the level of systemic risk embodied in the interbank market or the prediction of how infectious diseases will spread across a population, one has to resort to advanced reconstruction techniques identifying the most likely set of networks compatible with the partial information available. Network reconstruction is therefore rapidly becoming a key instrument in both science and policy making.

Graph combinatorics refers to various graph-related operations in discrete mathematics. These operations typically include, or are related to, the sampling of graphs with given topological properties, or their enumeration in some appropriate asymptotic limit. Imagine the set of, say, all simple undirected graphs with N vertices, i.e. graphs whose nodes can be connected by either no edge or a single edge that does not have a specified direction. Calculating how many such graphs have a given total

number L of links is straightforward. Similarly, sampling a certain number of such graphs can be done efficiently and uniformly for any value of L. However, counting how many graphs meet some more heterogeneous topological constraint can be a daunting task. An important example is the asymptotic enumeration and uniform sampling of graphs with given *degree sequence*, i.e. graphs where each vertex i is assigned a different number k_i of edges (this number is known as the *degree* of node i). At present, the counting problem is solved only under stringent conditions on the heterogeneity of the values in the degree sequence, while there is no general result allowing us to enumerate graphs with an arbitrary degree sequence. Similarly, sampling graphs with given degree sequence uniformly becomes more and more challenging as the heterogeneity in the degrees increases. Counting and sampling graphs with heterogeneous topology is therefore an active field of research.

* * *

The **message** of this book is that there are deep connections, of both theoretical and applied nature, between the three scientific challenges outlined above. As a consequence, it is possible to define certain problem-specific techniques within a common framework that facilitates a beneficial interaction between the three topics. In particular, two unifying notions will recur throughout the book: entropy maximization and topological complexity.

Entropy maximization is like an abstract, invisible and beautiful tree with many concrete, visible and juicy fruits and is therefore a powerful example for Lewin's golden statement that *"There is nothing more practical than a good theory"*. For the purposes of this book, the maximization of the entropy is introduced in order to construct ensembles of graphs whose topology is maximally random, apart from a controlled set of structural properties enforced as constraints. This leads us to the definition and analysis of maximum-entropy ensembles of networks with topological constraints. It turns out that such ensembles provide a common answer to the problems of the *maximally random* construction of null models for pattern detection, the *maximally unbiased* inference of networks from partial information and the *maximally uniform* sampling of graphs with given constraints. Our introduction of a general method to define constrained maximum-entropy ensembles of graphs will allow us to address each of these applications in a coherent fashion, thus emphasizing the relationships between the different topics covered by the book.

Topological complexity is, tautologically, what makes real-world networks 'complex'. It is loosely defined here as the combination of all the empirical topological properties, universally observed across most networks, that are not easily reproduced by simple homogeneous models. In particular, the widespread observation of heterogeneous (e.g. scale-free) and hierarchically structured (e.g. community-like) topologies mentioned above suggests that real networks may obey complex organizing principles. Therefore, any sensible model of real-world networks should feature an appropriately high degree of structural heterogeneity. Maximum-entropy models are no exception and they should be flexibly defined in terms of constraints that are distributed heterogeneously over the network. This is precisely what we will do in

this book, effectively combining entropy maximization and topological complexity together from the very beginning.

Importantly, besides calling for an appropriate theoretical framework for the general problem of network modelling, the topological complexity of real networks has also strong practical implications for each of the three topics discussed in this book. In fact, most pattern detection, network reconstruction and graph enumeration techniques work well when applied to networks with low topological complexity, while they are all seriously compromised by the high complexity of real-world networks. For instance, when applied to graphs with narrowly distributed degrees, these techniques are more amenable to analytical and numerical treatment; but when applied to real-world networks with sufficiently broad degree distributions, they either become unfeasible or require approximations that often allow for no control on the resulting errors and biases. Thus topological complexity, which in principle is 'only' a phenomenological property affecting the values of the structural quantities measured in real networks, turns out to be of fundamental nature and calls for new conceptual frameworks.

* * *

The **structure** of this book is organized as follows.

Chapter 2 illustrates the general maximum-entropy method which forms the core of the approach used in all the following chapters. The key quantities at the basis of this method, *entropy* and *likelihood*, are introduced and both maximized in order to obtain explicit probability distributions over ensembles of graphs with structural constraints taken as input from any real-world network. Importantly, the method is compatible with any level of empirical topological complexity from the outset. This is ensured by the use of exact maximum-entropy probability distributions which, unlike in most alternative approaches, we do not try to approximate and which we rigorously fit to real-world networks via the application of the maximum-likelihood principle. This methodology avoids any approximation based on questionable assumptions of 'simplicity' of the network. Two examples are worked out explicitly and several statistical criteria to discriminate between different models are explained.

Chapter 3 focuses on the detection of empirical patterns (such as degree correlations, the hierarchical character of a particular structure, the tendency to form reciprocal connections, etc.) in real-world networks. Here, the maximum-entropy method is used to define *null models* of a real-world network and patterns are sought for by looking for statistically significant discrepancies between the real network and the null model. Several null models are implemented and applied to many networks, ranging from economics to biology. Via these examples, we elaborate on selected applied topics such as the identification of topological patterns in food webs, the realistic modelling of the international trade network, the characterization of reciprocity of weighted networks and the detection of early-warning signals of the 2008 crisis in the Dutch interbank network.

Chapter 4 deals with the problem of network reconstruction from partial information. Here, the maximum-entropy method is used to infer the higher-order topological properties of real-world networks, with an emphasis on financial and economic ones,

starting from the knowledge of only a few pieces of empirical node-specific information imposed as constraints. We test various maximum-entropy models while looking for the empirical constraints that lead to satisfactorily reconstructed networks. We show that real-world networks can be reliably reconstructed only if some piece of topological information, heterogeneously distributed over nodes, is known and enforced. When this information is not known, one has to find ways to preliminary infer it. Remarkably, the maximum-entropy principle provides again a good recipe to perform this inference by predicting various functional relationships linking the unknown properties to the known ones. The result is a reliable reconstruction method with minimal input requirements.

Chapter 5 deals with graph combinatorics. Here we argue that, for any combinatorial problem with 'hard' constraints, e.g. counting how many graphs have exactly the same value of a given topological property, the maximum-entropy approach provides an explicit answer to the dual problem with 'softened' constraints, i.e. where the topological property is enforced only as an average over the ensemble of graphs. We show that, in regimes of strong topological heterogeneity for which the 'hard' enumeration problem is at present still unsolved, the dual 'soft' problem can sometimes be solved exactly. This naturally leads to the general question of whether ensembles with hard and soft constraints are asymptotically equivalent, an intriguing problem known as *ensemble (non)equivalence*. We show recent results proving that ensemble equivalence, while traditionally assumed to hold in fairly general circumstances, breaks down for ensembles of graphs with a number of constraints that grows with the number of nodes, e.g. for random graphs with given degree sequence. We discuss some important consequences for the enumeration and sampling of such graphs.

Chapter 6 looks at the main results of the book in retrospect and offers an overarching summary, a few take-home messages and some concluding remarks.

Chapter 2
Maximum-Entropy Ensembles of Graphs

Whereof one cannot speak, thereof one must be silent.

—Ludwig Josef Johann Wittgenstein, *Logisch-Philosophische Abhandlung*

Abstract In this chapter we describe the core method that will be used throughout the rest of the book, i.e. the construction of a *constrained maximum-entropy ensemble* of networks. This procedure requires the definition of the *entropy* of a network ensemble, the specification of structural properties to be enforced as *constraints*, the calculation of the resulting maximum-entropy *probability* of network configurations, and the maximization of the *likelihood*, given the empirical values of the enforced constraints. We describe this procedure explicitly, after giving some general motivations. In particular, we discuss the crucial importance of enforcing *local* constraints that preserve the (empirical) heterogeneity of node properties. The maximum-entropy method not only generates the exact probabilities of occurrence of any graph in the ensemble, but also the expectation values and the higher moments of any quantity of interest. Moreover, unlike most alternative approaches, it is applicable to networks that are either binary or weighted, either undirected or directed, either sparse or dense, either tree-like or clustered, either small or large. We also discuss various likelihood-based statistical criteria to rank competing models resulting from different choices of the constraints. These criteria are useful to assess the informativeness of different network properties.

2.1 Constructing Constrained Graph Ensembles: Why and How?

In Chap. 1 we already anticipated that various problems of great importance in network science may be (re)formulated in such a way that similar underlying concepts are invoked and a common toolkit is employed. In particular, we gave a series of motivations for addressing three specific problems that will be discussed in detail in

T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-69438-2_2

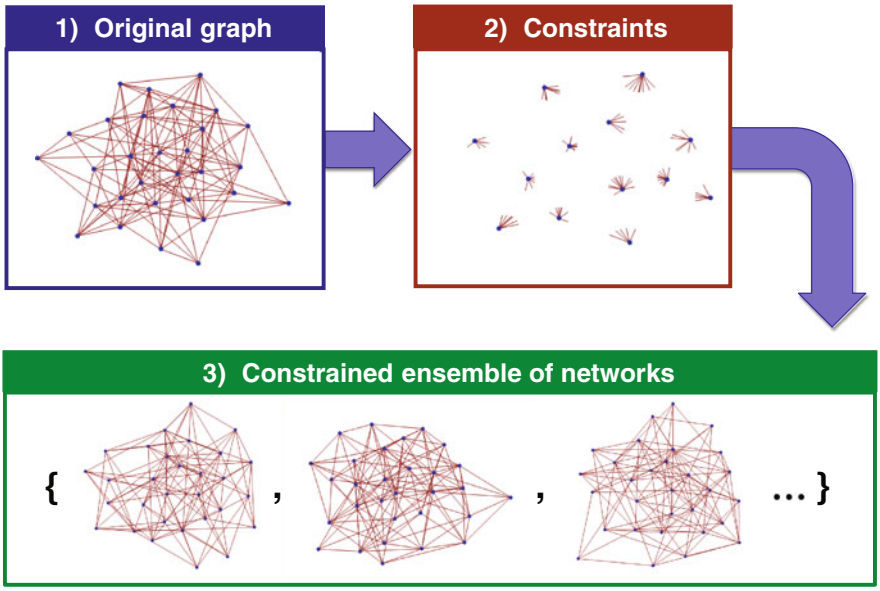

Fig. 2.1 Abstract construction of a constrained ensemble of networks. (1) First, a particular network (for instance, an observed real-world one) is considered. (2) Then, a set of topological properties (in the example shown, the different numbers of connections of nodes) is chosen as a constraint and measured on the network. (3) Finally, an ensemble of networks induced by the measured constraints is constructed according to some rule resulting in a probability distribution over the space of allowed configurations. In the problem of pattern detection (see Chap. 3), the average properties of the constrained ensemble are then compared to those of the original network in order to detect statistically significant patterns in the latter. In the problem of network reconstruction (see Chap. 4), one actually does not have empirical access to the original network, but only to a set of its properties; the procedure therefore starts at step 2) by treating these properties as constraints and then produces an ensemble of inferred possible configurations for the unknown network. Finally, in various problems in graph combinatorics (see Chap. 5), one is interested in correctly sampling and/or enumerating the configurations from the induced ensemble

the following chapters of this book, namely the detection of statistically significant structural patterns in real networks (Chap. 3), the reconstruction of networks from partial empirical information (Chap. 4) and the sampling or enumeration of graphs with specified topological properties (Chap. 5). These three different problems, while unrelated at first sight, require in fact a common framework: the construction of *an ensemble of random graphs with given constraints* [1–23]. In the case of pattern detection, the constraints represent null hypotheses used as a reference to identify empirical patterns. In the case of network reconstruction, they represent pieces of incomplete data used to infer missing information. In the case of graph combinatorics, they represent topological properties of the network configurations to be sampled or enumerated.

A pictorial representation of the construction of a constrained ensemble of graphs is given in Fig. 2.1. In general, the procedure may go through three steps: we may start

from a specific (real-world) network, then measure the topological properties we want to preserve, and finally impose these properties as a constraint in the construction of the ensemble. In all the cases considered in this book, we impose that the graphs in the ensemble all have exactly the same number of nodes as the original network. It should at this point be noted that, at least conceptually, we may skip the first step and start directly with the specification of the constraints themselves (in such a case, the number of nodes in the original network should also be known, if not already evident from the constraints themselves). Whether one can actually skip the first step depends on the particular technical implementation, not on the theoretical definition of the ensemble. For instance, in certain computational pattern-detection approaches that aim at iteratively randomizing a real-world network while preserving some of its properties (explicit examples are given below in Sect. 2.1.2), one has to start from the first step. By contrast, in other cases (e.g. when only partial information is available about the original network, as in the problems considered in Chap. 4), one is forced to start from the second step. This implies that, in order to be useful for multiple purposes, 'good' ensemble constructions should be able to take (only) the values of the chosen constraints as input. Of course, this requires that such values are *graphic*,[1] i.e. realizable in at least one graph. If the constraints come from the observation of some network (including the case when they are the only information available about some unknown underlying network), their graphicality is of course always guaranteed.

In general, the third step in the construction of an ensemble of constrained graphs, i.e. the specification of a (satisfactory) graph probability, is the most challenging one. The reason is twofold, as briefly explained below.

- Firstly, not all choices of the constraints lead to equally easy ways of constructing the resulting ensemble. In fact, the most important and useful constraints turn out to be *node-specific*, which implies that the local properties of nodes have to be preserved separately. This requirement complicates the construction of the probability distribution. This point is discussed in detail in Sect. 2.1.1.
- Secondly, not all probability distributions satisfying the chosen constraints are equally acceptable from a theoretical point of view. For instance, a key requisite is that they assign the same probability to all graphs that have the same value of the constraints, because there is no reason to prefer any one such graph over any other such graph. This point is illustrated in Sect. 2.1.2 for the case of computational methods and in Sect. 2.1.3 for the case of analytical methods.

In the rest of this chapter, we explain in detail the two points above, first by highlighting the importance of imposing local constraints (Sect. 2.1.1) and then by emphasizing how most computational (Sect. 2.1.2) and analytical (Sect. 2.1.3) methods proposed in the literature fail to correctly sample the resulting ensembles. Then, in Sect. 2.2 we introduce a rigorous methodology to produce a graph probability

[1] A topological property f, where $f(\mathbf{G})$ is the value of the property in graph $\mathbf{G}$, is said to evaluate to a *graphic* (or *graphical*) value $\tilde{f}$ if there exist at least one graph $\tilde{\mathbf{G}}$ that realizes such value, i.e. for which $f(\tilde{\mathbf{G}}) = \tilde{f}$.

meeting all the desired requirements. The methodology is based on the *maximization of the entropy* subject to a set of chosen constraints (this step fixes the *functional form* of the probability distribution) and the subsequent *maximization of the likelihood* (this step fixes the *numerical values* of the probability distribution). We will see that the maximum-entropy formulation solves all the highlighted problems in an elegant and mathematically explicit way, a result that will come as a relief. This procedure represents the core of the formalism that will be used repeatedly in this book.

2.1.1 Definition and Importance of Local Constraints

To characterize the structure of a given network, arbitrarily many topological properties can be defined. Among these, the simplest and most important properties are *local* quantities, i.e. functions of only the immediate neighbourhood of each node. Let us introduce some notation to define these local properties, before discussing their importance.

A *binary undirected graph*[2] with N vertices is completely specified by a symmetric $N \times N$ *adjacency matrix* $\mathbf{A}$. The entries of the latter are such that $a_{ij} = 1$ if the vertices i and j are connected and $a_{ij} = 0$ otherwise. For each node i, the *degree* $k_i(\mathbf{A}) = \sum_{j \neq i} a_{ij}$ is defined as the number of connections of that node, and is therefore a local node-specific property. The *degree sequence* $\mathbf{k}(\mathbf{A}) = \{k_i(\mathbf{A})\}_{i=1}^N$ is the N-dimensional vector of degrees of all nodes.

In case of *weighted*[3] undirected graphs, a network is specified by a symmetric $N \times N$ *weight matrix* $\mathbf{W}$ where the entry w_{ij} quantifies the intensity of the link connecting nodes i and j. This includes the case $w_{ij} = 0$ corresponding to nodes i and j being not connected. Besides the degree (which is still defined as the number of connections of a node, irrespective of their intensity), another local property that can be introduced in this case is the *strength* $s_i(\mathbf{W}) = \sum_{j \neq i} w_{ij}$, defined as the sum of the weight of all links of vertex i. The *strength sequence* $\mathbf{s}(\mathbf{W}) = \{s_i(\mathbf{W})\}_{i=1}^N$ is the N-dimensional vector of strengths of all nodes.

[2] An *undirected* graph (or network) is a graph where no direction is specified for the edges. An undirected graph is *binary* or *simple* if each pair of nodes i and j (with $i \neq j$) is connected by at most one edge, i.e. if there are *no multiple edges* between the same two nodes. We will also assume the absence of *self-loops* (edges starting and ending at the same node) throughout the book.

[3] A *weighted graph* (or network) is a graph where links may carry different intensities. When dealing with weighted networks, throughout the book we will assume non-negative integer link weights (i.e. $w_{ij} = 0, 1, 2 \cdots + \infty$) for simplicity. This corresponds to the assumption that an indivisible unit of measure of link weights has been preliminary specified. Under this assumption, a weighted network can also be regarded as a graph that is in general not simple, i.e. where multiple links of unit weight are allowed between the same two nodes. We will still exclude the possibility of self-loops. Ideally, one may think of link weights becoming continuous as the unit of measure is chosen to be vanishingly small.

In case of (either binary of weighted) *directed*[4] graphs, the matrices $\mathbf{A}$ and $\mathbf{W}$ are in general not symmetric, and each node admits an *in-degree* $k_i^{in}(\mathbf{A}) = \sum_{j \neq i} a_{ji}$, an *out-degree* $k_i^{out}(\mathbf{A}) = \sum_{j \neq i} a_{ij}$, an *in-strength* $s_i^{in}(\mathbf{W}) = \sum_{j \neq i} w_{ji}$ and an *out-strength* $s_i^{out}(\mathbf{W}) = \sum_{j \neq i} w_{ij}$. Correspondingly, we can introduce the *in-degree sequence* $\mathbf{k}^{in}(\mathbf{A}) = \{k_i^{in}(\mathbf{A})\}_{i=1}^N$, the *out-degree sequence* $\mathbf{k}^{out}(\mathbf{A}) = \{k_i^{out}(\mathbf{A})\}_{i=1}^N$, the *in-strength sequence* $\mathbf{s}^{in}(\mathbf{W}) = \{s_i^{in}(\mathbf{W})\}_{i=1}^N$ and the *out-strength sequence* $\mathbf{s}^{out}(\mathbf{W}) = \{s_i^{out}(\mathbf{W})\}_{i=1}^N$.

The degree(s) and strength(s) defined above are in some sense the immediate, first-order structural properties that can be measured in any network. For these reason, we will refer to the degree and strength sequences as the *local topological properties* of a network. To speak in general terms more easily, we will denote a generic sequence of such local constraints with the vector $\mathbf{C}(\mathbf{G})$, where $\mathbf{G}$ denotes a generic graph (either binary or weighted, either directed or undirected) and $\mathbf{C}$ denotes a generic sequence of constraints (e.g. $\mathbf{k}$ or $\mathbf{s}$) or a concatenation of more sequences (e.g. the concatenation of $\mathbf{k}^{out}$ and $\mathbf{k}^{in}$, or of $\mathbf{s}^{out}$ and $\mathbf{s}^{in}$).

The importance of local topological properties comes from the fact that, in most situations, they directly reflect the effects of 'size' or 'importance' of nodes. For instance, more popular people naturally have a higher degree in a social network, and more wealthy companies or countries naturally have a higher strength in an economic network. Clearly, one expects the size and/or importance of nodes to have a strong impact on the realized patterns of connections. For various reasons, one would like to characterize this effect quantitatively by constructing (ensembles of) networks that have the same local properties of a given real-world network. For instance, if one has empirical access only to the degrees and/or strengths of nodes of a network, then the best guess one can make about the unknown network is given by a suitable ensemble of graphs matching the empirical local properties. This is the problem of network reconstruction that will be treated extensively in Chap. 4. Another example is encountered when looking for higher-order patterns in a real network, i.e. for topological features that *cannot* be explained or replicated starting from the knowledge of only the local properties. In this case, which is the problem of pattern detection that will be treated extensively in Chap. 3, one requires a benchmark model constructed from only the local properties themselves. Both challenges require the introduction of ensembles of networks with given local properties.

Having clarified the importance of constructing graph ensembles tailored on the empirical values of the degrees and/or strengths of nodes, one might at this point wonder whether such values may be produced as random fluctuations around a common average value (in which case the model would only require the average value as a parameter, besides a choice of the probability distribution of the random fluctuations

[4] A *directed* graph is a graph where a direction is specified for each edge (self-loops are not allowed in this case as well). A directed graph is binary (or simple) if any two nodes i and j are connected in one of the following four mutually-exclusive ways: via only a directed link from i to j, via only a directed link from j to i, via both such links, or via no link at all. A directed graph is weighted if links can carry different intensities, including when they are pointing in opposite direction between the same two nodes. Again, we will assume non-negative integer weights.

around it) or whether more complicated and higher-dimensional models, controlling the local constraints for each node separately, are needed. The answer to this question has been given over decades of extensive empirical analyses which have conclusively shown that the empirical values of the degrees and the strengths observed in most real-world networks are in some sense 'irreducible' to the outcome of any simple homogeneous model. For instance, in most real-world networks both the empirical *degree distribution*[5] and the empirical *strength distribution*[6] turn out to be very broad, and typically with a right tail decaying as a *power law* of the form $P(x) \propto x^{-\gamma}$, with $2 < \gamma < 3$. In the abstract limit where the number of nodes becomes infinite, the variance of these distributions diverges while the mean remains finite, implying that the average value is not representative of the value of individual nodes. This signals the absence of a typical scale for the degree or strength of nodes. For this reason, most empirical networks are called *scale-free* [24]. The degree and strength distribution of these networks is much broader than would be obtained under a simple homogeneous network formation model with just a global constraint on e.g. the average degree or the average strength of nodes, even after including noise or stochasticity.

For instance, the oldest and most popular random graph model, the Erdős-Rényi (ER) model [25], constructs a simple binary random graph with N nodes by connecting each (distinct) pair of these nodes with a given probability p. Since each node has $N-1$ potential other nodes to connect to, and since the same value p of the probability is used for all pairs of nodes, it immediately follows that the expected[7] degree of each node i has the same value $\langle k_i \rangle = p(N-1)$. This is already an indication of the complete homogeneity of the ER model. Moreover, it is easy to show that, for each node i, the probability for the degree k_i taking a particular value k is distributed binomially in k around the above expected value $p(N-1)$. Since a binomial distribution is much narrower than typical empirical degree distributions, it is intuitively clear that the latter cannot be regarded as typical realizations of the ER model. This argument can be confirmed in various statistically rigorous ways, although we will not focus on this issue in this book. Note that the parameter p has a direct control on the expected total number of links $\langle L \rangle = pN(N-1)/2$, where $N(N-1)/2$ is the number of pairs of N nodes (i.e. the maximum possible number of edges). Therefore one can regard the ER model as an ensemble of random graphs with a global constraint on the expected total number $\langle L \rangle$ of links, or equivalently on the expected average degree[8] $\langle \bar{k} \rangle = 2\langle L \rangle / N = p(N-1)$. It is then clear that such a global, overall constraint would not produce realistic network configurations. This

[5]The *empirical degree distribution* is defined, for a given network, as the fraction $P(k)$ of nodes that have degree k.

[6]The *empirical strength distribution* is defined, for a given network, as the fraction $P(s)$ of nodes that have strength s.

[7]Throughout the book, by *expected value* (or *expectation*) of a topological property we mean the average of that property over the ensemble of random graphs under consideration. We denote expectation values with angular brackets $\langle \cdot \rangle$. The rigorous definition is given later in Eq. (2.7).

[8]The *average degree* in a simple undirected graph with N nodes is defined as $\bar{k} = N^{-1} \sum_{i=1}^{N} k_i$ and necessarily equals $2L/N$, where L is the total number of links.

calls for more complicated models where the (expected) degree of each node can be controlled independently of the degree of the other nodes.

An almost identical argument holds for the strengths. One can define the *weighted random graph model* (WRG) [19] as the weighted counterpart of the ER model, where the only constraint is now the expected total weight $\langle W \rangle$ of all links in the network, or equivalently the expected average strength[9] $\langle \bar{s} \rangle = 2\langle W \rangle / N$. It can be shown that this constraint can be implemented by going over all pairs of nodes and placing an edge of weight w according to a geometric probability distribution having the same parameter value for all node pairs. The resulting strength of all nodes is distributed according to a negative binomial distribution with the same expected value. Empirical strength distributions are therefore incompatible with typical realisations of the WRG. More complicated models of weighted networks, with separately controllable strenghts, are needed in order to restore compatibility with the heterogeneity of real-world networks.

The above discussion clarifies that, in order to construct ensembles of constrained networks that are both practically useful and theoretically sound, one should introduce a way of controlling each local property (i.e. each degree and/or strength) separately. It is useful at this point to look back at Fig. 2.1. We denote the particular initial graph (step 1) by $\mathbf{G}^*$ and the corresponding numerical value of the constraints (step 2) by $\mathbf{C}^* \equiv \mathbf{C}(\mathbf{G}^*)$. The third step will generate a collection of many graphs $\{\mathbf{G}\}$ which include $\mathbf{G}^*$ itself. It should be noted that the constraints $\mathbf{C}$ define the *sufficient statistics* of the problem: the construction of the ensemble should be possible by knowing only the value $\mathbf{C}^*$ (i.e. skipping step 1) and no other property of the graph $\mathbf{G}^*$. While this idea is conceptually simple, implementing it correctly is very challenging. Understanding the origins of this difficulty is a key step towards the appreciation of the maximum-entropy method that will be described in Sect. 2.2. For this reason, in the rest of this section we briefly review the problem by discussing various alternative attempts at the construction of ensembles of graphs with local constraints.

2.1.2 Computational Approaches

For concreteness, let us consider the case of binary undirected graphs, which is by far the most frequently explored situation. We will consider many other ensembles later in the book. The ensemble of binary undirected graphs with specified degree sequence $\mathbf{C}^* \equiv \mathbf{k}^*$ is known as the *binary configuration model* (BCM) [1, 2, 23].

Given a real-world binary undirected network $\mathbf{G}^* \equiv \mathbf{A}^*$, an entirely 'bottom-up' computational approach to the generation of the associated binary configuration model with degree sequence $\mathbf{k}^* \equiv \mathbf{k}(\mathbf{A}^*)$ consists in initially assigning each vertex i a number of 'edge stubs' equal to the target degree k_i^*. Then, pairs of stubs are

[9] The *average strength* in a weighted undirected graph with N nodes is defined as $\bar{s} = N^{-1} \sum_{i=1}^{N} s_i$ and necessarily equals $2W/N$, where W is the total weight of all links in the network.

randomly matched avoiding the formation of self-loops and multiple links, until all degrees reach their desired values (*edge stub connection*). Looking back at Fig. 2.1, this implementation has the desirable property that one can start from 'step 2' in the ensemble construction. Indeed, the edge stubs are precisely the half-edges portrayed inside the second box in the picture. Unfortunately, if the values of the degrees are too heterogeneous, this procedure is known to get stuck in configurations where vertices requiring additional connections have no more eligible partners [1, 2]. Typical realizations of the procedure share this problem, which therefore cannot be easily circumvented by simply aborting the unsuccessful realizations and starting over again.

A popular alternative method is based on a 'top-down' implementation where the entire real network $\mathbf{A}^*$ is taken as the initial configuration, and a family of randomized variants is generated by iteratively applying a *local rewiring algorithm* (LRA). In the LRA, two edges (A, B) and (C, D) are randomly selected and replaced by the two edges (A, D) and (C, B), if the latter are both not already present [1, 2] (see Fig. 2.2 for an illustration). Technically, the above procedure generates an ensemble where all randomized networks have exactly the same degree sequence as the original network. This method has been applied to various networks, including the Internet [2], cellular networks [3] and food webs [8], in order to detect higher-order patterns (such as clustering and motifs) not merely due to local constraints. Unfortunately, this approach is time-consuming since many (a number R much larger than the observed number of links L [1, 20], even if not rigorously specified) iterations of the LRA are required to obtain a *single* randomized network, and the entire process must be repeated several times to produce a large number M (again unspecified) of randomized networks, on each of which any topological property X of interest must be measured explicitly and averaged at the end to obtain an estimate for $\langle X \rangle$. The computational time required to obtain $\langle X \rangle$ is therefore of the order $O(M \cdot T_R \cdot R) + O(M \cdot T_X)$, where T_R is the average time required to perform a single successful rewiring step and T_X is that required to compute X on a single network in the randomized set. Moreover, even if the sufficient statistics of the problem is the degree sequence $\mathbf{k}(\mathbf{A}^*)$ alone, the above approach requires the entire original network $\mathbf{A}^*$ (or any other network with the same degree sequence, which is however difficult to obtain from scratch due to the problems discussed above) as the starting configuration, thus making use of much more information than required in principle.

Besides these practical limitations, the main problem of the LRA is the fact that it is *biased*, i.e. it does not sample the desired ensemble uniformly. This has been rigorously shown relatively recently [21, 26, 27]. For undirected networks, uniformity has been shown to hold, at least approximately, only when the degree sequence is such that [27]

$$k_{max} \cdot \overline{k^2}/(\overline{k})^2 \ll N \tag{2.1}$$

where k_{max} is the largest degree in the network, $\overline{k}$ is the average degree, $\overline{k^2}$ is the second moment, and N is the number of vertices. Clearly, the above condition sets an upper bound for the heterogeneity of the degrees of vertices, and is violated if the heterogeneity is strong. This is another indication that the available methods break

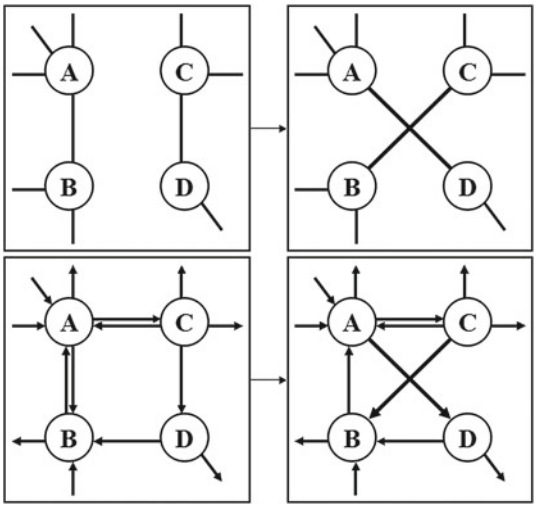

Fig. 2.2 An illustration of the local rewiring algorithm whose iteration allows to computationally explore the configuration model with sharp constraints (upper panel, for undirected networks; lower panel, for directed networks)

down for 'strongly heterogeneous' networks. As we discuss later, most real-world networks are found to fall precisely within this class.

For directed networks, where links are oriented and the constraints to be met are the numbers of incoming and outgoing links (in-degree and out-degree) separately, a condition similar to Eq. (2.1) holds, but there is also the additional problem that the LRA is non-ergodic, i.e. it is in general not able to explore the entire ensemble of networks [26]. The violation of uniformity and ergodicity in the LRA implies that the average quantities over the graphs it generates are biased, i.e. they do not correspond to the correct expectations.

It has been shown that, in order to restore ergodicity, it is enough to introduce an additional 'triangular move' inverting the direction of closed loops of three vertices [26]. However, in order to restore uniformity, one must do something much more complicated: at each iteration, the attempted 'rewiring move' must be accepted with a probability that depends on some complicated property of the current network configuration [21, 26, 27]. Since this property must be recalculated at each step, the resulting algorithm is extremely time consuming.

Other recent alternatives [28–30] rely on theorems, such as the Erdös-Gallai [31] one, that set necessary and sufficient conditions for a degree sequence to be *graphic*, i.e. realized by at least one graph. These 'graphic' methods exploit such (or related) conditions to define biased sampling algorithms in conjunction with the estimation of the corresponding sampling probabilities, thus allowing one to statistically reweight the outcome and sample the ensemble effectively uniformly [28–30]. Del Genio et al. [28] show that, for networks with power-law degree distribution of the form $P(k) \sim k^{-\gamma}$, the computational complexity of sampling *one* graph using their algorithm is $O(N^2)$ if $\gamma > 3$. However, when $\gamma < 3$ the computational complexity increases to $O(N^{2.5})$ if

$$k_{max} < \sqrt{N} \tag{2.2}$$

and to $O(N^3)$ if $k_{max} > \sqrt{N}$. The upper bound $\sqrt{N}$ is a particular case of the so-called 'structural cut-off' that we will discuss in more detail later. For the moment, it enough for us to note that Eq. (2.2) is another indication that, for strongly heterogeneous networks, the problem of sampling gets more complicated. As we will discuss later, most real networks violate Eq. (2.2) strongly.

So, while 'graphic' algorithms do provide a solution for every network, their complexity increases for networks of increasing (and realistic) heterogeneity. A more fundamental limitation is that they can only handle the problem of binary graphs with given degree sequence. The generalization of these methods to other types of networks and other constraints is not straightforward, as it would require the proof of more general 'graphicality' theorems, and *ad hoc* modifications of the algorithm.

For what concerns weighted networks, the available 'hard' algorithms regard each link weight as an integer multiple w of a fundamental unit of weight, transform each edge of weight w into w edges of unit weight and rewire the latter as in the unweighted case, now ensuring that the strength of each vertex is preserved. This means replacing a list of $L^* \leq N(N-1)/2$ weighted links, summing up to a total weight $W^* = \sum_{i<j} w^*_{ij}$, with $W^* \gg N(N-1)/2$ unweighed links. As real networks have broadly distributed weights summing up to a large W^*, this procedure becomes very time consuming as unfeasibly many rewiring steps per randomized variant must be performed. Moreover, much less is known about the potential bias produced by this algorithm in the case of weighted networks.

2.1.3 Analytical Approaches

In contrast with computational methods, analytical approaches seek to provide explicit mathematical expressions that directly estimate the ensemble averages of topological properties, without generating the ensemble computationally. Two main approaches exist. One makes use of generating functions for the relevant probability distributions [23]. For the binary configuration model, the key quantity is the generating function $g(z) = \sum_k z^k P(k)$ of the degree distribution. Unfortunately, this method assumes the network to be infinite and locally tree-like (even if in some cases this approximation turns out to perform unexpectedly well even beyond its formal range of applicability [32]), and is thus in general inappropriate if the size of the network is small and if the input degree distribution can only be realized by dense and/or clustered networks. In this approach, clustered or dense networks can only be generated by imposing additional constraints besides the degree sequence, such as the number of triangles attached to vertices [33], thus leading to a different ensemble which is not the one we are seeking to characterize. A different approach looks for an analytical expression for the probability p_{ij} that the vertices i and j are connected in the randomized ensemble [4]. Due to its probabilistic nature, this approach generates an ensemble with *soft* constraints, i.e. where graphs violating the constraints are present and assigned non-zero probabilities. The constraints are still realized on average, i.e. the expectation value $\langle \mathbf{C} \rangle$ of $\mathbf{C}$ is still equal to $\mathbf{C}^*$. The popular expression used for p_{ij} is

$$p_{ij} = \frac{k_i^* k_j^*}{2L^*} \tag{2.3}$$

where $L^* \equiv L(\mathbf{A}^*) = \sum_i k_i(\mathbf{A}^*)/2 = \sum_{i<j} a_{ij}^*$ is the total number of links. While the expected degree $\langle k_i \rangle = \sum_j p_{ij}$ generated by the above formula coincides (approximately, as we discuss below) with the desired degree k_i^*, the probability p_{ij} may exceed 1 for pairs of highly connected nodes such that $k_i^* k_j^* > 2L^*$. In general, only if the degree sequence is such that

$$k_i^* < \sqrt{2L^*} = \sqrt{\sum_j k_j^*} \quad \forall i \tag{2.4}$$

then using Eq. (2.3) on the real network $\mathbf{A}^*$ will not lead to the above problem. While the above condition is typically obeyed by networks with narrow degree distribution it is generally violated by scale-free networks displaying a power-law degree distribution $P(k) \sim k^{-\gamma}$, and this violation becomes stronger and stronger as the density of the network increases. In particular, it is easy to see that in order to ensure Eq. (2.4) the maximum degree $k_{max}^* = \max_i k_i^*$ in the network should not exceed the so-called *structural cut-off* $k_c \sim N^{1/2}$ [34]. This is particularly evident for sparse networks where the average degree $\bar{k} = \sum_i k_i/N = 2L/N$ remains constant as N increases, so that Eq. (2.4) remains valid only if $k_{max} < \sqrt{2L} \sim N^{1/2}$. By contrast, extreme value theory shows that in networks with degree distribution $P(k) \sim k^{-\gamma}$ the maximum degree scales as $k_{max}^* \sim N^{1/(\gamma-1)}$, so that if $\gamma < 3$ (as observed in most real-world scale-free networks) then $k_{max}^* > N^{1/2}$ which exceeds k_c.

Loosely speaking, the meaning of p_{ij} being larger than 1 for some pairs of vertices in Eq. (2.3) is that i and j should be connected by more than one undirected edge in order to actually realize the desired degree sequence. Also, since the desired equality $\langle k_i \rangle = k_i^*$ is only ensured if one lets the sum in $\sum_j p_{ij} = \langle k_i \rangle$ run over all vertices *including i itself*, one must allow the presence of self-loops in the randomized networks. Thus, even if this is not evident at a first glance, the ensemble generated by Eq. (2.3) does not only contain binary and loop-less undirected graphs and is thus not a proper null model for an empirical binary loop-less network $\mathbf{A}^*$ with degree sequence $\mathbf{k}^*$ violating Eq. (2.4), as is typically the case for real-world networks with broad degree distributions.

An elegant proof that the correct ensemble probability p_{ij} for loop-less graphs with no multiple connections differs from Eq. (2.3) has been proposed [5] and re-derived within the framework of maximum-entropy graph ensembles [10]. An independent proof of the inadequacy of Eq. (2.3) is that it does not generate the graph $\mathbf{A}^*$ with maximum likelihood [35]. These results show that the functional form of p_{ij} in Eq. (2.3) is intrinsically problematic and does not give highest likelihood to $\mathbf{A}^*$ and to all other graphs with the same degree sequence as $\mathbf{k}^*$.

We can briefly make a similar comment for weighted networks with given strength sequence $\mathbf{s}^*$, an ensemble known as *weighted configuration model* [11] (and discussed at length in Sect. 2.2.3 and Chaps. 3 and 4). A (naïve, yet widely used) generaliza-

tion [11, 36] of the (naïve, yet widely used) expression (2.3) states that the expected weight of the link between nodes i and j in this ensemble is

$$\langle w_{ij} \rangle = \frac{s_i^* s_j^*}{2W_{tot}^*} \tag{2.5}$$

where $W_{tot}^* \equiv W_{tot}(\mathbf{W}^*) = \sum_i s_i(\mathbf{W}^*)/2 = \sum_{i<j} w_{ij}^*$ is the total weight. The above expression has been shown to have as many limitations as its binary counterpart, and to be incorrect [18]. A simple signature of this inadequacy is the fact that, although Eq. (2.5) is treated as an expected value, there is no indication of the probability distribution from which it is derived. Therefore, it is impossible to derive the expected value of topological properties that are nonlinear functions of the weights.

Therefore, while the available analytical methods are useful to characterise artificially generated networks with special properties, they cannot be used to correctly describe ensembles of networks that are realistically small, clustered, or dense. Unfortunately, the above limitations are generally ignored, and Eqs. (2.3) and (2.5) are frequently used beyond their limits of applicability to estimate connection probabilities and expected link weights. Analogous problems exist in the case of directed networks.

2.2 The Maximum-Entropy Method

The discussion in the previous section highlights that none of the above implementations succeeds in obtaining the properties of ensembles of constrained networks such that two requests are met simultaneously:

- the method is general and works for any network, even if displaying small size, high density and large clustering;
- expected values across the ensemble are unbiased and can be computed analytically, without sampling the configuration space explicitly.

In this section, we introduce a different method that fulfills the above criteria. The method is based on the maximum-entropy principle and leads to exact expressions for the probability of occurrence of any graph. It therefore allows us to calculate, correctly and analytically, the expected topological properties of graphs in the ensemble. We first illustrate the methodology in full generality, i.e. by considering an abstract choice of topological constraints, and then work out two explicit examples in detail. More examples will be given throughout the rest of the book when needed to address specific problems.

Looking again at Fig. 2.1, let us denote by $\mathbf{G}$ a generic network in the ensemble, and by $\mathbf{G}^*$ the particular original network (we may think of it as the empirical network we need to randomize). The chosen constraint is $\mathbf{C}^* = \mathbf{C}(\mathbf{G}^*)$. The ensemble consists of all possible networks $\{\mathbf{G}\}$ with the same number N of nodes and of the same type (undirected/directed, binary/weighted, etc.) as $\mathbf{G}^*$, and includes $\mathbf{G}^*$ itself. Note that

$\mathbf{G}$ can always be thought of as a matrix with entries $\{g_{ij}\}$, where g_{ij} represents the (either binary or non-negative) weight of the edge (i, j). Any topological property X evaluates to $X(\mathbf{G})$ when measured on the particular network $\mathbf{G}$, i.e. it is an (arbitrarily complicated) function of the entries $\{g_{ij}\}$.

Each graph $\mathbf{G}$ in the ensemble has an occurrence probability $P(\mathbf{G})$ whose form is determined by the particular constraints enforced. This probability must always be such that

$$\sum_{\mathbf{G}} P(\mathbf{G}) = 1 \tag{2.6}$$

where the sum runs over all graphs in the ensemble. The expectation or mean value of any topological property X is the ensemble average

$$\langle X \rangle \equiv \sum_{\mathbf{G}} X(\mathbf{G}) P(\mathbf{G}). \tag{2.7}$$

At this point, we look for the probability distribution that maximizes the Shannon-Gibbs entropy

$$S(P) \equiv -\sum_{\mathbf{G}} P(\mathbf{G}) \ln P(\mathbf{G}) \tag{2.8}$$

subject to the normalization condition (2.6) and to the desired constraints $\mathbf{C}^*$. The entropy $S(P)$ is a measure of the level of uncertainty, or randomness, in the outcome of the random variable described by the probability distribution P. Variables that have a certain outcome, i.e. whose probability is one for such outcome and zero for all other outcomes, correspond to zero entropy. On the contrary, variables that are maximally uncertain, i.e. for which every possible outcome has exactly the same probability, yield the maximum value[10] of the entropy. Maximizing the entropy subject to constraints is widely used in statistical mechanics and information theory, and in general for problems with incomplete information [37–40]. The deep meaning of constrained entropy maximization is that, in absence of any information other than the knowledge of $\mathbf{C}^*$, the probability should make the outcome of the random variable ($\mathbf{C}$ in this case) maximally uncertain provided that the constraints are met. Otherwise, the probability would be favouring specific configurations, making them more predictable than others and introducing an unjustified bias. Now, the solution to the entropy maximization problem depends on whether we want the constraints $\mathbf{C}^*$ to be *hard* or *soft*.

Enforcing hard constraints means that we only allow (i.e. assign non-zero probability) the graphs that match the constraints *exactly*, i.e. such that $\mathbf{C}(\mathbf{G}) = \mathbf{C}(\mathbf{G}^*)$. This means that, in the above definition of entropy, we can restrict the sum to such configurations only. It is easy to see that the resulting maximum-entropy distribution, which is known as the *microcanonical ensemble* in statistical physics, is uniform over

[10] The maximum value of the entropy $S(P)$ depends on the total number of configurations over which the sum in Eq. (2.8) runs. This number can be rescaled to one for all probability distributions, upon normalizing $S(P)$ by the maximum value itself.

the set of graphs that match the hard constraints:

$$P_{\text{mic}}(\mathbf{G}) = \begin{cases} 1/\Omega_{\mathbf{C}^*} & \text{if } \mathbf{C}(\mathbf{G}) = \mathbf{C}(\mathbf{G}^*) \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

where $\Omega_{\mathbf{C}^*}$ denotes the number of graphs for which $\mathbf{C}(\mathbf{G}) = \mathbf{C}^*$. An intuitive picture of microcanonical ensembles of graphs is given in Fig. 2.3. Since $\Omega_{\mathbf{C}^*}$ is a combinatorial quantity, the above result establishes an important connection between statistical physics, probability theory and combinatorics. This connection will be explored in detail in Chap. 5. At this point, one should note that, while for simple constraints (such as the total number of links) it is easy to compute $\Omega_{\mathbf{C}^*}$, for more complicated constraints (including the degree sequence and the other local constraints we are interested in this book) this can become a very hard task. For instance, enumerating the number of graphs with a given degree sequence $\mathbf{k}^*$ is an open problem, and asymptotic expressions are known only in some restricted regime of density of the graph, i.e. under certain conditions that $\mathbf{k}^*$ must obey. For this reason, microcanonical graph ensembles are hard to deal with analytically and they are most often sampled computationally using the techniques we described in Sect. 2.1.2. However, as we discussed, these techniques are either computationally unfeasible or affected by the problem of bias, i.e. they do not sample the space of graphs according to the correct uniform probability (2.9). The computational difficulties are therefore related to the difficulties in calculating $\Omega_{\mathbf{C}^*}$ explicitly.

On the other hand, enforcing *soft* constraints means requiring that the desired value $\mathbf{C}^*$ is met only *on average* over the ensemble, or in other words that the constraint is $\langle \mathbf{C} \rangle = \mathbf{C}^*$. This requirement defines what is known as the *canonical ensemble* in statistical physics. However, unlike the traditional examples in physics, where the total energy is the only (scalar) constraint, for the cases of interest here the number of constraints grows linearly with the number of nodes in the system, since $\mathbf{C}$ is a vector of node-specific quantities. This important difference has enormous consequences, as we will discuss in Chap. 5. The form of the probability P_{can} in the canonical ensemble is found by requiring that, in addition to Eq. (2.6), the constraints are given by

$$\langle \mathbf{C} \rangle = \sum_{\mathbf{G}} \mathbf{C}(\mathbf{G}) P_{\text{can}}(\mathbf{G}) = \mathbf{C}^*. \tag{2.10}$$

It is easy to show [10] that the corresponding solution to the constrained entropy maximization problem is found by introducing a vector of Lagrange multipliers $\boldsymbol{\theta}$, one for each of the constraints in $\mathbf{C}$. The resulting conditional (on the value of $\boldsymbol{\theta}$) probability reads

$$P_{\text{can}}(\mathbf{G}|\boldsymbol{\theta}) = \frac{e^{-H(\mathbf{G},\boldsymbol{\theta})}}{Z(\boldsymbol{\theta})} \tag{2.11}$$

where $H(\mathbf{G}, \boldsymbol{\theta})$ is the so-called *graph Hamiltonian* defined as the linear combination

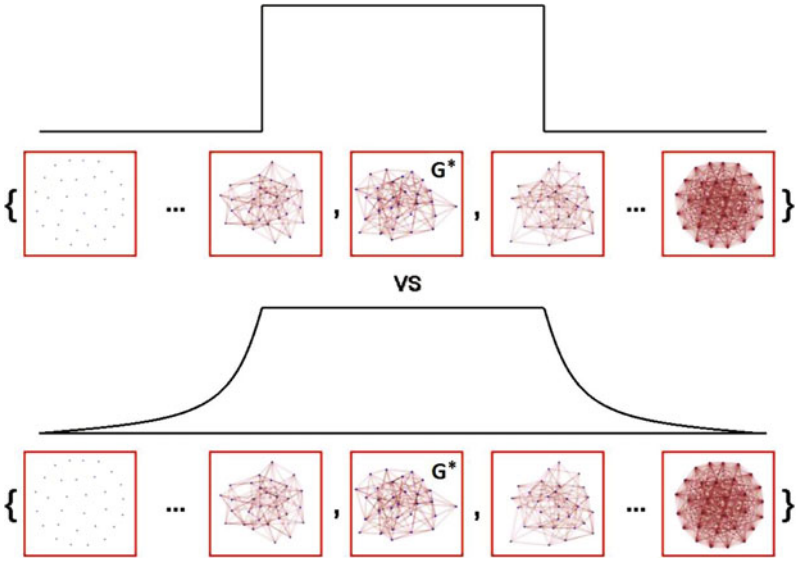

Fig. 2.3 Difference between microcanonical and canonical ensembles. Top: the microcanonical probability P_{mic} is non-zero only for the subset of graphs that realize the enforced constraints $\mathbf{C}^*$ exactly. Bottom: by contrast, the canonical probability P_{can} is non-zero for all graphs with the prescribed number of nodes, including those that violate the constraints (thus ranging from the empty graph to the complete graph), and has a constant value P_{can} for all graphs for which P_{mic} is non-zero. In general, P_{can} has the same value for all graphs that have the same value of $\mathbf{C}$

$$H(\mathbf{G}, \boldsymbol{\theta}) \equiv \sum_a \theta_a C_a(\mathbf{G}) = \boldsymbol{\theta} \cdot \mathbf{C}(\mathbf{G}) \tag{2.12}$$

and the normalizing quantity $Z(\boldsymbol{\theta})$ is the so-called *partition function*, defined as

$$Z(\boldsymbol{\theta}) \equiv \sum_{\mathbf{G}} e^{-H(\mathbf{G},\boldsymbol{\theta})}. \tag{2.13}$$

The above results show that the graph probability $P_{\text{can}}(\mathbf{G}|\boldsymbol{\theta})$ always depends on the value $\boldsymbol{\theta}$, which in turn depends on the constraints considered. As a consequence, we can rewrite Eq. (2.7) more explicitly as a function of $\boldsymbol{\theta}$:

$$\langle X \rangle_{\boldsymbol{\theta}} \equiv \sum_{\mathbf{G}} X(\mathbf{G}) P_{\text{can}}(\mathbf{G}|\boldsymbol{\theta}) \tag{2.14}$$

where $\langle \cdot \rangle_{\boldsymbol{\theta}}$ denotes that the ensemble average is evaluated at the particular parameter choice $\boldsymbol{\theta}$. The above expression clarifies that the expectation value of any topological property X depends on the specific enforced constraints through $\boldsymbol{\theta}$. Different choices of the constraints imply different values of $\boldsymbol{\theta}$, $P(\mathbf{G}|\boldsymbol{\theta})$ and $\langle X \rangle_{\boldsymbol{\theta}}$. Importantly, $P_{\text{can}}(\mathbf{G}|\boldsymbol{\theta})$ depends on $\mathbf{G}$ only through $\mathbf{C}(\mathbf{G})$. This automatically implies that

the canonical ensemble is unbiased, i.e. graphs with the same value of the constraints are assigned equal probability. A pictorial representation of this property is given in Fig. 2.3.

Now, a crucial difference between the microcanonical and canonical ensembles is that, if $\mathbf{C}^*$ is a vector of local topological constraints, P_{mic} cannot be exactly factorized into probabilities that involve distinct pairs of nodes, whereas P_{can} can. This implies that the exact computation of $\boldsymbol{\theta}^*$ is feasible even if that of $\Omega_{\mathbf{C}^*}$ is not. For these reasons, which will be illustrated in explicit examples later on, the canonical ensemble offers a viable and exact solution to the problem of constructing ensembles of graphs with local constraints. It will be the main tool we will use throughout the rest of the book. In statistical physics, results obtained within the canonical ensemble are generally expected to become equivalent to those that would be obtained within the microcanonical one in the so-called *thermodynamic limit*.[11] This notion is called *ensemble equivalence*. Whether or not ensemble equivalence holds also in the case of local constraints is an intriguing question, and its answer is postponed to Chap. 5. Here and in Chaps. 3 and 4, we assume that enforcing the constraint $\mathbf{C}^*$ softly is a perfectly acceptable strategy, for instance because its measured value may have been corrupted by noise or error, and we are therefore inclined to accept other values around $\mathbf{C}^*$ in the ensemble construction.

2.2.1 Maximum-Likelihood Parameter Estimation

Maximum-entropy graph ensembles generated by Eq. (2.11) have been used extensively to characterize mathematically networks with specified properties [5, 7, 10, 17, 18]. However, traditionally the Lagrange multipliers $\{\theta_a\}$ have been considered as free parameters, generally drawn from carefully chosen probability densities [10, 17, 18] that allow for analytical results, in terms of which the properties of the network model have been investigated. In most cases, the aim has been to explore the topological properties in the thermodynamic limit $N \to \infty$, where N is the number of vertices of the network. This means that only generic statistical properties of real networks, such as a power-law degree distribution with a certain exponent, have been used to generate the ensemble. However, this implies that the specific properties of a particular real network (such as deviations of individual vertices from the fitted degree distribution, the intrinsic finiteness of the system, etc.) have been ignored and, more importantly, that it has not been possible to establish any correspondence between the vertices of the real network and those of the model. Thus these approaches have not allowed maximum-entropy graph ensembles to be considered as null models of *a particular real network* in order to detect empirical topological patterns, or to

[11] In statistical physics, the *thermodynamic limit* is defined as the limit where the number of fundamental units that describe the microscopic configurations of the system diverges. In our graph ensembles, we regard the nodes as the units and their connections as the interactions.

reconstruct network topology from partial information, or even to enumerate graphs compatible with a specified vector of constraints.

Now, following [41], we make a step forward and construct, for a given choice of the constraints, the particular maximum-entropy graph ensemble representing the family of correctly randomized counterparts of a given real network $\mathbf{G}^*$. Explicitly, we consider a canonical ensemble of graphs with the same number N of vertices as the real network, and for a given choice of the constraints we fit the model defined by Eq. (2.11) to the empirical network $\mathbf{G}^*$. To this end, we exploit previous results [35] showing that maximum-entropy graph ensembles defined by Eq. (2.11) are a particular class of models for which the maximum-likelihood principle provides an excellent way to estimate parameters. In particular, it can be easily shown [35] that the log-likelihood

$$\mathscr{L}(\boldsymbol{\theta}) \equiv \ln P(\mathbf{G}^*|\boldsymbol{\theta}) = -H(\mathbf{G}^*, \boldsymbol{\theta}) - \ln Z(\boldsymbol{\theta}) \tag{2.15}$$

is maximized by the particular value $\boldsymbol{\theta}^*$ such that the ensemble average $\langle C_a \rangle_{\boldsymbol{\theta}^*}$ of each constraint C_a equals the empirical value $C_a(\mathbf{G}^*)$ measured on the real network:

$$\langle \mathbf{C} \rangle^* \equiv \langle \mathbf{C} \rangle_{\boldsymbol{\theta}^*} = \sum_{\mathbf{G}} \mathbf{C}(\mathbf{G}) P(\mathbf{G}|\boldsymbol{\theta}^*) = \mathbf{C}(\mathbf{G}^*) \tag{2.16}$$

where we have used $\langle \cdot \rangle^*$ as a shorthand notation to indicate the ensemble average $\langle \cdot \rangle_{\boldsymbol{\theta}^*}$ evaluated at the particular value $\boldsymbol{\theta}^*$. The above results means that *the maximum likelihood principle indicates, for maximum-entropy graph ensembles, precisely the parameter choice that ensures that the desired constraints are met*. This is not true in general: in other network models, tuning the average values of the topological properties of interest to their empirical values requires a parameter choice which in general does not maximize the likelihood to obtain the real network [35], thus introducing a bias in the analysis [42–44].

Solving the maximum-likelihood equations only takes a computational time T_E which is much shorter than the time required to measure any topological property of typical interest. Moreover, the time required to compute the expectation value $\langle X \rangle$ of a given property X analytically (formally corresponding to an average over a huge number of randomized configurations) is the same as the time T_X required to compute the same property on the single original network. The artificial generation of many randomized variants of the original network is no longer required. Therefore this method takes only a total time $O(T_E + T_X)$ to obtain $\langle X \rangle$ analytically, which is incredibly shorter than the aforementioned time $O(M \cdot T_R \cdot R) + O(M \cdot T_X)$ required by the LRA to obtain $\langle X \rangle$ only approximately. Importantly, T_E is independent of the complexity of the topological property X to measure, which means that for complicated properties $O(T_E + T_X) = O(T_X)$. Therefore for any topological property X which can be measured in a large but still reasonable time $O(T_X)$ on the real network, the computation of its expectation value $\langle X \rangle$ will require the same time $O(T_X)$. If the time required in order to obtain $\langle X \rangle$ is too large, it is because the time required to measure X is too large as well. In other words, the property X is too

complicated to be computed on the real network itself. In such a case, the problem is not due to the method, but to a demanding choice of X for that particular network.

Note that in Eqs. (2.14) and (2.16) the expectation values and the model parameters play inverted roles: while in Eq. (2.14) the expectation values are obtained as a function of the parameters $\boldsymbol{\theta}$ which can be varied arbitrarily, in Eq. (2.16) the observed constraints, which are measured on the particular real network and are therefore given as an input, are used to fix the model parameters to the values $\boldsymbol{\theta}^*$. Once the parameters solving the equations are found, they can be directly used to obtain the expectation value $\langle X \rangle$ and standard deviation $\sigma[X]$ of any topological property X of interest analytically (details on how to calculate standard deviations can be found in [41]). When useful, this also allows one to obtain a z-score representing the number of standard deviations by which the randomized value $\langle X \rangle$ differs form the observed value $X(\mathbf{A}^*)$. The possibility to obtain the standard deviations and z-scores is very important, because it allows one to assess which topological properties X are consistent with their randomized value $\langle X \rangle$ within a statistical error, and which deviate significantly from the null expectation. In the former case, one can conclude that the enforced constraints completely explain the higher-order property X. In the latter case, the observed property cannot be traced back to the constraints, and therefore requires additional explanations or generating mechanisms besides those required in order to explain the constraints themselves. We will discuss this procedure in more detail in the next chapter.

2.2.2 *A First Worked-Out Example: Binary, Undirected Networks with Constrained Degree Sequence*

In the binary, undirected case, each graph $\mathbf{G}$ is completely specified by its (symmetric) adjacency matrix $\mathbf{A}$. An important ensemble of binary undirected graphs is one where the constraint is the degree sequence [10]. This null model is also known as *configuration model* (CM). In our formalism this model is implemented by defining the following Hamiltonian:

$$H(\mathbf{A}) = \sum_i \theta_i k_i(\mathbf{A}) = \sum_i \sum_{j<i} (\theta_i + \theta_j) a_{ij} \tag{2.17}$$

and one can show [10] that this allows one to write the partition function as

$$Z(\boldsymbol{\theta}) \equiv \sum_{\mathbf{A}} e^{-H(\mathbf{A},\boldsymbol{\theta})} = \prod_i \prod_{j<i} (1 + x_i x_j) \tag{2.18}$$

and the graph probability as

$$P(\mathbf{A}) = \prod_i \prod_{j<i} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}} \tag{2.19}$$

where

$$p_{ij} = \frac{x_i x_j}{1 + x_i x_j} \tag{2.20}$$

(with $x_i \equiv e^{-\theta_i}$) is the probability that a link exists between vertices i and j in the maximum-entropy ensemble of binary undirected graphs characterized by the given degree sequence as the constraint.

The maximum-likelihood condition [41] prescribes to find the solution $\{x_i^*\}_{i=1}^N$ to the equations

$$\langle k_i \rangle = \sum_{j \neq i} \frac{x_i x_j}{1 + x_i x_j} = k_i(\mathbf{A}^*) \quad \forall i \tag{2.21}$$

by choosing the imposed constraint to be the empirical degree sequence $\{k_i(\mathbf{A}^*)\}_{i=1}^N$ of the particular real network $\mathbf{A}^*$ or, equivalently, by finding the values of the parameters $\{x_i^*\}_{i=1}^N$ that maximize the likelihood $P(\mathbf{A}^*)$ [35, 41]. Inserting the $\{x_i^*\}_{i=1}^N$ into Eq. (2.20) allows one to easily compute the expectation value $\langle X \rangle^*$ of any topological property X analytically, without generating the randomized networks explicitly [41].

Thus, Eq. (2.20) yields the exact value of the connection probability in the ensemble of randomized networks with the same average degree sequence as the empirical one and Eq. (2.21) shows that, by construction, the degrees of all vertices are special local quantities whose expected and empirical values are exactly equal: $\langle k_i \rangle^* = k_i(\mathbf{A}^*)$. It follows that the p_{ij} coefficients can be calculated by using any of the networks in the corresponding degree sequence-constrained microcanonical ensemble.

The expectation value of any higher-order topological property can be derived exploiting the fact that $\langle a_{ij} \rangle = p_{ij}$ and that different pairs of vertices are statistically independent, which implies $\langle a_{ij} a_{kl} \rangle = p_{ij} p_{kl}$ if (i, j) and (k, l) are distinct pairs of vertices, whereas $\langle a_{ij} a_{kl} \rangle = \langle a_{ij}^2 \rangle = \langle a_{ij} \rangle = p_{ij}$ if $(i, j) = (k, l)$.

2.2.3 *A Second Worked-Out Example: Weighted, Undirected Networks with Constrained Strength Sequence*

In the weighted, undirected case, each graph $\mathbf{G}$ is completely specified by its (symmetric) non-negative weight matrix $\mathbf{W}$ whose entries w_{ij} will be understood as integer-valued. The ensemble with local constraints is in this case known as *weighted configuration model* (WCM) [11] and specifies the strength sequence as the constraint. The Hamiltonian therefore reads

$$H(\mathbf{W}) = \sum_i \theta_i s_i(\mathbf{W}) = \sum_i \sum_{j<i} (\theta_i + \theta_j) w_{ij} \tag{2.22}$$

and one can show that this allows to write the partition function as [10]

$$Z(\boldsymbol{\theta}) \equiv \sum_{\mathbf{W}} e^{-H(\mathbf{W},\boldsymbol{\theta})} = \prod_i \prod_{j<i} (1 - x_i x_j)^{-1} \tag{2.23}$$

and the graph probability as [18]

$$P(\mathbf{W}) = \prod_i \prod_{j<i} q_{ij}(w_{ij}) \tag{2.24}$$

where

$$q_{ij}(w) = (x_i x_j)^w (1 - x_i x_j) \tag{2.25}$$

(with $x_i \equiv e^{-\theta_i}$) is the probability that a link of weight w exists between vertices i and j in the maximum-entropy ensemble of weighted, undirected graphs, subject to specifying the given strength sequence as the constraint.

If the latter is chosen to be the empirical strength sequence $\{s_i(\mathbf{W}^*)\}$ of the particular real network $\mathbf{W}^*$, then Eq. (2.25) yields the exact value of the connection probability in the ensemble of randomized weighted networks with the same average strength sequence as the empirical one, provided that the parameters $\{x_i\}_{i=1}^N$ are set to the values $\{x_i^*\}_{i=1}^N$ that maximize the likelihood $P(\mathbf{W}^*)$ [41]. These values are the solution of the following set of N coupled nonlinear equations:

$$\langle s_i \rangle = \sum_{j \neq i} \frac{x_i x_j}{1 - x_i x_j} = s_i(\mathbf{W}^*) \quad \forall i. \tag{2.26}$$

Once the values $\{x_i^*\}_{i=1}^N$ are found, they are inserted into Eq. (2.25) which allows to easily compute the expectation value $\langle X \rangle^*$ of any topological property X analytically. Equation (2.26) shows that, by construction, the strengths of all vertices are special local quantities whose expected and empirical values are exactly equal: $\langle s_i \rangle^* = s_i(\mathbf{W}^*)$.

The expectation value of any higher-order topological property can be derived exploiting the fact that $\langle w_{ij} \rangle = \sum_w w q_{ij}(w) = x_i x_j/(1 - x_i x_j)$, and that different pairs of vertices are statistically independent, which implies $\langle w_{ij} w_{kl} \rangle = \langle w_{ij} \rangle \langle w_{kl} \rangle$ if $(i - j)$ and $(k - l)$ are distinct pairs of vertices, whereas $\langle w_{ij} w_{kl} \rangle = \langle w_{ij}^2 \rangle$ if $(i - j)$ and $(k - l)$ are the same pair of vertices. The expected value of the power of the weight between vertices i and j is calculated as follows:

$$\langle w_{ij}^\alpha \rangle \equiv \sum_w w^\alpha q_{ij}(w) = (1 - x_i x_j) \mathrm{Li}_{-\alpha}(x_i x_j) \tag{2.27}$$

where $\mathrm{Li}_n(z)$ denotes the Polylogarithm function defined as

$$\mathrm{Li}_n(z) \equiv \sum_{l=1}^{\infty} \frac{z^l}{l^n}. \tag{2.28}$$

The adjacency matrix representing the existence of a link (irrespective of its intensity) between vertex i and vertex j is derived from the weight matrix by setting $a_{ij} = \Theta(w_{ij})$, where $\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ otherwise. The probability that vertices i and j are connected, irrespective of the edge weight, is now $\langle a_{ij} \rangle = p_{ij} \equiv 1 - q_{ij}(0) = x_i x_j$. In analogy with the expectation values of products of weights, we have $\langle a_{ij} a_{kl} \rangle = p_{ij} p_{kl}$ if $(i - j)$ and $(k - l)$ are distinct pairs of vertices, whereas $\langle a_{ij} a_{kl} \rangle = \langle a_{ij}^2 \rangle = \langle a_{ij} \rangle = p_{ij}$ if $(i - j)$ and $(k - l)$ are the same pair of vertices.

2.3 Comparing Models Obtained from Different Constraints

The two worked out examples considered above will be used extensively throughout this book, together with other models. When multiple models are applied to the same set of network data, one needs a rigorous statistical procedure to compare them and choose, loosely speaking, the 'best one'. In fact, judging a model purely on the basis of its performance in reproducing the observed trends represents a naïve way of proceeding exposed to many risks, the most dangerous one being that of preferring models that overfit the data via redundant parameters that have high inter-correlations and provide spurious information on the system [45, 46]. For instance, alternative models are often compared exclusively in terms of the values of their likelihood functions evaluated in their stationary points: the higher the value, the better the model is expected to describe the considered network. However, this procedure lacks a rigorous statistical justification and does not address the parsimony of the models, e.g. the number of parameters.

On the contrary, we would like to rely on a criterion able to unambiguously identify not only the most effective null model in explaining empirical data, but also the most statistically correct one. A more appropriate way of testing the effectiveness of two competing null models (say NM_i and NM_j, where NM_j contains extra parameters with respect to NM_i) is the *Likelihood Ratio Test* (LRT) [47], which prescribes to calculate the quantity

$$D_{NM_i/NM_j} \equiv -2(\mathscr{L}_{NM_i}(\boldsymbol{\theta}_i^*) - \mathscr{L}_{NM_j}(\boldsymbol{\theta}_j^*)) \tag{2.29}$$

(where the symbols $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ indicate the two different sets of Lagrange multipliers that maximize the likelihood of the two models) and compare it to some threshold value determined by some chosen significance level. If D_{NM_i/NM_j} is smaller than the threshold, then model NM_j should be rejected even though its log-likelihood is higher than that of NM_i.

However, the LRT suffers from some limitations [47]. One lies in the fact that the competing null models have to be nested: NM_i has to be a special case of NM_j. Another limitation has to do with the number of models that can be tested: only two alternative hypoteses at a time can be compared, thus making a global ranking of all the models in our set impossible.

So, we prefer a criterion which is suitable for more than two, possibly not nested, competing null models. The *Akaike Information Criterion* (AIC) [45, 46, 48] is one such criterion. It prescribes to calculate the quantity

$$\text{AIC}^*_{NM_i} \equiv 2K_{NM_i} - 2\mathscr{L}_{NM_i}(\boldsymbol{\theta}^*_i) \quad \forall i \tag{2.30}$$

for every null model in the set and then choose the model with the lowest value. Since the above quantity is (twice) the difference between the number of parameters K of null model NM_i and its log-likelihood evaluated in its maximum, such procedure satisfies all our requirements: it is likelihood-based, it discounts the number of model parameters and allows for a comparison among several (not necessarily nested) models.

However, whenever the number n of empirical observations becomes too small with respect to the number of parameters (a rule of thumb being $n/K_{NM_i} < 40$ [45, 46]) the modified quantity

$$\text{AICc}^*_{NM_i} \equiv \text{AIC}^*_{NM_i} + \frac{2K_{NM_i}(K_{NM_i}+1)}{n - K_{NM_i} - 1}, \tag{2.31}$$

providing an extra correction term further penalizing models with many parameters, should be used. When $n \gg K^2_{NM_i}$, AICc converges to AIC and the usual form is recovered. Notice that n has no subscript because the comparison between different null models has to be carried on the same set of observations: naturally, this holds true also for AIC and, generally speaking, for all model selection methods. More precisely, in all the cases of interest for us, our samples will be constituted by the entries of the adjacency matrix, i.e. $n = N(N-1)/2$ observations when dealing with undirected networks and $n = N(N-1)$ observations when dealing with directed networks.

Both AIC and AICc select the most effective model in explaining observations, avoiding (or, at least, strongly reducing) the risk of choosing overfitting models. However, to quantify the relative improvement brought about by the best model, the so called *Akaike weights* can be computed as follows:

$$w^{\text{AIC}}_{NM_i} \equiv \frac{e^{-\frac{\Delta_{NM_i}}{2}}}{\sum_{r=1}^{R} e^{-\frac{\Delta_{NM_r}}{2}}} \tag{2.32}$$

where $\Delta_{NM_i} \equiv \text{AIC}^*_{NM_i} - \min\{\text{AIC}^*_{NM_i}\}_{i=1}^R$, R being the total number of considered null models. The Akaike weight of a specific model is usually interpreted as the probability that that model is, in fact, the best one. Models with $\Delta \leq 2$

are given substantial statistical support, models with $4 \leq \Delta \leq 7$ are given less support and models with $\Delta > 10$ have essentially no support [45, 46, 48–50]. Confidence intervals can also be defined [45, 46, 48–50].

An alternative criterion to AIC, the *Bayesian Information Criterion* (BIC) [45, 46, 48–50], has also been proposed and the corresponding weights defined accordingly. The only, apparently simple but actually substantial, difference lies in the term to be discounted from the maximized likelihood:

$$\mathrm{BIC}^*_{NM_i} \equiv K_{NM_i} \ln n - 2\mathscr{L}_{NM_i}(\boldsymbol{\theta}^*_i) \quad \forall i. \tag{2.33}$$

The first addendum does not only account for the number of parameters, K_{NM_i}, but also for the cardinality of the sample, n. Since BIC discounts the sample cardinality from the very beginning, there is no need to define a corrected Bayesian Information Criterion analogous to AICc. The *Bayesian weights* are defined analogously to the Akaike weights:

$$w^{\mathrm{BIC}}_{NM_i} \equiv \frac{e^{-\frac{\Delta^B_{NM_i}}{2}}}{\sum_{r=1}^{R} e^{-\frac{\Delta^B_{NM_r}}{2}}} \tag{2.34}$$

where now $\Delta^B_{NM_i} \equiv \mathrm{BIC}^*_{NM_i} - \min\{\mathrm{BIC}^*_{NM_i}\}_{i=1}^R$, R being the total number of considered null models. Criteria to interpret BIC weights follow the same lines stated for AIC weights [45, 46, 48–50].

Generally speaking, because of the extra term $\ln n$, BIC is believed to be more restrictive than AIC, as the former favors models with a lower number of parameters than those favored by the latter [45]. However, which criterion performs best, and under which conditions, is still debated and other model-selection methods (such as *multimodel inference*, where some form of average over different models is performed [45]) have been proposed. In this book we will use both criteria and compare them when necessary.

References

1. S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks. Science **296**, 910–913 (2002)
2. S. Maslov, K. Sneppen, A. Zaliznyak, Detection of topological patterns in complex networks: correlation profile of the Internet. Physica A **333**, 529–540 (2004)
3. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks. Science **298**, 824–827 (2002)
4. F. Chung, L. Lu, "Connected components in random graphs with given expected degree sequences", Ann. Combin. **6**(125) (2002)
5. J. Park, M.E.J. Newman, "Origin of degree correlations in the Internet and other networks". Phys. Rev. E **68**, 026112 (2003)
6. M. Catanzaro, M. Boguná, R. Pastor-Satorras, Generation of uncorrelated random scale-free networks. Phys. Rev. E **71**, 027103 (2005)

7. D. Garlaschelli, M.I. Loffredo, Multispecies grand-canonical models for networks with reciprocity. Phys. Rev. E **73**, 015101(R) (2006)
8. D.B. Stouffer, J. Camacho, W. Jiang, L.A.N. Amaral, Evidence for the existence of a robust pattern of prey selection in food webs. Proc. R. Soc. B **274**, 1931–1940 (2007)
9. R. Guimerá, M. Sales-Pardo, L.A.N. Amaral, Classes of complex networks defined by role-to-role connectivity profiles. Nat. Phys. **3**, 63–69 (2007)
10. J. Park, M.E.J. Newman, Statistical mechanics of networks. Phys. Rev. E **70**, 066117 (2004)
11. M.A. Serrano, M. Boguná, "Weighted configuration model" AIP Conf. Proc. **776**(101) (2005)
12. M.A. Serrano, M. Boguná, R. Pastor-Satorras, Correlations in weighted networks. Phys. Rev. E **74**, 055101(R) (2006)
13. M.A. Serrano, Rich-club vs rich-multipolarization phenomena in weighted networks. Phys. Rev. E **78**, 026101 (2008)
14. A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. Proc. Nat. Acad. Sci. **101**, 3747–3752 (2004)
15. T. Opsahl, V. Colizza, P. Panzarasa, J.J. Ramasco, Prominence and Control: The Weighted Rich-Club Effect. Phys. Rev. Lett. **101**, 168702 (2008)
16. K. Bhattacharya, G. Mukherjee, J. Saramaki, K. Kaski, S.S. Manna, "The International Trade Network: weighted network analysis and modelling", J. Stat. Mech., P02002 (2008)
17. G. Bianconi, The entropy of network ensembles. Phys. Rev. E **79**, 036114 (2009)
18. D. Garlaschelli, M.I. Loffredo, Generalized Bose-Fermi statistics and structural correlations in weighted networks. Phys. Rev. Lett. **102**, 038701 (2009)
19. D. Garlaschelli, The weighted random graph model. New J. Phys. **11**, 073005 (2009)
20. R. Milo, N. Kashtan, S. Itzkovitz, M.E.J. Newman, U. Alon, "On the uniform generation of random graphs with prescribed degree sequences", http://arxiv.org/abs/cond-mat/0312028
21. Y. Artzy-Randrup, L. Stone, Generating uniformly distributed random networks. Phys. Rev. E **72**, 056708 (2005)
22. L. Tabourier, C. Roth, J.-P. Cointet, "Generating constrained random graphs using multiple edge switches", http://arxiv.org/abs/1012.3023
23. M.E.J. Newman, S.H. Strogatz, D.J. Watts, Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E **64**, 026118 (2001)
24. G. Caldarelli, *Scale-free Networks. Complex Webs in Nature and Technology* (Oxford University Press, 2007)
25. P. Erdös, A. Rényi, On random graphs. Publicationes Mathematicae Debrecen **6**, 290–297 (1959)
26. A.C.C. Coolen, A. De Martino, A. Annibale, Constrained Markovian dynamics of random graphs. J. Stat. Phys. **136**, 1035–1067 (2009)
27. E.S. Roberts, A.C.C. Coolen, Unbiased degree-preserving randomization of directed binary networks. Phys. Rev. E **85**, 046103 (2012)
28. C.I. Del Genio, H. Kim, Z. Toroczkai, K.E. Bassler, Efficient and exact sampling of simple graphs with given arbitrary degree sequence. PLoS ONE **5**(4), e10012 (2010)
29. H. Kim, C.I. Del Genio, K.E. Bassler, Z. Toroczkai, Constructing and sampling directed graphs with given degree sequences. New J. Phys **14**(2), 023012 (2012)
30. J. Blitzstein, P. Diaconis, A sequential importance sampling algorithm for generating random graphs with prescribed degrees. Internet Mathematics **6**(4), 489–522 (2011)
31. P. Erdös, T. Gallai, "Graphs with prescribed degree of vertices", Mat. Lapok **11**(477) (1960)
32. S. Melnik, A. Hackett, M.A. Porter, P.J. Mucha, J.P. Gleeson, The unreasonable effectiveness of tree-based theory for networks with clustering. Phys. Rev. E **83**, 036112 (2011)
33. M.E.J. Newman, Random graphs with clustering. Phys. Rev. Lett. **103**, 058701 (2009)
34. M. Boguná, R. Pastor-Satorras, A. Vespignani, Cut-offs and finite size effects in scale-free networks. Eur. Phys. J. B **38**, 205–209 (2004)
35. D. Garlaschelli, M.I. Loffredo, Maximum likelihood: Extracting unbiased information from complex networks. Phys. Rev. E **78**, 015101(R) (2008)
36. M.E.J. Newman, Analysis of weighted networks. Phys. Rev. E **70**, 056131 (2004)

37. J.W. Gibbs, *Elementary principles in statistical mechanics* (Charles Scribner's Sons, New York, 1902)
38. C. Shannon, A mathematical theory of communication. Bell System Tech. Jour. **27**(379–423), 623–656 (1948)
39. E.T. Jaynes, "Information theory and statistical mechanics", Phys. Rev. **106**(620) (1957)
40. E.T. Jaynes, "On the rationale of maximum-entropy methods", Proc. IEEE **70**(939) (1982)
41. T. Squartini, D. Garlaschelli, Analytical maximum-likelihood method to detect patterns in real networks. New J. Phys. **13**, 083001 (2011)
42. P. Holland, S. Leinhardt, Sociological Methodology, ed. by D. Heise. (Jossey-Bass, San Francisco, 1975), pp. 1–45
43. S. Wasserman, K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994)
44. T.A.B. Snijders, "Markov chain Monte Carlo estimation of exponential random graph models", J. Soc. Struct. **3**(2) (2002)
45. K.P. Burnham, D.R. Anderson, *Model selection and multi-model inference: a practical information-theoretic approach* (Springer, New York, 2002)
46. J.B. Johnson, K.S. Omland, Model selection in ecology and evolution. Trends Ecol. Evol. **9**, 101–108 (2004)
47. D.R. Cox, D.V. Hinkley, *Theoretical statistics* (Chapman and Hall, Boca Raton, 1974)
48. H. Akaike, A new look at the statistical model identification. IEEE Trans. Aut. Cont. **19**, 716–723 (1974)
49. K.P. Burnham, D.R. Anderson, Multimodel inference: understanding AIC and BIC in Model Selection. Soc. Met. Res. **33**, 261–304 (2004)
50. E.J. Wagenmakers, S. Farrell, AIC model selection using Akaike weights. Psych. Bull Rev. **11**, 192–196 (2004)

Chapter 3
Pattern Detection

In truth at first Chaos came to be, but next wide-bosomed Earth, the ever-sure foundation of all the deathless ones who hold the peaks of snowy Olympus, and dim Tartarus in the depth of the wide-pathed Earth, and Eros, fairest among the deathless gods, who unnerves the limbs and overcomes the mind and wise counsels of all gods and all men within them.

—Hesiod, $\Theta\varepsilon o\gamma o\nu\iota\alpha$

Abstract Here we show that the maximum-entropy method introduced in the previous chapter can be used to define various benchmarks (i.e. null models) to assess the presence of virtually any structural pattern of interest in a real network. Such patterns include assortativity, clustering, reciprocity, motifs and possibly the weighted counterparts of many of these properties. They are detected as statistically significant deviations of the real network from the random benchmark. We apply our toolkit to various biological, transportation, economic and financial networks. Importantly, since the maximum-entropy method allows for any level of density or clustering, the same methodology can be applied with equal and full reliability across all the networks considered. The results of this analysis may therefore differ from the corresponding results that would be obtained using alternative methods that do require some (often unverified) assumption of a sparse or tree-like topology. We indeed identify clear patterns, some of which turn out to be unaccessible or incorrectly interpretable using other methods, in the analysis of degree-degree and strength-strength correlations, anomalously dense connections in the neighborhoods of nodes and statistically surprising abundances of dyadic and triadic motifs. Incidentally, some of the detected motifs turn out to provide early-warning signals of the 2008 financial crisis in the Dutch interbank network and a measure of temporal (in)stability of economic networks, leading to a sort of 'equilibrium' and 'out of equilibrium' classification.

T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-69438-2_3

3.1 Detecting Assortativity and Clustering

We start with the problem of identifying and quantifying two important structural properties in real networks: *assortativity*, i.e. the tendency of nodes of similar degree to connect to each other, and *clustering*, i.e. the tendency of triples of nodes to connect together. We discuss the case of binary undirected networks in Sect. 3.1.1 and that of binary directed networks in Sect. 3.1.2. The case of weighted networks will be considered later in Sect. 3.4.1, as part of a separate discussion focusing on the generalization of various binary topological properties to the weighted case.

3.1.1 Undirected Networks

In social network analysis, *assortativity* is defined as the tendency of nodes that share similar features (e.g. age, gender, beliefs, culture, etc.) to be more likely connected with each other than with other nodes in the network. Assortativity therefore manifests itself as an increased abundance of links among nodes of similar type, as compared to a network where no such tendency is in place. The opposite tendency, known as *disassortativity* and corresponding to a decreased abundance of links among similar nodes, is also a possible property of real networks. Being one of the most studied topological properties of networks, (dis)assortativity has been defined and characterized in many different ways. Different definitions provide alternative quantifications of the (anti)correlation between the features of adjacent nodes. A feature of special importance, being already 'visible' in a network even if no external information about the properties of nodes (such as age, gender, etc.) is known, is the degree itself. The resulting notion of *(dis)assortativity by degree* indicates whether nodes of similar degree tend to have a preference for (not) being connected to each other. (Dis)assortativity by degree is found to be quite strong in real networks and has been shown to dramatically affect various structural and dynamical features [1]. Here we follow Ref. [2] and focus on a metric that provides a quantification of assortativity by degree for each node separately. In particular, given a real binary undirected network $\mathbf{A}^*$ we define the *average nearest neighbour degree* (ANND for short) of node i as

$$k_i^{nn}(\mathbf{A}^*) \equiv \frac{\sum_{j\neq i} a_{ij}^* k_j^*}{k_i^*} = \frac{\sum_{j\neq i}\sum_{k\neq j} a_{ij}^* a_{jk}^*}{\sum_{j\neq i} a_{ij}^*}, \tag{3.1}$$

where $k_i^* \equiv k_i(\mathbf{A}^*) = \sum_{j\neq i} a_{ij}^*$ is the degree of node i in the network $\mathbf{A}^*$. It should be noted that, while the degree is a first-order property that only depends on the number of links (topological paths of length one) entering a vertex, the ANND is a second-order property determined by paths of length two, i.e. the terms of the form $a_{ij}^* a_{jk}^*$ in the above formula.

Along the same line, one can consider third-order correlations by introducing topological properties involving paths of length three. *Clustering*, i.e. the tendency of triples of nodes to form triangles, is a clear and popular example of such properties. In social network analysis, clustering is also known as *triadic closure*, as it represents the tendency of triads of nodes to be 'closed', i.e. fully connected. Also this property is believed to be quite strong in real networks (e.g. pairs of your friends are typically friends of each other). Again, among the many possible definitions of clustering, we follow Ref. [2] and focus on the *local* (i.e. node-specific) *clustering coefficient* defined as

$$c_i(\mathbf{A}^*) \equiv \frac{\sum_{j\neq i}\sum_{k\neq i,j} a^*_{ij}a^*_{jk}a^*_{ki}}{k^*_i(k^*_i-1)} = \frac{\sum_{j\neq i}\sum_{k\neq i,j} a^*_{ij}a^*_{jk}a^*_{ki}}{\sum_{j\neq i}\sum_{k\neq i,j} a^*_{ij}a^*_{ki}}. \tag{3.2}$$

In the above formula, the numerator counts how many pairs of neighbours of node i are connected to each other, while the denominator counts how many pairs of neighbours of node i (irrespective of whether they are connected to each other) exist. So $c_i(\mathbf{A}^*)$ is a measure of the local link density, i.e. the fraction of links realized in the neighbourhood of node i.

A useful way of exploiting the local character of both $k^{nn}_i(\mathbf{A}^*)$ and $c_i(\mathbf{A}^*)$ is plotting their values versus that of $k_i(\mathbf{A}^*)$ for all nodes in a given network. In Fig. 3.1 we provide such plots for several real-world networks, namely the network of the 500 largest US airports [3], the synaptic network of the nematode *C. elegans* [4], the protein interaction network of the bacterium *H. pylori* [5] and an instance of the Italian interbank network [6]. These networks are among the most studied ones in the literature. Normally, the above plots are used in order to determine whether nontrivial correlations exist in the structure of real networks. In particular, the departure of the resulting trends from a 'flat' (i.e., degree-independent) behaviour is interpreted as a sign of interesting patterns, because such flat behavior is found for uncorrelated networks modeled as Erdős-Rényi random graphs (see Sect. 2.1.1).

However, as we discussed in Sect. 2.1.1, a comparison with a completely homogeneous model such as the Erdős-Rényi random graph fails to control for the effects of the obserbed heterogeneity of nodes in the real network considered. Such a null model destroys not only, as we desire, the empirical second- and third-order correlations we want to identify in the real networks, but also the first-order properties of nodes, such as their degree. As a consequence, the observed deviation of the real network from the model may be partly due to the heterogeneity of nodes, which should not be confused with distinct higher-order properties.

In order to disentagle first- from higher-order effects, one should therefore use the configuration model (see Sect. 2.1.1) as a more stringent null model where the degree of each node is set equal to its empirical value and the network is completely random otherwise. And this is where an incorrect implementation of the model may lead to highly biased results and incorrect interpretations.

We therefore compare the correlation structure of the original networks, as measured by the dependence of $k^{nn}_i(\mathbf{A}^*)$ and $c_i(\mathbf{A}^*)$ on $k_i(\mathbf{A}^*)$, with the expected values $\langle k^{nn}_i\rangle^*$ and $\langle c_i\rangle^*$ obtained analytically using the maximum-entropy method:

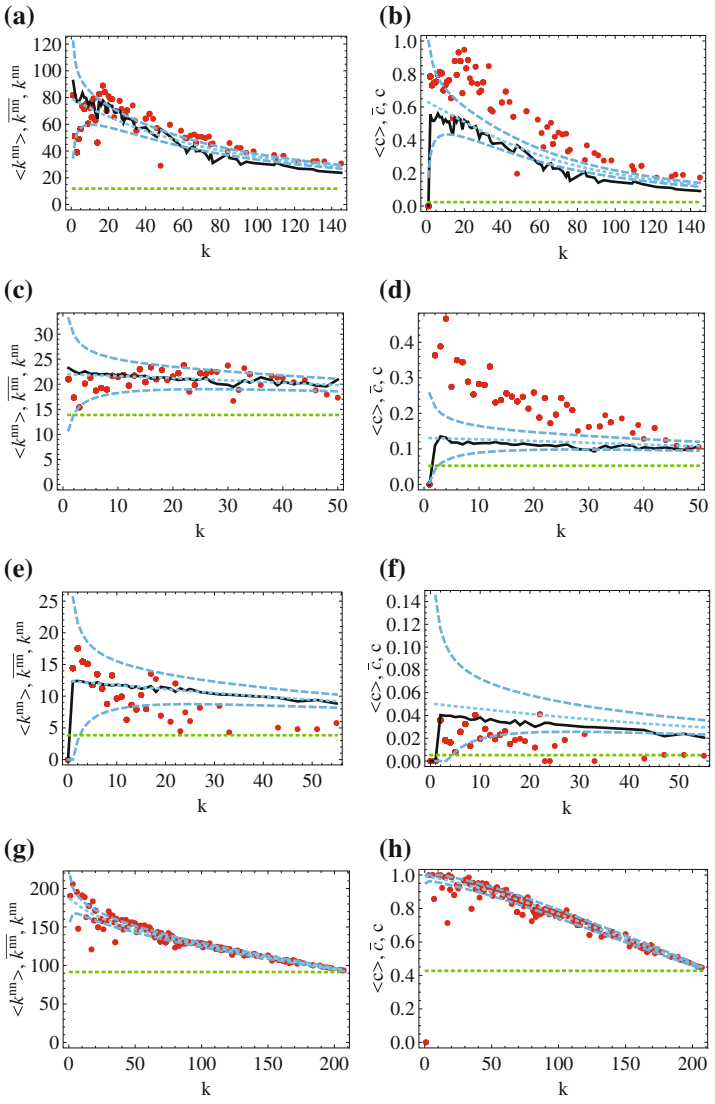

Fig. 3.1 Assortativity (left) and clustering (right) patterns in various real-world binary undirected networks. Red points are the empirical data, black solid curves are averages over the configuration model obtained using the local rewiring algorithm (see Sect. 2.1.2) and blue dashed curves are the analytical expectations ($\pm$ one standard deviation) obtained using the maximum-entropy method (see Sect. 2.1.2). Green curves are the flat expectations under the Erdös-Rényi random graph model (see Sect. 2.1.1), which highlights the average level of correlation in the random case. The panels show the average nearest neighbour degree k_i^{nn} versus k_i (left) and the local clustering coefficient c_i versus k_i (right) for: **a** and **b** the network of the largest US airports ($N = 500$ nodes) [3], **c** and **d** the synaptic network of *Caenorhabditis elegans* ($N = 264$ nodes) [4], **e** and **f** the protein-protein interaction network of *Helicobacter pylori* ($N = 732$ nides) [5], **g** and **h** the network of liquidity reserves exchanges between Italian banks in 1999 [6] ($N = 215$ nodes). (Adapted from Ref. [2])

$$\langle k_i^{nn}\rangle^* = \frac{\sum_{j\neq i} p_{ij}^* \langle k_j\rangle^*}{\langle k_i\rangle^*} = \frac{\sum_{j\neq i} p_{ij}^* k_j^*}{k_i^*} = \frac{\sum_{j\neq i}\sum_{k\neq j} p_{ij}^* p_{jk}^*}{\sum_{j\neq i} p_{ij}^*}, \tag{3.3}$$

$$\langle c_i\rangle^* = \frac{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}^* p_{jk}^* p_{ki}^*}{\langle k_i(k_i-1)\rangle^*} = \frac{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}^* p_{jk}^* p_{ki}^*}{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}^* p_{ki}^*}. \tag{3.4}$$

Note that we are averaging the values of $k_i^{nn}(\mathbf{A}^*)$ and $c_i(\mathbf{A}^*)$ over all vertices with the same degree: this makes our comparison with the values $\langle k_i^{nn}\rangle^*$ and $\langle c_i\rangle^*$ consistent, since both real and randomized quantities can be plotted using the same values $\langle k_i\rangle^* = k_i(\mathbf{A}^*)$ on the abscissa (we use the same strategy in what follows). We also highlight the region within one standard deviation around the average by plotting the curves $\langle k_i^{nn}\rangle^* \pm \sigma^*[k_i^{nn}]$ and $\langle c_i\rangle^* \pm \sigma^*[c_i]$.

For the sake of comparison, we also report the average values obtained sampling the microcanonical ensemble with the standard local rewiring algorithm [7, 8] (see Sect. 2.1.2) and the expected values obtained under the Erdős-Rènyi random graph model with the same number of links (see Sect. 2.1.1). As we mentioned in Sect. 2.1.2, the microcanonical method requires the generation of many randomized variants, many rewirings per variant and the measurement of k_i^{nn} and c_i on each variant separately, plus a final averaging. Moreover, it is subject to bias. By contrast, the canonical maximum-entropy method is unbiased and only requires the preliminary estimation of the $\{x_i^*\}_{i=1}^N$. Then the calculation of $\langle k_i^{nn}\rangle$ and $\langle c_i\rangle$ takes exactly the same time as that of the empirical values.

In this particular case, the two approaches yield very similar results. The reason lies in the low link density, making all the considered networks satisfy Eq. (2.1), i.e. reducing the bias affecting the microcanonical approach, for all practical purposes, to zero.

The above results allow us to interpret the effect of the degree sequence on higher-order properties. Firstly, the trends displayed by the CM are not flat as those expected in the random graph case. This confirms that residual structural correlations, simply due to the enforced constraint, are still present after the rewiring has taken place. The presence of these correlations does not require any additional explanation besides the existence of the constraints themselves. This is very different from the picture one would get by using the (wrong) expectation of Eq. (2.3) which would yield flat trends as well, naively suggesting that correlations can never be traced back to the degree sequence alone. Secondly, while the trends observed in all the networks considered are always decreasing, they unveil different correlation patterns when compared to the randomized trends. The real interbank data are almost indistinguishable from the randomized curves, meaning that structural constraints can fully explain the observed behaviour of higher-order network properties. Instead, in the airport network the randomized curves lie below the real data (except for an opposite trend of $\langle k_i^{nn}\rangle$ for low degrees). This means that the real network is more correlated than the baseline randomized expectation and indicates that additional mechanisms producing positive correlations must be present on top of structural effects. By contrast, in the *H. pylori*'s protein network the expected curves lie above the real data, suggesting the presence of mechanisms producing negative correlations. Therefore seemingly similar trends can

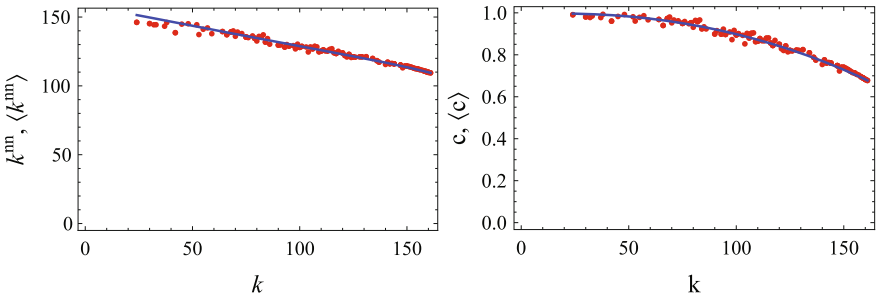

Fig. 3.2 Left panel: average nearest neighbor degree k_i^{nn} versus degree k_i in the 2002 snapshot of the real binary undirected WTW (red points) and corresponding average over the maximum-entropy ensemble with specified degrees (blue curve). Right panel: clustering coefficient c_i versus degree k_i (red points) and corresponding average over the maximum-entropy ensemble with specified degrees (blue curve). Adapted from Ref. [9]

actually correspond to very different types of structural organization. This confirms that measuring the topological properties alone is uninformative, making the comparison between real data and randomized ensembles essential. Thus the possibility to analytically and quickly characterize such ensembles is a remarkable advantage of the maximum-entropy approach.

The close agreement between the interbank network and the corresponding maximum-entropy model suggests that, for some networks, the knowledge of the degree sequence may be predictive of other higher-order properties. To further explore this hypothesis, we repeat the previous analysis on another economic network: the World Trade Web (WTW) or International Trade Network (ITN), whose nodes represent world countries and whose links represent (possibly directed and weighted) trade connections. In Fig. 3.2 we show the plot of the average nearest neighbour degree versus the degree for the WTW in its undirected and unweighted projection, where any two countries are connected if a trade relation in any direction exists between them. We observe a decreasing trend, indicating that on average countries trading with highly connected countries have a few trade partners, whereas countries trading with poorly connected countries have many trade partners. If we compare this trend with the one followed by the corresponding randomized quantity obtained under the binary configuration model, we find a very close agreement. This means that the empirical assortativity of this network, like the one of the interbank network considered above, is entirely explained by the degree sequence.

A similar result is found for the behavior of the *clustering coefficient* c_i. Again, we find a decreasing trend of c_i as a function of k_i (see Fig. 3.2). This means that trade partners of highly connected countries are poorly interconnected, whereas partners of poorly connected countries are highly interconnected. However, if this trend is compared with the one displayed by the randomized quantity we again find a very close agreement. This signals that in the WTW also the profile of the clustering coefficient is completely explained by the constraint on the degree sequence and does

not imply the presence of meaningful indirect interactions on top of a concatenation of direct interactions alone.

The above results show that the patterns observed in the binary undirected description of the interbank and trade networks considered do not require, besides the fact that different banks and countries have specific numbers of trade partners, the presence of higher-order mechanisms as an additional explanation. On the other hand, the fact that the degrees alone are enough to explain higher-order network properties means that the degree sequence is an important structural pattern in its own. This result suggests that it may be in principle possible to reconstruct the whole structure of various financial and economic networks from the knowledge of their degree sequence alone, an idea that will be explored in great detail in the next chapter. At the same time, from the perspective of economic modeling this result highlights the importance of including the degree sequence among the focuses of theories and models of banking and trade, which are instead currently oriented mainly at reproducing the weighted structure, rather than the topology of the network.

3.1.2 Directed Networks

When considering binary, directed networks the second-order topological properties are measured by the outward ANND and the inward ANND, which are defined as natural generalizations of Eq. (3.1):

$$k_i^{nn,out}(\mathbf{A}^*) \equiv \frac{\sum_{j\neq i} a_{ij}^* k_j^{out*}}{k_i^{out*}} = \frac{\sum_{j\neq i}\sum_{k\neq j} a_{ij}^* a_{jk}^*}{\sum_{j\neq i} a_{ij}^*}, \tag{3.5}$$

$$k_i^{nn,in}(\mathbf{A}^*) \equiv \frac{\sum_{j\neq i} a_{ji}^* k_j^{in*}}{k_i^{in*}} = \frac{\sum_{j\neq i}\sum_{k\neq j} a_{ji}^* a_{kj}^*}{\sum_{j\neq i} a_{ji}^*}. \tag{3.6}$$

Given a particular real network $\mathbf{A}^*$ and a measured topological property $X(\mathbf{A}^*)$, the maximum-entropy method allows to analytically obtain the expectation value $\langle X\rangle^*$ and standard deviation $\sigma^*[X]$ across the ensemble of binary directed graphs with, on average, the same directed degree sequences $\mathbf{k}^{out}(\mathbf{A}^*)$ and $\mathbf{k}^{in}(\mathbf{A}^*)$ as $\mathbf{A}^*$ (*directed configuration model*, DCM). In this case the method makes use of two N-dimensional vectors $\mathbf{x}, \mathbf{y}$ of auxiliary variables and requires that these parameters are set to the particular values $\mathbf{x}^*$, $\mathbf{y}^*$ that solve the following set of $2N$ coupled nonlinear equations:

$$\langle k_i^{out}\rangle = \sum_{j\neq i} \frac{x_i^* y_j^*}{1 + x_i^* y_j^*} = k_i^{out}(\mathbf{A}^*) \quad \forall i, \tag{3.7}$$

$$\langle k_i^{in}\rangle = \sum_{j\neq i} \frac{x_j^* y_i^*}{1 + x_j^* y_i^*} = k_i^{in}(\mathbf{A}^*) \quad \forall i. \tag{3.8}$$

The quantities $\mathbf{x}^*$, $\mathbf{y}^*$ allow to obtain $\langle X \rangle^*$ and $\sigma^*[X]$ analytically and quickly, outperforming the directed version of the LRA. Note that, as in the undirected case, the method only makes use of the sufficient statistics of the problem.

In Fig. 3.3 we plot the observed values $k_i^{nn,in}(\mathbf{A}^*)$ versus $k_i^{in}(\mathbf{A}^*)$ and $k_i^{nn,out}(\mathbf{A}^*)$ versus $k_i^{out}(\mathbf{A}^*)$, as well as the expectations $\langle k_i^{nn,in} \rangle^* \pm \sigma^*[k_i^{nn,in}]$ and $\langle k_i^{nn,out} \rangle^* \pm \sigma^*[k_i^{nn,out}]$ obtained using the DCM, for three real directed networks: the neural network of *C. elegans* [4] (now in its directed version), the metabolic network of *E. coli* [16] and the *Little Rock Lake* food web [17]. As before, we also show the microcanonical average obtained using the LRA and the expectation under the *directed random graph model* (DRG) with the same number of links. Again, we find a very good agreement between the two approaches, confirming that the maximum-entropy method yields the correct prediction in incredibly shorter time. For the *C. elegans* network (Fig. 3.3a, b), we also show the microcanonical standard deviations, which turn out to be indistinguishable from the canonical ones. We also confirm that while some networks (*C. elegans* and *E. coli*) are almost consistent with the null model, others (*Little Rock*) deviate significantly.

However, the most interesting point for the present analysis is that, while for the undirected networks considered above all randomized trends were decreasing, in this case we find that the three randomized trends behave in totally different ways. In the neural network, both $\langle k_i^{nn,in} \rangle^*$ and $\langle k_i^{nn,out} \rangle^*$ are approximately constant. This means that the baseline behavior for both quantities is flat and uncorrelated (as in the directed random graph, but at a different level). By contrast, in the metabolic network the expected curves are decreasing and, thus, the ensemble of randomized networks is disassortative as for the undirected graphs considered above. Finally, in the food web the constraints enforce unusual positive correlations and the randomized ensemble is even assortative. Interestingly, while it is expected that random networks with specified degrees display a disassortative behavior [8, 14], the assortative trend is totally surprising. This is because the maximum-entropy method extracts the hidden variables directly from the specific real world network, rather than drawing them from ad hoc distributions. The resulting values can be distributed in a very complicated fashion, invalidating the results obtained under other hypotheses. To further highlight this important point, we selected three more food webs characterized by a particularly small size (see Fig. 4.2). Small networks cannot be described by approximating the mass probability function of their topological properties (such as the degree) with a continuous probability density. Therefore in this case the difference between the expectations obtained by drawing the $\mathbf{x}$ and $\mathbf{y}$ values from analytically tractable continuous distributions and those obtained by solving Eq. (3.7) using the empirical degrees is particularly evident.

We now consider the binary, directed description of the WTW, with an interest in understanding whether the introduction of directionality changes the picture we have described so far. We consider the usual quantities

$$k_i^{in/in}(\mathbf{A}^*) \equiv \frac{\sum_{j\neq i} a_{ji}^* k_j^{in*}}{k_i^{in*}} = \frac{\sum_{j\neq i}\sum_{k\neq j} a_{ji}^* a_{kj}^*}{\sum_{j\neq i} a_{ji}^*}, \tag{3.9}$$

$$k_i^{out/out}(\mathbf{A}^*) \equiv \frac{\sum_{j\neq i} a_{ij}^* k_j^{out*}}{k_i^{out*}} = \frac{\sum_{j\neq i}\sum_{k\neq j} a_{ij}^* a_{jk}^*}{\sum_{j\neq i} a_{ij}^*} \tag{3.10}$$

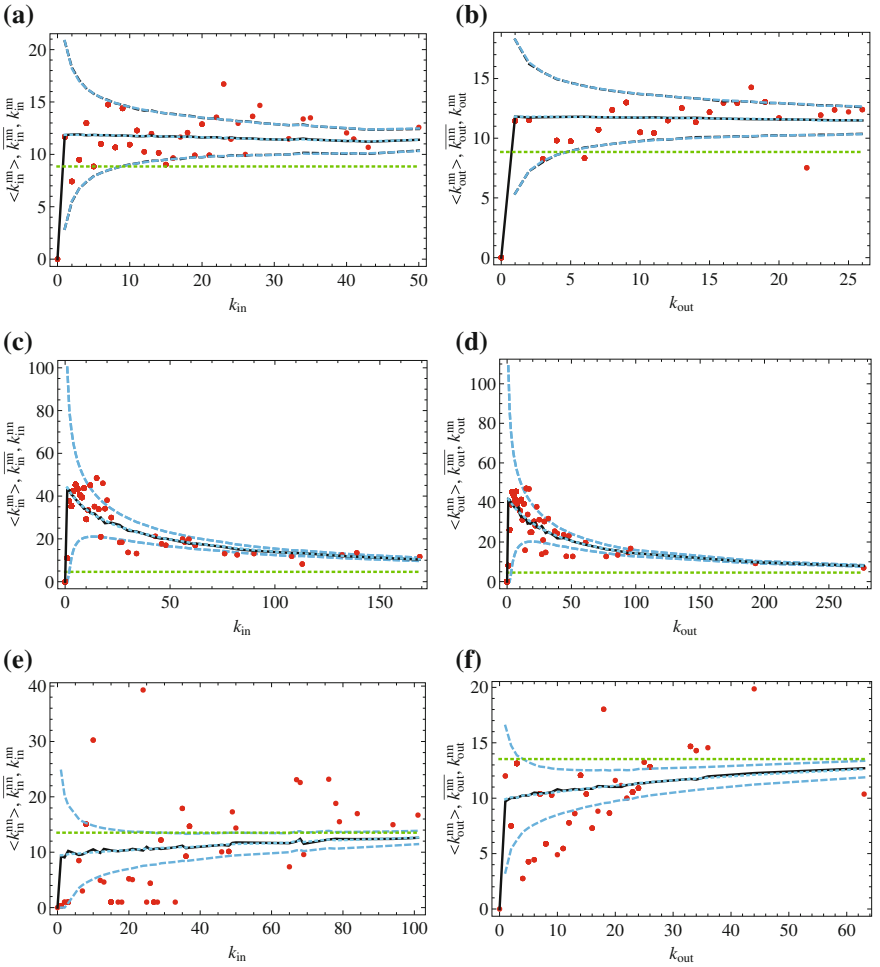

Fig. 3.3 Application of the maximum-entropy method to directed networks. Red points are the empirical data, the black solid curves are expectations under the directed configuration model using the local rewiring algorithm and the blue dashed curves are the exact expectations obtained using the maximum-entropy method (± one standard deviation). The green curves are the flat expectations under the directed version of the (Erdös-Rényi) RG model. The panels report $k_i^{nn,in}$ versus k_i^{in} (left) and $k_i^{nn,out}$ versus k_i^{out} (right) for: **a** and **b** the directed neural network of *Caenorhabditis elegans* ($N = 264$) [4], **c** and **d** the metabolic network of *Escherichia coli* ($N = 1078$) [16], **e** and **f** the *Little Rock Lake* food web ($N = 183$) [17]. For the *C. elegans* network, we also show the microcanonical standard deviations obtained using the LRA (black dotted curves), which are indistinguishable from the canonical ones. Adapted from Ref. [2]

plus the additional ones

$$k_i^{in/out}(\mathbf{A}^*) \equiv \frac{\sum_{j\neq i} a_{ji}^* k_j^{out*}}{k_i^{in*}} = \frac{\sum_{j\neq i}\sum_{k\neq j} a_{ji}^* a_{jk}^*}{\sum_{j\neq i} a_{ji}^*}, \tag{3.11}$$

$$k_i^{out/in}(\mathbf{A}^*) \equiv \frac{\sum_{j\neq i} a_{ij}^* k_j^{in*}}{k_i^{out*}} = \frac{\sum_{j\neq i}\sum_{k\neq j} a_{ij}^* a_{kj}^*}{\sum_{j\neq i} a_{ij}^*}. \tag{3.12}$$

The analysis of the four directed versions of the ANND is shown in Fig. 3.4. We immediately see that all quantities still display a disassortative trend, with some differences in the ranges of observed values. Again, all the four empirical behaviors are in striking accordance with the null model. This means that both the decreasing trends and the ranges of values displayed by all quantities are well reproduced by a collection of random graphs with the same in-degrees and out-degrees as the real network.

We might in principle consider various directed counterparts of the clustering coefficient which have been defined in literature [18, 19]. However, we will later consider all the possible triadic patterns of connectivity, which provide an even more refined description of the tendency to form directed triangles in a directed network. We therefore skip the discussion of directed clustering and move on to different topological patterns.

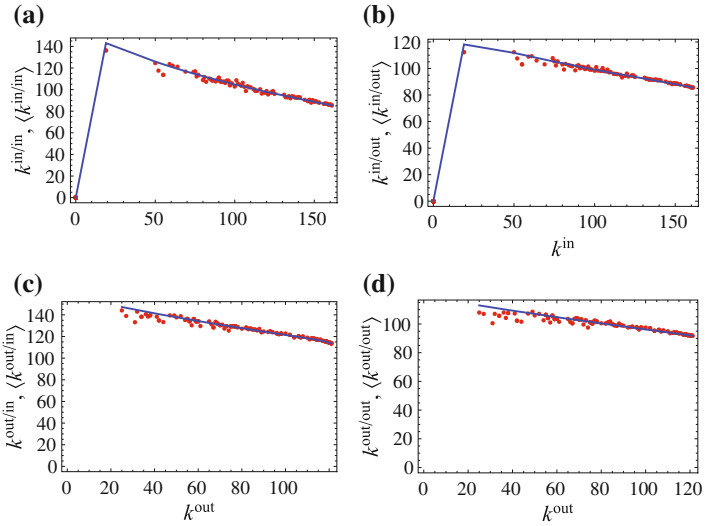

Fig. 3.4 Directed average nearest neighbor degrees versus vertex degrees in the 2002 snapshot of the real binary directed WTW (red points) and corresponding averages over the maximum-entropy ensemble with specified out-degrees and in-degrees (blue curves). **a** $k_i^{in/in}$ versus k_i^{in}. **b** $k_i^{in/out}$ versus k_i^{in}. **c** $k_i^{out/in}$ versus k_i^{out}. **d** $k_i^{out/out}$ versus k_i^{out}. Adapted from Ref. [9]

Fig. 3.5 The dyadic, binary, directed motifs: $L^{\rightarrow}$, $L^{\leftrightarrow}$ and $L^{\nleftrightarrow}$

3.2 Detecting Dyadic Motifs

For a binary network there are four possible dyadic configurations for each pair of nodes (see Fig. 3.5). In fact, node i can be linked to node j through a single link pointing from i to j, through a link pointing from j to i, through a pair of links pointing in opposite directions or through no link at all. In mathematical terms, it is possible to associate each of these four patterns, also called dyadic motifs, to one of the following binary quantities: $a_{ij}^{\rightarrow} \equiv a_{ij}(1-a_{ji})$, $a_{ij}^{\leftarrow} \equiv a_{ji}(1-a_{ij})$, $a_{ij}^{\leftrightarrow} \equiv a_{ij}a_{ji}$ and $a_{ij}^{\nleftrightarrow} \equiv (1-a_{ij})(1-a_{ji})$. Notice that when counting the global number of such patterns the first two configurations become actually the same: thus, at an aggregate level, we only speak of three non-isomorphic dyadic motifs.

The number of occurrences N_m of a particular dyadic motif m ($m = L^{\rightarrow}, L^{\leftrightarrow}, L^{\nleftrightarrow}$) can be written in two equivalent ways (see Table 3.1): the first one employs products of adjacency matrix elements, a_{ij} the second one employs the quantities $a_{ij}^{\rightarrow}$, $a_{ij}^{\leftarrow}$, $a_{ij}^{\leftrightarrow}$ and $a_{ij}^{\nleftrightarrow}$.

If N_m denotes the number of occurrences of a particular motif m (Fig. 3.5), the maximum-entropy method allows to calculate the expected number $\langle N_m \rangle^*$ (see Table 3.2) and standard deviation $\sigma^*[N_m]$ exactly and thus to obtain the z-score

$$z_m \equiv \frac{N_m(\mathbf{A}^*) - \langle N_m \rangle^*}{\sigma^*[N_m]} \tag{3.13}$$

analytically. If the observations were exactly reproduced by the null model, then the z-scores would be exactly zero. On the other hand, significantly large positive or negative z-scores indicate an over- or under-estimation of the motifs' empirical

Table 3.1 Classification and definitions of the dyadic motifs

Motif m	N_m: 1st definition	N_m: 2nd definition
$L^{\rightarrow}$	$\sum_i \sum_{j \neq i} a_{ij}(1-a_{ji})$	$\sum_i \sum_{j \neq i} a_{ij}^{\rightarrow}$
$L^{\leftrightarrow}$	$\sum_i \sum_{j \neq i} a_{ij}a_{ji}$	$\sum_i \sum_{j \neq i} a_{ij}^{\leftrightarrow}$
$L^{\nleftrightarrow}$	$\sum_i \sum_{j \neq i} (1-a_{ij})(1-a_{ji})$	$\sum_i \sum_{j \neq i} a_{ij}^{\nleftrightarrow}$

Table 3.2 Expectation values of the dyadic motifs

Motif m	$\langle N_m \rangle_{DCM}$
$\langle L^{\rightarrow} \rangle$	$\sum_i \sum_{j \neq i} p_{ij}(1-p_{ji})$
$\langle L^{\leftrightarrow} \rangle$	$\sum_i \sum_{j \neq i} p_{ij}p_{ji}$
$\langle L^{\nleftrightarrow} \rangle$	$\sum_i \sum_{j \neq i} (1-p_{ij})(1-p_{ji})$

abundance respectively. The meaning of the z-scores is well defined for normally distributed variables: in this case, the deviations can be nicely quantified in terms of probabilities, as the intervals $z_m = \pm 1, \pm 2, \pm 3$ select regions enclosing a probability of 68, 95 and 99.7%, respectively. Choosing a threshold allows the identification of significantly deviating patterns. While for non-normally distributed variables it is impossible to attach probabilities to z-scores, large values still highlight the most deviating patterns. Thus, the value of $z[N_m]$ indicates by how many standard deviations the observed and expected numbers of occurrences of motif m differ. Large values of $z[N_m]$ indicate motifs that are either over- or under-represented under the particular null model considered and that are therefore not explained by the lower-order constraints enforced.

We are now going to look at another example of financial network, namely the Dutch Interbank Network (DIN). This time we are interested in the dynamical evolution of the network, in particular in the identification of possible ongoing structural changes associated with the financial crisis of 2007-2008. To this end, we construct quarterly time series of the z-scores of the three non-isomorphic dyadic motifs, ranging from the beginning of 1998 to the end of 2008. When using the DRG as a null model, we find that from 1998 to 2007 these time series display moderate fluctuations around roughly stationary values, while in 2008 they all suddenly jump to different values (see top panel of Fig. 3.6). This means that, while the DRG does identify a global structural change provoked by the economic crisis (emphasizing that the critical configuration is 'anomalous' with respect to the previous decade), it does not provide any useful early-warning signal. Note that the fact that the DRG correctly identifies the 'crisis' only in terms of dyadic properties is in any case a fundamental result showing that there are clear signatures of the crisis in the DIN's topology. Without this preliminary observation, looking for early-warning signals in the evolution of the dyadic properties themselves would have no empirical justification.

Performing the same analysis under the DCM yields a completely different result [20]. The bottom panel of Fig. 3.6 shows that in this case the dyadic z-scores undergo a gradual evolution towards the collapsed configuration, thus providing an early-warning signal of the crisis. Remarkably, after a period of minor fluctuations, all the trends of the dyadic z-scores show a sudden inversion of sign at the beginning of 2005, thus backdating the beginning of the DIN's major structural change three years before its dramatic manifestation in 2008 [20].

This remarkable result, which we reprise also in the next section, shows that the maximum-entropy method can really extract hidden patterns from network data. While in the previous sections we had been concerned with static topological patterns, here we have identified a dynamical pattern, i.e. a regime shift.

At the same time, the above result indicates that only a careful choice of the constraints may be able to unveil the relevant pattern. In particular, controlling for the heterogeneous degrees of banks appears to be crucial. We will encounter a similar result in the next chapter when dealing with the problem of network reconstruction.

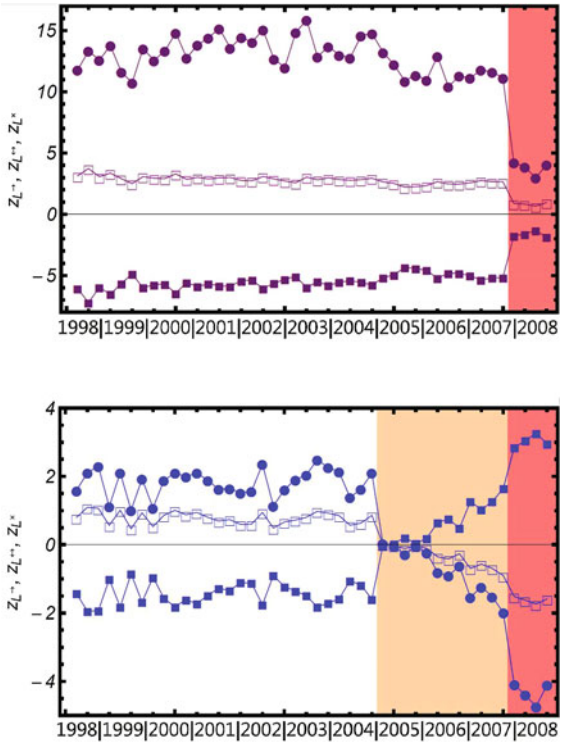

Fig. 3.6 z-scores of the 3 dyadic, binary, directed motifs (■ - $L^{\rightarrow}$, ● - $L^{\leftrightarrow}$, □ - $L^{\nrightarrow}$) for the 44 quarterly snapshots of the DIN under the DRG (top panel) and the DCM (bottom panel). The orange-shaded area corresponds to the 'pre-crisis' period, while the red-shaded area corresponds to the crisis period

3.2.1 Reciprocity

The dyadic pattern $m = L^{\leftrightarrow}$ has a special importance in the study of the *reciprocity* of directed networks. We therefore devote a separate discussion to this particular pattern. Indeed, the study of reciprocity has a long tradition in social science [21] as a way to quantify how many 'ties' (directed links) are reciprocated in a social network of 'actors' (vertices). The reciprocal link of a directed link pointing from i to j is a link pointing from j to i. A link is reciprocated if its reciprocal one is present in the network. In terms of the adjacency matrix of the graph, two reciprocated links are present between i and j if and only if $a_{ij} = a_{ji} = 1$. In a natural way, it follows that the *reciprocity* is the fraction of links having a reciprocal partner [22] and for the observed matrix $\mathbf{A}^*$ it is defined as

$$r(\mathbf{A}^*) \equiv \frac{\sum_i \sum_{j \neq i} a^*_{ij} a^*_{ji}}{L^*} \equiv \frac{L^{\leftrightarrow *}}{L^*}. \tag{3.14}$$

As for the other quantities defined so far, calculating the reciprocity is not enough to conclude whether its value is statistically significant. Can the measured value be the result of mere chance? To answer this question, let us implement the *directed*

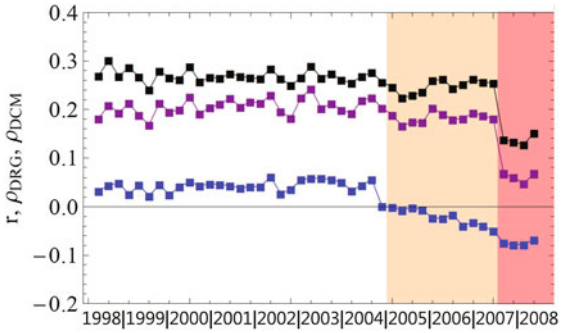

Fig. 3.7 Temporal evolution of the observed reciprocity r (black) of the Dutch Interbank Network in the time-interval 1998–2008 [20] and of ρ under the DRG (purple) and the DCM (blue). The orange-shaded area corresponds to the 'pre-crisis' period, while the red-shaded area corresponds to the crisis period

random graph model (DRG in what follows) and the DCM to compare the observed r with its expected value:

$$\langle r \rangle^* = \frac{\sum_i \sum_{j \neq i} p^*_{ij} p^*_{ji}}{\sum_i \sum_{j \neq i} p^*_{ij}} = \frac{\langle L^{\leftrightarrow} \rangle^*}{\langle L \rangle^*}; \tag{3.15}$$

in order to do this, let us calculate the ρ index [23], defined as

$$\rho \equiv \frac{r(\mathbf{A}^*) - \langle r \rangle^*}{1 - \langle r \rangle^*} \tag{3.16}$$

which automatically discounts for the effects of the imposed constraints. By definition, ρ ranges between 1 and -1: in fact, the denominator is always positive and, in magnitude, smaller than the numerator. It simply normalizes the index, not contributing to the sign of the quantity itself which, in turn, is decided only by the relative magnitude between the observed value r and its expectation. A positive sign indicates a stronger than expected tendency to reciprocate whereas a negative sign, a tendency weaker than expected to establish reciprocal links.

We can use the reciprocity to look again at the temporal evolution of the DIN. This is shown in Fig. 3.7. We confirm a behaviour similar to the one we observed when looking at the dyadic motifs. For the initial decade, the raw reciprocity shows an essentially constant trend, with small fluctuations around an average value of approximately 0.26, but the last four periods are characterized by an impressive decrease of the reciprocity value (approximately 40%): they lie almost 3 sigmas away from the sample average, clearly indicating that the DIN shows an anomalously low reciprocity value in those time periods already affected by the crisis.

The trends of ρ calculated under the DRG and the DCM are also shown in Fig. 3.7. The positive sign of the trend of ρ under the DRG (i.e. ρ_{DRG}) indicates that the

tendency of the network to reciprocate is stronger than expected under this model. This is intuitive by considering that $\langle r \rangle_{DRG}$ coincides with the network's density of links, which is also an average of all the single pair-specific probabilities. In fact,

$$\langle r \rangle_{DRG} = \frac{\sum_i \sum_{j \neq i} p^2}{\sum_i \sum_{j \neq i} p} = p = \frac{\sum_i \sum_{j \neq i} p_{ij}}{N(N-1)} = \frac{\langle L \rangle_{DRG}}{N(N-1)} \equiv \langle c \rangle_{DRG} \tag{3.17}$$

where c is the link density. Using the maximum-likelihood condition, the expected density is $\langle c \rangle^*_{DRG} = \frac{\langle L \rangle^*_{DRG}}{N(N-1)} = \frac{L(\mathbf{A}^*)}{N(N-1)} = c(\mathbf{A}^*)$ for any matrix $\mathbf{A}^*$ of the time period considered. Given the low value of the density c for this particular network, the DRG does basically a simple, small translation (and rescaling) of r towards lower values: $\rho_{DRG} = \frac{r-c}{1-c} \simeq r - c$.

Far more interesting is the trend of ρ_{DCM}. As a general comment, the network is more consistent with the DCM rather than with the DRG, as the smaller values of the respective ρ indices show. In more detail, the DCM highlights two opposite patterns. During the first twenty-eight periods, the reciprocity is higher than expected (similar to the DRG, but with the difference that ρ_{DCM} presents an almost constant trend, by showing smaller fluctuations than ρ_{DRG}): this implies that even the specification of the entire in- and out-degree sequences is not enough to fully account for the observed reciprocity, as the positive value of ρ_{DCM} witnesses. During the following sixteen quarters, the network inverts the tendency and tends to be less reciprocated than expected, showing an almost perfect monotonic decrease. This clear anti-reciprocal behavior, not detected by the DRG but revealed by the DCM (i.e. not encoded in the total number of links, but partially encoded in the degree sequences) is an early signature of the upcoming crisis, as the nodes start avoiding mutual exchanges two years before the 2008.

3.3 Detecting Triadic Motifs

In principle, third-order properties can be studied by introducing generalizations of the clustering coefficient [19, 24]. However, there is a proliferation of possible third-order patterns due to the directionality of links. For this reason, a more complete analysis consists in counting (across the entire network) all the possible directed *motifs* [25] involving two and three vertices and comparing the empirical abundances with the expected ones under the null model (see Fig. 3.8 for an illustration and Table 3.3 for the mathematical definition).

Before presenting our results, we note however that directionality makes the possible specifications of the null model proliferate as well. In particular, besides the DCM considered above, a more refined way to randomize directed networks includes the possibility to enforce additional constraints on the reciprocity structure [26, 27] (*reciprocal configuration model*, RCM). In other words, it is possible (and important in many applications [26, 28]) to preserve not only the total numbers k_i^{in} and k_i^{out}

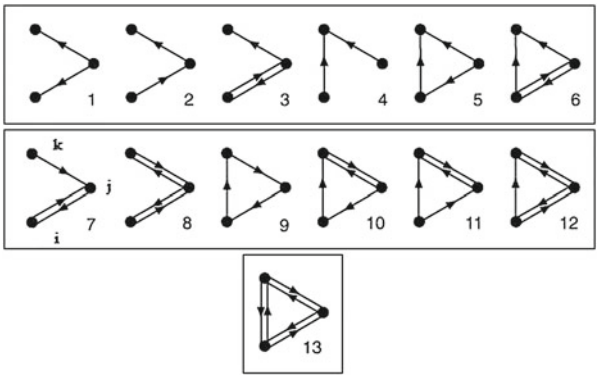

Fig. 3.8 The triadic, binary, directed motifs

of incoming and outgoing links of each vertex, but also the number of reciprocated links (pairs of links in both directions) [23, 29].

Given a real binary directed network $\mathbf{A}^*$, this specification is equivalent to enforcing, for each vertex i, the three directed-degree sequences [23, 27] $\mathbf{k}^{\rightarrow}(\mathbf{A}^*)$, $\mathbf{k}^{\leftarrow}(\mathbf{A}^*)$ and $\mathbf{k}^{\leftrightarrow}(\mathbf{A}^*)$ where

$$k_i^{\rightarrow}(\mathbf{A}^*) \equiv \sum_{j \neq i} a_{ij}^*(1 - a_{ji}^*),\ k_i^{\leftarrow}(\mathbf{A}^*) \equiv \sum_{j \neq i} a_{ji}^*(1 - a_{ij}^*),\ k_i^{\leftrightarrow}(\mathbf{A}^*) \equiv \sum_{j \neq i} a_{ij}^* a_{ji}^* \tag{3.18}$$

Table 3.3 Classification and definitions of the triadic motifs (sums are taken over terns of vertices)

Motif m	N_m: 1^{st} definition	N_m: 2^{nd} definition
1	$\sum_{i\neq j\neq k}(1-a_{ij})a_{ji}a_{jk}(1-a_{kj})(1-a_{ik})(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftarrow} a_{jk}^{\rightarrow} a_{ik}^{\leftrightarrow}$
2	$\sum_{i\neq j\neq k} a_{ij}(1-a_{ji})a_{jk}(1-a_{kj})(1-a_{ik})(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\rightarrow} a_{jk}^{\rightarrow} a_{ik}^{\leftrightarrow}$
3	$\sum_{i\neq j\neq k} a_{ij}a_{ji}a_{jk}(1-a_{kj})(1-a_{ik})(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftrightarrow} a_{jk}^{\rightarrow} a_{ik}^{\leftrightarrow}$
4	$\sum_{i\neq j\neq k}(1-a_{ij})(1-a_{ji})a_{jk}(1-a_{kj})a_{ik}(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftrightarrow} a_{jk}^{\rightarrow} a_{ik}^{\rightarrow}$
5	$\sum_{i\neq j\neq k}(1-a_{ij})a_{ji}a_{jk}(1-a_{kj})a_{ik}(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftarrow} a_{jk}^{\rightarrow} a_{ik}^{\rightarrow}$
6	$\sum_{i\neq j\neq k} a_{ij}a_{ji}a_{jk}(1-a_{kj})a_{ik}(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftrightarrow} a_{jk}^{\rightarrow} a_{ik}^{\rightarrow}$
7	$\sum_{i\neq j\neq k} a_{ij}a_{ji}(1-a_{jk})a_{kj}(1-a_{ik})(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftrightarrow} a_{jk}^{\leftarrow} a_{ik}^{\leftrightarrow}$
8	$\sum_{i\neq j\neq k} a_{ij}a_{ji}a_{jk}a_{kj}(1-a_{ik})(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftrightarrow} a_{jk}^{\leftrightarrow} a_{ik}^{\leftrightarrow}$
9	$\sum_{i\neq j\neq k}(1-a_{ij})a_{ji}(1-a_{jk})a_{kj}a_{ik}(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftarrow} a_{jk}^{\leftarrow} a_{ik}^{\rightarrow}$
10	$\sum_{i\neq j\neq k}(1-a_{ij})a_{ji}a_{jk}a_{kj}a_{ik}(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftarrow} a_{jk}^{\leftrightarrow} a_{ik}^{\rightarrow}$
11	$\sum_{i\neq j\neq k} a_{ij}(1-a_{ji})a_{jk}a_{kj}a_{ik}(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\rightarrow} a_{jk}^{\leftrightarrow} a_{ik}^{\rightarrow}$
12	$\sum_{i\neq j\neq k} a_{ij}a_{ji}a_{jk}a_{kj}a_{ik}(1-a_{ki})$	$\sum_{i\neq j\neq k} a_{ij}^{\leftrightarrow} a_{jk}^{\leftrightarrow} a_{ik}^{\rightarrow}$
13	$\sum_{i\neq j\neq k} a_{ij}a_{ji}a_{jk}a_{kj}a_{ik}a_{ki}$	$\sum_{i\neq j\neq k} a_{ij}^{\leftrightarrow} a_{jk}^{\leftrightarrow} a_{ik}^{\leftrightarrow}$

are, respectively, the (vertex-specific) number of non-reciprocated outgoing links, the number of non-reciprocated incoming links and the number of reciprocated links. Equivalently,

$$k_i^{\rightarrow}(\mathbf{A}^*) \equiv \sum_{j\neq i} a_{ij}^{\rightarrow *},\ k_i^{\leftarrow}(\mathbf{A}^*) \equiv \sum_{j\neq i} a_{ij}^{\leftarrow *},\ k_i^{\leftrightarrow}(\mathbf{A}^*) \equiv \sum_{j\neq i} a_{ij}^{\leftrightarrow *} \tag{3.19}$$

where $a_{ij}^{\rightarrow}$, $a_{ij}^{\leftarrow}$, $a_{ij}^{\leftrightarrow}$ and $a_{ij}^{\nleftrightarrow}$ are the dyadic quantities defined in the preceding section (Table 3.1).

The randomization procedure starts by writing the hamiltonian, as

$$H(\mathbf{A}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_i [\alpha_i k_i^{\rightarrow}(\mathbf{A}) + \beta_i k_i^{\leftarrow}(\mathbf{A}) + \gamma_i k_i^{\leftrightarrow}(\mathbf{A})]. \tag{3.20}$$

Explicitly rewriting the hamiltonian in terms of the mutually excluding variables $a_{ij}^{\rightarrow}$, $a_{ij}^{\leftarrow}$ and $a_{ij}^{\leftrightarrow} \equiv a_{ij}a_{ji}$ allows to analytically calculate the partition function as in [27]: $Z(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i<j}(1 + e^{-\alpha_i-\beta_j} + e^{-\alpha_j-\beta_i} + e^{-\gamma_i-\gamma_j})$. As a consequence, the graph probability factorizes as a pair-specific product of four factors

$$P(\mathbf{A}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_i \prod_{j<i} (p_{ij}^{\rightarrow})^{a_{ij}^{\rightarrow}} (p_{ij}^{\leftarrow})^{a_{ij}^{\leftarrow}} (p_{ij}^{\leftrightarrow})^{a_{ij}^{\leftrightarrow}} (p_{ij}^{\nleftrightarrow})^{a_{ij}^{\nleftrightarrow}} \tag{3.21}$$

and, by setting $x_i \equiv e^{-\alpha_i}$, $y_i \equiv e^{-\beta_i}$ and $z_i \equiv e^{-\gamma_i}$ [27], the coefficients's expressions are found to be

$$p_{ij}^{\rightarrow} \equiv \frac{x_i y_j}{1 + x_i y_j + x_j y_i + z_i z_j},\ p_{ij}^{\leftarrow} \equiv \frac{x_j y_i}{1 + x_i y_j + x_j y_i + z_i z_j}, \tag{3.22}$$

$$p_{ij}^{\leftrightarrow} \equiv \frac{z_i z_j}{1 + x_i y_j + x_j y_i + z_i z_j},\ p_{ij}^{\nleftrightarrow} \equiv \frac{1}{1 + x_i y_j + x_j y_i + z_i z_j}. \tag{3.23}$$

By maximizing the likelihood $\mathscr{L}(\mathbf{x}, \mathbf{y}, \mathbf{z})$, the values $\mathbf{x}^*$, $\mathbf{y}^*$, $\mathbf{z}^*$ corresponding to the point of maximum can be found by solving the following system

$$\langle k_i^{\rightarrow} \rangle = \sum_{j\neq i} \frac{x_i^* y_j^*}{1 + x_i^* y_j^* + x_j^* y_i^* + z_i^* z_j^*} = k_i^{\rightarrow}(\mathbf{A}^*) \quad \forall i, \tag{3.24}$$

$$\langle k_i^{\leftarrow} \rangle = \sum_{j\neq i} \frac{x_j^* y_i^*}{1 + x_i^* y_j^* + x_j^* y_i^* + z_i^* z_j^*} = k_i^{\leftarrow}(\mathbf{A}^*) \quad \forall i, \tag{3.25}$$

$$\langle k_i^{\leftrightarrow} \rangle = \sum_{j\neq i} \frac{z_i^* z_j^*}{1 + x_i^* y_j^* + x_j^* y_i^* + z_i^* z_j^*} = k_i^{\leftrightarrow}(\mathbf{A}^*) \quad \forall i. \tag{3.26}$$

The expectation value of any topological property, as well as its standard deviation, can now be calculated analytically in terms of the vectors $\mathbf{x}^*$, $\mathbf{y}^*$, $\mathbf{z}^*$.

As for the dyadic motifs, the number of occurrences N_m of a particular triadic motif m ($m = 1 \dots 13$) can be written either as a product of adjacency matrix elements, a_{ij} or employing the quantities $a_{ij}^{\rightarrow}$, $a_{ij}^{\leftarrow}$, $a_{ij}^{\leftrightarrow}$ and $a_{ij}^{\nleftrightarrow}$ (see Table 3.4). Notice that the RCM can be employed to detect only the abundance of triadic motifs, since the dyadic motifs would be exactly reproduced by definition (in fact, $\langle k_i^{\rightarrow} \rangle = k_i^{\rightarrow}$ implies $\langle L^{\rightarrow} \rangle = \langle \sum_i k_i^{\rightarrow} \rangle = \sum_i \langle k_i^{\rightarrow} \rangle = \sum_i k_i^{\rightarrow} = L^{\rightarrow}$ and the same holds true for the other two dyadic motifs.

Since the values of z_m are sensitive to the number of nodes, when it is necessary to compare the z-scores of networks with different size, or of differently sized snapshots of the same network, a size-independent measure is needed. For this reason, it is customary to normalize the z-scores by introducing the *significance profile* [26] defined as

$$SP_m \equiv \frac{z_m}{\sqrt{\sum_{m=1}^{13} z_m^2}} \tag{3.27}$$

and measuring the *relative* importance of each motif with respect to the other ones. While the z-scores are unbounded quantities, SP_m lies between -1 and $+1$.

In Fig. 3.9 we show the z-scores for all the possible 13 non-isomorphic connected motifs with three vertices in 8 real food webs, for both null models. We also show the two lines $z = \pm 2$ to highlight the region within 2 standard deviations from the model's expectations. This analysis is similar to that of Ref. [28], but is made much simpler by the maximum-entropy method which does not require to randomize the webs through a computational algorithm preserving the (reciprocal) degree sequences. The food webs considered here are from different ecosystems (lagoons, marshes,

Table 3.4 Expectation values of the triadic motifs (sums are taken over triples of vertices)

Motif m	$\langle N_m \rangle_{DCM}$	$\langle N_m \rangle_{RCM}$
1	$\sum_{i\neq j\neq k}(1-p_{ij})p_{ji}\,p_{jk}(1-p_{kj})(1-p_{ik})(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\leftarrow}\,p_{jk}^{\rightarrow}\,p_{ik}^{\nleftrightarrow}$
2	$\sum_{i\neq j\neq k} p_{ij}(1-p_{ji})p_{jk}(1-p_{kj})(1-p_{ik})(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\rightarrow}\,p_{jk}^{\rightarrow}\,p_{ik}^{\nleftrightarrow}$
3	$\sum_{i\neq j\neq k} p_{ij}\,p_{ji}\,p_{jk}(1-p_{kj})(1-p_{ik})(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\leftrightarrow}\,p_{jk}^{\rightarrow}\,p_{ik}^{\nleftrightarrow}$
4	$\sum_{i\neq j\neq k}(1-p_{ij})(1-p_{ji})p_{jk}(1-p_{kj})p_{ik}(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\nleftrightarrow}\,p_{jk}^{\rightarrow}\,p_{ik}^{\rightarrow}$
5	$\sum_{i\neq j\neq k}(1-p_{ij})p_{ji}\,p_{jk}(1-p_{kj})p_{ik}(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\leftarrow}\,p_{jk}^{\rightarrow}\,p_{ik}^{\rightarrow}$
6	$\sum_{i\neq j\neq k} p_{ij}\,p_{ji}\,p_{jk}(1-p_{kj})p_{ik}(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\leftrightarrow}\,p_{jk}^{\rightarrow}\,p_{ik}^{\rightarrow}$
7	$\sum_{i\neq j\neq k} p_{ij}\,p_{ji}(1-p_{jk})p_{kj}(1-p_{ik})(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\leftrightarrow}\,p_{jk}^{\leftarrow}\,p_{ik}^{\nleftrightarrow}$
8	$\sum_{i\neq j\neq k} p_{ij}\,p_{ji}\,p_{jk}\,p_{kj}(1-p_{ik})(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\leftrightarrow}\,p_{jk}^{\leftrightarrow}\,p_{ik}^{\nleftrightarrow}$
9	$\sum_{i\neq j\neq k}(1-p_{ij})p_{ji}(1-p_{jk})p_{kj}\,p_{ik}(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\leftarrow}\,p_{jk}^{\leftarrow}\,p_{ik}^{\rightarrow}$
10	$\sum_{i\neq j\neq k}(1-p_{ij})p_{ji}\,p_{jk}\,p_{kj}\,p_{ik}(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\leftarrow}\,p_{jk}^{\leftrightarrow}\,p_{ik}^{\rightarrow}$
11	$\sum_{i\neq j\neq k} p_{ij}(1-p_{ji})p_{jk}\,p_{kj}\,p_{ik}(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\rightarrow}\,p_{jk}^{\leftrightarrow}\,p_{ik}^{\rightarrow}$
12	$\sum_{i\neq j\neq k} p_{ij}\,p_{ji}\,p_{jk}\,p_{kj}\,p_{ik}(1-p_{ki})$	$\sum_{i\neq j\neq k} p_{ij}^{\leftrightarrow}\,p_{jk}^{\leftrightarrow}\,p_{ik}^{\rightarrow}$
13	$\sum_{i\neq j\neq k} p_{ij}\,p_{ji}\,p_{jk}\,p_{kj}\,p_{ik}\,p_{ki}$	$\sum_{i\neq j\neq k} p_{ij}^{\leftrightarrow}\,p_{jk}^{\leftrightarrow}\,p_{ik}^{\leftrightarrow}$

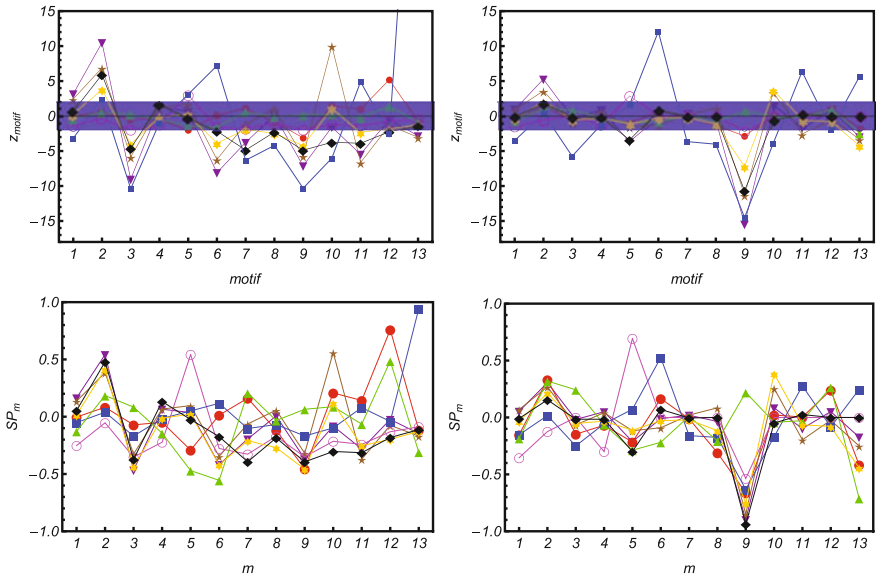

Fig. 3.9 Legend: ● - Chesapeake Bay, ■ - Little Rock Lake, ▲ - Maspalomas Lagoon, ▼ - Florida Bay, ✳ - St Marks Seagrass, ⋆ - Everglades Marshes, ○ - Grassland, ♦ - Ythan Estuary. Application of the maximum-entropy method to the analysis of directed motifs and significance profiles in 8 real food webs. Left column: z-scores and significance profiles obtained enforcing only the in-degree and out-degree sequences (DCM). Right column: z-scores and significance profiles obtained enforcing also the reciprocal degree sequence (RCM). The shaded area correspond to the region $z_m \in [-2, +2]$, enclosing a probability of 95%

lakes, bays, estuaries, grasses), with a prevalence of aquatic habitats. The presence of (intrinsically directed) predator-prey relationships implies that reciprocity is a very important quantity in food webs [28]. Thus the RCM should fluctuate less than the DCM. Indeed, this is confirmed by our analysis. The z-scores for the motifs $m = 2, 3, 13$ are significantly reduced from the DCM to the RCM. Also, while the motifs $m = 1, 6, 10, 11$ display large values of z with opposite signs across different webs under the DCM, the signs of all statistically surprising motifs (i.e. when $|z| \gtrsim 2$) become consistent with each other under the RCM (except for $m = 13$).

As a consequence, under the RCM all networks display a very similar pattern and the most striking features of real webs become the over-representation of motifs $m = 2, 10$ (plus $m = 6, 11, 13$ for the Little Rock Lake web) and the under-representation of motifs $m = 5, 9, 13$ (plus $m = 3, 7, 8$ for Little Rock Lake). In particular, the under-representation of motif $m = 9$ (the 3-loop) is the most common pattern across all webs and becomes stronger as the reciprocity of the web increases. Also note that in a network with no reciprocated links, the number of motifs with at least a pair of reciprocated links is zero. Under the RCM, the expected number of these motifs remains zero. By contrast, their expected number under the DCM is always positive. Thus we confirm that the upgrade to the RCM is necessary, as its stricter constraints

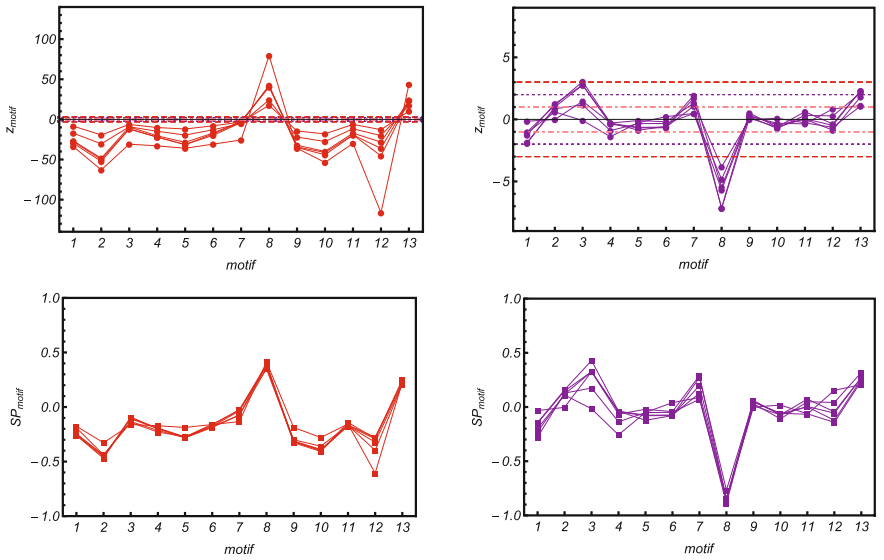

Fig. 3.10 z-scores (first and second panel) and significance profiles (third and fourth panel) of the 13 triadic, binary, directed motifs for the WTW in the years 1950, 1960, 1970, 1980, 1990 and 2000, under the DCM (■, first and third panel) and the RCM (●, second and fourth panel). The dashed, red lines represent the values $z = \pm 3$, the dotted, purple lines the values $z = \pm 2$ and the dot-dashed, pink lines the values $z = \pm 1$

allow us to analyze 3-vertices motifs once 2-vertices motifs (i.e. all possible dyadic patterns) are correctly accounted for. The possibility to treat the RCM analytically using the maximum-entropy method is therefore an important step forward.

We now perform a similar analysis on the WTW, where the different instances of the network represent different temporal snapshots of its long-term evolution. The initial number (85) of countries roughly doubles during the time period considered (1950–2000), mainly because of many colonies becoming independent and the Soviet Uniot disgregating into many states. This expansion of the network and the simultaneous globalization process have caused a significant increase in the number of links [30], as well as considerable variations in the nodes' degrees. This circumstance makes the WTW an ideal example for testing whether an economic network undergoes a quasi-equilibrium evolution driven by the dynamics of the local properties.

The results of the anaysis of the z-scores are shown in Fig. 3.10. Under the DCM, the z-scores indicate large deviations between observations and expectations and the agreement worsens as the network evolves. These results confirm that, while some higher-order properties of the WTW were previously found to be well-reproduced by constraining the nodes' degrees [9], the triadic patterns are irreducible to the in- and out-degrees themselves [25].

By contrast, the agreement improves substantially under the RCM: now, all the z-scores (with the only exception of motif 8) lie within the error bars $z_m = \pm 3$. This indicates that, once the number of reciprocated and non-reciprocated links of each node are separately controlled for, the triadic structure of the network is almost completely explained. Moreover, the shape of the profiles is more stable than under the DCM. All these findings indicate that the RCM should be preferred to the DCM, the reciprocity structure playing a strong role in shaping the topology of the WTW [23, 25].

The panels of Fig. 3.10 also show the significance profiles for all 13 motifs, as defined in Eq. (3.27). We find that discounting the effect of the increasing size of the network makes the curves of the 6 different snapshots collapse to a single profile. This effect is obviously more evident under the DCM, since under the RCM the z-scores of the different snapshots were already largely overlapping.

So, even if in absolute terms many structural quantities change (the number of nodes, the number of links, the degrees, etc.), under both null models the significance profiles are extremely stable, clearly pointing out that the deviating patterns are systematic and the relative importance of each motif remains constant.

We now show the results of the reciprocity analysis of the WTW is almost completely consistent with a quasi-equilibrium network driven by the local (non-)reciprocated degrees $k_i^{\leftrightarrow}$, $k_i^{\leftarrow}$ and $k_i^{\rightarrow}$. In other words, even if the network undergoes major changes under the effect of complicated economic and political processes (such as the creation of new independent states, globalization and the establishment of reciprocated relationships), once these processes are reabsorbed into the evolution of the local constraints, the quasi-equilibrium character of the network becomes manifest.

3.4 Some Extensions to Weighted Networks

In the rest of this chapter we move to the case of weighted networks and the corresponding generalization of some of the quantities considered so far for binary graphs. It should be noted that generalizing binary topological properties to weighted graphs is to some extent arbitrary, as no unique choice exist [31–34]. Therefore the definitions we consider here are necessarily only some examples of many possible generalizations.

3.4.1 Weighted Assortativity and Clustering

We start with the analysis of the completely aggregated WTW network. We start with the weighted counterpart of the average nearest neighbor degree (ANND), i.e. the *average nearest neighbor strength* (ANNS) of vertex i, defined as

$$s_i^{nn}(\mathbf{W}^*) \equiv \frac{\sum_{j\neq i} a_{ij}^* s_j^*}{k_i^*} = \frac{\sum_{j\neq i}\sum_{k\neq j} a_{ij}^* w_{jk}^*}{\sum_{j\neq i} a_{ij}^*} \tag{3.28}$$

The ANNS measures the average strength of the neighbors of a given vertex. Similarly to the ANND, the ANNS involves indirect interactions of length two, however (as happens for most weighted quantities) mixing both weighted and purely topological information: in particular, terms of the type $a_{ij}w_{jk}$ appear in the definition. The correlations between the strength of neighboring countries can be inspected by plotting s_i^{nn} versus s_i. This is shown in Fig. 3.11: even if the points are now significantly more scattered, we find a decreasing trend as previously observed for the corresponding binary quantities (see Fig. 3.2). This trend signals that highly trading countries trade typically with poorly trading ones (and vice versa), confirming on a weighted basis the disassortative character observed at the binary level. However, in this case the null model behaves in a completely different way: over the randomized ensemble with specified strength sequence (i.e. the Weighed Configuration Model introduced in Sect. 2.2.3), the expectation value

$$\langle s_i^{nn}\rangle^* = \frac{\sum_{j\neq i} p_{ij}^* \langle s_j\rangle^*}{\langle k_i\rangle^*} = \frac{\sum_{j\neq i} p_{ij}^* s_j^*}{\langle k_i\rangle^*} = \frac{\sum_{j\neq i}\sum_{k\neq j} p_{ij}^* \langle w_{jk}\rangle^*}{\sum_{j\neq i} p_{ij}^*} \tag{3.29}$$

of the ANNS decreases over a much narrower range (see Fig. 3.11) and is always different from the observed value. This important results implies that, even if we observe disassortativity in both cases (binary and weighted), we find that in the binary case this property is completely explained by the degree sequence, whereas in the weighted case it is not explained by the strength sequence. This has implications for economic models of international trade: while no theoretical explanation is required in order to explain why poorly connected countries trade with highly connected ones on a binary basis (once the number of trade partners is specified), additional

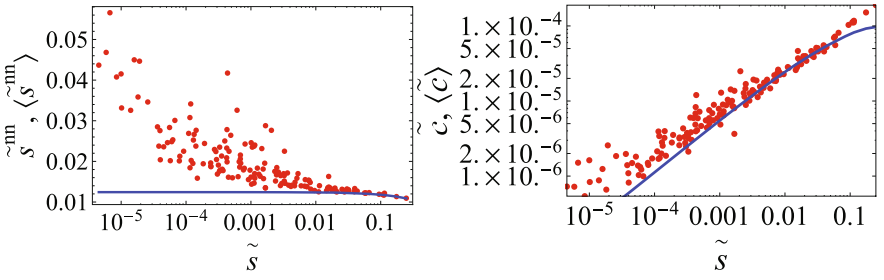

Fig. 3.11 Left panel: average nearest neighbor strength versus strength in the 2002 snapshot of the real weighted undirected WTW (red points) and corresponding average over the maximum-entropy ensemble with specified strengths (blue curve). Right panel: weighted clustering coefficient versus strength. All the quantities have been divided by the total weight of all links in the network and are hence denoted by a tilde

explanations are required in order to explain the same phenomenon at a weighted level, even after controlling for the total trade volumes of all countries.

We now consider the weighted version of the clustering coefficient. In particular, we choose the definition proposed in Ref. [18], which has a more direct extension to the directed case [19]. According to that definition, the *weighted clustering coefficient* c_i represents the intensity of the triangles in which vertex i participates:

$$c_i^w(\mathbf{W}^*) \equiv \frac{\sum_{j\neq i}\sum_{k\neq i,j}(w_{ij}^* w_{jk}^* w_{ki}^*)^{1/3}}{k_i^*(k_i^*-1)} = \frac{\sum_{j\neq i}\sum_{k\neq i,j}(w_{ij}^* w_{jk}^* w_{ki}^*)^{1/3}}{\sum_{j\neq i}\sum_{k\neq i,j} a_{ij}^* a_{ik}^*} \tag{3.30}$$

Note that c_i^w takes into account indirect interactions of length three, corresponding to products of the type $w_{ij}w_{jk}w_{ki}$ appearing in the above formula. In Fig. 3.11 we plot c_i versus s_i. This time we find an increasing trend of c_i^w as a function of s_i, indicating that countries with larger total trade participate in more intense trade triangles. We also show the trend followed by the randomized quantity

$$\langle c_i^w\rangle^* = \frac{\sum_{j\neq i}\sum_{k\neq i,j}\langle(w_{ij}w_{jk}w_{ki})^{1/3}\rangle^*}{\langle k_i(k_i-1)\rangle^*} = \frac{\sum_{j\neq i}\sum_{k\neq i,j}\langle w_{ij}^{1/3}\rangle^*\langle w_{jk}^{1/3}\rangle^*\langle w_{ki}^{1/3}\rangle^*}{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}^* p_{ik}^*} \tag{3.31}$$

which is found to approximately reproduce the empirical data. Despite the partial accordance between the clustering profile of real and randomized networks, the total level of clustering of the real WTW is however larger than its randomized counterpart [9, 10].

The above results confirm that, unlike the binary case, the properties of the weighted undirected version of the WTW are not completely reproduced by simply controlling for the local properties. The presence of higher-order mechanisms is required as an explanation for the onset and evolution of the observed patterns. This result holds across different years and is enhanced as lower levels of commodity aggregation are considered. This shows that a weighted network approach to the analysis of international trade conveys additional information with respect to traditional economic studies that describe trade in terms of local properties alone (total trade, openness, etc.) [36]. Interestingly, a major deviation between the real network and the null model is in the topology implied by local constraints. This confirms, from a different point of view, that in order to properly understand the structure of the international trade system is essential to reproduce its binary topology, even if one is interested in a weighted description. We will use this result in the next chapter, when aiming at defining reliable network reconstruction methods.

3.4.2 Weighted Reciprocity

We now follow [35] and discuss an interesting weighted generalization of the binary reciprocity introduced previously in Sect. 3.2.1. Let us consider a directed, weighted

network specified by the weight matrix $\mathbf{W}$, where the entry w_{ij} indicates the weight of the directed link from vertex i to vertex j, including the case $w_{ij} = 0$ indicating the absence of such link. As Fig. 3.12 shows, we can always decompose each pair (w_{ij}, w_{ji}) of reciprocal links into a bidirectional (fully reciprocated) interaction, plus a unidirectional (non reciprocated) interaction. Formally, we can define the *reciprocated* weight between i and j (the symmetric part) as

$$w_{ij}^{\leftrightarrow} \equiv \min[w_{ij}, w_{ji}] = w_{ji}^{\leftrightarrow} \tag{3.32}$$

and the *non-reciprocated weight* from i to j (the asymmetric part) as

$$w_{ij}^{\rightarrow} \equiv w_{ij} - w_{ij}^{\leftrightarrow} \tag{3.33}$$

Note that if $w_{ij}^{\rightarrow} > 0$ then $w_{ji}^{\rightarrow} = 0$, which makes the unidirectionality manifest. We can also define

$$w_{ij}^{\leftarrow} \equiv w_{ji} - w_{ij}^{\leftrightarrow} = w_{ji}^{\rightarrow} \tag{3.34}$$

as the *non-reciprocated weight* from j to i and restate the unidirectionality property in terms of the fact that $w_{ij}^{\rightarrow}$ and $w_{ij}^{\leftarrow}$ cannot be both nonzero. Thus any dyad (w_{ij}, w_{ji}) can be equivalently decomposed as $(w_{ij}^{\leftrightarrow}, w_{ij}^{\rightarrow}, w_{ij}^{\leftarrow})$. If the network is binary, all the above variables are either 0 or 1 and our decomposition coincides with the one already described in Sect. 3.2.

From the above fundamental dyadic quantities it is possible to define reciprocity measures at the more aggregate level of vertices. We recall that the out- and in-strength of a vertex i are defined as the sum of the weights of the out-going and in-coming links respectively:

$$s_i^{out}(\mathbf{W}^*) \equiv \sum_{j \neq i} w_{ij}^*, \; s_i^{in}(\mathbf{W}^*) \equiv \sum_{j \neq i} w_{ji}^*. \tag{3.35}$$

Using Eqs. (3.32)–(3.34), we can split the above quantities into their reciprocated and non-reciprocated contributions. We first define the *reciprocated strength*

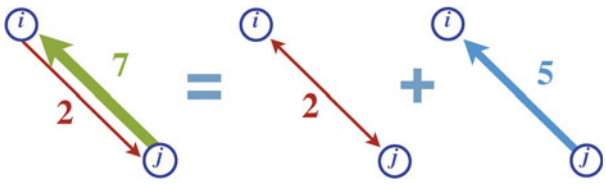

Fig. 3.12 Basic decomposition of any two dyadic fluxes (in the example shown, $w_{ij} = 2$ and $w_{ji} = 7$) into a fully reciprocated component ($w_{ij}^{\leftrightarrow} = 2$) and a fully non-reciprocated component ($w_{ij}^{\leftarrow} = 5$, which implies $w_{ij}^{\rightarrow} = 0$) [35]

$$s_i^{\leftrightarrow}(\mathbf{W}^*) \equiv \sum_{j \neq i} w_{ij}^{\leftrightarrow *} \tag{3.36}$$

which measures the overlap between the in-strength and the out-strength of vertex i, i.e. the portion of strength of that vertex which is fully reciprocated by its neighbours. Then we define the *non-reciprocated out-strength* as

$$s_i^{\rightarrow}(\mathbf{W}^*) \equiv \sum_{j \neq i} w_{ij}^{\rightarrow *} = s_i^{out} - s_i^{\leftrightarrow *} \tag{3.37}$$

and the *non-reciprocated in-strength* as

$$s_i^{\leftarrow}(\mathbf{W}^*) \equiv \sum_{j \neq i} w_{ij}^{\leftarrow *} = s_i^{in} - s_i^{\leftrightarrow *}. \tag{3.38}$$

The last two quantities represent the non-reciprocated components of s_i^{out} and s_i^{in} respectively, i.e. the out-going and in-coming fluxes which exceed the inverse fluxes contributed by the neighbours of vertex i.

Finally, we introduce weighted measures of reciprocity at the global, network-wide level. Recall that the total weight of the network is

$$W_{tot}^* \equiv \sum_i \sum_{j \neq i} w_{ij}^* = \sum_i s_i^{out}(\mathbf{W}^*) = \sum_i s_i^{in}(\mathbf{W}^*); \tag{3.39}$$

similarly, we denote the *total reciprocated weight* as

$$W^{\leftrightarrow *} \equiv \sum_i \sum_{j \neq i} w_{ij}^{\leftrightarrow *} = \sum_i s_i^{\leftrightarrow}(\mathbf{W}^*). \tag{3.40}$$

Extending the definition used for binary graphs (see Sect. 3.2.1), we can then define the *weighted reciprocity* of a weighted network as

$$r^w(\mathbf{W}^*) \equiv \frac{W^{\leftrightarrow} *}{W_{tot}^*} \tag{3.41}$$

If all fluxes are perfectly reciprocated (i.e. $W^{\leftrightarrow *} = W_{tot}^*$) then $r^w(\mathbf{W}^*) = 1$, whereas in absence of reciprocation (i.e. $W^{\leftrightarrow *} = 0$) then $r^w(\mathbf{W}^*) = 0$ [40–43].

Just like its binary counterpart, Eq. (3.41) is informative only after a comparison with a null model, NM, is made, i.e. with a value $\langle r^w \rangle_{NM}$ expected for a network having some property in common (e.g. the number of vertices N and/or the total weight W_{tot}) with the observed one. As a consequence, networks with different empirical values of such quantities cannot be consistently ranked in terms of the measured value of r^w. As in the binary case [23] the problem has been solved by introducing a transformed quantity that we generalize to the present setting as

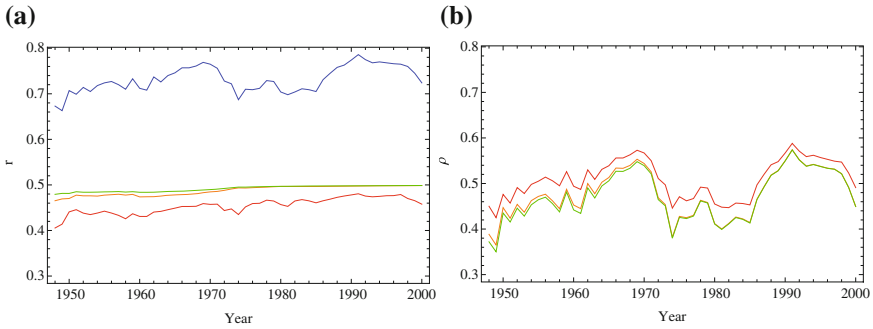

Fig. 3.13 Temporal evolution of the reciprocity of the World Trade Web during the 53 years from 1948 to 2000: **a** observed value of r^w (blue) and its expected values $\langle r^w \rangle$ under the Weighted Configuration Model (red), the Balanced Configuration Model (green) and the Weighted Random Graph (orange); **b** evolution of ρ^w under the same three null models as above

$$\rho^w \equiv \frac{r^w(\mathbf{W}^*) - \langle r^w \rangle^*_{NM}}{1 - \langle r^w \rangle^*_{NM}}. \tag{3.42}$$

The sign of ρ^w is directly informative of an increased, with respect to the null model, tendency to reciprocate ($\rho^w > 0$) or to avoid reciprocation ($\rho^w < 0$). If ρ^w is consistent with zero, then the observed level of reciprocity is compatible with what merely expected by chance under the null model.

In the weighted, directed case, each graph $\mathbf{G}$ is completely specified by its non-negative (integer-valued) weight matrix $\mathbf{W}$, which now is in general not symmetric. The constraints specified in the randomization method [2] are now the joint in-strength and out-strength sequence (*directed weighted configuration model*, DWCM): $\{C_a\} = \{s_i^{in}(\mathbf{W})\}_{i=1}^N$ and $\{s_i^{out}(\mathbf{W})\}_{i=1}^N$. The exact values of the connection probabilities in the ensemble of randomized (directed, weighted) graphs are found by solving the following set of $2N$ coupled non-linear equations [44]:

$$\langle s_i^{out} \rangle = \sum_{j \neq i} \frac{x_i y_j}{1 - x_i y_j}, = s_i^{out}(\mathbf{W}^*) \quad \forall i \tag{3.43}$$

$$\langle s_i^{in} \rangle = \sum_{j \neq i} \frac{x_j y_i}{1 - x_j y_i} = s_i^{in}(\mathbf{W}^*) \quad \forall i \tag{3.44}$$

where the in-strengths and out-strengths of all vertices are special local quantities whose expected and empirical values are exactly equal: $\langle s_i^{in} \rangle = s_i^{in}(\mathbf{W}^*)$ and $\langle s_i^{out} \rangle = s_i^{out}(\mathbf{W}^*)$.

Since ρ^w consistently ranks the reciprocity of networks with different properties, it can also track the evolution of reciprocity in a network that changes over time. For this reason, in our dataset we have included 53 yearly snapshots of the World Trade Web, from year 1948 to 2000 [45, 46]. In Fig. 3.13 we show the evolution of r^w, $\langle r^w \rangle$ and ρ^w under three null models: the DWCM, the *balanced configuration model*

(BCM, i.e. the same as the DWCM but further imposing $s_i^{in} = s_i^{out}$) and the *directed weighted random graph model* (WRG). The plots confirm that, unlike ρ^w, r^w is not an adequate indicator of the evolution of reciprocity, since the baseline expected value $\langle r^w \rangle$ (under every null model) also changes in time as a sort of moving target as shown in Fig. 3.13a.

Note that $\langle r^w \rangle_{WCM}$ fluctuates much more than $\langle r^w \rangle_{WRG}$ and $\langle r^w \rangle_{BCM}$ and its fluctuations resemble those of the observed value r^w (see Fig. 3.13a). This is due to the fact that, while all snapshots of the network are characterized by 'static' fluctuations of the empirical strengths of vertices around the balanced flux condition $s_i^{in} = s_i^{out}$ $\forall\, i$, these fluctuations have different entities in different years. Changes in the size of 'static' fluctuations produce the 'temporal' fluctuations observed in the evolution of $\langle r^w \rangle_{WCM}$ and partly also in the observed value r^w, confirming the important role of node (im)balances. After controlling for the time-varying entity of node imbalances (using the WCM), we indeed find that the fluctuations of ρ^w_{WCM} are less pronounced than those of ρ^w_{BCM} and ρ^w_{WRG} (see Fig. 3.13b). However, the fluctuations of r and $\langle r^w \rangle_{WCM}$ do not cancel out completely and their resulting net effect (the trend of ρ^w_{WCM}) is still significant, indicating the strongest level of reciprocity across the three null models.

While a binary analysis of the WTW [45, 47] detected an almost monotonic increase of the reciprocity, with a marked acceleration in the 90's, we find that the weighted reciprocity has instead undergone a rapid decrease over the same decade: this counter-intuitive result confirms that the information conveyed by a weighted analysis of reciprocity is nontrivial and irreducible to the binary picture.

The above results confirm that, unlike the binary case, the knowledge of local weighted properties conveys only limited information about the actual structure of the network: higher-order properties, in fact, are not explained by local constraints. Moreover, simple purely topological properties such as link density are not reproduced by the null model. This result confirms that, even in weighted analyses, the binary structure plays an important role, being responsible for major departures of the empirical network from the null model.

3.4.3 A Model for Weighted Networks with Reciprocity

We conclude this chapter by defining a maximum-entropy model that generalizes the RCM introduced in Sect. 3.3 to the weighted case. This model can generate weighted networks with given in-, out-, and reciprocated strengths sequences, and can thus accurately replicate the weighted reciprocity of any real network. The model goes under the name of Weighted Reciprocal Configuration Model (WRCM) and has been introduced in [35]. This implies a slight generalization of the formulas already shown. The graph hamiltonian becomes

$$H(\mathbf{W}|\boldsymbol{\theta}) = \sum_i [\alpha_i s_i^{\rightarrow}(\mathbf{W}) + \beta_i s_i^{\leftarrow}(\mathbf{W}) + \gamma_i s_i^{\leftrightarrow}(\mathbf{W})] \tag{3.45}$$

where, now, $\boldsymbol{\theta} \equiv \{\boldsymbol{\alpha},\ \boldsymbol{\beta},\ \boldsymbol{\gamma}\}$ and

$$s_i^{\rightarrow}(\mathbf{W}^*) \equiv \sum_{j\neq i} w_{ij}^{\rightarrow *},\ s_i^{\leftarrow}(\mathbf{W}^*) \equiv \sum_{j\neq i} w_{ij}^{\leftarrow *},\ s_i^{\leftrightarrow}(\mathbf{W}^*) \equiv \sum_{j\neq i} w_{ij}^{\leftrightarrow *} \tag{3.46}$$

with obvious meaning of the symbols (defined above). The partition function now becomes

$$Z(\boldsymbol{\theta}) = \prod_i \prod_{j<i} \frac{(1 - x_i x_j y_i y_j)}{(1 - x_i y_j)(1 - x_j y_i)(1 - z_i z_j)} \equiv \prod_i \prod_{j<i} Z_{ij}^{WRCM}(\boldsymbol{\theta}) \tag{3.47}$$

and the likelihood is

$$\begin{aligned} \ln P(\mathbf{W}^*|\boldsymbol{\theta}) = \sum_i \sum_{j<i} [&(w_{ij}^{\rightarrow})^* \ln(x_i y_j) + (w_{ij}^{\leftarrow})^* \ln(x_j y_i) + (w_{ij}^{\leftrightarrow})^* \ln(z_i z_j) + \\ &- \ln Z_{ij}^{WRCM}(\boldsymbol{\theta})]. \end{aligned} \tag{3.48}$$

The solution to this optimization problem, with respect to $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, can be found by solving the following system:

$$\langle s_i^{\rightarrow} \rangle^* = \sum_{j\neq i} \langle w_{ij}^{\rightarrow} \rangle^* = s_i^{\rightarrow}(\mathbf{W}^*), \quad \forall i, \tag{3.49}$$

$$\langle s_i^{\leftarrow} \rangle^* = \sum_{j\neq i} \langle w_{ij}^{\leftarrow} \rangle^* = s_i^{\leftarrow}(\mathbf{W}^*), \quad \forall i, \tag{3.50}$$

$$\langle s_i^{\leftrightarrow} \rangle^* = \sum_{j\neq i} \langle w_{ij}^{\leftrightarrow} \rangle^* = s_i^{\leftrightarrow}(\mathbf{W}^*), \quad \forall i \tag{3.51}$$

where

$$\langle w_{ij}^{\rightarrow} \rangle^* = \frac{x_i^* y_j^* (1 - x_j^* y_i^*)}{(1 - x_i^* y_j^*)(1 - x_i^* x_j^* y_i^* y_j^*)}, \tag{3.52}$$

$$\langle w_{ij}^{\leftarrow} \rangle^* = \frac{x_j^* y_i^* (1 - x_i^* y_j^*)}{(1 - x_j^* y_i^*)(1 - x_i^* x_j^* y_i^* y_j^*)}, \tag{3.53}$$

$$\langle w_{ij}^{\leftrightarrow} \rangle^* = \frac{z_i^* z_j^*}{1 - z_i^* z_j^*}. \tag{3.54}$$

With the definition of the WRCM model all the vertex-level, strength sequences are exactly reproduced, impyling that the reciprocity is reproduced at a *local* level. The WRCM is powerful enough to allow for the analysis of the *weighted motifs* and for community detection on weighted networks, especially for those networks where the reciprocity plays an important role in shaping their structure.

References

1. M.E.J. Newman, The structure and function of complex networks. SIAM Rev. **45**, 167 (2003)
2. T. Squartini, D. Garlaschelli, Analytical maximum-likelihood method to detect patterns in real networks. New J. Phys. **13**, 083001 (2011)
3. V. Colizza, R. Pastor-Satorras, A. Vespignani, Nat. Phys. **3**, 276–282 (2007)
4. K. Oshio, Y. Iwasaki, S. Morita, Y. Osana, S. Gomi, E. Akiyama, K. Omata, K. Oka, K. Kawamura, Tech. Rep. of CCeP (Keio University) **3** (2003)
5. http://dip.doe-mbi.ucla.edu/dip/Main.cgi
6. G. De Masi, G. Iori, G. Caldarelli, A fitness model for the Italian interbank money market. Phys. Rev. E **74**, 066112 (2006)
7. S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks. Science **296**, 910–913 (2002)
8. S. Maslov, K. Sneppen, A. Zaliznyak, Detection of topological patterns in complex networks: correlation profile of the Internet. Physica A **333**, 529–540 (2004)
9. T. Squartini, G. Fagiolo, D. Garlaschelli, Randomizing world trade. I. A binary network analysis. Phys. Rev. E **84**(4), 046117 (2011)
10. T. Squartini, G. Fagiolo, D. Garlaschelli, Randomizing world trade. II. A weighted network analysis. Phys. Rev. E **84**(4), 046118 (2011)
11. D. Garlaschelli, M.I. Loffredo, Structure and evolution of the world trade network. Physica A **355**, 138–144 (2005)
12. G. Fagiolo, S. Schiavo, J. Reyes, World-trade web: topological properties, dynamics, and evolution. Phys. Rev. E **79**, 036115 (2009)
13. M.A. Serrano, M. Boguná, Topology of the world trade web. Phys. Rev. E **68**, 015101(R) (2003)
14. J. Park, M.E.J. Newman, Origin of degree correlations in the Internet and other networks. Phys. Rev. E **68**(2), 026112 (2003)
15. R. Pastor-Satorras, A. Vázquez, A. Vespignani, Dynamical and correlation properties of the Internet. Phys. Rev. Lett. **87**(25), 258701 (2001)
16. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, The large scale organization of metabolic networks. Nature **407**, 651 (2000)
17. N.D. Martinez, Effects of resolution on the Little Rock Lake food web. Ecol. Monogr. **61**, 367–392 (1991)
18. J. Saramaki, M. Kivela, J.P. Onnela, K. Kaski, J. Kertesz, Generalizations of the clustering coefficient to weighted complex networks. Phys. Rev. E **75**(2), 027105 (2007)
19. G. Fagiolo, Clustering in complex directed networks. Phys. Rev. E **76**, 026107 (2007)
20. T. Squartini, I. van Lelyveld, D. Garlaschelli, Early-warning signals of topological collapse in interbank networks. Sci. Rep. **3**, 3357 (2013)
21. S. Wasserman, K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994)
22. M.E.J. Newman, S. Forrest, J. Balthrop, Email networks and the spread of computer viruses. Phys. Rev. E **66**, 035101(R) (2002)
23. D. Garlaschelli, M.I. Loffredo, Patterns of link reciprocity in directed networks. Phys. Rev. Lett. **93**(26), 268701 (2004)
24. S.E. Ahnert, T.M.A. Fink, Clustering signatures classify directed networks. Phys. Rev. E **78**, 036112 (2008)
25. T. Squartini, D. Garlaschelli, Triadic motifs and dyadic self-organization in the World Trade Network. Lec. Notes Comp. Sci. **7166**, 24–35 (2012)
26. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks. Science **298**, 824–827 (2002)
27. D. Garlaschelli, M.I. Loffredo, Multispecies grand-canonical models for networks with reciprocity. Phys. Rev. E **73**, 015101 (2006)
28. D.B. Stouffer, J. Camacho, W. Jiang, L.A.N. Amaral, Evidence for the existence of a robust pattern of prey selection in food webs. Proc. R. Soc. B **274**, 1931–1940 (2007)

29. V. Zlatic, H. Stefancic, Influence of reciprocal links in social networks. Phys. Rev. E **80**, 016117 (2009)
30. M. Barigozzi, G. Fagiolo, D. Garlaschelli, Multinetwork of international trade: a commodity-specific analysis. Phys. Rev. E **81**, 046104 (2010)
31. M.E.J. Newman, Analysis of weighted networks. Phys. Rev. E **70**, 056131 (2004)
32. S.E. Ahnert, D. Garlaschelli, T.M. Fink, G. Caldarelli, Ensemble approach to the analysis of weighted networks. Phys. Rev. E **73**, 015101(R) (2006)
33. J. Saramaki, M. Kivela, J.-P. Onnela, K. Kaski, J. Kertész, Generalizations of the clustering coefficient to weighted complex networks. Phys. Rev. E **75**, 027105 (2007)
34. A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. Proc. Nat. Acad. Sci. **101**(11), 3747–3752 (2004)
35. T. Squartini, F. Picciolo, F. Ruzzenenti, D. Garlaschelli, Reciprocity of weighted networks. Sci. Rep. **3**, 2729 (2013)
36. R.C. Feenstra, *Advanced International Trade: Theory and Evidence* (Princeton University Press, Princeton, 2004)
37. L.A. Meyers, M.E.J. Newman, B. Pourbohloul, Predicting epidemics on directed contact networks. J. Theor. Biol. **240**, 400–418 (2006)
38. M. Boguná, M.A. Serrano, Generalized percolation in random directed networks. Phys. Rev. E **72**, 016106 (2005)
39. P. Holland, S. Leinhardt, in *Sociological Methodology*, ed. by D. Heise pp. 1–45 (1975)
40. L. Kovanen, J. Saramaki, K. Kaski, Reciprocity of mobile phone calls. Dyn. Socio-Econ. Syst. **2**(2), 138–151 (2011)
41. G. Fagiolo, Directed or undirected? A new index to check for directionality of relations in socio-economic networks. Econ. Bull. **3**(34), 1–12 (2006)
42. C. Wang, O. Lizardo, D. Hachen, A. Strathman, Z. Toroczkai, N. Chawla, A dyadic reciprocity index for repeated interaction networks. Network Science **1**(1), 31–48 (2013)
43. L. Akoglu, P.O.S. Vaz de Melo, C. Faloutsos, Quantifying reciprocity in large weighted communication networks. Lec. Notes Comp. Sci. **7302**, 85–96 (2012)
44. D. Garlaschelli, M.I. Loffredo, Maximum likelihood: extracting unbiased information from complex networks. Phys. Rev. E **78**, 015101(R) (2008)
45. D. Garlaschelli, M.I. Loffredo, Structure and evolution of the world trade network. Physica A **355**(1), 138–144 (2005)
46. K.S. Gleditsch, Expanded trade and GDP data. J. Conflict Resol. **46**(5), 712–724 (2002)
47. F. Ruzzenenti, D. Garlaschelli, R. Basosi, Complex networks and symmetry II: reciprocity and evolution of world trade. Symmetry **2**(3), 1710–1744 (2010)

Chapter 4
Network Reconstruction

When they propose to establish the universal from the particulars by means of induction, they will effect this by a review of either all or some of the particulars. But if they review some, the induction will be insecure, since some of the particulars omitted in the induction may contravene the universal; while if they are to review all, they will be toiling at the impossible, since the particulars are infinite and indefinite.

—Sextus Empiricus, $\Pi\upsilon\rho\rho\omega\nu\varepsilon\iota\omega\iota\ \upsilon\pi\omega\tau\upsilon\pi\omega\sigma\varepsilon\iota\varsigma$

Abstract In this chapter we show that the maximum-entropy method has important and beneficial implications for the problem of network reconstruction. In general, network reconstruction covers many aspects, e.g. the inference of missing links, missing nodes, etc. Our formalism naturally defines an improved method of inference of unknown link probabilities and link weights from partial node-specific information. An important application is the reconstruction of privacy-protected financial networks from public bank-specific data and the subsequent estimation of the level of systemic risk in a financial system. The method allows us to infer the unobserved network structure with great accuracy even if only the strengths and/or the degrees of nodes are known. We present a detailed discussion of this problem and a validation of the method on many real-world examples. We start by illustrating the limitations of purely weighted null models which, by enforcing only the strength sequence, perform very poorly in reproducing the binary topology of real networks. These limitations can be overcome by adding the degree sequence as a new piece of information which is usually not included in traditional models. Indeed, when constraining both the degree and the strength sequence, the agreement between the observed and the expected higher-order properties of real-world networks improves dramatically. We then discuss how the observational requirements can be further relaxed while keeping the reconstruction method reliable, as imposed by the empirical inaccessibility of the degree sequence in many cases.

T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-69438-2_4

4.1 Reconstructing Network Properties from Partial Information

In this section we mainly follow Ref. [1] in describing our approach to the network reconstruction problem. As a preliminary observation and motivation for addressing this problem, it should be noted that how nodes are connected to each other determines not only the structure of a network, but also the dynamics of processes taking place on it. Indeed, a range of phenomena of critical importance, from the spread of infectious diseases to the diffusion of opinions and the propagation of financial crises, turn out to be highly sensitive to the topology of the underlying network that mediates the interactions [2]. This sensitivity implies that, whenever it is not possible to have a complete empirical knowledge of the network, one should make an optimal use of the partial information available and try to reconstruct the most likely network, or rather an ensemble of likely networks, in the least biased way. In the Big Data era, this kind of problem is becoming more and more important given the ever-increasing availability of data that, for privacy issues, are often of aggregate nature [3, 4].

Among the possible types of incomplete topological information (e.g. missing links, missing nodes, etc.), one of the most frequently encountered situations is when only a *local* knowledge of the network is available [5–10]. For instance, in binary networks knowing the *number* of links of each node is typically much easier than knowing the *identity* of all neighbours. Similarly, in weighted networks knowing the *total intensity* of all the links connected to each node is much easier than knowing the identity of all neighbours and the intensity of all links separately.

A typical example is that of interbank networks, where it is relatively easy to know the total exposures of each bank, while privacy issues make it much more difficult to know *who* is lending to whom, and *how much* [5, 6, 8, 9]. Similarly, a huge amount of information is continuously collected about individuals [3, 4]. In that case as well, privacy issues are becoming increasingly important, and methods that are able to give detailed predictions from aggregated data, while at the same time respecting the privacy of individuals, are therefore becoming more and more desirable.

In our approach, network reconstruction can be formally regarded as a constrained entropy maximization problem, where the constraints represent the available information and the maximization of the entropy ensures that the reconstructed ensemble of networks is maximally random, given the enforced constraints [11, 12]. When the available information is just local, one only knows $O(N)$ quantities (e.g. the degrees of all nodes) instead of the total $O(N^2)$ ones (e.g. all entries of the adjacency matrix) fully describing the network. This makes the network reconstruction problem very challenging, since the number of missing variables is still $O(N^2)$, i.e. of the same order of the total number (Fig. 4.1).

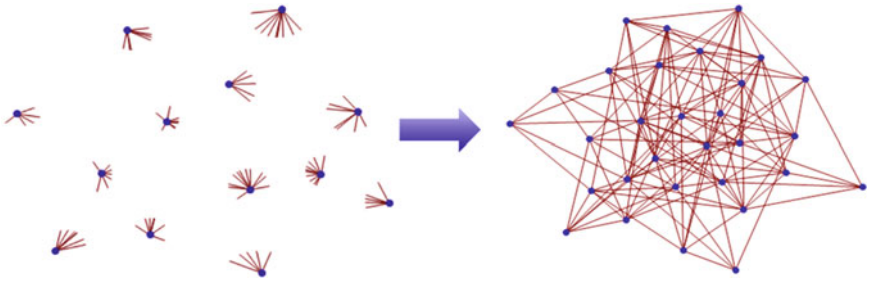

Fig. 4.1 Pictorial representation of the process of *network reconstruction from local topological properties*. Given a set of known node-specific properties (in the example considered, the degrees of all nodes), the goal is that of inferring the complete structure of the network. In general, since there are many network configurations compatible with the known constraints, the outcome of the reconstruction process is actually a *probability distribution* over the space of networks compatible with the constraints. This leads to the construction of a microcanonical or canonical ensemble (see Fig. 2.3), if the constraints are treated as 'hard' or 'soft' respectively

4.1.1 Reconstruction of Binary Networks

If we consider binary networks, the simplest and most extensively used null model is the CM that we have introduced in Subsect. 2.2.2 [11, 12]. From some of the results we have illustrated in Sect. 3.1, it turns out that some binary networks, e.g. the Italian Interbank network, the World Trade Web and some biological networks can be reconstructed remarkably well from the knowledge of their degree sequence alone. As additional evidence of this, in Fig. 4.2 we consider two foodwebs (note that these networks are directed) and show the results of an analysis similar to the one of Fig. 3.3. The figure shows that the empirical structure of the two webs is largely consistent with the null model.

Additionally, we consider a set of undirected networks of different type, including the Italian Interbank network in year 1999 [13], three 'classic' social networks collected in [14], seven food webs from [15] and finally the aggregated WTW in year 2002 [16]. For these networks, we provide a compact description of the agreement between the empirical structure and the reconstructed one in Fig. 4.3. This confirms that the assortativity and clustering patterns are correctly reproduced for most nodes of all networks.

4.1.2 Naive Extension to Weighted Networks and Its Limitations

In the rest of this chapter we address the problem of the effective reconstruction, from local properties alone, of *weighted* networks. We first show that, in contrast with what is generally believed, the reconstruction of weighted networks does not merely

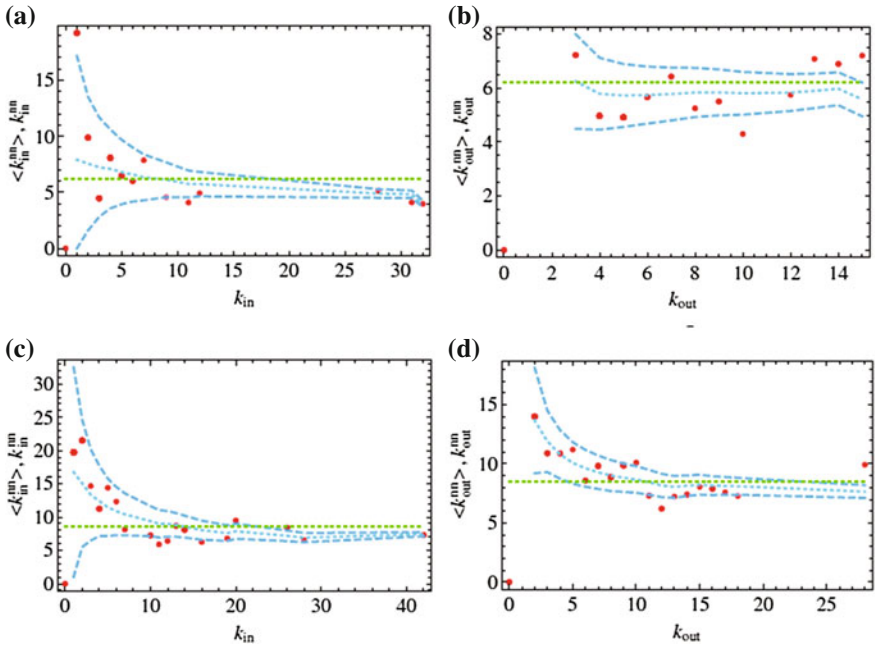

Fig. 4.2 Reconstruction of second-order properties of two directed food webs from the knowledge of the nodes degrees. Red points are the empirical data and the blue dashed curves are the exact expectations ($\pm$ one standard deviation) under the directed configuration model obtained using the maximum-entropy method. The green curves are the flat expectations under the directed version of the (Erdös-Rényi) RG model. The panels report $k_i^{nn,in}$ versus k_i^{in} (left) and $k_i^{nn,out}$ versus k_i^{out} (right) for: **a** and **b** the Narragansett Bay web ($N = 35$) [15], **c** and **d** the Mondego Estuary web ($N = 46$) [15]

involve a one-to-one mapping of the corresponding methodology that works well for binary networks. Specifically, we are going to show that inferring the structure of a weighted network only from the knowledge of its *strength sequence* can lead to a very bad reconstruction, even for the networks that, at a binary level, can be reproduced extremely well from their degree sequence [1, 12, 16, 17].

As shown in the previous chapters, the most natural generalization of the CM to weighted networks is a reconstructed ensemble with given *strength sequence* (i.e. the WCM) [12, 18, 19]. The latter is widely used both as a reconstruction method and as the most important null model to detect communities. In both cases, if s_i denotes the strength of node i and N is the number of nodes, the expected weight of the link between nodes i and j predicted by the WCM is routinely written as shown by Eq. (2.5) or in a slightly different way if the network is directed. Equation (2.5) represents one of the standard procedures to infer interbank linkages from the total exposures of individual banks [5], or the fundamental null model used by most algorithms aimed at detecting densely connected *communities* in weighted networks [20]. Unfortunately, despite its widespread use, such a recipe for weights estimation

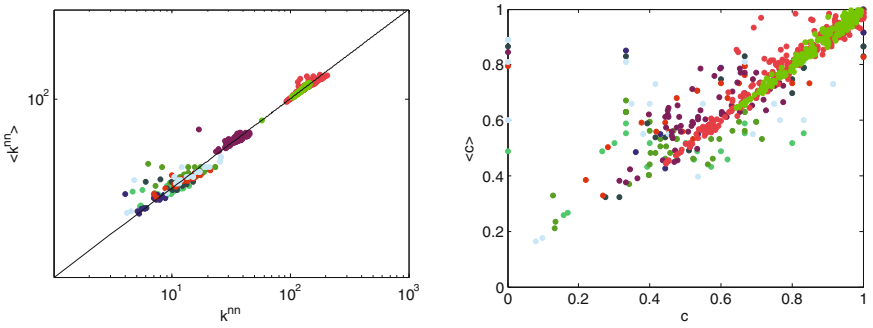

Fig. 4.3 Reconstruction of various undirected binary networks from node degrees using the CM, showing that purely binary local properties are very informative. In each panel we compare the reconstructed (y axis) and real (x axis) value of a node-specific network property, for all nodes of the following 12 networks: Office social network (●), Research group social network (●), Fraternity social network (●), Maspalomas Lagoon food web (●), Chesapeake Bay food web (●), Crystal River (control) food web (●), Crystal River food web (●), Michigan Lake food web (●), Mondego Estuary food web (●), Everglades Marshes food web (●), Italian Interbank network in year 1999 (●), aggregated World Trade Web in year 2002 (●). Left: average nearest neighbour degree (k_i^{nn}). Right: binary clustering coefficient (c_i)

is however incorrect and differs from the unbiased expression derived within our rigorous maximum-entropy approach [12, 21, 22].

But a more profound limitation of the WCM persists even when it is correctly implemented. It should be noted that the motivation for using the WCM as the natural generalization of the CM to weighted networks is the implicit assumption that the strength is an improved node-specific property, superior to the degree because it encapsulates the extra information provided by link weights. However, while the *complete* knowledge of a weighted network conveys of course more information than the complete knowledge of just its binary projection, the strength sequence is often surprisingly less informative than the degree sequence [12, 16, 17, 23]. In particular, several *purely topological* properties of real weighted networks turn out to be reproduced much better by applying the CM to the binary projection of the graph, than by applying the WCM to the original weighted network [12, 16, 17].

Figure 4.4 confirms and extends these non-obvious findings to various networks of different nature. We compare, for all networks in the sample, the empirical and reconstructed values of various structural properties, including both purely topological properties and their weighted counterparts. If the full weighted matrix is denoted by $\mathbf{W}$ (where w_{ij} is the weight of the link between node i and node j), the purely topological quantities are calculated on the binary projection $\mathbf{A}$ (adjacency matrix) of $\mathbf{W}$, with entries $a_{ij} = 1$ if $w_{ij} > 0$ and $a_{ij} = 0$ if $w_{ij} = 0$. The binary quantities we choose are again the average nearest neighbor degree (ANND) and the binary clustering coefficient introduced in Sect. 3.1.1. The corresponding weighted quantities are the average nearest neighbor strength (ANNS) and the weighted clustering coefficient introduced in Sect. 3.4.1. [16, 24]. In each panel of Fig. 4.4, we show

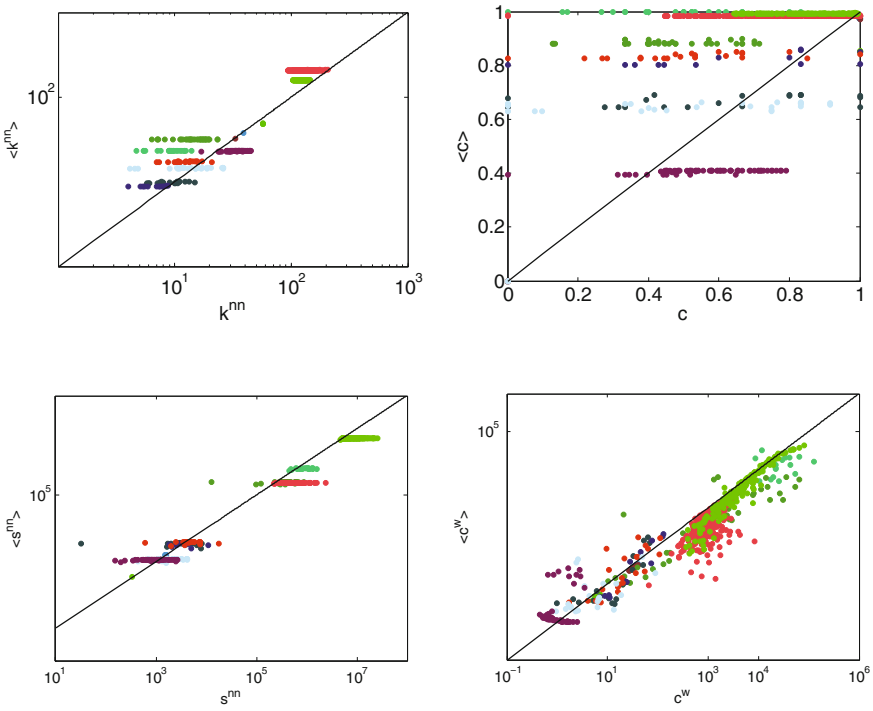

Fig. 4.4 Naïve network reconstruction from node strengths (WCM), showing that purely weighted local properties are poorly informative. In each panel we compare the reconstructed (y axis) and real (x axis) value of a node-specific network property, for all nodes of the following 12 networks: Office social network (●), Research group social network (●), Fraternity social network (●), Maspalomas Lagoon food web (●), Chesapeake Bay food web (●), Crystal River (control) food web (●), Crystal River food web (●), Michigan Lake food web (●), Mondego Estuary food web (●), Everglades Marshes food web (●), Italian Interbank network in year 1999 (●), aggregated World Trade Web in year 2002 (●). Top left: average nearest neighbour degree (k_i^{nn}). Top right: binary clustering coefficient (c_i). Bottom left: average nearest neighbour strength (s_i^{nn}). Bottom right: weighted clustering coefficient (c_i^w)

the measured value of one of the quantity defined above, for all nodes and for all networks, and we compare it with the corresponding reconstructed value predicted by the WCM. In this type of plot, every point is a node: therefore the target of a good reconstruction method is that of placing all the points along the identity. By contrast, in most cases we find that the reconstructed values for all nodes of a given network lie along horizontal lines, i.e. they are nearly equal to each other and totally unrelated to the 'target' real values.

The typical interpretation of a result like the above one is that the reconstruction of networks from local node-specific information is intrinsically problematic, presumably because of higher-order mechanisms involved in the formation of real networks, thus taking a difference between real data and the WCM as an important

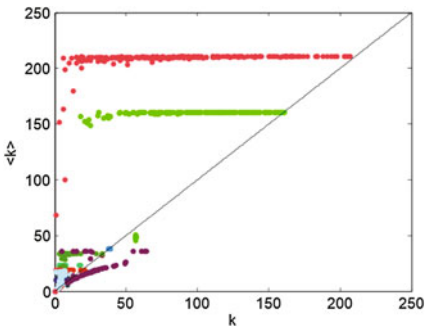

Fig. 4.5 Reconstruction of node degrees from node strengths (WCM), showing that purely weighted local properties are poorly informative. We compare the reconstructed (y axis) and real (x axis) value of the degree, for all nodes of the following 12 networks: Office social network (●), Research group social network (●), Fraternity social network (●), Maspalomas Lagoon food web (●), Chesapeake Bay food web (●), Crystal River (control) food web (●), Crystal River food web (●), Michigan Lake food web (●), Mondego Estuary food web (●), Everglades Marshes food web (●), Italian Interbank network in year 1999 (●), aggregated World Trade Web in year 2002 (●)

signature of non-local patterns [12, 18, 19]. In what follows, we propose a different interpretation of the above findings.

We conjecture that the reason for the bad agreement between observations and the WCM prediction lies in the fact that the strength sequence gives a very bad prediction of purely topological properties, and particularly the degrees, as Fig. 4.5 clearly shows: in fact, out of the many, possible ways to redistribute the strength of each node among the remaining vertices irrespectively of the number of new links created, the WCM selects those predicting much denser networks than the real ones [16].

4.2 The Enhanced Configuration Model

The above discussion leads us to the expectation that the poor reconstruction achieved by the WCM might be largely due to fact that the strength sequence discards purely topological information, in particular the degrees. Reversing the point of view, we keep following [1] and conjecture that the degrees are to be considered a 'fundamental' local structural property of weighted networks, irreducible to the knowledge of the strengths and thus at least as important as the latter. We should at this point clarify that by 'irreducible' we do not refer to the *numerical values* of strengths and degrees, but to the different *functional roles* that the two quantities play in determining or constraining the network structure: in fact, strengths and degrees are typically highly correlated in real networks [25], which means that we can reasonably infer the values of one quantity from those of the other (in this sense, strengths and degrees are 'reducible' to each other). However, this is, generally speaking, only true from

an *empirical* point of view. What is of interest to us is a deeper form of irreducibility, encountered when the joint specification of strengths and degrees *constrains the network in a fundamentally different way* than the specification of only one of the two. As an example, nothing guarantees that their observed correlation is preserved by the randomization procedure (i.e. that $s_i \propto f(k_i)$ implies $\langle s_i \rangle \propto f(\langle k_i \rangle)$), as proved by the bad performance of the WCM in reproducing the degree sequence of the networks considered in this chapter.

So, our conjecture leads us to believe that the reconstruction of weighted networks would be significantly enhanced by the specification of *both* strengths *and* degrees. For simplicity, we will refer to the ensemble of networks with given strengths and degrees as the *enhanced configuration model* (ECM) [1]. Early attempts to generate the ECM were either based on computational randomizations [26] or on theoretical arguments [19]. However, analytical calculations later showed that these approaches are statistically biased [22].

Formally, an ensemble of weighted networks with N nodes can be characterized by a collection $\{\mathbf{W}\}$ of $N \times N$ matrices and by an appropriate probability $P(\mathbf{W})$ [22]. We look for a probability that maximizes Shannon entropy with a constraint on the expected degree and strength sequences $\langle \mathbf{k} \rangle$, $\langle \mathbf{s} \rangle$ [22]. The fundamental result [22] of this constrained maximization is the probability

$$P(\mathbf{W}|\mathbf{x}, \mathbf{y}) = \prod_{i<j} q_{ij}(w_{ij}|\mathbf{x}, \mathbf{y}) \tag{4.1}$$

where $\mathbf{x}$ and $\mathbf{y}$ are two N-dimensional Lagrange multipliers controlling for the expected degrees and strengths respectively (with $x_i \geq 0$ and $0 \leq y_i < 1 \; \forall \, i$), and

$$q_{ij}(w|\mathbf{x}, \mathbf{y}) = \frac{(x_i x_j)^{\Theta(w)} (y_i y_j)^w (1 - y_i y_j)}{1 - y_i y_j + x_i x_j y_i y_j} \tag{4.2}$$

is the probability that a link of weight w exists between nodes i and j. In the above expression, $\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ otherwise. Equation (4.2) defines the 'mixed' Bose-Fermi distribution [22] where, due to the presence of $\Theta(w)$, the establishment of a link of unit weight between two nodes requires a different (higher if $x_i x_j > 0$) 'cost' than the reinforcement (by a unit of weight) of an already existing link. This feature is due to the mixed binary and weighted constraints and makes the ECM potentially very appropriate to model real networks.

We consider a particular real weighted network $\mathbf{W}^*$, whose only degrees $k_i^* \equiv k_i(\mathbf{W}^*)$ and strengths $s_i^* \equiv s_i(\mathbf{W}^*)$ are known. The log-likelihood of the ECM defined by Eqs. (4.1) and (4.2) reads

$$\begin{aligned} &\mathscr{L}(\mathbf{x}, \mathbf{y}) \equiv \ln P(\mathbf{W}^*|\mathbf{x}, \mathbf{y}) = \textstyle\sum_{i<j} \ln q_{ij}(w_{ij}^*|\mathbf{x}, \mathbf{y}) = \\ &\textstyle\sum_{i=1}^{N} \left(k_i^* \ln x_i + s_i^* \ln y_i \right) + \sum_{i<j} \ln \left(\frac{1 - y_i y_j}{1 - y_i y_j + x_i x_j y_i y_j} \right). \end{aligned} \tag{4.3}$$

We now look for the specific parameter values $\mathbf{x}^*, \mathbf{y}^*$ that maximize $\mathscr{L}(\mathbf{x}, \mathbf{y})$. A direct calculation shows that $\mathbf{x}^*, \mathbf{y}^*$ can be obtained as the real solution to the following system of $2N$ equations [12, 27]:

$$\langle k_i \rangle = \sum_{j \neq i} \frac{x_i x_j y_i y_j}{1 - y_i y_j + x_i x_j y_i y_j} = k_i(\mathbf{W}^*) \quad \forall i, \tag{4.4}$$

$$\langle s_i \rangle = \sum_{j \neq i} \frac{x_i x_j y_i y_j}{(1 - y_i y_j)(1 - y_i y_j + x_i x_j y_i y_j)} = s_i(\mathbf{W}^*) \quad \forall i. \tag{4.5}$$

Therefore, we find that the likelihood-maximizing values $\mathbf{x}^*, \mathbf{y}^*$ are precisely those ensuring that the expected degree and strength sequences coincide with the observed sequences $\mathbf{k}^*$ and $\mathbf{s}^*$, thus solving our initial problem. The values $\mathbf{x}^*, \mathbf{y}^*$ contain all the information necessary to reconstruct the network: consistently with our problem, either solving the $2N$ Eqs. (4.4) and (4.5) or, equivalently, maximizing the function $\mathscr{L}(\mathbf{x}, \mathbf{y})$ of $2N$ variables only requires the knowledge of the observed strengths and degrees, and not that of the entire network $\mathbf{W}^*$.

Once the solutions $\mathbf{x}^*$ and $\mathbf{y}^*$ are found, they can be used to obtain the reconstructed (ensemble-averaged) network properties analytically: for most topological properties of interest, this involves calculating the expected product of (powers of) distinct matrix entries, which simply reads

$$\left\langle \sum_{i \neq j \neq k, \ldots} w_{ij}^{\alpha} \cdot w_{jk}^{\beta} \cdot \ldots \right\rangle = \sum_{i \neq j \neq k, \ldots} \langle w_{ij}^{\alpha} \rangle \cdot \langle w_{jk}^{\beta} \rangle \cdot \langle \ldots \rangle \tag{4.6}$$

with the generic term given by

$$\langle w_{ij}^{\gamma} \rangle^* = \sum_{w=0}^{+\infty} w^{\gamma} q_{ij}(w|\mathbf{x}^*, \mathbf{y}^*) = \frac{x_i^* x_j^* (1 - y_i^* y_j^*) \mathrm{Li}_{-\gamma}(y_i^* y_j^*)}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*} \tag{4.7}$$

where $\mathrm{Li}_n(z)$ is the nth polylogarithm of z. The simplest and most useful cases $\gamma = 1$ and $\gamma = 0$ yield the expected weight $\langle w_{ij} \rangle$ and the connection probability $p_{ij} = \langle \Theta(w_{ij}) \rangle = \langle a_{ij} \rangle$, respectively. Therefore the reconstructed value $\langle X \rangle$ can be calculated in the same time as that required to calculate the real (if known) value $X(\mathbf{W}^*)$ (i.e. the shortest possible time), by simply replacing w_{ij}^{γ} with $\langle w_{ij}^{\gamma} \rangle$ in the definition of $X(\mathbf{W})$.

We can now apply our general methodology to the reconstruction of real-world networks. We consider again the assortativity and clustering properties. The result is illustrated in Fig. 4.6 for all the networks shown previously in Fig. 4.4. We clearly see that the enhanced method achieves a dramatic improvement over the standard approach. Now most points lie in the vicinity of the identity, meaning that the method is able to successfully reconstruct, for each vertex, the structure of the network two and three steps away from it. Note that the noisiest property is the binary clustering

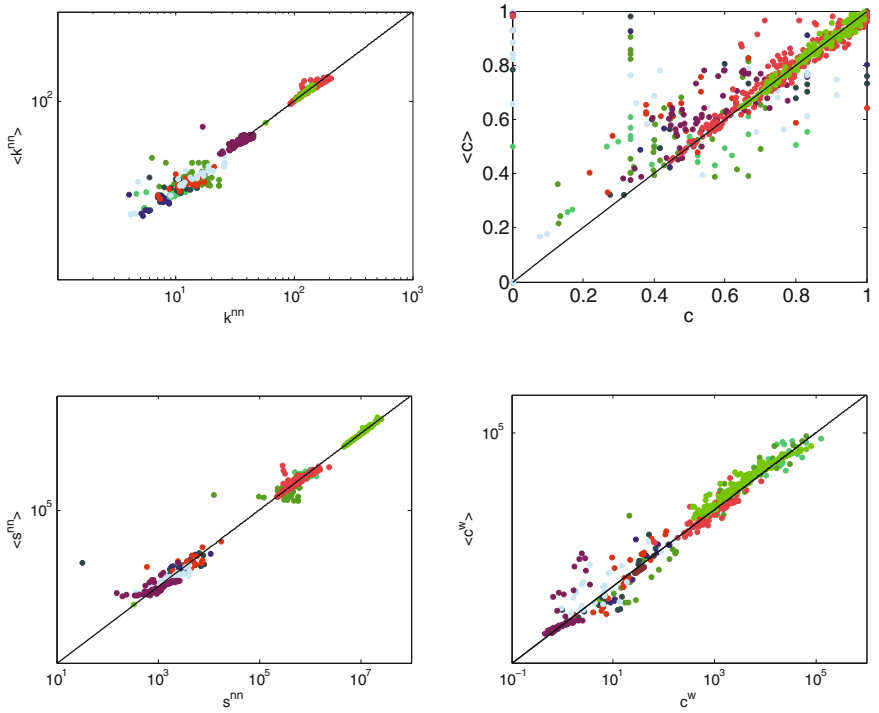

Fig. 4.6 Enhanced network reconstruction from strengths and degrees, showing dramatic improvements over the standard approach. In each panel we compare the reconstructed (y axis) and real (x axis) value of a node-specific network property, for all nodes of the following 12 networks: Office social network (●), Research group social network (●), Fraternity social network (●), Maspalomas Lagoon food web (●), Chesapeake Bay food web (●), Crystal River (control) food web (●), Crystal River food web (●), Michigan Lake food web (●), Mondego Estuary food web (●), Everglades Marshes food web (●), Italian Interbank network in year 1999 (●), aggregated World Trade Web in year 2002 (●). Top left: average nearest neighbour degree (k_i^{nn}). Top right: binary clustering coefficient (c_i). Bottom left: average nearest neighbour strength (s_i^{nn}). Bottom right: weighted clustering coefficient (c_i^w)

coefficient; however if we compare our results with the naive ones we find that the improvement achieved for this quantity is perhaps the most significant one.

The above findings completely reverse the conclusions one would draw from the previous interpretation of the naive results. First, network reconstruction from purely local properties is now shown to be possible to a highly satisfactory level, at least for the networks considered here. Second, the assortativity and clustering properties of these networks turn out to be well explained by purely local, even if augmented, properties. So, there is no need to invoke non-local mechanisms in order to explain such properties in these networks. We similarly expect that, if one considers the ECM as an improved null model to detect communities or other higher-order patterns,

the result will be dramatically different from what is obtained by using the WCM prediction in the definition of the modularity [20].

We now confirm the superiority of the method using the rigorous statistical tests illustrated in Sect. 2.3, comparing the performance of the WCM and ECM in reproducing the whole network. At the same time, this approach will automatically allow us to test our initial conjecture that the degrees are irreducible to the strengths. Indeed, both problems can be equivalently stated within a model selection framework, where one is interested in determining not only which of the two models achieves the best fit to the data, but also whether the introduction of the degrees as extra parameters in the ECM is really non-redundant.

To start with, we need to compare the likelihood of the ordinary WCM with that of ECM. Note that the WCM can be obtained as a particular case of the MCM by setting $\mathbf{x} = \mathbf{1}$. The log-likelihood of the WCM is therefore the reduced function $\mathscr{L}(\mathbf{1}, \mathbf{y})$ of N variables, and is maximized by a new vector $\mathbf{y}^{**} \neq \mathbf{y}^{*}$ which is also the solution of Eq. (4.5) with $\mathbf{x} = \mathbf{1}$. The predictions of the WCM are still obtained as in Eqs. (4.6) and (4.7), by replacing x_i^* with 1 and y_i^* with y_i^{**} in the latter. This is how the reconstructed properties in Fig. 4.4 were computed. Now, if we simply compare the maximized likelihoods of the two reconstruction methods, we trivially obtain $\mathscr{L}(\mathbf{x}^*, \mathbf{y}^*) \geq \mathscr{L}(\mathbf{1}, \mathbf{y}^{**})$ since the ECM always improves the fit to the real network $\mathbf{W}^*$, given that it includes the WCM as a particular case and has extra parameters. Thus, for our two competing null models, the most appropriate information-theoretic criteria are the LRT and AIC. The latter reads

$$\mathrm{AIC}^*_{ECM} \equiv 4N - 2\mathscr{L}(\mathbf{x}^*, \mathbf{y}^*), \tag{4.8}$$

$$\mathrm{AIC}^*_{WCM} \equiv 2N - 2\mathscr{L}(\mathbf{1}, \mathbf{y}^{**}) \tag{4.9}$$

and the optimal model is the one minimizing AIC. In our case, Akaike weights [28] read

$$w^{\mathrm{AIC}}_{ECM} \equiv \frac{e^{-\frac{\mathrm{AIC}_{ECM}}{2}}}{e^{-\frac{\mathrm{AIC}_{ECM}}{2}} + e^{-\frac{\mathrm{AIC}_{WCM}}{2}}}, \; w^{\mathrm{AIC}}_{WCM} \equiv 1 - w^{\mathrm{AIC}}_{ECM}. \tag{4.10}$$

since the addendum accounting for the minimum AIC in the terms Δ_{ECM} and Δ_{WCM} can be simplified (being present both at the numerator and at the denominator).

The AIC weights of the two reconstruction methods are shown in Table 4.1 for all networks [29]. We see that, apart from two social networks, the enhanced method is always superior to the naïve one, achieving unit probability (within machine precision) of being the best among the two models. Moreover, the LRT response is the same of AIC, at both 5 and 1% significance levels. A closer inspection of the two networks for which the opposite result holds reveals that they are (almost) fully connected. This explains why the specification of the degree sequence, which in this case is close to the almost fully connected prediction of the WCM, is redundant for these networks. In such cases, the relevant local constraints effectively reduce to the strength sequence, so the 'standard' WCM is preferable. Our method correctly identifies this situation. However, whenever the topology is nontrivial (as in most

Table 4.1 AIC weights for the considered null models (AICc and BIC weights give exactly the same results)

Network	w^{AIC}_{WCM}	w^{AIC}_{ECM}
Office social network [14]	1	0
Research group social network [14]	1	0
Fraternity social network [14]	0	1
Maspalomas Lagoon food web [15]	0	1
Chesapeake Bay food web [15]	0	1
Crystal River (control) food web [15]	0	1
Crystal River food web [15]	0	1
Michigan Lake food web [15]	0	1
Mondego Estuary food web [15]	0	1
Everglades Marshes food web [15]	0	1
Italian interbank network (1999) [13]	0	1
World Trade Web (2000)[16]	0	1

real-world networks), the local constraints are irreducible to the strength sequence alone and the degrees must be separately specified. We should therefore expect that, for the vast majority of real-world networks, the degree sequence is irreducible to the strength sequence. In such cases, the inclusion of degrees is non-redundant, explaining why the ECM retrieves significantly more information.

4.2.1 *In-Depth Example: Reconstructing the World Trade Web*

In sect. 3.4.1 we have shown that, at the binary level, the degree correlations and clustering structure of the WTW are excellently reproduced by the CM, i.e., using only the knowledge of the degree sequence. By contrast, when the WCM is implemented as a *natural* extension of the CM for valued graphs, the binary quantities and also the corresponding weighted quantities are very different from the predicted counterparts. These results are very robust and hold true over time and for various resolutions (i.e. for different levels of aggregation of traded commodities) [23].

In Fig. 4.7 we show the higher-order binary quantities (ANND and clustering) plotted versus the node degree and the higher-order weighted quantities (ANNS and clustering) plotted versus the node strength, for the 2002 snapshot of the World Trade Web. We plot together the observed values (red points), the corresponding quantities predicted by the WCM (green points) and by the ECM (blue points): while, on the one hand, there is a very close agreement between the observed values and the expected ones computed on the maximum-entropy ensemble generated by the ECM, on the other the ECM and the WCM predictions strongly differ; the expected values for the

binary and the weighted quantities under the WCM are similar to those for a fully connected topology (i.e. $\langle k_i \rangle_{WCM} \simeq N - 1$):

$$\langle k_i^{nn} \rangle_{WCM}^* \simeq N - 1, \ \langle c \rangle_{WCM}^* \simeq 1 \tag{4.11}$$

$$\langle s_i^{nn} \rangle_{WCM}^* \simeq \frac{\sum_{j \neq i} p_{ij}^* s_j^*}{\langle k_i \rangle^*} \simeq \frac{\sum_i s_i^*}{N - 1} \simeq \frac{2W_{tot}^*}{N - 1} \tag{4.12}$$

where N stands for the number of nodes in the network, while W_{tot}^* is total trade volume for the considered year (predictions (4.11) are represented, respectively, by the black dashed line in Fig. 4.7a, b while (4.12) corresponds to the black top line in Fig. 4.7c). This outcome, here emerging from the comparison between red and green

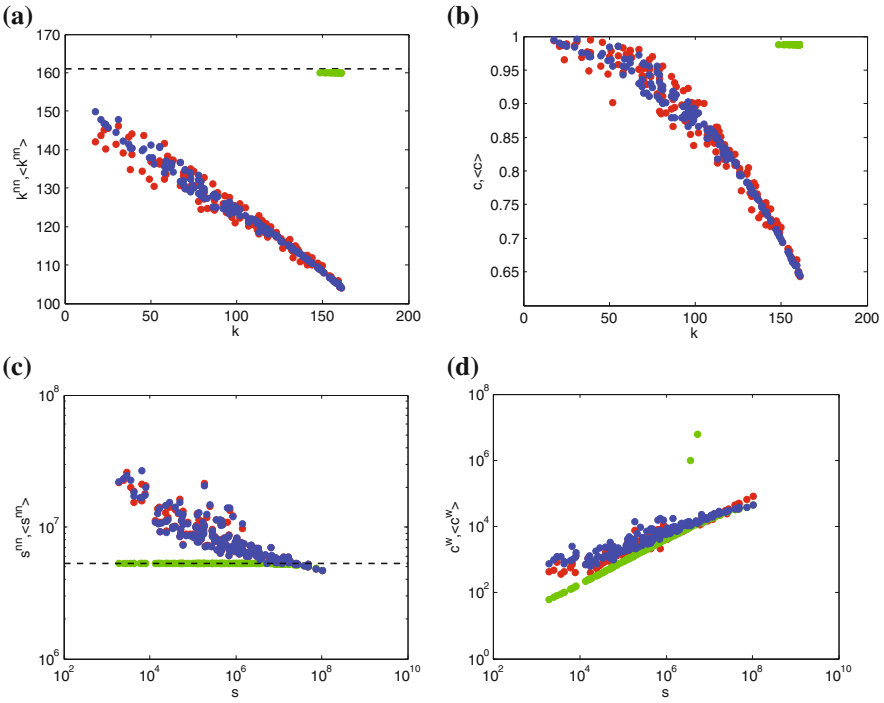

Fig. 4.7 Comparison between the observed undirected binary and weighted properties (red points) and the corresponding ensemble averages of the WCM (green points) and the ECM (blue points) for the aggregated WTW in the 2002 snapshot. Note that the red and the green points correspond to the values already shown in Fig. 3.11 apart from an overall rescaling constant. **a** average nearest neighbor degree k_i^{nn} versus degree k_i.; **b** binary clustering coefficient c_i versus degree k_i; **c** average nearest neighbor strength s_i^{nn} versus strength s_i; **d** weighted clustering coefficient c_i^w versus strength s_i

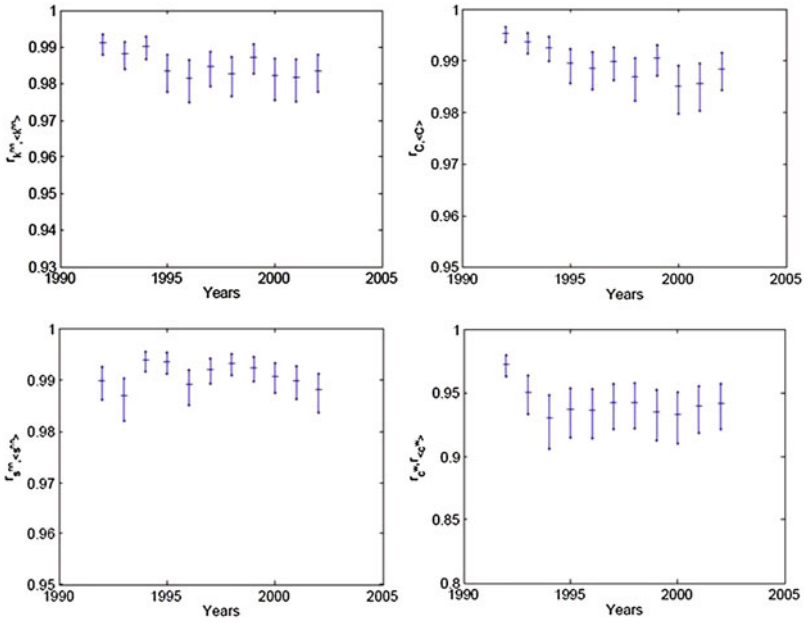

Fig. 4.8 Temporal evolution of the correlation coefficient between k_i^{nn} and $\langle k_i^{nn} \rangle$ (top left panel), c_i and $\langle c_i \rangle$ (top right panel), s_i^{nn} and $\langle s_i^{nn} \rangle$ (bottom left panel), c_i^w and $\langle c_i^w \rangle$ (bottom right panel) in the 1992–2002 snapshots of the observed undirected WTW and of the corresponding maximum-entropy ensembles with specified degrees and strengths. Red points stands for observed values, blue for the randomized ones; the 95% confidence intervals are represented as vertical bars

points, perfectly illustrates that the naïve expectation that weighted quantities are *per se* more informative than the corresponding binary ones is fundamentally incorrect.

On the contrary, the ECM performs excellently both for the binary and weighted versions of the WTW. Firstly, it reveals a slightly improved agreement for the binary trends. In fact, the monotonic trend predicted by the degree sequence only now follows, in a closer way, the observed cloud of points, while the prediction by WCM are concentrated far from the observed (red) points. Secondly, we also find a much better agreement between the observed and the randomized weighted trends. This implies that the weighted structure alone does not allow a deep understanding of the topology, representing an irreducible piece of information to be accounted for from the beginning.

In accordance with the existing literature we find a disassortative pattern for the WTW and a decreasing trend of c_i versus k_i. This confirms that it is very likely to find nodes with many trade partners connected with nodes with small degree (and vice-versa), while trade partners of poorly connected nodes are highly interconnected. Similar considerations hold true when we introduce weights, indeed s_i^{nn} and c_i^w are related with indirect paths of length two and three, respectively, but now they summarize mixed information about topology and weights (the terms $a_{ij}w_{jk}$

Table 4.2 The 14 most relevant commodity classes (plus aggregate trade) in year 2003 and the corresponding total trade value (USD), trade value per link (USD), and share of world aggregate trade. From Ref. [30]

HS Code	Commodity	Value (USD)	% of aggregate trade
84	Nuclear reactors, boilers, machinery and mechanical appliances; parts thereof	5.67×10^{11}	11.37
85	Electric machinery, equipment and parts; sound equipment; television equipment	5.58×10^{11}	11.18
27	Mineral fuels, mineral oils & products of their distillation; bitumin substances; mineral wax	4.45×10^{11}	8.92
87	Vehicles, (not railway, tramway, rolling stock); parts and accessories	3.09×10^{11}	6.19
90	Optical, photographic, cinematographic, measuring, checking, precision, medical or surgical instruments/apparatus; parts & accessories	1.78×10^{11}	3.58
39	Plastics and articles thereof.	1.71×10^{11}	3.44
29	Organic chemicals	1.67×10^{11}	3.35
30	Pharmaceutical products	1.4×10^{11}	2.81
72	Iron and steel	1.35×10^{11}	2.70
71	Pearls, precious stones, metals, coins, etc.	1.01×10^{11}	2.02
10	Cereals	3.63×10^{10}	0.73
52	Cotton, including yarn and woven fabric thereof	3.29×10^{10}	0.66
9	Coffee, tea, mate & spices	1.28×10^{10}	0.26
93	Arms and ammunition, parts and accessories thereof	4.31×10^{9}	0.09
ALL	Aggregate (all 97 commodities)	4.99×10^{12}	100.00

are determinant in this sense). Again, by plotting these quantities versus the strength we gather signals that countries highly involved in the WTW are connected with poorly trading countries, confirming a disassortative pattern (even if less prominent) for the weighted network. Interestingly, these patterns are perfectly reproduced by the quantities predicted using the ECM. This implies that the knowledge of both the number of trade partners of each node and the total amount of trade flowing through each country is maximally informative about the higher-order and non local dynamics of the whole network.

To further investigate this issue, we explore the evolution of the same properties over time. To this aim, for each network property we take the series of observed values, e.g. $\{k_i^{nn}\}_{i=1}^N$, and the series of its expected values, e.g. $\{\langle k_i^{nn}\rangle\}_{i=1}^N$; then, we compute the correlation coefficient between the analyzed property and the related constraint, e.g. k_i (for assumption $k_i \equiv \langle k_i\rangle$). The associated 95% confidence intervals are plotted as well. Figure 4.8 shows results perfectly in line with the outcome of the CM for the same data-set [23]. This implies that by simultaneously preserving degrees and strength, the ECM does not affect the ability of the CM to predict the topology of the WTW. We obtain a very close agreement between observed and expected values over time as confirmed by the correlation coefficient around $r \simeq 1$. Also for the weighted network properties we observe an excellent agreement between observed quantities and the corresponding averages over the ECM ensemble for the whole period. Indeed the correlation coefficients between observed and randomized properties is almost $r \simeq 1$ for all times.

We complete our analysis of the WTW as an undirected network by studying whether the picture changes when one considers the individual networks formed by imports and exports of single commodities. This application allows us also to gather information about the ECM ability to predict different networks according to their level of sparseness. Indeed, we know that the undirected WTW is a highly dense network (density $\sim$0.5) and we have already observed that some randomization techniques work only under specific conditions. Indeed, the commodities have been chosen and ordered according to the intensity of trade and level of aggregation.[1] For brevity we just show the scatter plot between binary and weighted higher-order properties and the related constraints, respectively $\{k_i\}_{i=1}^N$ and $\{s_i\}_{i=1}^N$ (Figs. 4.9 and 4.10).

We find that the results obtained in our aggregated study also hold for individual commodities, independently on the level of aggregation. We recognize a small improvement in the prediction according to the increase of network density (this is especially true for the weighted case), nevertheless the agreement is always very good. From an economic point of view, we can just point out a slight growth of dissortativity when less traded commodities are considered. Moreover, we confirm that the agreement between the model and the real network is still pronounced for the weighted quantities (Figs. 4.11, 4.12).

While the binary results confirm again the outcome shown in the previous chapter, the excellent agreement between observed and randomized weighted properties also for the commodity-specific case is surprising. The case of weighted clustering coefficient is really interesting in this sense. Indeed, in the aggregated case also the WCM

[1] In particular, among the 97 commodity classes, we have focused on the 14 particularly relevant commodities identified in [30] and reported in Table 4.2. These 14 commodities include the 10 most traded commodities in terms of total trade value (following the ranking in year 2003 [30]), plus 4 classes which are less traded but more relevant in economic terms. We selected the two least traded commodities in the set ($c = 93, 9$), two intermediate ones ($c = 39, 90$), the most traded one ($c = 84$), plus the network formed by combining all the top 14 commodities. The last sub-network represents an intermediate level of aggregation between single commodities and the completely aggregated data ($c = 0$).

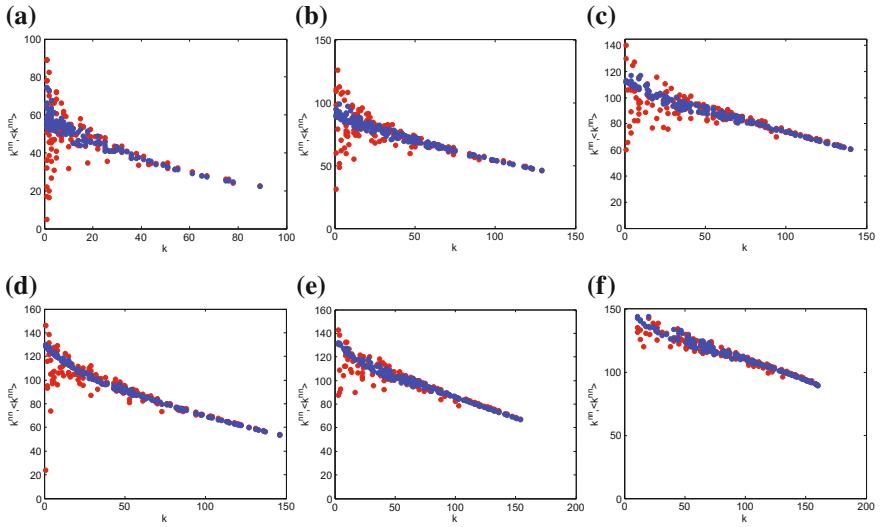

Fig. 4.9 Average nearest neighbor degree k_i^{nn} versus degree k_i in the 2002 snapshots of the commodity-specific (disaggregated) versions of the observed binary undirected WTW (red points), and corresponding average over the maximum entropy ensemble with specified degrees and strengths (blue points): **a** commodity 93; **b** commodity 09; **c** commodity 39; **d** commodity 90; **e** commodity 84; **f** aggregation of the top 14 commodities (see Table 4.1 for details). From **a** to **f**, the intensity of trade and level of aggregation increases

seemed to show a good prediction of this quantity, but this outcome is not robust to disaggregation. On the contrary, Fig. 4.12 shows that the ECM is not affected by this limit neither for c_i^w nor for any other network quantities.

The last step of our analysis consists in the comparison between the new enhanced model and the ordinary WCM in term of trade-off between accuracy of the results and parsimony in the use of constraints. Indeed, even if it is evident that the ECM performs better than the WCM in replicating WTW properties, we want to check if the former over-fits the network, i.e. if the introduction of degrees is redundant. As we mentioned, the WCM can be obtained as a particular case of the ECM by setting $x_i^* = 1 \; \forall \, i$, by 'switching' off the Lagrange parameters controlling for the degrees. The log-likelihood of the WCM is therefore the reduced function $\mathscr{L}(\mathbf{1}, \mathbf{y})$ of N variables, and is maximized by a new vector $\mathbf{y}^{**} \neq \mathbf{y}^*$, where $(\mathbf{x}^*, \mathbf{y}^*)$ stands for the solution of the ECM and $\mathbf{y}^{**}$ the solution of the WCM for the same observed network. For our two competing null models, following the suggestion in [30], we implement the AICc criterion, i.e.

$$\text{AICc}^*_{ECM} \equiv 4N + \frac{8N(2N+1)}{N^2 - 5N - 2} - 2\mathscr{L}(\mathbf{x}^*, \mathbf{y}^*), \tag{4.13}$$

$$\text{AICc}^*_{WCM} \equiv 2N + \frac{4N(N+1)}{N^2 - 3N - 2} - 2\mathscr{L}(\mathbf{1}, \mathbf{y}^{**}). \tag{4.14}$$

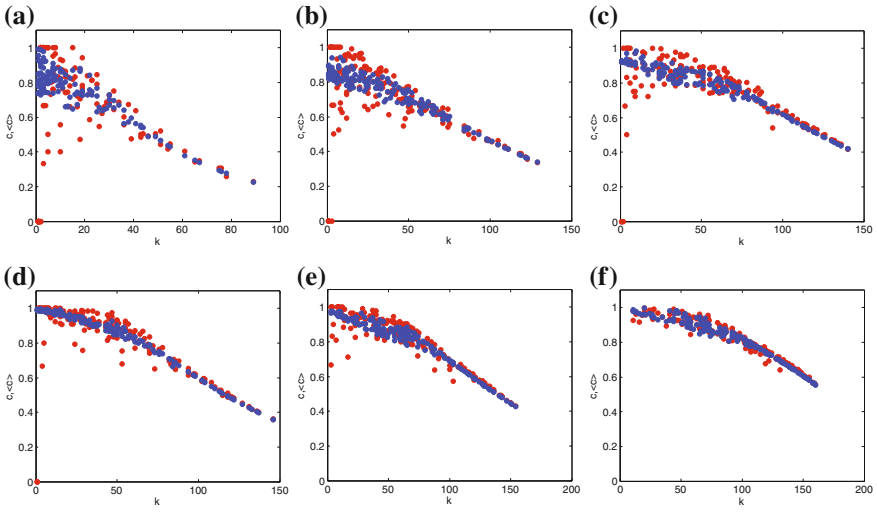

Fig. 4.10 Binary clustering coefficient c_i versus degree k_i in the 2002 snapshots of the commodity-specific (disaggregated) versions of the observed binary undirected WTW (red points), and corresponding average over the maximum entropy ensemble with specified degrees and strengths (blue points): **a** commodity 93; **b** commodity 09; **c** commodity 39; **d** commodity 90; **e** commodity 84; **f** aggregation of the top 14 commodities (see Table 4.1 for details). From **a** to **f**, the intensity of trade and level of aggregation increases

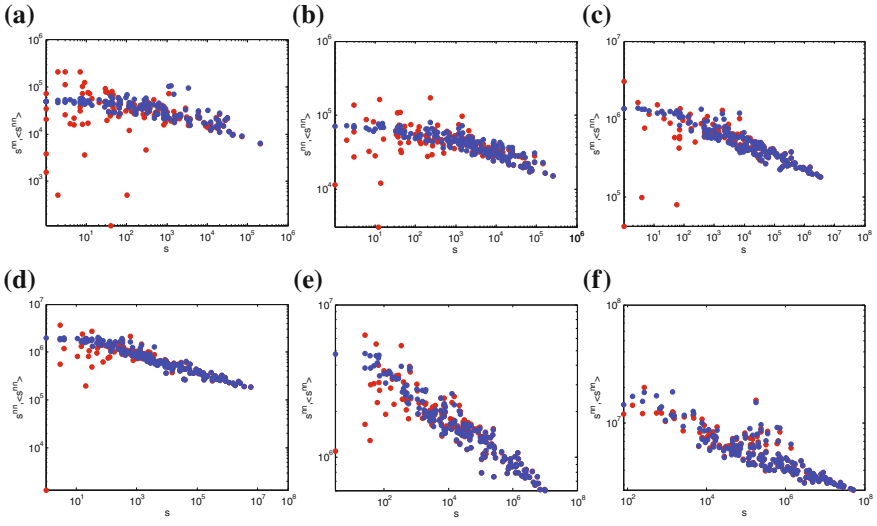

Fig. 4.11 Average nearest neighbor strength s_i^{nn} versus degree k_i in the 2002 snapshots of the commodity-specific (disaggregated) versions of the observed binary undirected WTW (red points), and corresponding average over the maximum entropy ensemble with specified degrees and strengths (blue points): **a** commodity 93; **b** commodity 09; **c** commodity 39; **d** commodity 90; **e** commodity 84; **f** aggregation of the top 14 commodities (see Table 4.1 for details). From **a** to **f**, the intensity of trade and level of aggregation increases

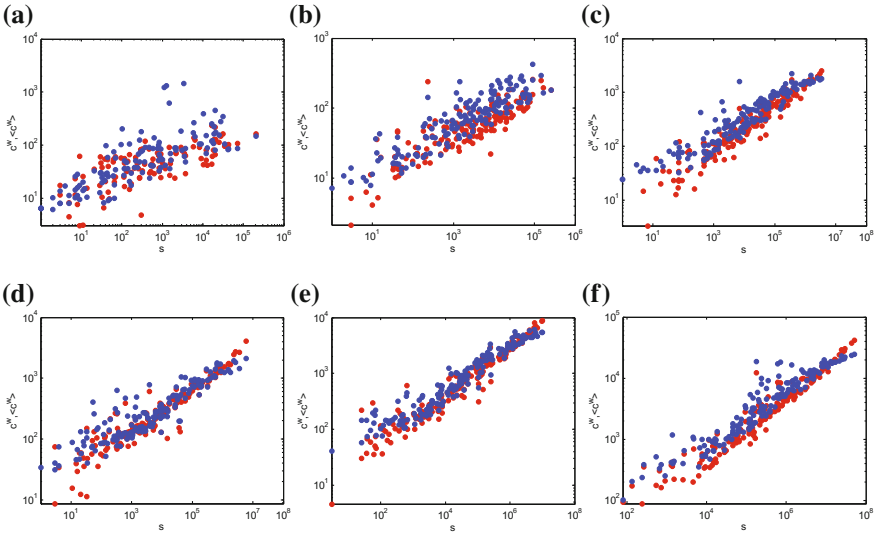

Fig. 4.12 Weighted clustering coefficient c_i^w versus degree k_i in the 2002 snapshots of the commodity-specific (disaggregated) versions of the observed binary undirected WTW (red points), and corresponding average over the maximum entropy ensemble with specified degrees and strengths (blue points): **a** commodity 93; **b** commodity 09; **c** commodity 39; **d** commodity 90; **e** commodity 84; **f** aggregation of the top 14 commodities (see Table 4.1 for details). From **a** to **f**, the intensity of trade and level of aggregation increases

The additional term is necessary because $n/K < 40$ (recall that n is the sample cardinality and K is the number of parameters of the model with more parameters, i.e. the ECM [30]). Akaike weights in our case read

$$w_{ECM}^{\text{AICc}} \equiv \frac{e^{-\frac{\text{AICc}_{ECM}}{2}}}{e^{-\frac{\text{AICc}_{ECM}}{2}} + e^{-\frac{\text{AICc}_{WCM}}{2}}}, \; w_{WCM}^{\text{AICc}} \equiv 1 - w_{ECM}^{\text{AICc}}. \tag{4.15}$$

Given a real network, a low value of w_{ECM}^{AICc} will indicate that the addition of the degree sequence is redundant (the relevant local constraints effectively reduce to the strength sequence, so the 'standard' WCM is preferable), while a high value of w_{ECM}^{AICc} will indicate that the local constraints are irreducible to the strength sequence (so the degrees must be separately specified). We stress that the result of this procedure is not predictable a priori (it depends on the numerical values of $\{s_i\}_{i=1}^N$ and $\{k_i\}_{i=1}^N$) and can only be achieved after a comparison with the ECM. Thus, even in cases when the WCM turns out to be the best model, our introduction of the ECM is still a necessary step making the whole approach self-consistent.

In Table 4.3 we show the results for the two competing model. We also used the Bayesian Information Criterion (BIC) [30]. Both criteria confirm that addition of the degree sequence to the WCM is non-redundant and extremely informative for the prediction of the WTW properties.

Table 4.3 AICc and BIC values, AICc and BIC weights for the considered null models applied to the WTW in 2002

	AICc	BIC	w^{AICc}	w^{BIC}
WCM	209.972	211.179	0	0
ECM	165.731	168.137	1	1

4.3 Further Reducing the Observational Requirements

Several analyses suggest that the probability for any two nodes to interact can be explicitly written in terms of non-structural quantities, which are typical of the system under analysis. This assumption rests upon the hypothesis that the activity of each node can be summed up by an 'intrinsic' quantity, g_i, called *fitness*, the latter being related to the corresponding Lagrange multiplier through a relation as $x_i = f(g_i)$. In the case of the World Trade Web such a relation reads $x_i = \sqrt{z}\,\text{GDP}_i\ \forall\, i$ [27].

This line of reasoning leads to a generalized procedure to make inference, whose defining equation can be assumed to be

$$p_{ij} = \frac{f(g_i)f(g_j)}{1 + f(g_i)f(g_j)}. \tag{4.16}$$

As noticed elsewhere [33], this generalization allows one to use of external as well as structural properties as fitnesses, provided that the aforementioned functional form is satisfied. This has profound implications on the kind of information that is necessary to know to accurately reconstruct a network. As we will show in the rest of the chapter, if the existing correlations between nodes strengths and nodes degrees are used, the amount of required information can be significantly reduced.

4.3.1 Bootstrap Method

Exploiting the correlations between strengths and degrees amounts at identifying the nodes fitnesses with the nodes strengths, i.e. $g_i = s_i\ \forall\, i$ and choosing a functional dependence relating x_i and s_i. Although this choice strongly depends on the particular system under analysis, in what follows we focus on a functional form which has been verified to hold for the economic and financial systems considered throughout this book: $x_i = \sqrt{z}\,s_i(\mathbf{W}^*)\ \forall\, i$ [31]. The nodes strengths $\{s_i(\mathbf{W}^*)\}_{i=1}^N$ are, thus, assumed to play the role of structural fitnesses controlling for the nodes degrees via a universal parameter z. Following this line of reasoning, Eq. (4.16) can be rewritten as

$$p_{ij} = \frac{z\,s_i(\mathbf{W}^*)s_j(\mathbf{W}^*)}{1 + z\,s_i(\mathbf{W}^*)s_j(\mathbf{W}^*)}. \tag{4.17}$$

Although Eq. (4.17) makes the knowledge of the whole degree sequence unnecessary, some kind of topological information is, however, still required in order to estimate the only unknown z. Upon resorting on the likelihood prescription mentioned in [10], the only equation to be solved is

$$\langle L \rangle = \sum_i \sum_{j<i} \frac{z\, s_i(\mathbf{W}^*) s_j(\mathbf{W}^*)}{1 + z\, s_i(\mathbf{W}^*) s_j(\mathbf{W}^*)} = L(\mathbf{A}^*). \tag{4.18}$$

The power of this approach lies in the fact that an accurate estimation of the whole degree sequence can be obtained at the negligible computational cost required to solve Eq. (4.18). Once the parameter z has been found, in fact, the unknown degrees can be easily estimated as $\tilde{k}_i = \sum_{j \neq i} \frac{z\, s_i(\mathbf{W}^*) s_j(\mathbf{W}^*)}{1 + z\, s_i(\mathbf{W}^*) s_j(\mathbf{W}^*)} \; \forall\, i$.

The algorithm we have described further clarifies the role that strengths have in the whole process of reconstruction. In particular, the information encoded into nodes strengths is not *per se* at a 'lower level' with respect to the information encoded into the degrees. What our method points out is the way to make a correct use of it: strengths should not be used to directly reconstruct the network but to, first, estimate the degrees and only then to be enforced as complementary constraints. In this sense, estimating the degrees from the strengths is equivalent to 'bootstrap' the ECM and provides an easy, yet very effective, recipe to obtain the degree sequence whenever the latter is not directly observable. Once the 'bootstrap' step has been carried on, we can rest upon our usual ECM estimation technique, enforcing the estimated degrees as genuine topological constraints:

$$\langle k_i \rangle = \sum_{j \neq i} \frac{x_i x_j y_i y_j}{1 - y_i y_j + x_i x_j y_i y_j} = \tilde{k}_i \quad \forall i, \tag{4.19}$$

$$\langle s_i \rangle = \sum_{j \neq i} \frac{x_i x_j y_i y_j}{(1 - y_i y_j)(1 - y_i y_j + x_i x_j y_i y_j)} = s_i(\mathbf{W}^*) \quad \forall i. \tag{4.20}$$

It is important to remark that the applicability of our 'bootstrapped' ECM crucially depends on the accuracy of the choice of the function $f(g_i)$ or, whenever strengths are used as fitnesses, whether the induced degrees are able to provide good estimates for the unknown degrees. This is indeed the case for the WTW and the Electronic Interbank Italian Market (e-MID), as Fig. 4.13 shows. The effectiveness of the aforementioned model in reproducing basic network properties is shown in Fig. 4.14.

4.3.2 The Degree-Corrected Gravity Model

Although the ECM (both in its complete and 'bootstrapped' version) represents a very accurate reconstruction model, solving Eqs. (4.20) and (4.19) can be, sometimes,

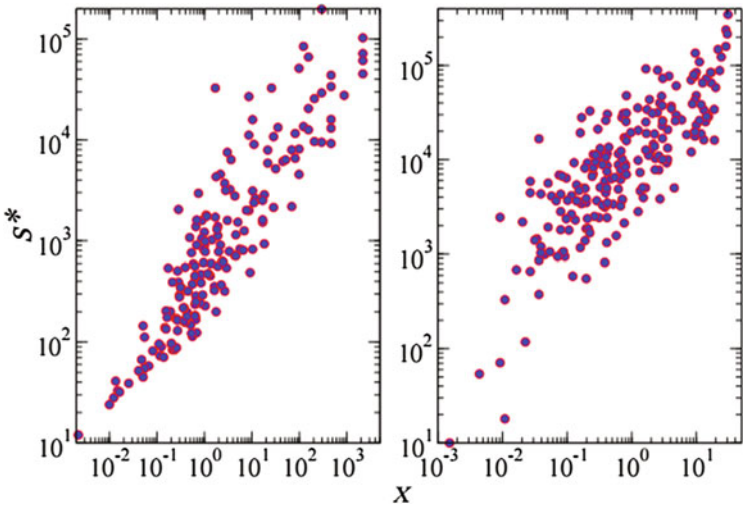

Fig. 4.13 Relation between node strengths $\{s_i(\mathbf{W}^*)\}_{i=1}^N$ and their degree-induced Lagrange multipliers $\{x_i\}_{i=1}^N$, obtained by solving the CM. The linearity of such a relation lies at the basis of our assumption that $x_i \propto s_i(\mathbf{W}^*)\ \forall\, i$ accurately describes the binary network topology. Left panel refers to the WTW, right panel to e-MID

computationally demanding. For this reason, a simpler version of the 'bootstrapped' ECM can be devised which is often as accurate as the aforementioned one.

Once more, economics provides the main inspiration for such a model: in particular, although the gravity model has been proved to provide a poor performance in reproducing the WTW topological structure, the observed trade exchanges between countries are, on the other hand, nicely reproduced by it [32]. In network terms, the gravity model prescription can be rephrased as

$$\langle w_{ij}\rangle = \frac{s_i(\mathbf{W}^*)s_j(\mathbf{W}^*)}{2W^*_{tot}} \tag{4.21}$$

where W^*_{tot} is the observed total weight. However, as already noticed in Chap. 2, the recipe provided by Eq. (4.21) suffers from a number of limitations, the most evident being that it induces a fully-connected network configuration. A straightforward way to both retain the explanatory power of the gravity model and avoid ending up with a complete network is provided by the following recipe

$$w_{ij} = \begin{cases} 0 & \text{with probability } 1 - p_{ij}, \\ \frac{s_i(\mathbf{W}^*)s_j(\mathbf{W}^*)}{2W^*_{tot}} p_{ij} & \text{with probability } p_{ij} \end{cases} \tag{4.22}$$

with $p_{ij} = \frac{z\, s_i(\mathbf{W}^*)s_j(\mathbf{W}^*)}{1+z\, s_i(\mathbf{W}^*)s_j(\mathbf{W}^*)}$. Each entry of the network adjacency matrix is 'weighted' by its probability of appearance, thus ensuring that the estimation provided by

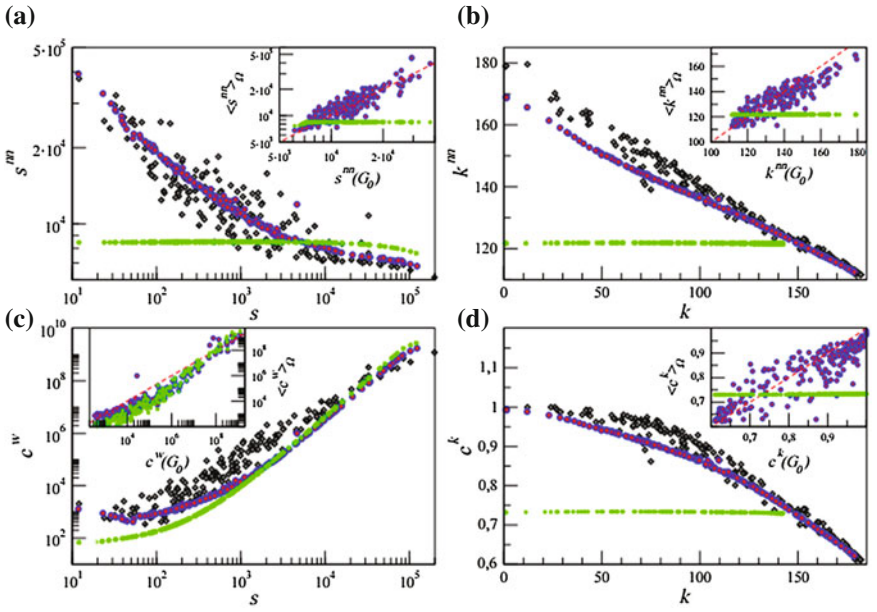

Fig. 4.14 Scatter plots of the average nearest neighbor strength s_i^{nn} versus strength s_i **a** average nearest neighbor degree k_i^{nn} versus degree k_i **b** weighted clustering coefficient c_i^w versus strength s_i **c** and binary clustering coefficient c_i versus degree k_i **d** for the real quantities (dark points), those estimated by our method (purple points) and those computed by a WCM-based reconstruction (green points). Insets: scatter plots of the observed values versus the expected values for the same quantities

Eq. (4.21) is preserved. The latter is appealing because provides a very simple recipe, yet able to satisfy the weighted constraints ($\langle s_i \rangle = s_i(\mathbf{W}^*) = \sum_j w_{ij}^* \ \forall \ i$). Equations (4.22) define the *degree-corrected gravity model* [33].

It should be noticed that, in order to ensure that the strengths are correctly reproduced, the sum above must run over all indices, including $i = j$. Terms of this kind describe nodes self-interactions which, however, are highly unrealistic. Excluding diagonal terms from our analysis implies that our sums must run over different indices, i.e. $j \neq i$, thus causing the expectations coming from the degree-corrected gravity model to need an extra-term to restore the correct value . More explicitly,

$$\langle s_i \rangle = \sum_{j \neq i} \langle w_{ij} \rangle = \frac{s_i(\mathbf{W}^*)(2W_{tot}^* - s_i(\mathbf{W}^*))}{2W_{tot}^*} = s_i(\mathbf{W}^*) - \frac{s_i^2(\mathbf{W}^*)}{2W_{tot}^*}, \tag{4.23}$$

and the missing term to be added up to our expectations is precisely the diagonal term, i.e. $\langle w_{ii} \rangle$. Following [34], we adopt the prescription of redistributing the diagonal term $\langle w_{ii} \rangle$ across the i-th row and the i-th column. As discussed in [34], redistributing the diagonal terms across the corresponding rows and columns amounts to redistribute

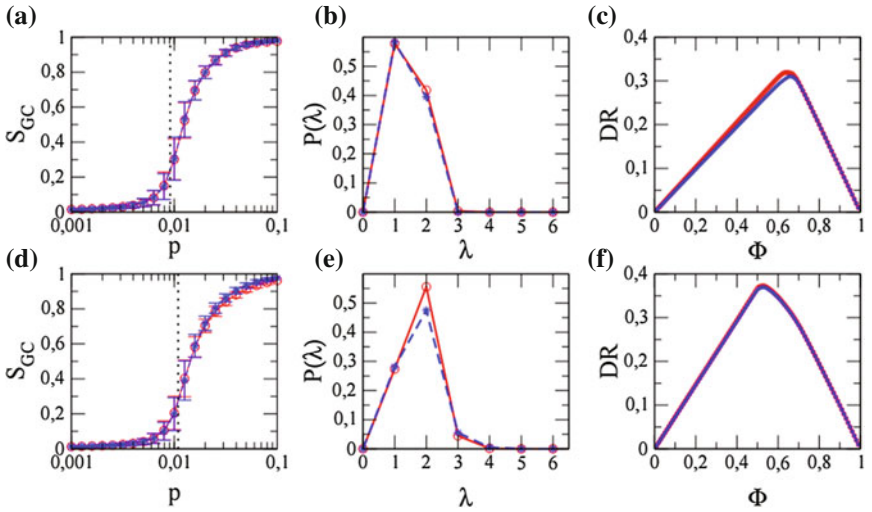

Fig. 4.15 Properties of real and synthetic networks. Left panels (**a**, **d**): dependence of the size of the giant component on the occupation probability p. Central panels (**b**, **e**): empirical probability distribution of the directed shortest path length λ. Right panels (**c**, **f**): dependence of the DebtRank index on the initial distress Φ [33]. Top panels refer to the WTW, bottom panels to e-MID

the strengths of the following matrix on the entries equal to 1

$$\begin{array}{ccccc|c} 0 & 1 & 1 & 1 & \dots & \frac{s_1^2(\mathbf{W}^*)}{2W_{tot}^*} \\ 1 & 0 & 1 & 1 & \dots & \frac{s_2^2(\mathbf{W}^*)}{2W_{tot}^*} \\ 1 & 1 & 0 & 1 & \dots & \frac{s_3^2(\mathbf{W}^*)}{2W_{tot}^*} \\ 1 & 1 & 1 & 0 & \dots & \frac{s_4^2(\mathbf{W}^*)}{2W_{tot}^*} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \frac{s_1^2(\mathbf{W}^*)}{2W_{tot}^*} & \frac{s_2^2(\mathbf{W}^*)}{2W_{tot}^*} & \frac{s_3^2(\mathbf{W}^*)}{2W_{tot}^*} & \frac{s_4^2(\mathbf{W}^*)}{2W_{tot}^*} & \dots & \end{array} \tag{4.24}$$

This problem can be solved by implementing a procedure inspired to the iterative proportional fitting (IPF) algorithm [35], defined by the following iterations

$$\begin{cases} w_{ij}^{(n)} = \frac{s_i^2(\mathbf{W}^*)}{2W_{tot}^*} \left(\frac{w_{ij}^{(n-1)}}{\sum_{k(\neq i)} w_{ik}^{(n-1)}} \right) \\ w_{ij}^{(n+1)} = \frac{s_j^2(\mathbf{W}^*)}{2W_{tot}^*} \left(\frac{w_{ij}^{(n)}}{\sum_{k(\neq j)} w_{kj}^{(n)}} \right) \end{cases} \tag{4.25}$$

and upon setting the matrix defined by $w_{ij}^{(0)} = 1 \; \forall \, i \neq j$ as the initial configuration. The probabilistic recipe defining the degree-corrected gravity model thus becomes

$$w_{ij} = \begin{cases} 0 & \text{with probability } 1 - p_{ij}, \\ \left(\frac{s_i(\mathbf{W}^*) s_j(\mathbf{W}^*)}{2W^*_{tot}} + w^{(\infty)}_{ij} \right) \frac{1}{p_{ij}} & \text{with probability } p_{ij}; \end{cases} \tag{4.26}$$

the analytical functional form of the first three IPF algorithm iterations read:

$$\begin{aligned} w^{(1)}_{ij} &= \frac{s_i^2(\mathbf{W}^*)}{2W^*_{tot}} \left[\frac{1}{N-1} \right]; \\ w^{(2)}_{ij} &= \frac{s_i^2(\mathbf{W}^*)}{2W^*_{tot}} \left[\frac{s_j^2(\mathbf{W}^*)}{\sum_{l \neq j} s_l^2(\mathbf{W}^*)} \right]; \\ w^{(3)}_{ij} &= \frac{s_i^2(\mathbf{W}^*)}{2W^*_{tot}} \left[\frac{s_j^2(\mathbf{W}^*)}{\sum_{l \neq j} s_l^2(\mathbf{W}^*)} \right] \left[\frac{1}{\sum_{k \neq i} \frac{s_k^2(\mathbf{W}^*)}{\sum_{m \neq k} s_m^2(\mathbf{W}^*)}} \right]. \end{aligned} \tag{4.27}$$

A directed version of the degree-corrected gravity model has been defined as well [33] whose performance is show in Fig. 4.15.

References

1. R. Mastrandrea, T. Squartini, G. Fagiolo, D. Garlaschelli, Enhanced reconstruction of weighted networks from strengths and degrees. New J. Phys. **16**, 043022 (2014)
2. A. Barrat, M. Barthlemy, A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008)
3. C. Lynch, Big data: How do your data grow? Nature **455**(7209), 28–29 (2008)
4. S. Lohr, The Age of Big Data, New York Times **11** (2012)
5. S. Wells, Financial interlinkages in the United Kingdom's interbank market and the risk of contagion, Bank of England Working Paper **230**/2004 (2004)
6. L. Bargigli, M. Gallegati, Random digraphs with given expected degree sequences: a model for economic networks. J. Econ. Behav. Organ. **78**(3), 396–411 (2011)
7. N. Musmeci, S. Battiston, G. Caldarelli, M. Puliga, A. Gabrielli, Bootstrapping topological properties and systemic risk of complex networks using the fitness model. J. Stat. Phys. **151**(3–4), 720–734 (2013)
8. G. Caldarelli, A. Chessa, F. Pammolli, A. Gabrielli, M. Puliga, Reconstructing a credit network. Nat. Phys. **9**(3), 125–126 (2013)
9. I. Mastromatteo, E. Zarinelli, M. Marsili, Reconstruction of financial networks for robust estimation of systemic risk. J. Stat. Mech. **03**, P03011 (2012)
10. D. Garlaschelli, M.I. Loffredo, Structure and evolution of the world trade network. Physica A **355**, 138–144 (2005)
11. J. Park, M.E.J. Newman, Statistical mechanics of networks. Phys. Rev. E **70**(6), 066117 (2004)
12. T. Squartini, D. Garlaschelli, Analytical maximum-likelihood method to detect patterns in real networks. New J. Phys. **13**(8), 083001 (2011)
13. G. De Masi, G. Iori, G. Caldarelli, Fitness model for the Italian interbank money market. Phys. Rev. E **74**(6), 066112 (2006)
14. P.D. Killworth, H.R. Bernard, Informant accuracy in social network data. Hum. Organ. **35**(3), 269–286 (1976)
15. http://vlado.fmf.uni-lj.si/pub/networks/data/bio/foodweb/foodweb.htm
16. T. Squartini, G. Fagiolo, D. Garlaschelli, Randomizing world trade. II. A weighted network analysis. Phys. Rev. E **84**(4), 046118 (2011)

17. G. Fagiolo, T. Squartini, D. Garlaschelli, Null models of economic networks: the case of the world trade web. J. Econ. Interact. Coord. **8**(1), 75–107 (2012)
18. M.A. Serrano, M. Boguná, Weighted configuration model. AIP Conference Proceedings **776**(101) (2005)
19. M.A. Serrano, M. Boguná, R. Pastor-Satorras, Correlations in weighted networks. Phys. Rev. E **74**, 055101(R) (2006)
20. S. Fortunato, Community detection in graphs. Phys. Rep. **486**(3), 75–174 (2010)
21. G. Bianconi, Entropy of network ensembles. Phys. Rev. E **79**(3), 036114 (2009)
22. D. Garlaschelli, M.I. Loffredo, Generalized bose-fermi statistics and structural correlations in weighted networks. Phys. Rev. Lett. **102**(3), 038701 (2010)
23. T. Squartini, G. Fagiolo, D. Garlaschelli, Randomizing world trade. I. A binary network analysis. Phys. Rev. E **84**(4), 046117 (2011)
24. G. Fagiolo, Clustering in complex directed networks. Phys. Rev. E **76**(2), 026107 (2007)
25. A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. Proc. Nat. Acad. Sci. **101**(11), 3747–3752 (2004)
26. K. Bhattacharya, G. Mukherjee, J. Saramäki, K. Kaski, S.S. Manna, The international trade network: weighted network analysis and modelling. J. Stat. Mech. **02**, P02002 (2008)
27. D. Garlaschelli, M.I. Loffredo, Maximum likelihood: extracting unbiased information from complex networks. Phys. Rev. E **78**(1), 015101 (2008)
28. H. Akaike, A new look at the statistical model identification. IEEE Trans. Aut. Cont. **19**, 716–723 (1974)
29. We also used the Bayesian Information Criterion (BIC) that puts a higher penalty on the number of parameters. We found that BIC weights are identical to the AIC ones (within machine precision) for all networks in our samples
30. K.P. Burnham, D.R. Anderson, Multimodel inference: understanding AIC and BIC in Model Selection. Soc. Met. Res. **33**, 261–304 (2004)
31. G. Cimini, T. Squartini, A. Gabrielli, D. Garlaschelli, Estimating topological properties of weighted networks from limited information. Phys. Rev. E **92**, 040802 (2015)
32. A. Almog, R. Bird, D. Garlaschelli, Enhanced Gravity Model of trade: reconciling macroeconomic and network models, (2015), arXiv:1506.00348
33. G. Cimini, T. Squartini, A. Gabrielli, D. Garlaschelli, Systemic risk analysis in reconstructed economic and financial networks. Sci. Rep. **5**, 15758 (2015)
34. T. Squartini, G. Cimini, A. Gabrielli, D. Garlaschelli, Network reconstruction via density sampling, Applied Network Science, 2(1), 3 (2016), arXiv:1610.05494
35. M.Y. Bishop, S.E. Fienberg, W.P. Holland, *Discrete Multivariate Analysis: Theory and Practice* (Springer Science & Business Media, Berlin, 2007)

Chapter 5
Graph Combinatorics

The heavier the burden, the closer our lives come to the earth, the more real and truthful they become. Conversely, the absolute, absence of burden causes man to be lighter than air, to soar into, heights, take leave of the earth and his earthly being, and become only half real, his movements as free as they are, insignificant. What then shall we choose? Weight or lightness?

—Milan Kundera, *Nesnesitelná Lehkost Bytí*

Abstract In this chapter we go back to a formal level and discuss the connection between the maximum-entropy ensembles of constrained graphs considered so far and various combinatorial problems in the asymptotic limit of an infinite number of nodes. This seemingly mysterious connection is actually a natural consequence of the fact that, for any discrete combinatorial problem where we need to sample or enumerate the (microcanonical) configurations compatible with a given 'hard' constraint, there exists a dual (canonical) problem induced by the 'soft' version of the same constraint. Thus, if the microcanonical and canonical ensembles are asymptotically equivalent, one can operate in the canonical ensemble and, up to finite-size corrections, extend the results to the macrocanonical one, which is otherwise very hard to deal with. It is therefore intriguing to relate the feasibility of combinatorial problems to the property of ensemble equivalence. We show that, while graphs with a single constraint on the total number of links are ensemble-equivalent, graphs with given degree sequence are not. Unlike other examples in statistical physics, where the lack of ensemble equivalence arises from long-range interactions or non-additivity, here the novel mechansim is the extensivity of the number of constraints. We discuss important implications for graph combinatorics and for the choice of the correct ensemble in practical situations. The final result is an explicit connection between the solution of a combinatorial problem and the (non)equivalence between the two associated microcanonical and canonical ensembles.

T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-69438-2_5

5.1 A Dual Route to Combinatorics?

As in the above reflection in Milan Kundera's "*The unbearable lightness of being*", which revolves around the dilemma of choosing between 'light' and 'heavy', in this chapter we ponder the duality of 'soft' and 'hard'. In Sect. 2.2 we have clarified that, in general, a given set of structural properties $\mathbf{C}^*$ can be implemented either as a 'hard' constraint, thus giving rise to the microcanonical ensemble, or as a 'soft' constraint, thus giving rise to the canonical ensemble. From that point onwards, we have decided to choose the canonical ensemble as our approach, due to the fact that, when local node-specific constraints are imposed on graphs, the canonical ensemble is analytically tractable while the microcanonical one is not. This has indeed led us to a series of successes that would have been impossible to achieve microcanonically.

Combinatorics, a branch of discrete mathematics that studies the construction, enumeration, and existence of discrete structures [1], is on the other hand an inherently 'microcanonical' discipline. For instance, in the context of graph theory (with which combinatorics strongly overlaps [2]), typical combinatorial problems are *graph sampling*, i.e. generating and mathematically characterizing instances of graphs that exactly match some criterion, and the closely related one of *graph enumeration*, i.e. counting how many graphs match a given set of constraints. Combinatorial problems are difficult precisely because of the 'hard' nature of the imposed constraints. Apparently, the knowledge we have accumulated about the canonical ensemble seems of no help in this case. But is this completely true?

In this chapter, we address some problems in graph combinatorics. As for the previous topics treated in this book, our aim is not that of reviewing the established field-specific approaches to the problem. In fact, combinatorics is a huge branch of discrete mathematics and it has its own well-developed toolkit of mathematical techniques, which we are not going to discuss here. Rather, we aim at emphasizing the fact that the maximum-entropy approach can give its unique contribution to certain combinatorial problems as well, highlighting some aspects that are otherwise not apparent. We are not going to consider the sampling and enumeration of graphs for any specific application, but we want to understand some of the properties of these combinatorial problems at a formal level. Interestingly, we will see in the end that, despite its abstract nature, this theoretical investigation has (somewhat unexpected) concrete consequences for how the practical problems discussed in Chaps. 3 and 4 have to be dealt with, and how the approach we used so far should be interpreted.

Our strategy is the following. In Sect. 5.2, we argue that, given a combinatorial problem with hard constraints $\mathbf{C}^*$, one can formally define a dual problem with 'softened' constraints. This can be achieved by reformulating the original microcanonical problem in the conjugate canonical ensemble defined by the same $\mathbf{C}^*$. In Sect. 5.3, using both simple examples and general arguments, we then characterize not only the exact canonical solution, but also the 'divergence' between the canonical ensemble and the corresponding microcanonical one in the thermodynamic limit (recall the definition in footnote 11). After considering alternative definitions for such a divergence, we rigorously quantify it in terms of the Kullback-Leibler divergence,

or relative entropy, between the canonical and microcanonical probability distributions $P_{\text{can}}(\mathbf{G})$ and $P_{\text{mic}}(\mathbf{G})$. Along the way, quantifying the difference between the canonical and microcanonical ensembles leads us to a beautiful journey at the fundamentals of statistical physics, where the notion of *ensemble equivalence* we have briefly anticipated in Sect. 2.2 is crucial for the calculation of many quantities. In Sect. 5.4 we show that ensembles of graphs with a single global constraint are equivalent, wheras those characterized by a set of local constraints are not. Besides proving non-equivalence, we also quantify it rigorously to leading order. In Sect. 5.5 we try to relate the unknown microcanonical solution to a combination of the exact canonical one and the calculated divergence between the two. Our end result will be that, for constraints that lead to ensemble equivalence, the canonical ensemble can ideed provide explicit solutions, corrected for the finite size of the system, to problems in graph combinatorics. On the other hand, for non-equivalent problems, combinatorial expressions will depend explicilty on the relative entropy between the canonical and microcanonical ensemble. This novel result establishes an enchanting, explicit connection between the solution of a combinatorial enumeration problem and the (non)equivalence between the two associated dual maximum-entropy ensembles. We conclude with a discussion, in Sect. 5.6, of the consequences of our results for the choice of the ensemble to work with in practical situations.

5.2 'Soft' Combinatorial Enumeration

Let us consider the canonical ensemble defined by the soft constraint $\mathbf{C}^*$ and assume that the Lagrange multipliers $\boldsymbol{\theta}$ have already been fixed to their maximum-likelihood values $\boldsymbol{\theta}^*$ that ensure $\langle \mathbf{C} \rangle = \mathbf{C}^*$ (see Sect. 2.2.1). For ease of the notation, we drop the symbol $\boldsymbol{\theta}^*$ from $P_{\text{can}}(\mathbf{G}|\boldsymbol{\theta}^*)$ and simply write $P_{\text{can}}(\mathbf{G})$. Let us now denote by $Q_{\text{can}}(\mathbf{C})$ the probability that, in a graph $\mathbf{G}$ sampled with canonical probability $P_{\text{can}}(\mathbf{G})$, the constraints take the particular value $\mathbf{C}$. Note that $Q_{\text{can}}(\mathbf{C})$ depends on $\boldsymbol{\theta}^*$, but this dependence has been omitted here as well. Of course, $\mathbf{C}$ is a random variable which is a function $\mathbf{C}(\mathbf{G})$ of the random variable $\mathbf{G}$. Now, since all graphs $\mathbf{G}$ that have the same value $\mathbf{C}(\mathbf{G})$ of the constraints are assigned the same probability $P_{\text{can}}(\mathbf{G})$, and since there is an (unknown) number $\Omega_{\mathbf{C}}$ of such graphs, we have

$$Q_{\text{can}}(\mathbf{C}) = \Omega_{\mathbf{C}}\, P_{\text{can}}(\mathbf{G}). \tag{5.1}$$

Inverting and setting $\mathbf{C} = \mathbf{C}^*$, we get

$$\Omega_{\mathbf{C}^*} = \frac{Q^*_{\text{can}}}{P^*_{\text{can}}} \tag{5.2}$$

where we have defined

$$P^*_{\text{can}} \equiv P_{\text{can}}(\mathbf{G}^*), \qquad Q^*_{\text{can}} \equiv Q_{\text{can}}(\mathbf{C}^*), \tag{5.3}$$

$\mathbf{G}^*$ being any of the 'microcanonical' graphs for which $\mathbf{C}(\mathbf{G}^*) = \mathbf{C}^*$.

Equation (5.2) establishes a simple but important connection between *a microcanonical, purely combinatorial quantity* (left hand side) and *two canonical, purely probabilistic quantities* (right hand side). Moreover it shows that, since we can calculate the value of P^*_{can} exactly, the calculation of $\Omega_{\mathbf{C}^*}$ reduces to that of Q^*_{can} only. The latter can be visualized as the height of the 'peak' of the probability $Q_{\text{can}}(\mathbf{C})$ at the value $\mathbf{C}^*$. Since $Q_{\text{can}}(\mathbf{C})$ must sum up to one, Q^*_{can} must become closer to 1 as $Q_{\text{can}}(\mathbf{C})$ becomes more peaked. Indeed, if we define a corresponding quantity $Q_{\text{mic}}(\mathbf{C})$ as the probability of observing the value $\mathbf{C}$ in the microcanonical ensemble, we have $Q^*_{\text{mic}} \equiv Q_{\text{mic}}(\mathbf{C}^*) = 1$, because $\mathbf{C}$ is a deterministic variable in that ensemble. This observation suggests that, if the ensembles become asymptotically equivalent in the thermodynamic limit, then $Q^*_{\text{can}} \to 1$ and hence $\Omega_{\mathbf{C}^*} \to 1/P^*_{\text{can}}$, which can be calculated exactly. Thus deriving an asymptotic expression for P^*_{can} allows one to do the same for $\Omega_{\mathbf{C}^*}$. On the other hand, we also expect that if the ensembles are not asymptotically equivalent, then the value of the constraints in the canonical ensemble will keep fluctuating around the value $\mathbf{C}^*$, even in the thermodynamic limit. We therefore expect that Q^*_{can} does not converge to 1, but to a value dictated by the amplitude of the fluctuations, i.e. by the breadth of the distribution $Q_{\text{can}}(\text{C})$. We will confirm these expectations rigorously in Sect. 5.5.

5.3 Quantifying Ensemble (non)equivalence

In statistical physics, calculating the equilibrium properties of a system with a given total energy requires averaging over the so-called *microcanonical ensemble* [3, 4], i.e. the set of all configurations having exactly the same value of the energy. Apart from trivial cases, this is a mathematically challenging task. Moreover, it is difficult to physically realize a condition when there is really no uncertainty in the value of the total energy. Therefore, one often prefers to work within the *canonical ensemble* [4], i.e. an extended set of configurations including those with the 'wrong' energy but such that the average energy over the ensemble matches the specified value. This matching is achieved through the selection of an appropriate temperature, mathematically arising as the Lagrange multiplier enforcing the average total energy.

Starting from Gibbs [4], the canonical and microcanonical ensembles have been shown to be equivalent in the thermodynamic limit for simple examples like ideal gases or other non-interacting systems. The original argument is that in a canonical ensemble at fixed temperature the energy fluctuations should be negligible with respect to the total energy, so that in the thermodynamic limit the ensemble is effectively microcanonical with a unique value of the energy. Today, most textbooks still convey the message that the equivalence of ensembles holds universally for every system, justifying the use of energy and temperature as two different parameters giving an equivalent description of the equilibrium properties of basically every system.

However, over the last couple of decades various studies have highlighted that, for certain many-body systems encountered in models of fluid turbulence [5, 6], quantum phase separation [7–9], star formation [10, 11], nuclear fragmentation [12] and spin interaction [13], ensemble equivalence breaks down. Physically, it is believed that one of the main causes of ensemble nonequivalence is the presence of long-range interactions. Mathematically, the problem has been approached in various ways [14, 15].

In particular, microcanonical and canonical ensembles are said to be *thermodynamically equivalent* [9] when the entropy and the free energy are one-to-one related by a Legendre transform. The ensembles are instead said to be *macrostate equivalent* [14] when the set of equilibrium values of the macrostate (energy, magnetization, etc.) predicted by the two ensembles are the same. Finally, a recent and mathematically appealing definition is that of *measure equivalence* [15], according to which the ensembles are said to be equivalent when the canonical probability distribution converges to the micro-canonical probability distribution in the thermodynamic limit. Under certain conditions, the three definitions have been shown to be equivalent [15]. Moreover, large deviations theory [16] shows that the ensembles are nonequivalent on all three levels when the microcanonical entropy function is nonconcave as a function of the energy density in the thermodynamic limit [15]. This is an interesting insight, because for a long time physicists had believed that the entropy is always a strictly concave function.

In Sect. 5.3.3 we adopt the third definition and study ensemble (non)equivalence for our ensembles of graphs with given topological constraints, following [17]. Before doing that, however, we consider (non)equivalence from the point of view of its possible manifestations on the marginal probabilities (Sect. 5.3.1) and on the canonical fluctuations of the constraints (Sect. 5.3.2).

5.3.1 Marginal Equivalence

Let us start by comparing the *marginal* probabilities of the microcanonical and canonical distributions. We stress that, for mathematically solvable choices of the constraints, the maximum-entropy method introduced in Chap. 2 directly provides 'from the beginning' the explicit values of the probabilities p_{ij}^{can} that a link from i to j is there. The superscript stands for 'canonical' and the probability is evaluated at the parameter values that maximize the likelihood. Let us start by considering some simple graphs. We shall only focus on unweighted networks for simplicity.

The Random Graph Model

Let us start with simple undirected graphs with a fixed number of links, i.e. $\mathbf{C} \equiv L$. Writing $L = \lambda M$ where $M \equiv N(N-1)/2$ is the number of pairs of vertices and λ is the fraction of realized links, we have for the microcanonical ensemble

$$\Omega_{L^*} = \binom{M}{L^*} = \binom{M}{\lambda^* M}. \tag{5.4}$$

The canonical distribution (2.11) can be obtained setting $H(\mathbf{G}, \theta) = \theta L(\mathbf{G})$ and $p^* \equiv \frac{e^{-\theta^*}}{1+e^{-\theta^*}} = \lambda^*$ [17]. This produces an Erdős-Rényi random graph where each pair of nodes is connected with probability $p^* = L^*/M$:

$$P_{\text{can}}(\mathbf{G}) = (p^*)^{L(\mathbf{G})}(1-p^*)^{M-L(\mathbf{G})}. \tag{5.5}$$

Marginal distributions describe the behavior of the single random variables, disregarding the correlations: since we are dealing with binary networks, marginals are Bernoulli distributions. We can, thus, focus on just one coefficient. In the microcanonical case, the probability p_{ij}^{mic} that the single-link random variable a_{ij} is 1 requires the knowledge of the number of configurations where that link is indeed present. This number can be obtained by repeating the counting process carried out for the whole network, while keeping the value of the specific entry fixed. Thus, the generic microcanonical marginal reads

$$p_{ij}^{\text{mic}} = \frac{\Omega_{\mathbf{L}^*|a_{ij}=1}}{\Omega_{\mathbf{L}^*}} = \binom{\frac{N(N-1)}{2}-1}{L-1} \Big/ \binom{\frac{N(N-1)}{2}}{L} = \frac{2L}{N(N-1)} = p^* = p_{ij}^{\text{can}}, \tag{5.6}$$

showing that the two marginal probabilities coincide.

The 'sparsest graph' model

We now consider all graphs characterized by N vertices (with N being an even number) and degree sequence equal to $k_i = 1$, $\forall i$ (i.e. all nodes have a degree which is exactly 1). In order to realize such constraints, nodes must be linked in pairs, in order to ensure that (just) $N/2$ links are present. The number of such configurations is

$$\Omega_{\mathbf{k}^*} = \frac{N!}{2^{N/2}\binom{N}{2}!} \tag{5.7}$$

which can be intuitively justified by considering that while a permutation of node-labels produces a different network structure, two operations exist that leave it unaltered: exchanging the labels of any two linked nodes and permuting the links labels. The microcanonical marginals can be computed by imagining to keep a given pair of nodes connected while reconnecting the other ones:

$$p_{ij}^{\text{mic}} = \frac{\Omega_{\mathbf{k}^*|a_{ij}=1}}{\Omega_{\mathbf{k}^*}} = \frac{(N-2)!}{2^{(N-2)/2}\binom{N-2}{2}!} \cdot \frac{2^{N/2}\binom{N}{2}!}{N!} = \frac{1}{N-1}. \tag{5.8}$$

In the canonical framework, a graph of this kind can be formally reconduced to the Random Graph Model, described by the probability distribution

$$P_{\text{can}}(\mathbf{G}) = \left(\frac{1}{N-1}\right)^{\frac{N}{2}} \left(1-\frac{1}{N-1}\right)^{\binom{N}{2}-\frac{N}{2}} \tag{5.9}$$

where $p^* = \frac{1}{N-1} = p_{ij}^{\text{can}} = p_{ij}^{\text{mic}}$, again showing equivalence of the marginal probabilities.

The name of this particular class of graphs comes from the observation that the link density of any graph of this kind reads $c = \frac{1}{N-1}$; interestingly, in order for any graph with N vertices (N being an even number) to have a lower density, a number of links $L < N/2$ should be observed: this, however, would allow for the presence of disconnected vertices.

The h-star model

Our third and last example of graphs concerns h-star graphs, characterized by a fully connected 'core' of h hubs and a 'periphery' of $N - h$ 'leaves', each with a single connection to one of the hubs. The periphery is divided into h sets of equal size $(N-h)/h$, such that every leaf in each of the sets is connected to the same hub. The microcanonical number of these graphs is given by the multinomial coefficient

$$\Omega_{\mathbf{k}^*} = \left(\frac{N-h}{h}, \frac{N-h}{h} \dots \frac{N-h}{h}\right)! = \frac{(N-h)!}{\left(\frac{N-h}{h}!\right)^h}; \tag{5.10}$$

on the other hand, the number of microcanonical configurations where a given link is present can be estimated upon keeping the corresponding leave-node label fixed while permuting the labels of the remaining leaves. The corresponding microcanonical marginal reads

$$p_{CP}^{\text{mic}} = \frac{\Omega_{\mathbf{k}^*|a_{CP}=1}}{\Omega_{\mathbf{k}^*}} = \frac{(N-h-1)!}{\left(\frac{N-h}{h}!\right)^{h-1}\left(\frac{N-h}{h}-1!\right)} \cdot \frac{\left(\frac{N-h}{h}!\right)^h}{(N-h)!} = \frac{1}{h}. \tag{5.11}$$

Now, the canonical probability is still described by Eqs. (2.19), (2.20) and (2.21). Noting that the degree of a peripheral node is $k_P = 1$ and that of a core node is $k_C = h - 1 + \frac{N-h}{h}$, Eq. (2.21) reduces to only two independent equations

$$h\, p_{CP}^* + (N-h-1)\, p_{PP}^* = 1, \tag{5.12}$$

$$(h-1)\, p_{CC}^* + (N-h)\, p_{CP}^* = (h-1) + \frac{N-h}{h}, \tag{5.13}$$

where p_{CC}^*, p_{PP}^* and p_{CP}^* denote the connection probability (evaluated at maximum likelihood) between two core nodes, two peripheral nodes, and a core node and a peripheral node respectively. The implicit solution is

$$p_{CC}^* = 1, \qquad p_{PP}^* = 0, \qquad p_{CP}^* = h^{-1} \tag{5.14}$$

thus guaranteeing that $p_{CP}^{\text{mic}} = p_{CP}^{\text{can}}$.

A real-world example

When coming to real-world examples, marginal equivalence can only be tested numerically. In particular, while the canonical approach is still analytical, the micro-

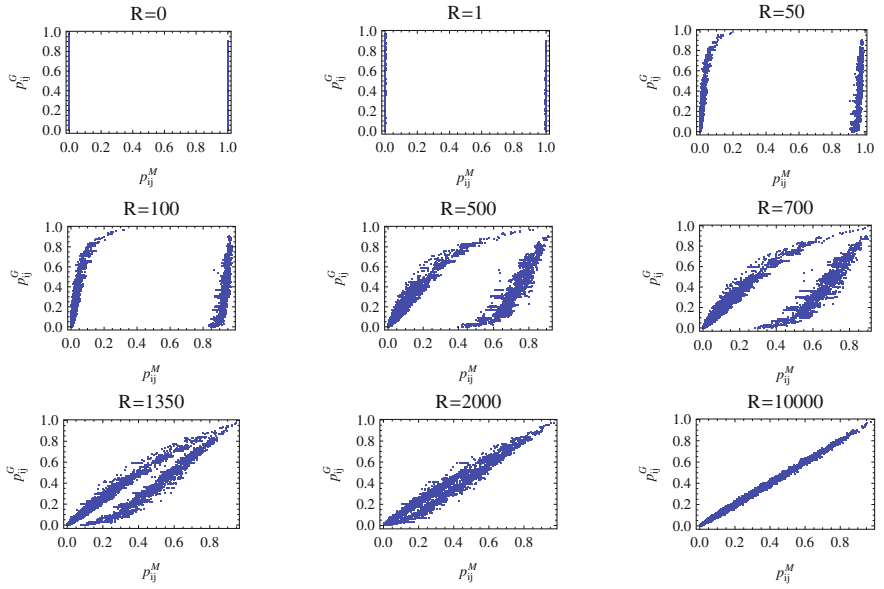

Fig. 5.1 Convergence of the microcanonical connection probability p_{ij}^{mic} (measured using the local rewiring algorithm described in Sect. 2.1.2) to the canonical probability p_{ij}^{can} (obtained using our maximum-entropy method) as the number R of local rewiring moves per network increases. The network being used is the Little Rock Lake food web. From [17]

canonical approach samples the configuration space iteratively and the microcanonical probability p_{ij}^{mic} can only be evaluated as the frequency of occurrence of the link over many randomizations.[1] As the number of randomized networks increases, this frequency converges to p_{ij}^{mic}. However this asymptotic value will also depend on the number R of elementary rewiring steps used to obtain a single randomized network in the Local Rewiring Algorithm described in Sect. 2.1.2.

To see this, let us focus on the Little Rock Lake network and consider the trivial case $R = 0$. As no rewiring takes place, all the 'randomized' networks will in fact coincide with the original network. If the adjacency matrix of the latter has elements $\{a_{ij}\}$, this means that $p_{ij}^{\text{mic}} = a_{ij}$. If R is nonzero but still very small, p_{ij}^{mic} will not change substantially. Only if R is large enough then p_{ij}^{mic} will approach p_{ij}^{can}. This is shown explicitly in Fig. 5.1, where we plot p_{ij}^{can} as a function of p_{ij}^{mic} for all pairs of vertices (i, j) by taking the *Little Rock Lake* food web (see Chap. 3 for a description) as the starting network. As R increases from $R = 0$ to $R = 10000$, the double-peaked shape (corresponding to $p_{ij}^{\text{mic}} = a_{ij}$ independently of p_{ij}^{can}) evolves towards the identity $p_{ij}^{\text{mic}} = p_{ij}^{\text{can}}$.

[1] Since, in this particular case, the low value of the network density guarantees the LRA not to be biased, we can safely use it to generate several randomized versions of the actual network structure.

Figure (5.1) shows that the canonical approach allows the values p_{ij}^{can}, to which the microcanonical p_{ij}^{mic} will converge only after several iterations, to be obtained 'from the beginning'. Notably, the number R of rewiring steps required for p_{ij}^{mic} to converge to p_{ij}^{can} acceptably is not known a priori and without the knowledge of p_{ij}^{can} itself. This problematic aspect of the microcanonical approach highlights another advantage of the canonical one.

5.3.2 Fluctuating Constraints

The marginal equivalence shown in the previous section has interesting consequences. One concerns the expected value of functions which show a *linear* dependence on our random variables. Since the canonical and microcanonical expectations of the latter coincide, the expected values of linear functions coincide as well. A simple example is provided by the total number of links, for which

$$\langle L \rangle = \sum_{i<j} \langle a_{ij} \rangle = \sum_{i<j} p_{ij}^{\text{can}} = \sum_{i<j} p_{ij}^{\text{mic}} = \overline{L} = L \tag{5.15}$$

(with obvious meaning of the symbols—$\overline{L}$ indicates the arithmetic mean of L taken over the microcanonical ensemble). Another relevant example is provided by degrees. As in the previous case,

$$\langle k_i \rangle = \sum_{j \neq i} \langle a_{ij} \rangle = \sum_{j \neq i} p_{ij}^{\text{can}} = \sum_{j \neq i} p_{ij}^{\text{mic}} = \overline{k_i} = k_i \quad \forall\, i. \tag{5.16}$$

Although the expected value of the constraints defining the two ensembles coincide, their fluctuations do not. While, in fact, constraints fluctuations do vanish in the microcanonical case, this is no longer true in the canonical case, where fluctuations differ from zero: standard deviations of links and degrees, in fact, read $\sigma[L] = \sqrt{M\lambda^*(\lambda^* - 1)}$ and $\sigma[k_i] = \sqrt{\sum_{j \neq i} p_{ij}^{\text{can}}(1 - p_{ij}^{\text{can}})}$, respectively.

It is instructive to recall that a fundamental quantity in statistical physics, the total energy, is characterized by (relative) fluctuations that vanish in the thermodynamic limit. In fact

$$\frac{\sigma[E]}{E} \propto \frac{1}{\sqrt{N}} \tag{5.17}$$

a ratio that tends to zero as the number of particles constituting the system tends to infinity. On the other hand, network models exist predicting values for the nodes degrees which remain finite even when the size of the network under analysis tends to infinity. As an example, let us consider the model defined by $p_{ij} = p \simeq c/N$ with c being a constant: the relative fluctuation of any degree reads, in this case,

$$\frac{\sigma[k_i]}{k_i} \simeq \frac{1}{c} \tag{5.18}$$

which does not vanish as the size of the system increases. Since the claim that the microcanonical and the canonical ensembles are equivalent rests upon the asymptotically vanishing value of (relative) fluctuations, the evidence that network models exist for which this is not true indicates that any trace of (non)equivalence has to be encoded into the constraints fluctuations.

In the general case, the formula for computing the relative fluctuations of node degrees read

$$\frac{\sigma[k_i]}{k_i} = \sqrt{\frac{1}{k_i} - \frac{\sum_{i \neq j} p_{ij}^2}{(\sum_{i \neq j} p_{ij})^2}} \tag{5.19}$$

and it is possible to derive an analogous expression for the relative fluctuations of node strengths. These calculations have been carried out in [27] for the Binary Configuration Model, the Weighed Configuration Model and the Enhanced Configuration Model. Upper or lower bounds for the relative fluctuations have also been obtained. The behaviour of these quantities is illustrated in Fig. 5.2 by fitting the models to a couple of real-world networks.

5.3.3 Measure Equivalence

The fact that the microcanonical and canonical marginal probabilities assume progressively closer values does not represent a sufficient condition for concluding that the microcanonical and canonical ensemble distributions are equivalent. In fact, marginal probabilities do not contain any information about the dependence between different pairs of vertices. While in the canonical approach these correlations are absent, and different pairs of vertices are always statistically independent, in the microcanonical approach these correlations arise from the microcanonical constraint of matching the degree sequence (or other contraints) exactly.

In order to measure the difference between microcanonical and canonical probabilities in a way that incorporates all possible dependencies between pairs of nodes, we use the KL divergence, which allows the *whole* amount of information encoded into the two aforementioned distributions to be compared at once. In the rest of this chapter, we mainly follow [26] and explore the idea that the microcanonical and canonical ensembles are equivalent if and only if their specific Kullback-Leibler (KL) divergence is zero. Since we are considering discrete systems, we need to use the following discrete version of the KL divergence:

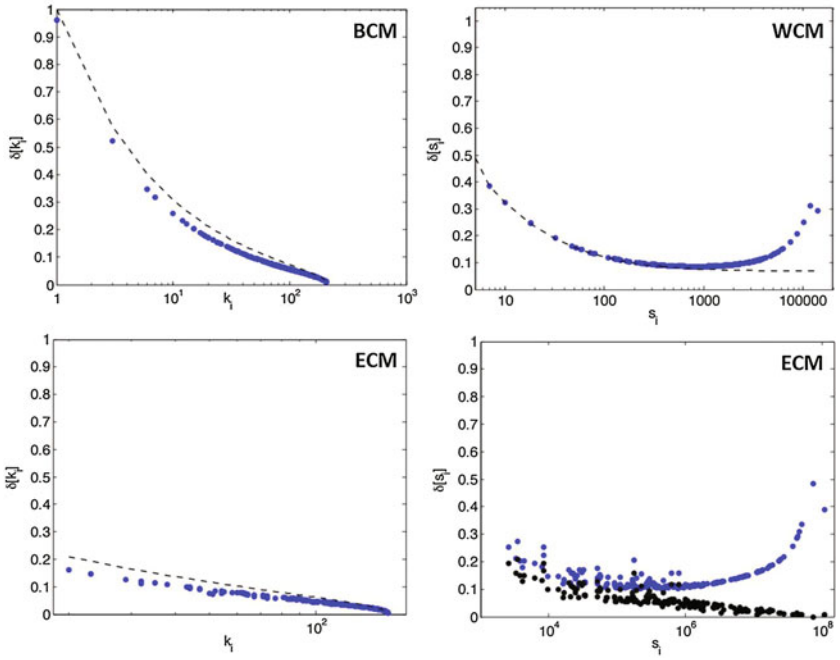

Fig. 5.2 Top left panel: relative fluctuations of the degree of each node in the Binary Configuration Model with degree sequence taken from the network of liquidity reserves exchanges between Italian banks in year 1999. The blue points are the exact values in Eq. (5.19), while the dashed curve is an upper bound (the lower bound is the abscissa). Top right panel: relative fluctuations of the strength of each node in the Weighted Configuration Model with strength sequence taken from the same network. The blue points are again the exact values, given by a formula analogous to Eq. (5.19), while the dashed curve is now a lower bound (the upper bound exceeds 1 and extends beyond the region shown). Bottom panels: relative fluctuations of the degree (left) and strength (right) of each node in the Enhanced Configuration Model with degree and strength sequences taken from the World Trade Web in year 2000. The blue points indicate the exact values, while the dashed curve (left) is an upper bound and the black points (right) are a reference value, typically exceeded. Adapted from [27]

$$S_N(P_{\text{mic}}||P_{\text{can}}) \equiv \sum_{\mathbf{G}} P_{\text{mic}}(\mathbf{G}) \ln \frac{P_{\text{mic}}(\mathbf{G})}{P_{\text{can}}(\mathbf{G})}. \tag{5.20}$$

The KL divergence is an information-theoretic quantity measuring how much of the information contained in the microcanonical ensemble is lost if we use the canonical one instead. The information loss can be understood in terms of the higher uncertainty encoded in the canonical distribution, where the value of the constraints is not sharp. Also note that, for the models we are considering, the canonical probability is the product of the marginal single-edge probabilities considered so far, i.e. is the product of the marginals of the microcanonical probability. Therefore the above KL divergence also coincides with the mutual information of the edges in the

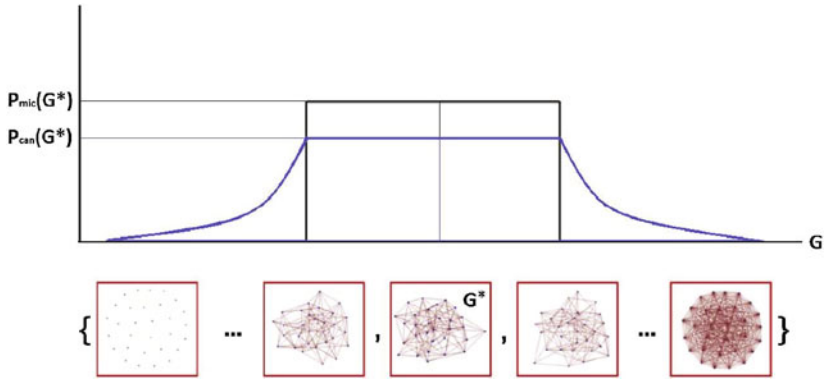

Fig. 5.3 Equation (5.25) shows that, in discrete systems, ensemble equivalence reduces to an extremely simple, local condition involving the ratio $P^*_{\rm mic}/P^*_{\rm can}$ of the microcanonical and canonical probabilities of a single 'microcanonical' configuration $\mathbf{G}^*$ (any one of those realizing the constraints sharply)

microcanonical ensemble. It precisely quantifies the microcanonical dependencies among different edges of the network. The independence of all edges is therefore the signature property of the canonical ensemble.

Note that, for finite systems, canonical and microcanonical probabilities are always different, and the KL divergence is always strictly positive. We therefore define the relative entropy *density*

$$s \equiv \lim_{N\to\infty} \frac{S_N(P_{\rm mic}||P_{\rm can})}{N} \tag{5.21}$$

and, following [15], we say that the ensembles are (measure) equivalent if and only if

$$s = 0. \tag{5.22}$$

Since $S_N(P_{\rm mic}||P_{\rm can})$ cannot be negative, the ensembles are necessarily nonequivalent if and only if

$$s > 0. \tag{5.23}$$

We now make a simple but crucial observation. From Eq. (5.3), and noting from the form of $H(\mathbf{G},\boldsymbol{\theta})$ that $P_{\rm can}(\mathbf{G}_1) = P_{\rm can}(\mathbf{G}_2)$ if $\mathbf{C}(\mathbf{G}_1) = \mathbf{C}(\mathbf{G}_2)$ (the canonical probability is the same for all configurations having the same values of the constraints), we rewrite Eq. (5.20) as

$$S_N(P_{\rm mic}||P_{\rm can}) = \ln \frac{P_{\rm mic}(\mathbf{G}^*)}{P_{\rm can}(\mathbf{G}^*)} = \ln \frac{P^*_{\rm mic}}{P^*_{\rm can}} \tag{5.24}$$

where $\mathbf{G}^*$ is *any* configuration in $\mathcal{G}_N$ such that $\mathbf{C}(\mathbf{G}^*) = \mathbf{C}^*$. The condition for equivalence, Eq. (5.22), then becomes

$$\lim_{N\to\infty} \frac{1}{N} \ln \frac{P^*_{\text{mic}}}{P^*_{\text{can}}} = 0 \tag{5.25}$$

or equivalently,

$$\lim_{N\to\infty} \frac{1}{N} \ln P^*_{\text{mic}} = \lim_{N\to\infty} \frac{1}{N} \ln P^*_{\text{can}} \tag{5.26}$$

Equation (5.25) shows that, in discrete systems, ensemble equivalence reduces to an extremely simple, local condition involving the ratio of the microcanonical and canonical probabilities of a single 'microcanonical' configuration $\mathbf{G}^*$ (any one of those realizing the constraints sharply). Equation (5.26) shows that the breaking of ensemble equivalence concides with $P_{\text{mic}}(\mathbf{G}^*)$ and $P_{\text{can}}(\mathbf{G}^*)$ obeying different large deviations properties [16]. Again, it is striking that this criterion involves only a single configuration $\mathbf{G}^*$. Besides its theoretical importance, this result also greatly simplifies the mathematical calculations.

5.4 Breaking of Equivalence Between Ensembles

Let us now use the above definitions in order to quantify the degree of (non) equivalence in the specific examples introduced in the previous section. The first model was an ensemble of simple undirected graphs with a fixed number of links, i.e. $\mathbf{C} \equiv L$, whose microcanonical and canonical ensembles are determined by Eqs. (5.4) and (5.5) respectively.

We can now compute the relative entropy

$$\begin{aligned} S(P_{\text{mic}}||P_{\text{can}}) &= -\ln \binom{M}{\lambda^* M} \\ &\quad - \lambda^* M \ln \lambda^* - (1-\lambda^*) M \ln(1-\lambda^*) \\ &= \ln \sqrt{2\pi M \lambda^*(1-\lambda^*)} + O(1/M), \end{aligned} \tag{5.27}$$

where we have used Stirling's formula $n! \approx \sqrt{2\pi n}(n/e)^n[1 + O(1/n)]$ [22]. We then get

$$s = \lim_{N\to\infty} \frac{\ln \sqrt{2\pi M \lambda^*(1-\lambda^*)}}{N} = 0, \tag{5.28}$$

which proves ensemble equivalence in this case.

For the next examples we consider four ensembles of networks with given degree sequence $\mathbf{C} \equiv \mathbf{k} = (k_1, k_2 \ldots k_N)$, i.e. three specific cases of the binary configuration

model considered throughout the previous chapters. Since the exact microcanonical number $\Omega_{\mathbf{k}}$ is unknown in general, we have to consider specific cases.

Our first three examples are in the sparse regime where the maximum degree behaves as

$$k_{max} = o(\sqrt{N}) \quad \text{with} \quad N \ll 2L, \tag{5.29}$$

where the inequality means that $\overline{k} \gg 1$. Using the combinatorial results valid in this regime [23, 24], we find

$$\Omega_{\mathbf{k}^*} = \frac{\sqrt{2}(\frac{2L}{e})^L}{\prod_{i=1}^n k_i^*!} \cdot e^{-f(\mathbf{k}^*)+\frac{1}{4}+o\left(\frac{(\overline{k^*})^3}{N}\right)}, \tag{5.30}$$

where $\overline{k} = \sum_{i=1}^N k_i/N$ (average degree), $L = N\overline{k}/2$ (number of links), $f(\mathbf{k}) \equiv (\overline{k^2}/2\overline{k})^2$ and $\overline{k^2} = \sum_{i=1}^N k_i^2/N$. On the other hand, as described in Sect. 2.2.2, the canonical ensemble is defined by the N equations [17]

$$\sum_{j\neq i} \frac{e^{-\theta_i^*-\theta_j^*}}{1+e^{-\theta_i^*-\theta_j^*}} = k_i^* \quad \forall i. \tag{5.31}$$

Setting $p_{ij}^* \equiv e^{-\theta_i^*-\theta_j^*}/(1+e^{-\theta_i^*-\theta_j^*})$, the canonical probability of the CM reads

$$P_{\text{can}}(\mathbf{G}) = \prod_{i=1}^N \prod_{j<i} (p_{ij}^*)^{g_{ij}} (1-p_{ij}^*)^{1-g_{ij}} \tag{5.32}$$

where g_{ij} is the i, j entry of the adjacency matrix of the graph $\mathbf{G}$. Equation (5.29) ensures that $k_{max} \ll \sqrt{2L}$, a condition under which Eq. (5.31) is solved by [17]

$$p_{ij}^* = \frac{e^{-\theta_i^*-\theta_j^*}}{1+e^{-\theta_i^*-\theta_j^*}} \approx e^{-\theta_i^*-\theta_j^*} = \frac{k_i^* k_j^*}{2L^*} \ll 1. \tag{5.33}$$

Note that this implies $\theta_i^* = -\ln \frac{k_i^*}{\sqrt{2L^*}}$ and $\ln(1-p_{ij}^*) \approx -p_{ij}^*$. Using these expressions, Eq. (5.32) leads to

$$\begin{aligned}\ln P_{\rm can}(\mathbf{G}^*) &= \sum_{i=1}^{N}\sum_{j<i}[g_{ij}^*\ln p_{ij}^* + (1-g_{ij}^*)\ln(1-p_{ij}^*)]\\ &= -\sum_{i=1}^{N}\sum_{j<i}[g_{ij}^*(\theta_i^*+\theta_j^*) + \ln(1-p_{ij}^*)]\\ &\approx -\sum_{i=1}^{N}k_i^*\theta_i^* - \sum_{i=1}^{N}\sum_{j<i}p_{ij}^*\\ &= \sum_{i=1}^{N}k_i^*\ln k_i^* - L^*\ln(2L^*) - L^*. \end{aligned} \tag{5.34}$$

Combining Eqs. (5.30) and (5.34), we obtain

$$\begin{aligned}S(P_{\rm mic}||P_{\rm can}) &= -L^*\ln(2L^*) + L^* + \sum_{i=1}^{N}\ln(k_i^*!)\\ &\quad + f(\mathbf{k}^*) - 1/4 + o\left((\overline{k^*})^3/N\right)\\ &\quad - \sum_{i=1}^{N}k_i^*\ln k_i^* + L^*\ln(2L^*) + L^*\\ &= \sum_{i=1}^{N}[\ln(k_i^*!) - k_i^*\ln k_i^* + k_i^*]\\ &\quad + f(\mathbf{k}^*) - 1/4 + o\left((\overline{k^*})^3/N\right). \end{aligned} \tag{5.35}$$

Defining $\xi \equiv \lim_{N\to\infty} f(\mathbf{k}^*)/N$, and using bars to denote averages over nodes (intended as a limit), we finally get

$$s = \overline{\ln(k^*!)} - \overline{k^*\ln k^*} + \overline{k^*} + \xi \geq \overline{\ln\sqrt{2\pi k^*}} + \xi \tag{5.36}$$

where the inequality is valid for every $k^* \geq 1$ [22]. Since $\xi \geq 0$, the above expression is strictly positive, proving that in this case the microcanonical and canonical ensembles are, more surprisingly, *not* equivalent.

Our first example of graphs with given degree sequence is the class of k^*-regular graphs (where every node has the same degree $k_i = k^*$) with $k^* = o(\sqrt{N})$. In this case $\xi = 0$ and $\overline{\ln k^*} = \ln k^*$. If k^* is finite, Eq. (5.36) becomes

$$s = \ln(k^*!) - k^*\ln k^* + k^* \geq \ln\sqrt{2\pi k^*}, \tag{5.37}$$

while, if k^* grows with N, s diverges like $\ln k^*$. The non-equivalence of regular graphs can be related to the fact that the canonical versions of the two ensembles considered so far, defined by $\mathbf{C} = L$ and $\mathbf{C} = \mathbf{k}$ respectively, coincide via the identification

$p^* = L^*/M = k^*/(N-1)$. Since the two microcanonical ensembles are different, while the canonical ones are the same, only one (if any) of the two cases can show ensemble equivalence. As we proved ensemble equivalence for $\mathbf{C} = L$, there cannot be equivalence for $\mathbf{C} = \mathbf{k}$ as well.

Our second example of networks with given degree sequence is the important class of sparse 'uncorrelated' scale-free networks [25], defined by a truncated power-law degree distribution of the form $F(k) = Ak^{-\gamma}$ for $1 \leq k \leq k_c$ (with $2 < \gamma < 3$ like in most real-world networks) and $F(k) = 0$ otherwise. The cut-off $k_c \sim \sqrt{N}$ [25] is needed to ensure that the maximum degree k^*_{max} does not exceed the bounds in Eq. (5.29), so that Eq. (5.33) is valid. Approximating $F(k)$ with a continuous distribution, it is easy to see that normalization implies $A = (\gamma-1)/(1-k_c^{1-\gamma})$. Calculating $\overline{k^a}$ as $\lim_{N\to\infty} \int_1^{k_c} dk F(k) k^a$, we get

$$f(\mathbf{k}^*) = \left(\frac{\gamma-2}{2(3-\gamma)} \cdot \frac{k_c^{3-\gamma}-1}{1-k_c^{2-\gamma}} \right)^2 \sim N^{3-\gamma}, \tag{5.38}$$

which leads to $\xi = 0$. Similarly, it is easy to show that

$$\overline{\ln k^*} = \lim_{N\to\infty} \int_1^{k_c} dk F(k) \ln k = \frac{1}{\gamma-1}. \tag{5.39}$$

Thus Eq. (5.36) becomes in this case

$$s \geq \overline{\ln \sqrt{2\pi k^*}} = \frac{1}{2(\gamma-1)} + \ln \sqrt{2\pi} \tag{5.40}$$

showing that the canonical and microcanonical ensembles of uncorrelated scale-free networks are not equivalent.

Our third example is the 'sparsest' graph model introduced in Sect. 5.3.1. Upon re-expressing its microcanonical probability as

$$P_{\text{mic}}(\mathbf{G}^*) = \prod_{i=0}^{N-1} \frac{1}{N-2i+1} = \frac{(N-1)!!}{N!} \tag{5.41}$$

the relative entropy is

$$S(P_{\text{mic}}||P_{\text{can}}) = \ln \left[\frac{(N-1)^{\binom{N}{2}} (N-1)!!}{N!(N-2)^{\binom{N}{2}-\frac{N}{2}}} \right] \tag{5.42}$$

and the KL divergence reads

$$s = 1. \tag{5.43}$$

Our fourth and last example of graphs with given degree sequence is the class of h-star graphs considered in Sect. 5.3.1, for which Eq. (5.29) is violated and therefore Eqs. (5.30) and (5.36) cannot be used. The microcanonical number of such graphs $\Omega_{\mathbf{k}^*} = \frac{(N-h)!}{(\frac{N-h}{h}!)^h}$ can thus be approximated by applying Stirling's approximation:

$$P_{\text{mic}}(\mathbf{G}^*) = h^{h-N}\,[2\pi(N-h)]^{\frac{1}{2}(k-1)}\,h^{-\frac{h}{2}}. \tag{5.44}$$

The canonical probability, instead, reads

$$P_{\text{can}}(\mathbf{G}^*) = \left(\frac{1}{h}\right)^{N-h}\left(1-\frac{1}{h}\right)^{(h-1)(N-h)}, \tag{5.45}$$

the relative entropy is

$$\begin{aligned} S(P_{\text{mic}}||P_{\text{can}}) &= (N-h)(h-1)\ln\left(\frac{h}{h-1}\right) \\ &\quad + (k-1)\ln\sqrt{2\pi(N-h)} - h\ln\sqrt{h}, \end{aligned} \tag{5.46}$$

and the specific KL divergence is

$$s = (h-1)\ln\left(\frac{h}{h-1}\right). \tag{5.47}$$

So, although microcanonical marginals match, we again find nonequivalence for all $h > 1$ (in the 'deterministic' case $h = 1$, both ensembles admit only one star-like configuration and are therefore equivalent).

The proof of the breakdown of ensemble equivalence in graphs with given degree sequence provides a theoretical explanation for some otherwise anomalous phenomena that have been recently observed, namely the fact that the canonical and microcanonical entropies of random regular graphs are different even in the thermodynamic limit [19] and the non-vanishing of canonical fluctuations in the configuration model [27]. Moreover, these results show that, while it is generally believed that ensemble nonequivalence is associated with long-range interactions, it can naturally arise also in systems with multiple (local) constraints.The proof of ensemble nonequivalence has been recently extended to random with given degrees and arbitrary modular structure [28].

5.5 Implications of (non)equivalence for Combinatorics

Looking at Eq. (5.2), we note that Eq. (5.24) can be reformulated as

$$S_N(P_{\text{mic}}||P_{\text{can}}) = \ln \frac{P^*_{\text{mic}}}{P^*_{\text{can}}} = \ln \frac{\Omega^{-1}_{\mathbf{C}^*}}{P^*_{\text{can}}} = -\ln Q_{\text{can}}(\mathbf{C}^*), \tag{5.48}$$

which is an important result showing that the degree of nonequivalence is entirely determined by the probability that, in the canonical ensemble, the constraint $\mathbf{C}$ takes the particular microcanonical value $\mathbf{C}^*$. We also note that Eq. (5.21) can be rewritten as

$$S_N(P_{\text{mic}}||P_{\text{can}}) = sN + o(N), \tag{5.49}$$

where $o(N)$ indicates a quantity that, if divided by N, vanishes in the limit of infinite N. Combining Eqs. (5.48) and (5.49), we get

$$\Omega_{\mathbf{C}^*} = \frac{e^{-Ns-o(1/N)}}{P^*_{\text{can}}}. \tag{5.50}$$

The above equation is an important result. It connects the number of microcanonical configurations, which is the result of a combinatorial enumeration problem, to the relative entropy density between the canonical and microcanonical ensembles. From the above equation, it follows that, for constraints that lead to ensemble equivalence ($s = 0$), the number of microcanonical configurations can be enumerated as

$$\Omega_{\mathbf{C}^*} = \frac{e^{-o(1/N)}}{P^*_{\text{can}}}. \tag{5.51}$$

Importantly, P^*_{can} can be calculated exactly using the method we described in Sect. 2.2, so the above formula is powerful for equivalent ensembles. On the other hand, for constraints that lead to nonequivalence ($s > 0$), the full expression (5.50) applies. For these systems, one should be warned that neglecting the effects of ensemble nonequivalence may naively result in the inappropriate use of Eq. (5.51). This would determine an error in the extimation of $\Omega_{\mathbf{C}^*}$ that is exponential in N.

One can check that Eq. (5.50) is verified in the previous examples where P^*_{mic}, and hence $\Omega_{\mathbf{C}^*}$, has been calculated explicitly. In particular, the Erdős-Rènyi model obeys Eq. (5.51), whereas the 'sparsest graph' and 'core-periphery' models obey Eq. (5.50).

In general, the above findings contribute to the theoretical understanding of nonequivalence in discrete systems, by showing that it can manifest itself in combinatorial enumeration problems and that it is directly related to a non-vanishing difference between the canonical and microcanonical large deviation properties of a single microstate.

5.6 *"What Then Shall We Choose?"* Hardness or Softness?

We provided evidence that, for graphs with local node-specific constraints, the canonical and microcanonical ensembles are not equivalent. This result implies that choosing between microcanonical and canonical approaches to the sampling of network ensembles is not only a matter of (computational) convenience, but also a theoretical issue that should be addressed more formally. To this end, we recall that microcanonical ensembles describe isolated systems that do not interact with an external 'heat bath' or 'reservoir'. In ordinary statistical physics, this means that there is no exchange of energy with the external world. In our setting, this means that microcanonical approaches do not contemplate the possibility that the network interacts with some external 'source of error', i.e. that the value of the enforced constraints might be affected by errors or missing entries in the data. When present, such errors (e.g. a missing link, implying a wrong value of the degree of two nodes) are propagated to the entire collection of randomized networks, with the result that the 'correct' network is not included in the microcanonical collection of graphs on which inference is being made.

By contrast, besides being unbiased and mathematically tractable, the canonical approach is also the most appropriate choice if one wants to account for possible errors in the data, since canonical ensembles appropriately describe systems in contact with an external reservoir (source of errors) affecting the value of the constraints. While in presence of even small errors microcanonical methods assign zero probability to the 'uncorrupted' configuration and to all the configurations with the same value of the constraints, the canonical method assigns these configurations a probability which is only slightly smaller than the (maximum) probability assigned to the set of configurations consistent with the observed ('corrupted') one. These considerations suggest that, given its simplicity, elegance, and ability to deal with potential errors in the data, the use of the canonical ensemble should be preferred to that of the microcanonical one.

References

1. J.H. van Lint, R.M. Wilson, *A course in combinatorics* (Cambridge university press, 2001)
2. J.M. Harris, J.L. Hirst, M.J. Mossinghoff, *Combinatorics and graph theory* (Springer, New York, 2008)
3. L. Boltzmann, Wiener Berichte **2**(76), 373–435 (1877)
4. J.W. Gibbs, *Elementary Principles of Statistical Mechanics* (Yale University Press, New Haven, 1902 - reprinted by Dover, New York, 1960)
5. R.S. Ellis, K. Haven, B. Turkington, J. Stat. Phys. **101**, 999 (2000)
6. R.S. Ellis, K. Haven, B. Turkington, Nonlinearity **15**, 239 (2002)
7. M. Blume, V.J. Emery, R.B. Griffiths, Phys. Rev. A **4**, 1071 (1971)
8. J. Barré, D. Mukamel, S. Ruffo, Phys. Rev. Lett. **87**, 030601 (2001)
9. R.S. Ellis, H. Touchette, B. Turkington, Physica A **335**, 518 (2004)
10. D. Lynden-Bell, Physica A **263**, 293 (1999)
11. P.-H. Chavanis, Astron. & Astrophys. **401**, 15 (2003)

12. M. D'Agostino, Phys. Lett. B **473**, 219 (2000)
13. J. Barré, B. Goncalves, Physica A **386**, 212–218 (2007)
14. H. Touchette, R.S. Ellis, B. Turkington, Physica A **340**, 138–146 (2004)
15. H. Touchette, Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels. J. Stat. Phys. **159**, 987–1016 (2015), arXiv:1403.6608
16. F. den Hollander, *Large Deviations, Fields Institute Monographs 14* (American Mathematical Society, Providence RI, 2000)
17. T. Squartini, D. Garlaschelli, New Journal of Physics **13**, 083001 (2011)
18. J. Park, M.E.J. Newman, Phys. Rev. E **70**, 066117 (2004)
19. K. Anand, G. Bianconi, Physical Review E **80**(4), 045102 (2009)
20. E.T. Jaynes, Phys. Rev. **106**, 4 (1957)
21. D. Garlaschelli, M.I. Loffredo, Phys. Rev. E **78**, 015101(R) (2008)
22. H. Robbins, Amer. Math. Monthly **62**, 26–29 (1955)
23. E.A. Bender, Discrete Math. **10**, 217–223 (1974)
24. B.D. McKay, N.C. Wormald, Combinatorica **11**, 369–382 (1991)
25. M. Boguñá, R. Pastor-Satorras, A. Vespignani, Eur. Phys. J. B **38**, 205–209 (2004)
26. T. Squartini, J. de Mol, F. den Hollander, D. Garlaschelli, Breaking of ensemble equivalence in networks. Phys. Rev. Lett. **115**, 268701 (2015)
27. T. Squartini, R. Mastrandrea, D. Garlaschelli, New J. Phys. **17**, 023052 (2015)
28. D. Garlaschelli, F. den Hollander, A. Roccaverde, Ensemble nonequivalence in random graphs with modular structure. Journal of Physics A: Mathematical and Theoretical **50**, 015001 (2017)

Chapter 6
Concluding Remarks

It lay thickly drifted on the crooked crosses and headstones, on the spears of the little gate, on the barren thorns. His soul swooned slowly as he heard the snow falling faintly through the universe and faintly falling, like the descent of their last end, upon all the living and the dead.

– James Joyce, *The Dubliners*

Now that we have reached the end of this book, we can look at its main contents in retrospect and try and make some overarching summary and remarks. Here we take the liberty of offering some personal perspective and take-home message, mostly based on our own experience with the scientific content of the book.

Throughout the book we have focused on the construction and use of ensembles of networks with given topological properties. We have discussed some general motivations, theoretical foundations, and practical applications. We have particularly emphasized three different challenges that, while being apparently unrelated, turn out to share common underlying concepts and are amenable to a unified treatment within the same theoretical framework. These practical applications are the problems of *pattern detection*, i.e. the identification of structural features that cannot be simply traced back to the properties of individual nodes in a real network (Chap. 3), *network reconstruction*, i.e. the probabilistic inference of the whole structure of a real-world network from partial node-specific information (Chap. 4), and *graph combinatorics*, i.e. certain combinatorial operations such as the sampling and enumeration of graphs with given local topological properties (Chap. 5).

In Chap. 2 we have introduced and adopted a specific methodology to construct the graph ensembles needed to address these three challenges, namely the principle of maximum entropy. As in many other applications in different branches of science, most prominently in statistical physics [1], the principle prescribes that the most unbiased (i.e. maximally random) expectation about the microscopic configuration (i.e. the state of all fundamental units) of a system is achieved by the probability dis-

T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-69438-2_6

tribution that maximizes Shannon's entropy, subject to a specified set of macroscopic constraints (e.g. the total energy). In the case of interest for this book, the principle translates into the construction of maximum-entropy ensembles of networks with given topological constraints. The resulting methodology was first proposed in a seminal paper [2] where the authors of this book combined the abstract formalism of maximum-entropy graph ensembles [1, 3, 4], also known as *exponential random graphs* in the social science literature [5, 6], with the technique of exact likelihood maximization [7]. This combination is a crucial step that allows significant progress towards the applications considered in this book.

Indeed, without likelihood maximization, the abstract use of maximum-entropy graph ensembles would result in treating the Lagrange multipliers as free parameters to be drawn from suitable probability distributions inducing certain classes of network topologies (e.g. regular graphs, scale-free networks, etc.) [8]. This approach would have not allowed us to calibrate the maximum-entropy model to a specific empirical network (or partial information about it) and would have therefore been of limited use in addressing the three challenges discussed here.

Similarly, the use of exponential random graphs in the social science literature mainly relies on approximate techniques, such as Markov Chain Monte Carlo [5, 6], that avoid the exact calculation of the partition function and the consequent exact maximization of the likelihood. This approach would have prevented us from achieving another goal we set ourselves in this book: namely, the possibility of calculating *analytically* the probability distribution over graphs in the ensemble and the resulting expectation values and higher moments of the quantities of interest. The analytic control over the ensemble conferred by our exact approach not only reduces the computational complexity of the calculation of the expected properties dramatically, but also allows us to carry out otherwise impossible operations, such as the canonical sampling of graphs with given properties, as described in Chap. 5.

In Chap. 3 we have discussed the application of the maximum-entropy method to the problem of pattern detection. Historically, this was the main initial motivation for the introduction of our method in the original paper [2]. Since pattern detection is the identification of nontrivial structural properties in a real-world network, it requires a comparison of such network with a suitable null model used as a reference or benchmark. As we have discussed, the latter is chosen to be a random graph model that preserves the local topological properties (i.e. the degrees and/or the strengths) of the real network and is otherwise completely random.

We have indeed illustrated the usefulness of maximum-entropy ensembles of graphs with local constraints in identifying a variety of nontrivial structural patterns in several real networks. These patterns include assortativity, clustering, reciprocity, and more in general all possible dyadic and triadic motifs. We have considered different types of graphs, i.e. binary, weighted, directed, and undirected graphs, and the corresponding maximum-entropy ensembles.

We have emphasized the importance of such ensembles being accurately tailored on the real-world network: depending on the particular values of the empirical local properties, i.e. the observed degrees or strengths of nodes, certain trends that would be superficially classified as assortative, disassortative or clustered may turn out to

actually deserve a different—sometimes opposite—interpretation. Even properties that are believed to be characteristic of the maximum-entropy ensembles themselves, e.g. the expectation that ensembles of unipartite graphs with given degrees would naturally exhibit a decreasing (disassortative) relationship between the average degree of the neighbours of a node and the degree of the node itself [8], are in some cases found to be completely reversed when the local node properties are taken as input from an empirical network, rather than sampled from an ad hoc distribution.

From a practical point of view, the use of our (canonical) analytical method allows a fast and exact calculation of expectation values and statistical deviations, such as the z-scores, that would otherwise require the numerical sampling of many graphs from the (microcanonical) ensemble of graphs with given constraints. The numerical techniques proposed to perform such a sampling, which include the popular Local Rewiring Algorithm [9], turn out to be biased and may therefore lead to misleading results.

In Chap. 4 we have discussed the application of the maximum-entropy method to the problem of network reconstruction. Network reconstruction employs purely local topological information to infer the higher-order structural properties of a real-world network [10] This problem arises whenever the complete structure of a network is not known (for instance, due to confidentiality or privacy issues), but local properties are. An example relevant for the epidemiology of sexually transmitted diseases is the network of sexual contacts among people, for which only aggregate information (the total number of contacts with different partners) can be typically surveyed in a population. In such cases, optimal inference about the network can be achieved by maximising the entropy subject to the known (local) constraints, which again leads to the ensembles with fixed degrees considered here.

In Chap. 5 we have used the canonical maximum-entropy method as a 'soft' alternative for 'hard' combinatorial problems. We have found that the canonical and microcanonical ensembles of random graphs with local constraints are *not* equivalent. Besides that, we have quantified the degree of nonequivalence in terms of the relative entropy between the two associated probability distribution. We have finally related the asymptotic solution to hard enumeration problems to (computable) properties of the dual soft ensemble and to the relative entropy itself. These results provide new perspectives into a probabilistic route to the solution of combinatorial problems.

The aforementioned applications, along with similar ones, make use of random graphs with local constraints. Our proof of nonequivalence of the corresponding ensembles have the following important implication. While for ensemble-equivalent models it makes practically no difference whether a microcanonical or canonical implementation is applied to large networks, for nonequivalent models different choices of the ensemble lead to asymptotically different results. As a consequence, while for applications based on ensemble-equivalent models the choice of the working ensemble can be arbitrary or be done on mathematical convenience (as usually done), for those based on nonequivalent models the choice should be principled, i.e., dictated by a theoretical criterion that indicates a priori which ensemble is the appropriate one.

Among the possible criteria, we stress again that the canonical ensemble represents the most appropriate choice whenever the available data are subject to (even small) errors, i.e. when the measured value of the constraint entering as input in the construction of the random graph ensemble is only the best available estimate for some unknown 'true' (error-free) value. In this situation, we want that possible small deviations of the constraint from its true value result in small deviations of the ensemble probability from the one we would have obtained using the true value. Using the microcanonical ensemble implies that even small deviations from the true value will imply a zero probability for the true graph and for all the configurations having the true value of the constraint. This means that even small initial measurement errors would severely bias the entire inference procedure. On the other hand, using the canonical ensemble implies that small deviations of the constraint from the true value imply only a slight difference in the constructed probability distribution. So, besides being the mathematically simpler option, we argue that canonical ensembles are also the most appropriate principled choice whenever one wants to account for possible uncertainties in the measured values of the constraints.

From a general point of view one should note that, in contrast with the traditional case in statistical physics, the constraints imposed on the most important network ensembles considered here are *mesoscopic* rather than macroscopic. Indeed, examples of macroscopic constraints (analogous to the total energy of a physical system) are the total number of links $L(\mathbf{A}) = \sum_{i,j} a_{ij}$ and the total weight $W(\mathbf{W}) = \sum_{i,j} w_{ij}$, for binary and weighted graphs respectively. These quantities are defined as sums over all the pairs of nodes in the network. By contrast, the most important ensembles we have used in the book enforce the degree $k_i(\mathbf{A}) = \sum_j a_{ij}$ and/or the strength $s_i(\mathbf{W}) = \sum_j w_{ij}$ of each node i separately. These quantities are local sums over nodes and therefore stand at an intermediate, mesoscopic level in between the microscopic level of individual edges, which are the true "degrees of freedom" of the system, and the macroscopic level where all such degrees of freedom are aggregated into an overall total value. Enforcing mesoscopic, as opposed to macroscopic, constraints is the key requirement that allows the graph ensembles to accurately replicate the observed heterogeneity observed in real-world networks. At the same time, unlike traditional examples in statistical physics, this property also leads to the breaking of the equivalence between canonical and microcanonical ensembles of graphs, with the aforementioned far-reaching consequences.

References

1. E.T. Jaynes, Phys. Rev. **106**, 4 (1957)
2. T. Squartini, D. Garlaschelli, Analytical maximum-likelihood method to detect patterns in real networks. New J. Phys. **13**, 083001 (2011)
3. L. Boltzmann, Wiener Berichte **2**(76), 373–435 (1877)
4. J.W. Gibbs, Yale University Press, Yale, C.T. (1902)
5. P. Holland, S. Leinhardt, in "Sociological Methodology", D. Heise Ed., 1–45 (1975)
6. S. Wasserman, K. Faust, *Social Network Analysis* (Cambridge University Press, 1994)

7. D. Garlaschelli, M.I. Loffredo, Phys. Rev. E **78**, 015101(R) (2008)
8. J. Park, M.E.J. Newman, Phys. Rev. E **70**, 066117 (2004)
9. S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks. Science **296**, 910–913 (2002)
10. R. Mastrandrea, T. Squartini, G. Fagiolo, D. Garlaschelli, Enhanced reconstruction of weighted networks from strengths and degrees. New J. Phys. **16**, 043022 (2014)

Index

T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks*,
SpringerBriefs in Complexity, https://doi.org/10.1007/978-3-319-69438-2