Maximum entropy models for financial systems

Almog, A.

Maximum Entropy Models for Financial Systems

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op vrijdag 13 januari 2017
klokke 12.30 uur

door

Assaf Almog

geboren te Eilat (Israel)
in 1982

Promotores:	Prof. Dr. E.R. Eliel
	Prof. Dr. Ir. W. van Saarloos
Co-promotor:	Dr. D. Garlaschelli

Promotiecommissie:	Prof. Dr. H.E. Stanley (Boston University, USA)
	Prof. Dr. I. van Lelyveld (VU Amsterdam, Netherlands)
	Prof. Dr. W.Th.F. den Hollander
	Prof. Dr. J.H. Meijer
	Prof. Dr. J.M. van Ruitenbeek
	Dr. P.J.H. Denteneer

This work was supported by the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, the Netherlands) with funds from beneficiaries of Duyfken Trading Knowledge BV, Amsterdam, the Netherlands.

The cover shows a visualisation of the international trade network for the year 2011. The different colours represent densely connected clusters of countries detected by a community detection algorithm. The figure was created with VOSviewer (http://www.vosviewer.com/).

Casimir PhD series Delft-Leiden 2016-38
ISBN 978-90-8593-283-3

To Liam, Daan and Tomer

Contents

Introduction

Following the 2008 crisis, there has been a soaring interest in using ideas from different disciplines to make sense of economic and financial markets. The near-collapse of the financial system could not be explained, even more so predicted, by the traditional economic models. Ignoring properties like the complex network of interactions and non-linear relations in the system, these economic models are based on very restrictive assumptions. Confronting this gap, in recent years an alternative view has been developed using network theory and tools from statistical physics. A notable example of this shift of perspective is the move from traditional measures of “risk” of individual financial entities to new measures of “systemic risk,” defined as the risk of collapse of an entire system. Nevertheless, the pursuit of physicists to characterize financial and economic systems with empirical laws and “simple” global models started already two decades before, giving rise to the controversial field of ‘Econophysics.’ The interaction between economists and physicists, although immersed in frictions and resistance, led to some fruitful results and attracted major interest by central banks, regulators, and policy makers. Nonetheless, despite the various “real world” applications and implications, this field has to problems of a great social relevance, it requires the toolkit of theoretical physics and, in particular, statistical physics. This stimulating scientific context is partially reflected in the environment wherein the research described in this Ph.D. has been conducted. This study combined theoretical work and data analysis at the Lorentz Institute for Theoretical Physics, alongside important interactions with practitioners in finance (Duyfken Trading Knowledge) and bank supervisors (The Dutch National Bank).

Employing concepts from physics or mathematics in the field of economics is by no means a new phenomenon. In fact, Leiden University provides some remarkable examples of scientists with a background in physics, who left a significant impact on the field of economics. The most famous one is Jan Tinbergen, the first recipient of the Nobel Memorial Prize in Economics in 1969, who obtained his Ph.D. in physics at Leiden University under the supervision of Paul Ehrenfest in 1929. The title of his thesis was "Minimum Problems in Physics and Economics". In the early sixties, Tinbergen proposed the so-called Gravity Model of International Trade, presumably inspired by his physics training. The model predicts the bilateral trade flows between two countries by a formula similar to Newton's law of gravitation, and is still being used by economists. As we explain later, part of this thesis focuses on extensions of the Gravity Model. Another great example is Tjalling Koopmans, which was a student of Jan Tinbergen in 1933. In 1936, Koopmans graduated with a Ph.D. from the faculty of mathematical and physical sciences at Leiden University, with a thesis entitled "Linear regression analysis of economic time series". Time series analysis is another core topic that we address in this thesis. Later in 1975, Koopmans was awarded the Nobel Memorial Prize in Economics (jointly with Leonid Kantorovich) for his contributions to the field of resource allocation, specifically the theory of the optimal use of resources. Looking back at these great scientists from a modern standpoint, they highlight the advantages of interdisciplinary research in tackling major challenges.

Coming back to the present, the research of complexity in economics has been steadily growing, gradually encompassing different scales: from 'microscopic' networks of financial assets, through 'mesoscopic' networks of firms, banks, and institutions, to 'macroscopic' networks of countries and economic sectors. In general, the dynamics of these complex financial systems is highly random and noisy. Nevertheless, they carry critical information. The 'universal' challenge, across the different scales, is the extraction of meaningful information regarding the state of the system from the observable (empirical) data. This problem is immensely complicated by the temporal heterogeneity, i.e. different dynamics in different time periods, and the structural heterogeneity, i.e. complex topology, in the systems. Most current models are much more homogeneous and cannot account for, or explain these complex properties. This takes us to the main research question of the thesis, where we want to introduce a new class of statistical models which enforce partial empirical information that accurately controls for the heterogeneity in the system. A very promising approach to this problem is the use of maximum-entropy ensembles. Maximum-entropy models can be used in different settings, and typically serve as a reference to identify non-random patterns or properties. The power of the maximum-entropy approach is that it applies to very different fields and systems, from neuroscience to social network analysis. In this work, we review various maximum-entropy models and their powerful applications to financial systems. As a by-product, we also apply our framework to brain data. This was possible due to a collaboration with Leiden University Medical Center (LUMC).

The thesis is divided into three independent chapters, where each chapter covers results from multiple scientific publications addressing a particular system or problem. For a better comprehensibility, in each chapter we start by introducing the theoretical models and framework, and later proceed to the different applications of our models to real-world systems. In Chapter 1 we aim at characterizing and quantifying the information encoded within the so-called binary projections (i.e. the signs of the increments) of financial time series. We introduce maximum-entropy ensembles of binary matrices that represent projections of single and multiple binary time series, subject to a set of desired constraints defined as simple empirical observables. Our approach leads to a family of analytically solved null models that allow us to quantify the amount of information encoded in the chosen constraints, i.e. the selected observables of the binary projections of real-time series. Lastly, we show that our approach is able to reproduce and mathematically characterize certain empirical non-linear relationships between binary and non-binary properties of real time series.

In Chapter 2 we focus on economic networks, in particular, the International Trade Network (ITN). The network describes the exchange of capital, goods, and services between countries, and plays a significant role in the propagation of shocks. Modelling this complex system has been tackled by different disciplines, starting in the early sixties with the aforementioned Gravity Model. However, the empirical topology of the ITN is much more heterogeneous than the one predicted by the Gravity Model. The complete characterization of the ITN via a simple, yet accurate, model is still an open problem. We propose two different GDP-driven models which reconcile the different approaches of macroeconomics and network theory. Specifically, one model is a maximum-entropy generalization of the popular Gravity Model that embeds the latter in a realistic network topology. Thus, it represents significant improvement with respect to current models.

In Chapter 3 we discuss the identification of functional structure from correlation matrices measured from empirical time series driven by a common nonstationary trend. We discuss a recent community detection method and generalize it using a complete maximum-entropy framework that builds on the results from the previous chapter 1. In this setting, we introduce a null model serving as a random benchmark for the identification of non-random patterns in the correlation matrix. We apply the method to various real-world financial markets, examining the emergent functional structure generated by financial time series and their binary projections. We show that the simple binary representation can replicate to a large degree the complex structure which is induced by the full weighted time series. Next, we show that our method also has a great potential in a biological setting, specifically in an empirical detection of functional brain organization. In collaboration with LUMC, we apply the method to the biological clock of mice (suprachiasmatic nucleus). This is a very small brain region that can be represented as a complex network of oscillating neurons. While other methods failed, our method consistently revealed a core-periphery structure associated with two populations of neurons, a result that has been cross-checked with independent analysis.

Finally, we end this thesis with some concluding remarks, reviewing the key findings presented in this work.

Chapter 1

Financial Time Series

The dynamics of complex systems, from financial markets to the brain, can be monitored in terms of time series of activity of their fundamental elements (such as stocks or neurons respectively). While the main focus of time series analysis is on the magnitude of temporal increments, a significant piece of information is encoded into the binary projection (i.e. the sign) of such increments. In this chapter we provide further evidence of this by showing strong nonlinear relationships between binary and non-binary properties of financial time series. We then introduce an information-theoretic approach to the analysis of the binary signature of single and multiple time series. Through the definition of maximum-entropy ensembles of binary matrices, we quantify the information encoded into the simplest binary properties of real time series and identify the most informative property given a set of measurements. Our formalism is able to replicate the observed binary/non-binary relations very well, and to mathematically characterize them. Moreover, we identify distinct regimes in the collective behaviour of groups of stocks, corresponding to different levels of coordination that only depend on the average return of the binary time series. This approach also allows us to connect simple stochastic processes to specific ensembles of time series inferred from partial information.

The results presented in this chapter have been published in the following reference: A. Almog and D. Garlaschelli *New Journal of Physics* 16: 093015 (2014).

1.1 Introduction

In large systems, the observed dynamics or activity of each unit can be represented by a discrete time series providing a sequence of measurements of the state of that unit. One of the main challenges researchers are faced with is that of extracting meaningful information from the high-dimensional (multiple) time series characterizing all the elements of a complex system [1, 2, 3, 4, 5, 6, 7, 8, 9]. Traditionally, the main object of time series analysis is the characterization of patterns in the amplitude of the increments of the quantities of interest. Given a signal $s_i(t)$ where i denotes one of the N units of the system and t denotes one of the T observed temporal snapshots, the generic increment or 'return' $r_i(t)$ can be defined as

$$r_i(t) \equiv s_i(t+1) - s_i(t) \qquad i = 1, N \quad t = 1, T \tag{1.1}$$

and generates a new time series.

While a time series of increments encapsulates all the relevant information about the amplitude of the fluctuations of the original signal, a significant part of this information is encoded in the purely 'binary' projection of $r_i(t)$, i.e. its sign

$$x_i(t) \equiv \mathrm{sign}[r_i(t)] = \begin{cases} +1 & r_i(t) > 0 \\ 0 & r_i(t) = 0 \\ -1 & r_i(t) < 0 \end{cases} \tag{1.2}$$

Previous analyses, mainly in the field of finance, have indeed documented various forms of statistical dependency between the sign and the absolute value of fluctuations, e.g. sign-volume correlations [10, 11] and the leverage effect [12, 13, 14, 15]. Other studies have also documented that the binary projections of various financial [16] and neural [17] time series exhibit nontrivial dynamical features that resemble those of the original data. All these results suggest that binary projections indeed retain a non-trivial piece of information about the original time series, and call for a deeper analysis of the problem.

Being binary, the sign of the increments is much more robust to noise than the increments themselves. Moreover, it is scale-invariant (i.e. independent of the chosen unit of increments) and does not depend on whether the original data have been preliminarily rescaled or log-transformed (as usually done e.g. for financial time series). Binary time series can also be analyzed with the aid of much simpler mathematical models than required by non-binary data (several examples of such models will be provided in this chapter). Finally, as we show later on, in multiple financial time series the total binary increment of a given cross section measures the instantaneous level of synchronization (i.e. the number of stocks moving in the same direction) of the market, while the total non-binary increment does not carry this piece of information. For all the above reasons, it is important to further investigate whether the full 'weighted' or 'valued' information can, in some circumstances, be somehow mapped to the binary one, thus providing a robust, highly simplified, more easily modeled, and informative representation of the system.

Motivated by the above considerations, in this chapter we further study, both empirically and theoretically, the relationship between weighted time series and their binary projections. We first provide robust empirical evidence of novel relationships between binary and non-binary properties of real financial time series. To this end, we use the daily closing prices of all stocks of three markets (S&P500, FTSE100 and NIKKEI225) over the period 2001-2011. We show that the average daily increment and average daily coupling of an empirical set of multiple time series are strongly and non-linearly related to the corresponding average increment of the binary projections of the same time series. These empirical relations quantify in a novel way the strong correlations existing between the increments of individual stocks and the overall level of synchronization among all stocks in the market.

Building on this evidence, we then introduce a formalism to analytically characterize random ensembles of single and multiple time series with desired constraints. Specifically, we follow Jaynes' interpretation and re-derivation of statistical physics as an inference problem from partial macroscopic information to the unobservable microscopic configuration [18, 19]. We define statistical ensembles of matrices that maximize Shannon's entropy [20], subject to a set of desired constraints. This maximum-entropy approach is widely used in many areas, from neuroscience [49] to social network analysis [25] (and more recently network science in general [21]), where it is known under the name of ERG (Exponential Random Graph) formalism. In the case of interest here, we introduce ensembles of maximum-entropy binary matrices that represent projections of single and multiple binary time series, subject to a set of desired constraints defined as simple empirical measurements. We discuss the main differences between our matrix ensembles and other techniques in time series analysis, including other ensembles of random matrices encountered in random matrix theory [26, 27, 28, 29, 30].

Our approach leads to a family of analytically solved null models that allow us to quantify the amount of information encoded in the chosen constraints, i.e. the selected observed properties of the binary projections of real time series. Different choices of the constraints lead to different stochastic processes, a result that allows us to relate known stochastic processes to the corresponding 'target' empirical properties defining the ensemble of time series spanned by the process itself. After applying the approach to the financial time series in our analysis, we compare the informativeness of various measured properties and show that different properties are more relevant for different time series and temporal windows. We also identify distinct regimes in the behaviour of multiple stocks and give the most likely explanation (endogenous, exogenous, or mixed) for the observed level of coordination or 'market mode', given the measured binary return at a given point in time. Finally, and most importantly, we show that our approach is able to reproduce and mathematically characterize the observed nonlinear relationships between binary and non-binary properties of real time series.

The rest of the chapter is organized as follows. In sec. 1.2 we describe the data and provide empirical evidence of the relationships that motivate our work. In sec. 1.3 we introduce our theoretical formalism in its general form. In sec. 1.4 we apply the formalism to single time series, while in sec. 1.5 we apply it to single cross-sections (temporal snapshots) of multiple time series. Finally, in sec. 1.6 we consider our method in its full extent and apply it to entire spans of multiple time series, for different financial markets around the globe. We end with our conclusions in sec. 1.7.

1.2 Empirical results

1.2.1 Data

We use daily closing prices, for the 10-years period ranging from 24/10/2001 to 18/10/2011, of all stocks from the indices S&P500, FTSE100 and NIKKEI225. For each index, we restrict our sample to the maximal group of stocks that are traded continuously throughout the selected period. This results in 445 stocks for the S&P500, 78 stocks for the FTSE100 and 193 stocks for the NIKKEI225.

We take logarithms of daily closing prices to obtain time series of *log-prices* that represent our original 'signal' $s_i(t)$, where i labels stocks and t labels days in the sample. Correspondingly, we construct time series of *log-returns* where each entry represents the increment $r_i(t)$ as defined in eq.(1.1). Finally, we take the sign $x_i(t)$ of each log-return $r_i(t)$ to obtain an additional, binarized set of time series as in eq.(1.2). We will refer to the binarized time series as the *binary projection* of the original time series. In fig. 1.1 we show a simple example of a weighted time series, along with the corresponding binary projection. The (multiple) time series of $r_i(t)$ and $x_i(t)$ are the main objects of our analysis throughout the chapter. Note that, while the use of log-returns rather than simple returns (i.e. price differences) in finance is an important step that allows to remove overall trend effects over long time spans [5], the binary signature is actually independent of whether the original prices have been logarithmically transformed.

The main reason for choosing the daily frequency is to achieve an optimal level of structural compatibility between the data and the models we introduce later. As we discuss in detail in sec. 1.3, our models are binary, i.e. they only allow the two values ± 1 depending on whether the increment of the original time series is positive or negative. An increment of 0 is not admitted in the models: consistently, we choose a frequency for which zero increments are extremely rare in the data. In financial markets, this is the case for daily (or lower) frequency. Indeed, a zero return value occurs in less than 0.2 % of the cases in our daily data (when this happens, we randomly switch the corresponding binary increment to either $+1$ or -1 with equal probability). Higher-frequency data feature an increasing percentage of zero returns, a property that calls for an extension of the models considered here.

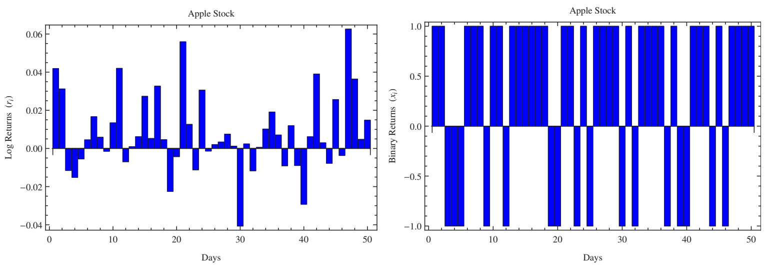

Figure 1.1: **Binary signature of financial time seires.** 'Weighted' (left) versus 'binary' (right) time series of log-returns of the Apple stock over a period of 50 days starting from 7/5/2011.

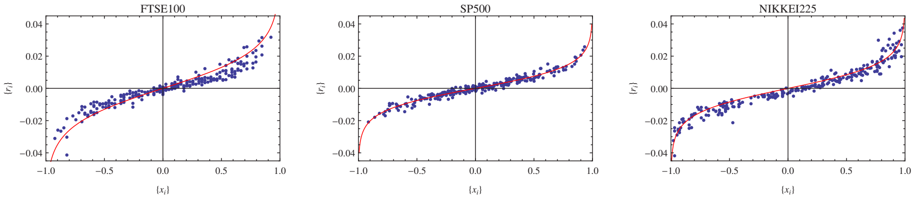

Figure 1.2: **Empirical relations between binary and non-binary properties in financial time series.** Nonlinear relationship between the average daily increment (weighted return) and the average daily sign (binary return) over all stocks in the FTSE100 (left), S&P500 (center) and NIKKEI225 (right) in various years (2003, 2007, and 2004 respectively). Here each point corresponds to one day in the time interval of 250 trading days (approximately one year). The red line represents the best fit with the function $y = a \cdot \mathrm{artanh} x$, whose use is theoretically justified later in sec. 1.6.

It should be noted that other types of binary time series, different from the ±1 type considered here, can also be defined. Most notably, 0/1 binary time series can indicate the occurrence of an event in a time period, i.e. whether the event happened (1) or not (0). Financial examples include time series of recession indicators [52, 53] or of 'switching points' in stock returns. For such 0/1 binary time series, correlations may not be very informative when measuring a dependence between the dichotomous variables. To confront this gap, in recent years new methods were introduced, like the auto-persistence function and auto-persistence graph. In these methods, the dependence structure among the observations is described in terms of conditional probabilities, rather than correlations. Although throughout this chapter we will be entirely focusing on ±1 binary time series that naturally descend from the original *signed* time series of fluctuations, it is interesting to notice that our approach can be extended, with slight modifications, to 0/1 time series as well. To this end, one needs to re-express all quantities in terms of a 0/1 binary variable $y \equiv (x+1)/2$, where x is our ±1 binary variable, and adapt our approach accordingly.

1.2.2 Nonlinear binary/non-binary relationships

We now come to the main empirical findings that motivate our research. For each index and for each day t in the sample, we first calculate the average (over all stocks) weighted return, that we denote as $\{r_i(t)\}$ and define as

$$\{r_i(t)\} \equiv \frac{1}{N}\sum_{i=1}^{N} r_i(t). \tag{1.3}$$

Note that the above expression does not depend on the particular stock i, but it does depend on time t. Our unconventional choice of the symbol $\{\cdot\}$ to denote an average over stocks is to avoid confusion with temporal averages, that will be denoted by the more usual bar $(\bar{\cdot})$ later in the chapter. Similarly, we calculate the corresponding average binary return $\{x_i(t)\}$, defined as

$$\{x_i(t)\} \equiv \frac{1}{N}\sum_{i=1}^{N} x_i(t) \tag{1.4}$$

In fig.1.2 we plot $\{r_i(t)\}$ as a function of $\{x_i(t)\}$ for all days of various 1-year intervals and for the three indices separately. We find a strong nonlinear dependency between the two quantities. Note that the average binary return is bound between -1 and $+1$ by construction, but the average weighted return is unbounded from both sides. While there are in principle infinite values of $\{r_i(t)\}$ that are consistent with the same value of $\{x_i(t)\}$, we observe a tight relationship between the two quantities. This relationship can be fitted by a one-parameter curve of the form

$$\{r_i(t)\} = a \cdot \operatorname{artanh}\big[\{x_i(t)\}\big] = \frac{a}{2}\ln\frac{1+x_i(t)}{1-x_i(t)} \tag{1.5}$$

(the theoretical justification for this functional form will be given in sec. 1.6), where a is in general different for different years and different indices. Still, as we show later, for a given year and market the average weighted return of any day t is to a large extent predictable (out of sample) from the average binary return of the same day, once a is known (for instance by fitting the above curve to the data for a past time window). In sec. 1.6 we will also show that the nonlinear character of the observed relations is a genuine signature of correlation in the data, as an uncorrelated null model shows a completely linear behaviour.

There is another empirical relationship, involving a higher-order quantity. For each index and for each day t in the sample, we calculated what we will call the average 'coupling' over the $N(N-1)/2$ distinct pairs of stocks:

$$\{r_i(t)r_j(t)\} \equiv \frac{2\sum_{i<j} r_i(t)r_j(t)}{N(N-1)} \tag{1.6}$$

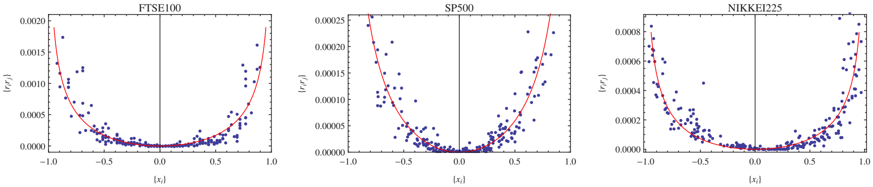

Figure 1.3: **Empirical relations between binary and second-order non-binary properties in financial time series.** Nonlinear relationship between the average daily coupling (weighted coupling) and the average daily sign (binary return) over all stocks in the FTSE100 (left), S&P500 (center) and NIKKEI225 (right) in various years (2003, 2007, and 2004 respectively). Here each point corresponds to one day in the time interval of 250 trading days (approximately one year). The red line represents the best fit with the function $y = b \cdot (\text{artanh} x)^2$, whose use is theoretically justified later in sec. 1.6.

(so now the symbol $\{\cdot\}$ indicates an average over *pairs* of stocks). In fig. 1.3 we plot $\{r_i(t)r_j(t)\}$ as a function of the average binary return $\{x_i(t)\}$, for the same data as in fig. 1.2. Again, we find a strong nonlinear dependency, where for a given value of the average binary return of day t there is a typical value of the average coupling among all stocks in the same day. The relationship can be fitted by a one-parameter curve that diverges at $\{x_i\} = \pm 1$. As we show in sec. 1.6, an uncorrelated null model would yield a different, parabolic curve with no divergences. Again, this means that the empirical trend is due to genuine correlations, whose nature will be clarified later on in the chapter.

There are even more examples of dependencies that we can find between binary and non-binary properties in the data. However, in one way or another all these relationships, including that shown in fig. 1.3, ultimately derive from eq.(1.5). For this reason, we refrain from showing redundant results and focus on the empirical findings discussed so far.

The above analysis indicates that the binary signature of financial time series contains relevant information about the original data. While the binary signature is *a priori* a many-to-one projection involving a significant information loss, we empirically find that there are properties (namely the average return and average coupling) for which the projection is virtually a one-to-one 'quasi-stationary' transformation (on appropriate time scales, as we show in sec. 1.6), allowing to reconstruct the corresponding original, weighted properties to a great extent. Rather than exploring the practical aspects of this possibility of reconstruction of the original signal from its binary projection, in this chapter we are interested in understanding the origin of this behaviour and providing a simple data-driven model of it. This will be ultimately achieved in sec. 1.6, where we also show that the binary/non-binary relations we have documented are a novel quantification of the fact that extreme price increments occur more often when most stocks move in the same direction. This is an important type of correlation between the magnitude of log-returns of individual time series and the level of synchronization (common sign) of the increments of all stocks in the market.

1.3 Maximum-entropy matrix ensembles

Having established that the binary projections of real time series contain nontrivial information, in the rest of the chapter we introduce a theory of binary time series aimed, among other things, at reproducing the observed nonlinear relationships showed in figs. 1.2 and 1.3. In our approach, we regard a synchronous set of binary time series as a ±1 matrix and we introduce an ensemble of such matrices via the maximization of Shannon's entropy, subject to the constraint that some specified properties of the ensemble match their observed values. An analogous approach is widely used e.g. in network analysis and known under the name of Exponential Random Graphs [21]. Moreover, we provide an analytical maximum-likelihood method to find the optimal values of the paramaters governing the ensembles, which is again similar in spirit to a method that has been recently introduced for networks [45, 46, 47]. Finally, we describe Akaike's information criterion (AIC) [23], which we will use to rank and compare the performance of different null models when fitted to the same data.

Being entropy-based, our approach automatically allows us to measure the amount of information encoded into the observed properties chosen as constraints, i.e. how much information is gained about the original (set of) time series once those properties are measured. It also allows us to identify, given a set of measured properties, which ones are more informative and which ones can be discarded, as we show on specific financial examples. Our framework turns out to reproduce the observed nonlinear relationships very well, thus providing a simple mathematical explanation and functional form for the plots shown in the previous section. Moreover, we are able to identify, as a function of the binary return only, distinct regimes in the collective behaviour of stocks, namely a 'coordinated' regime dominated by market-wide interactions, an 'uncoordinated' regime dominated by stock-specific noise and an 'intermediate' regime where both market-wide and stock-specific information is relevant.

We incidentally note that, despite the available variety of refined and advanced techniques in time series analysis [54], how one can quantify (in the sense of statistical ensembles) how much information is actually encoded into any given, measurable property of a time series is still not fully understood. While most studies, starting from the celebrated work by Kolmogorov about the algorithmic complexity of sequences of symbols [48], have addressed the quantification of the information content of a single time series, much less is known about the information encoded in the measured value of a given time series property (which, necessarily, involves the idea of an entire *ensemble* of time series consistent with the measured value itself). Our approach can provide an answer to such a question, by associating an absolute level of uncertainty (entropy) to each observable of an empirical (set of) time series. In relative terms, this also allows us to compare the information content of different properties of a time series, thereby indicating which measured property is the most informative about the original time series.

As a final consideration, it is worth mentioning that the maximum-entropy matrix ensembles that we introduce are clearly related to (and, depending on the specification, potentially overlapping with) some ensembles that are well studied by random matrix theory [37, 38, 39, 40, 41, 42]. However, our approach is different since we generate ensembles of matrices whose probability distributions are determined by the kind of partial information (empirically measured constraint) about the real system. In this approach the maximization of Shannon's entropy, given some real-world available information, yields the least biased probability distribution (over the space of possible matrices) consistent with the data. This formalism allows us to relate the probabilistic structure of each matrix ensemble with the choice of the original observed property, or constraint. Similarly, since our matrices represent (multiple) time series, we are able to connect the various ensembles to simple stochastic processes induced by the associated matrix probabilities and, again, to the chosen empirical property specifying the ensembles themselves.

1.3.1 Exponential random matrices

We first analytically characterize the properties of families of randomized matrices. More generally, we introduce a matrix ensemble that maximizes Shannon's entropy, while enforcing a set of observed constraints (selected time series properties). This procedure is analogous to e.g. that leading to the definition of Exponential Random Graphs in network theory [21]. However, we will modify it to accommodate ±1 matrices, as opposed to 0/1 or non-negative matrices that describe binary and weighted networks respectively. The resulting ensemble can thus be denoted as the 'Maximum-Entropy Matrix' (MEM) ensemble or equivalently 'Exponential Random Matrices' (ERMs) model.

Let us consider the ensemble of all ±1 matrices with dimensions $N \times T$. Each such matrix can represent N synchronous time series, all of duration T (for instance, if applied to a set of multiple financial time series, N refers to the number of stocks and T to the number of time steps). Let $\mathbf{X}$ denote a generic matrix in the ensemble, and $x_i(t)$ its entry ($1 \leq i \leq N$, $1 \leq t \leq T$). Let $\mathbf{X}^*$ be the particular real matrix that we observe. In other words, our ensemble is composed of all possible matrices $\mathbf{X}$ of the same type as $\mathbf{X}^*$, and includes $\mathbf{X}^*$ itself. For any data-dependent property R, we will consider the value $R(\mathbf{X})$ obtained when R is measured on the particular matrix $\mathbf{X}$. For each matrix $\mathbf{X}$ in the ensemble, we will assign an occurrence probability $P(\mathbf{X})$. The expectation value (ensemble average) of a property R can be expressed as

$$\langle R\rangle = \sum_{\mathbf{X}} R(\mathbf{X})P(\mathbf{X}) \tag{1.7}$$

where the sum runs over all matrices in the ensemble.

At this point, we introduce a set of constraints denoted by the vector $\vec{C}$. The constraints are meant to ensure that a given set of observed properties $\vec{C}(\mathbf{X}^*)$ in the real matrix $\mathbf{X}^*$ is reproduced by the ensemble itself. In our method we will enforce 'soft' constraints by requiring that their expectation value $\langle\vec{C}\rangle$ equals the observed one. The resulting ensemble is a *canonical* one where each matrix $\mathbf{X}$ is assigned a probability $P(\mathbf{X})$ that maximizes Shannon's entropy

$$S \equiv -\sum_{\mathbf{X}} P(\mathbf{X}) \ln P(\mathbf{X}) \tag{1.8}$$

subject to the normalization constraint

$$\sum_{\mathbf{X}} P(\mathbf{X}) = 1 \tag{1.9}$$

and to the chosen vector of constraints

$$\langle\vec{C}\rangle = \sum_{\mathbf{X}} C(\mathbf{X})P(\mathbf{X}) = \vec{C} \tag{1.10}$$

that we are enforced in order to reproduce the desired set of observed quantities.

The solution to the above constrained maximization problem is standard (see for instance [21] for a recent derivation in the context of networks). We first introduce the Lagrange multipliers α and $\vec{\theta}$, enforcing eqs.(1.9) and (1.10) respectively, and then require that the functional derivative of Shannon's entropy (plus the constraining terms) vanishes:

$$\frac{\partial}{\partial P(\mathbf{X})}\Big\{S + \alpha\Big[1 - \sum_{\mathbf{X}} P(\mathbf{X})\Big] + \sum_i \theta_i\Big[C_i - \sum_{\mathbf{X}} C(\mathbf{X})P(\mathbf{X})\Big]\Big\} = 0$$

This yields

$$\ln P(\mathbf{X}) + 1 + \alpha + \sum_i \theta_i C_i(\mathbf{X}) = 0 \tag{1.11}$$

for any matrix $\mathbf{X}$. Using a notation that makes the dependence of all quantities on $\vec{\theta}$ explicit, we then obtain

$$P(\mathbf{X}|\vec{\theta}) = \frac{e^{-H(\mathbf{X},\vec{\theta})}}{Z(\vec{\theta})} \tag{1.12}$$

where $H(\mathbf{X},\vec{\theta})$ is the *Hamiltonian*

$$H(\mathbf{X},\vec{\theta}) \equiv \vec{\theta}\cdot\vec{C}(\mathbf{X}) = \sum_i \theta_i C_i(\mathbf{X}), \tag{1.13}$$

which is a linear combination of the constraints, and $Z(\vec{\theta})$ is the *partition function*

$$Z(\vec{\theta}) \equiv e^{\alpha+1} = \sum_{\mathbf{X}} e^{-H(\mathbf{X},\vec{\theta})}, \tag{1.14}$$

which is the normalizing constant for the probability. Consistently, we can rewrite eq. (1.7) more explicitly as a function of $\vec{\theta}$:

$$\langle R\rangle_{\vec{\theta}} \equiv \sum_{\mathbf{X}} R(\mathbf{X})P(\mathbf{X}|\vec{\theta}) \tag{1.15}$$

where $\langle\cdot\rangle_{\vec{\theta}}$ indicates that the ensemble average is evaluated at the particular parameter value $\vec{\theta}$.

Equations (1.12) to (1.14) define the MEM or ERM model. Specifically, the model yields the probability distribution over a specified ensemble of matrices, which maximizes the entropy under a set of generic constraints. The guiding principle is that the probability distribution (over microscopic states) which have maximum entropy, subject to observed (macroscopic) properties, provides the most unbiased representation of our knowledge of the state of a system [19]. To put it in a more physical frame, this is analogous to the Gibbs-Boltzmann distribution over the microstates of a large system at a well defined temperature, given the thermodynamic (macroscopic) observables such as the total energy.

1.3.2 Maximum-likelihood parameter estimation

The above derivation shows that the expectation value of any property of the ensemble depends *functionally* on the specific enforced constraints $\vec{C}$ through the resulting structure of $P(\mathbf{X}|\vec{\theta})$. Of course, it also depends *numerically* on the measured values $\vec{C}(\mathbf{X}^*)$ of the constraints themselves, through the particular parameter value (that we denote by $\vec{\theta}^*$) required in order to enforce that the expected and observed values of $\vec{C}$ match:

$$\langle\vec{C}\rangle_{\vec{\theta}^*} = \vec{C}(\mathbf{X}^*). \tag{1.16}$$

We now show that the value $\vec{\theta}^*$ that satisfies eq.(1.16) coincides with the value that maximizes the likelihood to generate the empirical data, as in the corresponding Maximum Likelihood (ML) approach to network ensembles [22, 45].

We start by writing the log-likelihood function of an observed matrix $\mathbf{X}^*$ generated by the parameters $\vec{\theta}$:

$$\lambda(\vec{\theta}) \equiv \ln P(\mathbf{X}^*|\vec{\theta}) = -H(\mathbf{X}^*,\vec{\theta}) - \ln Z(\vec{\theta}) \tag{1.17}$$

We then look for the particular value $\vec{\theta}^*$ that maximizes $\lambda(\vec{\theta})$, i.e.

$$\vec{\nabla}\lambda(\vec{\theta}^*) = \left[\frac{\partial\lambda(\vec{\theta})}{\partial\vec{\theta}}\right]_{\vec{\theta}=\vec{\theta}^*} = \vec{0} \tag{1.18}$$

(it is easy to check that the higher-order derivative confirm that $\vec{\theta}^*$ is a point of maximum). This leads to

$$\vec{\nabla}\lambda(\vec{\theta}^*) = \left[-\vec{C}(\mathbf{X}^*) - \frac{1}{Z(\vec{\theta})}\frac{\partial Z(\vec{\theta})}{\partial\vec{\theta}}\right]_{\vec{\theta}=\vec{\theta}^*} = \vec{0} \tag{1.19}$$

the solution for that yields the ML condition

$$\vec{C}(\mathbf{X}^*) = \sum_{\mathbf{X}} \frac{\vec{C}(\mathbf{X})e^{-H(\mathbf{X},\vec{\theta}^*)}}{Z(\vec{\theta}^*)} = \langle\vec{C}\rangle_{\vec{\theta}^*}. \tag{1.20}$$

which coincides with eq.(1.16). Thus the likelihood of the real matrix $\mathbf{X}^*$ is maximized by the specific parameter choice such that the ensemble average of each constraint equals its empirical value measured on $\mathbf{X}^*$, automatically ensuring that the desired constraints are met.

1.3.3 Model selection

We finally show how we can use Akaike's information criterion (AIC) to rank the performance of different models, i.e. different choices of the constraints, in reproducing the same data. The AIC is an information-theoretic measure of the relative goodness of fit of a model, as compared to a set of alternative models all used to explain the same data [23]. It offers a relative measure of the information lost when the given model is used to describe reality. The power of AIC (and other similar criteria [24]) lies in the possibility to rank a set of models in terms of their achieved trade-off between accuracy (good fit to the data) and parsimony (low number of free parameters) [24]. In general, for the k-th model in a set of selected models, AIC is defined as

$$\mathrm{AIC}_k = 2n_k - 2\lambda_k^* \tag{1.21}$$

where n_k is the number of free parameters in the k-th model and λ_k^* is the maximized log-likelihood of the data under the same model. The above expression effectively discounts the number n_k of parameters (complexity) from the maximized likelihood λ_k^* (accuracy). The model with the lowest value of AIC_k (let us denote this value by AIC_{min}) is the 'best' model in the considered set, achieving the optimal trade-off [24].

In the ERM/MEM family of models we have introduced, a model is uniquely specified by the choice of the constraints $\vec{C}$. Given a $N \times T$ data matrix $\mathbf{X}^*$ and a set $\{\vec{C}_1, \ldots, \vec{C}_m\}$ of m possible choices of constraints, each of the resulting m models has an AIC value

$$\mathrm{AIC}_k = 2n_k - 2\ln P_k(\mathbf{X}^*|\vec{\theta}_k^*) \qquad k = 1, m \tag{1.22}$$

where n_k is the dimensionality of the vector $\vec{C}_k$, $\ln P_k(\mathbf{X}^*|\vec{\theta}_k^*)$ is the maximized log-likelihood of model k, and $\vec{\theta}_k^*$ is the parameter value maximizing such log-likelihood. Within our framework, AIC identifies which measured property $\vec{C}_k(\mathbf{X}^*)$ is most informative about the entire time series $\mathbf{X}^*$.

In order to understand whether models with values of AIC larger than but close to AIC_{min} are still competitive, it is customary to define the so-called 'AIC weights' which provide a normalized strength of evidence for a model [24]. For each model k in the set of m models, one first calculates the difference $\Delta_k = AIC_k - AIC_{min}$ and then defines the AIC weight

$$w_k \equiv \frac{e^{-\Delta_k/2}}{\sum_{r=1}^{m} e^{-\Delta_r/2}}. \tag{1.23}$$

The AIC weight w_k represents the probability that the k-th model is the best one among the m selected models. For instance, an AIC weight of $w_k = 0.75$ indicates that, given the data, model k has a 75% chance of being the best model among the m candidate ones. If two or more models have comparable AIC weights (e.g. $w_1 = 0.6$, $w_2 = 0.4$ or $w_1 = 0.35$, $w_2 = 0.25$, $w_3 = 0.4$), then there is no evidence that the model with the highest AIC weight (lowest AIC value) is clearly outperforming the other ones. All the models with comparable weights should be considered as competing alternatives, in principle leading to the problem of multi-model inference [24].

1.4 Single time series

In this section we consider the first family of specifications of our general approach outlined in sec. 1.3. We focus on the simple case of single time series ($N = 1$), where the ensemble of $N \times T$ matrices reduces to an ensemble of $1 \times T$ matrices, or equivalently of T-dimensional row vectors. Each such vector will still be denoted by $\mathbf{X}$. We assume long time series, i.e. $T \gg 1$.

This first specification of our abstract formalism is not meant to provide realistic models for the evolution of the binary increments of real financial time series. Rather, it allows us to make different sorts of considerations. On one hand, it allows us to introduce our formalism using simpler examples first, establishing the basis for the more general cases (leading to the main results of this chapter) that will be introduced later. On the other hand, it emphasizes that different and well known (one-dimensional) stochastic processes are found as particular examples of maximum-entropy ensembles defined by specific constraints that are otherwise obscure. Identifying these 'driving constraints' underlying common stochastic processes will help us interpret such processes in the light of the empirical properties being reproduced. Finally, our approach allows us to identify, given the data and given a set of simple properties, which of these properties is encoding the largest amount of information about the original binary signature.

Let $\mathbf{X}$ denote a single time series with entries $x(t)$, where $1 \leq t \leq T$, each representing a temporal increment. We will denote the average increment (first moment) as

$$M_1(\mathbf{X}) \equiv \overline{x(t)} = \frac{1}{T}\sum_{t=1}^{T} x(t). \tag{1.24}$$

Note that the second moment is always

$$M_2(\mathbf{X}) \equiv \overline{x^2(t)} = \frac{1}{T}\sum_{t=1}^{T} x^2(t) = 1, \tag{1.25}$$

so the sample variance is

$$M_2(\mathbf{X}) - M_1^2(\mathbf{X}) = 1 - \overline{x(t)}^2. \tag{1.26}$$

We also define the τ-delayed product (with $0 \leq \tau \leq T$)

$$B_\tau(\mathbf{X}) \equiv \overline{x(t) \cdot x(t+\tau)} = \frac{1}{T}\sum_{t=1}^{T} x(t) \cdot x(t+\tau) \tag{1.27}$$

where we have introduced periodic boundary conditions:

$$x(T+\tau) \equiv x(\tau) \quad \text{with} \quad 0 \leq \tau \leq T \tag{1.28}$$

The above periodicity condition in inessential, since we could have used a definition avoiding its introduction, but it makes some expressions simpler in what follows. Periodicity implies that the normalized (between -1 and $+1$) autocorrelation function (with delay τ) can be defined as

$$\begin{aligned} A_\tau(\mathbf{X}) &\equiv \frac{\overline{x(t) \cdot x(t+\tau)} - \overline{x(t)} \cdot \overline{x(t+\tau)}}{\overline{x^2(t)} - \overline{x(t)}^2} \\ &= \frac{B_\tau(\mathbf{X}) - M_1^2(\mathbf{X})}{1 - M_1^2(\mathbf{X})} \end{aligned} \tag{1.29}$$

Since a (± 1) binary time series can also be regarded as a chain of classical spins pointing either up or down, it is straightforward to consider simple, analytically solved spin models as the starting point, since these models are defined in terms of a ‘physical’ Hamiltonian that has precisely the same structure of our ‘information-theoretic’ Hamiltonian defined in eq. (1.13). In what follows, we introduce various

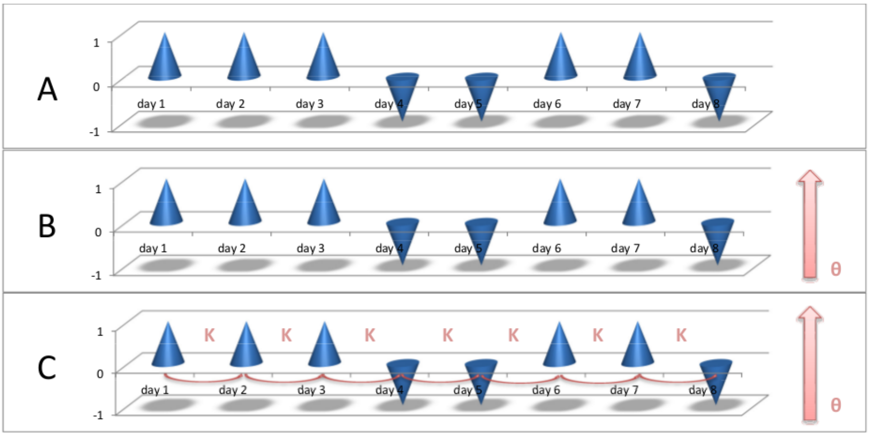

Figure 1.4: **Single financial time series as a chain of spins.** Illustration of our mapping from single binary time series to spin models. Each time series is regarded as a chain of ±1 spins, where the value of the spin indicates if the daily return of the stock is positive $(+1)$ or negative (-1) . In each model we enforce different constraints, that imply different spin models and different stochastic processes. Given the same time series, we consider three possible models. A) we enforce no constraint, which translates into a chain of non-interacting spins without external field (uniform random walk). B) we enforce the total temporal increment, which translates into a chain of non-interacting spins with external field (biased random walk). C) we enforce both the total increment and the one-lagged autocorrelation, which translates into a chain of spins with first-neighbour interactions and external field (markov process).

model specifications. For each model, we introduce the constraints that we enforce and the resulting Hamiltonian as described in sec. 1.3.1. Different constraints correspond to different spin models and lead to different stochastic processes. This is pictorially illustrated in fig. 1.4. The free parameters conjugated to the constraints will be fitted according to the Maximum Likelihood principle described in sec. 1.3.2. Different models will be ranked according to the AIC weights introduced in sec. 1.3.3.

1.4.1 Uniform random walk

The most trivial model is one where we enforce no constraint, i.e. there is no free parameter and the Hamiltonian is

$$H(\mathbf{X}) = 0. \tag{1.30}$$

Physically, the above Hamiltonian describes a gas of T non-interacting 'spins' in vacuum, i.e. in absence of an external magnetic field. This model is discussed in the Appendix. The probability of occurrence of a time series $\mathbf{X}$ is completely uniform over the ensemble of all binary time series of length T. All the T elements of $\mathbf{X}$ are mutually independent and identically distributed. This results in a completely uniform random walk with zero expected value for each increment:

$$\langle x(t)\rangle = 0 \tag{1.31}$$

While the (ensemble) variance of each increment equals

$$\mathrm{Var}[x(t)] \equiv \langle x^2(t)\rangle - \langle x(t)\rangle^2 = 1. \tag{1.32}$$

This trivial model generates a symmetric random walk. Since the expected return is zero, and the uncertainty is maximal, the variance is also maximal (for a ±1 binary random variable). Financially, the model assumes that the stock fluctuates randomly, with no memory, and with no overall 'price drift'. This is the most basic model of price dynamics that has been considered in the financial literature since the pioneering work of Bachelier [1], here adapted to the case of binary time series.

The model can be used as a basic benchmark for checking the performance of our other models. This comparison will be studied in sec. 1.4.4. Since here the likelihood is independent of any parameter, the AIC of the model can be calculated using eq. (1.22) where the probability is given by eq. (A.3) (see Appendix) and the number of parameters is $n_k = 0$.

1.4.2 Biased random walk

We now consider the total increment as the simplest non-trivial (one-dimensional) constraint:

$$C(\mathbf{X}) = T \cdot M_1(\mathbf{X}) = T \cdot \overline{x(t)} \tag{1.33}$$

This leads to the Hamiltonian

$$H\left(\mathbf{X}, \theta\right) = \theta \sum_{t=1}^{T} x(t), \tag{1.34}$$

which coincides with the physical Hamiltonian for a gas of T non-interacting 'spins' in a common external 'magnetic field' $-\theta$.

As we show in the Appendix, this model generates a *biased* random walk where the probability $P_t(x|\theta)$ of a given increment $x = \pm 1$ at time t is

$$P_t(x|\theta) = \frac{e^{-\theta x}}{e^{-\theta} + e^{+\theta}}. \tag{1.35}$$

The expected return is the hyperbolic tangent

$$\langle x(t) \rangle_\theta = -\tanh\theta, \tag{1.36}$$

while the variance is

$$\mathrm{Var}[x(t)] = 1 - \tanh^2\theta. \tag{1.37}$$

Financially, this model still assumes no memory in the fluctuations of a given stock, but it introduces a 'price drift' in terms of a non-zero expected return.

The maximum likelihood condition (1.16), fixing the value θ^* of the parameter θ given a real time series $\mathbf{X}^*$, leads to

$$\theta^* = -\frac{1}{2}\ln\left[\frac{1+\overline{x^*(t)}}{1-\overline{x^*(t)}}\right]. \tag{1.38}$$

The maximized likelihood for the model is

$$P(\mathbf{X}^*|\theta^*) = \prod_{t=1}^{T} P_t\big(x^*(t)|\theta^*\big) \tag{1.39}$$

which, using eq. (1.22) with $n_k = 1$, can be used to measure the AIC (see sec. 1.3.3) of the model, based on the observed data. This will be done in sec. 1.4.4.

1.4.3 One-lagged model

Let us now explore a more complex model of collective behaviour. The models considered so far were non-interacting, i.e. each return in the time series was independent of the previous outcomes. Now we consider a model where, besides the constraint on the total increment specified in eq. (1.33), we enforce an additional constraint on the time-delayed (lagged) quantity $T \cdot B_1(\mathbf{X})$, where $B_1(\mathbf{X})$ is defined in eq. (1.27) with $\tau = 1$. Financially, this amounts to enforce the average return *and* the average one-step temporal autocorrelation of the time series. In order words, besides a price drift, we also introduce a short-term memory.

The resulting 2-dimensional constraint can be written as

$$\vec{C}(\mathbf{X}) = \begin{pmatrix} C_1(\mathbf{X}) \\ C_2(\mathbf{X}) \end{pmatrix} = T \cdot \begin{pmatrix} M_1(\mathbf{X}) \\ B_1(\mathbf{X}) \end{pmatrix}. \tag{1.40}$$

If we write the corresponding Lagrange multiplier as

$$\vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = -\begin{pmatrix} I \\ K \end{pmatrix}, \tag{1.41}$$

then the Hamiltonian reads

$$H(\mathbf{X}, I, K) \quad = \quad -I\sum_{t=1}^{T} x(t) - K\sum_{t=1}^{T} x(t)x(t{+}1), \tag{1.42}$$

where we consider a periodicity condition as in eq. (1.28) with $\tau = 1$, i.e. $x(T+1) \equiv x(1)$. Note that, when $\mathbf{X}$ is a real binary time series of length T, this condition can be always enforced by adding one last (fictious) timestep $T+1$ and a corresponding increment $x(T+1)$ chosen equal to $x(1)$. For long time series (large T), the effects induced by this addition are negligible.

The above Hamiltonian coincides with that for the one-dimensional Ising model with periodic boundary conditions [55], which is a model of interacting spins under the influence of an external 'magnetic' field I. The model is analytically solvable (see Appendix for the complete derivation), which allows us to apply it to real time series in our formalism. In our setting, each time step t is seen as a site in an ordered chain of length T, and each value $x(t) = \pm 1$ is seen as the value of a spin sitting at that site. 'First-neighbour interactions' along the chain of spins are here interpreted as one-lagged memory effects. As a result of these interactions, the model generates time series according to a Markov process where the probability of an increment $x(t+1)$ depends on the realized increment $x(t)$ at the previous time step t. This is evident from the solution of the model, see e.g. eq. (A.32) in the Appendix.

The solution of the model yields the following expectation values

$$\langle M_1 \rangle_{I,K} \quad = \quad \frac{e^{2K}\sinh I}{\sqrt{1+e^{4K}\sinh^2 I}} \tag{1.43}$$

$$\langle B_\tau \rangle_{I,K} \quad = \quad \frac{e^{4K}\sinh^2 I + (\lambda_1/\lambda_2)^\tau}{1+e^{4K}\sinh^2 I} \tag{1.44}$$

(see Appendix) where

$$\lambda_1 \quad = \quad e^K\cosh I + \sqrt{e^{2K}\sinh^2 I + e^{-2K}}, \tag{1.45}$$

$$\lambda_2 \quad = \quad e^K\cosh I - \sqrt{e^{2K}\sinh^2 I + e^{-2K}}. \tag{1.46}$$

The resulting expected value of the normalized autocorrelation defined in eq. (1.47) is simply

$$\langle A_\tau \rangle_{I,K} = \left(\frac{\lambda_1}{\lambda_2}\right)^\tau . \tag{1.47}$$

The above expressions allow us to calculate all the relevant expected properties of the time series generated by the model, once the parameters I and K are set to the values I^* and K^* maximizing the likelihood $P(\mathbf{X}^*|I,K)$ of the observed time series $\mathbf{X}^*$. These values are the solutions of the coupled equations

$$M_1(\mathbf{X}^*) \quad = \quad \langle M_1 \rangle_{I^*,K^*} \tag{1.48}$$

$$B_1(\mathbf{X}^*) \quad = \quad \langle B_1 \rangle_{I^*,K^*}. \tag{1.49}$$

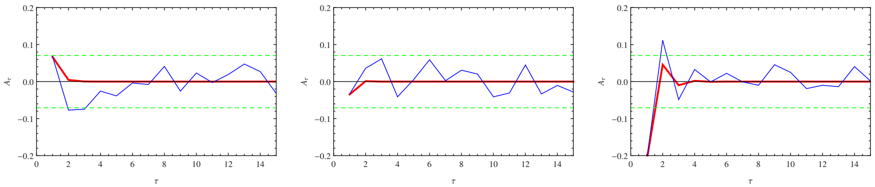

Figure 1.5: **Measured versus expected autocorrelation of single time series.** Measured autocorrelation (blue) of three different S&P500 stocks (Qcom, USB, and MJN respectively) over a period of 800 trading days (approximately 3.5 years), and comparison with the predicted autocorrelation $\langle A_\tau \rangle_{I,K}$ generated by the one-lagged (one-dimensional Ising) model (red). The green lines represent the noise level, calculated as ±2 standard deviations of the Fisher-transformed autocorrelation.

where $M_1(\mathbf{X}^*)$ and $B_1(\mathbf{X}^*)$ are the empirical values measured on the real data $\mathbf{X}^*$. The maximized likelihood of the model can be calculated as $P(\mathbf{X}^*|I^*, K^*)$, where $P(\mathbf{X}|I, K)$ is given by eq. (A.32) in the Appendix. From the maximized likelihood, the AIC can be easily obtained using eq. (1.22) with $n_k = 2$.

Note that the values I^* and K^* are such that the first point of the expected autocorrelation function, $\langle A_1 \rangle_{I^*,K^*}$, is necessarily equal to the observed value $A_1(\mathbf{X}^*)$. Based on this first value alone, the model will provide the full expected autocorrelation $\langle A_\tau \rangle_{I^*,K^*}$ as follows:

$$\big\langle A_\tau \big\rangle_{I^*,K^*} = \left(\frac{\lambda_1}{\lambda_2}\right)^{\tau}_{I^*,K^*} = \big[A_1(\mathbf{X}^*)\big]^{\tau}. \tag{1.50}$$

Comparing the above expression, for $\tau > 1$, with the observed autocorrelation function $A_\tau(\mathbf{X}^*)$ is an important test of the model. Note that, since $-1 \leq A_1(\mathbf{X}^*) \leq +1$, the absolute value of the autocorrelation function $\big\langle A_\tau \big\rangle_{I^*,K^*}$ is necessarily decreasing. If $A_1(\mathbf{X}^*) > 0$ then $\big\langle A_\tau \big\rangle_{I^*,K^*}$ will be positive (and exponentially decreasing) for all values of τ. By contrast, if $A_1(\mathbf{X}^*) < 0$ then $\big\langle A_\tau \big\rangle_{I^*,K^*}$ will be an oscillating function (modulated by a decreasing exponential), and will take negative values when τ is odd and positive values when τ is even.

In fig. 1.5 we compare the measured autocorrelation, eq. (1.29), with the predicted one, eq. (1.50), for three different S&P500 stocks (USB, Qcom, and MJN) over a period of 800 trading days (approximately 3.5 years). As expected, we see that the first point (one-lagged autocorrelation) is always reproduced exactly. We also confirm that, depending on the sign of the first point, the predicted trend is either exponentially decreasing (e.g. for the USB stock on the left) or oscillating (e.g. the Qcom and MJN stocks). The dashed lines indicate the noise level, that we arbitrarily fixed at two standard deviations of the Fisher-transformed [1] auto-

[1] For a set of T independent and identically distributed pairs of random variables $\{x_i, y_i\}_{i=1}^{T}$, the

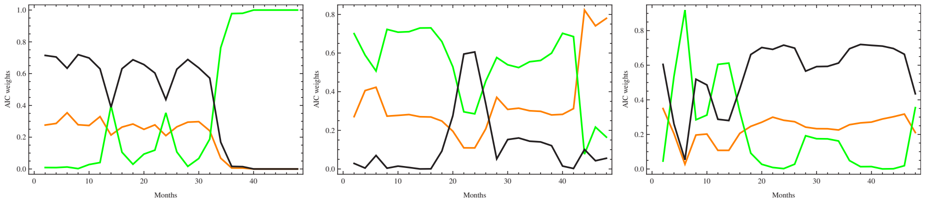

Figure 1.6: **Model performance for single time series.** Measured AIC weights for the three models (black: uniform random walk, orange: biased random walk, green: one-lagged model) calculated for three different S&P500 stocks, as a function of the time horizon T. The latter represents the number of months elapsed backwards from October 2011: for all stocks, all time series used to calculate the AIC weights have the same endpoint T_0 =31 October 2011, and a variable startpoint $T_0 - T$.

correlation. The behaviour of the USB and Qcom stocks is representative of the vast majority of stocks, with the autocorrelation within the noise level already at the minimum delay ($\tau = 1$). This is in good agreement with what we know about financial time series (no dependencies for daily frequency, the typical time scale for autocorrelation being of the order of minutes). We also found that the first point, the autocorrelation between two successive days, is small but negative for most stocks in our data set. In the rightmost panel (MJN stock) we observe a rare dynamic, where the one-lagged autocorrelation is breaching the noise level and then rapidly oscillates to zero.

As clear from the figure, our model reproduces well the observed autocorrelation in all these different cases, and gives a single mathematical explanation for both the exponentially decaying (from positive one-lagged autocorrelation) and the oscillating (from negative one-lagged autocorrelation) behaviour. Moreover, the generic feature of the one-dimensional Ising model, i.e. the absence of a phase transition characterized by a diverging length (here, time) scale [55], explains why in real-world time series the memory is always found to be short-ranged.

1.4.4 Comparing the three models on empirical financial time series

As we illustrated in sec. 1.3.3 in the general case, once we have more than one model for the same data $\mathbf{X}^*$, we can use the AIC weights to rank all models in terms of the achieved trade-off between accuray (good fit to the data) and parsimony (small number of parameters). The AIC weight w_k of a specific model

Pearson correlation coefficient $\rho_{x,y}$ is distributed around zero, but in a non-Gaussian way. However, the quantity $\phi_{x,y} \equiv \text{artanh}(\rho_{x,y})$, known as the *Fisher tranformation*, is normally distributed around zero, with standard deviation $\sigma = (T-3)^{-1/2}$. The interval $-2\sigma < \phi_{x,y} < +2\sigma$, representing a 95% confidence interval for $\phi_{x,y}$, can then be mapped back to the interval $-\tanh(2\sigma) < \rho_{x,y} < +\tanh(2\sigma)$ to obtain a 95% confidence interval for $\rho_{x,y}$ around zero.

k represents the probability that the model is the 'best' one, among the candidate models.

We applied this procedure to the three models discussed so far (uniform random walk, biased random walk, one-lagged model). As an example, in fig. 1.6 we show the values of the AIC weights for three different S&P500 stocks. We can see that the performance of the models is wildly fluctuating and different across stocks. This suggests that the informativeness of the measured properties is dependent on different factors, which are not entirely revealed to us. However, it is clear that in all cases the time horizon T plays a key role in the performance of the models. This means that the outcome depends on how many time steps are included in the analysis. For instance, we see that in some cases (Citigroup Inc. stock) the small T regime is oscillatory, while the large T regime appears to set a preference for a definite model. In other cases (United Health Group), the three models alternate over quite long periods of time. Most likely, this very irregular behaviour is due to the strong non-stationarity of financial markets: extending the analysis over longer time horizons does not necessarily improve the statistics, because for large T the underlying price (and return) distributions change in an uncontrolled way.

We stress again that the AIC weight indicates which property, among the constraints defining all models, can better characterize the stock, given the observed data. In other words, it highlights the *measured property* that is most informative about the original data. Despite the fact that the models considered so far are extremely simplified (and are by no means intended to be accurate models of financial time series), this approach can always identify, in relative terms, the most useful empirical quantity characterizing an observed time series.

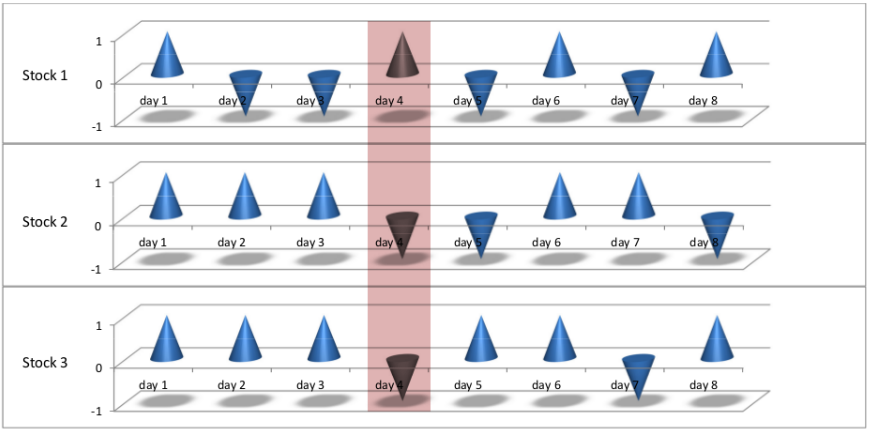

Figure 1.7: **Single cross-section of multiple time series as a chain of spins.** An example of cross section (highlighted in red) of a set of $N = 3$ multiple time series. Each cross section is a $N \times 1$ matrix (column vector) where each element is the instantaneous binary return of a different stock. For example, the highlighted cross section is the vector for day $t = 4$.

1.5 Single cross-sections of multiple time series

In the previous section we considered models for single time series, where $N = 1$ and T is large. Here we consider, as a second specification of our general formalism, the somewhat 'opposite' case of single cross-sections of N multiple time series, which represent a daily snapshot of the market dynamics. For clarity, fig. 1.7 portrays a single cross-section of a set of multiple time series. In this case, $T = 1$ and we assume $N \gg 1$. So the matrix $\mathbf{X}$ has dimensions $N \times 1$, i.e. it is an N-dimensional column vector. The entries of a cross-section $\mathbf{X}$ will be denoted by x_i, where $1 \leq i \leq N$, each representing the daily increment of a different asset.

Using again the symbol $\{\cdot\}$ to denote an average over stocks (as in sec. 1.2.2), we now define the average increment (first moment) of $\mathbf{X}$ as

$$M_1(\mathbf{X}) \equiv \{x_i\} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1.51}$$

and the second moment as

$$M_2(\mathbf{X}) \equiv \{x_i^2\} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 = 1. \tag{1.52}$$

Therefore the sample variance is

$$M_2(\mathbf{X}) - M_1^2(\mathbf{X}) = 1 - \{x_i\}^2. \tag{1.53}$$

We also define the total 'coupling' between stocks (for a specific cross section $\mathbf{X}$) as

$$D(\mathbf{X}) \equiv \sum_{i<j} x_i x_j = \{x_i x_j\} \frac{N(N-1)}{2}, \tag{1.54}$$

where now, as in eq. (1.6), $\{\cdot\}$ denotes an average over all pairs of stocks.

In what follows, we will consider various models for single cross-sections. The main difference with respect to the models of single time series considered in sec. 1.4 is that the interaction between time steps for a given stock is now replaced by the interaction between different stocks for a given time step. As well known, in real financial markets the interactions among stocks (as measured e.g. via cross-correlations) are much stronger than inter-temporal autocorrelations. This makes the cross-sectional properties significantly different from those of the dynamics of single time series, once inter-stock interactions are enforced in the model. Yet, in simple models without interaction, we recover similar expected properties.

1.5.1 Uniform random walk

As in sec. 1.4.1, we first consider a trivial model without constraints (see Appendix), defined by the Hamiltonian

$$H(\mathbf{X}) = 0. \tag{1.55}$$

The probability of occurrence of a cross section $\mathbf{X}$ is completely uniform over the ensemble of all binary cross sections of N stocks. Again, this 'gas of non-interacting spins in vacuum' model results in a uniform random walk, where all the N elements of $\mathbf{X}$ are mutually independent and identically distributed.

In the financial setting, this model assumes that all stocks fluctuate independently of each other (where the 'fluctuations' are intended as ensemble ones, since we are now considering a single cross section), and under the effect of no common factor. Each stock has zero expected value

$$\langle x_i \rangle = 0 \tag{1.56}$$

and maximum variance

$$\mathrm{Var}[x_i] \equiv \langle x_i^2 \rangle - \langle x_i \rangle^2 = 1. \tag{1.57}$$

In sec. 1.5.4, we will compare the performance of this trivial benchmark to that of the other models we are about to introduce. To this end, the AIC value can be calculated from eq. (1.22) choosing $n_k = 0$ and using the (constant) likelihood given by eq. (A.46) in the Appendix.

1.5.2 Biased random walk

In this model, which is analogous to that defined in sec. 1.4.2, the constraint is chosen as the total daily increment of the cross section $\mathbf{X}$:

$$C(\mathbf{X}) = N \cdot M_1(\mathbf{X}) = N \cdot \{x_i\}, \tag{1.58}$$

where $M_1(\mathbf{X})$ is defined by eq. (1.51). The Hamiltonian is then

$$H\left(\mathbf{X}, \theta\right) = \theta \sum_{i=1}^{N} x_i. \tag{1.59}$$

Similarly to its counterpart for single time series, this is a model of non-interacting spins under the effect of a common external field, and leads to a biased random walk (see Appendix). The financial interpretation is however different: in this model, all stocks are assumed to fluctuate (again, in an 'ensemble' sense) under the effect of a common market-wide factor, but are conditionally independent of each other, given the market-wide factor itself. In the econophysics literature, the overall tendency of all stocks to move together is generally referred to as the 'market mode' [2]. When applied to the data, this extremely simple model interprets the observed market mode as the consequence of an external factor (e.g. news), and not of direct interactions among stocks.

The probability $P_i(x|\theta)$ of a given increment $x = \pm 1$ for stock i is

$$P_i(x|\theta) = \frac{e^{-\theta x}}{e^{-\theta} + e^{+\theta}}, \tag{1.60}$$

the expected value of the i-th increment x_i is

$$\langle x_i \rangle_\theta = -\tanh\theta, \tag{1.61}$$

and the variance is

$$\mathrm{Var}[x_i] = 1 - \tanh^2\theta. \tag{1.62}$$

The maximum likelihood condition (1.16), fixing the value θ^* of the parameter θ given a real cross section $\mathbf{X}^*$, leads to

$$\theta^* = -\frac{1}{2}\ln\left[\frac{1 + \{x_i^*\}}{1 - \{x_i^*\}}\right], \tag{1.63}$$

where $\{x_i^*\}$ is the measured average increment of the observed cross section $\mathbf{X}^*$. We will apply this model to real financial data in secs. 1.5.4 and 1.6. The AIC of the model is given by eq. (1.22) where $n_k = 1$ and where the maximized likelihood is given by $P(\mathbf{X}^*|\theta^*)$, with $P(\mathbf{X}|\theta)$ given by eq. (A.53) (see Appendix).

1.5.3 Mean field model

We now consider a more complex model, with interactions among *all* stocks, which is suitable for financial cross-sections. Besides the constraint on the total increment, we enforce an additional constraint on the average coupling between stocks. The resulting 2-dimensional constraint can be written as

$$\vec{C}(\mathbf{X}) = \left(\begin{array}{c} C_1(\mathbf{X}) \\ C_2(\mathbf{X}) \end{array} \right) = \left(\begin{array}{c} N \cdot M_1(\mathbf{X}) \\ D(\mathbf{X}) \end{array} \right) \tag{1.64}$$

where $M_1(\mathbf{X})$ is given by eq. (1.51) and $D(\mathbf{X})$ by eq. (1.54). If we write the corresponding Lagrange multiplier as

$$\vec{\theta} = \left(\begin{array}{c} \theta_1 \\ \theta_2 \end{array} \right) = - \left(\begin{array}{c} h \\ J \end{array} \right) \tag{1.65}$$

then the Hamiltonian reads

$$H(\mathbf{X}, h, J) = -h \sum_{i=1}^{N} x_i - J \sum_{i<j} x_i x_j . \tag{1.66}$$

Like the one-lagged model for single time series (see sec. 1.4.3), this model is formally analogous to an Ising model of interacting spins under the influence of an external ‘magnetic’ field (here denoted by h). However, the big difference is that, whereas in the one-lagged model each increment $x(t)$ interacts *only with the next temporal increment* $x(t+1)$ *of the same stock*, here each increment x_i interacts *with all the other increments* x_j *of the same cross section* $\mathbf{X}$, i.e. with all other stocks in the market. As a model of spin systems, the above model is generally known as the mean-field Ising model [55]. In the Appendix we provide the analytical solution of the model, adapted to our setting.

In the financial setting, this model allows us to separately consider the effects of the external field, i.e. a common factor affecting all stocks in the market, from those of the average interaction among all stocks. This market-wide interaction can also cause all stocks to correlate, but has the different interpretation of a collective effect, i.e. the tendency of stocks increments to ‘align’ with each other as a result of direct interactions, rather than of a common influence. This is a sort of ‘herd effect’ at the coarse-grained level of attractive ($J > 0$) inter-stock interactions. So, the model can generate the ‘market mode’ either as the result of a common external influence such as news (in which case all stocks are still conditionally independent given the common factor), or as a collective effect due to mutual interactions (in which case all stocks are conditionally dependent given the common factor).

While the model can in principle simulate synthetic time series under a combination of the above two effects by varying the two parameters h and J independently, a problem arises when it is fitted to the data. The mathematical root of the problem is the well known fact that $H(\mathbf{X}, h, J)$ can be rewritten as a linear combination of $M_1(\mathbf{X})$ and $M_1^2(\mathbf{X})$. As we show in the Appendix, this implies that, when the maximum likelihood principle is used to fit the model to the data $\mathbf{X}^*$, the variance of $M_1(\mathbf{X})$ becomes zero. In other words, the model degenerates to one where $M_1(\mathbf{X})$ is no longer a random variable. This also implies that the two equations fixing the values of the parameters J^* and h^* become identical (see Appendix). Therefore it is no longer possible to uniquely fix the values of both parameters, and the problem is over-constrained. For this reason, we need to eliminate one parameter and consider the model only in the two extreme cases $h = 0$ and $J = 0$. These two cases can be treated separately.

The case $J = 0$ coincides with the biased random walk model already considered in sec. 1.5.2, where $\theta = -h$. Using eq. (1.63), we therefore specify this model using the two parameter values

$$h^* = \frac{1}{2} \ln \left[\frac{1 + \{x_i^*\}}{1 - \{x_i^*\}} \right], \quad J^* = 0 \tag{1.67}$$

where $\{x_i^*\}$ is the observed average increment of the empirical cross section $\mathbf{X}^*$. This model interprets the market mode as arising *only* from a common external factor.

The case $h = 0$ leads us instead to a novel model where the market mode is interpreted *only* as a collective effect arising from inter-stock interactions. Using the analytical results reported in the Appendix, and in particular eq. (A.78), we find that the parameter values are in this case

$$h^* = 0, \quad J^* = \frac{1}{2\{x_i^*\}(N-1)} \ln \left[\frac{1 + \{x_i^*\}}{1 - \{x_i^*\}} \right]. \tag{1.68}$$

In what follows, when using the 'mean-field' model, we will always refer to the parameter specification defined by eq. (1.68). The other specification, eq. (1.67), will instead still be denoted as the 'biased random walk' model.

In fig. 1.8 we plot the value of J^* as a function of $\{x_i^*\}$, as defined by eq. (1.68). We note however that eq. (1.68) is undefined for $\{x_i^*\} = \pm 1$ and $\{x_i^*\} = 0$. The breakdown for $\{x_i^*\} = \pm 1$ simply means that, in order to align *all* returns (in either direction), J^* should diverge to $+\infty$. The breakdown for $\{x_i^*\} = 0$ is instead more profound. For infinitesimal (both positive and negative) values of $\{x_i^*\}$, J^* admits the finite limit

$$\lim_{\{x_i^*\} \to 0^+} J^* = \lim_{\{x_i^*\} \to 0^-} J^* = \frac{1}{N-1} \tag{1.69}$$

However, at the very point $\{x_i^*\} = 0$, J^* is actually indeterminate.

The above effect is due to the well-known phase transition of the mean-field Ising model. In the traditional physical setting, the phase transition occurs at a critical temperature (here reabsorbed in the value of the parameters h and J).

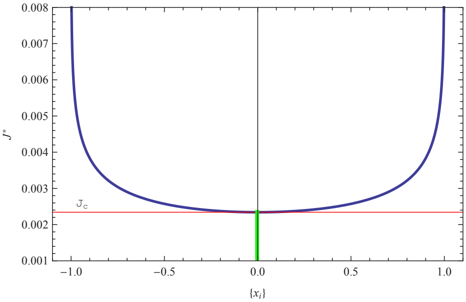

Figure 1.8: **"Phase diagram" of the fitted parameter J^*.** The value of the fitted parameter J^* as a function of the measured average binary return $\{x_i^*\}$ (blue curve) for a group of $N = 428$ stocks (as in our S&P sample). The curve shows a one-to-one relationship for $\{x_i\} \neq 0$. While $\lim_{\{x_i^*\}\to 0} J^* = J_c \equiv (N-1)^{-1}$, for $\{x_i^*\} = 0$ the value of J^* is actually indeterminate, as there is an infinity of values of J^* (namely all values $-\infty < J^* \leq J_c$, see vertical green line) that are possible solutions of the model. The value of J_c is indicated by the horizontal red line.

When $h = 0$, the critical value is obtained by setting $(N-1)J = 1$, because for $(N-1)J < 1$ eq. (A.76) (see Appendix) has the single solution $\langle M_1 \rangle = 0$, corresponding to a phase with no macroscopic magnetization, while for $(N-1)J > 1$ there are three solutions, one of which is still $\langle M_1 \rangle = 0$ (which is now unstable) and the other two ones being the stable solutions $\langle M_1 \rangle = \pm m$ (corresponding to the onset of a macroscopic magnetization $|m| > 0$ where most spins point in the same direction). In our financial setting, since the magnetization is fixed by the data through the relation $\langle M_1 \rangle = \{x_i^*\}$, the condition $(N-1)J^* = 1$ implies that the phase transition occurs at the critical value

$$J_c = \frac{1}{N-1} \tag{1.70}$$

of the control parameter J^*. We can therefore rewrite eq. (1.69) as

$$\lim_{\{x_i^*\}\to 0} J^* = J_c \tag{1.71}$$

For $J^* > J_c$ we get a 'magnetized' phase where most stock prices move in the same direction (aligned returns), while for $J^* < J_c$ we get a non-magnetized phase where there is no collective alignment of stock increments, and $\{x_i^*\} = 0$. We therefore conclude that the reason why the value of J^* is indeterminate for $\{x_i^*\} = 0$ is because there is an infinity of values of J^* (namely all values $-\infty < J^* \leq J_c$) that are possible solutions of the model.

It should be noted that the case $\{x_i^*\} = 0$ is never practically encountered in reality, since the empirical $\{x_i^*\}$ can be abritrarily small, but is generally not really zero. While this 'protects' the model from the indeterminacy discussed above, it raises another problem of arbitrariness, which can however be solved very effectively using the information-theoretic criteria that we have introduced in sec. 1.3.3. The problem is that the mean-field model will always interpret even the tiniest empirical deviations from $\{x_i^*\} = 0$ as the result of direct interactions among stocks, and attach a value $J^* > 0$ to this interpretation. This will also apply to e.g. most realizations of a purely uniform random walk: even if for such a model one knows that the theoretical expected return is zero, most realizations will be such that $\{x_i^*\}$ is small but non-zero. So the only phase of the mean-field model that can be explored is the 'magnetized' phase dominated by collective effects. This implies that even a pure effect of noise will be interpreted as the presence of interactions. However, this problem will be solved in the next section, where we show that an information-theoretic comparison between the mean-field model, the uniform random walk, and the biased random walk is able to discriminate the most parsimonious model, thus allowing us to trust the mean-field model only when $\{x_i^*\}$ is distant enough from zero.

1.5.4 Comparing the three models on empirical financial cross sections

We can now combine the three models together and use the AIC weights (see sec. 1.3.3) to determine which model achieves the optimal trade-off between accuracy and parsimony. This will immediately provide us with an indication of whether the observed market mode, as reflected in the empirical aggregate increment $\{x_i^*\}$, should be interpreted e.g. as a common exogenous factor, as a collective endogeneous effect, or even only as the sheer outcome of chance.

The fact that the likelihoods of the biased random walk and the mean-field model depend only on $\{x_i^*\}$ and N, plus the fact that the likelihood of the uniform random walk is constant, allows us to obtain the AIC values for the three models as functions of $\{x_i^*\}$ and N only. In fig. 1.9 we show the calculated AIC weights of the three models as a function of the observed value $\{x_i^*\}$, for $N = 428$ S&P500 stocks. Each point represents a different cross section, i.e. a different day of trade, for a total of 100 randomly sampled days. It is important to note that the empirical value of the average increment only determines which point(s) of the curves are actually visited, but the curves themselves are universal.

The figure reveals us a remarkable fact, namely the presence of three distinct regimes in the behaviour of the group of stocks. For $0 \leq |\{x_i^*\}| \lesssim 0.2$, we find that the best performing model is the uniform random walk, which displays an AIC weight practically equal to one (indicating that the model is almost surely the best one among the three models considered, see sec. 1.3.3). This means that, in this 'noisy' regime, the most parsimonious explanation of the market mode, as reflected in the measured value of $\{x_i^*\}$, is that of a pure outcome of chance.

For $0.2 \lesssim |\{x_i^*\}| \lesssim 0.5$, we find that the uniform random walk is almost surely *not* the best model, while the biased random walk and mean field models are competing. We observe an almost equal performance of the two models for

$|\{x_i^*\}| \approx 0.2$, and an increasing preference for the mean field model as $|\{x_i^*\}|$ increases towards 0.5. Despite this preference, we cannot reject the mean field model, meaning that in this 'mixed' regime the most likely explanation for the market mode is a combination of exogenous and endogenous effects.

Finally, for $0.5 \lesssim |\{x_i^*\}| \lesssim 1$, the mean field model achieves practically unit probability to be the best model. In this 'endogenous' regime, the most likely explanation for the market model is uniquely in terms of a collective effect of direct influence among stocks.

We can summarize the above findings as follows:

$$\begin{cases} \text{Uncoordinated (noisy) regime:} & 0 \leq |\{x_i^*\}| \lesssim 0.2 \\ \text{Mixed (endogenous + exogenous) regime:} & 0.2 \lesssim |\{x_i^*\}| \lesssim 0.5 \\ \text{Coordinated (endogenous) regime:} & 0.5 \lesssim |\{x_i^*\}| \leq 1 \end{cases}$$

where we recall that the values of $|\{x_i^*\}|$ delimiting the various regimes have been calculated for $N = 428$.

While the qualitative finding that larger values of $|\{x_i^*\}|$ are better explained in terms of collective effects might appear intuitive, the possibility to quantitatively identify the value $|\{x_i^*\}| \approx 0.5$ above which this intuition is fully supported by statistical evidence is a non-obvious output of the above approach. The same consideration applies to the identification of the other two regimes, and of a mixed phase where there is not enough statistical evidence in favour of a single interpretation of the market mode. Moreover, the fact that the mean field model starts being statistically significant only for $|\{x_i^*\}| \gtrsim 0.2$ solves the aforementioned problem of an otherwise problematic interpretation of even tiny values of $|\{x_i^*\}|$ as the result of inter-stock interactions. The AIC analysis shows that, for values below 0.2, one should not trust the mean field model, and consequently the value $J^* > 0$ that the model itself indicates. When $|\{x_i^*\}| \lesssim 0.2$, the best model is actually the uniform random walk, which effectively corresponds to $J^* = 0$. This is a highly non-trivial result.

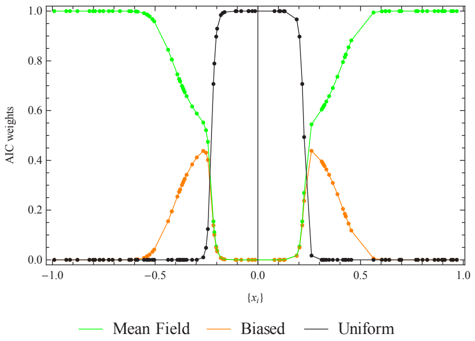

Figure 1.9: **Model performance for single cross-sections.** The calculated AIC weights of the three cross-sectional models (uniform random walk, biased random walk, mean field model) as a function of the measured average daily binary return $\{x_i^*\}$, for $N = 428$ S&P500 stocks, each studied for 100 days of trade.

1.6 Ensembles of matrices of multiple time series

In this section, as our third and final specification of the abstract formalism introduced in sec. 1.3, we extend the previous results to the general case where the observed data is a full $N \times T$ matrix $\mathbf{X}^*$ representing a set of multiple binary time series for N stocks, each extending over T timesteps. We recall that the entries of a generic such matrix $\mathbf{X}$ are denoted by $x_i(t)$, where i labels the stock and t labels the time step. We assume that N and T are both large, i.e. $N \gg 1$ and $T \gg 1$. Before introducing an explicit model, we need to make some important considerations.

We had already anticipated that the purpose of the models introduced in the previous sections was not that of introducing realistic models of financial time series. For instance, it is well known that the simple stochastic processes considered in sec. 1.4 are far too simple to reproduce some key stylized facts observed in real financial time series, such as volatility clustering [43, 44] or a bursty behaviour [50]. Moreover, being entirely binary, the above examples cannot address other well established properties characterizing the amplitude of fluctuations, e.g. the ‘fat’ (power-law) tails of the empirical distributions of price returns.

Nonetheless, there is a simple argument that legitimates us to use a proper extension of the above modelling approach, especially that introduced in sec. 1.5, provided that we adequately calibrate such extension on the observed set of multiple time series. The argument is basically the realization that we can properly model the binary signature of a time series, using temporal iterations of even the simplistic models we have introduced in sec. 1.5, if we assume that some aggregated information measured on the original ‘weighted’ time series $r_i(t)$

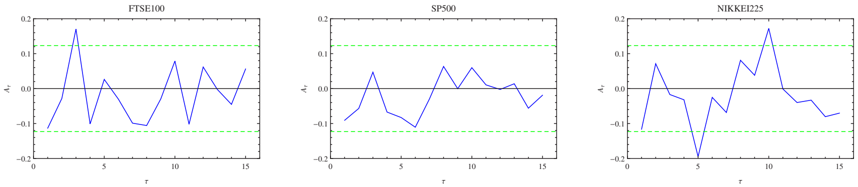

Figure 1.10: **Autocorrelation between daily cross-sections.** The measured autocorrelation of the average binary daily return $\{x_i^*(t)\}$ for the three indices in year 2006. The green lines represent the noise level, calculated as ±2 standard deviations of the Fisher-transformed autocorrelation.

$(1 \leq i \leq N)$ can be used as a proxy of the driving factor defining the model itself. We will show that this simple assumption is actually verified in the data. In particular, we will show that a sequence of temporal iterations of the biased random walk model, which assumes that the binary time series is driven by an 'external' field, can be 'bootstrapped' on the real data by assuming that the field can be replaced by a function of the (endogenous) observed aggregate increment of the original weighted time series, i.e. the empirical value $\{r_i^*\}$ of the quantity $\{r_i\}$ defined in eq. (1.3). In such a way, we do not need a model generating a realistic dynamics of $\{r_i\}$ (or of the individual stock-specific increments) in order to model the behaviour of $\{x_i\}$, because the time series of $\{r_i\}$ is taken from the data.

As a result, we will obtain an accurate model for the dynamics of the aggregate binary increment $\{x_i(t)\}$, given the observed dynamics of $\{r_i(t)\}$. This model will reproduce with great accuracy, and mathematically characterize, the empirical nonlinear relation between these two quantities that we have illustrated in sec. 1.2.2. We will finally test the temporal robustness and predictive power of the model, and conclude with discussion of the relatedness of our approach and more traditional 'factor models' in finance.

1.6.1 Temporal dependencies among cross sections

In order to execute the above plan, we first analyze the correlations between single cross sections of the market. We need this preliminary analysis in order to determine whether the temporal extension of the models defined in sec. 1.5 should incorporate dependencies among different snapshots.

Based on extensive financial literature, we expect no correlation (on a daily frequency) among the returns of different cross sections. However, most analyses focus on the autocorrelation of *individual* stocks, based on their *weighted* returns. So, to check our hypothesis we perform an explicit analysis of the temporal autocorrelation of the observed time series of the *aggregate*, *binary* return $\{x_i^*(t)\}$. This analysis is shown in fig.1.10 for the three indices, using daily data for year

2006. We confirm that the observed autocorrelation is not statistically significant, since (apart for a few points) it lies within the range of random noise (calculated by imposing a threshold of two standard deviations on the Fisher-transformed autocorrelation). This type of uncorrelated dynamics is observed throughout our dataset. This means that, in line with other analyses of autocorrelation, the memory of the aggregate binary return of real markets, if any, is much shorter than a day.

Going back to the result illustrated in fig. 1.9, we can then conclude that there is no significant correlation in the trajectories of the daily points populating the curves. In other words, given the knowledge of the position of the market in the AIC curves in a given day, we cannot predict where the market will move the next day, even if of course we know that it will move to another point in the curves themselves.

1.6.2 Reproducing the observed binary/non-binary relationships

The previous result sets the stage for our next step, where we consider an explicit extension of the models considered in sec. 1.5 to an ensemble of multiple time series, as introduced in sec. 1.3 in the general case. The absence of autocorrelation implies that we can define the Hamiltonian of the full $N \times T$ matrix $\mathbf{X}$ as a sum of T non-interacting Hamiltonians, each describing a single cross section of N stocks.

Next, we need to choose the model to extend. We want the final model to establish (among other things) an expected relationship between the binary and the weighted aggregate returns, so that we can test this prediction against the empirical relationships illustrated in sec. 1.2.2. This implies that we need to input the measured weighted return $\{r_i^*\}$ as a driving parameter of the binary model. Among the three models, only the biased random walk and the mean field model have parameters that can be related to $\{r_i^*\}$. In sec. 1.5 we treated those models as giving competing interpretations of the market model in terms of exogenous and endogenous effects respectively. However, it should be noted that this is no longer possible as soon as the parameters of these models are made dependent on the observed return. For instance, if we assume that the parameter θ of the biased random walk depends on $\{r_i^*\}$ (which is a property of the data), we can no longer interpret θ as an external field, since it has been somehow 'endogenized'. Determining whether θ can be interpreted as endogenous or exogeneous is now entirely dependent on whether $\{r_i^*\}$ itself can be interpreted as endogenous or exogeneous. This tautology does not prevent us from determining a relationship between $\{r_i^*\}$ and $\{x_i^*\}$ in their full range of variation, because such relationship is independent on the optimal (endogenous or exogenous) interpretation of both quantities.

We also note that the choice of the model to calibrate on $\{r_i\}$ is now completely independent of the relative performance of the various models that we have determined in the case of free parameters, including their AIC weights shown in fig.1.9. Indeed, apart from an initial calibration, the parameters will no longer be fitted using the maximum likelihood principle, making the AIC analysis no longer appropriate. In other words, ranking the 'free' models and endogenizing their parameters are two completely different problems. In particular, the low AIC weight of the biased random walk throughout most of fig.1.9 does not impede us from using this model in our next analysis. We will indeed 'bootstrap' the biased random walk on the real data, by looking for a relationship between $\{r_i\}$ and the parameter θ. We prefer this model over the mean field one because, while it is natural to think of (a function of) $\{r_i\}$ as a proxy of the 'field' θ affecting the market in the biased random walk model (notably, $\{r_i\}$ has a definition similar to that of a market index), it is less natural to think of the same quantity as a proxy of the inter-stock interaction J in the mean field model (although, as we said before, this would be technically possible).

Combining all the above considerations, we finally generalize the biased random walk model defined by eq. (1.59) to the matrix case as follows:

$$H(\mathbf{X},\vec{\theta}) = \sum_{t=1}^{T}\theta(t)\sum_{i=1}^{N}x_i(t) \tag{1.72}$$

where $\vec{\theta}$ it a T-dimensional vector with entries $\theta(t)$. Note that, while the models we introduced in sec. 1.4 have time-independent parameters and therefore correspond to time series at statistical equilibrium (for example a model with constant volatility), we are now considering more general models with time-dependent parameters. Relating $\theta(t)$ to $\{r_i(t)\}$ will allow us to incorporate any observed degree of non-stationarity of the data into the model itself.

As a preliminary calibration, we now look for an empirical relation between $\{r_i(t)\}$ and $\theta(t)$. To this end, we first treat the latter as a free parameter and look for the optimal value $\theta^*(t)$ maximizing the likelihood of the observed binary time series $\mathbf{X}^*$. Since the Hamiltonians for different timesteps are non-interacting, it is easy to show that $\theta^*(t)$ is given again by eq. (1.63) where $\{x_i^*\}$ is replaced by $\{x_i^*(t)\}$:

$$\theta^*(t) = -\frac{1}{2}\ln\left[\frac{1+\{x_i^*(t)\}}{1-\{x_i^*(t)\}}\right]. \tag{1.73}$$

In fig. 1.11 we compare the resulting value of $\theta^*(t)$ with the corresponding observed weighted return $\{r_i^*(t)\}$, for the three indices separately. Each point in the plot corresponds to a different day, and we considered 250 days (approximately one year) for each index. We find a strong linear relation between the two quantities. This relation can be fitted by the one-parameter curve

$$\{r_i^*(t)\} = -c\cdot\theta^*(t) \tag{1.74}$$

where $c > 0$. This finding is very important. It confirms that the parameter $\theta^*(t)$, defined through eq. (3.14) as a time-varying 'field' driving the observed binary

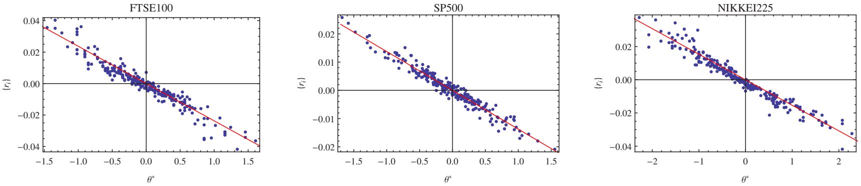

Figure 1.11: **Observed linear relation between fitted free parameter $\theta^*(t)$ and weighted property $\{r_i^*(t)\}$.** The most likely value of the driving field $\theta^*(t)$ calculated applying the biased random walk model to the projected binary signature of day t, compared with the measured average weighted return $\{r_i^*(t)\}$ of the same day, for 250 trading days (approximately one year) in the FTSE100 (left), S&P500 (center) and NIKKEI225 (right) in various years (2003, 2007, and 2004 respectively). We also show the linear fit $\{r_i(t)\} = -c\,\theta^*(t)$ with $c > 0$.

increment $\{x_i^*(t)\}$ with maximum likelihood, is an excellent proxy for the observed non-binary 'market index' $\{r_i^*(t)\}$. This result holds up to a negative factor c which, on the time scale considered, is constant for each market (in sec. 1.6.3 we will provide a more detailed analysis of the stability of c over different time scales). Since $\{r_i^*(t)\}$ is a property measured on the stock increments themselves, it reflects both external influences and internal dependencies. Therefore $\theta^*(t)$ cannot be (entirely) interpreted as an external field. This confirms our interpretation of the biased random walk as a model agnostic to the (endogenous or exogenous) nature of the driving field in the present setting.

Combining eqs.(1.73) and (1.74) together, we finally obtain a mathematical expression for the expected relationship between $\{r_i^*\}$ and $\{x_i^*\}$ in our model:

$$\{r_i^*(t)\} = \frac{c}{2}\ln\left[\frac{1+\{x_i^*(t)\}}{1-\{x_i^*(t)\}}\right] = c\cdot\operatorname{artanh}\{x_i^*(t)\}. \tag{1.75}$$

Inverting, we have

$$\{x_i^*(t)\} = \tanh\frac{\{r_i^*(t)\}}{c}. \tag{1.76}$$

We can now test the above expressions against the data shown previously in fig.1.2. In that figure, we already showed that the observed relationship between $\{r_i^*\}$ and $\{x_i^*\}$ can be fitted very well by a curve of the form given by eq. (1.75). We have just provided a theoretical justification for the otherwise arbitrary use of such expression. Moreover, now we can fit the value of c using eq. (1.74), which is independent of eq. (1.75). Once we obtain c in this way, we can use eq. (1.75) to predict $\{r_i^*(t)\}$ given $\{x_i^*(t)\}$, or *vice versa*, without fitting any parameter. In fig.1.12 we show the result of this operation. We confirm that the prediction of our model matches the empirical relationship very well.

We also consider a null model where we randomly shuffle the increments of each of the N time series independently. This results in a set of randomized time series, with elements $r'_i(t)$, where the total increment $\sum_{t=1}^{T} r'_i(t)$ for each stock is preserved, but the returns of all stocks in a given day are uncorrelated. From $r'_i(t)$, we obtain the binary signature $x'_i(t)$ as for the real data. As shown in fig.1.12, this randomized benchmark overlaps with the empirical trend only in a very narrow, linear regime. We will now try to understand this result.

The reason why the shuffled data result in a linear trend is the following. For each value of $\{x'_i\}$, there is a definite number N_{up} of 'up' stocks and a definite number $N_{down} = N - N_{up}$ of 'down' stocks, according to the relation

$$\{x'_i\} = \frac{N_{up} - N_{down}}{N} = \frac{2N_{up} - N}{N} = 2\frac{N_{up}}{N} - 1. \tag{1.77}$$

Conditional on the above value of $\{x'_i\}$, the expected value of $\{r'_i\}$ (over multiple shufflings) is

$$\langle\{r'_i\}\rangle = \frac{r^*_+ N_{up} + r^*_- N_{down}}{N} \approx \frac{r^*_+ N_{up} - r^*_+ N_{down}}{N} = r^*_+\left[2\frac{N_{up}}{N} - 1\right] = r^*_+\{x'_i\}. \tag{1.78}$$

where $r^*_+ > 0$ is the average positive increment (over all T time steps and all N time series) and $r^*_- < 0$ is the average negative increment. Note that both values coincide with the corresponding quantities in the original data, and have been denoted by a star accordingly. Assuming approximately symmetric log-return distributions for each of the N time series as typically observed, we have set $r^*_- \approx -r^*_+$. Given the overlap between real and shuffled data around zero returns in fig.1.12, we can linearize eq. (1.76) around zero and compare it with eq. (1.78) to get

$$c \approx r^*_+. \tag{1.79}$$

The above expression suggests that the value of c strongly depends on the original log-return distribution. Therefore, we expect that the stability of c is determined by that of r^*_+. In sec. 1.6.3 we will study the stability of c in more detail.

The above simple argument shows that, for shuffled data, we indeed expect a linear relationship between $\{r'_i(t)\}$ and $\{x'_i(t)\}$. This is a striking difference with respect to real data, where $\{r^*_i(t)\}$ virtually diverges as $|\{x^*_i(t)\}|$ approaches one. This 'divergence' indicates that, when most stocks are aligned in real markets ($|\{x^*_i(t)\}| \approx 1$), the observed log-returns are much larger than the typical positive increment ($|\{r^*_i(t)\}| \gg r^*_+$). In other words, extreme log-returns are more often observed when stocks are synchronized. This means that there is a strong correlation between the magnitude of log-returns of individual time series and the degree of coordination of all stocks in the market.

While for infinite realizations of the shuffling procedure we would observe eq. (1.78) extending to the full range $-1 \leq \{x'_i\} \leq +1$, for finite realizations we

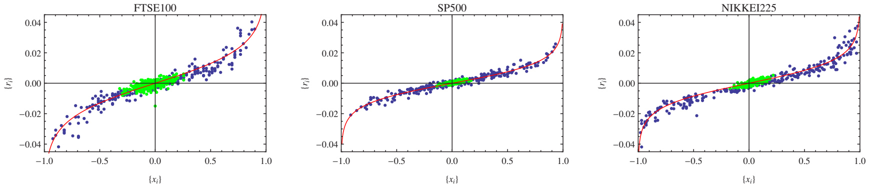

Figure 1.12: **Characterization of empirical relations between binary and non-binary properties in financial time series.** Nonlinear relationship between the average daily increment (weighted return) and the average daily sign (binary return) over all stocks in the FTSE100 (left), S&P500 (center) and NIKKEI225 (right) in various years (2003, 2007, and 2004 respectively). Here each point corresponds to one day in the time interval of 250 trading days (approximately one year). The red curve is our non-parametric prediction based on the fit shown in fig.1.11, and the green points are the same properties measured on the shuffled data.

observe a much narrower span of values (see fig.1.12). This is due to the absence of correlations among stocks, resulting in significantly lower values of both $\{r_i'\}$ and $\{x_i'\}$ with respect to the observed quantities $\{r_i^*\}$ and $\{x_i^*\}$. Interestingly enough, for the S&P500 index the randomized data span the range $|\{x_i'\}| \lesssim 0.2$, which coincides precisely with the regime we identified in fig.1.9 for a completely noisy-driven system with the same number of stocks. This confirms that the AIC analysis correctly pinpoints the boundaries outside which one should expect the observed value $\{x_i^*\}$ to be inconsistent with a typical realization of N purely random variables.

The above results also provide an explanation for the second empirical nonlinear relation that we had documented in sec. 1.2.2, i.e. the one between $\{r_i^*(t)r_j^*(t)\}$ and $\{x_i^*(t)\}$ (see fig. 1.3). In general, we can write $\{r_i r_j\}$ as

$$\{r_i r_j\} = \frac{1}{N(N-1)} \sum_{i \neq j} r_i r_j = \frac{1}{N(N-1)} \left[\left(\sum_{i=1}^{N} r_i \right)^2 - \sum_{i=1}^{N} r_i^2 \right]. \tag{1.80}$$

The term $\sum_{i=1}^{N} r_i^2$ is of order N, and vanishes for large markets when divided by $N(N-1)$. We are therefore left with

$$\{r_i r_j\} \approx \frac{1}{N(N-1)} \left(\sum_{i=1}^{N} r_i \right)^2 \approx \{r_i\}^2. \tag{1.81}$$

Using eq.(1.75), we get

$$\{r_i^*(t) r_j^*(t)\} \approx \{r_i^*(t)\}^2 = c^2 \text{artanh}^2 \{x_i^*(t)\} \tag{1.82}$$

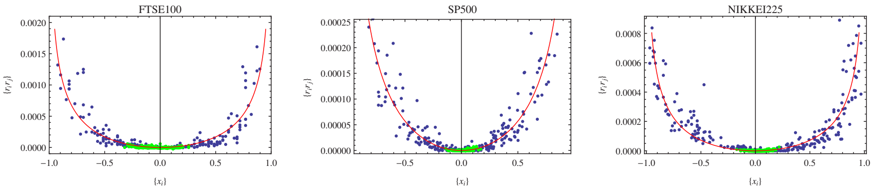

Figure 1.13: **Characterization of empirical relations between binary and second-order non-binary properties in financial time series.** Nonlinear relationship between the average daily coupling (weighted coupling) and the average daily sign (binary return) over all stocks in the FTSE100 (left), S&P500 (center) and NIKKEI225 (right) in various years (2003, 2007, and 2004 respectively). Here each point corresponds to one day in the time interval of 250 trading days (approximately one year). The red curve is our non-parametric prediction based on the fit shown in fig.1.11, and the green points are the same properties measured on the shuffled data.

which theoretically justifies the fitting function we had used in fig.1.3. Again, rather than fitting that curve on the data, we can use the value of c determined from the (independent) fit shown in fig.1.11. This results in the non-parametric plot shown in fig. 1.13. We confirm that, for each of the three indices, we can reproduce the observed relationship very well.

As before, we also show the relationship between $\{r'_i(t)r'_j(t)\}$ and $\{x'_i(t)\}$ for randomly shuffled data. The linearity of eq. (1.78) now translates into an expected parabolic relationship:

$$\langle\{r'_i r'_j\}\rangle \approx \{r'_i\}^2 = (r^*_+)^2\{x'_i\}^2. \tag{1.83}$$

Again, real data strongly deviate from the above 'uncorrelated' parabolic expectation, because extreme events make the empirical coupling $\{r^*_i r^*_j\}$ virtually 'diverge' when stocks are highly synchronized ($|\{x^*_i\}| \approx 1$).

1.6.3 Stability of the parameter c

Once we have mathematically characterized the observed nonlinear relations, an unavoidable question arises: in a given market, how stable are those relations? Since c is the only parameter in the above analysis, the question simply translates into the stability of c. We have already noted that c is related to the average positive return r^*_+, which we expect to be relatively stable. In order to study the stability of c in more detail, we now consider several yearly and monthly time windows, and explore the time evolution of the fitted parameter for the three indices.

In fig. 1.14 (upper panels) we plot the values of the parameter c (with error bars) for 11 yearly snapshots (2001-2010). It is clear that there are periods during which the yearly values are relatively stable, and periods when they fluctuate wildly. Thus, in most cases the fitted value of c in a given year does not allow to make predictions about the value of c int the next year.

However, we can also consider a monthly frequency. In the bottom panels of fig. 1.14 we show the result of our analysis, when carried out on the 12 monthly snapshots of year 2006. We choose this particular year because, in the yearly trends shown above, it represents very different points for different markets: the end of a stable period for the FTSE100, an exceptional jump for the S&P500, and the middle of an increasing trend for the NIKKEI225. Despite these differences, we find that in all three markets the monthly dynamics is much more stable than the yearly one. In particular, the trends for FTSE100 and NIKKEI225 are almost constant, and for the S&P500 there are only two deviating points from an otherwise stable trend (despite the large fluctuation that 2006 represents in the yearly trend for this index). This implies that, in most cases, one might even use the monthly value of c out of sample, in order to predict the future relationship between $\{x_i\}$ and $\{r_i\}$ based on a past observation. We should however stress that the aim of our method is to characterize such relationship, and not to predict it. Indeed, we cannot imagine any situation in which only the binary (or only the non-binary) information is available.

The above results show that there is a trade-off between short and long periods of time. For short (e.g. monthly) periods there are less points to calculate c through a fit of the type shown in fig.1.11. This explains why the monthly trends in fig.1.14 have bigger error bars than the yearly trends in the same figure. By contrast, for longer (e.g. yearly) periods each individual fit is better, but there are more fluctuations in the temporal evolution of the parameter c, because the data are less stationary. In general, we expect that in each market, and for a specific period of time, there is a different 'optimal' frequency to consider.

1.6.4 Relation to factor models

We would like to conclude this chapter with a discussion of the relationship between some of our findings and the popular *factor models* in the financial literature [3]. As a basic consideration, we stress that factor models can only be applied to the original (non-binary) increments (it is impossible to decompose a binary signal into a nontrivial combination of binary signals), while our models only apply to the binary projections. We should bear this irreducible difference in mind in what follows. However, due to the mapping between binary and non-binary increments that we have documented, we can try to indeed relate the two approaches.

First, let us consider the shuffled (uncorrelated) data, where the original log-returns are randomly permuted within each of the N time series. It is well known that the total temporal increment (over T time steps) of any empirical time series of price increments is generally close to zero (due to market efficiency), and that the distribution of log-returns is mostly symmetric around this value. This is especially true if each of the N original time series has been separately standardized,

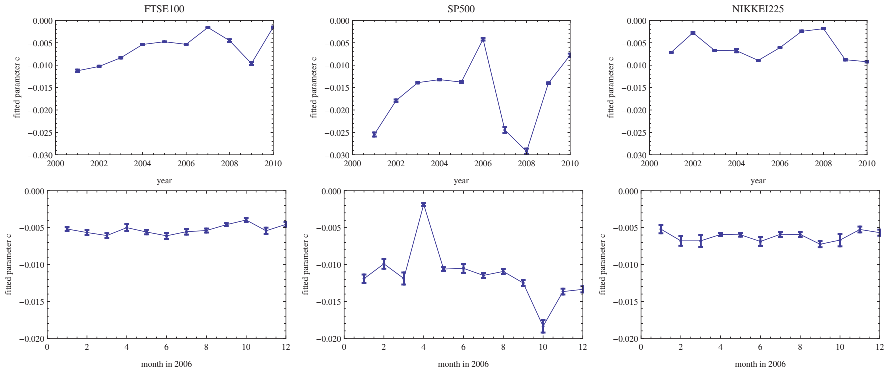

Figure 1.14: **Stability analysis of the parameter c.** Parameter c fitted as in fig.1.11 on various yearly (top panels) and monthly (bottom panels) snapshots of the market, for the FTSE100 (left), S&P500 (center) and NIKKEI225 (right).

i.e. the i-th temporal average has been subtracted from each increment of the i-th time series, and the result has been divided by the i-th standard deviation. In such a case, the N log-return distributions become also very similar to each other, because their support is the same and their values are comparable. This means that, after the shuffling, the time series are sequences of independent and almost identically distributed variables with zero mean. We denote the corresponding increments as

$$r_i(t) = \epsilon_i(t) \quad \forall i, \tag{1.84}$$

where the ϵ_i's are random variables. In a traditional factor analysis, the above scenario takes the form of a 'zero-factor' model. Under this model, the aggregate increment over N stocks is expected to be narrowly distributed around

$$\{r_i(t)\} = \frac{1}{N}\sum_{i=1}^{N} \epsilon_i(t) \approx 0. \tag{1.85}$$

When $\{r_i(t)\}$ takes small values around zero, we know from fig. 1.12 that $\{x_i(t)\}$ also takes small values around zero. Indeed, shuffled time series are in the linear regime that spans the range where the binary increment $\{x_i(t)\}$ is consistent with a uniform random walk (see fig.1.9). Therefore we find that the zero-factor model (for the non-binary returns) and the uniform random walk (for the binary returns) are consistent with each other in the linear regime. In other words, when in our analysis we measure a value of $\{x_i(t)\}$ that is consistent with a uniform random walk, we know that the original log-returns are consistent with a zero-factor model.

Next, we consider a one-factor model, where there is one dominant underlying factor assumed to control the dynamics of all the time series. In such a case, each return can be decomposed as

$$r_i(t) = \alpha_i \Phi_0(t) + \epsilon_i(t) \quad \forall i, \tag{1.86}$$

where α_i is the 'factor loading' of the i-th time series with the dominant factor $\Phi_0(t)$. When referring to stocks, the factor $\Phi_0(t)$ is attributed to the market mode. It is known that, during crisis times when the markets are highly correlated, a one-factor model can describe the dynamics quite well. Under this model, the aggregate increment is

$$\{r_i(t)\} = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \Phi_0(t) + \frac{1}{N}\sum_{i=1}^{N} \epsilon_i(t) \approx \{\alpha_i\}\Phi_0(t) \tag{1.87}$$

where $\{\alpha_i\} \equiv \frac{1}{N}\sum_{i=1}^{N} \alpha_i$ is the average loading, which is independent of both i and t. This result implies that, when the market is well described by a one-factor model, the average increment $\{r_i(t)\}$ that we measure in our analysis is proportional to the factor $\Phi_0(t)$ itself. We note that the one-factor model is somehow similar to our biased random walk model, as it assumes a common drive for all the stocks. However, since $\Phi_0(t)$ is fitted on the data, the one-factor model cannot distinguish between an endogenous or exogenous nature of the common drive. This situation is similar to when we use the observed value of $\{r_i(t)\}$ as the driving field of the biased random walk (see sec. 1.6.2).

In financial analysis, the factor model can be used to filter the original time series and remove the one-factor component from them. When the model is a good approximation to the real market, the filtered returns are $r_i(t) \approx \epsilon_i(t)$, leading us back to eq. (1.84) and the related considerations. In such a scenario, there is no correlation among the stocks, and each stock is acting as an i.i.d. variable. We therefore expect that, if we remove the market mode from the original time series, then (in periods where the market is indeed dominated by a single factor) we would obtain results similar to the shuffled case, and we would find the system in the uncoordinated phase of fig.1.9.

However, despite the fact that in certain conditions the one-factor model can generate the market behaviour, the model is too simplistic [3]. In reality the dynamics is more complex and can be attributed to many factors, that sometimes overlap with industrial (sub)sectors. Generally the different factors are identified by the largest, non-random eigenvalues of the empirical cross-correlation matrix, where the market mode relates to the highest eigenvalue [3]. The presence of many deviating eigenvalues is an indication of the fact that the one-factor model should be rejected. A more realistic, M-factor model is

$$r_i(t) = \sum_{j=0}^{M} \alpha_{ij} \Phi_j(t) + \epsilon_i(t) \quad \forall i, \tag{1.88}$$

where $j = 0$ denotes a common market-wide factor as above, while $j > 0$ denotes sector-specific factors. In such a case, our measured value of $\{r_i(t)\}$ is

$$\{r_i(t)\} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\alpha_{ij}\Phi_j(t) + \frac{1}{N}\sum_{i=1}^{N}\epsilon_i(t) \approx \sum_{j=1}^{M}\{\alpha_{ij}\}\Phi_j(t) \tag{1.89}$$

which is a linear combination of the multiple factors controlling the market dynamics.

It should be noted that factor models cannot distinguish between an endogenous and exogenous origin for the factors $\Phi_j(t)$ themselves, even if we invoke some information-theoretic criterion to rank different specifications of these models. By contrast, our binary models allow us to discriminate among these multiple scenarios, as we have shown in fig.1.9 and related discussions. Moreover, while our approach allows us to relate binary and non-binary increments of real time series and replicate the observed relationships among them (see figs.1.12 and 1.13), factor models cannot lead to a similar result, because they do not allow for a binary description.

1.7 Conclusions

In this chapter we presented a novel method for the analysis of single and multiple binary time series. Our information-theoretic approach allowed us to extract and quantify the amount of information encoded in simple, empirically measured properties. This resulted in the possibility to associate an entropy value to a time series given its measured properties, and to compare the informativeness of different measured properties.

By employing our formalism, we have identified distinct regimes in the collective behaviour of groups of stocks, corresponding to different levels of coordination that only depend on the average return of the binary time series. In each regime the market exhibits a dominant character: the market mode can be interpreted as an exogenous factor, as pure noise, or as a combination of endogenous and exogenous components. Moreover, each regime is characterized by the most informative property.

Finally and more importantly, we were able to replicate the observed nonlinear relations between binary and non-binary aggregate increments of real multiple time series. We have mathematically characterized these relations accurately, and interpreted them as the result of the fact that very large log-returns occur more often when most stocks are synchronized, i.e. when their increments have a common sign. Our findings suggest that the binary signatures carry significant information, and even allow to measure the level of coordination in a way that is unaccessible to standard non-binary analyses.

Bibliography

[1] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlation and Complexity in Finance* (Cambridge University Press, Cambridge, U.K., 1999).

[2] Sitabhra Sinha, Arnab Chatterjee, Anirban Chakraborti, Bikas K. Chakrabarti Econophysics: An Introduction (Wiley-VCH 2010)

[3] J.-P Bouchaud and M. Potters, *Theory of Financial Risk: From Statistical Physics to Risk Management* (Cambridge University Press, Cambridge, England, 2000).

[4] J. P. Bouchaud, Power laws in economics and finance: some ideas from physics. *Quantitative Finance* **1**, 105 (2001).

[5] R. N. Mantegna *et al.*, Nature, **376**, 46 (1995); R. N. Mantegna *et al.*, Nature, **383**, 587 (1996); V. Plerou *et al.*, Nature,**421**, 130 (2003); X. Gabaix *et al.*, Nature, **423**, 267 (2003);

[6] Schneidman, E., Berry, M.J., Segev, R., and Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012 (2006).

[7] Dal'Maso Peron, T. & Rodrigues, F. Collective behavior in financial markets. *EPL (Europhysics Letters)* **96** , 48004,(2011).

[8] Mantegna, R. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems* **11**, 193-197, (1999).

[9] Onnela, J., Chakraborti, A., Kaski, K., and Kertesz, J. Dynamic asset trees and black Monday. *Physica A: Statistical Mechanics and its Applications* **324**, 247-252 (2003).

[10] La Spada, G., Farmer,J., and Lillo, F. The non-random walk of stock prices:The long-term correlation between signs and sizes *Eur. Phys. J. B* **64**, 607-614, (2008).

[11] Alexander M. Petersen, Fengzhong Wang, Shlomo Havlin, and H. Eugene Stanley Quantitative law describing market dynamics before and after interest rate change *Physical Review E* **81**, 066121, (2010).

[12] F. Black, Studies of stock price volatility changes *Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section*, 177, (1976).

[13] T. G. Andersen, T. Bollerslev, F. X. Diebold, and C. Vega, Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange. *Amer. Econ. Rev.* **93**, 38 (2003).

[14] G. Bekaert and G. Wu, Asymmetric volatility and risk in equity markets. *Rev. Fin. Stud.* ,13:1, (2000).

[15] J.-P. Bouchaud, A. Matacz, and M. Potters. Leverage Effect in Financial Markets: The Retarded Volatility Model. *Physical Review Letters*, **87** (22):228701, (2001).

[16] M. Boguna, M. A. Serrano. Generalized percolation in random directed networks. *Physical Review E*, **72** 016106, (2005).

[17] J. Hertz, Y. Roudi, and J. Tyrcha. Ising Models for Inferring Network Structure From Spike Data. Xiv:1106.1752 [q-bio.QM].

Neural Comput. 2003 Feb;15(2):309-29. The dynamic neural filter: a binary model of spatiotemporal coding. Quenet B1, Horn D.

[18] Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **106**, 620 (1957).

[19] Jaynes, E.T. Information Theory and Statistical Mechanics 2. *Phys. Rev.* **108**, 171-190 (1957).

[20] C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal* ,**27** ,379-423,623-656 (1948)

[21] J.Park and M.E.J.Newman. The statistical mechanics of networks *Phys. Rev. E* **70** 066117 (2004)

[22] Diego Garlaschelli and Maria I. Loffredo. Maximum likelihood: Extracting unbiased information from complex networks. *Physical Review E* **78** 015101(R),(2008)

[23] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, (6): 716-723. DOI:10.1109/TAC.1974.1100705. MR 0423716.(1974)

[24] Burnham, Kenneth P., Anderson, David R. Model selection an multi model inference, 2nd edition,(Springer 2002)

[25] Stanley Wasserman. Social Network Analysis: Methods and Applications. (Cambridge University Press 1994)

[26] M. L. Metha, Nucl. Phys., 18 395 (1960); M. L. Mehta and F. J. Dyson, J. Math. Phys. 4, 713 (1963); M. L. Metha, Commun. Math. Phys., 20 245 (1971);

[27] M. L. Mehta, Random Matrices (Academic Press, Boston, 1991).

[28] E. P. Wigner, Ann. Math. 53, 36 (1951); E. P. Wigner, Proc. Cambridge Philos. Soc. 47, 790 (1951).

[29] F. J. Dyson, J. Math. Phys. 3, 140 (1962); F. J. Dyson and M. L. Mehta, J. Math. Phys. 4, 701 (1963).

[30] A. M. Sengupta and P. P. Mitra, Distributions of singular values for some random matrices *Phys. Rev. E*, **60**, 3 (1999).

[31] L. Laloux, P. Cizeau, J-P. Bouchaud, and M. Potters, Noise Dressing of Financial Correlation Matrices *Phys. Rev. Lett.* **83**, 1467 (1999)

[32] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E.Stanley, Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series *Phys. Rev. Lett.* **83** 1471 (1999).

[33] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, A random matrix theory approach to financial cross-correlations *Physica A* **287**, 374 (2000).

[34] V. Plerou, P. Gopikrishnan, B. Rosennow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, A Random Matrix Approach to Cross-Correlations in Financial Data *Phys. Rev. E* **64**, 066126 (2002).

[35] P. Gopikrishnan, B. Rosenow, V. Plerou, and H. E. Stanley, Quantifying and interpreting collective behavior in financial markets *Phys. Rev. E* **64**, 035106 (2001).

[36] B. Rosenow, P. Gopikrishnan, V. Plerou, and H. E. Stanley, Random magnets and correlation of stock price fluctuations *Physica A* **314**, 762-767 (2002).

[37] L. Laloux, P. Cizeau, J-P. Bouchaud, and M. Potters, Noise Dressing of Financial Correlation Matrices *Phys. Rev. Lett.* **83**, 1467 (1999)

[38] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E.Stanley, Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series *Phys. Rev. Lett.* **83** 1471 (1999).

[39] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, A random matrix theory approach to financial cross-correlations *Physica A* **287**, 374 (2000).

[40] V. Plerou, P. Gopikrishnan, B. Rosennow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, A Random Matrix Approach to Cross-Correlations in Financial Data *Phys. Rev. E* **64**, 066126 (2002).

[41] P. Gopikrishnan, B. Rosenow, V. Plerou, and H. E. Stanley, Quantifying and interpreting collective behavior in financial markets *Phys. Rev. E* **64**, 035106 (2001).

[42] B. Rosenow, P. Gopikrishnan, V. Plerou, and H. E. Stanley, Random magnets and correlation of stock price fluctuations *Physica A* **314**, 762-767 (2002).

[43] A. Sato , H. Takayasu Dynamic numerical models of stock market price: from microscopic determinism to macroscopic randomness *Physica A* **250**, 231-252 (1998).

[44] A. Krawieckia,b, J. A. Ho lysta,b, and D. Helbingb *Phys. Rev. Lett.* **89**, 158701 (2002).

[45] T. Squartini and D. Garlaschelli Analytical maximum-likelihood method to detect patterns in real networks *New Journal of Physics* **13**, 083001 (2011).

[46] T Squartini, G Fagiolo, D Garlaschelli Randomizing world trade. I. A binary network analysis *Phys. Rev. E* **84**, 046117 (2011).

[47] T Squartini, G Fagiolo, D Garlaschelli Randomizing world trade. II. A weighted network analysis *Phys. Rev. E* **84**, 046118 (2011).

[48] A. N.Kolmogorov Randomizing world trade. II. A weighted network analysis *lEEE Trans.Inform. TheoryI* **IT14**,662 (1965).

[49] Borst, A.; Theunissen, F.E Information theory and neural coding *Nat. Neurosci* **2**,947–957 (1999).

[50] V. Gontis and B. Kaulakys Multiplicative point process as a model of trading activity *Physica A* **343**,. 505-514, (2004).

[51] R. Startz Binomial autoregressive moving average models with an application to U.S. recession *J. Bus. Econom. Statist.* **26**,. 1-8, (2008).

[52] R. Startz Binomial autoregressive moving average models with an application to U.S. recession *J. Bus. Econom. Statist.* **26**,. 1-8, (2008).

[53] H. Kauppi and P. Saikkonen Predicting U.S. recessions with dynamic binary response models *Rev. Econ. Stat.* **90(4)**, 777-791, (2008).

[54] A. Bunde, S. Havlin Fractals in science (Springer, Berlin 1994).

[55] Rodney J. Baxter Exactly Solved Models in Statistical Mechanics (Dover publications 2007).

Chapter 2

Economic Networks

The International Trade Network (ITN) is a complex network formed by the bilateral trade relations between world countries. The network complex structure reflects important economic processes such as globalization, and the propagation of shocks and instabilities. Both from the perspective of network theory and macroeconomics, understanding the structure of the ITN is of paramount importance: in particular, understanding which economic quantities play a role in shaping the ITN structure is crucial. Traditional macroeconomics has mainly used the Gravity Model to characterize the magnitude of trade volumes, using macroeconomic properties such as GDP and geographic distance. On the other hand, recent maximum-entropy network models successfully reproduce the complex topology of the ITN, but provide no information about trade volumes. In this chapter, we first discuss the role played by the countries GDP in determining both the presence and the amount of trade exchanges between world countries. Next, we make an effort to integrate these two currently incompatible approaches via the introduction of two GDP driven models. We introduce a novel ingredient that we denote as 'topological invariance', i.e. the invariance of the expected topology under an arbitrary change of units of trade volumes. Via this unified and principled mechanism, which is transparent enough to be generalized to any economic network, the models provide a new econometric framework wherein trade probabilities and trade volumes can be separately controlled by macroeconomic variables. The models successfully reproduce both the topology and the weights of the ITN, finally reconciling the conflicting approaches.

The results presented in this chapter have been published in the following references:
A. Almog, T. Squartini, D. Garlaschelli *New Journal of Physics*, **17**: 013009 (2015).
A. Almog, T. Squartini, D. Garlaschelli *Int. J. Computational Economics and Econometrics*, in press, arXiv:1512.02454 (2016).
A. Almog, R. Bird, D. Garlaschelli *arXiv:1506.00348* (2016).

2.1 Introduction

The bilateral trade relationships existing between world countries form a complex network known as the International Trade Network (ITN). The observed complex structure of the system is at the same time the outcome and the determinant of a variety of underlying economic processes, including economic growth, integration, and globalization. Moreover, recent events such as the financial crisis clearly pointed out that the interdependencies between financial markets can lead to cascading effects which, in turn, can severely affect the real economy. International trade plays a significant role among the possible channels of interaction among countries [1, 2, 3, 4], thereby possibly further propagating these cascading effects worldwide and adding one more layer of contagion. Characterizing the global networked economy is, therefore, an important open problem and modelling the ITN represents a crucial step of this challenge [5, 11, 13, 16, 17, 24, 25].

Historically, macroeconomic models have mainly focused on modelling the trade volumes between countries. The Gravity Model, which was introduced in the early 60's by Jan Tinbergen [30], serves as a robust empirical model that aims at predicting the bilateral trade flow between any two (trading) countries based on the knowledge of their Gross Domestic Product (GDP) and mutual geographic distance. Although the model has been upgraded, over the years, to include other possible factors of macroeconomic relevance, like common language and trade agreements, GDP and distance remain the two factors with the largest explanatory power. The gravity model can reproduce the observed trade volumes between countries satisfactorily. However, at least in its simplest and most widespread implementation, the model cannot account for zero volumes, therefore predicting a fully-connected trade network. This outcome is entirely inconsistent with the observed, heterogeneous, topology of the ITN, which represents the backbone on which trade is made. Subsequent refinements of the gravity model allowing for zero trade flows succeeded only in reproducing the total number of missing links, not their position in the trade network, thereby producing sparser but still non-realistic topologies [14, 15].

The sharp contrast between the observed topological complexity of the ITN and the homogeneity of the network structure generated by the GM (including its recent extensions) calls for significant improvements in the modelling approach. Important steps have been made using network models [7, 38, 39, 23, 26], among which maximum-entropy techniques [18, 19, 20] have been proven to be particularly advantageous. Maximum-entropy models aim at reproducing higher-order structural properties of real-world networks using lower-order information (more precisely, node-specific), which is constrained to be reproduced [27, 28, 29]. Important examples of local properties that can be chosen as constraints are the *degree*, i.e. the number of links of a node (in the ITN case, this is the number of trade partners of a country) and the *strength*, i.e. the total weight of the links of a node (in the ITN case, this is the total trade volume of a country). Examples of higher-order properties that the method aims at reproducing are the *clustering coefficient*, which refers to the fraction of realised triangles around nodes, and the *degree correlations*. The studies have focused on both binary and weighted representations of the ITN, i.e. the two representations defined by the *existence* and by the *magnitude* of trade exchanges among countries, respectively. In principle, depending on which local properties are chosen as constraints, maximum-entropy models can either fail or succeed in replicating the higher-order properties of the ITN.

Importantly, the use of maximum-entropy models has lead to a counter-intuitive result about the structure of the trade network: contrary to what naive economic reasoning would predict, controlling for *purely binary* local properties (node degrees) is more informative than controlling for the corresponding weighted properties (node strengths). In other words, while the binary network reconstructed only from the knowledge of the degrees of all countries is topologically very similar to the real ITN, the weighted network reconstructed only from the strengths of all countries is very different (typically much denser) from the real network [21, 22, 19]. This is somewhat surprising, given that economic theory assumes that weighted properties (the total trade volume of a country) are *per se* more informative than the corresponding binary ones (the number of trade partners of a country). This empirical puzzle calls for a theoretical explanation and has generated further interest in the challenge of finding a unique mechanism predicting link probabilities and link weights simultaneously. In this chapter, we will propose different models that successfully implement such mechanism. The models can reproduce the observed properties of the ITN and finally highlight a clear mathematical explanation for the observed binary/weighted asymmetry.

Our approach builds on previous theoretical results. Recently, an improved reconstruction approach [32], based on an analytical maximum-likelihood estimation method [18], has been proposed in order to define more sophisticated maximum-entropy ensembles of weighted networks. This approach exploits previous mathematical results [34] characterizing a network ensemble where both the degree and the strength sequences are constrained. The graph probability is the so-called generalized Bose-Fermi distribution [34], and the resulting network model goes under the name of Enhanced Configuration Model (ECM) [32]. When used to reconstruct the properties of several empirical networks, the ECM shows a significant improvement with respect to the case where either only the degree sequence (Binary Configuration Model, BCM for short) or only the strength sequence (Weighted Configuration Model, WCM for short) is constrained. One, therefore, expects that combining the knowledge of strengths and degrees is precisely the ingredient required to reproduce the ITN from purely local information successfully. Indeed, a more recent study has shown that, when applied to international trade data (both aggregated and commodity-specific), the method successfully reproduces the key properties of the ITN, across different years and for different levels of aggregation (i.e. for various commodity-specific layers) [33].

However, in itself, the ECM is a general network reconstruction method, rather than a true structural model. To transform it into a proper network model for the ITN structure, it would be necessary to find a macroeconomic interpretation for the underlying variables involved in the method. This operation would correspond to what has already been separately performed in different studies. For the binary level, a strong relationship between the GDP and the variable controlling the degree of a country in the BCM [7, 36] was identified. At the purely weighted level, a similar relationship was found between the GDP and the variable controlling the strength of a country in the WCM [23], in the same spirit of the gravity model. Here, we aim at generalizing the results, to one model that is able to generate both strengths and degrees.

We start with a summarized review of previous maximum-entropy approaches to the characterization of the ITN, thus explaining the mathematical building blocks on which we build our unifying approach. The rest of the chapter is organized as follows. In section 2.4 we discuss the macroeconomic approach to model the ITN, in particular, the Gravity Model of Trade. We also present various empirical relations existing between the GDP and a range of country-specific properties. These results suggest a justification for the use of GDP as an empirical fitness to be used in maximum-entropy models. In section 2.5 we introduce a novel, GDP-driven, two-step model that successfully reproduces the binary *and* the weighted higher-order properties of the ITN simultaneously. In section 2.6 we introduce what we call the Enhanced Gravity Model (EGM) of trade, which represents a new, generalized model combining maximum-entropy network models with economically established gravity-like models. The model overcomes the limitations of the existing approaches and retains the power of the popular GM in reproducing trade volumes via geographic distances and GDPs. While at the same time dramatically improving its network properties by reproducing both first-order properties, such as node degree and node strength, and higher-order properties, such as assortativity and clustering (in both binary and weighted representations of the network). Finally, in section 2.7 we summarize our results and provide some conclusions.

2.2 Data

We have used data from the Gleditsch database which spans over the years 1950-2000 [9], and from the United Nations Commodity Trade Database (UN COMTRADE) [10] from the year 1992 to 2004. We use yearly bilateral data on exports and imports Here we analyze the aggregated level, which results in yearly temporal snapshots of undirected total trade flows. The data sets are available in the form of weighted matrices of bilateral trade flows w_{ij}, the associated adjacency matrices a_{ij} and vectors of GDPs. There are approximately 200 countries in the data set covering the considered 51 years; the GDP is measured in U.S. dollars.

This data set was the subject of many studies exploring both purely the binary representation, and its full weighted representation [18, 32, 33]. Another data set which is widely used to represent the ITN network is the trade data collected by Gleditsch [9]. The data contain the detailed list of bilateral import and export volumes, for each country in the period 1950-2000.

2.3 Maximum-entropy approaches to the international trade network

Since our results are a generalization of previous maximum-entropy approaches to the characterization of the ITN, in this section we first briefly review the main results of those approaches, while our new findings are presented in the next section. In doing so, we gradually introduce the mathematical building blocks of our analysis and illustrate our main motivations. Moreover, since previous studies have used different data sets, we also recalculate the quantities of interest on the same data set that we will use later for our own investigation. This allows us to align the results of previous approaches and properly compare them with our new findings.

2.3.1 Binary structure

If one focuses solely on the binary undirected projection of the ITN, then the Binary Configuration Model (BCM) represents a very successful maximum-entropy model. In the BCM, the local knowledge of the number of trade partners of each country, i.e. the degree sequence, is specified. It has been shown that higher-order properties of the ITN can be simply traced back to the knowledge of the degree sequence [21]. This result adds considerable information to the standard results of traditional macroeconomic analyses of international trade. In particular, it suggests that the degree sequence, which is a purely topological property, needs to be considered as an important target quantity that international trade models, in contrast with the mainstream approaches in economics, should aim at reproducing [19].

Let us first represent the observed structure of the ITN as a weighted undirected network specified by the square matrix $\mathbf{W}^*$, where the specific entry w^*_{ij} represents the weight of the link between country i and country j. Then, let us represent the binary projection of the network in terms of the binary adjacency matrix $\mathbf{A}^*$, with entries defined as $a^*_{ij} = \Theta[w^*_{ij}]$, $\forall\, i, j$, where Θ is a Heaviside step function. A maximum-entropy ensemble of networks is a collection of graphs where each graph is assigned a probability of occurrence determined by the choice of some constraints. The BCM is a maximum-entropy ensemble of binary graphs, each denoted by a generic matrix $\mathbf{A}$, where the chosen constraint is the degree sequence. In the canonical formalism [18], the latter can be constrained by writing the following Hamiltonian:

$$H(\mathbf{A}) = \sum_{i=1}^{N} \theta_i k_i(\mathbf{A}) \tag{2.1}$$

where the degree sequence is defined as $k_i(\mathbf{A}) = \sum_{j\neq i}^{N} a_{ij} = \sum_{j\neq i}^{N} \Theta[w_{ij}]$, $\forall\, i$, and θ_i are the free parameters (Lagrange multipliers) [18]. As a result of the constrained maximization of the entropy [18], the probability of a given configuration $\mathbf{A}$ can be written as

$$P(\mathbf{A}) = \frac{e^{-H(\mathbf{A})}}{\sum_{\mathbf{A}'} e^{-H(\mathbf{A}')}} = \prod_{i<j} \left[\frac{z_i z_j}{1+z_i z_j}\right]^{a_{ij}} \left[\frac{1}{1+z_i z_j}\right]^{1-a_{ij}} \tag{2.2}$$

where $z_i \equiv e^{-\theta_i}$ and $p_{ij} \equiv \frac{z_i z_j}{1+z_i z_j}$. The latter represents the probability of forming a link between nodes i and j, which is also the expected value

$$\langle a_{ij} \rangle = \frac{z_i z_j}{1+z_i z_j} = p_{ij}. \tag{2.3}$$

According to the maximum-likelihood method proposed in [18], the vector of unknowns $\vec{z}$ can be numerically found by solving the system of N coupled equations

$$\langle k_i \rangle = \sum_{j\neq i}^{N} p_{ij} = k_i(\mathbf{A}^*) \quad \forall i \tag{2.4}$$

where the expected value of each degree k_i is matched to the observed value $k_i(\mathbf{A}^*)$ in the real network $\mathbf{A}^*$. The (unique) solution will be indicated as $\vec{z}^*$. When inserted back into eq.(2.3), this solution allows us to analytically describe the binary ensemble matching the observed constraints. Being the result of the maximization of the entropy, this ensemble represents the least biased estimate of the network structure, based only on the knowledge of the empirical degree sequence.

In fig. 2.3 we plot some higher-order topological properties of the ITN as a function of the degree of nodes, for the 2002 snapshot. These properties are the so-called average nearest neighbour degree and the clustering coefficient. For both quantities, we plot the observed values (red points) and the corresponding expected values predicted by the BCM (blue points). The exact expressions for both empirical and expected quantities are provided in the Appendix. We see that the expected values are in very close agreement with the observed properties. These results replicate recent findings [19] based on the same UN COMTRADE

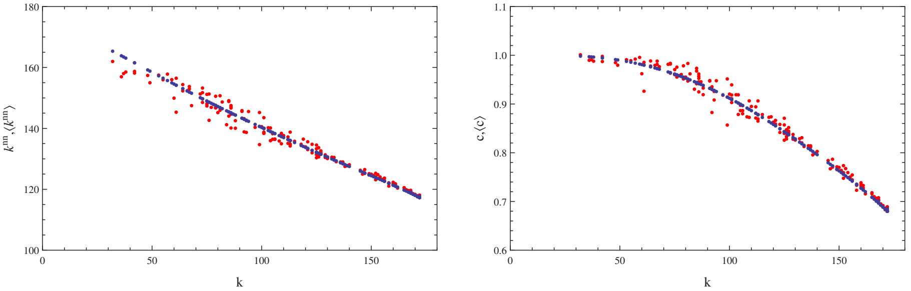

Figure 2.1: **Binary Configuration Model, reconstruction of higher-order properties.** between the observed undirected binary properties (red points) and the corresponding ensemble averages of the BCM (blue points) for the aggregated ITN in the 2002 snapshot. Left panel: Average Nearest Neighbor Degree k^{nn} versus degree k_i. Right panel: Binary Clustering Coefficient C_i versus degree k_i. The figure shows that the expected values are in very close agreement with the observed properties.

data. They show that at a binary level, the degree correlations (disassortativity) and clustering structure of the ITN are excellently reproduced by the BCM. As we also confirmed in the present analysis, these results were found to be very robust, as they hold true over time and for various resolutions (i.e., for different levels of aggregation of traded commodities) [19].

Relation with the fitness Model

It should be noted that eq.(2.3) can be thought of as a particular case of the so-called *Fitness Model* [35], which is a popular model of binary networks where the connection probability p_{ij} is assumed to be a function of the values of some 'fitness' characterizing each vertex. Indeed, the variables $\vec{z}^*$ can be treated as fitness parameters [7, 36] which control the probability of forming a link. A very interesting correlation between a fitness parameter of a country (assigned by the model) and the GDP of the same country has been found [36]. This relation is replicated here in fig. 2.4, where the rescaled GDP of each country ($g_i \equiv \frac{GDP_i}{\sum_i GDP_i}$) is compared to the value of the fitness parameter z_i^* obtained by solving eq.(2.4). The red line is a linear fit of the type

$$z_i = \sqrt{a} \cdot g_i. \tag{2.5}$$

This leads to a more economic interpretation where the fitness parameters can be replaced (up to a proportionality constant) with the GDP of countries, and used to reproduce the properties of the network. This procedure, first adopted in [7], can give predictions for the network based only on macroeconomic properties

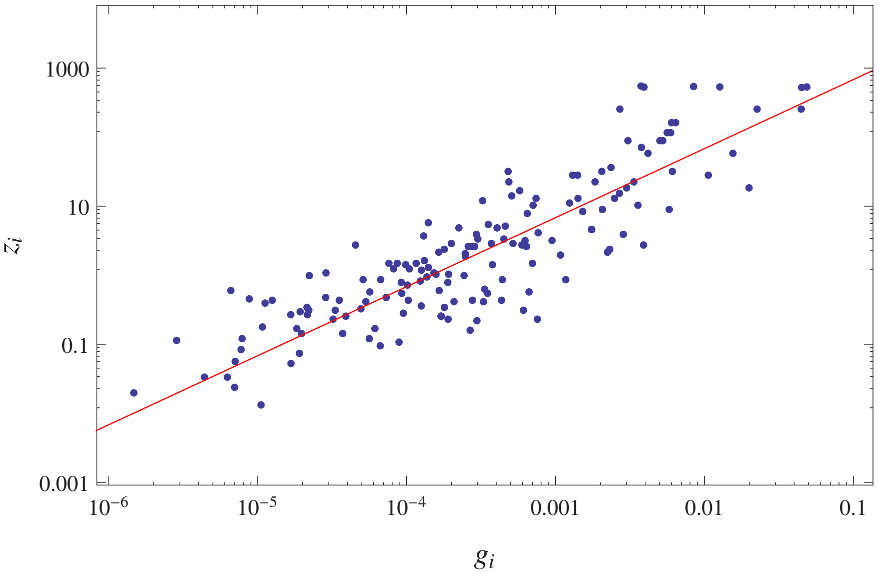

Figure 2.2: **Correlation between Lagrange multipliers and countries GDP.** The calculated z_i, compared with the g_i (re-scaled GDP) for each country for the undirected binary aggregated ITN in the 2002 snapshot, with the linear fit $z_i = \sqrt{a} \cdot g_i$ (red line).

of countries, and reveals the importance of the *GDP* to the binary structure of the ITN. Importantly, this observation was the first empirical evidence in favour of the fitness model as a powerful network model [7]. Likewise, other studies have shown that the observed topological properties turn out to be important in explaining macroeconomics dynamics [1, 2].

2.3.2 Weighted structure

Despite the importance of the topology, the latter is only the backbone over which goods are traded, and the knowledge of the volume of such trade is critical. To be able to give predictions about the weight of connections, one needs to switch from an ensemble of binary graphs to one of weighted graphs.

The simplest weighted counterpart of the BCM is the WCM, which is a maximum-entropy ensemble of weighted networks where the constraint is the strength sequence, i.e. the total trade of each country in the case of the ITN. In the canonical formalism [18], the latter can be constrained by writing the following Hamiltonian:

$$H(\mathbf{W}) = \sum_{i=1}^{N} \theta_i s_i(\mathbf{W}) \tag{2.6}$$

where the strength sequence is defined as $s_i(\mathbf{W}) = \sum_{j \neq i}^{N} w_{ij}$, $\forall\ i$, and θ_i are the free parameters (Lagrange multipliers) [18]. As a result of the constrained maximization of the entropy [18], the probability of a given configuration $\mathbf{W}$ can be written as

$$P(\mathbf{W}) = \frac{e^{-H(\mathbf{W})}}{\sum_{\mathbf{W}'} e^{-H(\mathbf{W}')}} = \prod_{i<j} (y_i y_j)^{w_{ij}} (1 - y_i y_j) \tag{2.7}$$

where $y_i \equiv e^{-\theta_i}$.

Recent studies have shown that the higher-order binary quantities predicted by the WCM, as well as the corresponding weighted quantities, are very different from the observed counterparts [18, 19]. More specifically, the main limitation of the model is that of predicting a mostly homogeneous and very dense (sometimes fully connected) topology. Roughly speaking, the model excessively ‘dilutes’ the total trade of each country by distributing it to almost all other countries. This failure in correctly replicating the purely topological projection of the real network is the root of the bad agreement between expected and observed higher-order properties.

Relation with the Gravity Model

Just like the BCM has been related to the Fitness Model [7], a variant of the WCM has been related to the Gravity Model [23]. The variant is actually a continuous version of the WCM, where the strength sequence is constrained, and the weights are real numbers instead of integers. When applied to the ITN, the model gives the following expectation for the weight of the links:

$$\langle w_{ij} \rangle = T \cdot g_i g_j \qquad \forall i, j \tag{2.8}$$

where T is the total strength in the network, and g_i is the re-scaled GDP as before [23]. In essence, the above expression identifies again a relationship between the GDP and the hidden variable (analogous to the fitness in the binary case) specifying the strength of a node.

Equation (2.8) coincides with eq.(2.9) where $\beta = 1$ and $\gamma = 0$. The model, therefore, corresponds to a particularly simple version of the Gravity Model. Indeed, the model reproduces reasonably well the observed non-zero weights of the ITN [23]. However, just like the Gravity Model, the model predicts a complete graph where $a_{ij} = 1\ \ \forall i, j$, and dramatically fails in reproducing the binary architecture of the network. This effect can be easily shown by realizing that the continuous nature of edge weights, which can take non-negative real values in the model, implies that there is a zero probability of generating zero weights (i.e. missing links). We will show the prediction of this model in comparison with our results later on in the chapter (when compared to the two-step model).

2.4 Macroeconomic approaches to the international trade network

In this section, after covering the maximum-entropy approaches, we briefly review the macroeconomic approaches to the characterization of the ITN, mainly focusing on the popular Gravity model of trade. Next, we discuss the role played by the countries GDP in determining both the presence and the amount of trade exchanges between world countries. Identifying the specific relations between the GDP and network properties, will enable us later to introduce models, which converge the two approaches.

2.4.1 The gravity model of trade

Traditionally, macroeconomic models have mainly focused on the weighted representation, because economic theory perceives the latter as being *a priori* more informative than the purely binary representation. The focus is on the expected volume of trade between two countries, given certain dyadic and country-specific macroeconomic properties. Jan Tinbergen, the physics-educated[1] Dutch economist who was awarded the first Nobel memorial prize in economics introduced the so-called Gravity Model (GM) of trade [41]. The GM aims at inferring the volume of trade between any two (trading) countries from the knowledge of their Gross Domestic Product, geographic distance, and other possible dyadic quantities of macroeconomic relevance (such as common currency, trade agreements, bordering conditions, common language, etc.) [58]. In one of its simplest forms, the gravity model predicts that the expected volume of trade between countries i and j is

$$\langle w_{ij} \rangle = \alpha \; GDP_i^{\beta} \; GDP_j^{\beta} \; R_{ij}^{\gamma}, \tag{2.9}$$

where GDP_k is the Gross Domestic Product of country k, R_{ij} is the geographic distance between countries i and j, and α, β, γ are free parameters to be estimated by fitting the model to the data [30, 15, 42, 43].[2] More complicated variants of

[1] Jan Tinbergen studied physics in Leiden, where he carried out a Ph.D. under the supervision of the famous theoretical physicist Paul Ehrenfest. Tinbergen defended his thesis in 1929, and then started a career as an economist. He was awarded the first Nobel memorial prize in economics in 1969.

[2] Note that eq.(2.9), by assuming for simplicity the same exponent β for the GDPs of both i and j, predicts $\langle w_{ij} \rangle = \langle w_{ji} \rangle$ and should therefore be interpreted as a model for the undirected version of the network. In this representation, the trade from country i to country j and the trade from country j to country i are combined into a single value of bilateral trade. Given the highly symmetric structure of the ITN at the aggregate level (i.e. when all traded products are combined), this simplification

eq.(2.9) use additional explanatory factors (with associated free parameters) either favoring or resisting trade. These additional factors can be country-specific like the GDP (e.g. population) or dyadic like the geographic distances (e.g. common currency, trade agreements, etc.). In general, if we collectively denote with $\mathbf{n}_i$ the set of *node-specific* factors and with $\mathbf{D}_{ij}$ the set of *dyad-specific* factors used, eq.(2.9) can be generalized to

$$\langle w_{ij} \rangle = F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij}), \tag{2.10}$$

where, in general, the functional form of $F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij})$ need not be of the same type as in eq.(2.9), and each component of $\mathbf{n}_i$ and $\mathbf{D}_{ij}$ may have a corresponding free parameter to be fitted. In fact, despite the fact we are focusing on the gravity model applied to the international trade network as our main application, our discussion applies to many other models of (socio-economic) networks as well. For instance, the recent Radiation Model (RM) [45], which improves the predictions of the GM when applied to mobility (rather than trade) networks, is also described by eq.(2.32), where $\mathbf{n}_i$ and $\mathbf{D}_{ij}$ are certain geographical and demographical variables. Our following discussion applies to both the GM and the RM, as well as any more general model described by eq.(2.10).

Equation (2.10) refers to the expected value of w_{ij}. The full probability distribution from which this expected value is calculated depends on the particular implementation of the model. In the GM case, this distribution can be Gaussian (implying that the expected weights can be fitted to the observed ones via a simple linear regression [46, 47]), log-normal (requiring a linear regression of log-transformed weights [49]), Poisson [49], or more sophisticated [48] (see [42] for a review). The non-uniqueness of the weight distribution already highlights a fundamental arbitrariness in the model. This is only one of the limitations of the GM and similar models.

The GM can successfully reproduce the observed trade volume between trading countries. However, at least in its simplest and most widespread implementation, the model cannot generate zero volumes and therefore predicts a fully connected network. From a model fitting perspective, this means that the GM can be fitted only to the *non-zero* weights, i.e. the strictly positive volumes existing between pairs of *connected* countries. Therefore the model effectively disregards the empirical topology of the network, thus making predictions on the basis of incomplete data. Operatively, the GM can predict a realistic trade volume only *after* the presence of the trade relation itself has been established independently [42]. This problem is particularly critical, since, depending on the datasets, up to approximately half of the possible links in the real ITN are *not* realized [21, 22, 7, 19]. If the total trade is disaggregated into commodity-specific trade flows, the resulting commodity-specific networks are even sparser [50, 51]. Clearly, the same problem

retains all the basic network properties of the system [38, 21, 22, 44, 59, 60]. Throughout the chapter, we will stick to this undirected (bilateral trade) representation, but the extension to the directed case is straightforward.

holds in general for the RM and other models of the form (2.10).

While there are variants and extensions of the GM that do generate zero weights (e.g. the so-called Poisson pseudo-maximum likelihood models [49] and 'zero-inflated' gravity models [48]), these variants can mainly reproduce the empirical link density (the realized fraction of connections), but not the observed topology [42, 43]. Indeed, even in these generalized forms, the GM predicts a largely homogeneous topology, while the empirical topological backbone of the ITN is much more heterogeneous and complex [15, 42]. For instance, the distribution of node degree (number of connections of a country) is very broad, as for the distribution of node strength (total trade volume of a country) [5, 38, 11, 39, 13, 14, 15]. A small number of rich countries dominate the trade patterns and account for most of the trade. Clustering and mixing patterns exhibit the rich-club phenomenon [55, 56], where well-connected nodes also connected to each other. The higher-order correlations are disassortative in the binary representation (nodes of low degree are more likely to be connected to nodes of a higher degree than expected by chance [21]), but assortative in the weighted representation (nodes are more likely to be connected to nodes of similar strength than expected by chance[22]). These structural properties are remarkably stable over time: despite the fourfold increase in trade volume over the last 65 years, the overall topology of international trade has remained largely constant [57].

In the next section, we will move forward, by trying to detect similarities between the two approaches. We explore the various empirical relations between the GDP of a country and specific country (network) properties like degree and strength. This simple empirical analysis reveals the GDP as a "macroeconomic fitness", i.e. a powerful predictor of the number and strength of country's trade relations.

2.4.2 The GDP as macroeconomic fitness

Let us start with an empirical analysis of the GDP. We first define new rescaled quantities of the GDP: g_i and $\tilde{g}_i$

$$g_i \equiv \frac{\text{GDP}_i}{\sum_j \text{GDP}_j}, \, \forall\, i \qquad \tilde{g}_i \equiv \frac{\text{GDP}_i}{\text{GDP}_{mean}}, \, \forall\, i, \tag{2.11}$$

where $\text{GDP}_{mean} \equiv \frac{\sum_i^N \text{GDP}_i}{N}$ is the average GDP for an observed year. The two quantities adjust the values of the countries GDPs for both the size of the network and the growth, and are a connected by a simple relation $\tilde{g}_i = N \cdot g_i$. We use the two quantities of the rescaled GDP throughout our analysis, mainly using g_i for the reason that the quantity is bounded $0 \leq g_i \leq 1$ which coincides with our model.

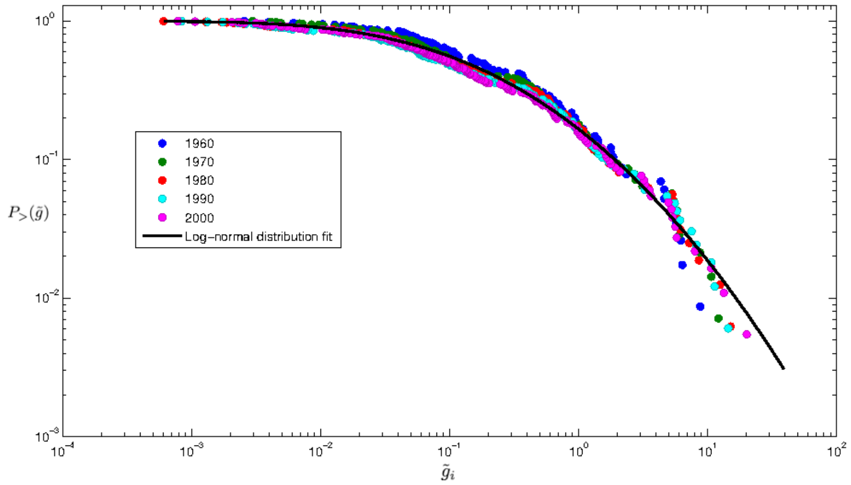

Figure 2.3: **Cumulative Distribution of countries GDP for different decades.** Empirical cumulative distributions $P_>(\tilde{g})$ of the GDP rescaled to the mean, for different years. The curve is a log-normal distribution fitted to the data.

In fig. 2.3 we plot the cumulative distribution of the rescaled GDP $\tilde{g}_i$ with i indexing the countries for the different decades collected into our data set. What emerges is that the distributions of the rescaled GDPs can be described by log-normal distribution characterized by similar values of the parameters. The log-normal curve is fitted to all the values (from the different decades). This suggests that the rescaled GDPs are quantities which do not vary much with the evolution of the system, thus potentially representing the (constant) hidden macroeconomic fitness ruling the entire evolution of the system itself. This, in turn, implies understanding the functional dependence of the key topological quantities on the countries rescaled GDP.

As already pointed out by a number of results [19], the topological quantities which play a major role in determining the ITN structure are the countries degrees (i.e. the number of their trading partners) and the countries strengths (i.e. the total volume of their trading activity). Thus, the first step to understanding the role of the rescaled GDP in shaping the ITN structure is quantifying the dependence of degrees and strengths on it. Since we want to analyse each snapshot at a time (correction for size is not needed), here we will use the bounded rescaled GDP g_i. Moreover, this form of the rescaled GDP coincides with a bounded macroeconomic fitness value, which is consistent with the models presented in the next sections.

To this aim, let us explicitly plot k_i versus g_i and s_i versus g_i for a particular decade, as shown in fig. 2.4. The red points represent the relations between the two pairs of observed quantities for the 2000 snapshot. Interestingly, the rescaled GDP is directly proportional to the strength (on a log-log scale), thus indicating that *the wealth of countries is strongly correlated to the total volume of trade they participate in*. Such evidence provides the empirical basis for the definition of the gravity model, stating that *the trade between any two countries is directly proportional to the (product of the) countries GDP*.

On the other hand, the functional dependence of the degrees on the g_i values is less simple to decipher. Generally speaking, the relation is monotonically increasing, and this means that countries with high GDP also have a high degree, i.e. are strongly connected with the others; coherently, countries characterized by a low value of the GDP have also a low degree, i.e. are less connected to the rest of the world. Moreover, while for low values of the GDP there seems to exist a linear relation (on a log-log scale) between k_i and g_i, as the latter rises a saturation effect is observed (in correspondence of the value $k_{max} = N - 1$), due to the finite size of the network under analysis. Roughly speaking, richest countries lie on the vertical trait of the plot, while poorest countries lie on the linear trait of the same plot: in other words, *the degree of countries represents a purely topological indicator of the countries wealth.*

To sum up, fig. 2.4 shows that countries GDP plays a double role in shaping the ITN structure: first, it controls for the number of trading channels each country establishes; second, it controls for the volume of trade each country participates in, via the established connections. The blue points in fig. 2.4, instead, represent the relation between $\langle k_i \rangle$ versus g_i and $\langle s_i \rangle$ versus g_i, where the quantities in brackets are the predicted values for degrees and strengths generated by our model, which we will discuss later.

The obvious question that arises from these findings is can we extend the result shown in section 2.3 to create a GDP-driven model for both the binary and the weighted representation of the ITN. In the next section, we tackle exactly this problem using a recent maximum-entropy model [33] which is able to reproduce both representations.

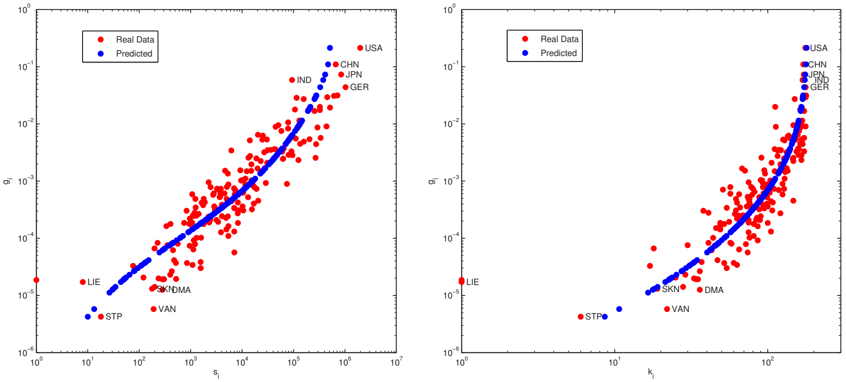

Figure 2.4: **The relation of the GDP with local network properties.** Comparison between observed (red points) and expected (blue points) degrees and strengths for the aggregated ITN in the 2000 snapshot. Right panel: degree k_i versus normalized GDP g_i and expected degree $\langle k_i \rangle$ versus normalized GDP g_i. Left panel: strength s_i versus normalized GDP g_i and expected strength $\langle s_i \rangle$ versus normalized GDP g_i.

2.5 A GDP-driven model of the ITN

Motivated by the challenge to satisfactorily model both the topology and the weights of the ITN, the ECM has been recently proposed as an improved model of this network [33]. The ECM focuses on weighted networks, and can enforce the degree and strength sequence simultaneously [32]. It builds on the so-called generalized Bose-Fermi distribution that was first introduced as a null model of networks with coupled binary and weighted constraints [34].

In the ECM, the degree and strength sequence can be constrained by writing the following Hamiltonian:

$$H(\mathbf{W}) = \sum_{i=1}^{N} [\alpha_i k_i(\mathbf{W}) + \beta_i s_i(\mathbf{W})] \tag{2.12}$$

where the strength sequence is defined as $s_i(\mathbf{W}) = \sum_{j \neq i}^{N} w_{ij}$, $\forall\, i$ and the degree sequence as $k_i(\mathbf{W}) = \sum_{j \neq i}^{N} a_{ij} = \sum_{j \neq i}^{N} \Theta[w_{ij}]$, $\forall\, i$. As a result, the probability of a given configuration $\mathbf{W}$ can be written as

$$P(\mathbf{W}) = \frac{e^{-H(\mathbf{W})}}{\sum_{\mathbf{W}'} e^{-H(\mathbf{W}')}} = \prod_{i<j} \frac{(x_i x_j)^{a_{ij}} (y_i y_j)^{w_{ij}} (1 - y_i y_j)}{1 - y_i y_j + x_i x_j y_i y_j} \tag{2.13}$$

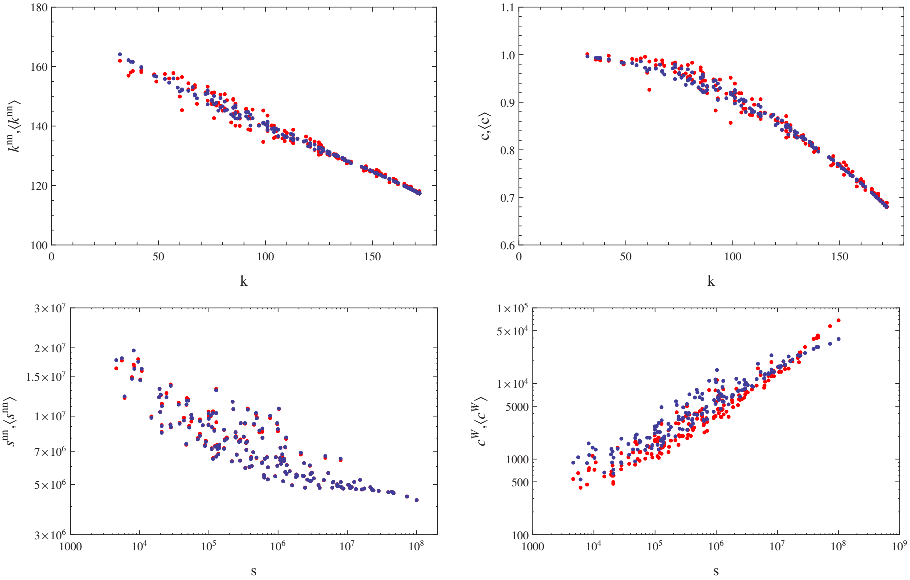

Figure 2.5: **Enhanced Configuration Model, reconstruction of higher-order properties.** Comparison between the observed undirected binary and weighted properties (red points) and the corresponding ensemble averages of the the ECM (blue points) for the aggregated ITN in the 2002 snapshot. Top left panel: Average Nearest Neighbour Degree k^{nn} versus degree k_i; Top right panel: Binary Clustering Coefficient C_i versus degree k_i ; Bottom left panel: Average Nearest Neighbour Strength s^{nn} versus strength s_i ; Bottom right panel: Weighted Clustering Coefficient C^W versus strength s_i .

with $x_i \equiv e^{-\alpha_i}$ and $y_i \equiv e^{-\beta_i}$. The ECM gives the following predictions about the probability of a link ($\langle a_{ij} \rangle$) and the expected weight of the link ($\langle w_{ij} \rangle$):

$$\langle a_{ij} \rangle = \frac{x_i x_j y_i y_j}{1 - y_i y_j + x_i x_j y_i y_j} = p_{ij} \tag{2.14}$$

$$\langle w_{ij} \rangle = \frac{x_i x_j y_i y_j}{(1 - y_i y_j + x_i x_j y_i y_j)(1 - y_i y_j)} = \frac{p_{ij}}{1 - y_i y_j}. \tag{2.15}$$

According to the maximum-likelihood method proposed in [32], the vectors of unknowns $\vec{x}$ and $\vec{y}$ can be numerically found by solving the system of $2N$ coupled equations

$$\langle k_i(\mathbf{W})\rangle = \sum_{j\neq i}^{N} p_{ij} = k_i(\mathbf{W}^*) \qquad \forall\, i \tag{2.16}$$

$$\langle s_i(\mathbf{W})\rangle = \sum_{j\neq i}^{N} \langle w_{ij}\rangle = s_i(\mathbf{W}^*) \qquad \forall\, i \tag{2.17}$$

and will be indicated as $\vec{x}^*$ and $\vec{y}^*$. These unknown parameters can be treated as fitness parameters which control the probability of forming a link and the expected weight of that link simultaneously.

The application of the ECM to various real-world networks shows that the model can accurately reproduce the higher-order empirical properties of these networks [32]. When applied to the ITN in particular, the ECM replicates both binary and weighted empirical properties, for different levels of disaggregation, and for several years (temporal snapshots) [33]. Indeed, in fig. 2.5 we show the higher-order binary quantities (average nearest neighbour degree and clustering coefficient) as well as their weighted ones (average nearest neighbour strength and weighted clustering coefficient) for the 2002 snapshot of the ITN. We compare the observed values (red points) and the corresponding quantities predicted by the ECM (blue points). The mathematical expressions for all these quantities are provided in the Appendix. We find a very good agreement between data and model, confirming the recent results in [33] for the data set we are using here. We also confirmed that these results are robust for several temporal snapshots [33].

2.5.1 From Lagrange multipliers to macroeconomic properties

Considering the promising results of the ECM and the results from section 2.4.2, we now make a step forward and check whether the hidden variables x_i and y_i, which effectively reproduce the observed ITN, can be thought of as 'fitness' parameters having a clear economic interpretation. This amounts to checking whether the relation shown previously in fig. 2.4 for the purely binary case can be generalized in order to find a macroeconomic interpretation to the abstract fitness parameters in the general weighted case as well.

In fig. 2.6 we show the relationship between the two parameters x_i and y_i and the rescaled GDP (g_i) for each country of the ITN in the 2002 snapshot. We find strong correlations between these quantities. The fitness parameter x_i turns out to be in a roughly linear relation with the rescaled GDP g_i, fitted by the curve

$$x_i = \sqrt{a}\cdot g_i \tag{2.18}$$

where $\sqrt{a}$ is the fitted constant, and $g_i = \frac{GDP_i}{\sum_i GDP_i}$ (all the GDPs are relative to that specific year). It should be noted that this relation is similar to that

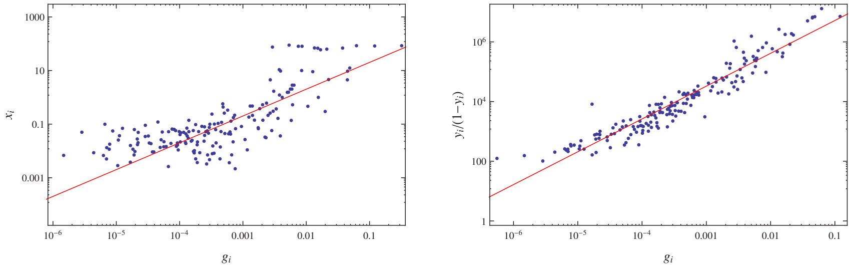

Figure 2.6: **Relation between Lagrange multipliers of the ECM and countries GDP.** calculated x_i (left panel) and $\frac{y_i}{1-y_i}$ (right panel) compared with the g_i (rescaled GDP) for each country for the undirected binary aggregated ITN in the 2002 snapshot, with the linear fits (in log-log scale) $x_i = \sqrt{a} \cdot g_i$, and $\frac{y_i}{1-y_i} = b \cdot g_i^c$ (red lines), where a, b, and c are the fitted constant parameters per year.

found between z_i and g_i in the BCM and shown previously in fig.2.4, but less accurate. This observation will be useful later. By contrast, since the *GDP* is an unbounded quantity, while the fitness parameter y_i is bounded between 0 and 1 (this is a mathematical property of the model [34, 32]), the relation between y_i and g_i is necessarily highly nonlinear. A simple functional form for such a relationship is given by

$$y_i = \frac{b \cdot g_i^c}{1 + b \cdot g_i^c}. \tag{2.19}$$

Indeed, fig. 2.6 confirms that the above expression provides a very good fit to the data.

We checked that the above results hold systematically over time, for each snapshot of the ITN in our data set. This implies that, in a given year, we can insert eqs.(2.18) and (2.19) into eqs.(2.38) and (2.37) to obtain a GDP-driven model of the ITN structure for that year. Such a model highlights that the *GDP* has a crucial role in shaping both the binary and the weighted properties of the ITN. While this was already expected on the basis of the aforementioned results obtained using the BCM and the WCM (or the corresponding simplified gravity model) separately, finding the appropriate way to explicitly combine these results into a unified description of the ITN has remained impossible so far. Rather than exploring in more detail the predictions of the GDP-driven model in the form described above, we first make some considerations leading to a simplification of the model itself.

2.5.2 Reduced two-step model

At this point, it should be noted that we arrived at two seemingly conflicting results. We showed that both the BCM and ECM give a very good prediction for the binary topology of the network. However, eqs.(2.3) and (2.38), which specify the connection probability p_{ij} in the two models, are significantly different. The comparable performance of the BCM and the ECM at a binary level (see figs.2.3 and 2.5) makes us expect that, when the specific values $\vec{z}^*$ and $\vec{x}^*$ are inserted into eqs.(2.3) and (2.38) respectively, the values of the connection probability become comparable in the two models, despite the different mathematical expressions.

In fig. 2.7 we compare the the two probabilities for the ITN in the 2002 snapshot. Note that each point refers to the probability of creating a link between a pair of countries, which results in $\frac{N(N-1)}{2}$ points. Indeed, we can see that the values are scattered along the identity line, confirming the expectation that the connection probability has similar value in the two models.

The above result allows us to make a remarkable simplification. In eqs.(2.38) and (2.37), we can replace the expression for p_{ij} provided by the ECM with that provided by the BCM in eq.(2.3). To avoid confusion, we denote the new probability with p_{ij}^{ts}, where ts stands for 'two-step', for a reason that will be clear immediately. This results in the following equations for the expected network properties:

$$\langle a_{ij}\rangle^{ts} = \frac{z_i z_j}{1+z_i z_j} \equiv p_{ij}^{ts}, \tag{2.20}$$

$$\langle w_{ij}\rangle^{ts} = \frac{p_{ij}^{ts}}{1-y_i y_j}. \tag{2.21}$$

where the z_i's, and therefore the p_{ij}^{ts}'s, depend only on the degrees through eq.(2.4), while the y_i's and the $\langle w_{ij}\rangle^{ts}$'s depend on both strengths and degrees through eqs.(2.16) and (2.17).

In this simplified model the connection probability, which fully specifies the topology of the ensemble of networks, no longer depends on the strengths as in the ECM, while the weights still do. This implies that we can specify the model via a two-step procedure where we first solve the N equations determining p_{ij}^{ts} via the degrees, and then find the remaining variables determining $\langle w_{ij}\rangle^{ts}$ through the ECM. For this reason, we denote the model as the Two-Step (TS) model.

The probability of a configuration $\mathbf{W}$ reads

$$P^{ts}(\mathbf{W}) = \prod_{i<j} q_{ij}^{ts}(w_{ij}) \tag{2.22}$$

where

$$q_{ij}^{ts}(w_{ij}) = \frac{(z_i z_j)^{a_{ij}}(y_i y_j)^{w_{ij}-a_{ij}}(1-y_i y_j)^{a_{ij}}}{1+z_i z_j} \tag{2.23}$$

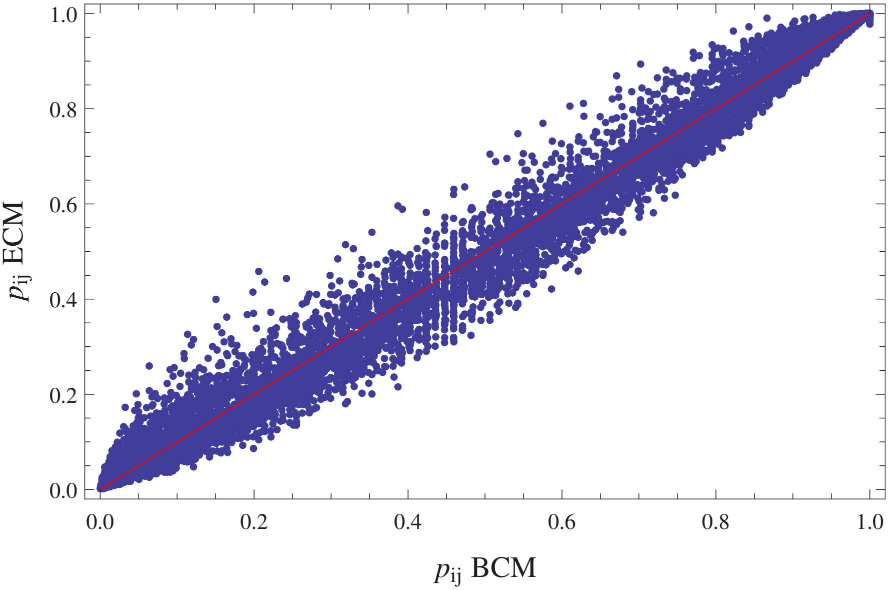

Figure 2.7: **Link formation probability for the different models.** The probability of forming a link in the ECM p_{ij} *ECM* compared to the probability in the BCM p_{ij} *BCM* for the undirected binary aggregated ITN in the 2002 snapshot. The red line describes the identity line.

is the probability that a link of weight w_{ij} connects the nodes (countries) i and j. The above probability has the same general expression as in the original ECM [32], but here z_i comes from the estimation of the simpler BCM. It is instructive to rewrite (2.23) as

$$q_{ij}^{ts}(0) = \frac{1}{1+z_i z_j} = (1-p_{ij}^{ts}); \qquad (2.24)$$

$$q_{ij}^{ts}(w) = p_{ij}^{ts}(y_i y_j)^{w-1}(1-y_i y_j), \; \forall\, w>0 \qquad (2.25)$$

to highlight the random processes creating each link. As a first step, one determines whether a link is created or not with a probability p_{ij}^{ts}. If a link (of unit weight) is indeed established, a second attempt determines whether the weight of the same link is increased by another unit (with probability $y_i y_j$) or whether the process stops (with probability $1-y_i y_j$). Iterating this procedure, the probability that an edge with weight w is established between nodes i and j is given precisely by $q_{ij}^{ts}(w)$ in eq.(2.25). The expected weight $\langle w_{ij}\rangle^{ts}$ is then correctly retrieved as $\sum_{w=0}^{+\infty} w \cdot q_{ij}^{ts}(w)$.

Using the relations found in eqs.(2.5) and (2.19), we can input the g_i as the fitness parameters into eqs.(2.20) and (2.21) to get the following expressions that mathematically characterize our GDP-driven specification of the TS model:

$$\langle a_{ij}\rangle^{ts} = \frac{ag_i g_j}{1+ag_i g_j} \equiv p_{ij}^{ts} \tag{2.26}$$

$$\langle w_{ij}\rangle^{ts} = p_{ij}^{ts}\frac{(1+bg_i^c)(1+bg_j^c)}{(1+bg_i^c+bg_j^c)}. \tag{2.27}$$

The above equations can be used to reverse the approach used so far: rather than using the $2N$ free parameters of the ECM ($\vec{x}$ and $\vec{y}$) or of the TS model ($\vec{z}$ and $\vec{y}$) to fit the models on the observed values of the degrees and strengths, we can now use the knowledge of the GDP of all countries to obtain a model that only depends on the three parameters a, b, c. Assigning values to these parameters can be done using two techniques: maximization of the likelihood function and non-linear curve fitting. Since the model is a two-step one, we can first assign a value to the parameter a, and only in the second step (once a is set) we fit the parameters b and c.

We chose to fix a by maximizing the likelihood function [36], which results in constraining the expected number of links to the observed number ($\langle L\rangle = L$), as in [7]. Fixing the values of b and c is slightly more complicated. Since the model uses the approximated expressions of the TS model, rather than those of the original ECM model, maximizing the likelihood function in the second step no longer yields the desired condition $\langle T\rangle = T$, where T is the total strength in the network. Similarly, extracting the parameters from the fit as shown in fig. 2.6 does not maintain the total strength in the network. In absence of any a-priori preference, we chose the latter procedure, due to its relative numerical simplicity with respect to the former one.

In fig. 2.8 we show a comparison between the higher-order observed properties of the ITN in 2002 and their expected counterparts predicted by the GDP-driven TS model. Again, the mathematical expressions of these properties are provided in the Appendix. As a baseline comparison, we also show the predictions of the *GDP*-driven WCM model with continuous weights described by eq.(2.8) [23], which coincides with a simplified version of the gravity model as we mentioned.

We see that the GDP-driven TS model reproduces the observed trends very well. Of course, as expected, the predictions in fig.2.8 (which use only three free parameters) are noisier than those in fig.2.5 (which use $2N$ free parameters). This is due to the fact that eqs.(2.5) and (2.19) describe fitting curves rather than exact relationships. Importantly, our model performs significantly better than the WCM/gravity model in replicating both binary and weighted properties. Again, the drawback of these models lies in the fact that they predict a fully connected topology and a relatively homogeneous network.

It should also be noted that the plot of average nearest neighbour strength (s^{nn}) predicted by our model is slightly shifted with respect to the observed points. This effect is due to the fact that, as we mentioned, the total strength T (hence the average trend of the s^{nn}) is only approximately reproduced by our model, as a result of the simplification from the ECM to the TS model.

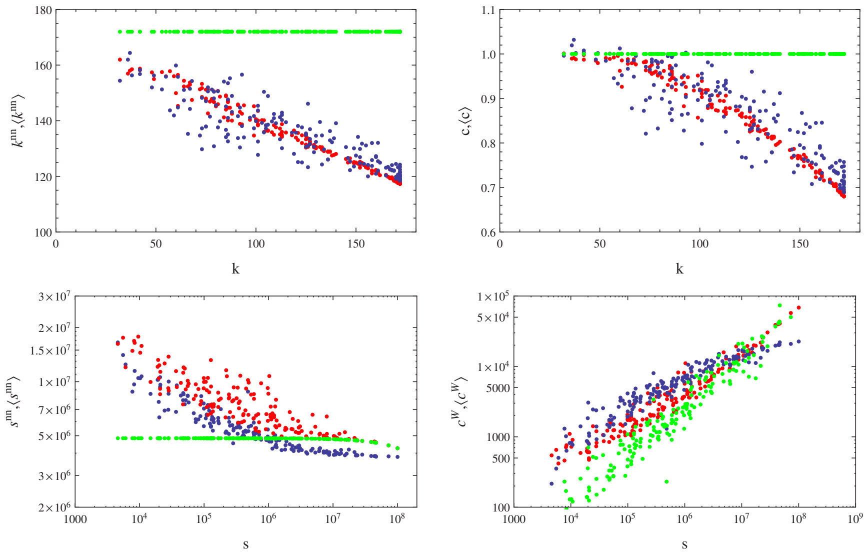

Figure 2.8: **Two-Step Model, reconstruction of higher-order properties.** Comparison between the observed properties (red points), the corresponding ensemble averages of the *GDP*-driven two-steps model (blue points) and the *GDP*-driven WCM model (green points), of the aggregated ITN in 2002. Top left: Average Nearest Neighbour Degree k_i^{nn} versus degree k_i. Top right: Binary Clustering Coefficient c_i versus degree k_i. Bottom left: Average Nearest Neighbour Strength s_i^{nn} versus strength s_i. Bottom right: Weighted Clustering Coefficient c_i^w versus strength s_i. The GDP-driven TS model reproduces the empirical trends very well with respect to the WCM.

As for all the other results in this chapter, we checked that our findings are robust over the entire time span of our data set. We, therefore, conclude that the ECM model, as well as its simplified TS variant, can be successfully turned into a fully GDP-driven model that simultaneously reproduces both the topology and the weights of the ITN.

The success of the TS model has a meaningful interpretation. Looking back at eqs.(2.20) and (2.26), we recall that the effect of the TS approximation is the fact that the connection probability p_{ij}^{ts} can be estimated separately from the weights $\langle w_{ij} \rangle^{ts}$, using only the knowledge of the degree sequence if eq.(2.20) is used, or the GDP and total number of links if eq.(2.26) is used, while discarding that of the strengths. By contrast, the estimation of the expected weights $\langle w_{ij} \rangle^{ts}$ cannot be carried out separately, as it requires that the connection probability p_{ij}^{ts} appearing in eqs.(2.21) and (2.27) is estimated first. This asymmetry of the model means that the topology of the ITN can be successfully inferred without any information about the weighted properties, while the weighted structure cannot be inferred without topological information. The expressions defining the TS model provide a mathematical explanation for this otherwise puzzling effect that has already been documented in previous analyses of the ITN [19, 33].

2.6 The enhanced gravity model

In the previous section, we have shown that the two-step model can reproduce the higher-order properties of both representations of the ITN. However, despite the vast improvements the model represent, it still has some definite limitations. Firstly, the model does not allow to introduce additional macroeconomic parameters like geographic distance or other pair-countries information like trade relations or common borders. This type of information is frequently being used in macroeconomic models. Furthermore, the model cannot reproduce the specific weights w_{ij} of the network. In this section, using a different methodology than before, we start with the gravity model and reformulate it according to maximum-entropy principles. This maximum-entropy generalization is aimed at creating a model that combines the advantages of the gravity model (accurate link expectation) with the benefits of the maximum-entropy models (topology and higher-order reconstruction).

As we discussed before, the gravity model in eq.(2.10) (which includes eq.(2.9) as a particular case) is successful in reproducing link weights only *after* the existence of the links themselves has been preliminarily established. This means that eq.(2.10) in actually incorrect and should by rather reformulated as a model for the *conditional expectation* of the weight w_{ij}, given that $w_{ij} > 0$.

To do so, we need to introduce $q_{ij}(w)$ as the probability that the volume of trade between countries i and j takes a value w, with w being, without loss of generality, a non-negative integer number (the event $w = 0$ indicates the absence of a trade link). The probability $q_{ij}(w)$ is the fundamental quantity that fully specifies the model. In particular, it controls both the topology and the link weights of the network. Our aim is to find the form of $q_{ij}(w)$ that produces the desired gravity-like conditional expectation for link weights, as well as a realistically expected topology. The search for the form of $q_{ij}(w)$ will be guided by the important requirement that *the expected topology should not depend on the (arbitrary) units of measure chosen to measure the link weights.* The latter requirement will be referred to as *topological invariance.*

Our first requirement is that $q_{ij}(w)$ produces eq.(2.10), once the latter has been rewritten as an expression for the conditional weights. We perform this rewriting first. Note that the probability p_{ij} that countries i and j are connected (irrespective of the volume of trade) is given by

$$p_{ij} = 1 - q_{ij}(0) = \langle a_{ij} \rangle, \tag{2.28}$$

where $a_{ij} = \Theta(w_{ij})$ is the entry of the binary adjacency matrix of the network (equal to 1 if a link between countries i and j exists and 0 otherwise). Note that p_{ij} does not depend on $q_{ij}(w)$ for $w > 0$. By contrast, the expected trade volume (irrespective of whether a connection is established) is given by

$$\langle w_{ij} \rangle \equiv \sum_{w>0} w \ q_{ij}(w), \tag{2.29}$$

which does not depend on $q_{ij}(0)$. Apparently, the fact that $\langle a_{ij} \rangle$ and $\langle w_{ij} \rangle$ depend on different quantities implies that they can be defined separately, thus allowing one to reproduce the topology and the weights simultaneously. However, we will show that this is not the case: to enforce topological invariance, an explicit dependence between $\langle a_{ij} \rangle$ and $\langle w_{ij} \rangle$ should be introduced.

To rewrite eq.(2.10) as a conditional expectation, we define the *conditional expected weight* of the link between nodes i and j as

$$\langle w_{ij} | a_{ij} = 1 \rangle \equiv \sum_{w>0} w \ q_{ij}(w | a_{ij} = 1) = \frac{\langle w_{ij} \rangle}{p_{ij}} \tag{2.30}$$

where

$$q_{ij}(w | a_{ij} = 1) = \frac{q_{ij}(w)}{\sum_{u>0} q_{ij}(u)} = \frac{q_{ij}(w)}{p_{ij}} \tag{2.31}$$

is the (correctly renormalized) conditional probability that w_{ij} equals w given that $a_{ij} = 1$ (i.e., given that the link is realized). We can now replace Eq. (2.10) with the intended expression

$$\langle w_{ij} | a_{ij} = 1 \rangle = F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij}). \tag{2.32}$$

Our next requirement is that $q_{ij}(w)$ enforces topological invariance. To ensure that this is done without making *ad hoc* assumptions and using only empirical information, we are going to formulate the problem within a maximum-entropy framework. In doing so, we will generalize previous maximum-entropy formulations of the GM by making them manifestly topologically invariant. For clarity, in the next section we first briefly review these previous formulations, before providing our extension.

2.6.1 Maximum-entropy reformulation of the gravity model

Per se, the standard GM is not a micro-founded model. However, various micro-founded models admit a gravity-like relation as their equilibrium outcome [64, 65, 66, 67]. Notably, the GM can be obtained from the maximum-entropy principle [68], and this result has been reformulated recently in the context of maximum-entropy ensembles of networks [23]. The maximum-entropy framework is in some sense the most general (i.e. requiring the minimal set of assumptions) context

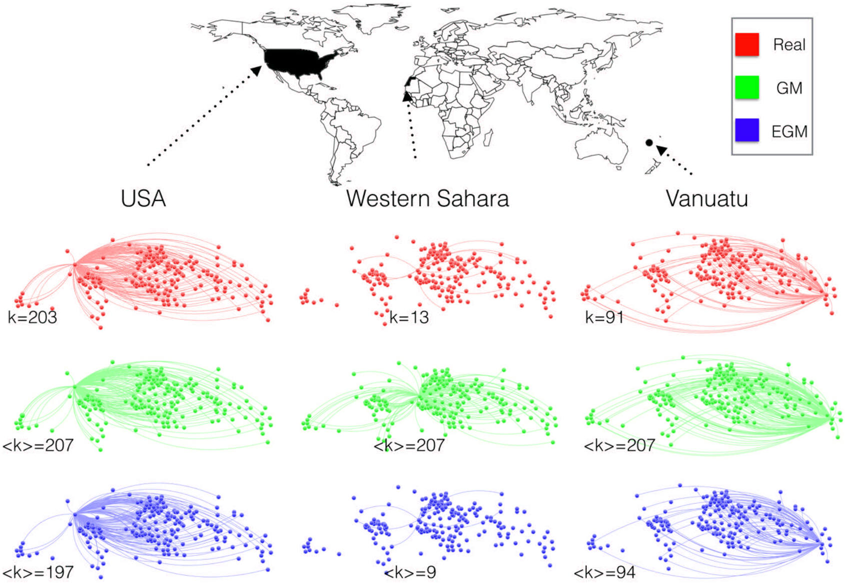

Figure 2.9: **The EGM enhanced reconstruction of the network topology.** The heterogeneous binary topology of the observed ITN (red points) and the expected values from the EGM (blue points). Left: the ANND as a function of the observed degree k. Right: the clustering coefficient C as a function of the observed degree k.

wherein the GM emerges naturally as an outcome. It also leads to the least biased predictions, as it makes no other hypothesis than consistence with a certain aggregation of the data. We first shortly review this approach, then slightly modify it in a form that assumes discrete rather than continuous trade volumes, and finally, generalize it to a novel model that fixes the main issue of the GM.

The maximum-entropy approach starts by considering the space of all networks with N nodes, where in our case N is the number of countries.

Generalizing the results in [18, 22], we preliminarily require that all the empirical edge weights $\{w_{ij}\}$ are reproduced on average by a maximum-entropy model. This leads to the graph probability $P(\mathbf{W})$ that maximizes Shannon's entropy $S \equiv -\sum_{\mathbf{W}} P(\mathbf{W}) \ln P(\mathbf{W})$ (where the sum runs over all possible weighted networks with the same number of nodes as the real network).

If the weight of a link between two nodes is denoted (in some units) as the entry w_{ij} of a non-negative integer matrix $\mathbf{W}$, then the corresponding entry in the (purely binary) adjacency matrix $\mathbf{A}$ of the graph is

$$a_{ij} \equiv \mathbf{\Theta}(w_{ij}), \tag{2.33}$$

where $\mathbf{\Theta}$ is a Heaviside step function.

The result is $P(\mathbf{W}) = \prod_{i<j} y_{ij}^{w_{ij}}/Z$, where y_{ij} is now a 'dyadic' (as opposed to node-specific) hidden variable and where Z is the normalizing constant, or partition function [18]. This model leads to expected weights of the form $\langle w_{ij}\rangle = \frac{y_{ij}}{1-y_{ij}}$ and $\langle w_{ij}|a_{ij} = 1\rangle = \frac{1}{1-y_{ij}}$. Compared with previous maximum-entropy approaches to the ITN [18, 22, 19, 33], this model has the big advantage that y_{ij} can be chosen such that $\langle w_{ij}|a_{ij} = 1\rangle = F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij})$ where F is any function of dyadic ($\mathbf{D}_{ij}$) and node-specific ($\mathbf{n}_i$) variables, like in eq.(2.32). Note that $F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij})$ has the same units of measure as w_{ij}. At the same time, the model fixes the entire probability $q_{ij}(w)$ that nodes i and j are connected by a link of weight w, which here has the geometric [18, 22] form $q_{ij}(w) = y_{ij}^{w}(1 - y_{ij})$, thus removing the aforementioned undesired arbitrariness of the weight distribution in the GM. Therefore this model can be regarded as the 'canonical' specification of the GM within a maximum-entropy framework. Note in particular that fixing $\langle w_{ij}|a_{ij} = 1\rangle = F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij})$ automatically fixes all the other moments of the geometric weight distribution, eliminating the extra parameters introduced by the addition of additive or multiplicative noise to eq. (2.9).

Enforcing topological invariance means that, if we express the trade volumes $\{w_{ij}\}$ in terms of millions of dollars rather than dollars, we want to obtain the same expectations for the topology $\{a_{ij}\}$, even if the expected volumes $\{\langle w_{ij}\rangle\}$ should instead scale accordingly. Unfortunately, in the above model the connection probability $p_{ij} \equiv \langle a_{ij}\rangle = y_{ij} = \langle w_{ij}\rangle/(1 + \langle w_{ij}\rangle)$ is not invariant under a change of units of measure for w_{ij}. While $\langle w_{ij}\rangle$ scales with w_{ij} as desired, $\langle a_{ij}\rangle$ should not scale with $\langle w_{ij}\rangle$, because the observed a_{ij} is independent of the scale of w_{ij}: we do not want the expected number of 'zeroes' to be affected by the arbitrary units of measure of the non-zero weights. Incidentally, we note that choosing a realistically small unit (e.g. one dollar like in many trade databases) implies that w_{ij} (and therefore $\langle w_{ij}\rangle$) is a large number, which implies $p_{ij} \to 1$: as the unit becomes smaller, this model predicts a denser network, asymptotically complete like in the traditional GM. The dependence of the link density on arbitrary units of weight, as well as its asymptotic saturation to a unit value, is a general explanation for the failure of an entire class of weighted network models (like the GM itself) that do not target purely topological properties in their construction. Indeed, we argue that a simple and general theoretical reason for the empirical failure of previous models is their lack of topological invariance.

2.6.2 The complete model

Since the maximum-entropy framework requires minimal assumptions (as it does not postulate mechanistic or behavioral rules), it represents the most general and transparent setting to reformulate the GM in such a way that topological invariance is enforced as an additional ingredient, while keeping the other more

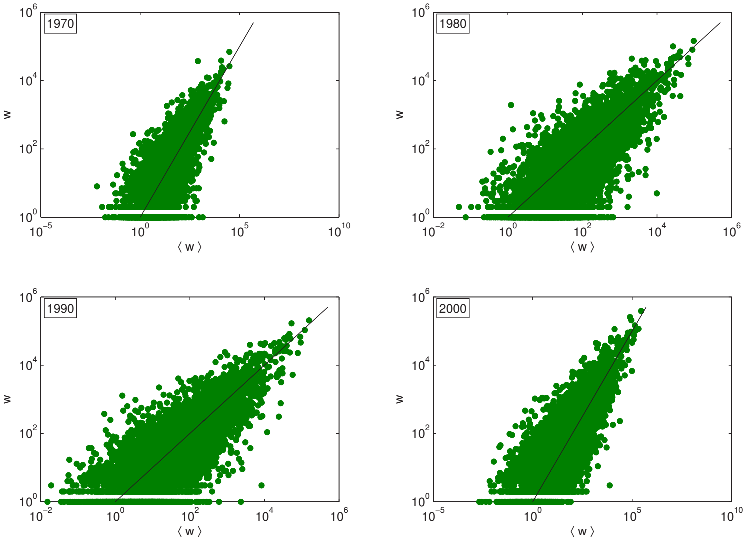

Figure 2.10: **The Gravity model, trade flow expectation.** Comparison between the observed weights of the ITN on the y-axis and the GM expected weights (green points) on the x-axis. The black line is the identity line.

traditional specifications unchanged.

We generalize a recent model [32], based on an analytical maximum-likelihood estimation method [18], that has been recently proposed in order to construct advanced maximum-entropy ensembles of weighted networks that significantly improve the fit to real data. The approach is particularly successful when applied to the ITN [33]. In its standard formulation, the model enforces the degree and strength sequence simultaneously [32]. It builds on the so-called generalized Bose-Fermi distribution that was first introduced as a null model of networks with coupled binary and weighted constraints [34].

Here, we reformulate the model more generally as a model that can flexibly reproduce the edge weights of a network using both dyadic and node-specific factors, while at the same time enforcing *topological invariance*, i.e. the desired invariance of the (expected) binary structure under a change in the (arbitrary) units of weight.

We therefore introduce a model that, besides the empirical weights $\{w_{ij}\}$, enforces the empirical topology $\{a_{ij}\}$, thus manifesting topological invariance from the very beginning. As shown in sec. Materials and Methods, this model yields a

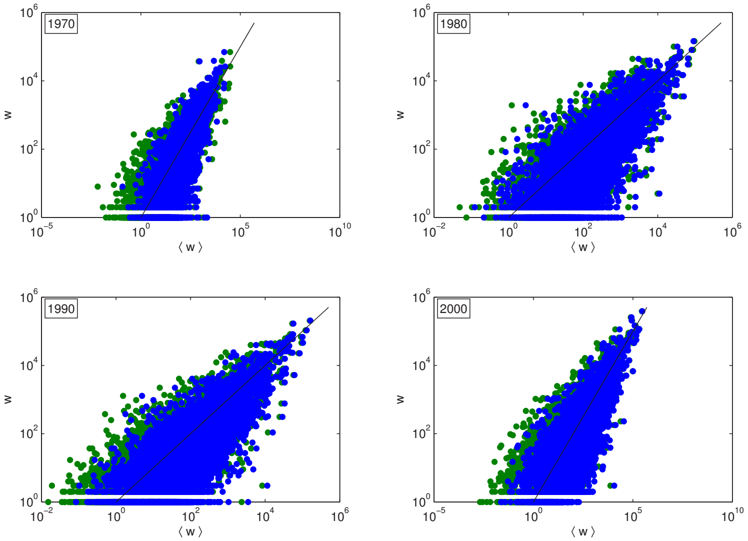

Figure 2.11: **The Enhanced Gravity model, trade flow expectation.** Comparison between the observed weights of the ITN on the y-axis and the EGM expected weights (blue points) on the x-axis, in green are the expectations of the original GM. The black line is the identity line.

weight probability

$$q_{ij}(w) \equiv \frac{(x_{ij})^{a_{ij}}\,(y_{ij})^{w_{ij}}\,(1-y_{ij})}{1-y_{ij}+x_{ij}y_{ij}}, \tag{2.34}$$

where $\{x_{ij}\}$ and $\{y_{ij}\}$ are two arrays of dyadic hidden variables. The conditional expected weight is

$$\langle w_{ij}|a_{ij}=1\rangle = \frac{1}{1-y_{ij}} \tag{2.35}$$

and the connection probability is

$$p_{ij} = 1 - q_{ij}(0) = \frac{x_{ij}y_{ij}}{1-y_{ij}+x_{ij}y_{ij}}. \tag{2.36}$$

As we illustrate below, now the presence of the extra variable x_{ij} allows to keep p_{ij} fixed (thus enforcing topological invariance) while varying y_{ij} as a result of a possible change of scale in the original weights $\{w_{ij}\}$.

The dyadic nature of the model allows us to combine the successful ingredients of the traditional GM, which satisfactorily reproduces the conditional weights of the ITN, with those of more recent network models, which accurately reproduce the topology. One one hand, we require that (2.35) has the generic structure of the GM as in eq.(2.32):

$$\langle w_{ij} | a_{ij} = 1 \rangle = \frac{1}{1 - y_{ij}} \equiv F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij}). \tag{2.37}$$

On the other hand, we require that eq.(2.36) has the structure of a binary maximum-entropy model reproducing the topology of the ITN [7, 36, 21, 31, 20]:

$$p_{ij} = \frac{x_{ij} y_{ij}}{1 - y_{ij} + x_{ij} y_{ij}} \equiv \frac{z_{ij}}{1 + z_{ij}}. \tag{2.38}$$

The above two expressions define our topology-enhanced model in general form. Importantly, the conditional weights are independent of the topology, while the opposite is not true. Also note that $F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij})$ depends on the chosen currency or money unit, while we require z_{ij} to be invariant (and dependent uniquely on the binary structure of the network). This implies that the general solution of the model is

$$y_{ij} = 1 - \frac{1}{F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij})} \tag{2.39}$$

$$x_{ij} = z_{ij} \frac{1 - y_{ij}}{y_{ij}} = \frac{z_{ij}}{F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij}) - 1} \tag{2.40}$$

In general, both $F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij})$ and z_{ij} allow for any combination of dyadic and country-specific properties. In what follows, we make (the simplest) particular choices for these quantities. When fitting the standard GM to empirical data, the typical result is that the main factor determining trade volume is GDP. Adding geographical distances improves the fit significantly, while adding other dyadic properties is generally a small refinement. For these reason, we choose the node specific variable $\mathbf{n}_i$ to be the GDP of country, rescaled to the total world GDP $\mathbf{n}_i = g_i \equiv \frac{GDP_i}{\sum_j GDP_j}$ (as shown in Figure 2.4). Next, we choose the dyadic variable $\mathbf{D}_{ij}$ to be solely the distance between the two countries $\mathbf{D}_{ij} \equiv R_{ij}$. Thus,

$$F(\mathbf{n}_i, \mathbf{n}_j, \mathbf{D}_{ij}) \equiv \alpha \ (g_i \ g_j)^{\beta} \ R_{ij}^{\gamma}. \tag{2.41}$$

In this formation, g_i is adimensional and does not depend on the chosen currency or unit of money. On the other hand, it has been shown [7, 31] that the binary structure of the ITN can be excellently reproduced by setting

$$z_{ij} \equiv \delta g_i g_j. \tag{2.42}$$

Although in principle one can add the geographic distances into z_{ij} as well, empirical evidence shows that, contrary to the weighted case, the effect of distances on the binary structure of the ITN is not very strong [53]. This result motivates our choice above. In any case, adding extra factors is straightforward.

Putting the pieces together, our model is fully specified by the weight probability (2.34) with parameters given by eqs.(2.39) and (2.40), which in turn depend only on GDP and distance through eqs.(2.41) and (2.42). As a result we get the following expected values:

$$p_{ij} = \langle a_{ij} \rangle = \frac{\delta g_i g_j}{1 + \delta g_i g_j}, \tag{2.43}$$

$$\langle w_{ij} | a_{ij} = 1 \rangle = \alpha \ (g_i \ g_j)^{\beta} \ R_{ij}^{\gamma}, \tag{2.44}$$

$$\langle w_{ij} \rangle = p_{ij} \langle w_{ij} | a_{ij} = 1 \rangle = \frac{\alpha\delta (g_i g_j)^{\beta+1}}{1 + \delta g_i g_j} \ R_{ij}^{\gamma}. \tag{2.45}$$

Note that the presence of p_{ij} in the expression for $\langle w_{ij} \rangle$ implies that the latter takes the standard gravity form (2.9) only for those pairs of countries with large GDP, which are surely connected in the model. For lower values of the GDP, the expression is instead different, and for pair of countries with very low GDP one gets $\langle w_{ij} \rangle \approx \alpha\delta (g_i g_j)^{\beta+1} \ R_{ij}^{\gamma}$, i.e. the exponent of $g_i g_j$ is increased by one.

The model can also be interpreted as constructed from two random processes. As a first step, one determines whether a link is present or not with a probability p_{ij}. If a link (of unit weight) is indeed established, a second attempt determines whether the weight of the same link is increased by another unit (with probability y_{ij}) or whether the process stops (with probability $1 - y_{ij}$). Iterating this procedure, the probability that an edge with weight w is established between nodes i and j is given precisely by $q_{ij}(w)$ in eq.(2.34). The presence of these two degrees of freedom allows us to make a parallel, like in [33], with the economic literature about the so-called extensive and intensive margins of trade [61, 62, 63], defined as the preference for the network to evolve by either establishing new connections or strengthening the intensity of existing ones respectively.

2.6.3 Results

In this section we will compare the performance of the two models, in reproducing the observed properties (low-order and high-order) of the real complex trade network. We start with the trading volumes between countries, i.e. the weights of the existing links in the network. This property has been the main focus of the empirical economic literature on international trade, and it is the only property the classical gravity model is designed to reproduce.

In Fig 2.10 we can see a typical log-log plot of the expected values versus the real observed weights of the GM (as defined in eq. (2.9)). The three parameters of the model, α, β, and γ, are first fitted to the data, then plotted with the identity line (black line). The picture shows the typical good agreement between the

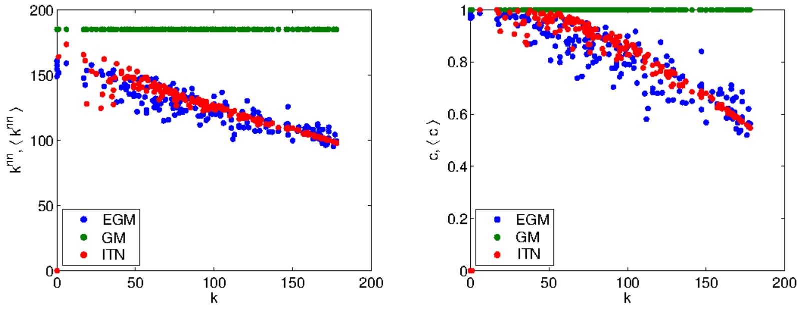

Figure 2.12: **The Enhanced Gravity Model, reconstruction of higher-order binary properties.** The heterogeneous binary topology of the observed ITN (red points) and the expected topology from the EGM (blue points) and the GM (green points). Left: the average nearest neighbour degree k^{nn} as a function of the observed degree k. Right: the clustering coefficient C as a function of the observed degree k.

predictions of the GM and the empirical non-zero trade volumes [15, 42, 43].

In Fig. 2.11 we compare the observed weights to both the gravity model (GM) and enhanced gravity model (EGM) conditional weights. Overall there is a very good agreement for both models; note that the EGM expectations are shifted to the right compared to the gravity model values, compensating for the higher expected number of zeroes. Despite the fact that in this study we use the classical gravity model in its simplest form, the EGM model can support more sophisticated models which incorporate additional dyadic information. Thus, the model has maximum flexibility when picking an expression for the conditional weights, and can be improved by using more refined models.

Binary Structure

In the binary representation, the main first-order property is the number of trade partners (connections) of each country, i.e. the degree sequence of the network. This simple yet important representation provides an added layer of considerable information to the standard results of traditional macroeconomic analyses of international trade. Recent studies have shown that higher-order binary properties, like the degree correlations (disassortativity) and clustering structure, of the ITN can be traced back to the knowledge of the degree sequence [21, 19]. This result indicates that the degree sequence, which is a purely topological property, needs to be considered as an important target quantity that international trade models, in contrast with the mainstream approaches in economics, should aim at reproducing.

In Fig. 2.12 we plot some higher-order topological properties of the ITN as a function of the degree of nodes, for the 2000 snapshot. These properties are the so-called average nearest neighbour degree and the clustering coefficient. For both quantities, we plot the observed values (red points) and the corresponding expected values predicted by the EGM (blue points) and the GM (green points). The exact expressions for both empirical and expected quantities are provided in sec. Materials and Methods. We see that the expected values of the EGM are in very close agreement with the observed properties, as opposed to the classical GM which consistently predicts a complete network. They show that at a binary level, the degree correlations (disassortativity) and clustering structure of the ITN are excellently reproduced by the EGM.

Weighted Structure

Despite the importance of the topology, the latter is only the backbone over which goods are traded, and the knowledge of the volume of such trade is imperative. The simplest weighted counterpart of the degree sequence is the strength sequence, i.e. the total trade of each country in the case of the ITN. Recent studies have shown that the higher-order binary quantities inferred from the strength sequence, as well as the corresponding weighted quantities, are very different from the observed counterparts [21, 22]. More specifically, the main limitation of models targeting only weighted properties, just like the gravity model, is that of predicting a mostly homogeneous and very dense (sometimes fully connected) topology. Roughly speaking, the models excessively ‘dilute’ the total trade of each country by distributing it to almost all other countries. This failure in correctly replicating the purely topological projection of the real network is the root of the bad agreement between expected and observed higher-order properties.

In fig. 2.13 we plot some higher-order weighted properties of the ITN as a function of the strength of nodes, for the 2000 snapshot. These properties are the so-called average nearest neighbour strength and the weighted clustering coefficient. For both quantities, we plot the observed values (red points) and the corresponding expected values predicted by the EGM (blue points) and the GM (green points). The exact expressions for both empirical and expected quantities are provided in sec. Materials and Methods. Again, we see that the expected values of the EGM are in very close agreement with the observed properties, as opposed to the classical GM. The bad performance of the traditional GM results directly from the unrealistically expected topology (complete network). This result highlights the importance of added structural information (degree sequence) to the models.

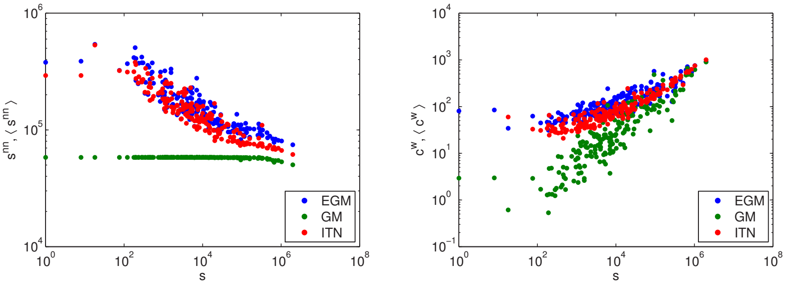

Figure 2.13: **The Enhanced Gravity Model, reconstruction of higher-order weighted properties.** Comparison between the observed weighted properties of the observed ITN (red points), the corresponding expected values of the EGM (blue points) and the gravity model (green points). Left: the average nearest neighbour strength s^{nn} as a function of the observed strength s. Right: the weighted clustering coefficient C^w as a function of the strength s.

2.7 Conclusions

In this chapter, we introduced a novel GDP-driven model which successfully reproduces both the binary and weighted properties of the ITN. The model uses the GDP of countries as a sort of macroeconomic fitness, and reveals the existence of strong relations between the GDP and the model parameters controlling the formation and the volume of trade relations. In the light of the limitations of the existing models (most notably the binary-only nature of the fitness model and the weighted-only nature of the gravity model), these results represent a promising step forward in the development of a unified model of the ITN structure. Later, we have introduced the EGM, which further improves these results and expand them by introducing additional macroeconomic parameters like geographic distance, aiming at bridging the gap between network-based and gravity-based approaches to the structure of international trade.

Theoretically, the EGM model originates within a maximum-entropy framework from a simple requirement of topological invariance under a change of money units. The maximum-entropy nature fixes the form of the weight distribution, thus removing an arbitrary ingredient of the original GM. Phenomenologically, the EGM allows us to reconcile two very different approaches that have remained incompatible so far: on one hand, the established GM which successfully reproduces non-zero trade volumes in terms of GDP and distance, while failing in predicting the correct topology [42]; on the other hand, network models which have been successful in reproducing the topology [7] but are more limited in their weighted structure [31]. Empirically, the EGM is the first model that can successfully reproduce the binary and the weighted empirical properties of the ITN simultaneously.

The EGM can be thought of as endowing the standard GM with a novel topologically invariant structure calibrated to replicate the binary properties of the ITN. Just like the standard GM, the EGM can accommodate additional economic factors in terms of extra-dyadic and country-specific properties.

Our results have strong implications for the theoretical foundations of trade models and the resulting policy implications. It is known that the traditional GM is consistent with a number of (possibly conflicting) micro-founded model specifications [64, 65, 66, 67]. For instance, a gravity-like relation can emerge as the equilibrium outcome of models of trade specialization and monopolistic competition with intra-industry trade [69, 70]. The empirical failure of the standard GM, which we ultimately traced back to its lack of topological invariance, implies a previously unrecognized limitation of these micro-founded models (and their policy implications) as well. At the same time, our results suggest a natural way to overcome this limitation via a topologically invariant reformulation of micro-founded models of trade, in such a way that a change in the units of trade volume has no impact on the resulting probability of trade among countries. How the policy implications of a model change as the mere result of this reformulation is an important point in the future research agenda. In general, we envisage the need for a new generation of micro-founded models that are consistent with the EGM. We, therefore, believe that the EGM can represent a novel benchmark supporting improved theories of trade and refined policy scenarios.

Bibliography

[1] R. Kali, J. Reyes (2007) 'The architecture of globalization: a network approach to international economic integration', *J. Int. Bus. Stud.* , Vol. 38, pp.595

[2] R. Kali, J. Reyes (2010) 'Financial contagion on the international trade networ', *Economic Inquiry* , Vol. 48, pp.1072

[3] S. Schiavo, J. Reyes, G. Fagiolo, (2010) 'International trade and financial integration: a weighted network analysis', *Quantitative Finance* , Vol. 10, pp.389

[4] F. Saracco, R. Di Clemente, A. Gabrielli, T. Squartini, (2015) 'Detecting the bipartite World Trade Web evolution across 2007: a motifs-based analysis', *arXiv:1508.03533* .

[5] A. Serrano, M. Boguna, (2003) 'Topology of the world trade web', *Phys. Rev. Lett.* , Vol. 68, pp.015101

[6] A. Serrano, M. Boguna, and A. Vespignani, *Patterns of dominant flows in the world trade web*, J. Econ. Interact. Coord.**2**, 111 (2007).

[7] D. Garlaschelli, M.I. Loffredo, (2004) 'Fitness-dependent topological properties of the World Trade Web', *Phys. Rev. Lett.* , Vol. 355, pp.188701

[8] D. Garlaschelli, M.I. Loffredo, (2005) 'Structure and Evolution of the World Trade Network', *Physica A* , Vol. 10, pp.138

[9] K.S. Gleditsch, (2002) 'Expanded Trade and GDP Data', *Journal of Conflict Resolution* , Vol. 46, pp.712

[10] http://comtrade.un.org/.

[11] A. Serrano, M. Boguna, A. Vespignani, (2007) 'Patterns of dominant flows in the world trade web', *J. Econ. Interact. Coord.*, Vol. 2, pp.111

[12] D. Garlaschelli, T. Di Matteo, T. Aste, G. Caldarelli, and M. Loffredo, (2007) 'Interplay between topology and dynamics in the World Trade Web', *Eur. Phys. J. B*, Vol. 57, pp.1434.

[13] G. Fagiolo, J. Reyes, S. Schiavo, (2008) 'On the topological properties of the world trade web: A weighted network analysis', *Physica A*, Vol. 387, pp.3868-3873

[14] G. Fagiolo, J. Reyes, S. Schiavo, (2009) 'World-trade web: Topological properties, dynamics, and evolution', *Phys. Rev. E*, Vol. 79, pp.036115

[15] G. Fagiolo, (2010) ‘The international-trade network: gravity equations and topological properties’, *J. Econ. Interact. Coord.*, Vol. 5, No. 5, pp.1-25

[16] M. Barigozzi, G. Fagiolo, D. Garlaschelli, (2010) ‘Multinetwork of international trade: A commodity-specific analysis’, *Phys. Rev. E*, Vol. 81, pp.046104

[17] L. De Benedictis, L. Tajoli, (2011) ‘The world trade network’, *The World Economy*, Vol. 34, pp.1417

[18] T. Squartini, D. Garlaschelli, *Analytical maximum-likelihood method to detect patterns in real networks* New J. Phys. **13**, 083001 (2011).

[19] G. Fagiolo, T. Squartini, D. Garlaschelli, (2013) ‘Null Models of Economic Networks: The Case of the World Trade Web’, *J. Econ. Interac. Coord.*, Vol. 8, No. 1, pp.75

[20] T. Squartini, R.Mastrandrea, D. Garlaschelli, (2015) ‘Unbiased sampling of network ensembles’, *New J. Phys.* , Vol. 17, pp.023052

[21] T. Squartini, G. Fagiolo, D. Garlaschelli, (2011) ‘Randomizing world trade. I. A binary network analysis’, *Phys. Rev.* , Vol. 84, pp.046117

[22] T. Squartini, G. Fagiolo, D. Garlaschelli, (2011) ‘Randomizing world trade. II. A weighted network analysis’, *Phys. Rev.* , Vol. 84, pp.046118

[23] A. Fronczak, P. Fronczak, J.A. Holyst, (2012) ‘Statistical mechanics of the international trade network’, *Phys. Rev. E*, Vol. 85, pp.056113

[24] M. Cristelli, A. Gabrielli, A. Tacchella, G. Caldarelli, L. Pietronero, (2013) ‘Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products’, *PLoS ONE*, Vol. 8, pp.0070726

[25] S. Sinha, A. Chatterjee, A. Chakraborti, B.K. Chakrabarti, (2010) ‘Econophysics: An Introduction’, *Wiley-VCH, Weinheim* .

[26] K. Bhattacharya, G. Mukherjee, J. Saramaki, K. Kaski, S.S. Manna, (2008) ‘The International Trade Network: weighted network analysis and modeling’, *J. Stat. Mech.*, pp.P02002

[27] Wells, S., (2004) ‘Financial interlinkages in the United Kingdom’s interbank market and the risk of contagion’, *Bank of England Working Paper*, No. 230/2004

[28] L. Bargigli, M. Gallegati, (2011) ‘Random digraphs with given expected degree sequences: A model for economic networks’, *J. Econ. Behav. & Organ.*, Vol. 78, pp.396

[29] N. Musmeci, S. Battiston, G. Caldarelli, M. Puliga, A. Gabrielli, (2013) 'Bootstrapping topological properties and systemic risk of complex networks using the fitness model', *J. Stat. Mech.*, Vol. 151, pp.720

[30] T. Squartini, D. Garlaschelli, (2014) 'an Tinbergen's legacy for economic networks: from the gravity model to quantum statistics', *Econophysics of Agent-Based Models*, Springer, pp.161-186

[31] A. Almog, T. Squartini, D. Garlaschelli, (2015) 'A GDP-driven model for the binary and weighted structure of the International Trade Network', *New J. Phys.*, Vol. 17, pp.013009

[32] R. Mastrandrea, T. Squartini, G. Fagiolo, D. Garlaschelli, (2014) 'Enhanced reconstruction of weighted networks from strengths and degrees', *New J. Phys.*, Vol. 16, pp.043022

[33] R. Mastrandrea, T. Squartini, G. Fagiolo, D. Garlaschelli, (2014) 'Reconstructing the world trade multiplex: the role of intensive and extensive biases', *Phys. Rev. E*, Vol. 90, pp.062804

[34] D. Garlaschelli, M.I. Loffredo, (2009) 'Generalized Bose-Fermi Statistics and Structural Correlations in Weighted Networks', *Phys. Rev. Lett.*, Vol. 102, pp.038701

[35] G. Caldarelli, A. Capocci, P. De Los Rios, M.A. Muñoz, (2002) 'Scale-free networks from varying vertex intrinsic fitness', *Phys. Rev. Lett.*, Vol. 89, pp.258702

[36] D. Garlaschelli, M.I. Loffredo, (2008) 'Maximum likelihood: Extracting unbiased information from complex networks', *Phys. Rev. E*, Vol. 78, No. 1, pp.015101

[37] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, D.R. White. *Economic networks: The new challenges*, Science **325(5939)**, 422 (2009).

[38] D. Garlaschelli and M. Loffredo, *Structure and Evolution of the World Trade Network*, Physica A **355**, 138 (2005).

[39] D. Garlaschelli, T. Di Matteo, T. Aste, G. Caldarelli, and M. Loffredo, *Interplay between topology and dynamics in the World Trade Web*, Eur. Phys. J. B **57**, 1434 (2007).

[40] T. Squartini and D. Garlaschelli, *In Self-Organizing Systems*, Lecture Notes in Computer Science 7166, pp. 24–35 (Springer, 2012).

[41] J. Tinbergen, *Shaping the World Economy: Suggestions for an International Economic Policy*, (The Twentieth Century Fund, New York, 1962).

[42] M. Duenas, G. Fagiolo, *Modeling the International-Trade Network: A Gravity Approach*, Journal Of Economic Interaction And Coordination **8** 155-178 (2013).

[43] L. De Benedictis, D. Taglioni, *The gravity model in international trade*, in The Trade Impact of European Union Preferential Policies, pp. 55-89 (Springer Berlin Heidelberg, 2011).

[44] D. Garlaschelli, M.I. Loffredo, *Patterns of link reciprocity in directed networks*, Physical Review Letters **93(26)**, 268701 (2004).

[45] F. Simini, M.C. González, A. Maritan, A. Barabási *A universal model for mobility and migration patterns* Nature **484**, 96–100 (2012).

[46] R. Glick and A.K. Rose *Does a Currency Union affect Trade? The Time Series Evidence*, NBER Working Paper No. 8396 (2001).

[47] A. K. ROSE and M.M. SPIEGEL *A Gravity Model of Sovereign Lending:Trade, Default, and Credit* IMF Staff Papers, Vol. 51, Special Issue, International Monetary Fund (2004)

[48] G.J. Linders and H.L.F. de Groot *Estimation of the Gravity Equation in the Presence of Zero Flows* Tinbergen Institute Discussion Paper No. 06-072/3 (2006).

[49] J. Silva and S.Tenreyro *The Log of Gravity* The Review of Economics and Statistics, **88**, issue 4, pages 641-658 (2006).

[50] V.Gemmetto, D. Garlaschelli *Multiplexity versus correlation: the role of local constraints in real multiplexes* Scientific Reports **5**, 9120. (2015).

[51] V. Gemmetto, T. Squartini, F. Picciolo, F. Ruzzenenti, D. Garlaschelli, *Multiplexity and multireciprocity in directed multiplexes* arXiv:1411.1282 (2016).

[52] Caldarelli, G., Chessa, A., Pammolli, F., Gabrielli, A., and Puliga, M *Reconstructing a credit network* Nature Physics, **9**, 125-126 (2013).

[53] F. Picciolo, T. Squartini, F. Ruzzenenti, R. Basosi, D. Garlaschelli, *The role of distances in the World Trade Web*, Proceedings of the Eighth International Conference on Signal-Image Technology & Internet-Based Systems (SITIS 2012), pp. 784-792 (edited by IEEE) (2013).

[54] G. Fagiolo, *Clustering in complex directed networks*, Phys. Rev. E **76**, 026107 (2007).

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

[56] V. Zlatic, G. Bianconi, A. D. Guilera, D. Garlaschelli, F. Rao, and G. Caldarelli, *On the rich-club effect in dense and weighted networks* Eur. Phys. J. B **67**, 271-275 (2009).

[57] J. Spitz, T. Kastelle, *Gains from Trade: the Impact of International Trade on National Economic Convergence; A Complex Network Analysis Approach*, Eastern Economic Association Annual Meeting, 2010.

[58] P. van Bergeijk, S. Brakman (eds.) *The gravity model in international trade* (Cambridge University Press, Cambridge, 2010).

[59] G. Fagiolo, *Directed or Undirected? A New Index to Check for Directionality of Relations in Socio-Economic Networks*, Economics Bulletin **3(34)**, 1-12 (2006).

[60] T. Squartini, F. Picciolo, F. Ruzzenenti, D. Garlaschelli, *Reciprocity of weighted networks*, Scientific Reports, 3 (2013).

[61] D. Ricardo, E. C. K. Gonner, and Q. Li, *The principles of political economy and taxation* (World Scientific, 1819).

[62] G. J. Felbermayr and W. Kohler, *Exploring the intensive and extensive margins of world trade*, Review of World Economics **142**, 642 (2006).

[63] L. De Benedictis and L. Tajoli, *The world trade network* The World Economy **34**, 1417 (2011).

[64] J. E. Anderson, *A Theoretical Foundation for the Gravity Equation*, American Economic Review, **69** (1), 106-16 (1979).

[65] J. Bergstrand, *The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence*, The Review of Economics and Statistics, vol. 67, issue 3, pages 474-81 (1985).

[66] A.V. Deardorff, *Determinants of Bilateral Trade:Does Gravity Work in a Neoclassical World?* NBER Working Paper, No. 5377, (1995).

[67] J.E. Anderson, E. Wincoop, *Gravity with Gravitas: A Solution to the Border Puzzle*, NBER Working Paper, No. 8079, (2001).

[68] Wilson, A. G. *The use of entropy maximising models in the theory of trip distribution, mode split and route split*, J. Transp. Econ. Policy 108-126 (1969).

[69] M.Fratianni, *Expanding RTAs, trade flows, and the multinational enterprise*, Journal of International Business Studies, Vol 40, **7**, 1206-1227, (2009).

[70] L.Benedictis and D. Taglioni *The Gravity Model in International Trade*, Springer, pp.55-89, (2011).

Chapter 3

Community Detection for Time Series

The mesoscopic organisation of complex systems, from financial markets to the brain, is the key intermediate level of organisation between the microscopic dynamics of individual units (stocks or neurons, in the mentioned cases), and the macroscopic dynamics of the system as a whole. The organisation is determined by "communities" of units whose dynamics, represented by time series of activity, is more strongly correlated internally than with the rest of the system. In the current literature, such emergent organisation is mainly detected through the measurement of cross-correlations among time series of nodes activity, the projection (usually via an arbitrary threshold) of these correlations to a network, and the subsequent search for denser modules (or so-called communities) in the network. It is well known that this approach suffers from an unavoidable information loss induced by the thresholding procedure. Another, less realized, limitation is the bias introduced by the use of network-based (as opposed to correlation-based) community detection methods. In this chapter we discuss an improved method for the identification of functional modules based on maximum-entropy. The method is threshold-free, correlation-based, and very powerful in filtering out both local unit-specific noise and global system-wide dependencies. The approach is guaranteed to identify mesoscopic functional modules that, relative to the global signal, have an overall positive internal correlation and negative mutual correlation.

The results presented in this chapter have been published in the following references:
A. Almog, F. Besamusca, M. MacMahon and D. Garlaschelli *PLoS ONE* **10** 7. e0133679 (2015).
A. Almog, O. Roethler, R. Buijink, J.H. Meijer, J.H.T. Rohling and D. Garlaschelli ,in preparation (2016).
R. Buijink, A. Almog, C.B. Wit, O. Roethler, D. Garlaschelli, J.H. Meijer, J.H.T. Rohling and S. Michel *PLoS ONE*, under second revision (2016).

3.1 Introduction

One of the most important properties of complex systems is community structure. Real-world systems are organized in a modular way, with clusters of units sharing similar dynamics or functionality. However, while the clusters themselves are internally cohesive, externally they can maintain contrasting dynamics. This emergent structure is typically resolved from the recorded activity time series of the system's fundamental units (such as stocks, neurons, etc.). The problem of resolving and identifying these mesoscopic structures, without any prior information, is extremely challenging. In financial markets, the mesoscopic scale corresponds to sets of stocks that share similar price dynamics. The knowledge of the market structure is highly valuable, and can assist in hedging risks and for better understanding of the market. Consequently, over the past years, scientists have deployed and developed many time series techniques to retrieve qualitative information regarding the hierarchy and structure of financial markets [1, 2, 3, 4].

A promising approach is that of employing community detection techniques, developed in network theory [5, 6], on empirical correlation matrices (constructed from multiple time series). However, The attempts made so far have basically replaced network data with cross-correlation matrices, and are not adapted to deal with correlation matrices. More specifically, the methods enforce a network representation on the empirical correlation matrix, and later apply a network-based (as opposed to correlation-based) community detection method. Such a process also involves some type of thresholding procedure, which results in significant information loss. These methodological problems rise significant limitations, when one comes to analyse empirical correlation matrices.

Recently, a novel method was proposed, which has been specifically designed to detect communities from correlation matrices of multiple time series [7]. The method is able to filter out both local unit-specific noise and global system-wide dependencies, using random matrix theory as the null model. When applied to financial time series, the method was able to capture the dynamical modularity of real financial markets. It is able to identify clusters of stocks which are correlated internally, but are anti-correlated with each other.

In this chapter we discuss this improved method and generalize it using a more complete framework of maximum-entropy. The generalized maximum-entropy approach is able to capture both the global dynamics of the system and the random noise induced by the different units. More importantly, this improved null model is more "data-driven" in contrast to the original null model, enabling us to use more empirical information to construct the community structure. In this setting, the null models help us identify the non-random properties in the system to a higher level of resolution.

The rest of the chapter is organized as follows. In section 3.3 we analyse financial markets, exploring the mesoscopic structures induced by both the original weighted time series and their corresponding binary signatures. We show that both the binary and weighted information yield very similar community structures, suggesting that the binary signatures carry significant structural information. In section 3.4 we apply the method to the suprachiasmatic nucleus of mice, which is a complex network of oscillating neurons, revealing a remarkable core periphery structure. This application to functional brain networks comes as a by-product, as time series of brain activity have in common with financial time series the potential of being driven by a strong signal and also being subject to noise. In fact our method could have a great impact on the field of functional brain networks. Finally, in sec 3.5 we summarize our results and provide some conclusions.

3.2 Maximum-entropy approach to community detection

In general, the problem of community detection in network science consists of finding clusters of nodes that have dense connections internally and sparser connections externally. Such clusters, or communities, represent sub-units of the network like families in social networks or brain regions in structural brain networks. Identifying these modules and their boundaries is of great importance, and over the last years, many methods have been developed to resolve this type of organisation [5]. However, here we are interested in identifying functional modules from a correlation matrix, generated by multiple time series, rather than a conventional network.

We now describe the so-called modularity-based community detection methods but adapted to correlation matrices. This restricts us to undirected networks, given the symmetry property of correlation matrices. Let us consider a network with N nodes. One can introduce a number of partitions of the N nodes into non-overlapping sets. The different partitions will be represented by an N-dimensional vector $\vec{\sigma}$ where the i-th component σ_i denotes the set in which node i is placed by that particular partition. Now, we introduce the modularity measure $Q(\vec{\boldsymbol{\sigma}})$ which indicates the quality of a particular choice of partition $\vec{\sigma}$ measured by a high degree of inter-community connectivity and a low degree of intra-community connectivity. So-called modularity optimization algorithms look for the specific partition that maximizes the value of $Q(\vec{\boldsymbol{\sigma}})$, the objective function. The latter is defined as

$$Q(\vec{\boldsymbol{\sigma}}) = \frac{1}{A_{tot}} \sum_{i,j} \left[A_{ij} - \langle A_{ij} \rangle\right] \delta(\sigma_i, \sigma_j) \tag{3.1}$$

where $\delta(\sigma_i, \sigma_j)$ is a delta function ensuring that only pairs of nodes within the same community contribute to the sum, and A_{ij} is the adjacency matrix that indicates whether a link exists between the nodes, ($A_{ij} = 1$) or not, ($A_{ij} = 0$). The pre-factor A_{tot} serves to normalize the value of $Q(\vec{\boldsymbol{\sigma}})$ between -1 and 1, where $A_{tot} \equiv \sum_{i,j} A_{ij} = 2L$ is twice the number of total links in the network. The term $\langle A_{ij} \rangle$ is vital to the outcome of the community detection process. It represents the expectation of whether a link exists or not, according to the specific null model chosen. So far the majority of the methods use null models (hypotheses) which are suited only for networks. For example, the configuration model is a null model that preserves the degree sequence of the network. It has been shown that such null models can introduce biases when applied to correlation matrices [7]. The reason is that, while a null model for a network assumes independent links, correlation matrices have different metric properties that imply dependencies among their entries, even under the null hypothesis. A recent method proposed a redefinition of the modularity, which does take into account the existence of known properties of correlation matrices (see sec. 3.3.1). Here, we take the maximum-entropy approach, providing a specification of an appropriate null model.

Instead of the previous adjacency matrix A_{ij}, we input the empirical correlation matrix C_{ij}. The method defines the modularity as

$$Q(\vec{\boldsymbol{\sigma}}) = \frac{1}{C_{norm}} \sum_{i,j} \left[C_{ij} - \langle C_{ij} \rangle_{me}\right] \delta(\sigma_i, \sigma_j) \tag{3.2}$$

where $\langle C_{ij} \rangle_{me}$ is a maximum-entropy null model that needs to identify the random properties of empirical correlation matrices.

In this approach, the empirical correlation matrix is first decomposed and then reconstructed using only the eigenvalues (and eigenvectors) that are not reproduced by the random null model. Thus in this chapter, contrary to chapter 1, we are interested in the correlation matrix spectrum which the random model (multiple time series) generates. Once compared with the observed spectrum of the empirical correlation matrix, the model will identify the non-random eigenvalues (by elimination). The non-random eigenvalues will be later used to generate the new filtered matrix. In the next sections, we introduce two null models, which will serve us as the "random benchmark" in this new definition of modularity.

3.2.1 Random time series

Let us go back to the same multiple time series formalism as in chapter 1, where the system describes N time series with length T ($N \times T$ matrix). We recall that the entries of such a general matrix $\mathbf{M}$ are denoted by $r_i(t)$, where i labels the stock and t labels the time step. Before introducing the specific models, we need to make some necessary changes. Here, we extend our null models to weighted

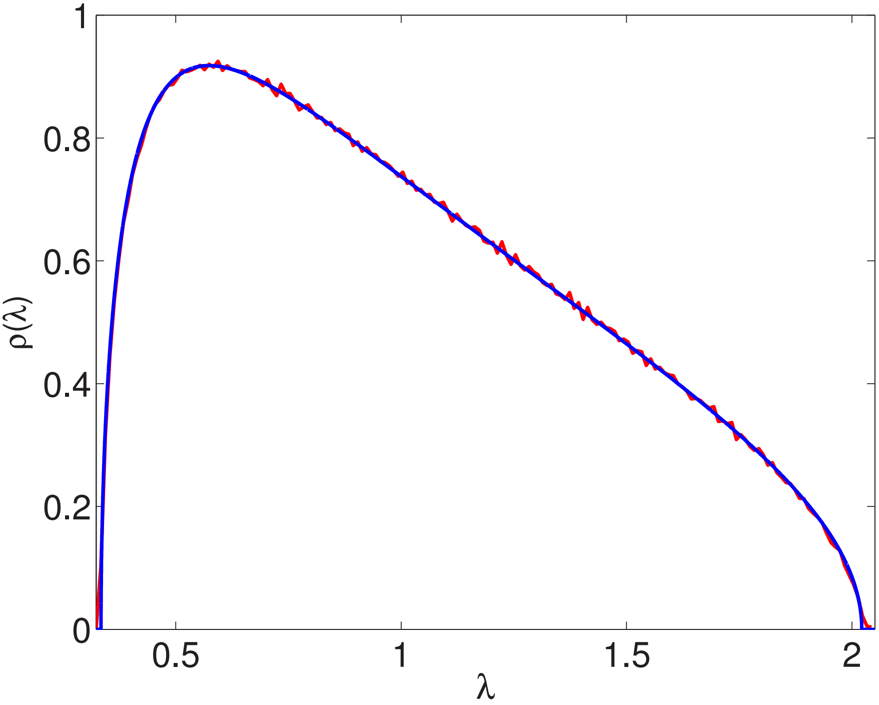

Figure 3.1: **Eigenvalues density distribution comparison: maximum-entropy versus Wishart matrix.** The eigenvalue density distribution of the correlation matrix of the time series generated by the solved maximum-entropy model for the $S\&P500$. In red is the sampled density distribution from 1000 runs, in blue is the theoretical Marchenko Pastur distribution. The figure present a good agreement between the maximum-entropy model and the known analytical curve of Wishart matrix.

returns, where $-\infty < r_i(t) < \infty$, and the constraints are enforced over the whole matrix. Moreover, since the system is weighted we want to apply a normalization condition on the distribution of the returns.

We start with the simplest case of N time series with no constraints. However, contrarily to the binary (± 1) case considered in chapter 1, here we are required to enforce one extra constraint over the second moment

$$\langle \frac{1}{NT} \sum_{i}^{N} \sum_{t}^{T} r_i^2(t) \rangle \equiv [r_i^*(t)]^2_{average} \tag{3.3}$$

to ensure that the probability distribution for $r_i(t)$ is normalised. The Hamiltonian reads

$$H(\mathbf{X},\alpha) = \alpha \sum_{i=1}^{N}\sum_{t=1}^{T} r_i^2(t) \tag{3.4}$$

with α as a single free parameter, which leads us to a normal distribution.

Using maximum likelihood estimation as before, one finds

$$\alpha^* = TN \left(2\sum_{i=1}^{N}\sum_{t=1}^{T} [r_i^*(t)]^2 \right)^{-1} \tag{3.5}$$

as the value of α.

Once solved, this model is a maximum-entropy equivalent to the generic case, where one measures the correlation between N independent random time series for T time steps (the observed period). In this specific case, the resulting correlation matrix would be an $N \times N$ Wishart matrix, whose statistical properties are well-known [19, 21]. In the limits where $N, T \to \infty$ and $T/N \geq 1$ the eigenvalues of the Wishart matrix are distributed according to a Marchenko-Pastur distribution

$$P(\lambda) = \frac{T}{N}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\lambda} \qquad \text{if} \qquad \lambda_- \leq \lambda \leq \lambda_+ \tag{3.6}$$

and $P(\lambda) = 0$ otherwise. The boundaries λ_+ and λ_- are dependent on the data size and given by

$$\lambda_\pm = \left[1 \pm \sqrt{\frac{N}{T}} \right]^2 . \tag{3.7}$$

This analytic curve represents the boundaries of the bulk eigenvalues, which predominantly represent noise, and so have little meaning assigned to them.

In Fig 3.1 we plot the eigenvalue density distribution of the correlation matrix of the time series generated by the maximum-entropy model for the $S\&P500$. In red is the empirical density sampled from 1000 runs, in blue is the theoretical Marchenko Pastur distribution eq.(3.6). It is clear that both distributions are almost identical, validating our claim of correspondence between the models when measuring the eigenvalues spectrum. We should note that although we use the parameter σ which is extracted by eq. (3.5), it does not play any role in the eigenvalue distribution. The distribution is invariant to different σ, which explains the agreement with the theoretical curve.

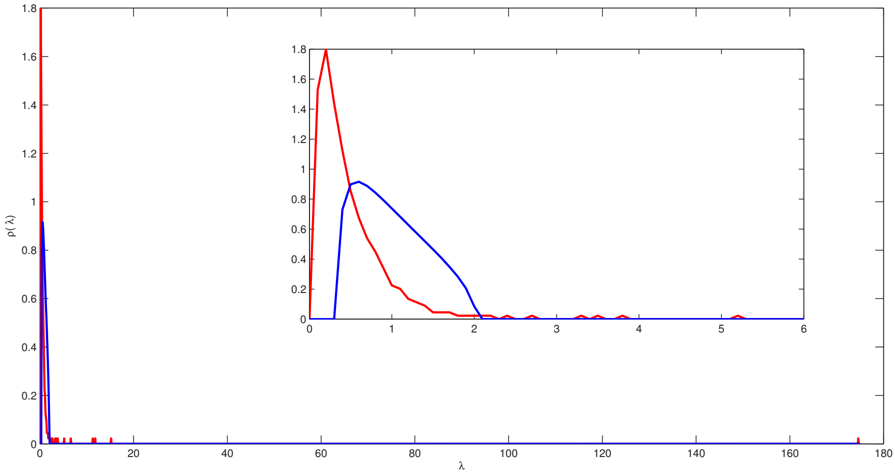

Figure 3.2: **Eigenvalues density distribution comparison: empirical correlation versus Wishart matrix.** The eigenvalue density distribution of the correlation matrix of the $S\&P500$. In red is the empirical eigenvalue distribution, in blue is the Marchenko-Pastur distribution. The figure shows the bad agreement of the Wishart matrix with respect to the empirical spectrum.

The empirical eigenvalues outside this range, however, have structural implications and relate to groups of correlated stocks [21]. As a result, any empirical correlation matrix C can be identified as a sum of of two matrices:

$$C = C^{(r)} + C^{(s)}, \tag{3.8}$$

where $C^{(r)}$ is the random part aggregated from the eigenvalues in the random spectrum ($\lambda_- \leq \lambda \leq \lambda_+$), i.e.

$$C^{(r)} \equiv \sum_{i:\lambda_- \leq \lambda_i \leq \lambda_+} \lambda_i |v_i\rangle\langle v_i|, \tag{3.9}$$

and $C^{(s)}$ is the "structured" component, which is composed from those eigenvalues above the boundary of the bulk eigenvalues. $\lambda > \lambda_+$.

Moving forward, in financial markets, it is well established that stocks typically move up or down together, an effect known as the "market mode". This effect is indicated by the presence of a very large eigenvalue λ_m, orders of magnitude greater than the rest. This can be observed in Fig 3.2 where we compare the eigenvalue density distribution (of the cross-correlation matrix) of the $S\&P500$ index to the Marchenko-Pastur distribution. Since this eigenvalue represents a common factor influencing all the stocks in a given market, from a structural perspective, the market mode eigenvalue signifies the presence of one single super-community, containing all the stocks in the market.

Thus, the other eigenvalues (not including the market mode), which deviate from the bulk, $\lambda_+ < \lambda_i < \lambda_m$ are the ones corresponding to mesoscopic clusters, i.e. groups of stocks with similar dynamics. This observation results in a further decomposition of the empirical correlation matrix

$$C = C^{(r)} + C^{(g)} + C^{(m)}, \tag{3.10}$$

where

$$C^{(m)} \equiv \lambda_m |v_m\rangle\langle v_m| \tag{3.11}$$

represents the market mode, and

$$C^{(g)} \equiv \sum_{i:\lambda_+ < \lambda_i < \lambda_m} \lambda_i |v_i\rangle\langle v_i| \tag{3.12}$$

represents the remaining correlated groups. These sub-groups of correlated stocks comprise the mesoscopic structure of the market. They are also referred as "group modes" in the literature [1, 21].

To conclude, this approach actively filters both the null model spectrum and the market mode (largest eigenvalue) and has been introduced in [7]. However, there are a few drawbacks to this approach. Firstly, the global mode is not generated by the null model itself, and we are required to filter it "manually". The reason is that the null model only considers N random variables, and the presence of the market model is purely an empirical fact which is not being account for in the null hypothesis. Next, the random bulk distribution is only dependent on the size of the data (T and N), and not on the data itself. One can imagine that different systems would have different dynamics which requires different types of spectral filtering. Lastly, the null model cannot reproduce the so-called sub-random eigenvalues, where $\lambda_i < \lambda_-$, as we can see in Fig 3.2. This is quite alarming since the sub-random spectrum contains the majority of the empirical eigenvalues. In the next section, we will introduce a new null model that can overcome these limitations.

3.2.2 Random time series with global mode

Here we introduce a more sophisticated maximum-entropy null model, which enforce added "structural information", trying to overcome the random model limitations. The model is designed to enforce also the average daily return, hence capturing the non-stationary nature of the system. Thus, the constraints are

$$\langle \frac{1}{NT}\sum_{i}^{N}\sum_{t}^{T} r_i^2(t)\rangle \equiv [r_i^*(t)]^2_{average} \qquad \langle \frac{1}{N}\sum_{i=1}^{N} r_i(t)\rangle \equiv \{r_i^*(t)\} \;\; \forall t \tag{3.13}$$

recalling that we denote $\{r_i(t)\} \equiv \frac{1}{N}\sum_i r_i(t)$ as the average over stocks.

This step is motivated by section 1.6, where we impose the information separately for each day.

Combining all the above considerations, we finally generalize the model defined by eq.(1.59) to the matrix case as follows:

$$H(\mathbf{X}, \alpha, \vec{\theta}) = \alpha \sum_{i=1}^{N}\sum_{t=1}^{T} r_i^2(t) + \sum_{t=1}^{T}\theta_t \sum_{i=1}^{N} r_i(t) = \sum_{it}\left[\alpha r_i^2(t) + \theta_t r_i(t)\right] \tag{3.14}$$

where $\vec{\theta}$ it a T-dimensional vector with entries $\theta(t)$.

Using maximum likelihood estimation, in this case one needs to solve $T+1$ coupled equations

$$\alpha^* = TN\left(2\sum_{i=1}^{N}\sum_{t=1}^{T}(r_i^*(t) - \{r_i^*(t)\})^2\right)^{-1} \tag{3.15}$$

$$\theta_t^* = -2\alpha^*\{r_i^*(t)\} \tag{3.16}$$

in order to find the value of α and $\vec{\theta}$.

Once solved, this maximum-entropy model corresponds directly to a one-factor model defined in sec. 1.6.4. If we now define σ and μ as follows

$$\sigma^* \equiv \sqrt{\frac{1}{2\alpha^*}} \quad , \quad \mu_t^* \equiv \{r_i^*(t)\} \tag{3.17}$$

the probability of a matrix (multiple time series) is the product of Gaussian distributed random variables p_{it}

$$P(\mathbf{X}) = \prod_{it}\underbrace{\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{\frac{-1}{2\sigma^2}\sum\limits_{\mathrm{it}}(\mathrm{r_i(t)} - \mu_\mathrm{t})^2}}_{p_{it}} \tag{3.18}$$

resulting from the integration (from $-\infty$ to ∞) of the exponential (partition function). The components of the multiple financial time series can thus be sampled as follows

$$r_i(t) \sim \mathbf{N}\left(\mu_t^*, \sigma^*\right), \tag{3.19}$$

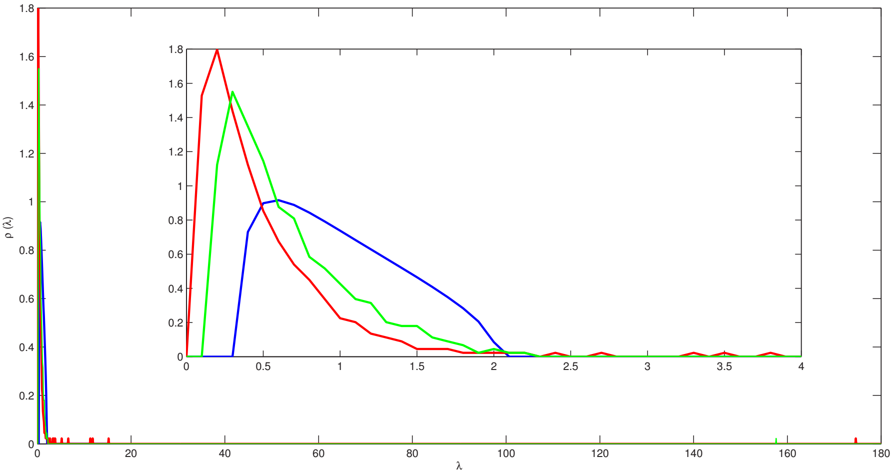

Figure 3.3: **Eigenvalues density distribution comparison: maximum-entropy model with global mode versus Wishart matrix.** The eigenvalue density distribution of the correlation matrix of the time series generated by the solved maximum-entropy model for the $S\&P500$. In red is the empirical eigenvalue distribution, in green is the sampled density distribution from 1000 runs (new null model), in blue is the Marchenko-Pastur distribution. The figure shows the spectrum of the improved null model with respect to the previous one, and how it can better replicate the random bulk and global mode that are observed in the empirical spectrum.

where $\mathbf{N}$ represents a random variable with normal distribution with mean μ_t^* specific to each time step and variance σ^*.
In Fig 3.3 we plot the eigenvalue density distribution (of the cross-correlation matrix) for the $S\&P500$ index. The green curve is the sampled eigenvalue distribution of the maximum-entropy model, and the red curve is the empirical eigenvalue distribution and the blue curve the Marchenko-Pastur distribution. We can see that the new null model can replicate the global mode (largest eigenvalue). Furthermore, there is a shift in the random bulk, which is in agreement with the empirical distribution. This results present an improved null model which is able to generate both the global mode and the random bulk of the empirical data,

$$\langle C \rangle_{me} = C^{(m)} + C^{(r)} \tag{3.20}$$

using partial information from the original data.

Returning to the modularity, we define the filtered empirical correlation matrix $C_{ij}^{(g)}$ filtering both the global mode $C^{(m)}$ and the random bulk $C^{(r)}$.

Once we input this result into the modularity eq.(3.2.2)

$$Q(\vec{\boldsymbol{\sigma}}) = \frac{1}{C_{norm}} \sum_{i,j} \left[C_{ij} - \langle C_{ij} \rangle_{me}\right] \delta(\sigma_i, \sigma_j) = \frac{1}{C_{norm}} \sum_{i,j} (C_{ij} - C_{ij}^{(r)} - C_{ij}^{(m)}) \delta(\sigma_i, \sigma_j)$$

we see that this leads to

$$Q(\vec{\boldsymbol{\sigma}}) = \frac{1}{C_{norm}} \sum_{i,j} C_{ij}^{(g)} \delta(\sigma_i, \sigma_j). \tag{3.21}$$

In other words, to clearly differentiate between the mesoscopic groups, one must subtract out the main drift of the system and the random correlation, using the maximum-entropy null model. The filtered matrix $C_{ij}^{(g)}$ constituted from the "non-random" eigenvalues $\lambda_+ < \lambda_i < \lambda_m$ and their corresponding eigenvectors v_i. The new method modified three modern community detection algorithms, customizing where necessary to be effective with correlation matrices [7]. The three algorithms we use in this paper are known as the Potts (or spin glass) method [12, 13], the Louvain method [14] and the spectral method [15]

3.3 Financial markets

Traditionally, the main object of time series analysis is the characterization of patterns in the amplitude of the increments of the quantities of interest (stock price in our case). The analysis requires a weighted description of the system, i.e. both the amplitude and the sign of the activity. Indeed, a time series of increments enclose complete information about the amplitude of the fluctuations of the original signal. However, a significant part of this information is encoded in the purely 'binary' projection of the time series, i.e. its sign. Recent studies have shown various forms of statistical dependency between the sign and the absolute value of fluctuations [8, 9, 10]. In chapter 1, we have shown a robust empirical relationship between binary and non-binary properties of real financial time series [11]. The results show that binary signatures, which retain only the sign of fluctuations, encode significant information regarding the full behaviour of the stock (both amplitude and direction). Motivated by these results, here we further explore the higher-order relations between financial time series and their corresponding binary signatures, in a more complex setting. In this section we study whether the binary signatures of assets can reproduce the same complex community organisation of financial markets, as the weighted information.

To this end, we use the daily closing prices of the stocks of three indexes (S&P500, FTSE100 and NIKKEI225) over the period 2001-2011. For each index, we restrict our sample to the maximal group of stocks that are traded continuously throughout the selected period. This results in 445 stocks for the S&P500, 78 stocks for the FTSE100 and 193 stocks for the NIKKEI225. Given a stock price

$P_i(t)$ where i denotes one of the N stocks in the index, and t denotes one of the T observed temporal snapshots (days), the log-return is defined as

$$r_i(t) \equiv \log\left[\frac{P_i(t)}{P_i(t-1)}\right], \tag{3.22}$$

where $2 < t < T$.
For each stock in the system we use the time series of it's log-returns for our analysis. This is the construct we refer to as the "weighted time series" throughout the rest of the chapter. In contrast, the "binary signatures" only reveal the direction of the fluctuation (sign) in the price and are defined as

$$x_i(t) \equiv \mathrm{sign}[r_i(t)] = \begin{cases} +1 & r_i(t) > 0 \\ 0 & r_i(t) = 0 \\ -1 & r_i(t) < 0 \end{cases}. \tag{3.23}$$

In Fig. 1.1 from chapter 1 we show a simple example of a weighted time series, along with the corresponding binary projection.

The two types of information are in fact different descriptions of the same system, and are used to construct cross-correlation matrices. In turn, we deploy three popular community-detection algorithms [12, 13, 14, 15] specifically adapted, where necessary, for the correct use of cross-correlation matrices [7]. We examine and quantify similarities and variations in the organisation of the markets for these two representations. This approach reveals some interesting results. First, we can quantify the level of information encoded within the binary signatures, with respect to the full weighted time series. Secondly, we observe that both the binary and weighted representations yield very similar structures, which indicates that most of the information regarding the structure of financial communities is already encoded within the sign of a stock.

3.3.1 Spectral analysis

In this section we analyse the eigenvalue density distribution of the cross-correlation matrices for the two representations of the data (binary and weighted). When plotting the density distribution one can identify specific spectral properties that have structural implications. In other words, it is possible to identify distinct eigenvalues in the spectrum, which correspond to correlated clusters of stocks, and typically indicate a non-trivial structure (as defined in sec. 3.2.2) .

Our focus here is to detect these non-random eigenvalues in the spectrum of both the binary and weighted data. Moreover, we want to explore the similarities and differences between the two spectra to inform us about the corresponding structures yielded by each type of data.

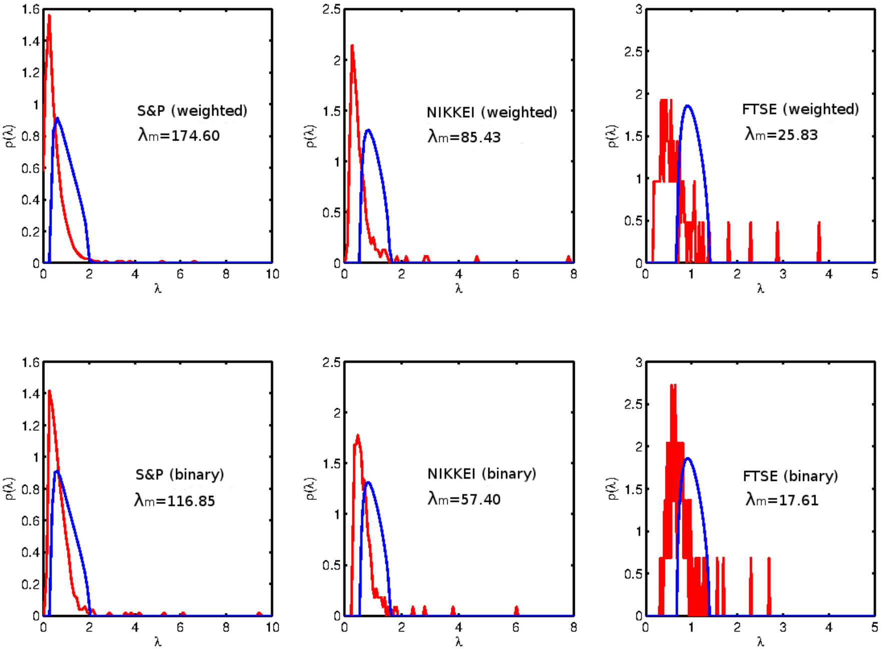

Figure 3.4: **Signatures of complex structure in the eigenvalue spectrum.**The eigenvalue density distribution (of the cross-correlation matrix) for the different indexes, where the upper panels are for the weighted series and the lower panels are for the binary series. The red curve is the empirical eigenvalue distribution and the blue curve the Marchenko-Pastur distribution. The largest empirical eigenvalue λ_m is not shown in the plots, but its value is reported in each panel. The figure shows that the complex structure, composed from global, group, and random modes, is also maintained in the binary representation.

In Fig. 3.4 we plot the eigenvalue density distribution for the three different indices. The top row corresponds to the weighted representation (log-returns), and the bottom row corresponds to the binary representation (binary signatures). We can observe the known structure of the financial markets in the weighted data, however this complex structure also exists in the binary data. This result is non trivial. We can observe a market mode, and several deviating eigenvalues also in the “simpler” binary data (with the same order of magnitude).

We also want to inspect whether both descriptions of the system function the same under randomization. The returns of each stock were separately permuted randomly, therefore preserving the total return of the stocks and destroying the daily correlation between the returns. Once the time series entries are shuffled, both binary and weighted correlation matrices end up as elementary random

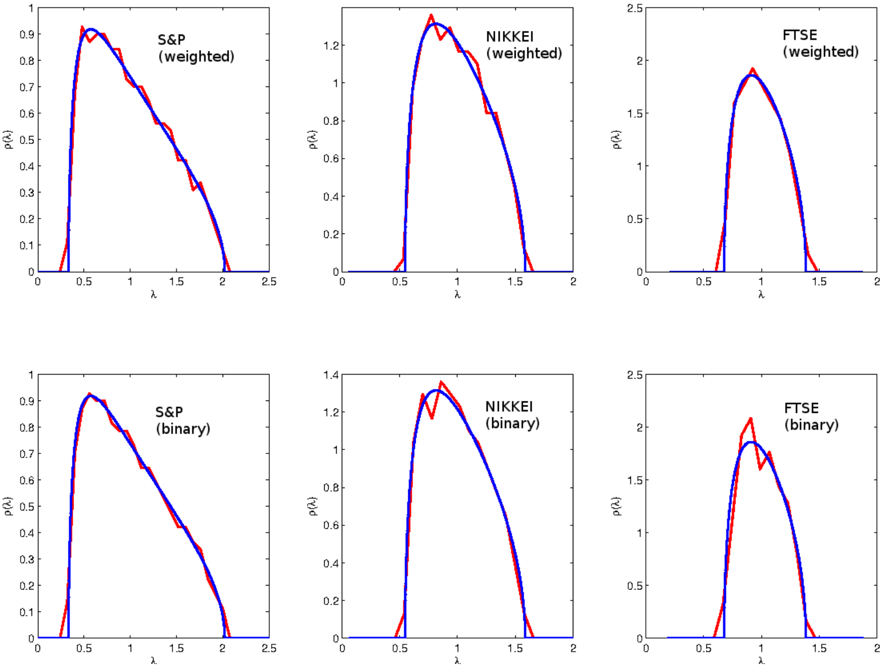

Figure 3.5: **Eigenvalues density distribution of randomized empirical data.** The eigenvalue density distribution of the Pearson correlation matrices where the upper panels are for the weighted series and the lower panels are for the binary series. The red curve is the empirical eigenvalue distribution and the blue curve the Marchenko-Pastur distribution. The figure shows the collapse to the random distribution once the original data is shuffled.

matrices. As discussed before, the eigenvalues of such matrices will be distributed with a Marchenko-Pastur distribution.

In Fig. 3.5 we plot the density distribution of the shuffled data for the three different indices. The top row corresponds to the weighted representation (log-returns), and the bottom row corresponds to the binary representation (binary signatures). As expected, in both cases we observed the known characteristics of a random matrix. The spectra of both representations collapsed to the known analytic curve.

To sum up this section, we identified a sub-group structure both in the weighted and the binary representation of the three indices. Each of the binary spectra we studied retain all the known properties of a “regular” (weighted) spectrum (random bulk, market mode and group modes). This result propels us to do a more refined analysis, and to further explore (and compare) the sub-group structure

Consumer Discretionary:	🟪	Consumer Staples:	🟦
Energy:	🟦	Financials:	🟩
Health Care:	🟥	Industrials:	🟧
Information Technology:	🟦	Materials:	🟨
Telecom. Services:	🟪	Utilities:	🟫

Table 3.1: The 10 industry sectors in the Global Industry Classification Standard (GICS), with the colour representation used to highlight the sectors in the following Figures.

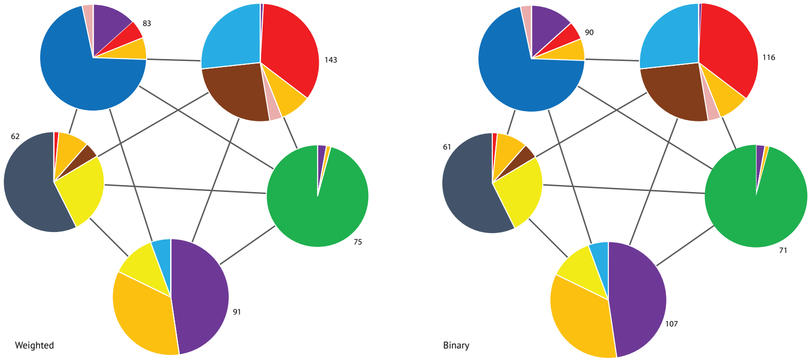

Figure 3.6: **S&P community structure.** Communities of the S&P 500 (daily closing prices from 2001 to 2011) generated using the modified Louvain algorithm [7]. Each community is labelled with the number of stocks and the pie chart represents the relative composition of each community based on the industry sectors of the constituent stocks (colour legend in Table 1). The inter-community link weights are negative, indicating that the communities are all residually anti-correlated. We can see that the binary partition is almost identical to the weighted one.

of the different indices. In the next section we will apply community detection algorithms to extract a more detailed structure for both representations, so that we can better quantify the similarities and the variations.

3.3.2 Community structure

In network theory, a community structure is the partition of the network into relatively dense sets of nodes, with respect to the rest of the network. More specifically, it is the organisation into clusters of nodes with dense connections internally, while the connections between the clusters are sparser. Community detection is the identification of such clusters of agents (nodes) in the system (network). In the last decade there has been a burst of research concerning this topic, across a myriad of different fields [5].

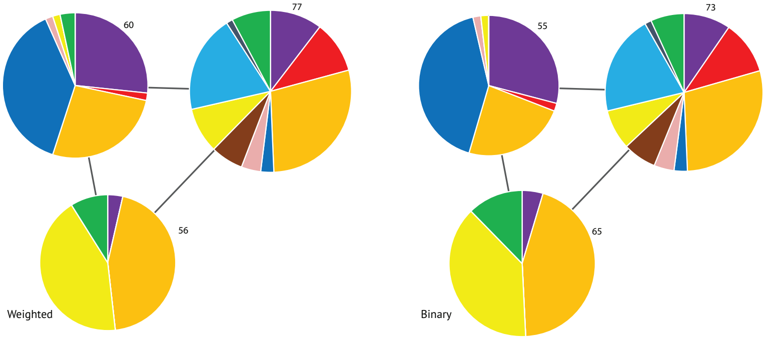

Figure 3.7: **Nikkei community structure.** Communities of the Nikkei 225 (daily closing prices from 2001 to 2011) generated using the modified Louvain algorithm [7]. Each community is labelled with the number of stocks, and the pie chart represents the relative composition of each community based on the industry sectors of the constituent stocks (colour legend in Table 1).The link weights are negative, indicating that the communities are all residually anti-correlated. We can see that the binary partition is almost identical to the weighted one.

This promising approach has also been applied to analyse time series data [22, 23, 24, 25] where the goal is to identify clusters of components with a similar dynamics. The attempts made so far have basically replaced network data with cross-correlation matrices as the input. However, this procedure suffers from a significant limitation. The null hypotheses used in the network-based algorithms are inconsistent with the properties of correlation matrices. As a result, these approaches can introduce an undesired bias when applied to the detection of communities in time series based networks.[7]. Here we adopt a new method [7], which is specifically shaped to deal with correlation matrices, based on the spectral properties we presented in the previous section. The method presents an improved and consistent way to cluster multiple time series, by leveraging a set of null models, specifically designed for use with correlation matrices.

Applying the new approach, we use three popular community detection algorithms, customizing where necessary to be effective with correlation matrices. The three algorithms we use in this chapter are known as the Potts (or spin glass) method [12, 13], the Louvain method [14] and the spectral method [15], and are modified in [7] to correctly deal with correlation matrices. The algorithms use a modularity optimization process, where modularity is a measure for how "optimal" your partition is. The algorithms attempt to choose a specific partitioning of the network into groups such that the corresponding value of the modularity is maximized (see Methods). The modularity implements a new null hypothesis, which is fitted to time series (correlation matrices). More specifically, the hypothesis considers the empirical correlation matrix as a superposition of modes eq.(3.10), and decomposes it accordingly. Both the random mode and market mode are filtered out, and we are left with only the informative group mode, which is then used to extract the market structure.

Thus, we end up with three community detection algorithms that are consistent with time series data and represent the counterparts of the most popular techniques used in network analysis. The method allows us to explore the mesoscopic structure of different financial indices, and more specifically compare the different community structures resulting from the different representations (binary and weighted).

First, we perform community detection on both the binary and the weighted time series, using all three community detection methods, for the full time period of the data (2001-2011). We pick the division that maximizes the modularity (for a specific representation and algorithm), and compare the results for the two types of information. In the case where several divisions maximize the modularity (different runs result in different divisions), we take the division with the most occurrences over 1000 runs (the highest probability). Here, we want to identify groups with similar dynamics over the ten year period, with the rationale that such a long period will reduce the noise.

To help further explore the communities resulting from the different passes, we label each of the stocks according to its industry sector from the Global Industry Classification Standard (GICS). This classification represents a more "traditional" frame of mind where the different sectors are comprised of stocks conforming to a particular, qualitative description of the industry they represent. Recent results show that real markets have a more complex structure [2, 26, 27, 16], where different sectors are mixed in different sub-groups, i.e. the communities are assembled out of stocks from different sectors. Furthermore, the classification helps us to compare the results of the two representations in a very clear and visual way.

In Figs. 3.6 and 3.7 we plot the community structure of the S&P 500 and Nikkei 225 (daily closing prices from 2001 to 2011) generated using the modified Louvain algorithm. Each community is labelled with the number of stocks and the pie chart represents the relative composition of each community based on the industry sectors of the constituent stocks (colour legend in Table 1). The links between the communities represent "residual" (i.e filtered) anti-correlation relations [7]. We can see that the binary partition is very similar to the weighted one. For both indexes about $7-8$ percent of the stocks switch community. In the next section we will give a more quantitative measure for the dissimilarities of the different partitions.

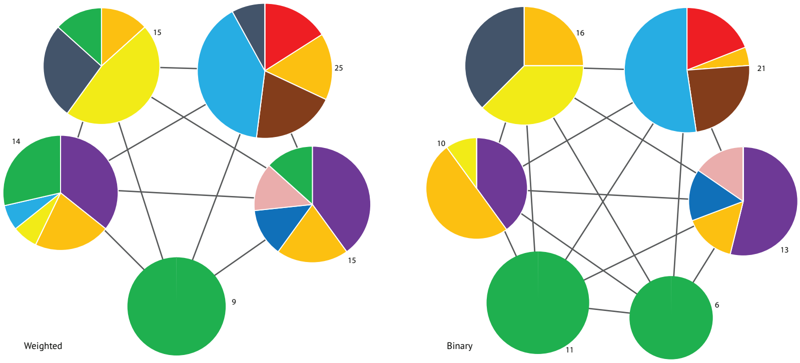

Figure 3.8: **FTSE community structure.** Communities of the FTSE 100 (daily closing prices from 2001 to 2011) generated using the modified Louvain algorithm [7]. Each community is labelled with the number of stocks, and the pie chart represents the relative composition of each community based on the industry sectors of the constituent stocks (colour legend in Table 1). The inter-community link weights are negative, indicating that the communities are all residually anti-correlated.

In Fig. 3.8 we observe a more complex result. Again, we plot the community structure of the FTSE100 (daily closing prices from 2001 to 2011) generated using the modified Louvain algorithm. Now, the binary representation consistently identifies one more community than the weighted representation (for all the algorithms). Later, we further explore these differences in community structure. Most notably, the binary information results in a cluster configuration where the Financials sector (green) was partitioned into two communities, whereas the weighted representation created only one sole Financials community, spreading the rest of the stocks among other clusters.

We should note that, purely from a community detection perspective, there is no "correct" partition. Each partition is generated from different data and so maximizing modularity should not be expected to yield the same partitions. We are comparing the end results of these processes, and in this setting (this chapter) we treat the weighted partition as the "truth" since it is using a priori more information to establish the partition. Thus, our aim is merely to examine the degree to which the "binary" community structure matches the "weighted" community structure, despite having less information for the algorithms to work with. That said, it would be interesting to determine if the binary information can yield different points of view, or added structural information with respect to its weighted counterpart. This would be useful in particular because binary time series are more robust to noise and errors in the data. Such a line of exploration is however beyond the scope of this chapter.

In this section we showed that the binary description leads us to a very similar market structure to the weighted description. This results suggest that the information regarding the community partition is mainly encoded in the binary signature of the fluctuations, i.e. just from the knowledge of the direction of movement, one can practically reproduce the "correct" structure. In the next section we will quantify the similarities and deviations of the different partitions for the different algorithms. Furthermore, we will explore the evolution in time of the variations between the two representations.

3.3.3 Variation of information analysis

Index	Method	**Q** weighted	**Q** binary	Frequent VI	Switching stocks (%)	Minimal VI	Switching stocks (%)
S&P	Potts	0.4035	0.4134	0.3543	8.81%	0.3198	6.74%
	Louvain	0.4070	0.4134	0.3477	8.09%	0.3192	7.64%
	Spectral	0.4006	0.3932	0.6955	61.57%	0.6955	61.57%
NIKKEI	Potts	0.4551	0.4525	0.3689	6.78%	0.2598	4.66%
	Louvain	0.4551	0.4525	0.3711	7.25%	0.2604	4.15%
	Spectral	0.4481	0.4424	0.4521	8.29%	0.4521	8.29%
FTSE	Potts	0.4641	0.4988	0.5031	28.42%	0.4026	26.14%
	Louvain	0.4635	0.4988	0.4995	21.79%	0.3981	17.95%
	Spectral	0.4597	0.4903	0.6919	69.23%	0.6919	69.23%

Table 3.2: The Variation of Information measured between the binary and weighted partitions, with the maximal modularity **Q**, for the period 2001-2011. "Frequent VI" is the variation of information between the most common partitions (that maximize the modularity), and "Minimal VI" is the variation of information between the most similar (that maximize the modularity). The "Switching stocks" is the percentage of stocks that moved to different communities.

Once we obtain the community structure using the different algorithms, our goal is to quantify the dissimilarities (or similarities) between the different partitions (binary and weighted). For this task we apply the Variation of Information (VI) measurement [28, 29]. The variation of information is an information-theoretic measure of the distance between two partitions. The different partitions $\vec{\boldsymbol{\sigma}}^1$ and $\vec{\boldsymbol{\sigma}}^2$ represent N-dimensional vectors where the i-th component σ_i denotes the set in which node i is placed by that particular partition.

The variation of information involves the mutual information $I(\vec{\boldsymbol{\sigma}}^1 : \vec{\boldsymbol{\sigma}}^2)$ which is defined as

$$I(\vec{\boldsymbol{\sigma}}^1 : \vec{\boldsymbol{\sigma}}^2) = \sum_{i=1}^{N}\sum_{j=1}^{N} p(\sigma_i^1, \sigma_j^2) \log \left(\frac{p(\sigma_i^1, \sigma_j^2)}{p(\sigma_i^1) p(\sigma_j^2)} \right), \tag{3.24}$$

where $p(\sigma_i^1, \sigma_i^2)$ is the joint probability distribution, and $p(\sigma_i^1)$ the marginal distribution of σ_i^1. The mutual information measures the overlap between the two partitions, however it is not a metric (does not obeys the triangle inequality) nor is it normalized. Thus, for this study we use the (normalized) variation of information which is defined as

$$VI(\vec{\boldsymbol{\sigma}}^1 : \vec{\boldsymbol{\sigma}}^2) = 1 - \frac{I(\vec{\boldsymbol{\sigma}}^1 : \vec{\boldsymbol{\sigma}}^2)}{H(\vec{\boldsymbol{\sigma}}^1 : \vec{\boldsymbol{\sigma}}^2)} \tag{3.25}$$

where $H(\vec{\boldsymbol{\sigma}}^1 : \vec{\boldsymbol{\sigma}}^2)$ is the joint entropy and is defined as

$$H(\vec{\boldsymbol{\sigma}}^1 : \vec{\boldsymbol{\sigma}}^2) = \sum_{i=1}^{N} \sum_{j=1}^{N} p(\sigma_i^1, \sigma_j^2) \log \left(p(\sigma_i^1, \sigma_j^2) \right) . \tag{3.26}$$

The variation of information ranges from 0 to 1, where 0 indicates two identical partitions, and 1 a complete dissimilarity between the partitions.

First, we measure the variation of information between two partition vectors, generated by the weighted and binary time series. Respectively, this approach enables us to quantify the difference in group structure (for 2001-2011), and compare the performances of the different algorithms.

In Table 3.2 we plot the measured VI between the different partitions, which resulted from the binary and weighted data. These measurements are for the community structures resulting from 10 years (2500 time steps) of data, for each of the three different indices. We run the algorithms 1000 times (for Louvain and Potts, while the spectral is deterministic) and extract the partitions that maximize the modularity. We measure the VI between two different partitions: the most frequent one, and the one that minimizes the VI (the most similar ones). One can consider this to be the best result (subject to the best partition). Again we should note that all the partitions maximize the modularity, and therefore are optimal. Since the VI is not linear (and not intuitive), we also included a simpler measurement of the percentage of stocks that are not occupying the same community in the different partitions.

We can observe that both the Potts and Louvain algorithms consistently perform better then the spectral method, i.e. they yield partitions with higher modularity. We should note that, one can only compare the value of the modularity for the same representation (binary or weighted), while there is no meaning in comparing modularity between the different representations. Generally, we can observe that the binary information and the weighted information result in very similar structures, as we showed in the previous section. The exception to this is the FTSE index, where the binary information consistently yields a greater number of communities. It is interesting to note that the binary information always results in either the same or a greater number of communities over the weighted time series.

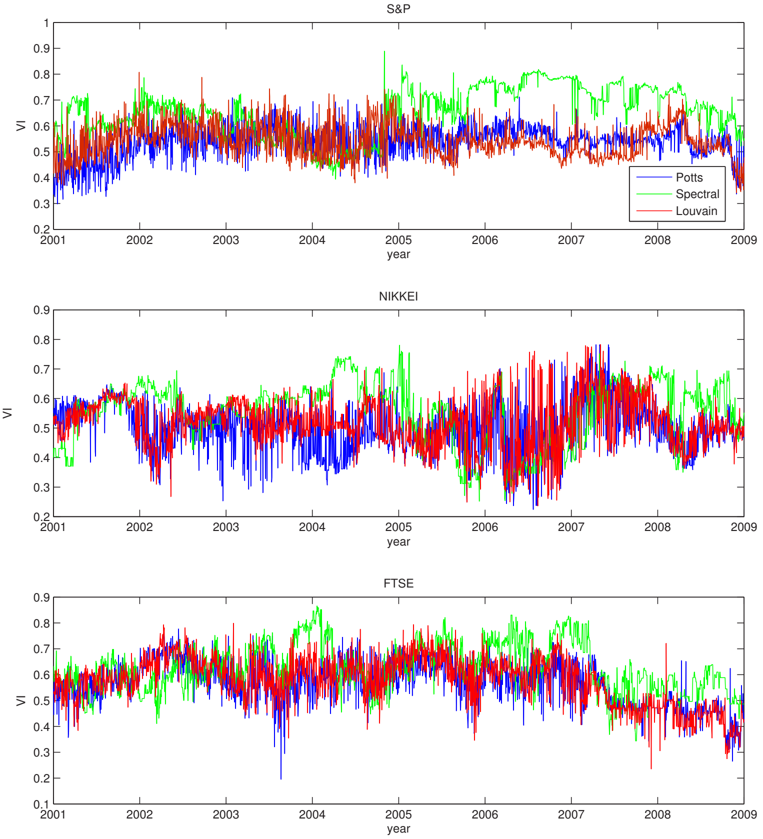

Figure 3.9: **Variation of information analysis for the different clustering algorithms.** variation of information between the binary and weighted partitions for a sliding window of 600 trading days (approximately 28 moths) starting at Q3 2001. The VI is measured between the frequent partitions for the different algorithms: Potts (blue), Louvain(red) and Spectral (green).

Next, we will explore the evolution in time of the VI between the different types of information. We considered a sub-period of 600 time steps (about two and a half years), and apply the same procedure as before. However, here we use a sliding window technique, where in each step we input a new day and ignore the previous information. This results in 1900 time steps (from the original 2500), where each point is the frequent VI calculated from the correlation matrix using the given 600 time step. In Fig. 3.9 we plot the different measurements for the different algorithms. The Potts and the Louvain methods present a more stable dynamics, while the Spectral method yield higher VI. Furthermore, there is no systematic effects of the financial crisis on the similarity between the binary and the weighted representations.

The analysis reveals that in financial markets both the binary and weighted information yield very similar structures, which also manifest themselves in similar spectral properties. The algorithms find complex mesoscopic structure of internally correlated clusters, which are residually anti-correlated with each other [7]. Moreover, the clusters are populated by stocks from various sectors. Remarkably, we show that the simple knowledge of the direction of increments of each stock can reproduce this complex structure very successfully.

3.4 Functional brain networks

In this section, we move from financial markets to brain systems. Remarkably, from a pure "network science perspective", these different systems share very similar features. Their dynamics is monitored in terms of multiple time series of activity of the constituent units, such as stocks or neurons respectively. In both systems, we have no complete information or exhaustive understanding about the underlying mechanisms at play, and we resort to information extraction from observations and empirical measurements. This illustrates why the use of correlation matrices is so popular both in financial systems and in neuroscience. Resolving the structure of brain networks from time series data (via the calculation an empirical correlation matrix), exactly as we did for financial markets in the previous section, is the object of the research field that goes under the name of 'functional brain networks.' In neuroscience, functional networks are defined by the correlated dynamical activity of neurons, as opposed to structural networks which are instead defined as the real physical connections between neurons [32, 33, 34].

In functional brain networks, the mesoscopic structure is the key intermediate level of organisation bridging the microscopic dynamics of individual neurons with the macroscopic dynamics of the brain as a whole. At this mesoscopic level, brain activity tends to be organized in a modular way [41], with functionally related units being positively correlated with each other, while at the same time being relatively less (or even negatively) correlated with dissimilar ones. Such emergent organisation is mainly detected through the aforementioned functional networks. In this setting, the network displays the correlated dynamical activity between pairs of units in the brain. Dependent on the type and resolution of the data, the units can vary from single neurons to brain regions. Over the last years, the statistical characterization of these functional networks developed into a very

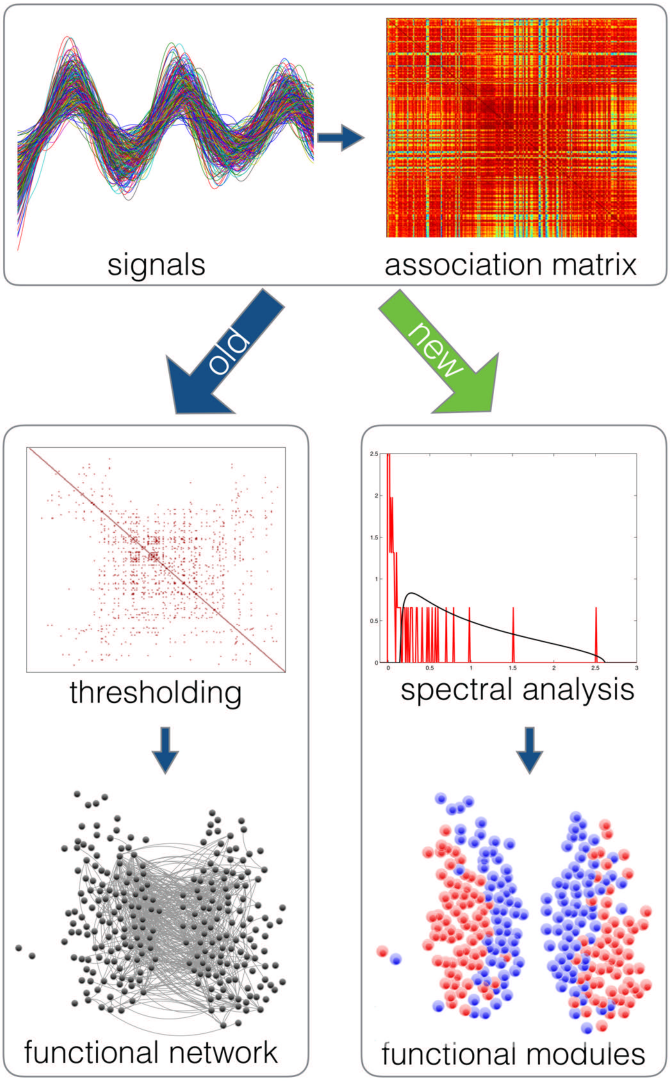

Figure 3.10: **Functional modules without functional network, a new method to identify emergent organisation.** A compression between the general steps of functional network construction and our new method. On the left scheme, we can see the introduction of a threshold procedure, once the majority of the data is neglected one can create a network representation. On the right scheme, we apply our spectrum filtering by comparing the empirical data to the random null model. By directly producing a partition of the original time series into communities, our new method bypasses the functional network projection, avoiding the use of a threshold procedure.

popular and powerful tool to study brain organisation [33, 35, 36].

Recent studies have used functional networks analysis to quantify global and local properties of brain networks [43, 44]. The use of functional networks yields very promising results. They are expected to encode the physiological mechanism for information processing and mental representations [37, 38, 39, 40]. Moreover, various case studies have shown that neurological and psychiatric disorders, such as Alzheimer's disease and schizophrenia, lead to variations in the network topology [45, 46]. Despite these encouraging results, it is well known that the current approaches suffer from specific methodological problems [42]. Firstly, the methods suffer from an inevitable information loss induced by the thresholding procedure. Another, less realized, limitation is the bias introduced by the use of network-based (as opposed to correlation-based) community detection methods [7]. Finally, the presence of (anti)correlated modules is obfuscated by the presence of a global mode of brain activity which imparts an overall positive correlation, a problem that becomes particularly evident when searching for communities in small brain regions, where the global signal is strong.

As we demonstrated in the previous section, our maximum-entropy approach can overcome precisely these methodological limitations. Furthermore, our method is not limited to a set of random walks with a common global factor. In financial markets, which we analysed in the previous section, the global factor translates to the similar reaction of stocks to news or events (the known market mode). However, once generalized, our new null model as described in section 3.2.2 can account for various systems with a global mode. In biological systems, the global mode can be a result of the units sharing a similar environment, or being affected by the same periodic signal (e.g. blood pressure or heartbeat rates). Remarkably, the removal of the global mode in our method entails by definition the removal of all common factors which affect the units of the system. This filtering capability demonstrates why our method has a great potential in the biological setting.

3.4.1 The suprachiasmatic nucleus

As we mentioned, we analyse the suprachiasmatic nucleus (SCN), located in the hypothalamus in the brain. The SCN is a complex structure comprising of a heterogeneous collection of genetically and electrically rhythmic neurons that synchronize their activity to the external world to form a central pacemaker. This complex network needs to produce stable output signals and at the same time remain flexible to changes in the environment. Thus, it is essential for the regulation of our daily and seasonal rhythms. The network is organized in a complex way for it to coordinate these daily rhythms at the tissue level [47], where the central pacemaker is a 24 hour rhythm that is synchronized to the external light-dark cycle. In response to a shift in the external cycle, neurons of the SCN resyn-

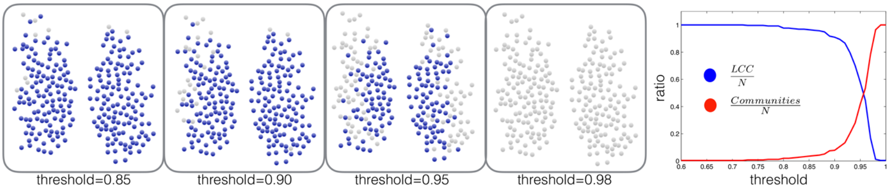

Figure 3.11: **The community structure of the SCN as resolved by a standard threshold approach.** On the left, we plot the community structure, resolved by the standard method, for different thresholds. In blue are the nodes that belong to the large cluster, while in gray are isolated nodes (communities composed out of one node). In the right panel, we plot the fraction of nodes in the largest connected component $\frac{LCC}{N}$ in blue, and the fraction of communities detected $\frac{Communities}{N}$ in red.

chronize with a different frequency. Recent studies have shown that the neuronal network organisation of the SCN changes in different seasons [48], however, the mechanisms behind these changes are still elusive. Currently, not much is understood about emergent organisation of the SCN, and if it is shaped by the external light-dark cycle. Here, we apply our community detection method to identify the functional community structure in different light-dark conditions.

This set-up investigates gene transcription activity, on a cellular scale, in SCN slices of male mice. This process is part of the molecular clock and is rhythmically expressed with a period of 24 hours. The target protein, which we monitor, is fused with the light emitting firefly luciferase. For each cell in the sample, the luminosity levels are recorded, generating the gene transcription "activity" time series of the units in the system. We should note that this "single cell resolution" of the data is unique for functional analysis. Currently, brain networks are most often derived from data acquisition techniques that do not have the possibility to perform recordings at the single cell level. Techniques such as Functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG) or Electroencephalography (EEG) use brain regions as nodes in the network and fiber bundles between these regions as edges. Our setup enables us to investigate a brain network at the micro-scale where nodes are single cells and edges are functional connections between the cells. The bioluminescence time series were obtained from individual cells (for exact method REF: [50]). In the next section, we apply the popular standard methods for constructing functional networks, highlighting their known limitations.

3.4.2 Standard approach to functional networks

Before we present the output of our method, we preliminary preform a standard analysis based in the mainstream method for detecting communities via functional networks. This is a useful reference as a comparison with our own method. In Figure 3.10 we present the standard ("old") approach versus our new approach. Once we measure the empirical correlation matrix, the conventional method requires a threshold value to exclude negative values and to neglect "weak" links simultaneously. Despite the significant loss of information, this thresholding procedure is an integral part of all current methods. Next, we adapt the thresholded matrix as an adjacency matrix, which corresponds to a network structure. The resulting network can either be a binary one, where all links are 1 or 0, or a weighted one, where the values from the correlation matrix remain as the weights of the links. Finally, we can apply various community detection methods [12, 13, 15, 14], to detect the community structure. In this analysis, we use the Louvain method [14].

In Figure 3.11 we present the community structure, resolved by the standard method, for different thresholds. In blue are the nodes that belong to the large cluster, while in gray are isolated nodes (communities composed out of one node). In the right panel, we plot the fraction of nodes in the largest connected component $\frac{LCC}{N}$ in blue, and the fraction of communities detected $\frac{Communities}{N}$ in red. It is evident that applying different thresholds essentially detaches isolated nodes from the large cluster, and there is no optimal value for the threshold. Therefore, the standard method can only observe a "radial gradient" of connectivity, and there is no large scale left-right symmetry, which is one of the signatures of functional as opposed to structural connectivity. This poor performance of the method is a known limitation when applied to very dense networks.

In the next section, we employ our new method described in section 3.2, which results in the new scheme portrayed in Figure 3.10. By directly producing a partition of the original time series into communities, our method bypasses the functional network projection, avoiding the use of a threshold procedure. Also, the method removes noise and filters out the common trend in the system.

3.4.3 SCN analysis

In Figure 3.12 we present the community structure detected by different recursive runs of the method. In the bottom panels are the partitions detected, where each community is marked with a different colour. In the top panels are the corresponding resolved filtered correlation matrices (for each run) displaying the resolved structure as a block matrix. In the left panel, we can observe that all the nodes in the system belong to one community when the global mode is not filtered out yet. Next, once we filtered out the global mode we found a consistent core-periphery structure. This result is robust to all the samples we have

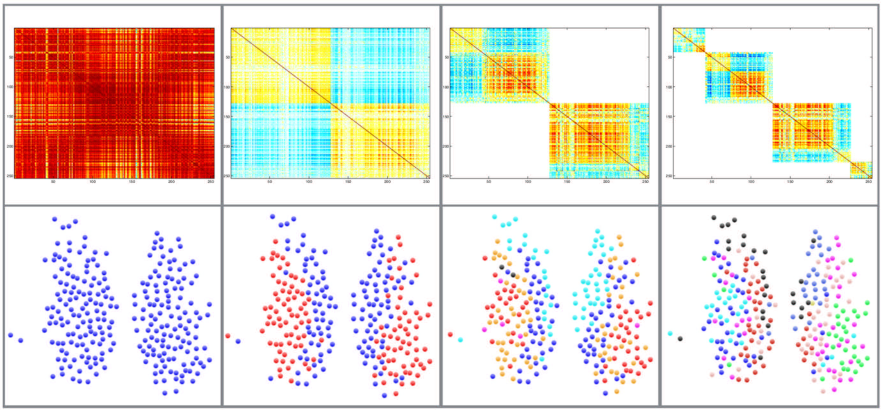

Figure 3.12: **The hierarchical community structure of the SCN as resolved by our method.** The community structure detected by different recursive runs of the method. In the bottom panels are the partitions detected, where each community is marked with a different colour. In the top panels are the corresponding resolved filtered correlation matrices (for each run) displaying the resolved structure as a block matrix.

analysed. Further iterations of our method within each of the identified modules, can detect hierarchy in the community structure. Note that for the second run the functional modules still retain the left-right symmetry, which implies a real functional structure. However, the hierarchical decomposition of the two clusters is difficult to interpret, because the resulting sub-clusters are small in size, which troubles reliable detection of significant differences between these sub-clusters. As a result, in this chapter, we will focus on the first partition that our method yields since it is providing the most impressive and consistent results.

The core-periphery structure confirms the ventral-dorsal distinction within the SCN. The ventral part of the SCN receives light input and adjusts quickly to changing light schemes, while the dorsal part of the SCN lags behind [51]. Our approach is able to identify the two communities in all the different experimental conditions, i.e. different the light-dark conditions simulating summer conditions (long days, short nights: L16D8) and winter day conditions (short days, long nights: L8D16). This results present a very robust functional structure, which is not dependent on external conditions. However, a more detailed analysis of the signals within the communities presented some variations in the distribution of the signals (within the communities) which depend on the light-dark cycles [50]. These findings motivate further research into this functional structure, and

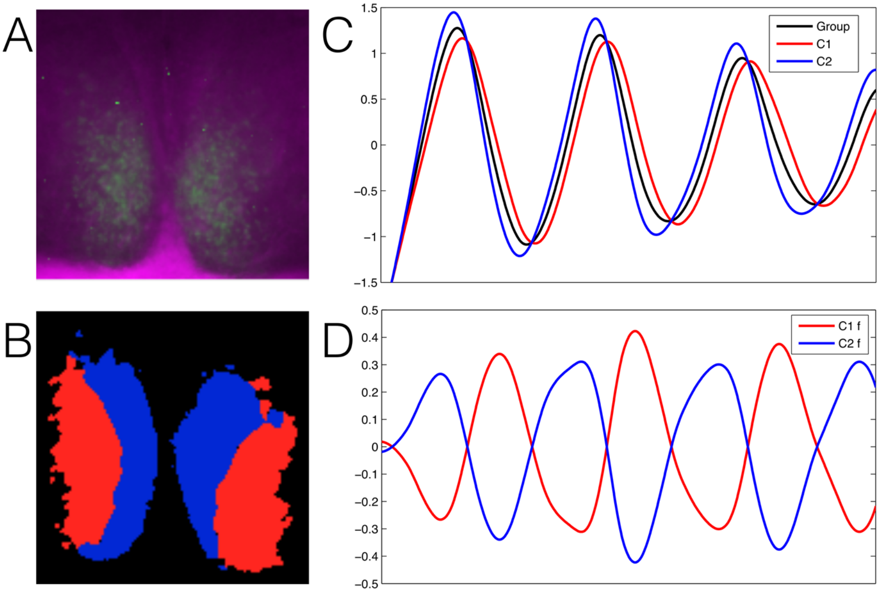

Figure 3.13: (A) The bioluminescence image of one SCN sample. (B) The plotted average partition over all the samples. (C) The plotted average signal of the whole system (in black) versus the mean signals of the two detected communities (in red and blue). (D) The plotted average residual signals of the two communities, once the global signal is subtracted.

possibly the next hierarchical partition.

We should note that regional analyses of the SCN has been performed by other groups. Evans and co-workers used a similar approach to identify single-cell-like regions of interest, but do not use clustering algorithms and choose the regions by hand [52]. Silver and co-workers also used regions of interest, called superpixels, but these were not necessarily identified as single cells. Based on these superpixels they use threshold methods to determine regional differences in the SCN [53, 54]. Abel and co-workers used a threshold method based on mutual information on single-cell-like regions of interest [55]. However, all studies encounter the known limitations as described in the previous section using the threshold method. They only find one large cluster and many isolated cell-like communities (in the shell or dorsal part).

In Figure 3.13 we present the overall findings our method provides. In panel A we can observe the experimental setup and the original sample with the illuminating neurons. In panel B we plot the average partition over all the samples (approximately 40); these results highlight the robustness of our findings and the clear core-periphery partition. In C we plot the average signal of the two communities (blue and red) with the mean signal of the whole system (black). In panel D we plot the deviations of each community signal from the common signal. We found a perfect anti-correlation with respect to the global mode. Our method is indeed guaranteed to resolve mutually anti-correlated modules, as the result of the maximization of the modularity. Note however, that the anti-correlation is defined in terms of the "residual" signals with respect to the average system-wide signal (see [7]).

3.5 Conclusions

Over the last few years community detection methods have revealed themselves as useful tools to study the structure of complex systems. In this chapter, we first have introduced a new approach aiming at analysing structural dependencies, which result from different descriptions (weighted and binary activity) of a complex system. Our approach enables us to quantify the level of "structural information" encoded within the binary projection of weighted time series, and measure variations and similarities between the different partitions.

Our findings suggest that the binary signatures of financial time series carry significant structural information. These results are far from trivial, as one might expect that the full knowledge of the amplitudes of price fluctuations is a key component in clustering the markets into correlated groups. However, here we explicitly showed that purely binary information can replicate the main features obtained from complete information. Thus, we conclude that the key features of the market structure are induced by the binary dynamics of the stocks. Even when the two representations differ by some extent, the binary description provides very sensible information (as exemplified by the Financials sector in the FTSE).

Typically, the binary signature of a time series is obtained as a projection from the full weighted information, and is therefore known only if the latter is also known. This means that there are not many practical situations in which the full weighted information is unknown, while the binary projection is known. However, what can occur quite often (and indeed typically occurs) is that the weighted amplitudes of time series increments, much more than their signs, are affected by noise or errors. For instance, in many financial institutions (e.g. hedge funds or investment banks) there are specific departments in charge of cleaning the data and verifying their reliability. Generally, errors in the data are detected in the form of increments with anomalously large or small absolute value. By contrast, the sign of the increment itself is very robust to errors. Therefore our results indicate that analyses based on binary projections are likely to be much more robust to noise than analyses based on the full original data. Since the extraction of the binary signatures is an extremely simple procedure, while the verification of the quality of weighted data can be very demanding and time consuming, our findings suggest that, at least for the purpose of identifying non-trivially correlated groups of stocks, the simpler procedure can be safely adopted.

Alongside the financial markets analysis, we have also applied our method to brain data. This step was possible due to the maximum-entropy generalization of the null model 3.2.2. By enforcing the total increments of each time step in the data, under the maximum-entropy formalism, we were able to expand our method to any system of signals with a common global factor, i.e. not be limited to a system with N random walks with an ad hoc market mode. Moreover, the new null model uses the optimal amount of information (eigenvalues) based on the particular data structure and not the data size (i.e. N and T) as before. In this general setting, the method resolves the partition of communities with maximum mutual anti-correlation, with respect to the global signal.

From a neuroscience perspective, our threshold-free method has a high potential to refine the search for functional modules in the brain. The fact that the approach does not require a "traditional" functional network representation is enabling the use of complete data sets and not just parts of it. Another aspect of our method that has significant implications for the research of brain networks is the multi-resolution quality. Since our method can recursively filter any global factors that are present in the system, we were able to detect single-cell-like regions of interest. This result pushes the resolution limit beyond the state-of-the-art methods, when identifying functional modules.

We applied our method to the SCN, revealing a core-periphery functional structure of two communities. Once the global signal is filtered from the system, the two clusters present a distinct anti-correlated dynamics. The global signal in this system corresponds to the common 24 hours cycle in all the cells, where the two communities describe a phase leading and a phase lagging community. This core-periphery structure confirms the ventral-dorsal distinction within the SCN. The ventral part of the SCN receives light input and adjusts quickly to changing light schemes, while the dorsal part of the SCN lags behind. Our community detection approach enhances the identification and the subsequent functional characterization of neuronal clusters in the SCN, possibly paving the way for more elaborate network analysis on the level of single cells in other brain regions.

Bibliography

[1] Sinha S, Chatterjee A, Chakraborti A, Chakrabarti BK. *Econophysics: An Introduction*, Wiley-VCH, Weinheim; 2010.

[2] Bouchaud JP, Potters M. *Theory of Financial Risk and Derivative Pricing*, Cambridge University Press, 2nd ed; 2003.

[3] Mantegna RN, Stanley HE. *Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press; 1999.

[4] Mantegna RN. *Hierarchical Structure in Financial Markets*, Eur Phys J B, 1999; 11, 193.

[5] Fortunato S. *Community detection in graphs*, Phys Rep. 2010; 486, 75.

[6] Newman MEJ. *Networks, an Introduction*, Oxford University Press; 2010.

[7] MacMahon M, Garlaschelli D. *Community detection for correlation matrices*, Phys Rev X. 2015; 5, 021006.

[8] La Spada G, Farmer J, Lillo F. *The non-random walk of stock prices: The long-term correlation between signs and sizes*, Eur Phys J B. 2008; 64, 607-614.

[9] Petersen AM, Wang F, Havlin S, Stanley HE. *Quantitative law describing market dynamics before and after interest rate change*, Phys Rev E. 2010; 81, 066121.

[10] Boguna M, Serrano MA. *Generalized percolation in random directed networks*, Phys Rev E. 2005; 72, 016106.

[11] Almog A, Garalaschelli D. *Binary versus non-binary information in real time series: empirical results and maximum-entropy matrix models*, New J Phys. 2014; 16, 093015.

[12] Reichardt J, Bornholdt S. *Detecting Fuzzy Community Structures in Complex Networks with a Potts Model*, Phys Rev Lett. 2004; 93, 218701.

[13] Reichardt J, Bornholdt S. Ntatistical, *Mechanics of Community Detection*, Phys Rev E. 2006; 74, 016110.

[14] Blondel JVD, Guillaume J, Lambiotte R, Lefebvre E. *Fast Unfolding of Communities in Large Networks*, Journal of Statistical Mechanics: Theory and Experiment. 2008; 10008.

[15] Newman MEJ. *Modularity and Community Structure in Networks*, Proc Natl Acad Sci USA. 2006; 103, 8577.

[16] Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE. *Random Matrix Approach to Cross Correlations in Financial Data*, Phys Rev E. 2002; 65, 066126.

[17] Wigner EP. *Characteristic Vectors of Bordered Matrices With Infinite Dimensions*, The Annals of Mathematics. 1955; 62, 548.

[18] Mehta ML. *Random Matrices*, Elsevier; 2004.

[19] Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Stanley HE. *Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series*, Phys Rev Lett. 1999; 83, 1471.

[20] Laloux L, Cizeau P, Bouchaud JP, Potters M. *Noise Dressing of Financial Correlation Matrices*, Phys Rev Lett. 1999; 83, 1467.

[21] Heimo T, Kumpula JM, Kaski K, Saramaki J. *Noise Dressing of Financial Correlation Matrices*, Phys Rev Lett. 1999; 83, 1467.

[22] Heimo T, Kumpula JM, Kaski K, Saramäki J. *Detecting Modules in Dense Weighted Networks with the Potts Method*, Journal of Statistical Mechanics: Theory and Experiment. 2008; 08, P08007.

[23] Fenn DJ, Porter MA, Mucha PJ, McDonald M, Williams S, Johnson NF, Jones NS. *Dynamical Clustering of Exchange Rates*, Quantitative Finance. 2012; 10, 1493.

[24] Isogai T. *Clustering of Japanese Stock Returns by Recursive Modularity Optimization for Efficient Portfolio Diversification*, J Complex Networks. 2014; 2, 557.

[25] Piccardi C, Calatroni L, Bertoni F. *Clustering Financial Time Series by Network Community Analysis*, Int J Mod Phys. 2011; 22, 35.

[26] Utsugi A, Ino K, Oshikawa M. *Random Matrix Theory Analysis of Cross Correlations in Financial Markets*, Phys Rev E. 2004; 70, 026110.

[27] Potters M, Bouchaud JP, Laloux L. *Financial Applications of Random Matrix Theory: Old Laces and New Pieces*, Acta Physica Polonica B. 2005; 36, 2767.

[28] Meila M. *Comparing Clusterings, An Information Based Distance*, Journal of Multivariate Analysis. 2007; 98, 873.

[29] Meila M. *Comparing Clusterings by the Variation of Information*, Learning Theory and Kernel Machines. 2003; 2777, 173.

[30] http://www.lorentz.leidenuniv.nl/ garlaschelli/.

[31] M. MacMahon, http://www.mathworks.com/matlabcentral/fileexchange/49011.

[32] Eguiluz VM, Chialvo DR, Cecchi GA, Baliki M, Apkarian AV. Scalefree *brain functional networks*, Phys Rev Lett. 2005; 94, 018102.

[33] Bullmore E, Sporns O. *Complex brain networks: graph theoretical analysis of structural and functional systems*, Nature Reviews Neuroscience. 2009; 10, 186–198.

[34] Park H, Friston k. *Structural and Functional Brain Networks: From Connections to Cognition*, Science. 2013; 342, 1238411-1238411.

[35] Singer, W. *Neuronal synchrony: a versatile code for the definition of relations?*, Neuron **24**, 49–65, (1999).

[36] Fries, P. *A mechanism for cognitive dynamics: neuronal communication through neuronal coherence*, Trends Cogn. Sci. **9**, 474–480 (2005).

[37] Bressler, S. L. *Large-scale cortical networks and cognition*, Brain Res. Brain Res. Rev. **20**, 288–304 (1995).

[38] Mesulam, M. M. *From sensation to cognition*, Brain **121**, 1013–1052 (1998).

[39] McIntosh, A. R. *Towards a network theory of cognition*, Neural Netw. **13**, 861–870 (2000).

[40] Buzsáki, G. *Rhythms of the Brain*, (Oxford Univ. Press, New York, 2006).

[41] Meunier, D., Lambiotte, R., Fornito, A., Ersche, K.D., Bullmore, E.T. *Hierarchical modularity in human brain functional networks*, Front. Neuroinformatics **3**, 37.

[42] RubinovaM. Sporns O. *Weight-conserving characterization of complex functional brain networks*, NeuroImag, **56**, Issue 4, 2068–2079, (2011).

[43] Stam, C.J., Reijneveld, J.C. *Graph theoretical analysis of complex networks in the brain*, Nonlinear Biomed. Phys. **1**, 3, (2011).

[44] Rubinov, M., Sporns, O. *Complex network measures of brain connectivity: Uses and interpretations* Neuroimage **52**, 1059–1069, (2010).

[45] Supekar, K., Menon, V., Rubin, D., Musen, M., Greicius, M.D. *Network analysis of intrinsic functional brain connectivity in Alzheimer's disease*, PLoS Comput. Biol. **4**, e1000100 (2008).

[46] Lynall, M.E., Bassett, D.S., Kerwin, R., McKenna, P.J., Kitzbichler, M., Muller, U., Bullmore,E. *Functional connectivity and brain networks in schizophrenia*, J. Neurosci. **30**, 9477–9487 (2010).

[47] Liu, A. C., Welsh, D. K., Ko, C. H., Tran, H. G., Zhang, E. E., Priest, A. A. *Intercellular coupling confers robustness against mutations in the SCN circadian clock network*, Cell, **129(3)**, 605-616 (2007).

[48] VanderLeest, H.T., Houben, T., Michel, S.,Deboer, T.,Albus, H., Vansteensel, J.M., Block, G.D, and Meijer, J.H. *Seasonal Encoding by the Circadian Pacemaker of the SCN*, Current Biology, **17**, Issue 5, 468-473 (2007).

[49] Rohling J.H.T., vanderLeest, H.T., Michel, S., Vansteensel, M.J., Meijer, J.H. *Phase resetting of the mammalian circadian clock relies on a rapid shift of a small population of pacemaker neurons*, PLoS ONE **6**, 2011, e25437.

[50] Buijink, R., Almog, A., Wit, C.B., Roethler, O., Garlaschelli, D., Meijer, J.H., Rohling, J.H.T, Michel, S. *Exposure to long photoperiods induces changes in coupling between single neurons of the mouse suprachiasmatic nucleus*, PLoS ONE, accepted pending (2016).

[51] Albus, H., Vansteensel, M.J., Michel, S., Block, G.D., Meijer, J.H. *A GABAergic mechanism is necessary for coupling dissociable ventral and dorsal regional oscillators within the circadian clock*, Curr Biol ,**10** :886-93 (2005).

[52] Evans J.A., Leise T.L., Castanon-Cervantes O., Davidson A.J. *Intrinsic regulation of spatiotemporal organization within the suprachiasmatic nucleus*, PLoS One. **6(1)**:522 e15869, (2011).

[53] Foley N.C., Tong T.Y., Foley D., Lesauter J., Welsh D.K., Silver R. *Characterization of orderly 536 spatiotemporal patterns of clock gene activation in mammalian suprachiasmatic nucleus*, Eur J Neurosci **33(10)**, 1851-65, (2011).

[54] Pauls S., Foley N.C., Foley D.K., LeSauter J., Hastings M.H., Maywood E.S. *Differential contributions of intra-cellular and inter-cellular mechanisms to the spatial and temporal architecture of the suprachiasmatic nucleus circadian circuitry in wild-type, cryptochrome-null and vasoactive intestinal peptide receptor 2-null mutant mice*, Eur J Neurosci, **542**, 2528-40 (2014).

[55] Abel J.H., Meeker K., Granados-Fuentes D., St John P.C., Wang T.J., Bales B.B. *Functional network inference of the suprachiasmatic nucleus*, Proc Natl Acad Sci U S A, **19** ,4512-7 (2016).

Concluding Remarks

In this thesis, we have reviewed various maximum-entropy models and their applications to different complex systems. We have demonstrated how flexible, and yet powerful, this approach is when applied to different systems. The foundations on which our maximum-entropy method is built come from deep within statistical physics. The models essentially describe canonical ensembles of different matrices representing time series, multiple time series, and various types of networks, which maximize Shannon's entropy given some constraints. As a result, for a given choice of constrains, the method characterizes each system with a unique probability distribution. Remarkably, dependent on the amount of partial information a model preserves, it can then be used in various ways, from statistical inference to empirical modelling and data filtration.

This takes us back to the main research question of the thesis, where we wanted to introduce a new class of statistical models which are as heterogeneous as real-world systems. We applied our models to complex financial systems covering different scales from the "microscopic" resolution of single stocks (time series) to the mesoscopic resolution of correlated communities of stocks, and finally to the macroscopic scale of entire economic networks. For each system, we identified the specific problems and challenges, which led us to a specific maximum-entropy approach in each setting.

In Chapter 1, we analysed financial time series and their corresponding binary projections. In this setting, the models enabled us to characterize and quantify the amount of information encoded in the binary signatures. We measured and compared the performance of the different models, indicating which property encodes more information, i.e. has the highest probability to replicate the original time series. When applied to cross-sections of financial time series, our method identified distinct regimes in the collective behaviour of groups of stocks, corresponding to different levels of coordination that only depend on the average return of the binary time series. Moreover, each regime is characterized by the most informative property. Finally, using the models we were able to mathematically characterize a universal relation between binary and non-binary properties for financial time series. These findings suggest that binary signatures of financial time series carry significant information, and present a new coordination measure in financial markets. The results in this chapter provided both the theoretical formalism and the motivation for the studies in chapter 3.

In chapter 2, we focused on the International Trade Network. We exploited the power of the maximum-entropy model in reproducing the complex large-scale topology of the empirical system and combined it with a more popular approach of macroeconomics models which focuses on individual link weights instead. Indeed, macroeconomic models have mainly concentrated on the expected volume of trade between two countries, given certain macroeconomic properties, and have disregarded the topology in which the system is embedded. As a result, the model's outcome is inconsistent with the observed, complex, topology of the ITN. Here, we assigned an accurate macroeconomic interpretation to the Lagrange multipliers which control for the number and weight of the links of each node in the maximum-entropy ensemble. In turn, this led to a new set of topologically invariant network models, which reformulate otherwise ill-defined economic models in such a way that the expected network topology does not depend on the arbitrary choice of the units of link weights. Lastly, this formula is general and can be applied to any economic network for which an empirically well-established econometric model for the link weights exists.

In chapter 3, maximum-entropy models were used as a random benchmark for filtering empirical correlation matrices. We have generalized a community detection method using the maximum-entropy formalism, resulting in an improved null model for the detection of non-random properties in the correlation matrix. Next, we applied the method to financial markets trying to uncover the community structure encoded within the binary signatures of financial time series. The analysis shows that in financial markets both the binary and weighted representations shared similar spectral properties, and formed very similar structures. Thus, indicate that the binary description of financial time series encodes significant structural information. The results motivate the use of binary projections, which are much more robust to noise than the original full data, for identifying non-trivially correlated groups of stocks. Finally, we tested the method in a biological setting applying it to the biological clock of mice, a complex network of oscillating neurons, uncovering a functional core-periphery structure which has been validated independently. We have shown that alternative state-of-the-art methods fail in detecting such core-periphery structure, as they only identify a radial gradient of connectivity decreasing from the centre towards the periphery, with no sharp boundary in between. Our approach enhances the identification of models and structure in functional brain networks, facilitating a more refine network analysis on the level of single neurons.

To conclude, the findings in this thesis emphasize the strength of the maximum-entropy approach when applied to complex systems. This research motivates further exploration of more sophisticated maximum-entropy models and the introduction of new tools to characterize heterogeneous and non-stationary systems. Such models have a great potential to make an impact in the fields of finance, economics, and neuroscience.

Appendix A

Appendix

A.1 Time series models

Models for single time series

We consider the case $N = 1$, i.e. when $\mathbf{X}$ is a $1 \times T$ matrix or equivalently a T-dimensional row vector. Let us denote the entries of $\mathbf{X}$ as $x(t)$.

Uniform random walk model

The trivial model is obtained when no constraints are enforced. In this case, there is no free parameter and the Hamiltonian has the form

$$H(\mathbf{X}) = 0 \tag{A.1}$$

As a result, the partition function is

$$Z = \sum_{\mathbf{X}} 1 = 2^T \tag{A.2}$$

which is nothing but the number of possible binary time series of length T. The probability of occurrence of a time series $\mathbf{X}$ is then

$$P\left(\mathbf{X}\right) = \frac{1}{Z} = 2^{-T} \tag{A.3}$$

and is completely uniform over the ensemble of all binary time series of length T. All the T elements of $\mathbf{X}$ are mutually independent and identically distributed with probability

$$P_t(x) \equiv \mathrm{Prob}\big(x(t) = x\big) = \begin{cases} 1/2 & x = -1 \\ 1/2 & x = +1 \end{cases} \tag{A.4}$$

This results in a completely uniform random walk with zero expected value for each increment:

$$\langle x(t)\rangle = 0 \tag{A.5}$$

While the (ensemble) variance of each increment equals

$$\mathrm{Var}[x(t)] \equiv \langle x^2(t)\rangle - \langle x(t)\rangle^2 = 1. \tag{A.6}$$

Biased random walk model

We now consider the total increment as the simplest non-trivial (one-dimensional) constraint:

$$C(\mathbf{X}) = T \cdot M_1(\mathbf{X}) = T \cdot \overline{x(t)} \tag{A.7}$$

If we denote the corresponding (scalar) Lagrange multiplier by θ, the Hamiltonian has the form

$$H(\mathbf{X}, \theta) = \theta \cdot T \cdot \overline{x(t)} = \theta \sum_{t=1}^{T} x(t). \tag{A.8}$$

The partition function is

$$\begin{aligned} Z(\theta) &= \sum_{\mathbf{X}} e^{-\theta \sum_{t=1}^{T} x(t)} = \sum_{\mathbf{X}} \prod_{t=1}^{T} e^{-\theta x(t)} \\ &= \prod_{t=1}^{T} \sum_{x=\pm 1} e^{-\theta x} = \prod_{t=1}^{T} \left[e^{-\theta} + e^{+\theta}\right] \\ &= \left[e^{-\theta} + e^{+\theta}\right]^T \end{aligned} \tag{A.9}$$

where, when interchanging the order of the sum and product, we have replaced the sum over all time series $\mathbf{X}$ with the sum over the two possible values $x = \pm 1$ of each individual entry.

The probability of the occurrence of a time series $\mathbf{X}$ is

$$\begin{aligned} P(\mathbf{X}|\theta) &= \frac{e^{-\theta \sum_{t=1}^{T} x(t)}}{\left[e^{-\theta} + e^{+\theta}\right]^T} = \prod_{t=1}^{T} \frac{e^{-\theta x(t)}}{e^{-\theta} + e^{+\theta}} \\ &= \prod_{t=1}^{T} P_t\big(x(t)|\theta\big) \end{aligned} \tag{A.10}$$

where we have introduced the probability $P_t(x|\theta)$ of a given increment $x = \pm 1$ at time t, which we identify as

$$P_t(x|\theta) = \frac{e^{-\theta x}}{e^{-\theta} + e^{+\theta}}. \tag{A.11}$$

The above expression shows that the stochastic process corresponding to this model is a biased random walk, as the two outcomes $x = \pm 1$ have a different probability, unless $\theta = 0$ (which leads us back to the uniform random walk model considered above).

The expected value of the t-th increment $x(t)$ (representing the bias of the random walk) is

$$\langle x(t) \rangle_\theta = \sum_{x=\pm 1} x P_t(x|\theta) = \frac{e^{-\theta} - e^{+\theta}}{e^{-\theta} + e^{+\theta}} = -\tanh\theta \tag{A.12}$$

and the variance is

$$\mathrm{Var}[x(t)] = \langle x^2(t) \rangle_\theta - {\langle x(t) \rangle_\theta}^2 = 1 - \tanh^2\theta. \tag{A.13}$$

The maximum likelihood condition (1.16), fixing the value θ^* of the parameter θ given a real time series $\mathbf{X}^*$, reads

$$T\langle\overline{x(t)}\rangle = \sum_{t=1}^{T} \langle x(t) \rangle = -T\tanh\theta = T \cdot \overline{x^*(t)} \tag{A.14}$$

Where $\overline{x^*(t)}$ is the measured average increment in the observed time series $\mathbf{X}^*$. This yields

$$-\tanh\theta^* = \overline{x^*(t)} \tag{A.15}$$

which gives a parameter value

$$\theta^* = -\mathrm{artanh}\left[\overline{x^*(t)}\right] = -\frac{1}{2}\ln\left[\frac{1+\overline{x^*(t)}}{1-\overline{x^*(t)}}\right] \tag{A.16}$$

One-dimensional Ising model

We now consider a model where, besides the constraint on the total increment specified in eq.(1.33), we enforce an additional constraint on the time-delayed (lagged) quantity $T \cdot B_1(\mathbf{X})$, where $B_1(\mathbf{X})$ is defined in eq.(1.27) with $\tau = 1$. This amounts to enforce the average one-step temporal autocorrelation of the time series. The resulting 2-dimensional constraint can be written as the column vector

$$\vec{C}(\mathbf{X}) = \begin{pmatrix} C_1(\mathbf{X}) \\ C_2(\mathbf{X}) \end{pmatrix} = T \cdot \begin{pmatrix} M_1(\mathbf{X}) \\ B_1(\mathbf{X}) \end{pmatrix}. \tag{A.17}$$

If we write the corresponding Lagrange multiplier as

$$\vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = -\begin{pmatrix} I \\ K \end{pmatrix}, \tag{A.18}$$

then the Hamiltonian reads

$$\begin{aligned} H(\mathbf{X}, I, K) &= \vec{\theta} \cdot \vec{C}(\mathbf{X}) = T\theta_1 M_1(\mathbf{X}) + T\theta_2 B_1(\mathbf{X}) \\ &= -I \sum_{t=1}^{T} x(t) - K \sum_{t=1}^{T} x(t)x(t+1), \end{aligned} \tag{A.19}$$

where we consider a periodicity condition as in eq.(1.28) with $\tau = 1$, i.e. $x(T+1) \equiv x(1)$. Note that, when $\mathbf{X}$ is a real binary time series of length T, this condition can be always enforced by adding one last (fictious) timestep $T+1$ and a corresponding increment $x(T+1)$ chosen equal to $x(1)$. For long time series, this has a negligible effect.

The above Hamiltonian coincides with that for the one-dimensional Ising model with periodic boundary conditions [55] (chapter 1). Each time step t is seen as a site in an ordered chain of length T, and each value $x(t) = \pm 1$ is seen as the value of a spin sitting at that site. The model is analytically solvable, which allows us to apply it to real time series in our formalism. For the readers familiar with time series analysis but not necessarily with the Ising model, we briefly recall the standard solution of the model, adapting it from ref. [55] (chapter 1).

Applying the periodicity condition of eq.(1.28) ensures that all sites (time steps) are statistically equivalent, i.e.:

$$\langle x(1) \rangle = \langle x(2) \rangle = \cdots = \langle x(T) \rangle \tag{A.20}$$

so that the system is translationally (here, temporally) invariant. The partition function is

$$Z(I, K) = \sum_{\mathbf{X}} \exp\left[I \sum_{t=1}^{T} x(t) + K \sum_{t=1}^{T} x(t)x(t+1) \right]$$

and can be rewritten as a product of terms involving only two successive time steps:

$$Z(I, K) = \sum_{\mathbf{X}} \prod_{t=1}^{T} V\big(x(t), x(t+1)\big), \tag{A.21}$$

where we have introduced the function $V(x, y)$ defined as

$$V(x, y) \equiv \exp\left(I \frac{x+y}{2} + Kxy \right). \tag{A.22}$$

We since both x and y can take only the values ±1, we can regard $V(x, y)$ as the element of a 2×2 matrix $\mathbf{V}$ called the *transfer matrix* [55] (chapter 1):

$$\mathbf{V} \equiv \begin{pmatrix} V(+1,+1) & V(+1,-1) \\ V(-1,+1) & V(-1,-1) \end{pmatrix} = \begin{pmatrix} e^{K+I} & e^{-K} \\ e^{-K} & e^{K-I} \end{pmatrix}. \tag{A.23}$$

This allows us to rewrite eq.(A.21) as

$$Z(I,K) = \mathrm{Tr}\big(\mathbf{V}^T\big). \tag{A.24}$$

Let $\vec{v}_1, \vec{v}_2$ denote the two eigenvectors of $\mathbf{V}$, and λ_1, λ_2 the corresponding eigenvalues, so that

$$\mathbf{V}\vec{v}_j = \lambda_j \vec{v}_j, \quad j = 1, 2. \tag{A.25}$$

The 2×2 matrix $\mathbf{Q} \equiv (\vec{v}_1, \vec{v}_2)$ (having column vectors $\vec{v}_1$ and $\vec{v}_2$) diagonalizes $\mathbf{V}$, i.e.

$$\mathbf{V} = \mathbf{Q} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \mathbf{Q}^{-1}, \tag{A.26}$$

where a direct calculation of the eigenvalues and eigenvectors yields

$$\lambda_1 = e^K \cosh I + \sqrt{e^{2K} \sinh^2 I + e^{-2K}} \tag{A.27}$$

$$\lambda_2 = e^K \cosh I - \sqrt{e^{2K} \sinh^2 I + e^{-2K}} \tag{A.28}$$

and

$$\mathbf{Q} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}, \tag{A.29}$$

with ϕ defined by

$$\cot 2\phi \equiv e^{2K} \sinh I. \tag{A.30}$$

It then follows that eq.(A.24) simply reduces to

$$Z(I,K) = \mathrm{Tr} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^T = \lambda_1^T + \lambda_2^T, \tag{A.31}$$

and the probability of occurrence of a time series $\mathbf{X}$ is

$$P(\mathbf{X}|I,K) = \frac{\prod_{t=1}^T V\big(x(t), x(t+1)\big)}{\lambda_1^T + \lambda_2^T}. \tag{A.32}$$

The above results allow us to analytically obtain expected values. That of $x(t)$ is

$$\langle x(t) \rangle = \sum_{\mathbf{X}} x(t) P(\mathbf{X}|I,K) = \frac{\mathrm{Tr}\big(\mathbf{S}\mathbf{V}^T\big)}{\lambda_1^T + \lambda_2^T}, \tag{A.33}$$

where we have introduced the diagonal matrix

$$\mathbf{S} \equiv \begin{pmatrix} S(+1,+1) & S(+1,-1) \\ S(-1,+1) & S(-1,-1) \end{pmatrix} = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix} \tag{A.34}$$

having elements

$$S(x,y) \equiv x\delta(x,y). \tag{A.35}$$

Similarly, for $0 < s-t < T$ the expected value of $x(t)x(s)$ is

$$\begin{aligned}\langle x(t)x(s)\rangle &= \sum_{\mathbf{X}} x(t)x(s)P(\mathbf{X}|I,K) \\ &= \frac{\mathrm{Tr}\big(\mathbf{S}\mathbf{V}^{s-t}\mathbf{S}\mathbf{V}^{T+t-s}\big)}{\lambda_1^T+\lambda_2^T}. \end{aligned} \tag{A.36}$$

In the limit $T \to \infty$ (corresponding to long time series in our case) with $s-t$ fixed, these expressions become

$$\langle x(t)\rangle = \cos 2\phi \tag{A.37}$$

$$\langle x(t)x(s)\rangle = \cos^2 2\phi + \sin^2 2\phi \left(\frac{\lambda_1}{\lambda_2}\right)^{s-t} \tag{A.38}$$

Now, we note that eqs.(A.33) and (A.36) manifestly show the translational (temporal) invariance of the model, as $\langle x(t)\rangle$ is independent of t and $\langle x(t)x(s)\rangle$ depends on t and s only through their difference $s-t$. This implies that, writing $\tau \equiv s-t$ and performing a temporal average,

$$\big\langle M_1\big\rangle = \cos 2\phi \tag{A.39}$$

$$\big\langle B_\tau\big\rangle = \cos^2 2\phi + \sin^2 2\phi \left(\frac{\lambda_1}{\lambda_2}\right)^{\tau}. \tag{A.40}$$

Using eq.(A.30) we can rewrite these expressions in terms of the model parameters, I and K, as

$$\big\langle M_1\big\rangle = \frac{e^{2K}\sinh I}{\sqrt{1+e^{4K}\sinh^2 I}} \tag{A.41}$$

$$\big\langle B_\tau\big\rangle = \frac{e^{4K}\sinh^2 I + (\lambda_1/\lambda_2)^\tau}{1+e^{4K}\sinh^2 I}. \tag{A.42}$$

The expected value of the autocorrelation defined in eq. (1.47) can be approximated as the ratio of two expected values as follows:

$$\big\langle A_\tau\big\rangle \equiv \left\langle \frac{B_\tau - M_1^2}{1-M_1^2}\right\rangle \approx \frac{\langle B_\tau\rangle - \langle M_1^2\rangle}{1-\langle M_1^2\rangle} = \left(\frac{\lambda_1}{\lambda_2}\right)^\tau. \tag{A.43}$$

Models for single cross-sections of multiple time series

For a single cross-section of a set of N multiple time series, $\mathbf{X}$ is a $N \times 1$ matrix or equivalently a N-dimensional column vector. We denote the entries of $\mathbf{X}$ as x_i.

Uniform random walk model

The uniform random walk is a simple modification of the same model that we considered for single time series, where $x(t)$ is replaced by x_i and T is replaced by N. This model is obtained when no constraints are enforced. The Hamiltonian is

$$H(\mathbf{X}) = 0 \tag{A.44}$$

and the partition function is simply the number of possible configurations for a single cross-section of N stocks:

$$Z = \sum_{\mathbf{X}} 1 = 2^N. \tag{A.45}$$

The probability of occurrence of a cross section $\mathbf{X}$ is

$$P\left(\mathbf{X}\right) = \frac{1}{Z} = 2^{-N} \tag{A.46}$$

and is completely uniform over the ensemble of all cross sections of N stocks. All the N elements of $\mathbf{X}$ are mutually independent and identically distributed with probability

$$P_i(x) \equiv \text{Prob}\big(x_i = x\big) = \begin{cases} 1/2 & x = -1 \\ 1/2 & x = +1 \end{cases} \tag{A.47}$$

This results in a completely uniform random walk with zero expected value

$$\langle x_i \rangle = 0 \tag{A.48}$$

and maximum variance

$$\text{Var}[x_i] \equiv \langle x_i^2 \rangle - \langle x_i \rangle^2 = 1. \tag{A.49}$$

Biased random walk model

Also this model is analogous to the corresponding model for single time series. We select the total daily increment of the cross section $\mathbf{X}$ as the constraint:

$$C(\mathbf{X}) = N \cdot M_1(\mathbf{X}) = N \cdot \{x_i\} \tag{A.50}$$

Let the corresponding Lagrange multiplier be denoted by θ. The Hamiltonian is

$$H\left(\mathbf{X}, \theta\right) = \theta \cdot N \cdot \{x_i\} = \theta \sum_{i=1}^{N} x_i \tag{A.51}$$

and the partition function is

$$\begin{aligned} Z(\theta) &= \sum_{\mathbf{X}} e^{-\theta \sum_{i=1}^{N} x_i} = \sum_{\mathbf{X}} \prod_{i=1}^{N} e^{-\theta x_i} \\ &= \prod_{i=1}^{N} \sum_{x=\pm 1} e^{-\theta x} = \prod_{i=1}^{N} \left[e^{-\theta} + e^{+\theta}\right] \\ &= \left[e^{-\theta} + e^{+\theta}\right]^N . \end{aligned} \tag{A.52}$$

The probability of the occurrence of a cross section $\mathbf{X}$ is

$$\begin{aligned} P(\mathbf{X}|\theta) &= \frac{e^{-\theta \sum_{i=1}^{N} x_i}}{\left[e^{-\theta} + e^{+\theta}\right]^N} = \prod_{i=1}^{N} \frac{e^{-\theta x_i}}{e^{-\theta} + e^{+\theta}} \\ &= \prod_{i=1}^{N} P_i\big(x_i|\theta\big) \end{aligned} \tag{A.53}$$

where we have introduced the probability $P_i(x|\theta)$ of a given increment $x = \pm 1$ for stock i, which we identify as

$$P_i(x|\theta) = \frac{e^{-\theta x}}{e^{-\theta} + e^{+\theta}}. \tag{A.54}$$

Just like the corresponding model for single time series, this model is a biased random walk, because the two outcomes $x = \pm 1$ have a different probability unless $\theta = 0$.

The expected value of the i-th increment x_i is

$$\langle x_i \rangle_\theta = \sum_{x=\pm 1} x P_i(x|\theta) = \frac{e^{-\theta} - e^{+\theta}}{e^{-\theta} + e^{+\theta}} = -\tanh\theta \tag{A.55}$$

and the variance is

$$\mathrm{Var}[x_i] = \langle x_i^2 \rangle_\theta - {\langle x_i \rangle_\theta}^2 = 1 - \tanh^2\theta. \tag{A.56}$$

The maximum likelihood condition (1.16), fixing the value θ^* of the parameter θ given a real cross section $\mathbf{X}^*$, reads

$$N\big\langle\{x_i\}\big\rangle = \sum_{i=1}^{N} \langle x_i \rangle = -N \tanh\theta = N \cdot \{x_i^*\} \tag{A.57}$$

where $\{x_i^*\}$ is the measured average increment of the observed cross section $\mathbf{X}^*$. This yields

$$-\tanh\theta^* = \{x_i^*\} \tag{A.58}$$

which gives a parameter value

$$\theta^* = -\mathrm{artanh}\left[\{x_i^*\}\right] = -\frac{1}{2}\ln\left[\frac{1 + \{x_i^*\}}{1 - \{x_i^*\}}\right]. \tag{A.59}$$

Mean-field Ising model

In this model, we enforce two constraints: the total increment and the total coupling between stocks. The resulting 2-dimensional constraint can be written as

$$\vec{C}(\mathbf{X}) = \begin{pmatrix} C_1(\mathbf{X}) \\ C_2(\mathbf{X}) \end{pmatrix} = \begin{pmatrix} N \cdot M_1(\mathbf{X}) \\ D(\mathbf{X}) \end{pmatrix}. \tag{A.60}$$

We can write the corresponding Lagrange multiplier as

$$\vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = - \begin{pmatrix} h \\ J \end{pmatrix} \tag{A.61}$$

and the Hamiltonian as

$$H(\mathbf{X}, h, J) = -h \sum_{i=1}^{N} x_i - J \sum_{i<j} x_i x_j. \tag{A.62}$$

Note that here we are not enforcing nearest-neighbor interactions as in the one-lagged model for single time series, but market-wide interactions among all stocks for the same time step (cross section). This is the result of the fact that, when considering cross sections, there is no natural notion of 'lattice sites' induced by e.g. a temporal ordering as in the one-lagged model. In other words, pairs of stocks in a cross section are neither 'close' nor 'distant'. We therefore assume a common interaction strength J among all stocks.

The above model, known as the mean-field Ising model, is analytically solvable. Here we adapt the derivation illustrated in ref. [55] (chapter 1). We first note that, since ${x_i}^2 = 1$ for all i, $H(\mathbf{X}, h, J)$ can be expressed as a function of $M_1(\mathbf{X})$ alone:

$$H(\mathbf{X}, h, J) = -hNM_1(\mathbf{X}) - \frac{J}{2}\big[N^2 M_1^2(\mathbf{X}) - N\big]. \tag{A.63}$$

This implies that the sum over configurations in the partition function can be replaced by a sum over the allowed values of $M_1(\mathbf{X})$, weighted by the number of configurations for each value. If we denote by r the number of increments that are negative ($x = -1$), and by $(N - r)$ the number of increments that are positive ($x = +1$), then we can write the Hamiltonian as a function of r alone through the expression

$$NM_1(\mathbf{X}) = N - 2r. \tag{A.64}$$

The partition function can therefore be calculated as

$$Z(h, J) \equiv \sum_{\mathbf{X}} e^{-H(\mathbf{X},h,J)} = \sum_{r=1}^{N} C_r \tag{A.65}$$

where

$$C_r \equiv \frac{N!}{r!(N-r)!} e^{h(N-2r)+\frac{J}{2}[(N-2r)^2-N]} \tag{A.66}$$

incorporates the binomial coefficient enumerating the configurations with given r. The expected increment is therefore

$$\langle M_1 \rangle = \left\langle 1 - \frac{2r}{N} \right\rangle = \frac{\sum_{r=1}^{N} \left(1-\frac{2r}{N}\right) C_r}{Z(h,J)} \quad \forall i. \tag{A.67}$$

When N is large, a traditional derivation [55] (chapter 1) shows that the sum at the numerator of eq.(A.67) is dominated by the single addendum corresponding to the maximum of C_r. The same applies to the partition function at the denominator. If r_0 denotes the value of r such that C_r is maximum, we then get

$$\langle M_1 \rangle \approx 1 - \frac{2r_0}{N}. \tag{A.68}$$

A further expansion [55] (chapter 1) finally shows that, given h and J, the expected value $\langle M_1 \rangle$ is the solution of the nonlinear equation

$$\langle M_1 \rangle = \tanh\big[(N-1)J\langle M_1 \rangle + h\big]. \tag{A.69}$$

From the above equation, one can infer the existence of a phase transition in the model, separating a regime where the expected 'magnetization' (here the average increment $\langle M_1 \rangle$) is zero from one where it is non-zero [55] (chapter 1). This transition is discussed in sec.1.5.3.

Before proceeding further, we note a peculiarity of the model, which has implications for the applicability of our maximum likelihood approach. An argument similar to that leading to eq.(A.68) implies that the second moment of $M_1(\mathbf{X})$ can be expressed as

$$\langle M_1^2 \rangle = \left\langle \left(1 - \frac{2r}{N}\right)^2 \right\rangle \approx \left(1 - \frac{2r_0}{N}\right)^2 \approx \langle M_1 \rangle^2. \tag{A.70}$$

This implies that

$$\mathrm{Var}[M_1] \equiv \langle M_1^2 \rangle - \langle M_1 \rangle^2 = 0, \tag{A.71}$$

or in other words that $M_1(\mathbf{X})$ is no longer a random variable. As a consequence, something unusual happens when we apply the maximum likelihood principle. From eq.(A.63), and recalling the general result embodied by eq.(1.20) in sec.1.3.2, it is clear that the parameter values h^* and J^* maximizing the likelihood can be found as the solution to the two coupled equations

$$\langle M_1 \rangle = M_1(\mathbf{X}^*) \tag{A.72}$$

$$\langle M_1^2 \rangle = M_1^2(\mathbf{X}^*) \tag{A.73}$$

However, eq. (A.70) implies that eq.(A.73) can be rewritten as

$$\langle M_1\rangle^2 = M_1^2(\mathbf{X}^*) \tag{A.74}$$

which coincides with eq.(A.72). So eqs. (A.72) and (A.73) are equivalent, and they cannot be used to uniquely determine the two unknown parameters h^* and J^*. This is the result of the fact that, when fitted to the data, the model is actually over-constrained: there are two parameters to fit the only constraint (M_1) on which the Hamiltonian depends. This aspect of the model is not manifest when M_1 is regarded as a function of h and J, as usually done when simulating spin systems.

The above consideration implies that we should drop one of the two parameters and consider the two cases $J = 0$ and $h = 0$ separately. The former case coincides with the biased random walk model that we already discussed, and we will not discuss it any further. The latter case will instead represent our genuine specification of the 'mean-field' model. Setting $h = 0$ implies

$$H(\mathbf{X}, 0, J) = -\frac{J}{2}\left[N^2 M_1^2(\mathbf{X}) - N\right] \tag{A.75}$$

and

$$\langle M_1\rangle = \tanh\left[(N-1)J\langle M_1\rangle\right]. \tag{A.76}$$

Applying the maximum likelihood principle to eq.(A.75) tells us to select J^* as the solution of eq.(A.73). However, we have seen that this condition leads to eq.(A.74), which is actually equivalent to eq.(A.72). Therefore, the value of J^* can be found by replacing $\langle M_1\rangle$ with the observed value $M_1(\mathbf{X}^*) = \{x_i^*\}$ in eq. (A.76), which leads to

$$\{x_i^*\} = \tanh\left[(N-1)J^*\{x_i^*\}\right]. \tag{A.77}$$

Note that in the traditional situation one is interested in finding the (expected) magnetization given a value of J, which implies that the transcendental eq. (A.76) should be solved numerically. Here, we are instead facing the inverse situation where we look for the value of J^* given the (observed) value of the magnetization. In this quite unusual case, it turns out that eq. (A.77) can be inverted to give the following analytical solution:

$$J^* = \frac{\operatorname{artanh}\{x_i^*\}}{\{x_i^*\}(N-1)} = \frac{1}{2\{x_i^*\}(N-1)}\ln\left[\frac{1+\{x_i^*\}}{1-\{x_i^*\}}\right]. \tag{A.78}$$

Once this value is calculated, it can be inserted into the probability

$$P(\mathbf{X}^*|0, J) = \frac{e^{-H(\mathbf{X}^*,0,J)}}{Z(0,J)} = \frac{e^{JN(N\{x_i^*\}^2-1)/2}}{\sum_{r=1}^{N}\frac{N!}{r!(N-r)!}e^{J[(N-2r)^2-N]/2}} \tag{A.79}$$

(where we have set $h = 0$) to obtain the maximized likelihood of generating the observed cross section $\mathbf{X}^*$ under the mean-field model.

In this appendix we give a summarized description of the binary and weighted network quantities which are studied in this paper. Specifically, we first show how the properties are measured over a real network, and then how the expected values under the ECM and the TS model are constructed.

A.2 Network models

Observed Properties

Let us note a weighted undirected network as a square matrix $\mathbf{W}$, where the specific entry w_{ij} represents the edge weight between country i and country j. The binary representation of the network is noted by a binary matrix $\mathbf{A}$, where the entries are $a_{ij} = \Theta[w_{ij}]$, $\forall\, i, j$.

We compute the Average Nearest Neighbor Degree as:

$$k_i^{nn}(\mathbf{W}) = \sum_{j \neq i} \frac{a_{ij} k_j}{k_i} = \frac{\sum_{j \neq i} \sum_{k \neq j} a_{ij} a_{jk}}{\sum_{j \neq i} a_{ij}}. \tag{A.80}$$

Its calculated as the arithmetic mean of the degrees of the neighbors of a specific node, which is a measure of correlation between the degrees of adjacent nodes.

The Binary Clustering Coefficient has the following expression:

$$c_i(\mathbf{W}) = \frac{\sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{jk} a_{ki}}{\sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{ki}}. \tag{A.81}$$

It is a measure of the tendency to which nodes in a graph form cluster together. More specifically, it counts how many closed triangles are attached to each node with respect to all the possible triangles.

The corresponding weighted properties are the Average Nearest Neighbor Strength and the weighted Clustering Coefficient. The Average Nearest Neighbor Strength, defined as:

$$s_i^{nn}(\mathbf{W}) = \sum_{j \neq i} \frac{a_{ij} s_j}{k_i} = \frac{\sum_{j \neq i} \sum_{k \neq j} a_{ij} w_{jk}}{\sum_{j \neq i} a_{ij}} \tag{A.82}$$

where $s_i = \sum_j w_{ij}$ is the strength (total flow) of a country. The s_i^{nn} measure the average strength of the neighbors for a specific node i. Like its binary counterpart, it gives the magnitude of activity of a specific node neighbors (weighted activity).

The weighted Clustering Coefficient [54] (chapter 2) is defined as:

$$c_i^W(\mathbf{W}) = \frac{\sum_{j\neq i}\sum_{k\neq i,j}(w_{ij}w_{jk}w_{ki})^{\frac{1}{3}}}{\sum_{j\neq i}\sum_{k\neq i,j}a_{ij}a_{ki}}. \tag{A.83}$$

The $c_i^W(\mathbf{W})$ is a measure of the weight density in the neighborhood of a node. It classify the tendency of a specific node to cluster in a triangle taking into account also the edge-values.

Now, the measured properties of the real network need to be compared with the reproduced properties of the different models. These reproduced properties are the expected values of the maximum entropy ensemble that each model id generating, and can be calculated analytically. The expected values can be obtained by simply replacing a_ij with the probability p_{ij} for the different models (p_{ij} is different to each model). This next step is what we will discuss in the next sections.

Expected values in the BCM and ECM

Since the BCM model is only dealing with the binary representation, we will have expected values just for the two binary higher-order properties. While the ECM gives expectations for the weighted counterparts of the binary properties.

For the binary higher-order properties, we replace a_ij with p_{ij} which is the probability of creating a link, and also the expected value of the edge $p_{ij} = \langle a_{ij}\rangle$. This simple procedure yields the analytic formula of the expected value for the properties. We compute the expected Average Nearest Neighbor Degree as:

$$\langle k_i^{nn}\rangle = \frac{\sum_{j\neq i}\sum_{k\neq j}p_{ij}p_{jk}}{\sum_{j\neq i}p_{ij}} \tag{A.84}$$

and the expected Binary Clustering Coefficient as:

$$\langle c_i\rangle = \frac{\sum_{j\neq i}\sum_{k\neq i,j}p_{ij}p_{jk}p_{ki}}{\sum_{j\neq i}\sum_{k\neq i,j}p_{ij}p_{ki}} \tag{A.85}$$

where for the BCM model we input $p_{ij} = \frac{z_i z_j}{1+z_i z_j}$, and for the ECM the more complex term $p_{ij} = \frac{x_i x_j y_i y_j}{1-y_i y_j+x_i x_j y_i y_j}$

In the weighted case (weighted higher-order properties), we are left only with the ECM. The expected Average Nearest Neighbor Strength is calculated as:

$$\langle s_i^{nn}\rangle = \frac{\sum_{j\neq i}\sum_{k\neq j}p_{ij}\langle w_{jk}\rangle}{\sum_{j\neq i}p_{ij}} \tag{A.86}$$

where $\langle w_{jk}\rangle = \frac{x_i x_j y_i y_j}{(1-y_i y_j+x_i x_j y_i y_j)(1-y_i y_j)}$ and we input the p_{ij} of the ECM model as before.

In the expected value of the c^W we should be more careful, since it is necessary to calculate the expected product of (powers of) distinct matrix entries

$$\langle c_i^W\rangle = \frac{\sum_{j\neq i}\sum_{k\neq i,j}\langle (w_{ij}w_{jk}w_{ki})^{\frac{1}{3}}\rangle}{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}p_{ki}}. \tag{A.87}$$

We know that

$$\langle \sum_{i\neq j\neq k} w_{ij}^{\alpha}\cdot w_{jk}^{\beta}\cdot ...\rangle = \sum_{i\neq j\neq k}\langle w_{ij}^{\alpha}\rangle\cdot\langle w_{jk}^{\beta}\rangle\cdot\langle ...\rangle \tag{A.88}$$

with the generic term for the ECM case

$$\langle w_{ij}^{\gamma}\rangle = \sum_{0}^{\infty} w^{\gamma} q_{ij}(w|\vec{x},\vec{y}) = \frac{x_i x_j(1-y_i y_j)Li_{\gamma}(y_i y_j)}{1-y_i y_j + x_i x_j y_i y_j} \tag{A.89}$$

where $Li_n(R) = \sum_{l=1}^{\infty}\frac{R^l}{l^n}$ is the nth polylogarithm of R. For a more comprehensive description please refer to [32] (chapter 2).

Expected values in the TS model

Here again we use the known expressions for the properties and replacing the terms p_{ij}^{ts} and w_{ij}^{ts} with the expected values $\langle a_{ij}\rangle$ and $\langle w_{ij}\rangle$ correspondingly. However, here the expected values are a function of the GDP of the countries, or more specifically the re-scaled GDP g_i. The expressions for the higher-order binary properties are as before :

$$\langle k_i^{nn}\rangle = \frac{\sum_{j\neq i}\sum_{k\neq j} p_{ij}^{ts}p_{jk}^{ts}}{\sum_{j\neq i} p_{ij}^{ts}} \tag{A.90}$$

and

$$\langle c_i\rangle = \frac{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}^{ts}p_{jk}^{ts}p_{ki}^{ts}}{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}^{ts}p_{ki}^{ts}} \tag{A.91}$$

where $p_{ij}^{ts} = \frac{a g_i g_j}{1+a g_i g_j}$.

In the weighted case,the expected Average Nearest Neighbor Strength is calculated as:

$$\langle s_i^{nn}\rangle = \frac{\sum_{j\neq i}\sum_{k\neq j} p_{ij}^{ts}\langle w_{jk}^{ts}\rangle}{\sum_{j\neq i} p_{ij}^{ts}} \tag{A.92}$$

where $\langle w_{ij}\rangle^{ts} = \frac{ag_ig_j}{1+ag_ig_j}\cdot\frac{(1+bg_i^c)(1+bg_j^c)}{(1+bg_i^c+bg_j^c)}$ and we input the p_{ij}^{ts} of the Two-Step model as before.

For convenience reasons we will write the expression for the weighted Clustering Coefficient c^W first as a function of the fitness parameters z_i and y_i, and later replaced them with the corresponding GDP terms. The expected value of the c^W in the Two-Step case is :

$$\langle c_i^W\rangle = \frac{\sum_{j\neq i}\sum_{k\neq i,j}\langle (w_{ij}w_{jk}w_{ki})^{\frac{1}{3}}\rangle^{ts}}{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}^{ts}p_{ki}^{ts}}. \tag{A.93}$$

As before we observe that

$$\langle \sum_{i\neq j\neq k} w_{ij}^{\alpha}\cdot w_{jk}^{\beta}\cdot ...\rangle = \sum_{i\neq j\neq k}\langle w_{ij}^{\alpha}\rangle\cdot\langle w_{jk}^{\beta}\rangle\cdot\langle ...\rangle \tag{A.94}$$

with the generic term for the Two-Step model

$$\langle w_{ij}^{\gamma}\rangle = \sum_{0}^{\infty} w^{\gamma}q_{ij}(w|\vec{z},\vec{y}) = \frac{z_iz_j(1-y_iy_j)Li_{\gamma}(y_iy_j)}{(1+z_iz_j)y_iy_j} \tag{A.95}$$

where $Li_n(R) = \sum_{l=1}^{\infty}\frac{R^l}{l^n}$ is the nth polylogarithm of R.

Once we input the expressions of z_i and y_i

$$\begin{aligned} z_i &= \sqrt{a}\cdot g_i, \\ y_i &= \frac{b\cdot g_i^c}{1+b\cdot g_i^c} \end{aligned} \tag{A.96}$$

equation (A.95) yields

$$\langle w_{ij}^{\gamma}\rangle = \frac{ag_ig_j}{1+ag_ig_j}\cdot\frac{1+cg_i^c+bg_j^c}{b^2g_i^cg_j^c}Li_{\gamma}\left(\frac{b^2g_i^cg_j^c}{(1+bg_i^c)(1+bg_j^c)}\right) \tag{A.97}$$

Samenvatting

Complexe systemen, variërend van financiële markten tot de hersenen, hebben heterogene structuren en vertonen niet-stationaire dynamica. Deze karakteristieken manifesteren zich in de diversiteit van de elementen in het systeem en in gedrag dat verandert in de tijd. Indien men er in slaagt om het heterogene karakter in adequate modellen te vangen en de verschijnselen daaruit te begrijpen, kan dat niet alleen in de natuurwetenschappen belangrijk zijn, maar ook ten behoeve van maatschappelijke uitdagingen zoals het voorkomen van de volgende financiële crisis of het ontwikkelen van geavanceerde afbeeldingstechnieken voor de hersenen. In dit proefschrift gebruiken we het concept van *maximale entropie* om een nieuwe klasse van statistische modellen in te voeren die de waargenomen heterogeniteit in structuur en/of tijd in zich bergen. Deze modellen worden toegepast op een aantal complexe systemen in de werkelijke wereld en zo worden verschillende problemen aangepakt.

In het eerste hoofdstuk onderzoeken we toepassingen van de maximale-entropie aanpak op de analyse van *financiële tijdreeksen.* We introduceren een nieuwe klasse van maximale-entropie ensembles voor tijdreeksen, die ons in staat stelt om de informatie te karakteriseren en te kwantificeren die bevat is in de *binaire representatie* (die slechts aangeeft of iets omhoog of omlaag gaat) van financiële tijdreeksen. Vervolgens gebruiken we een van de modellen om de waargenomen niet-lineaire empirische relatie tussen binaire en niet-binaire eigenschappen van financiële tijdreeksen wiskundig te verklaren. Deze analyse suggereert dat binaire representaties van financiële tijdreeksen significante informatie bevatten omtrent de volledige *gewogen* tijdreeksen waarvan ze zijn afgeleid. De analyse maakt ook de identificatie mogelijk van te meten eigenschappen die het meest informatief zijn over de originele tijdreeks.

In hoofdstuk 2 richten we ons op de modellering van *economische netwerken*, met name het *International Trade Network* (ITN). We gebruiken opnieuw een maximale-entropie aanpak, die zeer succesvol is in het reproduceren van de empirische complexe topologie van het daadwerkelijke netwerk. De methode wordt vervolgens geïntegreerd in een bekende macro-economische aanpak. Macro-economische modellen hebben als grootste beperking dat ze een niet-realistische uniforme topologie genereren. Wanneer de twee methoden worden gecombineerd ontstaat een nieuwe verzameling topologisch-invariante netwerkmodellen voor het ITN. Deze zijn een significante verbetering voor wat betreft de gehele structuur van het systeem. Het is opmerkelijk dat het mechanisme waarmee het model een complexe topologie oplegt, het principe van topologische invariantie, gegeneraliseerd kan worden naar elk economisch netwerkmodel.

In het laatste hoofdstuk wordt de maximale-entropie aanpak gebruikt om een nieuwe methode te ontwikkelen voor het zeven van empirische *correlatiematrices*. De maximale-entropie modellen geven in deze toepassing willekeurig verdeelde ijkmodellen, zgn. *null models*, die een deel van de afhankelijkheden in het systeem behouden. De eigenschappen die niet worden gereproduceerd door de null models worden als niet-willekeurig beschouwd en worden gebruikt om de empirische correlatiematrices te zeven. Onze methode kan functionele modules in een systeem detecteren, die intern gecorreleerd zijn, maar anti-gecorreleerd met elkaar. We passen onze methode met succes toe op financiële markten en op gegevens omtrent de functionaliteit van hersenen. Dit resulteert in informatie over de interne hiërarchische organisatie van deze systemen met een veel grotere resolutie dan tot nog toe mogelijk was.

Summary

Complex systems, from financial markets to the brain, exhibit heterogeneous structures and non-stationary dynamics. These characteristics manifest themselves in the diversity of the elements in a system, and in the changing behaviour over time. Capturing and understanding this heterogeneity via appropriate models, can have important implications not only for science, but also for societal challenges like predicting the next financial crisis or developing advanced brain imaging techniques. In this thesis, we use the maximum-entropy approach to introduce a new class of statistical models, which captures part of the observed structural and/or temporal heterogeneity in the system. The models are applied to various real-world complex systems, and are used to address different problems.

First, we explore some applications of our maximum-entropy approach to the analysis of financial time series. We introduce a new class of maximum-entropy time series ensembles, which enable us to characterize and quantify information encoded within the binary signatures, i.e., the signs, of the increments of financial time series. Next, using one of the models, we mathematically characterize the observed non-linear empirical relations between binary and non-binary properties in financial time series. The analysis suggests that binary signatures of financial time series encode significant information with respect to their complete weighted counterparts. It also allows to identify which measured properties are most informative about the original time series.

Second, we focus on modelling economic networks, in particular the International Trade Network. We adopt again a maximum-entropy approach, which in this case turns out to be powerful in reproducing the empirical complex topology of the real-world system, and integrate it with the known macroeconomic approach. Indeed, the main limitation of macroeconomic models is the fact that they generate an unrealistically uniform topology. Combining the two approaches together leads to a new set of topologically invariant network models for the ITN, which represent a significant improvement when modelling the entire structure of the system. Remarkably, the mechanism through which the model enforces a complex topology, i.e., the principal of topological invariance, can be generalized to any economic network model.

Finally, we use the maximum-entropy approach to develop a new technique to filter empirical correlation matrices. The maximum-entropy models represent random benchmarks in this case, which still preserve part of the dependencies in the system. The properties which are not reproduced by the null model are considered non-random and are used to filter the empirical correlation matrix. Our method is able to detect functional modules in the system, which are internally correlated and mutually anti-correlated. We successfully apply our method to financial markets and functional brain data to detect the internal hierarchical organization of these systems with a level of resolution previously unavailable.

List of publications

[1] *Binary versus non-binary information in real time series: empirical results and maximum-entropy matrix models*, A. Almog and D. Garlaschelli
New Journal of Physics **16** 093015 (2014).

[2] *A GDP-driven model for the binary and weighted structure of the International Trade Network*, A. Almog, T. Squartini and D. Garlaschelli
New Journal of Physics **17** 013009 (2015).

[3] *Mesoscopic Community Structure of Financial Markets Revealed by Price and Sign Fluctuations*, A. Almog, F. Besamusca, M. MacMahon and D. Garlaschelli
PLoS ONE **10** 7. e0133679 (2015).

[4] *The double role of GDP in shaping the structure of the International Trade Network*, A. Almog, T. Squartini and D. Garlaschelli
Int. J. Computational Economics and Econometrics, in press, arXiv:1512.02454 (2016) .

[5] *Enhanced Gravity Model of trade: reconciling macroeconomic and network models*, A. Almog, R. Bird and D. Garlaschelli
arXiv:1506.00348, *to appear.*

[6] *Functional modules without functional networks: community detection for complex brain networks*, A. Almog, O. Roethler, R. Buijink, J.H. Meijer, J.H.T. Rohling and D. Garlaschelli,
to appear.

[7] *Long photoperiod increases period variability and weakens intercellular coupling within the mammalian circadian clock*, R. Buijink, A. Almog, C.B. Wit, O. Roethler, D. Garlaschelli, J.H. Meijer, J.H.T. Rohling and S. Michel
PLoS ONE, in press.

Curriculum vitæ

I was born in Eilat, Israel, on the 22nd of April 1982. I have graduated from Begin high school in Rosh Hain, with specialization in physics and mathematics. After finishing high school, I served as a Marine officer in the Israeli Navy for almost six years and completed my service with a captain rank. In October 2006, after a fair share of travelling, I started my bachelor in physics at Ben-Gurion University (Israel). I graduated in January 2010, and later that year started my master in physics at TU Delft (Netherlands). I completed my masters degree in November 2012. In January 2013, I started my Ph.D. studies at the Lorentz Institute for Theoretical Physics in Leiden, under the supervision of Dr. Diego Garlaschelli. After graduation, I will continue as a post-doctoral researcher at the Lorentz institute.

Acknowledgements

First of all, I would like to thank my supervisor Diego Garlaschelli, with whom I have the pleasure of working over the last five years. Since my master's thesis, his expertise, high standards, and vast knowledge, shaped my academic experience significantly. In his own calm and peaceful way, he always has been there for any question, discussion, or advice I needed. I truly appreciate the scientific freedom and the optimal research conditions that were given to me throughout this long process.

I express my warm thanks to the partners and researchers at Duyfken Trading Knowledge for providing me not only with financial data but also with important insights and a professional and business-oriented point of view. I would like to thank Jan van Ruitenbeek, the chair of the Dutch Econophysics Foundation, for accommodating and supporting this project and this line of research. Special thanks goes to Iman van Lelyveld, for arranging and supervising my internship at the Dutch National Bank, and for his scientific guidance and many insightful discussions.

I am grateful to my collaborators who have contributed towards shaping this thesis. I would like to thank Jos Rohling, Renate Buijink and Joke Meijer, for their support and for providing me the great opportunity to work on brain networks. I extend my gratitude to Tiziano Squartini, for his valuable guidance, patience, and friendship. My warmest thanks go to Mel MacMahon, my collaborator, good friend, and probably the main reason for this doctorate.

I thank my colleagues and friends, Alex, Valerio, Elena, Ferry, Vasyl and Rhys for the stimulating discussions, for the countless friendly lunches, and for all the fun we have had in the last four years.

Finally, I would like to thank my wife and my parents for the support over all the years of my studies.