Statistical physics and information theory for systems with local constraints

Zhang, Q.

Statistical physics and information theory for systems with local constraints

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 1 december 2021
klokke 10.00 uur

door

Qi Zhang

geboren te Mianyang (Sichuan), China
in 1989

Promotors:	Dr. D. Garlaschelli Prof. dr. E. R. Eliel
Promotiecommissie:	Prof. dr. M. R. H. Mandjes (University of Amsterdam) Dr. R. Mastrandrea (IMT Advanced School Lucca, Lucca, Italy) Dr. T. Squartini (IMT Advanced School Lucca, Lucca, Italy) Prof. dr. J. Aarts Prof. dr. W. Th. F. den Hollander

Casimir PhD series, Delft-Leiden 2021-30
ISBN 978-90-8593-496-7
An electronic version of this thesis can be found at
https://openaccess.leidenuniv.nl

This work was conducted at the Lorentz Institute for theoretical physics, Leiden University. The author was supported by a state scholarship program funded by the China Scholarship Council (CSC) and the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, the Netherlands).

The cover shows a growing of constraint's number forming ensemble nonequivalence.

To My Family

Contents

Chapter 1

Introduction

Statistical physics is built to explain the macroscopic properties in physical systems from the probabilistic examination of underlying microscopic configurations [1]. As the initial research of statistical physics is focused on thermodynamic systems, most traditional examples in statistical physics are under global constraints, such as the fixed total energy and the fixed number of particles. However, recent research on complex systems shows that local constraints, which is implied by the heterogeneous spatial and temporal dependencies, general existed in natural systems [2–4]. Unlike the global constraint, which works on all units in the system with the same influence, the local constraints work on the different groups of units in the system with different influences. The local constraints are not only a new model, which can be used to describe the non-physical systems with heterogeneous dependencies. It also brings several new phenomenons for statistical physics. In particular, ensemble equivalence is broken in networks with fixed degree sequences by the extensive local constraints [5–7]. This breaking of ensemble equivalence exists in the whole parameter space of networks and even without the appearance of phase transitions, which is the essential condition of ensemble nonequivalence in traditional statistical physics. Ensemble equivalence as a basic assumption in statistical physics has been widely explored both practically and theoretically [8]. This breaking of ensemble equivalence in systems with local constraints will affect many fundamental assumptions and calculations in statistical physics. Therefore, further exploration of statistical physics needs a general theory to describe the system with local constraints and the possible appearance of ensemble nonequivalence.

Moreover, local constraints in complex systems also bring new problems to the information theory. For instance, the activities of neurons in nervous systems with heterogeneous spatial correlations give an information source with numerous interacting units [9]. The temporal dependent fluctuation in financial systems breaks the identical independent assumption of the information sources in the classical information theory [3]. The behaviour of those new information sources is impossible to be described by the random variables with finite outcomes as in classical information theory. To describe them, we need statistical ensembles with local constraints. Particularly, the possible appearance of ensemble nonequivalence that is caused by the extensive local constraints may even break the asymptotic equipartition property in the information theory and affect the information-theoretical bounds. Hence, a nontrivial generalization of the information theory for the system with local constraints is needed to study the information transmission and storage in complex systems.

This thesis is an exploration of systems with local constraints. The first part introduces two physical models with local constraints and studies the breaking of

ensemble equivalence of it. In the second part, these physical models are used to describe information sources and sequences with heterogeneous dependencies to find new information-theoretical bounds and the influence of ensemble nonequivalence in it.

The introduction has four sections. I will introduce the definition of statistical ensembles in section1.1. The basic conception of ensemble nonequivalence is introduced in section1.2. The classical Shannon information theory is introduced in section1.3. Section1.4 is the outline of this thesis.

1.1 Statistical ensembles

In the research on thermodynamic systems, to describe the behaviour of numerous random interacted particles is a difficult task, e.g., a glass of water can easily contain 10^{23} molecules. Classical mechanics can be used to describe the collision between the finite number of water molecules, but it is impossible to understand the process of collision with 10^{23} units analytically. This insurmountable problem prompts physicists to build new mechanics to describe the macroscopic behaviour of the system and still based on the physical law already know from the microscopic view. This requirement leads to the birth of statistical physics in the 19th century [1, 10].

Generally, the founding of statistical physics is credited to three physicists. Ludwig Boltzmann gives the fundamental definition of entropy by the collection of microstates [10]. James Clerk Maxwell applies the model of the probability distribution into the description of those microstates [11]. Josiah Willard Gibbs defines statistical ensembles and explains the laws of thermodynamic as the statistical properties of the ensembles [1]. This pattern also can be found later in the establishment of information theory. The initial study focuses on the analytical configurations, but the theory is established on probabilistic explanation.

The framework of statistical physics can be explained as the application of statistical methods and probability theory to large assemblies of microscopic entities and then using the mathematical tools for dealing with large populations and approximations to connect the behaviour of the microscopic entities with the macroscopic property. There are three basic postulates in statistical physics. The ergodic hypothesis shows that each state with the same energy has the equiprobable probability to appears in the system over long periods. The principle of indifference shows that we can only assign equal probabilities to each state when there is insufficient information to describe states in this system. The maximum information entropy presents that the correct probability distribution of the states in the system with limited information should maximum the Gibbs entropy (Information entropy) of it [1]. In the past decades, the principle in statistical physics is not only applied in the physics system but also generalized in chemistry, biology, economics, and even social science [12–14], to build the relationship between the microscopic entries and the macroscopic property of them.

The building of the statistical ensembles needs states in thermodynamic systems

to remain 'static'. This static is that the macroscopic observable variables in the system keep stable under the unpredictable internal particles' motion. These observable variables can be the total energy, the temperature, and the pressure [1]. The difference between macroscopic variables allows us to describe the systems by different ensembles.

- *Microcanonical ensemble* is used to describe the system with a fixed total energy and particle numbers. In this isolated system, the energy change between the system and the outside environment is forbidden. States in this ensemble have the same probability of appearing in the system. The value of the probability is decided by the total number of microscopic configurations in this system.

- The *Canonical ensemble* is used to describe the system with fixed particle numbers and temperature. This system contacts a heat bath, which has a precise temperature. There is an energy change between this system and the heat bath. The probability of each state appears in this system is decided by their total energy under the fixed temperature.

- The *Grand canonical ensemble* is used to describe the open system. Both the total energy and the number of particles are not fixed. Systems described by this ensemble will change energy and particles with the outside environment. The probability of each state in this ensemble is decided by their total energy and their total particles.

According to the energy isolation, the microcanonical ensemble and the canonical ensemble are under two different constraints. The microcanonical is under 'hard' constraints, as all states should have the same total energy. But the canonical ensemble is under 'soft' constraints, the total energy of each state in this ensemble is not the same, but the average value should fix.

As the way to describe the interactions are decided by the property of constraints of the statistical ensemble, the soft or hard constraints in the systems not only decide which statistical ensemble is suitable for their descriptions but also affect the way to describe the inner interactions [15, 16]

The traditional statistical ensemble is under global constraints like the total energy, the fixed temperature or the total number of particles in the system. But the information sources and the process of signal generating of the complex systems are all under heterogeneous interactions. It means the constraints in the information source or the signal generation need to be localized to describe those heterogeneous interactions. The localized constraint is not a new concept, it has been studied in networks theory, where the nodes in the networks always have different degrees, and this heterogeneous degree distribution will affect the dynamic and structure of the networks [16–20].

However, systems that need to model in the new information theory is more general than the binary network with fixed degree sequences. The interactions among units in information sources may have different degrees. Thus, using the weighted

network to model those heterogeneous dependencies is more reasonable than the binary networks. Furthermore, when the signal generated by the finite variables under temporal dependences in long periods needs to record, the data structure we need is not the adjacency matrix. This data structure is close to an $m \times n$ matrix, where m represents the number of finite variables and n is the increasing length. Therefore, to describe the systems mentioned above, new models with local constraints are needed.

1.2 Ensemble (non)equivalence

Normally, the three ensemble descriptions are unavoidable different when the system has a finite size. There are fluctuations in the macroscopic properties of states in the canonical and grand canonical ensemble. But when the system has numerous particles (in the thermodynamic limit), all the three ensembles tend to give an identical description. The fluctuation of macroscopic properties will vanish. This is a basic assumption in the tradition statistical physics, which is named as ensemble equivalence [1, 8, 21].

Specifically, in the microcanonical ensemble and the canonical ensemble, this equivalence appears in the systems with the same total number of particles but under 'soft' or 'hard' constraints (according to the energy isolation). Since the canonical ensemble is mathematically easy to calculate, the presence of the ensemble equivalence means that the replacing of the microcanonical ensemble with the canonical ensemble in the application of statistical physics is allowed [8]. The ensemble equivalence has three forms,

- *Thermodynamically equivalence*: When the entropy of the microcanonical ensemble and the free energy of the canonical ensemble is one-to-one related under the Legendre transform, the two ensembles are believed under thermodynamical equivalence.

- *Macrostate equivalence*: The macrostate equivalence is the equilibrium values of the macrostate predicted by the microcanonical ensemble is the same as the one of the canonical ensemble in the thermodynamic limit.

- *Measure level equivalence*: when the probability distribution of states in the canonical ensemble Gibbs distribution)converges to the probability distribution of the states in the microcanonical ensemble that is defined by the Boltzmann's equiprobability postulate, we believe the two ensembles are under measure level equivalence.

The presence of ensemble equivalence in the thermodynamic limit shows that although the specific action in the canonical ensemble varies from one microstate to another, most of the microstates are still roughly equiprobable.

However, this ensemble equivalence does not always hold. Recent researches on fluid turbulence [22, 23], star formation [24, 25] and networks [26] show that the ensemble equivalence will break at the critical point on the boundary of phase transition.

This ensemble nonequivalence is caused by the nonadditivity in the system with long-range interaction [13, 8]. More recently, in networks with fixed degree sequences, the breaking of ensemble equivalence even has been found in the whole parameter space under the complete absence of phase transitions [5, 7, 26]. It means the ensemble equivalence is not only a critical phenomenon but also an intrinsic property of the system with extensive local constraints [27, 28]. Therefore, there will always have a non-neglected difference between different ensembles in the ensemble nonequivalent systems.

The three forms of equivalence have three coincide ensemble nonequivalences, and those three forms already proved equivalent in [8]. For instance, the thermodynamical nonequivalence is the function of microcanonical entropy, not one-to-one relates to the *Legendre transform* of the canonical ensemble. It can be mapped by the difference between the prediction of macroscopic property for different ensembles. And it also can be detected by the difference between the probability distribution of the states in different ensembles [8].

The *measure level ensemble equivalence* is based on the difference between the probability distributions of different ensembles [8]. States in the microcanonical ensemble with the same total energy all belong to the conjugate canonical ensemble [15]. The measure-level ensemble equivalence between the two ensembles is that the probability distribution of the states in the canonical ensemble converges to the probability distribution of the states in the microcanonical ensemble. Therefore, under measure level ensemble nonequivalence, the difference between two probability distributions will not vanish, and it can be quantified by the relative entropy between the microcanonical and canonical ensemble as

$$S[P_{\mathrm{mic}}||P_{\mathrm{can}}] = \sum_{x\in\S} P_{\mathrm{mic}}(x) \ln \frac{P_{\mathrm{mic}}(x)}{P_{\mathrm{can}}(x)}, \tag{1.1}$$

where § is the collection of all the possible microscopic configurations of the system with n particles in it. Symbol $P_{\mathrm{mic}}(x)$ represents the probabilty of the state x in the microcanonical ensemble. Symbol $P_{\mathrm{can}}(x)$ represents the probability of the state x in the canonical ensemble.

The indicator of the measure ensemble nonequivalence is the specific relative entropy density [8, 5], which is defined as the limit of the relative entropy rescaled by the number of particle in it

$$s_\infty = \lim_{n\to\infty} \frac{1}{n} S[P_{\mathrm{mic}}||P_{\mathrm{can}}]. \tag{1.2}$$

When the value of s_∞ is equal to 0, we believe the system is under measure level ensemble equivalence. When $s_\infty > 0$, the system is believed to be under measure level ensemble nonequivalence [8].

Ensemble nonequivalence appears in the boundary of phase transition in the system with long-range interactions has been well studied theoretically and experimentally [13, 8]. But we know little awareness about the ensemble nonequivalence in the

system with local constraints. Compared to the one that only happens on the critical point of parameter space, the ensemble nonequivalence in the system with local constraints has a more general form. In the binary networks with degree sequences, the ensemble nonequivalence appears in the whole parameter space of it [5]. No matter how the degree distribution of the network change, the probability distribution of the networks' configuration in the microcanonical always has a non-neglected difference between the canonical ensemble. This difference will affect the generalization of statistical ensembles to complex systems with heterogeneous interaction, especially in the information theory, as the information source and the information transmission are all described by the probability theory. Thus, a more fundamental study about the ensemble nonequivalence in the system with local constraints is needed.

Classical information theory has a close relationship with statistical physics. The increasing length of the sequences in information theory is the same as the extension of the system's size in statistical physics. The identical probability distribution of the information sources and the independent signal generating are all coincided with the definition of the microscopic behaviour of the particles in statistical physics under global constraints [29, 5]. Thus, the information-theoretical bounds as the macroscopic properties of the information storage and transmission are decided by the statistical properties of microscopic configurations. The ensemble nonequivalence that may appear in the information source with heterogeneous interaction or the information sequences with temporal dependence has a distinct influence on the information storage and transmission. In other words, the generalization of the information theory for the complex system should be based on the *ensemble nonequivalent* systems.

1.3 Information theory

The information theory is built to describe the information transmission and storage in communication system [29]. The birth of information theory in the first half of the twentieth century is believed to be stimulated by the dramatic development of electronic communication systems. At that time, the worldwide electronic telegraph network continuous works more than a half-century. The birth of the telephone and television has already completely changed the daily life of humans. But there is still do not have a quantifiable definition of the information [30]. The earliest attempt to quantify information can be traced to 1924 when Nyquist introduced a theoretical speed of information transmission depending on the change of voltage in the line [31]. In 1928, Hartley first used the word 'information' to describe the stuff flows in the communication and generalized the definition of Nyquist into the whole communication systems as the number of possible states of the symbols transmits in this system [32]. Afterwards, in 1948, based on the probability theory, Shannon gives the first quantifiable definition of information (information entropy) and the information-theoretical bounds of the information storage and transmission [33].

Shannon divided the communication system into three parts: the information source, the channel and the receiver. These three parts can be described independently

by the probability theory [33]. Information generated by the information sources $\mathbf{x}$ can be quantified by the definition of information entropy $H(\mathbf{x})$, which is equal the expected value of each probability's logarithm as $H(\mathbf{x}) = -\sum_{x\in\mathbf{x}} p(x)\log p(x)$ [33].

This information theory solved the two main problems in the research on artificial communication systems at that time. One is the smallest space to store the information generated by the information source, and the other is the maximum speed of reliable information transmitted through a channel. Shannon found that the information generated by the information source $\mathbf{x}$ is carried in the sequences $\{x_1, x_2, \cdots, x_n\}$ to record the state of the sources in n times' activities. Therefore, the space to store the information generated by the information source is equivalent to the space to store those sequences. Simultaneously, Shannon also found that to store the information generated by the information sources only need to focus on the sequences in the typical set, as the sum of the probability of sequences in the typical set is close to 1. Therefore, the size of the typical set decides the size of the smallest space to store the information carried by those sequences. This typicality is based on the asymptotic equipartition property (AEP). According to the generalization of AEP into joint variables (source and receiver), Shannon also found the maximum speed of reliable information transmission through a channel is decided by the mutual information between the information source and receiver [33].

The rapidly developing information industry has demonstrated the effectiveness of classical information theory. And the application of it into biology, physics, and other disciplines also give new inspirations to solve the old problems or find new phenomenons in those systems [15, 34, 35]. Simultaneously, the applications also bring new issues back to the research of information theory. As in the systems like nervous systems with billions of neurons and social networks with billions of users [9, 36], the information sources are not a single variable but enormous interacted units. On the other hand, the process of signal generating by the information source is not independent. The probability of the information source to get different states is not identical, and it is constrained by the interactions from other units and the dependence from its past states [9]. Therefore, a new generalization of the information theory to deal with the heterogeneous dependencies in those systems is required.

To establish the new generalization information theory, we need to find new models to describe the information sources with enormous interacted units. And we also need to describe the process of single generating with heterogeneous dependence. The statistical ensembles from statistical physics are suitable for this duty [15], as they are build to describe the motion of particles with numerous interactions. The information sources with numerous interacted units need to be modelled by the statistical ensembles with local constraints. The temporal dependence in the process of the single generating that has broken the (identical independent distribution) i.i.d. assumption also can be modelled by the local constraints in the statistical ensemble. Therefore, this generalization is a combination of statistical physics and information theory for systems with local constraints.

1.4 Outline of this thesis

This thesis has two main parts: the first part is the study of systems with local constraints in statistical physics, which includes Chapter 2 and Chapter 3. The second part is the generalization of statistical ensembles with local constraints into information theory. It also has two chapters to introduce our contributions: chapter 4 and chapter 5.

In the second chapter, we focus on the weighted networks with core-periphery structures, which provides a possible model that has the phase transition and local constraint simultaneously, We find that relative fluctuation of constraints as a criterion to check the ensemble nonequivalence in traditional statistical physics vanish in the non-BEC phase while some of them do not disappear in the BEC phase. This result shows that fluctuations of constraints are sensitive to the phase transition. By contrast, the non-vanished relative entropy density for all positive temperatures shows that the extensive number of constraints breaks the ensemble equivalence. Only at zero temperature, where the effective number of constraints becomes finite, ensemble equivalence is broken by BEC in a subtle way. Therefore, in the presence of local constraints, the vanishing of relative fluctuations no longer guarantees ensemble equivalence.

In the third chapter, we extend the discussion of ensemble nonequivalence into a more general local constrained system, the $n \times m$ matrix ensemble $\mathbf{G}$. This matrix can be binary or weighted by setting the range of the value each entry archived. In this matrix ensemble $\mathbf{G}$, we can have the global constraint, one-side local constraints and two-side local constraints. In this general model, ensemble equivalence is still broken by the extensive number of constraints, both in the one-side local constraints (with n local constraints in it) and two-side local constraints (with $m + n$ local constraints in it). Surprisingly, when m is finite, the relative entropy of microcanonical and canonical ensemble has the same order as the canonical entropy, which is as strong as the one that appears in the boundary of phase transition caused by the non-activity. The result breaks the former empirical cognition that the ensemble nonequivalence should be 'strong' but 'restricted' as in the boundary of phase transition or 'general' but 'weak' as in the networks with degree sequences. Ensemble nonequivalence in the system with local constraints can be both 'general' and 'strong' when the units have a finite degree of freedom.

In the fourth chapter, the statistical ensembles with local constraints are used to describe the information source with numerous heterogeneous interacted units. We find the typical set of the microcanonical ensemble described information sources is a subset of the conjugate canonical ensemble described information sources. The classical information theory used to describe the multivariate information sources is a special case of the canonical ensemble description when the local constraints in the information sources are independent and identical. The microcanonical ensemble descriptions need less space to store the information generated by them. The extra sequences we counted in the canonical-typical set are those sequences, which have the same sum of *Hamiltonian* but different constraints comparing with the states in

the microcanonical ensemble. The size of the extra space is decided by the degree of ensemble nonequivalence between the two ensembles. When the information sources are under strong ensemble nonequivalence, the space we can save by using the microcanonical ensemble has the same order as we cost in canonical ensemble description (in classical information theory). The information-theoretical bounds are directly affected by the ensemble nonequivalence in it.

In the fifth chapter, we focus on signal generating with coupled dependencies. The activity of variables in information sources is affected by the interactions between other variables and their historical behaviours. Each unit in the information source has two different dependencies: the spatial correlations come from the interaction between other variables in the source, and the temporal dependencies come from the historical sampling of itself. But all the information generated by these information sources is still carried by the information sequences, which are independent of each other. Thus, to find the limit of information storage and transmission under this situation, we need to use statistical ensembles with local constraints to describe the information sequences, not focus on the information source as in traditional information theory. We find that the breaking of ensemble equivalence in the signal generating is determined by the extensive spatial variational dependence among the variables in the information source, not the finite temporal dependencies of each variable itself. This result also explains why Shannon's classical information theory is so powerful. As in the classical information theory, there is only one variable or several independent variables in the information source, so there is no spatial dependence in this i.i.d. process. The sequences described by the classical information theory are under ensemble equivalence. The temporal dependence realized by the hard or soft constraints is equivalent to each other. The finite number of temporal dependence is not enough to break the ensemble equivalence between the canonical ensemble descriptions and the microcanonical ensemble descriptions. Thus, the signal generating process described by the classical information theory will approach the actual signal generating process, following the length increasing of the information sequences.

The last chapter gives the conclusion and some open problems about systems with local constraints.

Chapter 2

Ensemble nonequivalence and Bose-Einstein condensation in weighted networks

Abstract

The asymptotic (non)equivalence of canonical and microcanonical ensembles, describing systems with soft and hard constraints respectively, is a central concept in statistical physics. Traditionally, the breakdown of ensemble equivalence (EE) has been associated with nonvanishing relative canonical fluctuations of the constraints in the thermodynamic limit. Recently, it has been reformulated in terms of a nonvanishing relative entropy density between microcanonical and canonical probabilities. The earliest observations of EE violation required phase transitions or long-range interactions. More recent research on binary networks found that an extensive number of local constraints can also break EE, even in absence of phase transitions. Here we study for the first time ensemble nonequivalence in weighted networks with local constraints.

Unlike their binary counterparts, these networks can undergo a form of Bose-Einstein condensation (BEC) producing a core-periphery structure where a finite fraction of the link weights concentrates in the core. This phenomenon creates a unique setting where local constraints coexist with a phase transition. We find surviving relative fluctuations only in the condensed phase, as in more traditional BEC settings. However, we also find a non-vanishing relative entropy density for all temperatures, signalling a breakdown of EE due to the presence of an extensive number of constraints, irrespective of BEC. Therefore, in the presence of extensively many local constraints, vanishing relative fluctuations no longer guarantee EE [1].

[1] This chapter is based on:
Qi Zhang, Diego Garlaschelli, "Ensemble nonequivalence and BEC in weighted networks" arXiv preprint arXiv:2012.09998 (2020)

2.1 Introduction

Statistical ensembles were introduced by Gibbs [1] to mathematically describe systems at thermodynamic equilibrium, i.e. where certain conserved macroscopic properties (such as the total energy) are constant, while the microscopic state (i.e. the state of all the microscopic constituents) is subject to fluctuations. For a system with n units and discrete degrees of freedom, a statistical ensemble is a probability distribution $P(\mathbf{W})$ over the collection $\mathcal{W}_n = \{\mathbf{W}\}$ of all the possible (unobserved) microscopic states of the system, given a set of measurable macroscopic properties. Clearly, $P(\mathbf{W}) \geq 0$ for all $\mathbf{W} \in \mathcal{W}_n$ and $\sum_{\mathbf{W}\in\mathcal{W}_n} P(\mathbf{W}) = 1$. This distribution conceptualizes the fact that, ideally, repeated observations of the microscopic state would retrieve different (and independent) outcomes. It can be viewed as the probability distribution that maximizes the Gibbs-Shannon entropy functional

$$S[P] \equiv -\sum_{\mathbf{W}\in\mathcal{W}_n} P(\mathbf{W}) \ln P(\mathbf{W}), \tag{2.1}$$

under a set of (macroscopic) constraints, therefore being maximally noncommittal with respect to missing (microscopic) information [15].

Depending on the choice of the macroscopic properties being constrained, different statistical ensembles can be constructed. The *microcanonical ensemble* $P_{\text{mic}}(\mathbf{W}|E)$ is used to describe systems with fixed total energy E (energetic isolation), while the *canonical ensemble* $P_{\text{can}}(\mathbf{W}|\beta)$ is used to describe systems with fixed (inverse) temperature β (thermal equilibrium) [1]. For a physical system, the inverse temperature β equals $1/kT$ where T is the absolute temperature and k is Boltzmann's constant. In both ensembles, the microscopic state $\mathbf{W}$ is random, but the randomness is governed differently by the two distributions $P_{\text{can}}(\mathbf{W}|\beta)$ and $P_{\text{mic}}(\mathbf{W}|E)$. In particular, while the microcanonical ensemble assigns each realized configuration $\mathbf{W}$ a constant (deterministic) value $E(\mathbf{W}) = E$ of the total energy (which can therefore be regarded as a 'hard' constraint corresponding to energetic isolation), the canonical ensemble assigns configurations a fluctuating (random) energy with a certain expected value $\langle E\rangle_\beta = \sum_{\mathbf{W}\in\mathcal{W}_n} P_{\text{can}}(\mathbf{W}|\beta)E(\mathbf{W})$ and a positive standard deviation $\sigma_\beta(E) > 0$ (i.e. the energy plays the role of a 'soft' constraint resulting from the contact with a heat bath at fixed inverse temperature β).

2.1.1 Conjugate ensembles

The two ensembles can be made conjugate to each other by choosing a specific value E^* and simultaneously setting the total energy $E(\mathbf{W})$ of each realized configuration $\mathbf{W}$ in the microcanonical ensemble equal to $E(\mathbf{W}) \equiv E^*$ and the inverse temperature β in the canonical ensemble to the corresponding value $\beta \equiv \beta^*$ such that the resulting average value $\langle E\rangle_{\beta^*}$ of the fluctuating total energy under the canonical probability $P_{\text{can}}(\mathbf{W}|\beta^*)$ equals E^*, i.e.

$$\langle E\rangle_{\beta^*} \equiv E^*. \tag{2.2}$$

For systems with finite size, the two conjugate ensembles are unavoidably different, because in the microcanonical ensemble the hardness of the constraint implies extra dependencies among the state of the microscopic constituents with respect to the canonical case. However, in the *thermodynamic limit* (i.e. when the number of units in the system goes to infinite) and under certain 'natural' circumstances, the two associated probabilistic descriptions are expected to become effectively equivalent (i.e. the canonical fluctuations and the microcanonical dependencies are both expected to play an asymptotically vanishing role) as a result of some form of the law of large numbers. This idea, which dates back to Gibbs himself [1] and has continued to attract a lot of interest until presently [13, 8, 5, 37], goes under the name of *ensemble equivalence* (EE). When EE holds, one can treat the two ensembles as asymptotically interchangeable, and hence use any of them based on mathematical or computational convenience. For instance, analytical calculations are signficantly easier in the canonical ensemble, while numerical randomizations of an initial configuration can be carried out more naturally in the microcanonical ensemble.

Most statistical physics textbooks still convey the message that EE is expected to hold in general as a sort of principle at the basis of ensemble theory. The possible breakdown of EE, also known as *ensemble nonequivalence* (EN), is still not discussed systematically in the literature. However, several observations of EN have been documented over the past decades [22, 38, 39, 23–25, 40–42, 21, 43, 13, 5]. These observations motivated various efforts aimed at elucidating both the possible physical mechanisms at the origin of EN and, in parallel, its proper mathematical definition(s).

2.1.2 Physical mechanisms for ensemble (non)equivalence

Traditionally, the 'natural' circumstances generally invoked to ensure EE mainly concern the presence of (loosely speaking) 'at most weak' interactions between the constituents of the system. This condition is automatically realized when the system consists of independent units or units with short-range interactions and sufficiently high temperature (to stay away from possible low-temperature phases with broken symmetries, for which the canonical average value of the energy is no longer the typical value). Indeed, violations of EE have been documented in presence of long-range interactions (e.g. in gravitational systems) or phase transitions (e.g. in interacting spin systems) [22, 38, 39, 23–25, 40–42, 13, 8].

However, recent research on complex systems encountered beyond the usual realm of physics has found an additional mechanism that can break EE, even in presence of weak (or no) interactions: namely, the presence of *an extensive number of local constraints* [5, 37]. This situation is frequently found in networks with constraints on the number of links (*degree*) of each node. More specifically, the *binary configuration model* [44] is a widely used null model of graphs with a given *degree sequence*, i.e. a given vector of node degrees. The model captures many properties found in real-world networks, because the local character of the degree constraint can accomodate the strong structural heterogeneity typically observed across nodes in real networks. Unlike the traditional thermodynamic example where the total energy (and possibly

a small, finite number of additional macroscopic properties) is a global and unique constraint for the system, networks with given node degrees are systems with as many constraints as the number of fundamental units, i.e. where constraints are *extensive in number and local in nature.* This situation has been found to break the equivalence of the corresponding canonical and microcanonical ensembles, even without long-range interactions or phase transitions [5, 26, 45]. Notably, since systems with local constraints are generic models for virtually any heterogenous system, the new mechanism significantly widens the range of real-world cases for which EE may break down. This novel result deserves further research.

2.1.3 Mathematical definitions of ensemble (non)equivalence

Besides the aforementioned advances in the study of the physical mechanisms at the origin of EN, significant progress has been made in the mathematical characterization of EN as well. Traditionally, the informal criterion [1] used to test whether two conjugate ensembles are equivalent is checking whether the *relative fluctuations* of the constraint in the canonical ensemble, e.g. the ratio $\sigma_{\beta^*}(E)/E^*$ of the canonical standard deviation to the average value of the energy, vanish in the thermodynamic limit. If this happens, then the canonical fluctuations of the total energy are negligible with respect to the total energy itself and, intuitively, the energy in the canonical ensemble can be though of as an effectively deterministic quantity, very much like in the conjugated microcanonical ensemble. Similarly, the extra dependencies among the microscopic constituents in the microcanonical ensemble are expected to play a smaller and smaller role, thus coming closer to the canonical case.

More recent approaches have considered different rigorous definitions of EE, which can be beautifully summarized [8] as the following three notions: *thermodynamic* equivalence (convexity of the microcanonical entropy density), *macrostate* equivalence (equality of the expected values of macroscopic quantities under the two ensembles), and *measure* equivalence (vanishing of the relative entropy density between the microcanonical and canonical probability distributions). Under mild conditions, these notions have been shown to be equivalent [8]. In this paper, we use measure equivalence as it is more transparently related to the ensemble probabilities.

As a useful result, research on the relationship between statistical physics and combinatorics has revealed that the relative entropy between the microcanonical and the canonical probability distributions is, under certain conditions, asymptotically proportional to the logarithm of the determinant of the covariance matrix of the *effective* constraints in the canonical ensemble [37]. The effective constraints are those that are neither redundant, i.e. trivially replicating other constraints, nor degenerate, i.e. deterministically restricting the canonical and microcanonical configurations in exactly the same way. For instance, formally imposing 'two' constraints where one is the total energy E and the other one is twice the total energy $2E$ is clearly a redundant choice: the effective number of constraints is just one in this case. As another example, if in addition to the energy E we impose its square value E^2, then for those values of the Lagrange multipliers such that $\langle E^2 \rangle = \langle E \rangle^2$ the variance of the

energy will be zero also in the canonical ensemble: E will become degenerate and deterministically equal to its imposed value in both ensembles, so not contributing any difference between the two (by contrast, for parameter values such that $\langle E^2\rangle > \langle E\rangle^2$ there are no allowed configurations in the microcanonical ensemble because the hard values of E^2 and E become conflicting, thereby breaking the equivalence with the canonical one). In general, if the problem is not ill-posed from the beginning, the number of effective constraints coincides with the number of enforced constraints. However, it may happen that some constraints become ineffective for certain degenerate values of the parameters. In any case, for a given parameter value the number of effective constraints coincides with the rank of the covariance matrix of all imposed constraints [37].

Since nonequivalence in the measure sense corresponds to the (super)extensivity of the relative entropy, studying the asymptotic behaviour of the determinant of the (effective) covariance matrix is enough in order to assess ensemble nonequivalence. It is worth noticing that, if there is a single constraint (say, the total energy E), then the determinant of the covariance matrix coincides with the corresponding variance $\sigma^2_{\beta^*}(E)$ and the relative entropy grows asymptotically (under the necessary hypotheses) as $\ln \sigma^2_{\beta^*}(E) = 2 \ln \sigma_{\beta^*}(E)$. On the other hand, since E is a global constraint, it is generally extensive in the number n of units of the system. Therefore the vanishing of the relative fluctuations, i.e. $\sigma_{\beta^*}(E) = o(E)$, implies that the relative entropy is subextensive, i.e. the relative entropy density vanishes in the thermodynamic limit. This suggests that, *in presence of a global constraint, the vanishing of the relative fluctuations implies ensemble equivalence.*[2] How this picture changes in presence of an extensive number of local constraints has not been investigated yet. In particular, whether the vanishing of relative fluctuations still implies ensemble equivalence remains an open question.

2.1.4 The contribution of this chapter

This chapter connects to both lines of research described above (physical mechanisms and mathematical definitions for EN) and its aim is therefore twofold. On the one hand, we aim at investigating for the first (to the best of our knowledge) time the phenomenon of EN in a model system that combines the presence of an extensive number of local constraints with a phase transition. On the other hand, we aim at understanding whether the intuitive criterion of vanishing relative fluctuations of the constraints still ensures EE in this more general setting.

Concretely, we consider the *weighted configuration model* [44], namely a model of *weighted* (as opposed to binary) networks with given *strength sequence*, i.e. with given values of the *strength* (sum of the weights of incident links) of each node. The weighted character of the model allows for the emergence of a phase transition that is

[2]Note that the converse is not necessarily true. However, observing ensemble equivalence and non-vanishing relative fluctuations simultaneously requires some rather uncommon circumstances: for instance, if $\sigma_{\beta^*}(E)$ grows like E^α with $\alpha \geq 1$ while the entropy still grows like E, then the relative entropy is still subextensive, while the relative fluctuations do not vanish.

impossible to observe in the corresponding binary configuration model, namely Bose-Einstein Condensation (BEC) [46, 19]. For the sake of clarity, it is worth mentioning here that, although a form of BEC in networks was identified for the first time in growing binary graphs [47], the notion we refer to here refers to static networks and as such can only occur in weighted networks [46]. Indeed, while the configuration model for weighted networks obeys Bose-Einstein statistics, the configuration model for binary networks obeys Fermi-Dirac statistics [46, 19, 20]. BEC can arise in our model by appropriately tuning the strength sequence. In particular, we are going to show that we can make the strength sequence temperature-dependent and generate BEC by picking a sufficiently low temperature, below a certain critical value. The simplest such setting is one where the network has a 'core-periphery' structure, with BEC appearing in the core.

We find that, for all temperatures and irrespective of whether BEC emerges, the canonical and microcanonical ensembles are always nonequivalent as signalled by a nonvanishing relative entropy density. On the other hand, the relative fluctuations of all the constraints vanish when BEC is absent, while some of them do not vanish when BEC is present. This shows that the relative fluctuations cannot distinguish between equivalence and nonequivalence of the ensembles in this more general case where multiple constraints are present. In fact, what they do is detecting the presence of BEC. Therefore the traditional criterion for EE based on the vanishing of the relative fluctuations is no longer valid in presence of an extensive number of constraints, even when applied simultaneously to all constraints. These results enrich our understanding of the phenomenology of EN and shed more light on its relationship with both the extensivity of the constraints and the presence of phase transitions.

The remainder of this chapter is organized as follows. In Sec. 2.2 we rigorously define the canonical and microcanonical ensembles of weighted networks with given strength sequence. In Sec. 2.3 we introduce two criteria for the (non)equivalence of the ensembles, one based on the relative entropy between the corresponding probability distributions (measure equivalence) and one based on the relative fluctuations of the constraints. In Sec. 2.4 we study in detail a model defined by the simplest family of strength sequences, driven by a temperature parameter, such that we can observe both a BEC and a non-BEC phase. In Sec. 2.5 we offer our conclusions. Finally, the Appendix contains useful calculations needed to establish the scaling of the relative entropy in all the regimes considered.

2.2 Canonical and microcanonical ensembles of weighted networks

In this section, we introduce the definition of weighted networks and of canonical and microcanonical ensembles of weighted networks with given strength sequence.

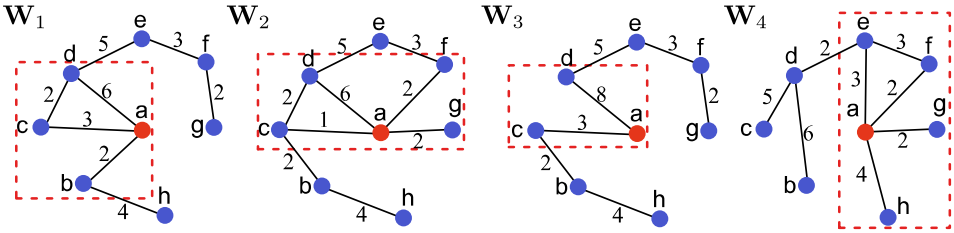

Figure 2.1. Example of four weighted networks $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{W}_4$ having the same number $n = 8$ of nodes (labelled from a to h) and the same strength sequence $\vec{s}(\mathbf{W}_i) = \vec{s}^* = (11, 6, 5, 13, 8, 5, 2, 4)$ $(i = 1, 4)$, but different structure ($\mathbf{W}_i \neq \mathbf{W}_j$ for all $i \neq j$). The dashed blocks highlight the links from node a to its neighbours: in different networks, node a can have different neighbours and, importantly, different distributions of link weights (more or less concentrated on specific neighbours). More homogeneous choices of $\vec{s}^*$ would result in less concentrated link weights, while more heterogeneous choices of $\vec{s}^*$ would result in more concentrated link weights.

2.2.1 Weighted network ensembles

Weighted networks are widely used to describe systems with a large number of components and heterogeneous patterns of interaction [17]. We represent a possible configuration of a weighted network with n nodes as an $n \times n$ weighted adjacency matrix $\mathbf{W}$. Each entry of the matrix w_{ij} $(1 \leq i \leq n, 1 \leq j \leq n)$ denotes the weight of the link between node i and node j, which is taken from the set $\mathbb{N}$ of natural numbers (including zero, which corresponds to the absence of a link between i and j). In this work, we only consider undirected networks without self-loops, thus the weighted matrix $\mathbf{W}$ is a symmetric matrix ($w_{ij} = w_{ji}$ for all i, j) and its diagonal elements are zero ($w_{ii} = 0$ for all i). The number of independent entries of each such matrix is therefore $\binom{n}{2} = n(n-1)/2$.

An ensemble of weighted networks on n nodes is the discrete (infinite) set $\mathcal{W}_n = \mathbb{N}^{n(n-1)/2}$ of all available configurations for the matrix $\mathbf{W}$ and a probability distribution $P(\mathbf{W})$ over $\mathcal{W}_n$ that is specified by a given vector $\vec{C}(\mathbf{W})$ of constraints, which can be enforced either as a soft constraint (canonical ensemble) or as a hard constraint (microcanonical ensemble) [44]. So the matrix $\mathbf{W}$ is a possible outcome of a random variable. We will consider the weighted configuration model, for which the constraints are the strengths of all nodes, i.e. the strength sequence $\vec{C}(\mathbf{W}) = \vec{s}(\mathbf{W})$, where the strength $s_i(\mathbf{W})$ of node i is a local sum of all the link weights that connect i to its neighbours in the particular network $\mathbf{W}$:

$$s_i(\mathbf{W}) = \sum_{j=1}^{n} w_{ij}, \qquad i = 1, n. \tag{2.3}$$

Clearly, the number of scalar constraints is n, which coincides with the number of nodes, so this model is a perfect example of a system subject to an extensive number

of local constraints. In fact, it is the weighted counterpart of the binary configuration model, where EN driven by local constraints was observed for the first time [5].

A crucial consequence of the presence of local constraints in weighted networks is illustrated in Fig.2.1, where we show an example of different networks with the same strength sequence. For a given choice of $\vec{s}^*$, in different realizations of the network each node i can have different neighbours and different distributions of weights on the links that connect it to those neighbours. In particular, the strength s_i^* can be more or less concentrated on specific neighbours (a property that is usually quantified by the so-called *disparity* [48]). However, more homogeneous choices of $\vec{s}^*$ unavoidably result in less concentrated link weights, while more heterogeneous choices of $\vec{s}^*$ impose more concentrated link weights. This fact will allow us to consider (in Sec. 2.4) different structural regimes ranging between two extreme limits: a constant (infinite-temperature) strength sequence implying that on average each node is connected to its neighbours in an equally strong way, and a 'step-like' (zero-temperature) strength sequence implying an extreme concentration of link weights among a small subset of the n nodes. In between these two limits, a certain critical temperature separates a 'non-condensed' (high-temperature) phase from a 'condensed' (low-temperature) phase featuring the properties of BEC.

2.2.2 Canonical weighted network ensemble

We first discuss how to implement the strength sequence constraint mathematically in the canonical ensemble (soft constraint) [44]. Recall that in the traditional canonical ensemble the inverse temperature β^* is the only (scalar) parameter of the canonical probability distribution $P_{\rm can}(\mathbf{W}|\beta^*)$, conjugate to a certain (scalar) total energy E^* in the corresponding microcanonical ensemble. Explicitly, $P_{\rm can}(\mathbf{W}|\beta^*)$ is the Boltzmann distribution $P_{\rm can}(\mathbf{W}|\beta^*) = e^{-\beta^* E(\mathbf{W})}/Z(\beta^*)$ with inverse temperature $\beta^* = 1/kT^*$. By contrast, in our setting the canonical distribution $P_{\rm can}(\mathbf{W}|\vec{\beta}^*)$ has to depend on an n-dimensional vector $\vec{\beta}^*$ of parameters, conjugate to the n-dimensional constraint $\vec{s}^*$ which, in turn, defines the microcanonical ensemble. This distribution is found by maximizing the Gibbs-Shannon entropy functional defined in Eq. (2.1) under the soft constraint

$$\langle \vec{s} \rangle_{\vec{\beta}^*} = \sum_{\mathbf{W}\in\mathcal{W}_n} P_{\rm can}(\mathbf{W}|\vec{\beta}^*)\vec{s}(\mathbf{W}) \equiv \vec{s}^* \tag{2.4}$$

which generalizes the conjugacy condition in Eq. (2.2). The solution to the maximization problem sees $\vec{\beta}^*$ play the role of a vector of Lagrange multipliers coupled to the strength sequence $\vec{s}^*$, and is given by [44]

$$P_{\rm can}(\mathbf{W}|\vec{\beta}^*) = \frac{e^{-H(\mathbf{W},\vec{\beta}^*)}}{Z(\vec{\beta}^*)}, \tag{2.5}$$

where the (network) *Hamiltonian* $H(\mathbf{W},\vec{\beta}) = \vec{\beta}\cdot\vec{s}(\mathbf{W})$ is the linear combination of the node strengths, the *partition function* $Z(\vec{\beta}) = \sum_{\mathbf{W}\in\mathcal{W}_n} e^{-H(\mathbf{W},\vec{\beta})}$ is a normaliza-

tion constant, and $\vec{\beta}^*$ is the unique parameter value realizing Eq. (2.4). Note that Eq. (2.5) has still the form of the Boltzmann distribution, with the important caution that *the inverse temperature has been reabsorbed into the Hamiltonian.* Therefore, to keep the parallel with the traditional physical situation, in our setting, the Hamiltonian should be thought of as the inverse temperature times the energy, and $\vec{\beta}^*$ as the inverse temperature times a vector of 'fields', each coupled to a different constraint. Clearly, since the probability in Eq. (2.5) must be dimensionless, the product $\vec{\beta} \cdot \vec{s}(\mathbf{W})$ must be dimensionless as well. In Sec. 2.4, we will notice that the Hamiltonian can be further reinterpreted as also incorporating a 'chemical potential' governing the expected weight of the links in the network [20].

Notably, $P_{\rm can}(\mathbf{W}|\vec{\beta}^*)$ depends on $\mathbf{W}$ only through $\vec{s}(\mathbf{W})$. In particular, it gives the same value $P_{\rm can}(\mathbf{W}^*|\vec{\beta}^*)$ to any network $\mathbf{W}^*$ such that $\vec{s}(\mathbf{W}^*) = \vec{s}^*$. Explicitly, given the definition of node strength in Eq. (2.3), the network Hamiltonian can be written for a generic value of $\vec{\beta}$ as

$$H(\mathbf{W}, \vec{\beta}) = \sum_{i=1}^{n} \sum_{i<j} (\beta_i + \beta_j) w_{ij}. \tag{2.6}$$

The partition function can be easily shown [44] to be

$$Z(\vec{\beta}) = \prod_{i=1}^{n} \prod_{i<j} \frac{1}{1 - e^{-(\beta_i + \beta_j)}} \tag{2.7}$$

provided that $\beta_i + \beta_j > 0$ for all i, j (otherwise, the model admits no solution). The canonical probability distribution therefore factorizes over pairs of nodes as

$$P_{\rm can}(\mathbf{W}|\vec{\beta}) = \prod_{i=1}^{n} \prod_{i<j} q_{ij}(w_{ij}|\vec{\beta}), \tag{2.8}$$

where

$$q_{ij}(w|\vec{\beta}) = \frac{e^{-(\beta_i + \beta_j)w}}{[1 - e^{-(\beta_i + \beta_j)}]^{-1}} \tag{2.9}$$

is the probability that the weight of the link between nodes i and j takes the particular value w. Therefore different pairs of nodes are statistically independent in the canonical ensemble (while they are not in the microcanonical one).

Note that $q_{ij}(w|\vec{\beta})$ is a geometric distribution [49, 44] with expected value

$$\begin{aligned} \langle w_{ij} \rangle_{\vec{\beta}} &= \sum_{w \in \mathbb{N}} w \, q_{ij}(w|\vec{\beta}) \\ &= \sum_{w \in \mathbb{N}} w \, \frac{e^{-(\beta_i + \beta_j)w}}{[1 - e^{-(\beta_i + \beta_j)}]^{-1}} \\ &= \frac{e^{-(\beta_i + \beta_j)}}{1 - e^{-(\beta_i + \beta_j)}} \end{aligned} \tag{2.10}$$

(representing the expected weight of the link connecting nodes i and j) and variance

$$\begin{aligned} \mathrm{Var}_{\vec{\beta}}(w_{ij}) &= \frac{e^{-(\beta_i+\beta_j)}}{[1-e^{-(\beta_i+\beta_j)}]^2} \\ &= \langle w_{ij}\rangle_{\vec{\beta}}(1+\langle w_{ij}\rangle_{\vec{\beta}}). \end{aligned} \tag{2.11}$$

As we will discuss in detail is Sec. 2.4, Eq. (2.10) has the form of Bose-Einstein statistics, where $\langle w_{ij}\rangle_{\vec{\beta}}$ plays the role of an expected occupation number for the state labeled by nodes i and j. In an appropriate 'low-temperature' regime, BEC can emerge into the model through the divergence of the occupation number $\langle w_{ij}\rangle_{\vec{\beta}}$ for one (possibly degenerate) 'ground state' (corresponding to $\beta_i+\beta_j \to 0^+$), while the occupation number for the other states remains finite [46, 44, 19]. This discussion requires a series of considerations that we leave for later. For the moment, we notice that Eq. (2.10) allows us to determine the special value $\vec{\beta}^*$ corresponding to the given strength sequence $\vec{s}^*$. Summing over all nodes $j \neq i$, the average value of the strength of node i is

$$\langle s_i\rangle_{\vec{\beta}} = \sum_{j\neq i}\langle w_{ij}\rangle_{\vec{\beta}} = \sum_{j\neq i}\frac{e^{-(\beta_i+\beta_j)}}{1-e^{-(\beta_i+\beta_j)}}, \tag{2.12}$$

whence we can reformulate Eq. (2.4) as

$$\sum_{j\neq i}\langle w_{ij}\rangle_{\vec{\beta}^*} = \sum_{j\neq i}\frac{e^{-(\beta_i^*+\beta_j^*)}}{1-e^{-(\beta_i^*+\beta_j^*)}} \equiv s_i^* \quad i=1,n \tag{2.13}$$

which fixes the unique parameter value $\vec{\beta}^*$. Notably, this value is also the one that maximizes the (log-)likelihood [18, 44], i.e.

$$\vec{\beta}^* = \mathrm{argmax}_{\vec{\beta}} \ln P_{\mathrm{can}}(\mathbf{W}^*|\vec{\beta}), \tag{2.14}$$

where, again, $\mathbf{W}^*$ is any configuration such that $\vec{s}(\mathbf{W}^*) = \vec{s}^*$. In general, it is not possible to write $\vec{\beta}^*$ explicitly as a function of $\vec{s}^*$. However, Eq. (2.13) or equivalently Eq. (2.14) can be efficiently solved numerically [49, 44] using various algorithms that have been coded for this purpose [50, 51]. In any case, a general property (that we will use later) is that for any two nodes i and j with the same expected strength ($s_i^* = s_j^*$), the corresponding parameters β_i^* and β_j^* obey the same equation in (2.13) and are therefore equal. In other words, $s_i^* = s_j^*$ implies $\beta_i^* = \beta_j^*$.

Once $\vec{\beta}^*$ is calculated, we can plug it back into Eq. (2.8) to finally obtain the probability distribution

$$P_{\mathrm{can}}(\mathbf{W}|\vec{\beta}^*) = \prod_{i=1}^{n}\prod_{i<j}\frac{e^{-(\beta_i^*+\beta_j^*)w_{ij}}}{[1-e^{-(\beta_i^*+\beta_j^*)}]^{-1}} \tag{2.15}$$

that characterizes the canonical ensemble entirely. For practical purposes, $\vec{\beta}^*$ can be inserted into Eq. (2.9) to obtain the link weight probability $q_{ij}(w|\vec{\beta}^*)$, from which

several expected network properties can be calculated very directly. For instance, besides the expected link weight $\langle w_{ij} \rangle_{\vec{\beta}^*}$, we can calculate the probability that nodes i and j are connected by a link, irrepective of the weight of the latter, as follows:

$$\begin{aligned} p_{ij}^* &\equiv \langle \Theta(w_{ij}) \rangle_{\vec{\beta}^*} \\ &= \sum_{w=1}^{\infty} w \, q_{ij}(w|\vec{\beta}^*) \\ &= 1 - q_{ij}(0|\vec{\beta}^*) \\ &= e^{-(\beta_i^* + \beta_j^*)} \\ &= \frac{\langle w_{ij} \rangle_{\vec{\beta}^*}}{1 + \langle w_{ij} \rangle_{\vec{\beta}^*}}, \end{aligned} \tag{2.16}$$

where $\Theta(x)$ denotes the Heaviside step function, defined as $\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ if $x \leq 0$. Note that, if i and j belong to the condensed state where the expected link weight $\langle w_{ij} \rangle_{\vec{\beta}^*}$ diverges ($\beta_i^* + \beta_j^* \to 0^+$), then they become deterministically connected, i.e. $p_{ij}^* \to 1^-$. By contrast, non-condensed states have $\langle w_{ij} \rangle_{\vec{\beta}^*} < \infty$, $\beta_i^* + \beta_j^* > 0$, and $p_{ij}^* < 1$. The analogy with BEC will be discussed in much more detail in Sec. 2.4.

Besides the structural properties, one of the key quantities that we will need in order to determine EE (or the lack thereof) is the resulting canonical entropy $S_{\rm can}^*$, obtained by inserting Eq. (2.15) into Eq. (2.1):

$$\begin{aligned} S_{\rm can}^* &\equiv S[P_{\rm can}(\mathbf{W}|\vec{\beta}^*)] \\ &= \langle H(\mathbf{W}, \vec{\beta}^*) \rangle + \ln Z(\vec{\beta}^*) \\ &= \vec{\beta}^* \cdot \langle \vec{s}(\mathbf{W}) \rangle + \ln Z(\vec{\beta}^*) \\ &= \vec{\beta}^* \cdot \vec{s}^* + \ln Z(\vec{\beta}^*) \\ &= -\ln P_{\rm can}(\mathbf{W}^*|\vec{\beta}^*). \end{aligned} \tag{2.17}$$

Note that the calculation of the canonical entropy $S_{\rm can}^*$ of the entire weighted network ensemble only requires the knowledge of the probability of one generic network $\mathbf{W}^*$ with strength sequence $\vec{s}^*$, which is in turn directly calculated through Eq. (2.15).

2.2.3 Microcanonical weighted network ensemble

We now come to the microcanonical ensemble. Its governing probability distribution $P_{\rm mic}(\mathbf{W}|\vec{s}^*)$ can be obtained by maximizing the Gibbs-Shannon entropy functional in Eq. (2.1) under the hard constraint

$$\vec{s}(\mathbf{W}) = \vec{s}^* \tag{2.18}$$

that applies to each network $\mathbf{W}$ realized (with positive probability) in the set $\mathcal{W}_n$. The solution is obviously the uniform probability distribution

$$P_{\rm mic}(\mathbf{W}|\vec{s}^*) = \left\{ \begin{array}{ll} \Omega_{\vec{s}^*}^{-1} & \vec{s}(\mathbf{W}) = \vec{s}^* \\ 0 & \vec{s}(\mathbf{W}) \neq \vec{s}^* \end{array} \right. \tag{2.19}$$

where $\Omega_{\vec{s}^*}$ is the number of networks for which the hard constraint in Eq. (2.18) is realized. An implicit assumption throughout this paper is that the particular strength sequence $\vec{s}^*$ is *graphic*, i.e. it can be realized by at least one network, so that $\Omega_{\vec{s}^*} > 0$. In this case as well, the (microcanonical) entropy is obtained by inserting Eq. (2.19) into Eq. (2.1):

$$\begin{aligned} S_{\rm mic}^* &\equiv S[P_{\rm mic}(\mathbf{W}|\vec{s}^*)] \\ &= \ln \Omega_{\vec{s}^*} \\ &= -\ln P_{\rm mic}(\mathbf{W}^*|\vec{s}^*), \end{aligned} \tag{2.20}$$

which is also known as Boltzmann entropy. Note that $\mathbf{W}^*$ has the same meaning as in Eq. (2.17), therefore both the canonical and microcanonical entropies are equal to minus the log of the corresponding probability, evaluated in any state $\mathbf{W}^*$ realizing the hard constraint in Eq. (2.18).

Note that, although the derivation of $P_{\rm mic}(\mathbf{W}|\vec{s}^*)$ is formally much more direct than that of $P_{\rm can}(\mathbf{W}|\vec{\beta}^*)$ in the conjugate canonical ensemble, its explicit calculation is more challenging, as it requires the combinatorial enumeration of all the $\Omega_{\vec{s}^*}$ weighted networks with strength sequence $\vec{s}^*$ (as a side remark, it is precisely the local nature of $\vec{s}^*$ that makes the calculation of $\Omega_{\vec{s}^*}$ daunting). Here, we will employ a recently proposed saddle-point asymptotic formula, for a generic discrete system under a K-dimensional vector $\vec{C}^*$ of *effective* (see Sec. 2.1.3) constraints, for the number $\Omega_{\vec{C}^*}$ of microcanonical configurations [37]. The formula uses only conjugate canonical quantities, namely the canonical entropy $S_{\rm can}^*$ and the $K \times K$ covariance matrix $\mathbf{\Sigma}^*$ among the K constraints in the canonical ensemble, and reads [37]

$$\Omega_{\vec{C}^*} = \frac{e^{S_{\rm can}^*}}{\sqrt{\det(2\pi\mathbf{\Sigma}^*)}} \prod_{k=1}^{K} [1 + O(1/\lambda_k^*)], \tag{2.21}$$

where $\{\lambda_k^*\}_{k=1}^K$ are the eigenvalues of $\mathbf{\Sigma}^*$. The symbol $O(x)$ indicates a quantity with a finite limit when divided by x as $n \to \infty$, i.e. $O(x)$ is asymptotically of the same order as x. Note that, since covariance matrices are positive semidefinite, $\lambda_k^* \geq 0$ for all k. Moreover, since the constraints are assumed to be non-redundant, then $\lambda_k^* > 0$ for all k [37] (if some of the constraints were redundant, there would be certain zero eigenvalues rendering the above equation inapplicable; that is why the formula should be applied to a maximal set of K non-redundant constraints). Finally, if these eigenvalues grow sufficiently fast as $n \to \infty$, then the product on the right hand side becomes irrelevant, in which case the knowledge of $S_{\rm can}^*$ and $\det(2\pi\mathbf{\Sigma}^*)$ is enough in order to characterize the asymptotics of $\Omega_{\vec{C}^*}$.

In our setting where $\vec{C}^* = \vec{s}^*$ and $K = n$ (node strengths are all mutually independent as it is not possible to guess any individual node strength from the knowledge of the other $n-1$ ones), we calculate the entries of $\mathbf{\Sigma}^*$ as

$$\begin{aligned} \Sigma^*_{ij} &\equiv \mathrm{Cov}_{\vec{\beta}^*}(s_i, s_j) \\ &= \langle s_i s_j \rangle_{\vec{\beta}^*} - \langle s_i \rangle_{\vec{\beta}^*} \langle s_j \rangle_{\vec{\beta}^*} \\ &= \left. \frac{\partial^2 \ln Z(\vec{\beta})}{\partial \beta_i \partial \beta_j} \right|_{\vec{\beta}=\vec{\beta}^*} \end{aligned} \tag{2.22}$$

where $Z(\vec{\beta})$ is given by Eq. (2.7). An explicit calculation gives

$$\begin{aligned} \Sigma^*_{ii} &= \mathrm{Var}_{\vec{\beta}^*}(s_i) \\ &= \sum_{j \neq i} \frac{e^{-(\beta_i^* + \beta_j^*)}}{[1 - e^{-(\beta_i^* + \beta_j^*)}]^2} \\ &= \sum_{j \neq i} \langle w_{ij} \rangle_{\vec{\beta}^*} (1 + \langle w_{ij} \rangle_{\vec{\beta}^*}) \end{aligned} \tag{2.23}$$

for the diagonal entries (i.e. the variances of the constraints) and

$$\begin{aligned} \Sigma^*_{ij} &= \mathrm{Cov}_{\vec{\beta}^*}(s_i, s_j) \\ &= \mathrm{Var}_{\vec{\beta}^*}(w_{ij}) \\ &= \frac{e^{-(\beta_i^* + \beta_j^*)}}{[1 - e^{-(\beta_i^* + \beta_j^*)}]^2} \\ &= \langle w_{ij} \rangle_{\vec{\beta}^*} (1 + \langle w_{ij} \rangle_{\vec{\beta}^*}) \quad (i \neq j) \end{aligned} \tag{2.24}$$

for the off-diagonal entries (i.e. the covariances between distinct constraints).

We finally obtain

$$\begin{aligned} S^*_{\mathrm{mic}} &= \ln \Omega_{\vec{s}^*} \\ &= S^*_{\mathrm{can}} - \ln \sqrt{\det(2\pi \mathbf{\Sigma}^*)} + \sum_{k=1}^{n} \ln[1 + O(1/\lambda_k^*)] \end{aligned} \tag{2.25}$$

where we have used $\sqrt{\det(2\pi\mathbf{\Sigma}^*)} = \prod_{k=1}^{n} \sqrt{2\pi\lambda_k^*}$. Note that the eigenvalues $\{\lambda_k^*\}_{k=1}^n$ are positive, the n node strengths being linearly independent constraints [37]. In principle, in order to compute (the leading term of) Eq. (2.25) explicitly, we need to specify a value for $\vec{s}^*$, calculate the resulting matrix $\mathbf{\Sigma}^*$, and the eigenvalues of the latter. However, in Sec. 2.3 we show that the knowledge of the diagonal elements of the covariance matrix is enough for our purposes. This result is then used in Sec. 2.4 when we consider specific choices of $\vec{s}^*$.

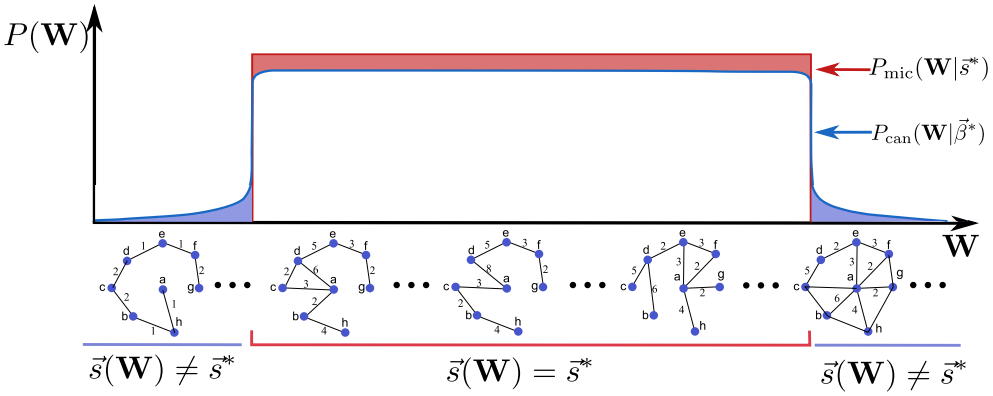

Figure 2.2. Illustration of the nonequivalence of ensembles of weighted networks with given strength sequence $\vec{s}^*$ in the measure sense. Schematically, the x axis represents all weighted networks $\mathbf{W} \in \mathcal{W}_n$. Here, $n = 8$ and $\vec{s}^* = (11, 6, 5, 13, 8, 5, 2, 4)$. The $\Omega_{\vec{s}^*}$ networks matching the particular strength sequence $\vec{s}^*$, i.e. those for which $\vec{s}(\mathbf{W}) = \vec{s}^*$, are represented in the middle. The y axis represents the canonical and microcanonical probabilities for each network. The microcanonical distribution $P_{\rm mic}(\mathbf{W}|\vec{s}^*)$ assigns zero probability to the networks for which $\vec{s}(\mathbf{W}) \neq \vec{s}^*$, and uniform probability $P_{\rm mic}(\mathbf{W}^*|\vec{s}^*) = \Omega^{-1}_{\vec{s}^*}$ to each network $\mathbf{W}^*$ for which $\vec{s}(\mathbf{W}^*) = \vec{s}^*$. By contrast, the conjugate canonical distribution $P_{\rm can}(\mathbf{W}|\vec{\beta}^*)$ assigns positive probability to all the networks in $\mathcal{W}_n$ and therefore has 'tails' that extend all over the x axis. Normalization implies that, while also the canonical probability gives a constant value $P_{\rm can}(\mathbf{W}^*|\vec{\beta}^*)$ to each network $\mathbf{W}^*$ with the strength sequence $\vec{s}^*$, this value is (for n finite) smaller (lower plateau) than the corresponding microcanonical one (upper plateau): $P_{\rm can}(\mathbf{W}^*|\vec{\beta}^*) < P_{\rm mic}(\mathbf{W}^*|\vec{s}^*)$. Indeed, the blue and red areas should be equal because of normalization. Intuitively, the two ensembles become equivalent in the thermodynamic limit if, sufficiently fast as $n \to \infty$, the canonical tails vanish (the blue areas disappear) and the canonical plateau 'catches up' with the microcanonical one (the red area disappears). Measure equivalence formalizes this 'sufficiently fast' rigorously, finding that EE corresponds to the condition $\lim_{n\to\infty}[\ln P_{\rm mic}(\mathbf{W}^*|\vec{s}^*) - \ln P_{\rm can}(\mathbf{W}^*|\vec{\beta}^*)]/n = 0$. It turns out that this is not the case for the model discussed here: canonical and microcanonical ensembles of weighted networks with given strength sequence are not equivalent.

2.3 Equivalence and nonequivalence of weighted network ensembles

In this section we use the knowledge of the canonical and microcanonical probability distributions derived in the previous section in order to establish two criteria for the equivalence of ensembles of weighted networks, namely the one based on the vanishing of the relative entropy density between the two distributions [8] and the traditional one based on the vanishing of the canonical relative fluctuations of the constraints [1].

2.3.1 Relative entropy density

As we have anticipated, EE can be stated mathematically using three different notions, namely *thermodynamic*, *macrostate* and *measure* equivalence [8]. These definitions turn out to be, under mild assumptions, essentially equivalent [8]. Here, we use the definition in the measure sense, which has been recently formulated explicitly for binary network ensembles [5, 26, 45, 37] and is based on the vanishing of a suitable *relative entropy density* between the microcanonical and canonical probability distributions. Our calculations generalize those results to the case of weighted networks, for which measure equivalence has not been studied yet.

In general, the relative entropy (or Kullback-Leibler divergence) between two distributions P and Q, both having support over a discrete set $\mathcal{W}_n = \{\mathbf{W}\}$ of configurations in analogy with Eq. (2.1), is defined as

$$D[P||Q] \equiv \sum_{\mathbf{W}\in\mathcal{W}_n} P(\mathbf{W}) \ln \frac{P(\mathbf{W})}{Q(\mathbf{W})} \tag{2.26}$$

and quantifies 'how far' the distribution P is from the reference distribution Q [52]. When P and Q represent the microcanonical and canonical distributions respectively, it can be shown [5, 37] that $D[P||Q]$ reduces to the difference between the canonical and microcanonical entropy, which can be both estimated on a single configuration realizing the hard constraint (as we have indeed shown in the previous section for our weighted network model). Moreover, its calculation asymptotically requires only the canonical covariance matrix $\mathbf{\Sigma}^*$ between the constraints.

We are now going to see how these general results apply to our specific case. A visual illustration of the idea behind measure equivalence for our ensembles of weighted networks with given strength sequence is provided in Fig. 2.2. Following Eq. (2.26), the relative entropy between $P_{\rm mic}(\mathbf{W}|\vec{s}^*)$ and $P_{\rm can}(\mathbf{W}|\vec{\beta}^*)$ is defined as

$$\begin{aligned} D^* &\equiv D[P_{\rm mic}(\mathbf{W}|\vec{s}^*)||P_{\rm can}(\mathbf{W}|\vec{\beta}^*)] \\ &= \sum_{\mathbf{W}\in\mathcal{W}_n} P_{\rm mic}(\mathbf{W}|\vec{s}^*) \ln \frac{P_{\rm mic}(\mathbf{W}|\vec{s}^*)}{P_{\rm can}(\mathbf{W}|\vec{\beta}^*)}. \end{aligned} \tag{2.27}$$

By inserting Eq. (2.19) into Eq. (2.27) and using the fact that $P_{\rm can}(\mathbf{W}|\vec{\beta}^*)$ has the same value for any network $\mathbf{W}^*$ matching the hard constraint $\vec{s}(\mathbf{W}^*) = \vec{s}^*$, we confirm

that D^* can be estimated pointwise on $\mathbf{W}^*$ as

$$D^* = \ln P_{\rm mic}(\mathbf{W}^*|\vec{s}^*) - \ln P_{\rm can}(\mathbf{W}^*|\vec{\beta}^*). \tag{2.28}$$

Moreover, using Eqs. (2.17) and (2.20), we also confirm that it reduces to the entropy difference

$$D^* = S^*_{\rm can} - S^*_{\rm mic}. \tag{2.29}$$

Now, Eq. (2.25) immediately allows us to obtain

$$D^* = \ln \sqrt{\det(2\pi\mathbf{\Sigma}^*)} + \sum_{k=1}^{n} \ln[1 + O(1/\lambda^*_k)] \tag{2.30}$$

which depends only on the eigenvalues of the canonical covariance matrix $\mathbf{\Sigma}^*$, whose diagonal and off-diagonal entries are given in Eqs. (2.23) and (2.24) respectively.

The definition of measure equivalence is the vanishing of the relative entropy *density*, i.e. of the ratio D^*/n, in the thermodynamic limit [8]. Explicitly, EE in the measure sense corresponds to

$$d^* \equiv \lim_{n\to\infty} \frac{D^*}{n} = 0 \tag{2.31}$$

or equivalently [37]

$$D^* = o(n), \tag{2.32}$$

where $o(n)$ denotes a quantity that, if divided by n, vanishes as $n \to \infty$. Equation (2.29) allows us to understand the above definition of macrostate EE as follows. The microcanonical entropy $S^*_{\rm mic}$ is the logarithm of the number of accessible configurations under hard constraints, while the canonical entropy $S^*_{\rm can}$ is the logarithm of the corresponding 'effective' number of configurations under soft constraints. EE requires that, as n increases, the typical configurations of the system under the two ensembles become the same, i.e. that the (effective) numbers of configurations in the two ensembles become closer to each other. This cannot happen if, as we keep adding one more unit to the system, the difference between the two entropies (i.e. the relative entropy D^*) keeps increasing by an arbitrary amount. The criterion in Eq. (2.31) establishes that if the entropy difference per node, i.e. D^*/n, does not vanish as n diverges, then the two ensembles cannot be equivalent.

Equation (2.32) implies that, in order to assess whether the system is under EE, we do not need the exact value of D^*, but only its leading order with respect to n. Then from Eq. (2.30) we see that, since the term $1 + O(1/\lambda^*_k)$ is at most of the same order as $\sqrt{2\pi\lambda^*_k}$, the presence of $O(1/\lambda^*_k)$ does not affect ensemble (non)equivalence:

$$D^* = O\left(\sum_{k=1}^{n} \ln \lambda^*_k\right) = O(\ln\det \mathbf{\Sigma}^*). \tag{2.33}$$

So, in order to check whether Eq. (2.32) holds, it is ultimately enough to check whether

$$\ln\det \mathbf{\Sigma}^* = o(n). \tag{2.34}$$

On the other hand, since our hypothesis of non-redundant constraints implies $\lambda_k^* > 0$ for all k (as discussed in Sec. 2.2.3), we see that the contribution of the term $\sum_{k=1}^{n} \ln[1 + O(1/\lambda_k^*)]$ to the relative entropy in Eq. (2.30) is at most $O(n)$. So if $\ln \sqrt{\det(2\pi\mathbf{\Sigma}^*)}$ grows faster than $O(n)$ then Eq. (2.30) reduces to the stronger result

$$D^* \approx \ln \sqrt{\det(2\pi\mathbf{\Sigma}^*)} \approx \frac{1}{2} \ln \det \mathbf{\Sigma}^*, \tag{2.35}$$

i.e. the leading term of the relative entropy (not only its leading order) can be calculated exactly from $\ln \det \mathbf{\Sigma}^*$ (throughout this paper, the symbol "$x \approx y$" indicates that the leading term of x and y is asymptotically the same, i.e. the two quantities differ by a quantity that vanishes if divided by either x or y as $n \to \infty$). Depending on the regimes considered later on in the paper, different techniques for calculating (the leading order of) the determinant of the covariance matrix $\mathbf{\Sigma}^*$ can be used. We will discuss these techniques when needed, and refer to the Appendix for explicit calculations. We will show that, except in a certain zero-temperature limit, the conditions ensuring Eq. (2.35) are met and the leading order of the relative entropy can be calculated exactly.

2.3.2 Relative fluctuations of the constraints

We now consider the relative fluctuations of the constraints, whose behaviour in the thermodynamic limit is, historically, the traditional criterion used to check whether statistical ensembles are equivalent [1]. In the standard situation, where the canonical and microcanonical ensembles are defined through a single scalar constraint on the total energy E, the relative fluctuations are captured by a single scalar quantity $r^* \equiv \sigma_{\beta^*}(E)/E^*$ representing the ratio of the canonical standard deviation of the energy to the mean energy itself. EE is then associated with the vanishing of r^* as $n \to \infty$. In statistics, r^* is called the *coefficient of variation* of the random variable E with expected value E^* and variance $\sigma^2_{\beta^*}(E)$

In the case of networks with local constraints, there are n coefficients of variation to consider. They have been calculated for both the binary and the weighted versions of the configuration model [49]. In extreme summary, those results show that, in the binary case (where there is a constraint on the degree k_i^* for each node $i = 1, n$), an *upper bound* $1/\sqrt{k_i^*}$ for the relative fluctuation $r_i^* = \sigma_{\beta^*}(k_i)/k_i^*$ can been established. By contrast, in the weighted case (corresponding to the model considered here with a constraint on the strength s_i^* for each node $i = 1, n$), the value $1/\sqrt{s_i^*}$ becomes a *lower bound* for the relative fluctuation $r_i^* = \sigma_{\beta^*}(s_i)/s_i^*$ [49]. Those results have two implications.

First, in the binary case the only regime for which general conclusions can be drawn about the vanishing of the relative fluctuations is the so-called *dense* regime where the expected degree of all nodes diverges, hence $r_i^* \to 0$. In the opposite *sparse* regime where the average degree of all nodes is finite, we only have a finite upper bound for the relative fluctuations, but their actual value depends on the specific network. In general, however, the decreasing behaviour of the upper bound for the relative

fluctuations in binary networks with increasing degrees is opposite to that of the relative entropy density, which increases as the expected degree increases [5, 26, 45].

Second, in the weighted case we have a somewhat opposite situation where we can only conclude that, in the sparse regime where the expected strengths are finite (apart from possible hubs), the relative fluctuations do not vanish. By contrast, in the dense case where the expected strengths diverge, the relative fluctuations can in principle pick any value. Confusingly, even if in the weighted case the lower bound for the relative fluctuations of the strengths goes to zero for nodes with diverging strength s_i^*, previous results seem to indicate that the realized value of r_i^* in networks with heterogeneous strength sequence actually *increases* for higher s_i^* [49]. This behaviour suggests that weighted networks, due to the many possible ways in which weight can accumulate on links, can behave very differently from binary networks. This observation requires further research and strengthens our motivation for studying the relative fluctuations in weighted networks in different scenarios (ranging from homogeneous to heterogenous concentrations of link weights), in conjunction with the relative entropy and, in general, ensemble (non)equivalence.

Using Eq. (2.23), we can immediately calculate the standard deviation of each constraint s_i around its expected value s_i^* as

$$\begin{aligned}
\sigma_{\vec{\beta}^*}(s_i) &= \sqrt{\mathrm{Var}_{\vec{\beta}^*}(s_i)} \\
&= \sqrt{\sum_{i\neq j} \frac{e^{-(\beta_i^*+\beta_j^*)}}{[1-e^{-(\beta_i^*+\beta_j^*)}]^2}} \\
&= \sqrt{\sum_{i\neq j} \langle w_{ij}\rangle_{\vec{\beta}^*}(1+\langle w_{ij}\rangle_{\vec{\beta}^*})} \\
&= \sqrt{s_i^* + \sum_{i\neq j} \langle w_{ij}\rangle_{\vec{\beta}^*}^2}.
\end{aligned} \tag{2.36}$$

The relative fluctuation of the strength is therefore

$$r_i^* = \frac{\sigma_{\vec{\beta}^*}(s_i)}{s_i^*} = \sqrt{\frac{1}{s_i^*} + \frac{\sum_{i\neq j}\langle w_{ij}\rangle_{\vec{\beta}^*}^2}{(\sum_{i\neq j}\langle w_{ij}\rangle_{\vec{\beta}^*})^2}}, \tag{2.37}$$

Since $\langle w_{ij}\rangle_{\vec{\beta}^*} \geq 0$ for all i, j, we have $(\sum_{i\neq j}\langle w_{ij}\rangle_{\vec{\beta}^*})^2 \geq \sum_{i\neq j}(\langle w_{ij}\rangle_{\vec{\beta}^*})^2$, showing that $1/\sqrt{s_i^*}$ is indeed a lower bound for r_i^*. When studying the asymptotic behaviour of the relative fluctuations in the thermodynamic limit, we will be interested in whether the limit

$$\rho_i^* \equiv \lim_{n\to\infty} r_i^* \tag{2.38}$$

is zero (vanishing relative fluctuations) or positive (nonvanishing relative fluctuations) for each node $i = 1, n$.

2.4 BEC in weighted networks

In physical systems composed of bosons, i.e. particles obeying Bose-Einstein statistics, BEC is a phase transition whereby, below a certain critical temperature, a finite fraction of the total number of particles condenses in the ground state, i.e. the state with lowest energy (or more generally in a finite number of states with lowest energy). BEC was theoretically predicted by Satyendra Nath Bose and Albert Einstein in 1924 [53], and it has since then been observed in various physical systems. Models of BEC have been studied in different statistical ensembles in the standard case with only global constraints (total energy and/or total number of particles) [54–59]. Although the detailed phenomenology exhibited by these models depends on the choice of the energy and the structure of the interactions, it is generally found that EE breaks down in the condensed (low-temperature) phase, as signalled by nonvanishing relative fluctuations of the constraints.

In this Section, we are going to show that a form of BEC, even if quite different from that found in more traditional physical settings, can also appear in our ensembles of weighted networks. The possible onset of BEC in our system creates an ideal situation where an EE-breaking phase transition can be studied in combination with an additional and unrelated mechanism for the breakdown of EE, i.e. the presence of local constraints, which is always active in both the condensed and the non-condensed phases. To illustrate our results, we first make some important clarifications in order to establish a rigorous link from weighted network ensembles to Bose-Einstein statistics and then study the different phases of the model.

2.4.1 Bose-Einstein statistics in weighted networks

As we have already recalled, weighted networks with a constraint on the strength sequence obey Bose-Einstein statistics, as opposed to binary networks that obey Fermi-Dirac statistics [46, 19, 20]. Indeed, inserting Eq. (2.6) into Eq. (2.5) we get the probability of a configuration for a gas of free particles in the *grandcanonical ensemble*[3], where the pair i, j labels an energy state, the weight w_{ij} is the number of particles in that state (occupation number), and the sum $\beta_i + \beta_j$ can be interpreted as

$$\beta_i + \beta_j = \frac{\epsilon_{ij} - \mu(T)}{kT}. \tag{2.39}$$

In the latter expression, ϵ_{ij} represents the energy of the state, $1/kT$ is the inverse temperature, and $\mu(T)$ is the *chemical potential* (required to fix the same expected overall number of particles for all values of T) [20]. Indeed, as we discussed in Sec. 2.2.2, in

[3]In the grandcanonical ensemble, both the total energy and the total number of particles are treated as soft constraints. With respect to the canonical ensemble, the appearance of the number of particles as an additional soft constraint requires the introduction of an extra Lagrange multiplier, the chemical potential.Interestingly, in the context of BEC a fourth (so-called 'Maxwell's Demon') ensemble has also been introduced where the total number of particles is soft while the total energy is hard [54].

our setting the energy and temperature (and in this case, the chemical potential as well) are all reabsorbed into $\vec{\beta}$. Therefore we can interpret the link weight w_{ij} as the number of 'elementary particles' of weight, i.e. the number of quanta of unit weight, populating the link between nodes i and j [46, 19]. The total number of such particles in the system is the total weight W of all links in the network:

$$W(\mathbf{W}) = \sum_{i=1}^{n} \sum_{j<i} w_{ij} = \frac{1}{2} \sum_{i=1}^{n} s_i(\mathbf{W}). \tag{2.40}$$

The 'weighted' property $w_{ij} \in \mathbb{N}$, which leads to Eqs. (2.7), (2.8) and (2.9), corresponds to the possibility that the same state (pair of nodes) is occupied by indefinitely many particles (subject to the average number dictated by the chemical protential), which is a property of bosons. By contrast, in binary networks one has to impose $w_{ij} \in (0, 1)$, which is a property of fermions [46]. An extensive treatment of the role of chemical potential and temperature in binary networks can be found in [20]. Here, to properly interpret what the weighted model is doing, we should give a series of clarifications.

First, we should make a clear distrinction beween the n 'units' of our system (i.e. the nodes of the network) and the W 'particles' of weight that, as a formal analogy, can be interpreted as populating the links of the network. The former are the real constituents of our physical system, while the latter are a mathematical abstraction used to represent the nature of the interactions (links) between such constituents. If we imagine doubling the size of our network, we should imagine doubling the number n of nodes: indeed, we can imagine the network 'growing in size' by adding one single node at a time, but we cannot imagine adding one single *pair* of nodes at a time, without actually adding n new pairs. One should also not be tempted to regard node pairs as the fundamental units by the fact that, mathematically, the $n(n-1)/2$ variables $\{w_{ij}\}$ involving different pairs of nodes are independent random variables: actually, this only occurs in the canonical ensemble and would in any case not be true for more general choices of the constraints. Moreover, even the $n(n-1)/2$ independent node pairs in the canonical ensemble cannot be assigned independent values of the parameters, since there are only n parameters corresponding to the Lagrange multipliers attached to each node. Explicit (and strong) consequences of this fact will be illustrated precisely in the context of BEC. Therefore the physical size of our system is n, and this is why in Eq. (2.31) we defined the relative entropy density as the relative entropy divided by n in the first place. How the total weight W varies with the system size n depends on a specific property, i.e. on how we make the entries of $\vec{s}^*$ (and the resulting value of $W^* = \sum_{i=1}^{n} s_i^*/2$) scale with n. For instance, we may choose to be in the sparse regime where $\vec{s}^*$ remains finite as $n \to \infty$, or in the dense regime where $\vec{s}^*$ diverges as $n \to \infty$. As we show below, the latter is the relevant case for BEC to emerge.

Second, we stress that, irrespective of the above, we always consider a hard number n of nodes, and this is why we compare (only) the canonical (soft value of $\vec{s}$) and microcanonical (hard value of $\vec{s}$) ensembles of networks, both for fixed n (which

sets the dimension of $\vec{s}$). We do *not* consider the grandcanonical ensemble *of network configurations* where n is soft. The grandcanonical ensemble introduced in the aforementioned analogy with systems of bosons is a different one; it may be denoted as an ensemble of *weight quanta* in a network with fixed n and originates from the fact that the Hamiltonian in Eq. (2.6), and consequently the total link weight (not n) in Eq. (2.40), is a fluctuating quantity in the canonical ensemble of network configurations. The fluctuations in the (rescaled) energy H (canonical ensemble of network configurations) are seen as fluctuations in the particle number W (grandcanonical ensemble of weight quanta) in the Bose-Einstein analogy. Fluctuations in the particle number have been the subject of many studies in the literature on BEC [54–59]. Note that, in both canonical and microcanonical ensembles, the individual link weights $\{w_{ij}\}$ are fluctuating quantities, despite the fact that the total link weight W^* is constant in the microcanonical ensemble. Therefore the numbers of 'weight particles' of individual links fluctuate in both ensembles.

Third, while we necessarily discuss the (non)equivalence of the canonical and microcanonical ensembles in the thermodynamic limit $n \to \infty$, the total weight W^* can (and, across the canonical ensemble, will in any case) be arbitrarily large even for finite n. Indeed, the phase transition that we are about to discuss (namely, BEC) does not *per se* require the limit $n \to \infty$, while it definitely requires the limit $W^* \to \infty$. Abstractly, these two limits (and the associated phenomena of EN and BEC respectively) may appear as mathematically unrelated. However, in practice they are physically related once the scaling of $\vec{s}^*$ with n is specified. In particular, we are going to show that, in order to observe BEC, we need be in a dense regime where $W^* = O(n^2)$. This ensures that, when taking the thermodynamic limit $n \to \infty$ in order to study ensemble (non)equivalence, we are automatically implying $W^* \to \infty$ so that we can check for BEC at the same time.

Last, we recall that $\vec{\beta} \cdot \vec{s}(\mathbf{W})$ has to be dimensionless in order to ensure that the probability is a number. Therefore, since w_{ij} is dimensionless, so are s_i^* and β_i^*. In turn, this implies that both sides of Eq. (2.39) must be dimensionless. On the other hand, when modelling a real system, the 'energy' ϵ_{ij} may represent any physical 'cost' associated to the link between nodes i and j (more precisely, it represents the cost of reinforcing w_{ij} by a unit of weight) and may therefore carry its own unit of measure (e.g. it may depend on some distance between nodes i and j). Necessarily, the chemical potential $\mu(T)$ carries the same units as the energy. As for the 'temperature' T, it may be chosen to be dimensionless as it merely represents a control parameter (this is the choice that we will make later); alternatively, it may carry the same units of the energy if it is useful that temperature and energy live on the same scale. Irrespective of this choice, in our setting the 'Boltzmann constant' k is simply a constant that takes care of all dimensional units of measure and makes the ratio on the right hand side of Eq. (2.39) dimensionless.

2.4.2 Core-periphery networks

With the above clarifications, we can finally go back to our model. In the traditional physical situation, in the canonical ensemble the energy ϵ_{ij} of each state i, j is a constant and the temperature T can be varied. Clearly, ϵ_{ij} is independent of T, while the chemical potential $\mu(T)$ is chosen, as a function of temperature, in order to realize the correct (T-independent) expected total number $\langle W \rangle^* \equiv W^*$ of particles for all values of T. In this 'direct problem', every state will therefore have an expected occupation number governed by ϵ_{ij}, T and $\mu(T)$. In our 'inverse' setting, T^* and μ^* are instead reabsorbed into $\vec{\beta}^*$, which in turn is induced by the chosen value of the strength sequence (rather than the other way around). We should therefore regard $\vec{s}^* = \vec{s}^*(T)$ and $\vec{\beta}^* = \vec{\beta}^*(T)$ as T-dependent, while W^* remains T-independent. This means that the chemical potential $\mu^*(T)$ should be such that

$$\sum_{i=1}^{n} s_i^*(T) = 2W^* \quad \forall T \geq 0. \tag{2.41}$$

BEC emerges when, below a certain critical temperature T_c, the occupation number of the state with minimum energy $\epsilon_{\min} = \min_{i,j}\{\epsilon_{ij}\}$ (ground state), or of a finite number of states with lowest energy, becomes so large that it reaches a finite fraction of the total number W^* of particles. Clearly, this requires the existence of at least two different energy levels (the ground state and at least one 'excited' state). Therefore the simplest way to obtain BEC in our model is by considering a strength sequence of the form

$$s_i^*(T) = \begin{cases} s_+^*(T) & i = 1, n_+ \\ s_-^*(T) & i = n_+ + 1, n \end{cases} \qquad s_+^*(T) \geq s_-^*(T), \tag{2.42}$$

i.e. by partitioning the n nodes into two classes, which we call *core* and *periphery*: the core has a finite number

$$n_+ = O(1) \tag{2.43}$$

of nodes, each having a 'large' strength $s_+^*(T)$, while the periphery has an extensive number

$$n_- = n - n_+ = O(n) \tag{2.44}$$

of nodes, each having a 'small' strength $s_-^*(T)$. What we mean precisely by 'small' and 'large' will be clarified below. For the moment, we notice that the BEC phase ($T < T_c$) corresponds to picking a 'condensed' value of $\vec{s}^*(T < T_c)$ such that, in the thermodynamic limit, the core takes up a finite fraction of the total weight W^* of all links in the network, despite having a finite size. In particular, in the zero-temperature limit *all* the total weight W^* is in the core. By contrast, the non-condensed phase $T > T_c$ is one where $\vec{s}^*(T > T_c)$ is such that no individual link receives a finite fraction of W^*. In particular, the infinite-temperature limit should be such that the energy difference between ground and excited states becomes ineffective, i.e.

$s^*_+(T\to\infty) = s^*_-(T\to\infty)$. The different phases can be efficiently monitored by introducing a temperature-dependent order parameter $Q^*(T)$, as we show below.

We stress that, since we are ultimately interested in the relative fluctuations of the canonical constraints and in the relative entropy that can be asymptotically calculated purely from canonical quantities according to Eq. (2.33), practically we only need to study the canonical ensemble. The only check we need to make is that, whenever we speak of the system being in a certain 'phase', this statement does not depend on the particular ensemble. In other words, we need to show that the order parameter has always the same value in the canonical and microcanonical ensembles.

Before studying the individual phases, let us make some general considerations, valid for all values of T. We first find the value of $\vec{\beta}^*(T)$ corresponding to the value of $\vec{s}^*(T)$ in Eq. (2.42). As we mentioned, $s^*_i(T) = s^*_j(T)$ implies $\beta^*_i(T) = \beta^*_j(T)$, therefore the entries of $\vec{\beta}^*(T)$ take only two values $\beta^*_+(T)$ and $\beta^*_-(T)$ such that

$$\beta^*_i(T) = \begin{cases} \beta^*_+(T) & i = 1, n_+ \\ \beta^*_-(T) & i = n_+ + 1, n \end{cases} \qquad \beta^*_+(T) \leq \beta^*_-(T). \tag{2.45}$$

These values solve the n equations in (2.13), which here reduce to the two independent equations

$$(n_+ - 1)w^*_+(T) + n_- w^*_0(T) \equiv s^*_+(T) \tag{2.46}$$
$$(n_- - 1)w^*_-(T) + n_+ w^*_0(T) \equiv s^*_-(T) \tag{2.47}$$

where we have defined

$$w^*_+(T) = \frac{e^{-2\beta^*_+(T)}}{1 - e^{-2\beta^*_+(T)}} \tag{2.48}$$

as the expected link weight $\langle w_{ij}\rangle_{\vec{\beta}^*(T)}$ between any two nodes in the core ($i, j = 1, n_+$),

$$w^*_-(T) = \frac{e^{-2\beta^*_-(T)}}{1 - e^{-2\beta^*_-(T)}} \tag{2.49}$$

as the expected link weight $\langle w_{ij}\rangle_{\vec{\beta}^*(T)}$ between any two nodes in the periphery ($i, j = n_+ + 1, n$), and

$$\begin{aligned} w^*_0(T) &= \frac{e^{-\beta^*_-(T)-\beta^*_+(T)}}{1 - e^{-\beta^*_-(T)-\beta^*_+(T)}} \\ &= \frac{\sqrt{w^*_+(T)w^*_-(T)}}{\sqrt{1 + w^*_+(T)}\sqrt{1 + w^*_-(T)} - \sqrt{w^*_+(T)w^*_-(T)}} \\ &= \frac{1}{\sqrt{1 + 1/w^*_+(T)}\sqrt{1 + 1/w^*_-(T)} - 1} \end{aligned} \tag{2.50}$$

as the expected link weight $\langle w_{ij} \rangle_{\vec{\beta}^*(T)}$ between any node in the core and any node in the periphery ($i = 1, n_+$ and $j = n_+ + 1, n$ or $j = 1, n_+$ and $i = n_+ + 1, n$). Note that $\beta_+^*(T) \leq \beta_-^*(T)$ implies $w_-^*(T) \leq w_0^*(T) \leq w_+^*(T)$.

Now, solving Eqs. (2.46) and (2.47), we obtain the explicit values of $\beta_+^*(T)$ and $\beta_-^*(T)$ appearing in Eq. (2.45):

$$\beta_\pm^*(T) = \frac{1}{2} \ln \frac{1 + w_\pm^*(T)}{w_\pm^*(T)} = \frac{1}{2} \ln \left(1 + \frac{1}{w_\pm^*(T)} \right). \tag{2.51}$$

Also note from Eq. (2.50) that

$$\beta_+^*(T) + \beta_-^*(T) = \ln \left(1 + \frac{1}{w_0^*(T)} \right). \tag{2.52}$$

Also, using Eq. (2.16) we can define

$$p_\pm^*(T) \equiv e^{-2\beta_\pm^*(T)} = \frac{w_\pm^*(T)}{1 + w_\pm^*(T)} \tag{2.53}$$

as the probability that a link exists (irrespective of its weight) between any two core-core (+) or any two periphery-periphery (−) nodes, and

$$p_0^*(T) \equiv e^{-\beta_+^*(T) - \beta_-^*(T)} = \frac{w_0^*(T)}{1 + w_0^*(T)} \tag{2.54}$$

as the probability that a link exists (irrespective of its weight) between a core node and a periphery node.

From Eq. (2.39), we notice that the existence of the two values above for the entries of $\vec{\beta}^*(T)$ implies that there are three energy levels, associated with the energies

$$\epsilon_+^* = \mu^*(T) + 2kT\beta_+^*(T), \tag{2.55}$$

$$\epsilon_-^* = \mu^*(T) + 2kT\beta_-^*(T), \tag{2.56}$$

$$\epsilon_0^* = \mu^*(T) + kT[\beta_+^*(T) + \beta_-^*(T)] = \frac{\epsilon_+^* + \epsilon_-^*}{2}, \tag{2.57}$$

where $\epsilon_+^* \leq \epsilon_0^* \leq \epsilon_-^*$ (we recall that all energy values are finite and independent of both T and n). The appearance of *three* distinct energy levels out of just *two* values of the fundamental Lagrange multipliers confirms the interpretation that the true units of the system are the nodes and not the node pairs: it would indeed be impossible for our system to exhibit exactly two energy states, or in general to engineer an arbitrary number of energy states for the node pairs, since the only arbitrary values are those that can be attached to nodes, not to node pairs. Also note that all the three levels above are degenerate: the $n_+(n_+ - 1)/2$ pairs of nodes in the core have the same expected link weight $w_+^*(T)$ and energy ϵ_+^*, the $n_-(n_- - 1)/2$ pairs of nodes in the periphery have the same expected link weight $w_-^*(T)$ and energy ϵ_-^*, and the $n_+ n_-$

pairs of nodes across core and periphery have the same expected link weight $w_0^*(T)$ and energy ϵ_0^*. Therefore the ground state has energy $\epsilon_{\min}^* = \epsilon_+^*$ and degeneracy $n_+(n_+-1)/2$. These degeneracies are dictated by the numbers of nodes in the two sets and cannot be assigned arbitrarily.

The occupation number of the ground state (with energy ϵ_+^*) coincides with the expected weight of all links between core nodes (total 'core-core' weight):

$$W_+^*(T) = \frac{n_+(n_+-1)}{2} w_+^*(T). \tag{2.58}$$

Similarly, the occupation number of the first excited state (with energy ϵ_0) coincides with the expected weight of all links between nodes across core and periphery (total 'core-periphery' weight):

$$W_0^*(T) = n_+ n_- w_0^*(T). \tag{2.59}$$

Finally, the occupation number of the second excited state (with energy ϵ_-) coincides with the expected weight of all links between periphery nodes (total 'periphery-periphery' weight):

$$W_-^*(T) = \frac{n_-(n_--1)}{2} w_-^*(T). \tag{2.60}$$

By writing W^* as the sum of its core-core, core-periphery and periphery-periphery components, we get

$$\begin{aligned} W^* &= W_+^*(T) + W_0^*(T) + W_-^*(T) \\ &= \frac{n_+(n_+-1)}{2} w_+^*(T) + n_+ n_- w_0^*(T) \\ &\quad + \frac{n_-(n_--1)}{2} w_-^*(T). \end{aligned} \tag{2.61}$$

Using Eq. (2.41), the total weight can also be expressed as

$$W^* = \frac{n_+ s_+^*(T) + n_- s_-^*(T)}{2} \tag{2.62}$$

which, through Eqs. (2.46) and (2.47), indeed reduces to Eq. (2.61).

We stress again that the chemical potential $\mu^*(T)$ appearing in Eqs. (2.55), (2.56) and (2.57) plays the role of a global Lagrange multiplier ensuring that, for all values of T, the total expected weight is W^*. Note that the T-independence of W^* allows us to conclude immediately that its value should be of order

$$W^* = O(n^2) \tag{2.63}$$

because, in particular, in the non-condensed phase all the $n(n-1)/2$ individual link weights $w_\pm^*$, w_0^* must be finite by definition. As we have anticipated, this result ensures that in the thermodynamic limit ($n \to \infty$) we automatically have $W^* \to \infty$, so that we can study ensemble (non)equivalence and BEC simultaneously, thereby

'physically' connecting two otherwise mathematically unrelated limits. We also note that, irrespective of temperature, the network is always in the dense regime. We can therefore introduce the average expected link weight

$$w^* \equiv \frac{2W^*}{n(n-1)} = O(1), \tag{2.64}$$

which is a T-independent, finite parameter of our model, controlling the overall link weight in the network. Clearly,

$$w^*_-(T) \leq w^* \leq w^*_+(T) \quad \forall T. \tag{2.65}$$

It is good to remark again that, in our 'inverse' problem (construction of the conjugate canonical and microcanonical ensembles), the parameters of the model are the values of the constraints, which here reduce to the two (diverging when $n \to \infty$) numbers $s^*_\pm(T)$. However, to allow consistent comparisons for different temperatures, not all strength sequences are allowed, but only those that can be obtained from one another by varying T. In particular the values $s^*_\pm(T)$ have to be specified for each value of T and be such that the total weight is always W^*. Indeed Eq. (2.62) shows that only two of the three quantities $s^*_\pm(T)$, W^* are independent. By contrast, the traditional 'direct' problem in physics sees the three energies $\epsilon^*_\pm$ and ϵ^*_0 (which do not depend on T) as the parameters of the model, plus either w^* or the chemical potential $\mu^*(T)$. However, Eq. (2.57) shows that $\epsilon^*_0 = (\epsilon^*_+ + \epsilon^*_-)/2$, indicating only two independent values of the energy (say $\epsilon^*_\pm$), as a result of the fact that there are only two types of nodes. Moreover, we may set the minimum energy $\epsilon^*_+ \equiv 0$ without loss of generality, because any overall energy shift can be reabsorbed into the chemical potential. We can therefore rename the only remaining independent value of the energy as $\epsilon^*_- \equiv \epsilon^* > 0$, and similarly $\epsilon^*_0 = \epsilon^*/2 > 0$. Using these replacements into Eqs. (2.55) and (2.56), and combining the two equations, we get

$$\mu^*(T) = -2kT\beta^*_+(T) = \epsilon^* - 2kT\beta^*_-(T) \tag{2.66}$$

which is a convenient expression for solving the 'direct' problem. Rearranging, we obtain

$$\epsilon^* = 2kT\left[\beta^*_-(T) - \beta^*_+(T)\right]. \tag{2.67}$$

and, using Eqs. (2.51) and (2.53),

$$\begin{aligned} e^{-\epsilon^*/kT} &= e^{2[\beta^*_+(T) - \beta^*_-(T)]} \\ &= \frac{p^*_-(T)}{p^*_+(T)} \\ &= \frac{w^*_-(T)[1 + w^*_+(T)]}{w^*_+(T)[1 + w^*_-(T)]}, \end{aligned} \tag{2.68}$$

which shows how the energy difference ϵ^* between periphery-periphery (+) and core-core (−) states is related to the corresponding connection probabilities $p^*_\pm(T)$ and

expected link weights $w^*_\pm(T)$. Therefore the most compact way of parametrizing the direct problem is by specifying only the two finite, positive and T-independent numbers ϵ^* and w^*, and explore the resulting network properties by finding $\mu(T)$ (as a function of ϵ^*, w^* and T) and varying T as a control parameter. This will indeed allow us to easily explore the different (high- and low-temperature) phases consistently.

In our model, BEC occurs below a critical temperature T_c such that a finite fraction of the total weight W^* condenses in the core, which remains of finite size (i.e. of a finite number n_+ of nodes) even when the size of the whole network diverges. This corresponds to requiring that, as $n \to \infty$, n_+ remains finite as dictated by Eq. (2.43), W^* diverges, and $W^*_+(T)$ takes up a finite fraction of W^*. Rigorously, we can define this fraction (for finite n) as

$$Q^*_n(T) \equiv \frac{W^*_+(T)}{W^*} \tag{2.69}$$

and use it to introduce the *order parameter* as

$$Q^*(T) \equiv \lim_{n\to\infty} Q^*_n(T) = \lim_{n\to\infty} \frac{W^*_+(T)}{W^*}. \tag{2.70}$$

We can then define the BEC phase as a phase emerging below a certain critical temperature T_c such that

$$Q^*(T<T_c) > 0. \tag{2.71}$$

By contrast, the non-BEC phase is such that

$$Q^*(T>T_c) = 0. \tag{2.72}$$

A visual anticipation of the qualitative behaviour that our system will exhibit is provided in Fig. 2.3.

In conjunction with BEC, we will investigate ensemble (non)equivalence. Therefore, in each phase of the model we will consider the relative entropy between the microcanonical and canonical ensembles and the relative fluctuations of the constraints. The criterion for measure equivalence is based on the relative entropy in Eq. (2.33), and useful techniques for the calculation of the determinant of the covariance matrix $\mathbf{\Sigma}^*$ in each phase are provided in the Appendix. Clearly, from Eqs. (2.23) and (2.45) we see that the diagonal entries Σ^*_{ii} of $\mathbf{\Sigma}^*$ take two possible values:

$$\Sigma^*_{ii}(T) = \sigma^2_{\vec{\beta}^*(T)}(s_i) = \begin{cases} \Sigma^*_+(T) & i = 1, n_+ \\ \Sigma^*_-(T) & i = n_+ + 1, n \end{cases} \tag{2.73}$$

where

$$\begin{aligned} \Sigma^*_\pm(T) &= \frac{(n_\pm - 1)e^{-2\beta^*_\pm(T)}}{\left[1 - e^{-2\beta^*_\pm(T)}\right]^2} + \frac{n_\mp e^{-\beta^*_+(T)-\beta^*_-(T)}}{\left[1 - e^{-\beta^*_+(T)-\beta^*_-(T)}\right]^2} \\ &= (n_\pm - 1)w^*_\pm(T)\left[1 + w^*_\pm(T)\right] \\ &\quad + n_\mp w^*_0(T)\left[1 + w^*_0(T)\right]. \end{aligned} \tag{2.74}$$

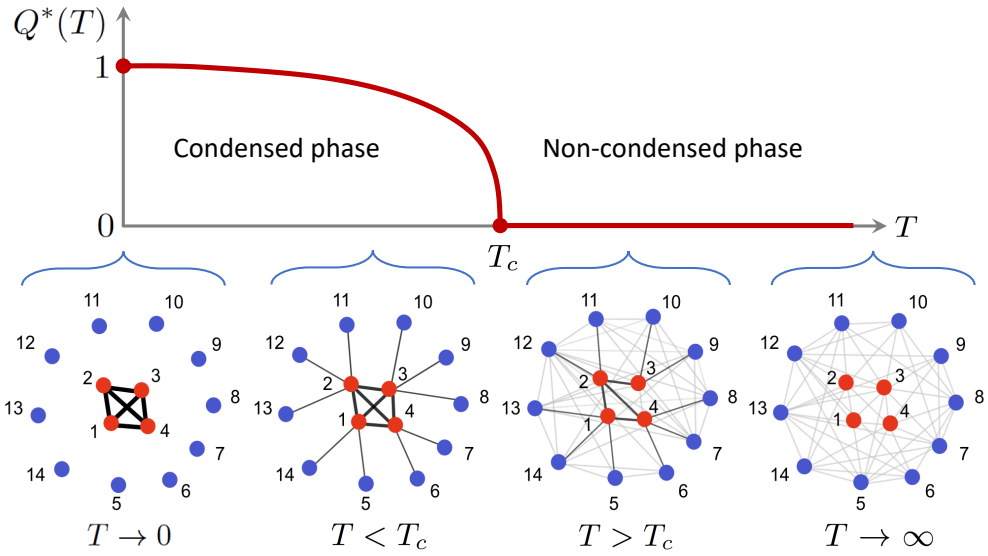

Figure 2.3. Illustration of possible realizations of a network as a function of temperature, from higher (right) to lower (left) values of T. As a schematic example, a network with $n = 14$ nodes, of which n_+ are core nodes and $n_- = 10$ are peripheral nodes, is considered. The order parameter $Q^*(T)$ is zero for temperatures above the critical temperature T_c, while it is positive for temperatures below T_c, increasing towards 1 at zero temperature. At infinite temperature ($T \to \infty$), there is no distinction between core and periphery: all links have the same probability of existing and the same expected weight. At lower but supercritical temperature ($T > T_c$), a quantitative (but not yet qualitative) distinction between core and periphery appears: core-core links have higher probability and expected weight than core-periphery links, which in turn have higher probability and expected weight than periphery-periphery links, however all these probabilities and expected link weights are of the same (finite) order $O(1)$. Below a certain critical temperature ($T < T_c$), the distinction between core and periphery becomes qualitative and more dramatic (the core forms a 'condensate'): the expected link weights are of order $O(n^2)$ for core-core links (with the corresponding connection probabilities approaching one) and $O(1)$ for core-periphery and periphery-periphery links. Finally, at zero temperature ($T \to 0$) the condensate decouples from the rest: peripheral nodes are completely disconnected and all links end up in the core, with an expected weight still of order $O(n^2)$.

Recalling Eq. (2.7), we remark that the canonical entropy $S^*_{\rm can}(T)$ can be easily calculated from Eq. (2.17) as the following sum of five terms:

$$\begin{aligned} S^*_{\rm can}(T) &= \vec{\beta}^*(T)\cdot\vec{s}^*(T) + \ln Z[\vec{\beta}^*(T)] \\ &= n_+\beta^*_+(T)s^*_+(T) \\ &\quad +n_-\beta^*_-(T)s^*_-(T) \\ &\quad +\frac{n_+(n_+-1)}{2}\ln\frac{1}{1-e^{-2\beta^*_+(T)}} \\ &\quad +\frac{n_-(n_--1)}{2}\ln\frac{1}{1-e^{-2\beta^*_-(T)}} \\ &\quad +n_+n_-\ln\frac{1}{1-e^{-\beta^*_+(T)-\beta^*_-(T)}}, \end{aligned} \tag{2.75}$$

while the microcanonical entropy $S^*_{\rm mic}(T)$ is in general hard to compute, as it requires an explicit enumeration. However, its leading order can be obtained combining Eqs. (2.25) and (2.75).

The relative fluctuations of the constraints take the form

$$r^*_i(T) = \begin{cases} r^*_+(T) & i = 1, n_+ \\ r^*_-(T) & i = n_+ + 1, n \end{cases} \tag{2.76}$$

where, from Eq. (2.37),

$$\begin{aligned} r^*_\pm(T) &= \frac{\sqrt{\Sigma^*_\pm(T)}}{s^*_\pm(T)} \\ &= \sqrt{\frac{1}{s^*_\pm(T)} + \frac{(n_\pm-1)[w^*_\pm(T)]^2 + n_\mp[w^*_0(T)]^2}{[s^*_\pm(T)]^2}} \\ &= \sqrt{\frac{1}{s^*_\pm(T)} + \frac{(n_\pm-1)[w^*_\pm(T)]^2 + n_\mp[w^*_0(T)]^2}{[(n_\pm-1)w^*_\pm(T) + n_\mp w^*_0(T)]^2}}. \end{aligned} \tag{2.77}$$

Therefore in the thermodynamic limit the relative fluctuations of the constraints, as defined in Eq. (2.38), take only the two possible limiting values

$$\rho^*_i(T) = \begin{cases} \rho^*_+(T) & i = 1, n_+ \\ \rho^*_-(T) & i = n_+ + 1, n \end{cases} \tag{2.78}$$

where

$$\rho^*_\pm(T) = \lim_{n\to\infty} r^*_\pm(T). \tag{2.79}$$

Armed with the above general results, we can now study each phase in detail.

2.4.3 Non-condensed phase

Let us start from the non-BEC phase ($T > T_c$). We first consider the finite-temperature case ($T_c < T < \infty$) and then the infinite-temperature limit ($T \to \infty$). As all the interesting phenomenology (in terms of both BEC and EN) occurs in the thermodynamic limit $n \to \infty$, we look for the asymptotic behaviour of all quantities in that limit.

Finite (supercritical) temperature: $T_c < T < \infty$

Since, by definition, when $T > T_c$ there is no concentration of 'particles' of weight on any of the links, all the expected link weights must be separately finite, i.e. $w_+^*(T > T_c)$, $w_-^*(T > T_c)$ and $w_0^*(T > T_c)$ are all $O(1)$. Consequently, from Eqs. (2.53) and (2.54) we see that all the connection probabilities $p_+^*(T > T_c)$, $p_-^*(T > T_c)$ and $p_0^*(T > T_c)$ are strictly smaller than one, i.e. missing links can occur anywhere in the network. Using this fact into Eqs. (2.46), (2.47), (2.58), (2.59) and (2.60), and using Eqs. (2.43) and (2.44), we immediately get the strength of nodes in the core, i.e.

$$\begin{aligned} s_+^*(T > T_c) &= (n_+ - 1) w_+^*(T > T_c) + n_- w_0^*(T > T_c) \\ &\approx n\, w_0^*(T > T_c) \\ &= O(n), \end{aligned} \tag{2.80}$$

and in the periphery, i.e.

$$\begin{aligned} s_-^*(T > T_c) &= (n_- - 1) w_-^*(T > T_c) + n_+ w_0^*(T > T_c) \\ &\approx n\, w_-^*(T > T_c) \\ &= O(n). \end{aligned} \tag{2.81}$$

Similarly, for $W_\pm^*(T > T_c)$, $W_0^*(T > T_c)$ we get

$$W_+^*(T > T_c) = n_+(n_+ - 1) w_+^*(T > T_c)/2 = O(1), \tag{2.82}$$

$$W_-^*(T > T_c) \approx n^2 w_-^*(T > T_c)/2 = O(n^2), \tag{2.83}$$

$$W_0^*(T > T_c) \approx n_+\, n\, w_0^*(T > T_c) = O(n), \tag{2.84}$$

from which we see that in this phase the total weight W^* is essentially all in the periphery, i.e.

$$W_-^*(T > T_c) = W^* - o(n^2) \approx W^*, \tag{2.85}$$

$$w_-^*(T > T_c) = w^* - o(1) \approx w^*. \tag{2.86}$$

We stress that the above result does not mean that the core is empty or that there are no connections between core and periphery. Rather, it indicates that the total weight of all core-core and core-periphery connections is asymptotically negligible with respect to the total weight located inside the periphery, simply because the number of periphery-periphery node pairs dominates the number of core-periphery and core-core

pairs. In particular, the finite parameter $w_+^*(T>T_c)$ can take an arbitrarily large value, without 'moving' the (finite and positive) value of the average link weight w^*. All positive values of $w_+^*(T>T_c)$ are therefore allowed. By contrast, $w_-^*(T>T_c)$ is forced to take (to leading order) only the value w^*.

Using Eqs. (2.81), (2.83) and (2.84), we write the order parameter as

$$\begin{aligned} Q^*(T>T_c) &\equiv \lim_{n\to\infty} \frac{W_+^*(T>T_c)}{W^*} \\ &= \lim_{n\to\infty} \frac{W^* - W_-^*(T>T_c) - W_0^*(T>T_c)}{W^*} \\ &= 1 - \lim_{n\to\infty} \frac{n^2\, w_-^*(T>T_c)}{2W^*} \\ &= 1 - \lim_{n\to\infty} \frac{n\, s_-^*(T>T_c)}{2W^*}, \\ &= 0, \end{aligned} \tag{2.87}$$

confirming the definition of non-condensed phase in Eq. (2.72) and showing that, since both $s_-^*(T>T_c)$ and W^* have by construction the same value in the canonical and microcanonical ensemble, the order parameter must be zero in both ensembles, for all values of $T>T_c$. Therefore, whenever one ensemble is in the non-condensed phase, the other ensemble is the non-condensed phase as well. Importantly, this allows us to refer to the conjugate canonical and microcanonical ensembles 'in the non-condensed phase' consistently.

To solve the 'inverse' problem, we use Eqs. (2.51) and (2.52) and invert Eqs. (2.80) and (2.81) to get

$$\begin{aligned} \beta_-^*(T>T_c) &= \frac{1}{2}\ln\left(1+\frac{1}{w_-^*(T>T_c)}\right) \\ &\approx \frac{1}{2}\ln\left(1+\frac{n}{s_-^*(T>T_c)}\right) \end{aligned} \tag{2.88}$$

and

$$\begin{aligned} \beta_+^*(T>T_c)+\beta_-^*(T>T_c) &= \ln\left(1+\frac{1}{w_0^*(T>T_c)}\right) \\ &\approx \ln\left(1+\frac{n}{s_+^*(T>T_c)}\right). \end{aligned} \tag{2.89}$$

Then, subtracting Eq. (2.88) from Eq. (2.89), we get

$$\begin{aligned} \beta_+^*(T>T_c) &= \frac{1}{2}\ln\left(1+\frac{1}{w_+^*(T>T_c)}\right) \\ &\approx \ln\frac{1+\frac{n}{s_+^*(T>T_c)}}{\sqrt{1+\frac{n}{s_-^*(T>T_c)}}}. \end{aligned} \tag{2.90}$$

Equations (2.88) and (2.90) express $\beta_{\pm}^*(T>T_c)$ as a function of the two (diverging) constraints $s_{\pm}^*(T>T_c)$, or equivalently as a function of the finite parameters $w_{\pm}^*(T>T_c)$, which have to be specified for all values of T.

To solve the 'direct' problem, we first note an important consequence of Eq. (2.86): in the large n limit, $w_-^*(T>T_c)$ and $\beta_-^*(T>T_c)$ are independent of temperature. Indeed, using Eq. (2.81) and (2.86) into Eq. (2.88), we get asympotically

$$\beta_-^*(T>T_c) \approx \frac{1}{2}\ln\left(1+\frac{1}{w^*}\right) \tag{2.91}$$

and, using Eq. (2.67),

$$\beta_+^*(T>T_c) \approx \frac{1}{2}\ln\left(1+\frac{1}{w^*}\right) - \frac{\epsilon^*}{2kT}. \tag{2.92}$$

When inserted into Eq. (2.66), the above expressions allow us to directly obtain the chemical potential as

$$\begin{aligned}\mu(T>T_c) &= \epsilon^* - 2kT\beta_-^*(T>T_c) \\ &\approx \epsilon^* - kT\ln\left(1+\frac{1}{w^*}\right) + o(T).\end{aligned} \tag{2.93}$$

As anticipated, the above result provides the solution to the direct problem in terms of the two finite constants ϵ^* and w^*, and allows us to explore the model by varying T throughout the non-condensed phase $T>T_c$.

We now consider ensemble (non)equivalence. Inserting Eqs. (2.80) and (2.81) into Eq. (2.74), we obtain the variance of the strength of nodes in the core, i.e.

$$\begin{aligned}\Sigma_+^*(T>T_c) &\approx n\,w_0^*(T>T_c)\,[1+w_0^*(T>T_c)] \\ &\approx s_+^*(T>T_c)\left[1+\frac{s_+^*(T>T_c)}{n}\right] \\ &= O(n),\end{aligned} \tag{2.94}$$

and in the periphery, i.e.

$$\begin{aligned}\Sigma_-^*(T>T_c) &\approx n\,w_-^*(T>T_c)\,[1+w_-^*(T>T_c)] \\ &\approx s_-^*(T>T_c)\left[1+\frac{s_-^*(T>T_c)}{n}\right] \\ &= O(n).\end{aligned} \tag{2.95}$$

As we show in Appendix 2.A.1, it is possible to show that the leading term of the determinant of the covariance matrix $\mathbf{\Sigma}^*(T>T_c)$ in this non-condensed phase is

$$\det[\mathbf{\Sigma}^*(T>T_c)] = \prod_{i=1}^{n}\Sigma_{ii}^*(T>T_c) + O(n^{n-2}). \tag{2.96}$$

Using Eqs. (2.74), (2.94) and (2.95) we obtain

$$\begin{aligned}\prod_{i=1}^{n}\Sigma^*_{ii}(T>T_c) &= [\Sigma^*_+(T>T_c)]^{n_+}[\Sigma^*_-(T>T_c)]^{n_-}\\ &= O(n^n).\end{aligned} \tag{2.97}$$

Inserting this result into Eq. (2.96), we find

$$\det[\mathbf{\Sigma}^*(T>T_c)] = O(n^n) + O(n^{n-2}) = O(n^n), \tag{2.98}$$

showing that the determinant is dominated by the diagonal entries of $\mathbf{\Sigma}^*(T>T_c)$. Taking the logarithm, we obtain

$$\ln\det\mathbf{\Sigma}^*(T>T_c) = O(n\ln n) \tag{2.99}$$

which, when compared with Eq. (2.34), shows that the system is under ensemble nonequivalence. We note that the $O(n\ln n)$ scaling of $\ln\det\mathbf{\Sigma}^*(T>T_c)$ ensures that Eq. (2.35) holds, so the leading order of the relative entropy can be calculated exactly as

$$D^*(T>T_c) \approx \frac{1}{2}\ln\det\mathbf{\Sigma}^*(T>T_c) \approx \frac{1}{2}n\ln n \tag{2.100}$$

where we have used Eqs. (2.94) and (2.95) into Eq. (2.96). The above result is in line with the scaling of the relative entropy found in the case of binary networks with a constraint on the node degrees in the dense regime [5, 26, 45]. Another similarity between the two models is the order of the canonical entropy:

$$S^*_{\rm can}(T>T_c) = O(n^2), \tag{2.101}$$

which can be easily seen from Eq. (2.75) using $\beta^*_\pm(T>T_c) = O(1)$ and $s^*_\pm(T>T_c) = O(n)$, as found in Eqs. (2.80), (2.81), (2.88) and (2.90). Then Eq. (2.25) also implies

$$S^*_{\rm mic}(T>T_c) = O(n^2). \tag{2.102}$$

Note that, even if the relative entropy is subleading with respect to the canonical and microcanonical entropies, it is still superextensive in the number n of units of the system, thereby breaking ensemble equivalence as in binary networks with fixed degrees. Therefore the result in Eq. (2.99) is another observation, for the first time in weighted networks, of the fact that ensemble equivalence can be broken by the presence of an extensive number of local constraints, even away from phase transitions.

Coming to the relative fluctuations of the constraints, we see from Eqs. (2.77) and (2.79) that

$$\begin{aligned}r^*_\pm(T>T_c) &= \frac{\sqrt{\Sigma^*_\pm(T>T_c)}}{s^*_\pm(T>T_c)}\\ &\approx \sqrt{\frac{n+s^*_\pm(T>T_c)}{n\,s^*_\pm(T>T_c)}}\\ &= O\left(\frac{1}{\sqrt{n}}\right)\end{aligned} \tag{2.103}$$

and

$$\rho_{\pm}^{*}(T>T_c) = 0. \tag{2.104}$$

We therefore observe that in the non-condensed phase the decay of the relative fluctuations of each constraint is of the same order $O(1/\sqrt{n})$ as generally observed for the global constraint (total energy) in a system with short-range interactions away from phase transitions. However, while in the traditional situation the vanishing of the relative fluctuations implies that the relative entropy is subextensive and that the relative entropy density vanishes in the thermodynamic limit (as discussed in Sec. 2.1.3), here the relative entropy density does not vanish and the ensembles are not equivalent. Therefore we find that, *in systems with an extensive number of local constraints, the vanishing of even all the relative fluctuations does not ensure ensemble equivalence.*

Infinite temperature: $T \to \infty$

The extreme regime of the non-condensed phase is the infinite-temperature case, which can be explored by taking the limit $T \to \infty$ in the solution to the 'direct' problem provided by Eq. (2.93). In such a limit, Eq. (2.39) implies that $\beta_{+}^{*}(T>T_c)$ and $\beta_{-}^{*}(T>T_c)$ converge to the same value β_{∞}^{*} given by

$$\beta_{+}^{*}(T\to\infty) = \beta_{-}^{*}(T\to\infty) = \beta_{\infty}^{*} \equiv \frac{1}{2}\ln\left(1+\frac{1}{w^*}\right). \tag{2.105}$$

Then, through Eqs. (2.48), (2.49), (2.50), (2.53) and (2.54), we get

$$w_{+}^{*}(T\to\infty) = w_{-}^{*}(T\to\infty) = w_{0}^{*}(T\to\infty) = w^*, \tag{2.106}$$

$$p_{+}^{*}(T\to\infty) = p_{-}^{*}(T\to\infty) = p_{0}^{*}(T\to\infty) = p^*, \tag{2.107}$$

i.e. all node pairs have the same expected link weight w^* and connection probability p^* given by

$$p^* = \frac{w^*}{1+w^*}. \tag{2.108}$$

This is the characteristic situation in the infinite-temperature limit of Bose-Einstein statistics, where each particle is equally likely distributed across all energy levels. Here, this situation translates in the graph becoming completely homogeneous: the distinction between core and periphery disappears as the finite difference between energy levels becomes entirely dominated by the diverging temperature. The expected strength of every node has the same value $s^* \equiv (n-1)w^*$:

$$s_{+}^{*}(T\to\infty) = s_{-}^{*}(T\to\infty) = s^* = (n-1)w^*, \tag{2.109}$$

i.e. the strength sequence becomes a constant vector.

Clearly, the above result does not change the value of the order parameter in Eq. (2.87):

$$Q^*(T\to\infty) = 0. \tag{2.110}$$

Similarly, the final results in eqs. (2.99) and (2.104) about the simultaneous breakdown of ensemble equivalence and the vanishing of all relative fluctuations carry over to the infinite-temperature limit, so in principle we do not have to further discuss this case. However, the fact that the strength sequence becomes a constant vector allows us to calculate many of the properties of the model exactly, so this example is a very transparent and instructive one. It is therefore worth considering it in some more detail, also because some of the following results will be useful in the (much less trivial) zero-temperature limit as well.

In particular, Eqs. (2.105) and (2.109) imply that Eqs. (2.23) and (2.24) can be rewritten as

$$\Sigma^*_{ii}(T\to\infty) = \frac{(n-1)e^{-2\beta^*_\infty}}{(1-e^{-2\beta^*_\infty})^2} = w^*(1+w^*)(n-1) \tag{2.111}$$

for all $i = 1, n$ and

$$\Sigma^*_{ij}(T\to\infty) = \frac{e^{-2\beta^*_\infty}}{(1-e^{-2\beta^*_\infty})^2} = w^*(1+w^*) \tag{2.112}$$

for all $i \neq j$ respectively. In Appendix 2.A.2 we show that the above expressions can be used to calculate the determinant of $\mathbf{\Sigma}^*(T\to\infty)$ exactly as

$$\det \mathbf{\Sigma}^*(T\to\infty) = 2(n-1)(n-2)^{n-1}\left[w^*(1+w^*)\right]^n, \tag{2.113}$$

from which we confirm, without having made the approximation in eq. (2.96), that

$$\ln\det \mathbf{\Sigma}^*(T\to\infty) = O(n\ln n) \tag{2.114}$$

and that

$$D^*(T\to\infty) \approx \frac{1}{2}\ln\det \mathbf{\Sigma}^*(T\to\infty) \approx \frac{1}{2}n\ln n. \tag{2.115}$$

Clearly, Eqs. (2.101) and (2.102) hold in this limit as well:

$$S^*_{\rm can}(T\to\infty) = O(n^2),\ S^*_{\rm mic}(T\to\infty) = O(n^2). \tag{2.116}$$

Finally, from eqs. (2.109) and (2.111) we see that eq. (2.77) leads in this case to a unique value for the coefficient of variation of all the strengths:

$$r^*_i(T\to\infty) = \frac{\sqrt{\Sigma^*_{ii}(T\to\infty)}}{s^*_i(T\to\infty)} = \sqrt{\frac{1+w^*}{(n-1)w^*}} \quad \forall i, \tag{2.117}$$

so that

$$\rho^*_i(T\to\infty) = 0 \quad \forall i, \tag{2.118}$$

in accordance with eq. (2.104).

2.4.4 Condensed phase

We now consider the BEC phase ($T<T_c$). We first derive general results and then discuss the finite-temperature case and the zero-temperature limit separately.

By the definition in Eq. (2.71), the condensed phase must be such that a positive fraction $Q^*(T<T_c)>0$ of the total weight lies in the core, i.e. (to leading order)

$$W_+^*(T<T_c) \approx Q_n^*(T<T_c)W^* \approx Q^*(T<T_c)W^* \tag{2.119}$$

which necessarily means

$$w_+^*(T<T_c) = O(n^2), \quad W_+^*(T<T_c) = O(n^2) \tag{2.120}$$

and $p_+^*(T<T_c) \approx 1$, i.e. the core does not have missing links (the presence of core-core links is no longer a random event, while the weight of such links is still a random variable). As expected, BEC corresponds to the divergence of $w_+^*(T<T_c)$, and we now see that the speed of this divergence is of order n^2 in our model. For convenience, we may define

$$\psi_+^*(T<T_c) = \lim_{n\to\infty} \frac{w_+^*(T<T_c)}{n^2} \tag{2.121}$$

which is finite and positive, so that

$$w_+^*(T<T_c) \approx \psi_+^*(T<T_c)n^2. \tag{2.122}$$

Combining Eqs. (2.50) and (2.120) we see that

$$w_0^*(T<T_c) \approx \frac{1}{\sqrt{1+1/w_-^*(T<T_c)}-1}, \tag{2.123}$$

which inserted into Eq. (2.61) shows that, to leading order,

$$w^* \approx n_+(n_+-1)\psi_+^*(T<T_c) + w_-^*(T<T_c), \tag{2.124}$$

an expression that relates the finite parameters of the model with each other in the condensed phase. Therefore we see from Eq. (2.119) that

$$w_-^*(T<T_c) \approx [1-Q_n^*(T<T_c)]w^* \tag{2.125}$$

and

$$W_-^*(T<T_c) \approx [1-Q_n^*(T<T_c)]W^*. \tag{2.126}$$

Inserting Eq. (2.125) into Eq. (2.123) yields

$$w_0^*(T<T_c) \approx \frac{1}{\sqrt{1+\frac{1}{[1-Q_n^*(T<T_c)]w^*}}-1} \tag{2.127}$$

and

$$W_0^*(T<T_c) \approx \frac{n_+ n_-}{\sqrt{1+\frac{1}{[1-Q_n^*(T<T_c)]w^*}}-1}. \tag{2.128}$$

The above expressions show that neither $w_-^*(T<T_c)$ nor $w_0^*(T<T_c)$ diverge, indicating that BEC occurs only in the ground state and that $p_-^*(T<T_c)<1$ and $p_0^*(T<T_c)<1$, i.e. there can be missing links in the periphery and between core and periphery. Moreover, we see that $W_0^*(T<T_c)$ is subleading with respect to both $W_+^*(T<T_c)$ and $W_-^*(T<T_c)$: although individual core-periphery links have an expected weight $w_0^*(T<T_c)$ larger than the expected weight $w_-^*(T<T_c)$ of individual periphery-periphery links, the number n_+n_- of core-periphery links is of smaller order with respect to the number $n_-(n_--1)/2$ of periphery-periphery links, and as a result the total weight of all core-periphery links is of smaller order as well.

To obtain $s_\pm^*(T<T_c)$ to leading order, we use Eqs. (2.46) and (2.47):

$$\begin{aligned} s_+^*(T<T_c) &= (n_+-1)w_+^*(T<T_c)+n_-w_0^*(T<T_c) \\ &\approx (n_+-1)\psi_+^*(T<T_c)n^2, \end{aligned} \tag{2.129}$$

$$\begin{aligned} s_-^*(T<T_c) &= (n_--1)w_-^*(T<T_c)+n_+w_0^*(T<T_c) \\ &\approx n\,w^*[1-Q_n^*(T<T_c)]. \end{aligned} \tag{2.130}$$

Now, combining the above expressions, we see that the order parameter defined in Eq. (2.70) can be written as

$$\begin{aligned} Q^*(T<T_c) &= \lim_{n\to\infty}\frac{W_+^*(T<T_c)}{W_+^*(T<T_c)+W_-^*(T<T_c)} \\ &= \lim_{n\to\infty}\frac{1}{1+\frac{n_-(n_--1)w_-^*(T<T_c)}{n_+(n_+-1)w_+^*(T<T_c)}} \\ &= \lim_{n\to\infty}\frac{1}{1+\frac{n_-s_-^*(T<T_c)}{n_+s_+^*(T<T_c)}} \\ &= \frac{1}{1+\frac{w_-^*(T<T_c)}{n_+(n_+-1)\psi_+^*(T<T_c)}} \\ &= \frac{n_+(n_+-1)\psi_+^*(T<T_c)}{w^*}>0. \end{aligned} \tag{2.131}$$

Besides quantifying the order parameter, the above calculation shows that, since the value of $Q^*(T<T_c)$ only depends on the values of $s_+^*(T<T_c)$ and $s_-^*(T<T_c)$ (which by construction are the same in the canonical and microcanonical ensembles), whenever one ensemble is in the BEC phase, the other ensemble is the BEC phase as well, for all temperatures $T<T_c$. As for the non-condensed, this allows us to refer

to the conjugate canonical and microcanonical ensembles being 'in the same phase' consistently. Inverting Eq. (2.131), we can also express the parameter $\psi_+^*(T<T_c)$ in terms of the order parameter as follows:

$$\psi_+^*(T<T_c) = \frac{w^* Q^*(T<T_c)}{n_+(n_+-1)}. \tag{2.132}$$

The 'inverse' problem is solved by inverting Eqs. (2.129) and (2.130) and using them into Eq. (2.51) to get

$$\begin{aligned}
\beta_+^*(T<T_c) &\approx \frac{1}{2}\ln\left(1+\frac{1}{n^2\psi_+^*(T<T_c)}\right) \\
&\approx \frac{1}{2n^2\psi_+^*(T<T_c)} \\
&\approx \frac{n_+-1}{2s_+^*(T<T_c)} \\
&\approx \frac{n_+(n_+-1)}{2n^2 w^* Q_n^*(T<T_c)},
\end{aligned} \tag{2.133}$$

$$\begin{aligned}
\beta_-^*(T<T_c) &= \frac{1}{2}\ln\left(1+\frac{1}{w_-^*(T<T_c)}\right) \\
&\approx \frac{1}{2}\ln\left(1+\frac{n}{s_-^*(T<T_c)}\right) \\
&\approx \frac{1}{2}\ln\left(1+\frac{1}{w^*\left[1-Q_n^*(T<T_c)\right]}\right).
\end{aligned} \tag{2.134}$$

The above equations solve the inverse problem by expressing $\beta_\pm^*(T<T_c)$ as a function of the constraints $s_\pm^*(T<T_c)$, which can in turn be expressed either in terms of the finite parameters $\psi_+^*(T<T_c)$ and $w_-^*(T<T_c)$ or in terms of $Q_n^*(T<T_c)$ and the temperature-independent parameter w^*.

Again, the 'direct' problem requires finding the chemical potential $\mu^*(T)$ as a function of ϵ^*, w^* and T. From Eq. (2.66) we get

$$\begin{aligned}
\mu^*(T<T_c) &= -2kT\beta_+^*(T<T_c) \\
&\approx -\frac{kT}{n^2\psi_+^*(T<T_c)} \\
&\approx -kT\frac{n_+(n_+-1)}{n^2 w^* Q_n^*(T<T_c)}.
\end{aligned} \tag{2.135}$$

We now consider the variance of the constraints. Inserting Eqs. (2.120), (2.125) and (2.127) into Eqs. (2.74), we immediately see that

$$\Sigma_+^*(T<T_c) \approx (n_+-1)[\psi_+^*(T<T_c)]^2 n^4, \tag{2.136}$$

$$\Sigma_-^*(T<T_c) \approx w_-^*(T<T_c)[1+w_-^*(T<T_c)]n. \tag{2.137}$$

Finite (subcritical) temperature: $0<T<T_c$

In this regime we have

$$0 < Q^*(0<T<T_c) < 1, \tag{2.138}$$

which, as clear from Eqs. (2.125) and (2.127), implies

$$w_-^*(0<T<T_c) \approx [1-Q^*(0<T<T_c)]w^*, \tag{2.139}$$

$$w_0^*(0<T<T_c) \approx \frac{1}{\sqrt{1+\frac{1}{[1-Q^*(0<T<T_c)]w^*}}-1}, \tag{2.140}$$

where both quantities are $O(1)$. Using these results, it is possible to show (see Appendix 2.A.3) that the leading term of the determinant of the covariance matrix between the constraints is

$$\det[\mathbf{\Sigma}^*(0<T<T_c)] = O(n^{n+3n_+}), \tag{2.141}$$

implying

$$\ln\det[\mathbf{\Sigma}^*(0<T<T_c)] = O(n\ln n), \tag{2.142}$$

which is the same scaling found in Eq. (2.99) for the non-condensed phase. The criterion for measure equivalence in Eq. (2.34) is again violated, showing that ensemble equivalence does not hold in the condensed case as well. The leading term of the relative entropy can still be calculated exactly from Eq. (2.35) and is the same as the one found in Eq. (2.100) for the non-condensed phase:

$$D^*(0<T<T_c) \approx \frac{1}{2}\ln\det\mathbf{\Sigma}^*(0<T<T_c) \approx \frac{1}{2}n\ln n. \tag{2.143}$$

Similarly, the canonical entropy is still of order $O(n^2)$, as can be seen by inserting Eqs. (2.129), (2.130), (2.133) and (2.134) into Eq. (2.75). We therefore retrieve

$$S^*_{\rm can}(0<T<T_c) = O(n^2), \tag{2.144}$$

$$S^*_{\rm mic}(0<T<T_c) = O(n^2). \tag{2.145}$$

Coming to the relative fluctuations, from Eqs. (2.77), (2.79), (2.136) and (2.137) we obtain

$$\rho^*_+(0<T<T_c) = \frac{1}{\sqrt{n_+-1}}, \tag{2.146}$$

$$\rho^*_-(0<T<T_c) = 0. \tag{2.147}$$

The above result can be interpreted as follows. The term $\sum_{i\neq j}(\langle w_{ij}\rangle_{\vec{\beta}^*})^2/(\sum_{i\neq j}\langle w_{ij}\rangle_{\vec{\beta}^*})^2$ in Eq. (2.37) is an inverse participation ratio, taking values in the range $[(n-1)^{-1},1]$ and quantifying the inverse of the effective number of link weights contributing to the strength of node i [49]. Here, for a node in the core, there is a finite number n_+-1

of dominant link weights, each equal to $w_+^* = O(n^2)$, while the remaining n_- weights are of smaller order. Taking the thermodynamic limit, these $n_+ - 1$ dominant weights lead to the value for ρ_+^* in Eq. (2.146). By contrast, for a node in the periphery, all the expected link weights are of the same order, so the inverse participation ratio, and consequently the value of ρ_-^* in Eq. (2.147), vanishes. It should be noted that, even if the expected strength of the core nodes is much bigger than that of the periphery nodes, the relative fluctuations of the core nodes do not vanish, while those of the peripheral nodes do.

The fact that BEC occurs necessarily among the core nodes confirms that the units of the system are the nodes, and not the node pairs: the 'ground state pairs' are necessarily all and only the pairs of 'ground state nodes'. Indeed, one cannot decide arbitrarily which node pairs form the degenerate ground state where condensation occurs. This would have been possible only if node pairs were the fundamental units, by assigning the same degenerate ground state energy value to any set of node pairs (including pairs not necessarily involving the same set of nodes). For instance, it would have been possible to include the pairs (i, j) and (i, k), without necessarily including the pair (j, k), in the degenerate ground state (which is instead unavoidable in our system).

Zero temperature: $T \to 0$

We finally consider the zero-temperature limit as the extreme case of the condensed phase. Importantly, we have to be careful how we approach the two limits $T \to 0$ and $n \to \infty$. Indeed, we are going to show that taking the limit $T \to 0$ while n is kept fixed leads to results that cannot be subsequently carried over to the thermodynamic limit by taking the limit $n \to \infty$. Since we are interested precisely in the thermodynamic limit, we have to take a different route. To show the difference, we consider the zero-temperature limit first in the case of finite n and then in the case of growing n.

If n is finite, the zero-temperature limit simply represents the situation where the only populated state is the degenerate ground state corresponding to the links in the core, i.e.

$$Q_n^*(T \to 0) = 1 \tag{2.148}$$

by construction. All other links are not present. As usual, if the ground state is not degenerate, then the microcanonical entropy is zero, while if the ground state is degenerate, then the microcanonical entropy approaches a value called *residual entropy* which, for a system of fixed size, is a constant that depends only on the degeneracy (these statements usually go under the names of *Third Law of Thermodynamics* or, somehow improperly, *Nernst Theorem*) [60]. In our setting the ground state is non-degenerate only if $n_+ = 2$, in which case the link between the two core nodes is the one with minimum energy. In the general case $n_+ > 2$, the ground state is degenerate and both the microcanonical and canonical entropies are strictly positive. In any case, for finite n the zero-temperature limit is characterized by the fact that, in both the canonical and microcanonical ensembles, all nodes in the periphery are *deterministically* isolated, i.e. necessarily isolated in all realizations of the network.

The periphery becomes completely disconnected, both internally and from the core. Note that this is one of the degenerate situations (mentioned in Sec. 2.1.3) where, even if the constraints are in principle all mutually independent, for certain degenerate parameter value(s) some of them become 'hard' in both ensembles, thereby not contributing any difference between the two ensembles. Note that if we simply take the ideal limit $n \to \infty$ starting from this zero-temperature state, we would be considering the degenerate situation where an infinite number of isolated 'peripheral' nodes are added to the fully connected core. These nodes are unavoidably disconnected in both ensembles, so their contribution to the system is purely formal. The only variability (hence the only possible source of nonequivalence) comes from the core, which keeps having a finite number n_+ of nodes: as an extreme signature of BEC, the condensate behaves as an effectively lower-dimensional object.

In order to access the thermodynamic limit, we therefore have to consider from the beginning the case where n can grow indefinitely. We are going to show that the main difference arises from the fact that the temperature can only correspond to graphical strength sequences, which on the other hand depend on n. Therefore one should expect a certain n-dependent temperature T_n. At that point, a temperature value $T_n > 0$ that is small but finite when n is finite (allowing for certain populated excited states besides the ground state) may actually approach zero as n diverges, i.e. $\lim_{n\to\infty} T_n = 0$. The corresponding excited states will effectively become part of the accessible configurations in the zero-temperature limit and contribute an extra residual entropy.

To study this scenario, we start from the consideration that if the two limits $T \to 0$ and $n \to \infty$ were taken simultaneously for quantities that depend on both Q_n^* and n, e.g. terms such as $n(1-Q_n^*)$, we would encounter indeterminate expressions. We therefore need to understand how, as T goes to zero, Q_n^* goes to one as a function of n, for n large. We recall that our starting point is always the value $\vec{s}^*(T)$ of the constraints. The temperature T is a parameter that allows us to vary $\vec{s}^*(T)$, while keeping it graphic, i.e. realizable in at least one configuration of the network. For large n, we therefore have to identify the possible states of the network, hence the values of $\vec{s}^*(T)$, closest to zero temperature, i.e. when $T \simeq 0$ (we will use the symbol '$\simeq$' to denote this near-zero-temperature behaviour of any quantity, thereby keeping the notation distinct from the symbol '$\approx$' that will still denote the leading order of any quantity for large n). This is easily done by realizing that, if we start from some $\vec{s}^*(T>0)$ and decrease the temperature towards zero, the lowest excited state accessible to the network (before all links condense in the core) is one where only the smallest possible number ΔW^* of the W^* particles of weight remain out of the core, while keeping the strength sequence $\vec{s}^*(T \simeq 0)$ in the form given by Eq. (2.42). This state is necessarily such that $s_-^*(T \simeq 0) = 1$ (which is the minimum non-zero value of the strength, recalling that the strength is a non-negative integer by construction) and can be realized in multiple ways: either by connecting the n_- peripheral nodes in pairs, thus creating $n_-/2$ periphery-periphery links of unit weight and energy ϵ^* (in which case $\Delta W^* = n_-/2$), or by connecting each peripheral node to a core node, thus creating n_- core-periphery links of unit weight and energy

$\epsilon^*/2$ (in which case $\Delta W^* = n_-$), or finally by combining both types of situations. Recalling the discussion in Sec. 2.4.2, in all cases the ΔW^* links outside of the core have collectively the same energy $\epsilon^* n_-/2$ while the links in the core have zero energy; indeed, all such configurations are equiprobable. If we also consider the next excited states with $s_-^*(T \simeq 0) = 2, 3, \ldots$, in general we will have $\Delta W^*(T \simeq 0) = n_- \ell/2$ where ℓ is a small (in a sense that will be clear in a moment) integer.

The above considerations imply that the possible values of Q_n^* close to zero temperature are of the form

$$\begin{aligned} Q_n^*(T \simeq 0) &= \frac{W^* - \Delta W^*(T \simeq 0)}{W^*} \\ &= 1 - \frac{n_- \ell/2}{W^*} \\ &\approx 1 - \frac{\ell}{n w^*}. \end{aligned} \tag{2.149}$$

Basically, the above expression makes it explicit that, since the strength is a discrete quantity, technically the temperature can only take discrete values in order to keep the strength sequence graphic, so the role of T is taken up by ℓ (which is an integer) and a low temperature corresponds to a 'small', i.e. finite or at most $o(n)$, value of ℓ. Indeed, in the thermodynamic limit we recover

$$Q^*(T \simeq 0) = \lim_{n \to \infty} Q_n^*(T \simeq 0) = 1, \tag{2.150}$$

confirming that $\ell = o(n)$ leads to the correct zero-temperature limit. At the same time, Eq. (2.149) shows that, to recover any finite-temperature value $Q^*(T > 0) < 1$, we would need ℓ to grow linearly in n in the thermodynamic limit (note that ℓ cannot grow faster than n, because ΔW^* cannot grow faster than n^2, which is the order of W^*). The $\ell = o(n)$ regime considered here is therefore genuinely different from the positive-temperature cases discussed so far.

Having characterized the zero-temperature limit in this way, we can calculate

$$\lim_{n \to \infty} n[1 - Q_n^*(T \simeq 0)] = \frac{\ell}{w^*}, \tag{2.151}$$

from which we can obtain various asymptotic expressions. Indeed, from Eqs. (2.125) and (2.126) we obtain

$$w_-^*(T \simeq 0) \approx [1 - Q_n^*(T \simeq 0)] w^* \approx \frac{\ell}{n} \tag{2.152}$$

and

$$W_-^*(T \simeq 0) \approx \frac{\ell n}{2}. \tag{2.153}$$

Similarly, expanding Eqs. (2.127) and (2.128) for Q_n^* close to 1, we get

$$w_0^*(T \simeq 0) \approx \sqrt{[1 - Q_n^*(T \simeq 0)] w^*} \approx \sqrt{\frac{\ell}{n}} \tag{2.154}$$

and

$$W_0^*(T \simeq 0) \approx n_+ \sqrt{\ell n}. \tag{2.155}$$

Since both $W_-^*(T \simeq 0)$ and $W_0^*(T \simeq 0)$ are subleading with respect to $W_+^*(T \simeq 0)$, we have $W_+^*(T \simeq 0) \approx W^*$ which can be rewritten as $w_+^*(T \simeq 0) n_+(n_+ - 1)/2 \approx w^* n(n-1)/2$. This implies

$$w_+^*(T \simeq 0) \approx \frac{w^* n^2}{n_+(n_+ - 1)} \tag{2.156}$$

and, from Eq. (2.121),

$$\psi_+^*(T \simeq 0) = \frac{w^*}{n_+(n_+ - 1)}, \tag{2.157}$$

consistently with Eq. (2.132).

We now note that, inserting Eqs. (2.152) and (2.156) into Eq. (2.68), we obtain the anticipated dependence of T_n (for $T_n \simeq 0$) on n:

$$e^{-\epsilon^*/kT_n} = \frac{w_-^*(T_n \simeq 0)[1 + w_+^*(T_n \simeq 0)]}{w_+^*(T_n \simeq 0)[1 + w_-^*(T_n \simeq 0)]} \approx \frac{\ell}{n}. \tag{2.158}$$

Inverting, we find how the temperature approaches zero as n grows:

$$T_n \approx \frac{\epsilon^*}{k \ln(n/\ell)} \approx \frac{\epsilon^*}{k \ln n}. \tag{2.159}$$

The above result, which is independent of ℓ, connects the thermodynamic limit $n \to \infty$ with the zero-temperature limit $T \to 0$ in our setting and confirms that it would be inappropriate to first identify the ground state are the core links by taking the limit $T \to 0$ and subsequently let n grow. On the contrary, the zero-temperature state turns out to be the entire set of configurations obtained displacing a certain number of units of weight out of the core and such that $\ell = o(n)$. Inserting Eqs. (2.152), (2.154) and (2.156) into Eqs. (2.53) and (2.54) we can characterize these configurations through the connection probabilities

$$p_+^*(T \simeq 0) \approx 1 - \frac{n_+(n_+ - 1)}{w^* n^2}, \tag{2.160}$$

$$p_-^*(T \simeq 0) \approx \frac{\ell}{n}, \tag{2.161}$$

$$p_0^*(T \simeq 0) \approx \sqrt{\frac{\ell}{n}}, \tag{2.162}$$

which in the thermodynamic limit behave as expected for the ground state, i.e. $p_+^*(T \to 0) = 1$, $p_-^*(T \to 0) = 0$ and $p_0^*(T \to 0) = 0$.

Using Eqs. (2.152) and (2.154), we can calculate the strengths from Eqs. (2.129) and (2.130) as

$$s_+^*(T\simeq 0) \approx (n_+-1)\psi_+^*(T\simeq 0)n^2 \approx \frac{w^*}{n_+}n^2 \tag{2.163}$$

$$s_-^*(T\simeq 0) \approx nw_-^*(T\simeq 0)+n_+w_0^*(T\simeq 0)\approx \ell. \tag{2.164}$$

The 'inverse' problem is solved by Eqs. (2.133) and (2.134), which now become

$$\beta_+^*(T\simeq 0) \approx \frac{n_+(n_+-1)}{2n^2w^*}, \tag{2.165}$$

$$\beta_-^*(T\simeq 0) \approx \frac{1}{2}\ln\frac{n}{\ell}\approx\frac{1}{2}\ln n. \tag{2.166}$$

By contrast, the solution to the 'direct' problem is given through the chemical potential, obtained inserting Eqs. (2.159) and (2.165) into Eq. (2.135):

$$\begin{aligned}\mu^*(T\simeq 0) &= -2kT\beta_+^*(T\simeq 0)\\ &\approx -\frac{\epsilon^* n_+(n_+-1)}{w^*n^2\ln n}.\end{aligned} \tag{2.167}$$

The variances of the constraints can be calculated inserting Eqs. (2.152) and (2.157) into Eqs. (2.136) and (2.137). This yields

$$\Sigma_+^*(T\simeq 0) \approx \frac{(w^*)^2}{n_+^2(n_+-1)}n^4, \tag{2.168}$$

$$\Sigma_-^*(T\simeq 0) \approx \ell. \tag{2.169}$$

Using the above relationships, it is possible to show (see Appendix 2.A.4) that the leading order of the determinant of the covariance matrix is

$$\det[\mathbf{\Sigma}^*(T\simeq 0)] = O(n^{4n_+}\ell^{n-n_+}). \tag{2.170}$$

This leads to

$$\ln\det[\mathbf{\Sigma}^*(T\simeq 0)] = O(n\ln\ell), \tag{2.171}$$

which is of smaller order compared to the scaling $O(n\ln n)$ found in Eqs. (2.99), (2.114) and (2.142) for all positive temperatures, but still signalling the breakdown of ensemble equivalence. Equation (2.171) implies that the requirements ensuring the validity of Eq. (2.35) are not met, therefore in this case the leading term of the relative entropy cannot be calculated exactly. However, the leading order is still given by Eq. (2.33)

$$D^*(T\simeq 0) = O\big(\ln\det\mathbf{\Sigma}^*(T\simeq 0)\big) = O(n\ln\ell). \tag{2.172}$$

The canonical entropy $S_{\text{can}}^*(T\simeq 0)$ can be calculated by using Eqs. (2.165) and (2.166) to rewrite the sum of the five terms appearing into Eq. (2.75) as the following sum:

$$\begin{aligned}S_{\text{can}}^*(T\simeq 0) &= \frac{n_+(n_+-1)}{2}+\frac{\ell}{2}n\ln n+n_+(n_+-1)\ln n\\ &\quad +\frac{\ell}{2}n+n_+\sqrt{n\ell}=O(\ell n\ln n),\end{aligned} \tag{2.173}$$

which, unless $\ell = O(n/\ln n)$ or bigger, is different from the scaling $O(n^2)$ obtained for finite temperatures in Eqs. (2.101) and (2.144). This slower increase of the canonical entropy with n confirms that in the zero-temperature limit the system behaves as a Bose-Einstein condensate, a phenomenon that determines a strong reduction in the dimensionality of the space of allowed configurations. Note that since the relative entropy is of order given by Eq. (2.171), which is smaller than the order of $S^*_{\rm can}(T \simeq 0)$, the leading term of the microcanonical entropy must be

$$S^*_{\rm mic}(T \simeq 0) \approx \frac{\ell}{2} n \ln n = O(\ell n \ln n). \tag{2.174}$$

Combining Eqs. (2.163)-(2.169), we finally obtain the relative canonical fluctuations

$$\rho^*_+(T \simeq 0) = \frac{\sqrt{n_+}}{n_+ - 1}, \tag{2.175}$$

$$\rho^*_-(T \simeq 0) = \frac{1}{\sqrt{\ell}}, \tag{2.176}$$

which differ from the results in Eqs. (2.146) and (2.147) obtained in the (subcritical) finite-temperature case. In particular, now both $\rho^*_-(T \simeq 0)$ and $\rho^*_-(T \simeq 0)$ (if ℓ is finite) are non-zero. Note that, while Eq. (2.147) can be formally retrieved from Eq. (2.176) by letting ℓ grow linearly in n, Eq. (2.146) cannot be retrieved from Eq. (2.175), because the assumption $\ell = o(n)$ has already been exploited in the derivation.

2.4.5 Critical temperature: $T = T_c$

Having characterized the model in all regimes, we can now discuss more easily what happens right at the critical temperature $T = T_c$. Clearly, a very interesting question is whether the phase transition is of first or second order. A first-order phase transition is obtained when the order parameter $Q^*(T)$ jumps discontinuously from zero to a strictly positive value as T is lowered through T_c. In such a case, the left and right limits of $Q^*(T)$ at $T = T_c$ are different:

$$Q^*(T \to T_c^-) > Q^*(T \to T_c^+) = 0. \tag{2.177}$$

By contrast, the phase transition is of second order if the order parameter increases continuously from zero to positive values as T is lowered through T_c:

$$Q^*(T \to T_c^-) = Q^*(T \to T_c^+) = 0. \tag{2.178}$$

In principle, in our setting we can engineer the order of the phase transition as we like: as clear from Eq. (2.131), the value of the order parameter for values slight below the critical temperature is governed by the value of $\psi^*_+(T \lesssim T_c)$ defined in Eq. (2.121). So, if we choose $\psi^*_+(T \to T_c^-) > 0$ the transition will be first-order, while if we choose $\psi^*_+(T \to T_c^-) = 0$ the transition will be second-order. While both choices are possible,

the case $\psi_+^*(T \to T_c^-) > 0$ is somewhat unnatural, since it would 'forbid' all those strength sequences $\vec{s}^*(T<T_c)$ that, while being both graphic and perfectly consistent with the definition of 'condensed' given in Sec. 2.4.4, are such that asymptotically $w_+^*(T<T_c) < n^2\psi_+^*(T \to T_c^-)$ or equivalently, by virtue of Eq. (2.124), such that $w_-^*(T<T_c) < w^* - n_+(n_+-1)\psi_+^*(T \to T_c^-)$. Note that the latter inequality implies that $w_-^*(T)$ would experience a finite jump from $w^* - n_+(n_+-1)\psi_+^*(T \to T_c^-) < w^*$ to w^* as T is raised from a value just below T_c to a value just above T_c: as discussed in Sec. 2.4.3, w^* is (to leading order) the only allowed value for $w_-^*(T)$ above the critical temperature.

Therefore we find more appropriate to choose $\psi_+^*(T)$ such that $\psi_+^*(T \to T_c^-) = 0$. In this way, all values of $w_-^*(T<T_c)$ in the range $[0, w^*]$ are allowed and there is no discontinuity for $w_-^*(T)$ at T_c: Eq. (2.124) implies that its left limit is $w_-^*(T \to T_c^-) \approx w^*$, which coincides with its right limit $w_-^*(T \to T_c^+) \approx w^*$ implied by Eq. (2.86). With this choice, we can locate the critical temperature T_c by equating the right and left limits of Eq. (2.68): since $\lim_{T \to T_c^-} e^{-\epsilon^*/kT} = \lim_{T \to T_c^+} e^{-\epsilon^*/kT} = e^{-\epsilon^*/kT_c}$, the right and left limits of

$$\frac{w_-^*(T)[1+w_+^*(T)]}{w_+^*(T)[1+w_-^*(T)]} \tag{2.179}$$

must coincide. This implies

$$\frac{1+w_+^*(T \to T_c^+)}{w_+^*(T \to T_c^+)} = \frac{1+w_+^*(T \to T_c^-)}{w_+^*(T \to T_c^-)} = 1 \tag{2.180}$$

since in the thermodynamic limit $w_+^*(T \to T_c^-) = \infty$. The above expression in turn implies $w_+^*(T \to T_c^+) = \infty$. Note that this is consistent with the fact that, as discussed in Sec. 2.4.3, $w_+^*(T>T_c)$ is finite but can be arbitrarily large, while not altering (to leading order) the average weight w^*. So, $w_+^*(T>T_c)$ can grow indefinitely as T decreases towards T_c, and take an infinite limit as $T \to T_c^-$, consistently with the fact that, for even lower temperatures, $w_+^*(T)$ diverges with speed n^2 as dictated by Eq. (2.120). Inserting Eq. (2.180) into Eq. (2.68), we get

$$e^{-\epsilon^*/kT_c} = \frac{w^*}{1+w^*} \tag{2.181}$$

as $n \to \infty$. We therefore obtain

$$T_c = \frac{\epsilon^*}{k \ln\left(1+\frac{1}{w^*}\right)}, \tag{2.182}$$

finally showing how the critical temperature depends on the expected link weight w^*, on the energy difference ϵ^* between periphery-periphery and core-core links, and on the constant k converting the units of the 'cost of links' (energy) to those of the temperature. These results explain much of the information anticipated previously in Fig. 2.3 and summarized therein.

Adopting the view that the order parameter is continuous through the critical value T_c, we notice that the 'direct' solutions in Eqs. (2.93) and (2.135), as well as the behaviour at the critical point $T = T_c$, can be combined into the general 'phenomenological' expression

$$\mu^*(T) \approx \epsilon^* - kT \ln \left(1 + \frac{1}{w^* \left[1 - Q_n^*(T) \right]} \right) + o(T) \tag{2.183}$$

which is valid for all values of the temperature. Indeed, when $T \geq T_c$ the order parameter is zero and the above expression reduces to Eq. (2.93), while when $T < T_c$ the order parameter takes the positive value in Eq. (2.131) and the above expression reduces to Eq. (2.135). The extreme limits $T \to \infty$ and $T \to 0$ can be retrieved from Eq. (2.183) as well.

2.5 Conclusions

We have investigated the breakdown of equivalence between canonical and microcanonical ensembles of weighted networks with local constraints on the strength of each node (*weighted configuration model* [44]). While ensemble nonequivalence in the corresponding *binary configuration model* (i.e. binary networks with given node degrees) had already been studied in detail [5, 26, 7, 45, 61], a similar analysis for weighted networks had not been carried out so far. As a unique and novel ingredient in the case considered here, weighted networks can undergo BEC, a phase transition that is impossible to observe in the unweighted case. BEC emerges when a finite fraction of the total weight of all links condenses in a finite number of links. We constructed the simplest model exhibiting such behaviour: a network with a finite core, an infinite periphery, and a temperature-dependent strength sequence. This setting allows us to combine for the first time, in a single model, two completely different mechanisms that can potentially destroy the equivalence of ensembles: a phase transition (a condition exhibited in the earliest observations of ensemble nonequivalence [22, 38, 39, 23–25, 40–42, 21, 43, 13]) and an extensive number of local constraints (an ingredient found in more recent investigations on network ensembles [5, 26, 45, 37]).

We have considered two criteria for ensemble equivalence: the traditional and intuitive one based on the vanishing of the relative canonical fluctuations of the constraints in the thermodynamic limit [1] and the more recent and rigorous one based on the vanishing of the relative entropy density between microcanonical and canonical probability distributions (*measure equivalence*) [8]. While in the standard situation (i.e. under only one or a finite number of global constraints) the vanishing of the relative fluctuations implies measure equivalence, the relationship between the two criteria had not been investigated in presence of an extensive number of local constraints yet. Technically, while the relative fluctuations can be calculated exactly (as they are purely canonical quantities), the relative entropy requires in principle unfeasible microcanonical calculations but it can still be calculated asymptotically via a recently

proposed saddle-point technique showing that its leading term is the logarithm of the determinant of the matrix of canonical covariances between the constraints.

We found that, for all positive temperatures, the relative entropy is $O(n \ln n)$ while the canonical and microcanonical entropies are $O(n^2)$. These behaviours mimick the corresponding ones found for the binary configuration model in the dense regime [5, 26, 45]. This result shows that, for all $T>0$ (including $T\to\infty$), the relative entropy is subleading with respect to the canonical and microcanonical entropies, but is still superextensive in the number of nodes n, which in all network models represents the number of units (physical size) of the system. In the zero-temperature limit, we found slower scalings for the canonical, microcanonical and relative entropies. This is due to the fact that, in both canonical and microcanonical ensembles, the peripheral nodes are asymptotically disconnected from all other nodes in each possible realization of the network. In this zero-temperature limit, the condensate effectively behaves as a lower-dimensional system, as commonly observed in the physics of BEC. Its entropy is the residual entropy resulting from the degeneracy of the ground state. The scaling of the relative entropy still indicates ensemble nonequivalence. We note that in the binary configuration model (which obeys Fermi-Dirac rather than Bose-Einstein statistics) the zero-temperature phase is one where the canonical and microcanonical ensembles are instead *identical*, because in both ensembles the pairs of nodes below a certain 'Fermi energy' (whose value coincides with the chemical potential) are surely connected, while those above it are surely disconnected [20]. Therefore we can conclude that, irrespective of BEC, at all temperatures ensemble equivalence is broken by the presence of an extensive number of local constraints, as in the binary configuration model (for which, however, BEC cannot occur). So the condensation phase transition (occuring at some critical temperature $T_c>0$) appears to have no effect on ensemble equivalence.

On the other hand, the calculation of the canonical relative fluctuations of the constraints shows that they are sensitive to the phase transition, while they cannot be used to characterize ensemble (non)equivalence as traditionally expected. Indeed, we found that in the non-condensed phase ($T>T_c$) the relative fluctuations of the strength of all the n nodes vanish in the thermodynamic limit. By contrast, in the condensed phase ($T<T_c$) the relative fluctuations of the strength of the n_+ nodes in the core do not vanish, while those for the n_- nodes in the periphery still do (except in the zero-temperature limit, for which not even the relative fluctuations for the peripheral nodes vanish). Therefore, as the temperature is lowered below the critical temperature, there is a sudden change in the relative fluctuations but no change in the scaling of the relative entropy. Conversely, as the temperature is further lowered to zero, there is a sudden change in the scaling of the relative entropy, while the relative fluctuations for the core nodes remain non-zero, albeit with a different value. These results show that, at least in the dense case studied here, the relative entropy and the relative fluctuations capture different aspects of the phenomenology of the proposed model, the former being sensitive to the presence of local constraints and the latter being sensitive to the phase transition. In any case, in presence of an extensive number of local constraints the vanishing of (even all) the canonical relative

fluctuations does not guarantee measure equivalence and is therefore no longer a valid criterion for ensemble equivalence as intuitively expected.

We stress that, while the network model presented here is deliberately simple from the structural point of view (a core-periphery network with local, but homogeneous, constraints), it could certainly serve as a reference for more complicated models (e.g. a core-periphery network with local and heterogeneous constraints). Indeed, ensemble nonequivalence will still be manifest in such a generalized model for all positive temperatures, because research on binary networks with given degrees has shown that nonequivalence is due to the locality of the constraints, and not to their specific value [5, 26, 45, 37]. Additionally, since a more heterogeneous choice of the constraints can only increase the number of states with different energy in the network, we expect that BEC will still emerge below some critical temperature. In general, we expect a qualitatively similar behaviour to the one found here, with only quantitative differences.

The concept of ensemble equivalence is central for the foundations of statistical physics, irrespective of the particular system being considered. The findings documented here shed new light on the breakdown of EE, on the (possibly misleading) criteria used to detect it, and on the (so far undocumented) interplay between different mechanisms producing it. We hope they can inspire future research on these subjects.

Appendix 2.A Determinant of the covariance matrix

$\mathbf{\Sigma}^*(T)$ is the $n \times n$ canonical covariance matrix between the strengths of all nodes, with entries

$$\begin{aligned} \Sigma^*_{ij}(T) &= \begin{cases} \mathrm{Var}_{\vec{\beta}^*(T)}(s_i) & i = j \\ \mathrm{Var}_{\vec{\beta}^*(T)}(w_{ij}) & i \neq j \end{cases} \\ &= \begin{cases} \Sigma^*_{ii}(T) & i = j \\ \langle w_{ij} \rangle_{\vec{\beta}^*(T)} \left[1 + \langle w_{ij} \rangle_{\vec{\beta}^*(T)}\right] & i \neq j \end{cases} . \end{aligned} \tag{2.184}$$

Combining Eqs. (2.23) and (2.24) with the results discussed in Sec. 2.4.2 for our core-periphery model, it is easy to see that, for all values of temperature, $\mathbf{\Sigma}^*(T)$ is a combination of four blocks

$$\mathbf{\Sigma}^*(T) = \begin{bmatrix} \mathbf{A}(T) & \mathbf{B}(T) \\ \mathbf{C}(T) & \mathbf{D}(T) \end{bmatrix}, \tag{2.185}$$

where $\mathbf{A}(T)$ is the $n_+ \times n_+$ submatrix of covariances between the strengths of nodes in the core, with entries

$$A_{ij}(T) = \begin{cases} \Sigma^*_+(T) & i = j \\ w^*_+(T)[1 + w^*_+(T)] & i \neq j \end{cases} , \tag{2.186}$$

$\mathbf{B}(T)$ is the $n_+ \times n_-$ submatrix of covariances between the strengths of nodes across core and periphery, with entries

$$B_{ij}(T) = w^*_0(T)[1 + w^*_0(T)] \quad \forall i, j, \tag{2.187}$$

$\mathbf{C}(T)$ is a $n_- \times n_+$ matrix equal to the transpose of $\mathbf{B}(T)$, and $\mathbf{D}(T)$ is the $n_- \times n_-$ submatrix of covariances between the strengths of nodes in the periphery, with entries

$$D_{ij}(T) = \begin{cases} \Sigma^*_-(T) & i = j \\ w^*_-(T)[1 + w^*_-(T)] & i \neq j \end{cases} . \tag{2.188}$$

Depending on the range of temperature values of interest, different techniques become useful in order to calculate the determinant of $\mathbf{\Sigma}^*(T)$. We therefore consider each regime separately below.

2.A.1 Non-condensed phase

In the regime of finite supercritical temperature ($T > T_c$) discussed in Sec. 2.4.3, it is possible to show that the asymptotic behaviour of $\det \mathbf{\Sigma}^*(T > T_c)$ can be decomposed as the product of the diagonal elements of $\mathbf{\Sigma}^*(T > T_c)$, plus a correction. Our rationale for this decomposition comes from the fact that, as noted in Sec. 2.4.3, $w^*_+(T > T_c)$,

$w_-^*(T > T_c)$ and $w_0^*(T > T_c)$ are all $O(1)$, i.e. the expected link weights are all of the same, finite order. Consequently, the block structure depicted in Eq. (2.185) does not identify any particular difference in the order of magnitude of the entries of $\mathbf{\Sigma}^*(T > T_c)$. Rather, an important property of $\mathbf{\Sigma}^*(T > T_c)$ in this regime is that its diagonal entries are, on average, n times bigger than its off-diagonal ones. Indeed, the off-diagonal entries are $O(1)$, while the diagonal ones are $O(n)$. Then the asymptotic behaviour of $\det \mathbf{\Sigma}^*(T > T_c)$ must be essentially dictated by the product of the diagonal entries of $\mathbf{\Sigma}^*(T > T_c)$.

To make this intuition more rigorous, we recall that if a $k \times k$ matrix $\mathbf{L}$ can be decomposed as

$$\mathbf{L} = \mathbf{M} + \varepsilon \mathbf{N}, \tag{2.189}$$

where $\mathbf{M}$ is a diagonal matrix with entries of finite order and $\varepsilon\mathbf{N}$ is a perturbation, then Jacobi's formula applies as follows:

$$\begin{aligned} \det \mathbf{L} &= \det(\mathbf{M} + \varepsilon \mathbf{N}) \\ &= \det \mathbf{M} + \varepsilon (\det \mathbf{M}) \mathrm{tr}(\mathbf{M}^{-1}\mathbf{N}) + O(\varepsilon^2). \end{aligned} \tag{2.190}$$

Moreover, if the diagonal elements of $\mathbf{N}$ are equal to 0, then the product $\mathbf{M}^{-1}\mathbf{N}$ is a $k \times k$ zero matrix and therefore

$$\mathrm{tr}(\mathbf{M}^{-1}\mathbf{N}) = 0. \tag{2.191}$$

Equation (2.191) then becomes

$$\begin{aligned} \det \mathbf{L} &= \det \mathbf{M} + O(\varepsilon^2) \\ &= \prod_{i=1}^{k} L_{ii} + O(\varepsilon^2). \end{aligned} \tag{2.192}$$

Turning to the matrix $\mathbf{\Sigma}^*(T > T_c)$, we note that the above hypotheses apply by setting

$$k \equiv n, \qquad \varepsilon \equiv \frac{1}{n}, \qquad \mathbf{L} \equiv \frac{\mathbf{\Sigma}^*(T > T_c)}{n} \tag{2.193}$$

and defining the entries of $\mathbf{M}$ and $\mathbf{N}$ as

$$M_{ij} = \delta_{ij} \Sigma_{ij}^*(T > T_c)/n, \tag{2.194}$$

$$N_{ij} = (1 - \delta_{ij}) \Sigma_{ij}^*(T > T_c), \tag{2.195}$$

where δ_{ij} is the Kronecker delta symbol. Equation (2.192) then becomes

$$\det\left(\frac{\mathbf{\Sigma}^*(T > T_c)}{n}\right) = \frac{1}{n^n} \prod_{i=1}^{n} \Sigma_{ii}^*(T > T_c) + O(n^{-2})$$

and, finally,

$$\begin{aligned} \det \mathbf{\Sigma}^*(T>T_c) &= n^n \det\left(\frac{\mathbf{\Sigma}^*(T>T_c)}{n}\right) \\ &= \prod_{i=1}^{n} \Sigma^*_{ii}(T>T_c) + O(n^{n-2}), \end{aligned} \tag{2.196}$$

proving Eq. (2.96) used in the main text.

2.A.2 Infinite-temperature limit

In the infinite-temperature limit discussed in Sec. (2.4.3), the determinant can be calculated exactly as follows. From Eqs. (2.111) and (2.112) we see that, if we introduce a $k \times k$ matrix $\mathbf{Z}_k$ defined as

$$\mathbf{Z}_k = \begin{pmatrix} k-1 & 1 & \cdots & 1 & 1 \\ 1 & k-1 & 1 & \cdots & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & \cdots & 1 & k-1 & 1 \\ 1 & 1 & \cdots & 1 & k-1 \end{pmatrix}, \tag{2.197}$$

then we can rewrite the covariance matrix as

$$\mathbf{\Sigma}^*(T\to\infty) = w^*(1+w^*)\mathbf{Z}_n. \tag{2.198}$$

Clearly, the calculation of $\det \mathbf{\Sigma}^*(T\to\infty)$ reduces to the calculation of $\det \mathbf{Z}_n$:

$$\det \mathbf{\Sigma}^*(T\to\infty) = [w^*(1+w^*)]^n \det \mathbf{Z}_n. \tag{2.199}$$

To compute $\det \mathbf{Z}_k$ for arbitrary k, we note that

$$\mathbf{Z}_k = (k-2)\mathbf{I}_k + \mathbf{u}_k^T\mathbf{u}_k = (k-2)\left(\mathbf{I}_k + \frac{\mathbf{u}_k^T\mathbf{u}_k}{k-2}\right), \tag{2.200}$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix and

$$\mathbf{u}_k = (1, \cdots, 1) \tag{2.201}$$

is the k-dimensional row vector with all unit entries. Then, using Sylvester's identity $\det(\mathbf{I}_k + \mathbf{XY}) = \det(\mathbf{I}_l + \mathbf{YX})$ (where $\mathbf{X}$ is a $k \times l$ matrix, $\mathbf{Y}$ is an $l \times k$ matrix, and $\mathbf{I}_k$ and $\mathbf{I}_l$ are $k \times k$ and $l \times l$ identity matrices respectively) with $l = 1$, $\mathbf{X} = \mathbf{u}_k^T$,

$\mathbf{Y} = \mathbf{u}_k$ and $\mathbf{I}_l = 1$, we get

$$\begin{aligned}
\det \mathbf{Z}_k &= (k-2)^k \det\left(\mathbf{I}_k + \frac{\mathbf{u}_k^T}{\sqrt{k-2}}\frac{\mathbf{u}_k}{\sqrt{k-2}}\right) \\
&= (k-2)^k \det\left(1 + \frac{\mathbf{u}_k}{\sqrt{k-2}}\frac{\mathbf{u}_k^T}{\sqrt{k-2}}\right) \\
&= (k-2)^k \left(1 + \frac{k}{k-2}\right) \\
&= 2(k-1)(k-2)^{k-1}. \qquad (2.202)
\end{aligned}$$

Combining Eqs. (2.199) and (2.202), and setting $k = n$, we obtain exactly Eq. (2.113) used in the main text.

2.A.3 Condensed phase

In the regime of subcritical temperature ($T < T_c$) discussed in Sec. 2.4.4, the block structure indicated in Eq. (2.185) becomes particularly relevant, as it captures the important differences in the order of magnitude of both diagonal and off-diagonal entries of $\mathbf{\Sigma}^*(T<T_c)$ calculated using Eqs. (2.122), (2.125), (2.127), (2.136) and (2.137). We first express each block conveniently and then proceed to the calculation of the determinant. Inserting Eqs. (2.122) and (2.136) into Eq. (2.186), we obtain

$$\mathbf{A}(T<T_c) \approx [\psi_+^*(T<T_c)]^2 n^4\, \mathbf{Z}_{n_+}, \qquad (2.203)$$

where $\mathbf{Z}_k$ is still the matrix defined in Eq. (2.197). Next, we note from Eq. (2.187) that

$$\mathbf{B}(T<T_c) = w_0^*(T<T_c)[1 + w_0^*(T<T_c)]\mathbf{u}_{n_+}^T \mathbf{u}_{n_-}, \qquad (2.204)$$

where $\mathbf{u}_k$ is still given by Eq. (2.201). Similarly,

$$\mathbf{C}(T<T_c) = w_0^*(T<T_c)[1 + w_0^*(T<T_c)]\mathbf{u}_{n_-}^T \mathbf{u}_{n_+}. \qquad (2.205)$$

Finally, inserting Eq. (2.137) into Eq. (2.188), we obtain

$$\mathbf{D}(T<T_c) \approx w_-^*(T<T_c)[1 + w_-^*(T<T_c)]\, \mathbf{Z}_{n_-}. \qquad (2.206)$$

Now, since $\mathbf{A}(T<T_c)$ is invertible, it is useful to exploit the block structure of $\mathbf{\Sigma}^*(T<T_c)$ by expressing its determinant as

$$\det \mathbf{\Sigma}^*(T<T_c) = \det \mathbf{A}(T<T_c)\, \det \overline{\mathbf{A}}(T<T_c), \qquad (2.207)$$

where

$$\begin{aligned}
\overline{\mathbf{A}}(T<T_c) \equiv\ & \mathbf{D}(T<T_c) \qquad (2.208) \\
& -\mathbf{C}(T<T_c)\mathbf{A}^{-1}(T<T_c)\mathbf{B}(T<T_c)
\end{aligned}$$

is the so-called Shur complement of $\mathbf{A}(T<T_c)$. To calculate $\det \mathbf{A}(T<T_c)$, we use Eq. (2.203) and immediately obtain

$$\begin{aligned} \det \mathbf{A}(T<T_c) &\approx [\psi_+^*(T<T_c)]^{2n_+} n^{4n_+} \det \mathbf{Z}_{n_+} \\ &= O(n^{4n_+}) \end{aligned} \tag{2.209}$$

where, using Eq. (2.202),

$$\det \mathbf{Z}_{n_+} = 2(n_+ - 1)(n_+ - 2)^{n_+ - 1}. \tag{2.210}$$

To calculate $\det \overline{\mathbf{A}}(T<T_c)$, we first use Eq. (2.203) and obtain

$$\mathbf{A}^{-1}(T<T_c) \approx [\psi_+^*(T<T_c)]^{-2} n^{-4} \mathbf{Z}_{n_+}^{-1}, \tag{2.211}$$

where, using Eq. (2.197), $\mathbf{Z}_k^{-1}$ is easily calculated by direct inversion of $\mathbf{Z}_k$ as

$$\begin{aligned} \mathbf{Z}_k^{-1} &= c_k \begin{pmatrix} 2k-3 & -1 & \cdots & -1 & -1 \\ -1 & 2k-3 & -1 & \cdots & -1 \\ \vdots & & \ddots & & \vdots \\ -1 & \cdots & -1 & 2k-3 & -1 \\ -1 & -1 & \cdots & -1 & 2k-3 \end{pmatrix} \\ &= \frac{\mathbf{I}_k}{k-2} - c_k \mathbf{u}_k^T \mathbf{u}_k \\ &= \frac{1}{k-2} \left[\mathbf{I}_k - \frac{\mathbf{u}_k^T \mathbf{u}_k}{2(k-1)} \right] \end{aligned} \tag{2.212}$$

with

$$c_k = \frac{1}{2(k-1)(k-2)}. \tag{2.213}$$

Inserting Eqs. (2.204),(2.205), (2.206) and (2.211) into Eq. (2.208), and noticing that

$$\begin{aligned} \mathbf{u}_{n_-}^T \mathbf{u}_{n_+} \mathbf{Z}_{n_+}^{-1} \mathbf{u}_{n_+}^T \mathbf{u}_{n_-} &= c_{n_+} n_+ (n_+ - 2) \mathbf{u}_{n_-}^T \mathbf{u}_{n_-} \\ &= \frac{n_+}{2(n_+ - 1)} \mathbf{u}_{n_-}^T \mathbf{u}_{n_-}, \end{aligned} \tag{2.214}$$

we can obtain the Shur complement of $\mathbf{A}(T<T_c)$ as

$$\begin{aligned} \overline{\mathbf{A}}(T<T_c) \approx\ & w_-^*(T<T_c)[1 + w_-^*(T<T_c)]\, \mathbf{Z}_{n_-} \\ & - \frac{n_+ [w_0^*(T<T_c)]^2 [1 + w_0^*(T<T_c)]^2}{2(n_+ - 1)[\psi_+^*(T<T_c)]^2 n^4} \mathbf{u}_{n_-}^T \mathbf{u}_{n_-} \end{aligned} \tag{2.215}$$

from which we can calculate $\det \overline{\mathbf{A}}(T<T_c)$. We have to distinguish the cases $0 < T < T_c$ and $T \simeq 0$, as they are characterized by different scalings of $w_-^*(T<T_c)$ and

$w_0^*(T<T_c)$. In the rest of this section we consider the case of finite temperature, while the zero-temperature limit is considered in the next section.

When $0<T<T_c$, we recall from Eqs.(2.139) and (2.140) that both $w_-^*(0<T<T_c)$ and $w_0^*(0<T<T_c)$ are $O(1)$. From Eq. (2.215) we therefore see that all the off-diagonal entries of $\overline{\mathbf{A}}(0<T<T_c)$ are $O(1)$, while all the diagonal ones are $O(n_-)$. This implies that we can use the decomposition in Eq. (2.189) where

$$k \equiv n_-, \qquad \varepsilon \equiv \frac{1}{n_-}, \qquad \mathbf{L} \equiv \frac{\overline{\mathbf{A}}(0<T<T_c)}{n_-}. \tag{2.216}$$

Equation (2.192) then implies

$$\begin{aligned} \det \overline{\mathbf{A}}(0<T<T_c) &= \prod_{i=1}^{n_-} \overline{A}_{ii}(0<T<T_c) + O\big(n_-^{n_- -2}\big) \\ &= O(n^{n_-}). \end{aligned} \tag{2.217}$$

Combining Eqs. (2.209) and (2.217) into Eq. (2.207), we finally obtain the full determinant of $\mathbf{\Sigma}^*(0<T<T_c)$. We are interested only in its scaling with n, which is

$$\det \mathbf{\Sigma}^*(0<T<T_c) = O(n^{4n_+ + n_-}) = O(n^{n+3n_+}), \tag{2.218}$$

proving Eq. (2.141) used in the main text.

2.A.4 Zero-temperature limit

In the zero-temperature limit, all calculations of the previous section remain valid until and including Eq. (2.215). The scaling of the entries of $\overline{\mathbf{A}}(T\simeq 0)$ will however be different. Indeed, we recall from Eqs.(2.152) and (2.154) that $w_-^*(T\simeq 0) \approx \ell/n$ and $w_0^*(T\simeq 0) \approx \sqrt{\ell/n}$. Inserted into Eq. (2.215), these expressions imply that all the diagonal entries of $\overline{\mathbf{A}}(T\simeq 0)$ are asymptotically equal to ℓ, while all the off-diagonal ones are $O(\ell/n)$. We can therefore use the decomposition in Eq. (2.189) where

$$k \equiv n_-, \qquad \varepsilon \equiv \frac{1}{n}, \qquad \mathbf{L} \equiv \overline{\mathbf{A}}(T\simeq 0). \tag{2.219}$$

Equation (2.192) then implies

$$\begin{aligned} \det \overline{\mathbf{A}}(T\simeq 0) &= \prod_{i=1}^{n_-} \overline{A}_{ii}(T\simeq 0) + O\big(n^{-2}\big) \\ &= \ell^{n_-} + O(n^{-2}). \end{aligned} \tag{2.220}$$

Combined with Eq. (2.209) into Eq. (2.207), the above result leads to the full determinant of $\mathbf{\Sigma}^*(T<T_c)$, whose scaling with n is

$$\det \mathbf{\Sigma}^*(T\simeq 0) = O(n^{4n_+}\ell^n), \tag{2.221}$$

proving Eq. (2.170) used in the main text.

Chapter 3

Strong ensemble nonequivalence in systems with local constraints

Abstract

The asymptotic equivalence of canonical and microcanonical ensembles is a central concept in statistical physics, with important consequences for both theoretical research and practical applications. However, this property breaks down under certain circumstances. The most studied violation of ensemble equivalence requires phase transitions, in which case it has a 'restricted' (i.e. confined to a certain region in parameter space) but 'strong' (i.e. characterized by a difference between the entropies of the two ensembles that is of the same order as the entropies themselves) form. However, recent research on networks has shown that the presence of an extensive number of local constraints can lead to ensemble nonequivalence even in the absence of phase transitions. This occurs in a 'weak' (i.e. leading to a subleading entropy difference) but remarkably 'unrestricted' (i.e. valid in the entire parameter space) form. Here we look for more general manifestations of ensemble nonequivalence in arbitrary ensembles of matrices with given margins. These models have widespread applications in the study of spatially heterogeneous and/or temporally nonstationary systems, with consequences for the analysis of multivariate financial and neural time-series, multi-platform social activity, gene expression profiles and other Big Data. We confirm that ensemble nonequivalence appears in 'unrestricted' form throughout the entire parameter space due to the extensivity of local constraints. Surprisingly, at the same time it can also exhibit the 'strong' form. This novel, simultaneously 'strong and unrestricted' form of nonequivalence is very robust and imposes a principled choice of the ensemble. We calculate the proper mathematical quantities to be used in real-world applications[1].

[1]This chapter is based on:
Qi Zhang, Diego Garlaschelli, "Strong ensemble nonequivalence in systems with local constraints" arXiv preprint arXiv:2107.04920 (2021)

3.1 Introduction

In statistical physics, systems with different constraints can be described by different ensembles. For example, systems with fixed energy can be described by the *microcanonical ensemble*, where all microscopic configurations have precisely the same value of the energy and are equiprobable, thereby modelling large isolated systems. In this case, the energy is treated as a 'hard' constraint enforced separately on each configuration. By contrast, systems with fixed temperature (which is the 'dual' thermodynamic quantity conjugated with the energy) can be described by the *canonical ensemble*, where individual microscopic configurations can have different values of the energy and are assigned different probabilities, but in such a way that the average value of the energy coincides with the one defining the corresponding microcanonical ensemble [10]. This ensemble represents systems that can exchange energy with their environment, and the energy is in fact treated as a 'soft' constraint which is enforced only as an ensemble average.

When the size of the system is finite, the two ensembles are necessarily different. However, in the simplest and most traditional situation, the microcanonical description as a function of the energy becomes equivalent with the canonical description as a function of the temperature in the *thermodynamic limit* (i.e., when the number of particles in the system tends to infinity). This phenomenon is called *ensemble equivalence* (EE) and is a basic concept in statistical mechanics as already established by Gibbs [1]. The property of EE justifies the replacement of the (typically unfeasible) asymptotic calculations in the microcanonical ensemble with the corresponding (much easier) calculations in the canonical ensemble, i.e. to choose the ensemble based on mathematical convenience.

However, over the past decades, the breakdown of EE has been observed in various physical systems, including models of gravitation, fluid turbulence, quantum phase separation, and networks [21, 43, 5]. When the system is under *ensemble nonequivalence* (EN), the microcanonical description can no longer be replaced by the canonical description in the thermodynamic limit. In this situation, many assumptions and calculations that are based on EE in statistical mechanics do not hold anymore. Thus checking for the breaking of EE is important for both practical applications and theoretical research. Quantitatively, EN can be defined as a nonvanishing relative entropy density between the microcanonical and canonical probability distributions of microscopic configurations [8, 5]. This is equivalent to a nonvanishing difference between the canonical and microcanonical entropy densities [37]. Technically, this is the so-called *measure-level* EN, which (under mild assumptions) has been shown to coincide with other definitions as well [8]. Importantly, the traditional criterion for EE based on the vanishing of the relative canonical fluctuations of the contraints has been recently found to break down when the latter are local in nature and extensive in number [28].

Indeed, for the most studied systems in statistical physics, the number of constraints defining the ensembles of interest is finite. Traditional physical examples are *global* constraints such as the total energy and the total number of particles. Non-

physical examples of systems under global constraints have also been considered, e.g. networks with given total numbers of edges and triangles [6]. In order to observe the breakdown of EE in these systems, one typically needs to introduce long-range interactions implying the non-additivity of the energy and possibly associated with the onset of phase transitions [13] (in the example of networks, the underlying mechanism is a sort of 'frustration' in the simultaneous realizability of the desired numbers of edges and triangles [6]). In this form of EN, the relative entropy between canonical and microcanonical ensembles is of the same order as the canonical entropy itself [8, 6]. This is what we will refer to as a 'strong' form of EN. At the same time, this form of EN is also 'restricted', because it is confined to a selected (e.g. critical) region of the space of parameters. Outside this region, EE is restored.

Recently, a new manifestation of EN has been observed in a different class of network ensembles, where a constraint is enforced on the *degree* (number of links) and/or the *strength* (total weight of all incident links) of each node [5, 26, 49, 28]: unlike systems with global constraints, in these models the number of constraints grows linearly in the number of nodes. This crucial difference implies that, at variance with the more 'traditional' situation described above, the onset of EN in this class of models is completely unrelated to phase transitions and is instead the result of the presence of an *extensive* number of *local* constraints [37, 7, 45]. This situation is by far less studied, because systems with local constraints are not the traditional focus of statistical physics and have attracted attention only recently as models of complex systems with built-in spatial heterogenetity [14] and/or temporal non-stationarity [62]. In this different form of EN, the relative entropy between microcanonical and canonical ensembles is, at least for the cases studied so far, of lower order (i.e. subleading) with respect to the canonical entropy. For this reason, we may refer to this situation as a 'weak' (i.e. weaker than the one found in the presence of phase transitions) form of EN. However, this form of EN is 'unrestricted', precisely because it is not confined to specific values of the control parameters and holds in the entire parameter space. Rather than a property of a phase (or a phase boundary), in this case EN appears to be an intrinsic property of the system itself. In these models, no parameter value can restore EE.

The above results indicate that, so far, EN has manifested itself either in a 'strong but restricted' form (under a finite number of global constraints, but in presence of phase transitions), or in a 'unrestricted but weak' form (under an extensive number of local constraints, but without phase transitions). Clearly, a number of questions remain open. How general is the manifestation of EN under local constraints, both in terms of the underlying mechanism and in terms of the strength of the resulting effect? Besides networks, can the breaking of EE be observed in additional systems characterized by local constraints? If so, do these systems necessarily exhibit only the weak form of EN, or can the strong form appear as well? Finally, is there a (possibly modified) way to exploit the canonical ensemble in order to bypass the challenge of unfeasible microcanonical calculations *even when EE breaks down*, i.e. even when the two ensembles can no longer be treated interchangeably according to mathematical convenience?

In this chapter, we will address these problems by exploring the effects of the presence of an extensive number of local constraints on more general ensembles than the ones that have been considered so far to model random networks with given node degrees [5, 26, 7, 63]. In particular, we consider the general setting where each of the n units of the system has a number m of 'state variables' (or 'degrees of freedom'), and where constraints are defined as sums over these state variables. Surprisingly, besides confirming the onset of an unrestricted form of EN in the thermodynamic limit where n diverges, we also find its simultaneous manifestation in strong form. This happens when each element of the system retains only a finite number m of degrees of freedom in the thermodynamic limit. For brevity, we will denote this situation as the 'strong and unrestricted' form of EN. To the best of our knowledge, this finding provides the first evidence that EN, *even in its strong form*, does not need phase transitions and can appear in the entire parameter space as an intrinsic property of the system, if the latter is subject to an extensive number of local constraints. This simultaneously 'strong and unrestricted' form of EN is the most robust among the ones studied so far. Spatial heterogeneity and temporal non-stationarity are simple candidate mechanisms that can lead to this phenomenon.

To emphasize the general and important consequences of this form of EN for a diverse range of practical applications, we consider generic ensembles of random matrices with fixed margins. These ensembles, which include matrices with 0/1 (or equivalently ± 1) and non-negative integer entries subject to global or local constraints, arise for instance in studies of multi-cell gene expression profiles [64], multiplex (online) social activity [65], multi-channel communication systems [29], complex networks [44], and multivariate time series in finance [62], neuroscience [66] or other disciplines. Our results imply that, in many practical situations, the assumption of EE is incorrect and leads to mathematically wrong conclusions. For the benefit of the aforementioned applications, we compensate for the 'disconnection' between the two ensembles by calculating explicitly the correct canonical and microcanonical quantities of interest via a generalized relationship that is either analytically computable or asymptotically determined by the covariance matrix of the constraints in the canonical ensemble. These calculations represent a practical tool for properly dealing with the consequences of EN in all real-world situations.

3.2 General formalism

3.2.1 Matrix ensembles

A discrete $n \times m$ matrix ensemble is a set $\mathcal{G}$ of available configurations for an $n \times m$ integer-valued matrix $\mathbf{G}$, endowed with a suitably chosen probability distribution $P(\mathbf{G})$ over such configurations, such that $\sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) = 1$. An entry of the matrix is denoted by g_{ij} (with $1 \leq i \leq n$, $1 \leq j \leq m$). We distinguish two main cases, the *binary* case where g_{ij} takes one of the two values $\{0, 1\}$ and the *weighted* case where g_{ij} takes a value in the set $\{0, 1, 2, \dots\}$ of non-negative integer values. The number n

of rows in each matrix represents the number of elements (i.e. the *size*) of the system being modelled. The number m of columns represents instead the number of state variables, or degrees of freedom, for each element.

In general, each matrix $\mathbf{G}$ can represent one of the possible states of a (large) real-world system. For instance, $\mathbf{G}$ may represent the realization of a multivariate time series, where n is the number of units (e.g. brain regions, financial stocks, etc.) emitting signals, and m is the number of time steps during which the signals are recorded. $\mathbf{G}$ may also represent a multi-cell array of gene expression profiles, where n is the number of cells and m the number of genes for which expression levels are being measured. Similarly, $\mathbf{G}$ may represent the state of a multi-channel communication systems, where n is the length of the sequences being transmitted from sender to receiver (in information theory, such length defines the 'size' of the communication process) and m is the number of channels. Finally, $\mathbf{G}$ may represent the adjacency matrix of a *bipartite graph*, where n is the number of nodes in the layer of interest (e.g. people in a co-affiliation network), while m is the number of possible dimensions where nodes can co-occur (e.g. work, family, sport, friendship, etc.). In the special case $m = n$, the network can also be interpreted as a (binary or weighted) *directed unipartite graph*, i.e. one where there is a single set of n nodes that can be linked to each other via directed edges (note that, by contrast, undirected unipartite graphs are associated with a symmetric adjacency matrix, a property that we do not enforce in this paper; the nonequivalence of ensembles of binary or weighted undirected graphs with given constraints has been studied previously in [5, 7, 45, 28]).

3.2.2 Global and local constraints

In each of the examples mentioned above, the 'true' microscopic configuration (or *microstate*) of the system can be uniquely represented by a specific 'empirical' matrix $\mathbf{G}^*$ in the set $\mathcal{G}$ of all possible states. A schematic illustration is shown in Fig. 3.1. As ordinary in statistical mechanics, when the size of the system is large one no longer focuses on the specific microstate $\mathbf{G}^*$ (which becomes not empirically accessible), but rather on the *macrostate* defined by a collection of microscopic configurations compatible with the empirical value $\vec{C}^* \equiv \vec{C}(\mathbf{G}^*)$ of a certain observable quantify $\vec{C}(\mathbf{G})$ playing the role of a constraint. The choice of $\vec{C}(\mathbf{G})$ determines the probability distribution $P(\mathbf{G})$ over $\mathcal{G}$ conditional on our knowledge of $\vec{C}^*$. In other words, it determines how our estimate of the microstate of the system concentrates around the compatible configurations once we observe $\vec{C}^*$. Intuitively, before anything is obserbed, $P(\mathbf{G})$ is uniform on $\mathcal{G}$.

In ordinary statistical physics, the quantity $\vec{C}(\mathbf{G})$ is typically scalar (e.g. the total energy) or at most low-dimensional (e.g. a vector containing the total energy and the total number of particles), reflecting a few global conservation laws applying to a large homogeneous system at thermodynamic equilibrium. However, in models of complex systems $\vec{C}(\mathbf{G})$ can be high-dimensional, as it may encode a large number of local constraints reflecting separate conservation laws imposed by spatial heterogeneity

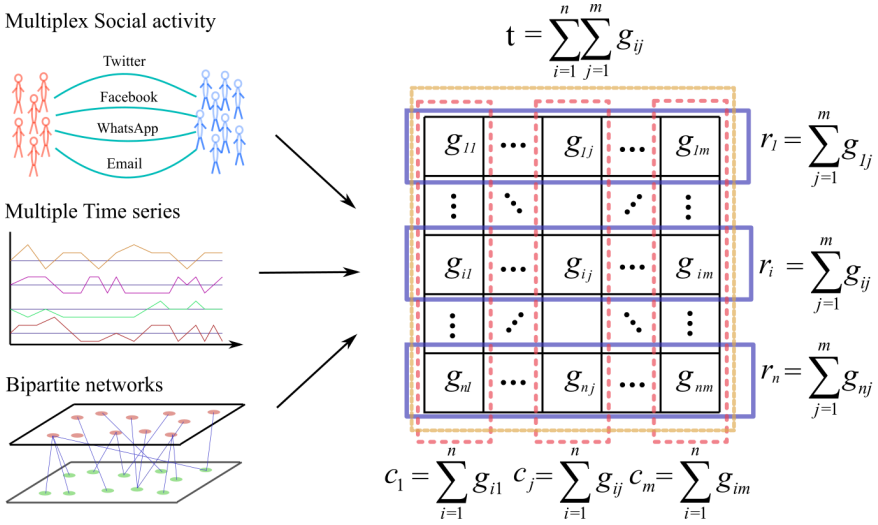

Figure 3.1. Schematic illustration of how the state of different real-world systems with n elements and m degrees of freedom can be represented in an $n \times m$ matrix $\mathbf{G}$. The collection of all possible states of the system is the set of all such matrices. Typically, real-world systems have a strong heterogeneity or nonstationarity. This empirical fact implies that their possible states are not sufficiently characterized by the knowledge of a single *global* constraint (t: solid orange box). More informative ensembles can be constructed by specifying one- ($\vec{r}$: solid blue boxes) or two-sided ($\vec{r}, \vec{c}$: solid blue and dashed red boxes) *local* constraints.

and/or temporal non-stationarity. For instance, if $\mathbf{G}^*$ is the observed configuration of a complex network with n nodes (i.e. the empirical $n \times n$ adjacency matrix), it is well known that the knowledge of purely global properties such as the overall number of links is insufficient in order to produce a statistical ensemble of networks with properties similar to those found in $\mathbf{G}^*$. Indeed, enforcing only the total number of links produces the popular Erdős-Rényi random graph model, whose topological properties are way too homogenous as compared with those of real-world networks. By contrast, if the number of links of each node is enforced separately (as in the so-called *configuration model*), the resulting ensemble of graphs is found to successfully replicate many higher-order empirical topological properties [44]. As another example, if $\mathbf{G}^*$ represents a set of synchronous time series produced by the n components of a non-stationary system observed over m time steps, then the statistical properties of these time series will change over time. As a result, overall time-independent constraints will not be enough in order to produce ensembles of multivariate time series with properties close to those of $\mathbf{G}^*$, and time-dependent (i.e. local in time) constraints will in general be needed [62].

In our setting, we consider the general case where the distribution $P(\mathbf{G})$ defining the (binary or weighted) matrix ensemble is induced by a K-dimensional vector $\vec{C}(\mathbf{G})$

of *constraints* imposed on the matrices. We will assume that the K constraints are all non-redundant, e.g. they are not trivial copies or linear combinations of each other [37]. We will consider both *global* and *local* constraints. As global constraint we will consider the scalar quantity $t(\mathbf{G})$ defined as the total sum of all the entries of the matrix $\mathbf{G}$, i.e. $t(\mathbf{G}) = \sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij}$. The number of constraints in this case is $K = 1$ and the 'empirical' value of t will be denoted as $t^* \equiv t(\mathbf{G}^*)$. As local constraints, we will consider two possibilities: *one-sided* local constraints and *two-sided* local constraints. A one-sided local constraint is the n-dimensional vector $\vec{r}(\mathbf{G})$ where the entry $r_i(\mathbf{G}) = \sum_{j=1}^{m} g_{ij}$ $(i = 1, n)$ represents the sum of the entries of the matrix $\mathbf{G}$ along its i-th row. The number of constraints is in this case $K = n$ and the empirical value of $\vec{r}$ will be denoted as $\vec{r}^* \equiv \vec{r}(\mathbf{G}^*)$. A two-sided local contraint is a pair of vectors $(\vec{r}(\mathbf{G}), \vec{c}(\mathbf{G}))$, where $\vec{r}(\mathbf{G})$ is still the n-dimensional vector representing the n row sums of $\mathbf{G}$, while $\vec{c}(\mathbf{G})$ is the m-dimensional vector representing the m column sums of $\mathbf{G}$, i.e. where each entry $c_j(\mathbf{G}) = \sum_{i=1}^{n} g_{ij}$ $(j = 1, m)$ is the sum of the entries of $\mathbf{G}$ along its j-th column. The number of constraints is therefore $K = n + m$ and the empirical value of the pair $(\vec{r}, \vec{c})$ will be denoted as $(\vec{r}^*, \vec{c}^*) \equiv (\vec{r}(\mathbf{G}^*), \vec{c}(\mathbf{G}^*))$. A visual illustration of these constraints for possible data structures of practical interest is shown in Fig. 3.1.

Purely global constraints lead to completely homogeneous expectations for the entries of the matrices in the ensemble. This result follows intuitively from symmetry arguments, and will be confirmed explicitly in the specific cases considered later. By contrast, local constraints lead to different expectations for entries in different rows and/or columns. Since, as we mentioned above, real-world complex systems are generally very heterogeneous in space and/or time, the only models that can capture the main features of such systems are those constructed from (one- or two-sided) local constraints. This is very important because, as we will show, it is precisely in presence of local constraints (of either type) that the property of EE breaks down. This result implies that spatial heterogeneity and/or temporal non-stationarity might be natural origins for the breaking of EE.

3.2.3 Soft constraints: canonical ensemble

Any constraint, whether global or local, can be enforced either as a soft constraint (canonical ensemble) or as a hard constraint (microcanonical ensemble). We start with the case of soft constraints, i.e. when one imposes that the ensemble average

$$\langle \vec{C} \rangle \equiv \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \vec{C}(\mathbf{G}) \tag{3.1}$$

is fixed to a specific value $\vec{C}^*$.

The *functional form* of the resulting canonical probability P_{can} over $\mathcal{G}$ is found by maximizing Shannon's entropy functional

$$S_n[P] \equiv -\sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \ln P(\mathbf{G}) \tag{3.2}$$

(where the subscript n indicates that the entropy is calculated for given n), subject to the condition $\langle \vec{C} \rangle = \vec{C}^*$. The result [15] of this constrained maximization problem is the parametric solution

$$P_{\rm can}(\mathbf{G}|\vec{\theta}) = \frac{e^{-H(\mathbf{G},\vec{\theta})}}{Z(\vec{\theta})}, \tag{3.3}$$

where $\vec{\theta}$ is a vector of Lagrange multipliers coupled to the constraint $\vec{C}$, the *Hamiltonian* $H(\mathbf{G},\vec{\theta}) = \vec{\theta} \cdot \vec{C}(\mathbf{G})$ is a linear combination of the constraints, and the *partition function* $Z(\vec{\theta}) = \sum_{\mathbf{G}\in\mathcal{G}} e^{-H(\mathbf{G},\vec{\theta})}$ is the normalization constant.

The *numerical values* of the canonical probability are found by setting

$$P^*_{\rm can}(\mathbf{G}) \equiv P_{\rm can}(\mathbf{G}|\vec{\theta}^*) \tag{3.4}$$

where $\vec{\theta}^*$ is the unique parameter value that realizes the 'soft' constraint

$$\langle \vec{C} \rangle_{\vec{\theta}^*} = \vec{C}^* \tag{3.5}$$

where the symbol $\langle \cdot \rangle_{\vec{\theta}}$ denotes an ensemble average with respect to $P_{\rm can}(\mathbf{G}|\vec{\theta})$, i.e.

$$\langle \vec{C} \rangle_{\vec{\theta}} = \sum_{\mathbf{G}\in\mathcal{G}} P_{\rm can}(\mathbf{G}|\vec{\theta})\, \vec{C}(\mathbf{G}). \tag{3.6}$$

Equivalently, the unique value $\vec{\theta}^*$ is the one that maximizes the log-likelihood function

$$\mathcal{L}^*(\vec{\theta}) \equiv \ln P_{\rm can}(\mathbf{G}^*|\vec{\theta}), \tag{3.7}$$

where $\mathbf{G}^*$ is the empirical configuration, or equivalently any configuration that realizes the empirical constraint exactly, i.e. such that $\vec{C}(\mathbf{G}^*) = \vec{C}^*$ [15]. The uniqueness of $\vec{\theta}^*$ follows whenever $\mathcal{L}^*(\vec{\theta})$ can be differentiated at least twice [37], as we confirm below for all the models considered in this paper.

Inserting Eq. (3.4) into Eq. (3.2), we obtain the value of the canonical entropy

$$S^*_{\rm can} \equiv S_n[P^*_{\rm can}] = -\mathcal{L}^*(\vec{\theta}^*) = -\ln P^*_{\rm can}(\mathbf{G}^*) \tag{3.8}$$

where we have omitted the dependence on n to simplify the notation. The last equality is very useful, as it show that $S^*_{\rm can}$ can be calculated by simply evaluating $P^*_{\rm can}(\mathbf{G})$ on the single configuration $\mathbf{G}^*$ [37].

3.2.4 Hard constraints: microcanonical ensemble

In the case of hard constraints, one requires that each individual configuration realizes the value $\vec{C}^*$. This means that the 'soft' constraint in Eq. (3.5) is replaced by the much stricter constraint

$$\vec{C}(\mathbf{G}) = \vec{C}^* \tag{3.9}$$

for each allowed configuration $\mathbf{G}$. The microcanonical probability $P_{\rm mic}$ is found by enforcing this stronger requirement, while still maximizing the entropy $S_n[P]$ defined in Eq. (3.2). The result is the uniform distribution

$$P^*_{\rm mic}(\mathbf{G}) = \begin{cases} \Omega^{-1}_{\vec{C}^*} & \vec{C}(\mathbf{G}) = \vec{C}^* \\ 0 & \vec{C}(\mathbf{G}) \neq \vec{C}^* \end{cases} , \tag{3.10}$$

where $\Omega_{\vec{C}^*}$ is the number of configurations in $\mathcal{G}$ realizing the 'hard' constraint in Eq. (3.9). The corresponding microcanonical entropy is obtained by inserting Eq. (3.10) into Eq. (3.2):

$$S^*_{\rm mic} \equiv S_n[P^*_{\rm mic}] = \ln \Omega_{\vec{C}^*}, \tag{3.11}$$

which is also known as Boltzmann entropy.

Crucially, in order to define the microcanonical ensemble it is necessary that $\Omega_{\vec{C}^*} > 0$, i.e. that there is at least one configuration realizing the constraint. In other words, the value of $\vec{C}^*$ should be realizable in (at least) one single configuration, and not only as an ensemble value. This requirement is not strictly necessary for the canonical ensemble (even though our interpretation of $\vec{C}^*$ as the 'empirical' value makes the requirement always natural). In any case, since in this paper we are going to study the (non)equivalence between the two ensembles, we need *both* of them to be well defined in order to be compared, for a given value of $\vec{C}^*$. Therefore we are going to assume that *the value of $\vec{C}^*$, irrespective of the ensemble considered, is always realizable by at least one configuration, i.e. such that $\Omega_{\vec{C}^*} > 0$.*

Notably, calculating $\Omega_{\vec{C}^*}$ (especially in presence of many constraints and because of the discrete nature of the problem of interest for us) can be a complicated enumeration problem. Therefore the microcanonical ensemble is typically much more difficult to deal with mathematically than the canonical ensemble. For this reason, if the property of EE holds, one prefers to operate in the canonical ensemble and work out its asymptotics in the limit of large system size, trusting that the result would return the correct asymptotics for the microcanonical ensemble as well. The above approach is at the core of many stardard calculations in statistical mechanics textbooks, where the property of EE is typically assumed to hold in general (at least in absence of phase transitions and long-range interactions). However, when EE breaks down, this approach will lead to mathematically incorrect results. We will study this problem in detail, for the ensembles considered, in the rest of the paper. To do so, we first need to define what we mean by *thermodynamic limit*.

3.2.5 The thermodynamic limit

We will consider the thermodynamic limit defined as $n \to +\infty$, i.e. when the size of the system diverges. However, the limit is not completely defined until we also specify how both m and $\vec{C}^*$ behave as n grows.

First of all, we consider two possibilities for the behaviour of m as n diverges:

- m remains finite as $n \to \infty$: in this case, we have $m = O(1)$ where $O(x)$ indicates a quantity that has a finite limit if divided by x as $n \to +\infty$, i.e. $O(x)$ is asymptotically of the same leading order[2] as x;

- m diverges as $n \to \infty$: in this case, for simplicity and realism we assume that m can diverge at most as fast as n, i.e. m is at most $O(n)$; it is indeed difficult to imagine a physical situation where the number m of state variables characterizing each of the n units grows faster than the number n of units themselves.

In simple words, the above assumptions mean that the number of state variables should be either (asymptotically) independent of the number n of units being added to the system (as in the case of 'intrinsic' observations, e.g. for multivariate time series) or at most proportional to n (as in the case of 'relational' observations, e.g. for networks). We will show that these two situations lead to very different asymptotic results in terms of the strength of EN. Importantly, the requirement that m grows at most proportionally to n implies that the number K of both one-sided ($K = n$) and two-sided ($K = n + m$) local constraints is always *extensive*, i.e. $K = O(n)$ which grows linearly in the size n of the system, irrespective of the behaviour of m.

A separate, equally important consideration concerns the scaling of the value of the constraint $\vec{C}^*$ in the thermodynamic limit $n \to +\infty$. Also here, we distinguish between two situations that we denote as the *sparse* and the *dense* regimes.

- We define *sparse matrices* those for which each of the m column sums (irrespective of whether such sums are chosen as constraints) is *finite* in the thermodynamic limit, i.e. $c_j^* = O(1)$ $(j = 1, \ldots, m)$. This implies that, in the canonical ensemble, the expected value of any entry g_{ij} of the matrix $\mathbf{G}$ is on average $O(1/n)$; correspondingly, in the microcanonical ensemble the allowed matrices are dominated by zeroes (whence the name 'sparse matrices'). Note that for the row sums one has $r_i^* = O(m/n)$ for all i. If m grew slower than n, these row sums would vanish as $n \to +\infty$, which would imply that asympotically no microcanonical configuration would realize the local constraints. Since we require $\Omega_{\vec{C}^*} > 0$ (see above), this means that in the sparse case we necessarily need $m = O(n)$. Consequently, $r_i^* = O(1)$, $r_i^*/m = O(1/n)$ $(i = 1, \ldots, n)$, $t^* = O(n)$, and $t^*/mn = O(1/n)$.

- By contrast, we define *dense matrices* those for which each of the m column sums (again, irrespective of whether they are chosen as constraints) *diverges* proportionally to n in the thermodynamic limit, i.e. $c_j^* = O(n)$ $(j = 1, \ldots, m)$. In the canonical ensemble, the expected value of g_{ij} is therefore $O(1)$, which makes the allowed matrices in the corresponding microcanonical ensemble 'dense'. The

[2]Note that the 'big-O' notation we use here is not always used with the same meaning throughout the literature: some authors prefer the 'big-Θ' notation $\Theta(x)$ to indicate a quantity that is of the same leading order as the argument x, and the 'big-O' notation to indicate only an upper bound for it.

row sums are now $r_i^* = O(m)$ and we have $r_i^*/m = O(1)$ (for all i) and $t^*/mn = O(1)$. In this case, we consider m as either remaining finite, in which case we have $m = O(1)$, $r_i^* = O(1)$ $(i = 1, \ldots, n)$ and $t^* = O(n)$, or diverging proportionally to n (see above), in which case we have $m = O(n)$, $r_i^* = O(n)$ $(i = 1, \ldots, n)$, and $t^* = O(n^2)$.

- Note that, in principle, in the weighted case we may even consider a sort of *superdense* regime where some of the individual entries of the matrix diverge in the thermodynamic limit. This possibility is related to a Bose-Einstein condensation concentrating a finite fraction of the total weight t^* of the matrix in a finite number of entries [28]. However, we will not consider this extreme case here for simplicity, as it would not arise in most real-world applications.

Combined with the scaling of $\vec{C}^*$, the behaviour of m as a function of n in the thermodynamic limit can determine different asymptotic regimes, and in particular lead to the weak or strong form of EN. The strong form, for the cases considered below, turns out to be possible in the regime where the matrices are *dense* and m is *finite* as $n \to +\infty$.

3.2.6 Ensemble (non)equivalence

There are various ways to mathematically define the property of ensemble (non)equivalence. These include the notions of EE in the *thermodynamic*, *macrostate* and *measure* sense which, under mild assumptions, can be proven to be equivalent [8]. We will adopt the definition in the measure sense, which states that the ensembles are equivalent if the relative entropy

$$S_n[P^*_{\text{mic}}||P^*_{\text{can}}] \equiv \sum_{\mathbf{G}\in\mathcal{G}} P^*_{\text{mic}}(\mathbf{G}) \ln \frac{P^*_{\text{mic}}(\mathbf{G})}{P^*_{\text{can}}(\mathbf{G})} \tag{3.12}$$

(which is the Kullback-Leibler divergence for given n between the microcanonical and canonical entropies and is guaranteed to be non-negative [52]), when rescaled by n, vanishes in the thermodynamic limit [8], i.e. if the *specific relative entropy* vanishes:

$$s[P^*_{\text{mic}}||P^*_{\text{can}}] \equiv \lim_{n\to+\infty} \frac{S_n[P^*_{\text{mic}}||P^*_{\text{can}}]}{n} = 0 \tag{3.13}$$

or equivalently

$$S_n[P^*_{\text{mic}}||P^*_{\text{can}}] = o(n), \tag{3.14}$$

where $o(x)$ indicates a quantity that goes to zero when divided by x as $n \to +\infty$.

Importantly, it can be shown [5, 37] that

$$S_n[P^*_{\text{mic}}||P^*_{\text{can}}] = \ln \frac{P^*_{\text{mic}}(\mathbf{G}^*)}{P^*_{\text{can}}(\mathbf{G}^*)} = S^*_{\text{can}} - S^*_{\text{mic}}. \tag{3.15}$$

The inequality $S_n[P^*_{\text{mic}}||P^*_{\text{can}}] \geq 0$, which is a general property of the relative entropy, implies therefore $S^*_{\text{can}} \geq S^*_{\text{mic}}$ and indicates the presence of an 'extra entropy' in

the canonical ensemble. This extra entropy is due to the fact that, while in the microcanonical ensemble the constraint $\vec{C}$ is a deterministic quantity fixed to the value $\vec{C}^*$ through the hard constraint introduced in Eq. (3.9), in the canonical ensemble it is a random variable fluctuating around the expected value $\vec{C}^*$ as dictated by the soft constraint defined in Eq. (3.5). With respect to the canonical ensemble, the hardness of the constraint in the microcanonical ensemble implies additional dependencies (i.e. smaller entropy) among the entries of $\mathbf{G}$. The definition of EE in Eq. (3.13) states that, if the extra entropy $S_n[P^*_{\text{mic}}||P^*_{\text{can}}]$, once divided by n, vanishes in the thermodynamic limit, then the ensembles are equivalent.

From Eq. (3.15) it is clear that Eq. (3.13) is equivalent to the condition

$$\lim_{n\to+\infty} \frac{S_n[P^*_{\text{can}}] - S_n[P^*_{\text{mic}}]}{n} = 0 \tag{3.16}$$

or in other words to the asymptotic (for n large) relation

$$S^*_{\text{mic}} = S^*_{\text{can}} - o(n). \tag{3.17}$$

This implies

$$\Omega_{\vec{C}^*} = e^{S^*_{\text{can}} - o(n)}, \tag{3.18}$$

i.e. $\Omega_{\vec{C}^*}$ is approximated by $e^{S^*_{\text{can}}}$ up to a subexponential (in n) correction factor. The above asymptotics is used in statistical mechanics textbooks whenever the property of EE is believed to hold, i.e. in absence of phase transitions or long-range interactions. When EE does not hold, Eq. (3.18) breaks down. In this case, the extra entropy in the canonical ensemble grows at least as fast as n. Recent research has shown that this breakdown can happen even in complete absence of phase transitions, hence also in situations where EE was typically believed to hold. Here we are going to show that, additionally, the breakdown can occur with previously undocumented strength, i.e. the extra entropy can grow as fast as the entropy itself.

Combining Eqs. (3.15) and (3.11), one obtains the following exact generalization of Eq. (3.18), valid irrespective of whether EE holds:

$$\Omega_{\vec{C}^*} = e^{S^*_{\text{can}} - S_n[P^*_{\text{mic}}||P^*_{\text{can}}]}. \tag{3.19}$$

Clearly, the above expression reduces to Eq. (3.18) in case of EE, i.e. when Eq. (3.14) holds. Although exact, Eq. (3.19) is not very useful unless one can calculate $S_n[P^*_{\text{mic}}||P^*_{\text{can}}]$ explicitly. An equivalent exact expression, which only requires the knowledge of P^*_{can} and is again valid even when EE does not hold, has been derived [37]:

$$\begin{aligned} \Omega_{\vec{C}^*} &= \sum_{\mathbf{G}\in\mathcal{G}} \int_{-\vec{\pi}}^{\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} e^{\mathrm{i}\vec{\psi}[\vec{C}^* - \vec{C}(\mathbf{G})]} \\ &= \int_{-\vec{\pi}}^{+\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} P^{-1}_{\text{can}}(\mathbf{G}^*|\vec{\theta}^* + \mathrm{i}\vec{\psi}) \end{aligned} \tag{3.20}$$

(where $\int_{-\vec{\pi}}^{+\vec{\pi}} \mathrm{d}\vec{\psi} \equiv \prod_{k=1}^{K} \int_{-\pi}^{+\pi} \mathrm{d}\psi_k$). We will confirm that the above expression provides the exact result in cases where the complex integral can be calculated explicitly and $\Omega_{\vec{C}^*}$ can be evaluated independently via combinatorial enumeration. Indeed, Eq. (3.20) highlights a beautiful connection between canonical and microcanonical probabilities through an extension to complex numbers.

When the integral in Eq. (3.20) cannot be calculated directly, it is still possible to use a saddle-point technique leading to [37]

$$\begin{aligned} \Omega_{\vec{C}^*} &= \frac{e^{S^*_{\mathrm{can}}}}{\sqrt{\det(2\pi\boldsymbol{\Sigma}^*)}} \prod_{k=1}^{K} [1 + O(1/\lambda_k^*)] \\ &= e^{S^*_{\mathrm{can}}} \prod_{k=1}^{K} \frac{1 + O(1/\lambda_k^*)}{\sqrt{2\pi\lambda_k^*}} \end{aligned} \tag{3.21}$$

where $\boldsymbol{\Sigma}^*$ is the covariance matrix of the K constraints in the canonical ensemble, whose entries are defined as

$$\Sigma^*_{ij} = \Sigma_{ij}\big|_{\vec{\theta}=\vec{\theta}^*} \tag{3.22}$$

with

$$\begin{aligned} \Sigma_{ij} &\equiv -\frac{\partial^2 \mathcal{L}^*(\vec{\theta})}{\partial\theta_i \partial\theta_j} \\ &= \frac{\partial^2 \ln Z(\vec{\theta})}{\partial\theta_i \partial\theta_j} \\ &= \langle C_i C_j \rangle_{\vec{\theta}} - \langle C_i \rangle_{\vec{\theta}} \langle C_j \rangle_{\vec{\theta}} \\ &= \mathrm{Cov}_{\vec{\theta}}[C_i, C_j] \end{aligned} \tag{3.23}$$

and $\{\lambda_k^*\}_{k=1}^K$ are the eigenvalues of $\boldsymbol{\Sigma}^*$. We recall that covariance matrices are positive-semidefinite, so all their eigenvalues are non-negative. If λ_k^* is finite, then the quantity $O(1/\lambda_k^*)$ in Eq. (3.21) cannot in general be calculated explicitly, although it generates a correction that does not change the leading order of $\Omega_{\vec{C}^*}$ and S^*_{mic}. If λ_k^* is infinite (i.e., if it diverges in the thermodynamic limit), then $O(1/\lambda_k^*)$ will vanish asymptotically and we have $1 + O(1/\lambda_k^*) = 1 + o(1)$. This implies that, if all the eigenvalues of $\boldsymbol{\Sigma}^*$ diverge, then Eq. (3.21), when inserted into certain expressions, can lead to an exact result. This includes the case of local constraints, for which K diverges in the thermodynamic limit. We will therefore discuss the asymptotic behaviour of the eigenvalues of $\boldsymbol{\Sigma}^*$ in each of the examples considered later.

Equation (3.21) generalizes Eq. (3.18) to the case where EE does not necessarily hold. Note that our initial assumption that the K constraints are non-redundant implies that $\lambda_k^* > 0$ for all k, i.e. $\boldsymbol{\Sigma}^*$ is positive-definite [37]. Keeping this assumption also in the thermodynamic limit (as ensured by our choice of both global and local constraints defined above), we note two consequences. First, since Eq. (3.23) shows that $\boldsymbol{\Sigma}^*$ is the Hessian matrix of second derivatives of $-\mathcal{L}^*(\vec{\theta})$, the fact that $\boldsymbol{\Sigma}^*$ is

positive-definite implies that $\vec{\theta}^*$ is a unique global maximum for $\mathcal{L}^*(\vec{\theta})$ [37], confirming what we had anticipated previously. Second, the product in Eq. (3.21) is at most of the same order as the denominator. Therefore, in full generality, we can exploit Eq. (3.21) to rewrite Eq. (3.19) as

$$\Omega_{\vec{C}^*} = e^{S^*_{\rm can} - O(\alpha_n)}, \tag{3.24}$$

where we have defined [37]

$$\alpha_n \equiv \ln \sqrt{\det(2\pi \mathbf{\Sigma}^*)} = \frac{1}{2} \sum_{k=1}^{K} \ln(2\pi \lambda_k^*). \tag{3.25}$$

We can now make three important considerations. First, Eq. (3.24) means that

$$S_n[P^*_{\rm mic}||P^*_{\rm can}] = O(\alpha_n), \tag{3.26}$$

showing that the speed of growth of $S_n[P^*_{\rm mic}||P^*_{\rm can}]$ with n can be calculated explicitly through Eq. (3.25) using the knowledge of $\mathbf{\Sigma}^*$, which requires only the canonical ensemble. This is useful when microcanonical calculations are unfeasible. Second, if K is finite, or if K diverges but all (except possibly a finite number of) the eigenvalues of $\mathbf{\Sigma}^*$ diverge, then the product inside Eq. (3.21) gives a subleading contribution to $S_n[P^*_{\rm mic}||P^*_{\rm can}]$, which therefore has the same asymptotic behaviour as α_n:

$$S_n[P^*_{\rm mic}||P^*_{\rm can}] = \alpha_n[1 + o(1)]. \tag{3.27}$$

This result, which is stronger than Eq. (3.26), means that in such a case one can obtain exact estimates of quantities that depend on $S_n[P^*_{\rm mic}||P^*_{\rm can}]$, using only the knowledge of α_n. Third, Eq. (3.26) shows that the definition of EE given by Eq. (3.14) coincides with

$$\alpha_n = o(n) \tag{3.28}$$

which, again, can be ascertained by evaluating only $\mathbf{\Sigma}^*$ and avoiding any microcanonical calculation. Indeed, Eq. (3.28) can be formulated as an equivalent definition of EE in the measure sense [37]. If α_n grows faster than $o(n)$, then the system is under EN.

3.3 Weak and strong ensemble nonequivalence

In this section we illustrate the main results, i.e. we identify systems for which the breaking of EE occurs in a form that is at the same time 'strong' and 'unrestricted' and we calculate the relative entropy in various such systems. To this end, we first make some general considerations leading to a rigorous definition of 'strong' EN and subsequently study specific examples within our matrix ensembles.

3.3.1 Relative entropy ratio

Equation (3.21) reveals that the asymptotic behaviour of $\Omega_{\vec{C}*}$ depends on that of K and of the eigenvalues of the covariance matrix $\mathbf{\Sigma}^*$. We can indeed convince ourselves of this fact by looking at results of previous studies from a novel perspective.

Specifically, if $K = o(n)$ and if we exclude phase transitions, then Eq. (3.21) leads to Eq. (3.18), i.e. the ensembles are equivalent. This includes the traditional situation where one has a finite number of constraints, as well as more complicated cases where the number of constraints is subextensive (e.g. random graphs with constraints on a subextensive subset of node degrees [63]). In order to break EE in this case, one needs phase transitions corresponding to singularities of the partition function [8]. For instance, in the case of graphs with fixed numbers of edges and triangles (or wedges) [67], there is a region in parameter space where one gets $S_n[P^*_{\text{mic}}||P^*_{\text{can}}] = O(n^2)$ and therefore $\Omega_{\vec{C}*} = e^{S^*_{\text{can}} - O(n^2)}$. Since also S^*_{can} and S^*_{mic} are $O(n^2)$ in this case, it follows that

$$S_n[P^*_{\text{mic}}||P^*_{\text{can}}] = O(S^*_{\text{can}}) \tag{3.29}$$

(note that in general $S^*_{\text{can}} \geq S^*_{\text{mic}}$ due to the non-negativity of the Kullback-Leibler divergence and to Eq. (3.15), therefore $O(S^*_{\text{can}})$ is necessarily the leading order). This is what we have previously referred to as a form of EN that is 'restricted' (i.e. valid only in a certain region in parameter space arising from a phase transition and outside which EE is restored) but 'strong' (i.e. where the relative entropy is of the same order as the entropy itself).

If $K = O(n)$, then Eq. (3.18) is in general no longer valid. For instance, in the case of *sparse* graphs with fixed degrees ($K = n$), all the eigenvalues of $\mathbf{\Sigma}^*$ are finite in the thermodynamic limit [5, 7]; one indeed obtains $S_n[P^*_{\text{mic}}||P^*_{\text{can}}] = O(n)$ [5] and hence $\Omega_{\vec{C}*} = e^{S^*_{\text{can}} - O(n)}$. Note that in this case the product in Eq. (3.21) (which in general cannot be calculated exactly) is of the same order as the denominator and should be taken into account. In the case of *dense* graphs with fixed degrees (again $K = n$), all the eigenvalues of $\mathbf{\Sigma}^*$ are instead $O(n)$ [7]; one indeed obtains $S_n[P^*_{\text{mic}}||P^*_{\text{can}}] = O(n \ln n)$ [7] and hence $\Omega_{\vec{C}*} = e^{S^*_{\text{can}} - O(n \ln n)}$. The product in Eq. (3.21) is in this case negliglible with respect to the denominator, which can be calculated exactly. In any case, since S^*_{can} and S^*_{mic} are still $O(n^2)$ for both sparse and dense networks with fixed degrees, these situations correspond to

$$S_n[P^*_{\text{mic}}||P^*_{\text{can}}] = o(S^*_{\text{can}}), \tag{3.30}$$

i.e. to what we have defined 'weak' EN. On the other hand, this type of EN is not associated with phase transitions (which are indeed absent in the mentioned examples of graphs with fixed degrees) and is therefore 'unrestricted', i.e. valid in the entire parameter space.

The above considerations suggest that, in order to rigorously define the strength of EN, we may define the ratio

$$R_n \equiv \frac{S_n[P^*_{\text{mic}}||P^*_{\text{can}}]}{S_n[P^*_{\text{can}}]} = 1 - \frac{S_n[P^*_{\text{mic}}]}{S_n[P^*_{\text{can}}]} \tag{3.31}$$

between the relative entropy and the canonical entropy, calculated for fixed n, and consider its limit as $n \to +\infty$, i.e.

$$\begin{aligned} R_\infty &\equiv \lim_{n\to\infty} R_n \\ &= \lim_{n\to\infty} \frac{S_n[P^*_{\rm mic}||P^*_{\rm can}]}{S_n[P^*_{\rm can}]} \\ &= 1 - \lim_{n\to\infty} \frac{S_n[P^*_{\rm mic}]}{S_n[P^*_{\rm can}]}. \end{aligned} \tag{3.32}$$

For brevity, we will call R_n the *relative entropy ratio* and R_∞ the *limiting relative entropy ratio*. Note that the inequality $S^*_{\rm can} \geq S^*_{\rm mic} \geq 0$ implies $0 \leq R_n \leq 1$ for all $n > 0$. The condition characterizing our notion of strong EN in Eq. (3.29) coincides with R_∞ being strictly positive. The value of R_∞ in that case quantifies exactly the asymptotic proportionality between $S_n[P^*_{\rm mic}||P^*_{\rm can}]$ and $S_n[P^*_{\rm can}]$, which is otherwise left unquantified by Eq. (3.29) alone. We will therefore adopt the strict inequality

$$R_\infty > 0 \tag{3.33}$$

(which in turns implies the breakdown of Eq. (3.28), the converse being in general not true) as our definition of *strong EN*. By contrast, the condition characterizing our notion of weak EN in Eq. (3.30) can be rephrased as the equality $R_\infty = 0$. Note that one may have $R_\infty = 0$ also in cases where the ensembles are equivalent. We will therefore adopt the condition $R_\infty = 0$, *in conjunction with the violation of Eq.* (3.28), as our definition of *weak EN*. Note that our discussion following Eq. (3.21) implies that, if all but at most a finite number of the eigenvalues of $\mathbf{\Sigma}^*$ diverge, then the exact value of R_∞ can be retrieved by replacing $S_n[P^*_{\rm mic}||P^*_{\rm can}]$ with α_n given by Eq. (3.25), i.e. using only the canonical covariances between the constraints, without microcanonical calculations.

Note that Eq. (3.19) implies

$$\Omega_{\vec{C}^*} = e^{S^*_{\rm can}(1-R_n)} = O\left(\left(e^{S^*_{\rm can}}\right)^{1-R_\infty}\right). \tag{3.34}$$

So, in presence of strong nonequivalence ($R_\infty > 0$), $\Omega_{\vec{C}^*}$ is of strictly smaller order compared with the ordinary estimate in Eq. (3.18). This is actually due to the canonical ensemble having much bigger entropy than the microcanonical one: indeed, Eq. (3.15) implies

$$S^*_{\rm mic} = S^*_{\rm can}(1 - R_n) \tag{3.35}$$

and, inverting,

$$S^*_{\rm can} = \frac{1}{1 - R_n} S^*_{\rm mic}. \tag{3.36}$$

Note that the factor $1/(1 - R_n)$ can be arbitrarily large since R_n can be arbitrarily close to 1.

Given the above definitions of 'weak' and 'strong' EN in terms of the limiting relative entropy ratio, in what follows we will consider the specific ensembles of matrices introduced in the previous section, under both global and local constraints, and calculate the value of α_n and R_∞ in each case.

3.3.2 Global constraints

As already discussed, ensembles of (binary or weighted) $n \times m$ matrices with a global constraint are defined by requiring that the single quantity $t(\mathbf{G}) = \sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij}$ takes, either 'hardly' or 'softly', a specific value $t^* \equiv t(\mathbf{G}^*)$. For this simple choice of the constraint, both $S^*_{\rm can}$ and $S^*_{\rm mic}$ can be calculated exactly. This allows us to check that the complex integral in Eq. (3.20) indeed provides the exact value of Ω_{t^*}. Moreover, we can confirm the correctness of the asymptotic formula in Eq. (3.21). All these approaches show that for both binary and weighted matrices with a global constraint the canonical and microcanonical ensembles are equivalent.

Binary matrices under a global constraint

Let us start with the case when the global constraint t^* is imposed on binary matrix ensembles characterized by $g_{ij} \in \{0, 1\}$. The calculation of the canonical entropy $S^*_{\rm can}$ is straightforward (see Appendix) by first calculating the likelihood

$$P_{\rm can}(\mathbf{G}^*|\theta) = \frac{e^{-\theta\, t^*}}{(1 + e^{-\theta})^{mn}} \tag{3.37}$$

and then looking for the value θ^* that maximizes $P_{\rm can}(\mathbf{G}^*|\theta)$ or, equivalently, realizes the soft constraint $\langle t \rangle_{\theta^*} = t^*$. The result is

$$\theta^* = \ln \frac{mn - t^*}{t^*}. \tag{3.38}$$

Using Eq. (3.8), we can then easily evaluate $S^*_{\rm can}$ from Eqs. (3.37) and (3.38) as

$$S^*_{\rm can} = -\ln P_{\rm can}(\mathbf{G}^*|\theta^*) = \ln \frac{(mn)^{mn}}{(t^*)^{t^*}\,(mn - t^*)^{mn-t^*}}. \tag{3.39}$$

The calculation of the microcanonical entropy $S^*_{\rm mic}$ is in this case even simpler than that of the canonical one, since the number Ω_{t^*} of configurations realizing the hard constraint $t(\mathbf{G}) = t^*$ is simply the number of ways in which t^* 'ones' can be placed in mn available positions, i.e. the binomial coefficient $\Omega_{t^*} = \binom{mn}{t^*}$. This implies

$$S^*_{\rm mic} = \ln \Omega_{t^*} = \ln \binom{mn}{t^*}. \tag{3.40}$$

Importantly, it is possible to confirm that, upon extending the argument of the likelihood to the complex domain and calculating $P_{\rm can}(\mathbf{G}^*|\theta^* + i\psi)$, the integral formula in

Eq. (3.20) returns a value of Ω_{t^*} that produces the exact value of the microcanonical entropy S^*_{mic} given in Eq. (3.40):

$$\begin{aligned} S^*_{\text{mic}} &= \ln \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi}{2\pi} P^{-1}_{\text{can}}(\mathbf{G}^*|\theta^* + \mathrm{i}\psi) \\ &= \ln \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi}{2\pi} \frac{(1+e^{-\theta^*-\mathrm{i}\psi})^{mn}}{e^{-(\theta^*+\mathrm{i}\psi)\,t^*}} \\ &= \ln \binom{mn}{t^*}, \end{aligned} \tag{3.41}$$

where the (instructive) calculation justifying the last equality is reported in the Appendix.

Combining the expressions for S^*_{mic} and S^*_{can} into Eq. (3.15), we obtain the relative entropy between the two ensembles:

$$S_n[P^*_{\text{mic}}||P^*_{\text{can}}] = \ln \frac{(mn)^{mn}}{\binom{mn}{t^*}(t^*)^{t^*}(mn-t^*)^{mn-t^*}}. \tag{3.42}$$

In this simple example, the inequality $S_n[P^*_{\text{mic}}||P^*_{\text{can}}] > 0$ clearly arises from the presence of dependencies among the entries of $\mathbf{G}$ in the microcanonical ensemble and the absence of such dependencies in the canonical one. Indeed, while in the microcanonical ensemble the hard constraint $t(\mathbf{G}) = t^*$ makes all the entries of $\mathbf{G}$ mutually dependent, in the canonical ensemble the soft constraint $\langle t \rangle_{\theta^*} = t^*$ leaves each entry g_{ij} independent and identically (Bernoulli-)distributed with probability

$$p(g_{ij}|\theta^*) = \frac{e^{-\theta^* g_{ij}}}{1+e^{-\theta^*}}, \quad g_{ij} \in \{0,1\} \tag{3.43}$$

(see Appendix). Consequently, while in the microcanonical ensemble the constraint $t(\mathbf{G})$ is a deterministic quantity fixed to the value t^*, in the canonical ensemble $t(\mathbf{G})$ is a random variable with expected value t^* and variance

$$\Sigma^* = \mathrm{Var}_{\theta^*}[t] = nm\frac{e^{-\theta^*}}{(1+e^{-\theta^*})^2} = t^*\left(1 - \frac{t^*}{mn}\right) \tag{3.44}$$

(see Appendix), where Σ^* is the only (recall that here $K = 1$) of the covariance matrix $\mathbf{\Sigma}^*$ introduced in Eq. (3.23).

As discussed in Subsection 3.2.6, $S_n[P^*_{\text{mic}}||P^*_{\text{can}}]$ and Σ^* are asymptotically related through Eq. (3.26), and the (non)equivalence of canonical and microcanonical ensembles is decided by the asymptotic behaviour of these two quantities. We will confirm both results in the particular case under consideration here. However, for compactness, we do this in conjunction with the weighted case, after introducing the latter below.

Weighted matrices under a global constraint

We now consider the case when the global constraint t^* is enforced on weighted matrices where g_{ij} is a non-negative integer. As we show in the Appendix, in the canonical ensemble the likelihood can be calculated as

$$P_{\rm can}(\mathbf{G}^*|\theta) = e^{-\theta\, t^*}(1-e^{-\theta})^{mn} \tag{3.45}$$

and is maximised by the parameter value

$$\theta^* = \ln \frac{mn+t^*}{t^*} \tag{3.46}$$

(note the change of sign with respect to the binary case) realizing the soft constraint $\langle t\rangle_{\theta^*} = t^*$. The canonical entropy is therefore

$$S^*_{\rm can} = -\ln P_{\rm can}(\mathbf{G}^*|\theta^*) = \ln \frac{(mn+t^*)^{mn+t^*}}{(t^*)^{t^*}(mn)^{mn}}. \tag{3.47}$$

In the microcanonical ensemble, the number Ω_{t^*} of configurations realizing the hard constraint $t(\mathbf{G}) = t^*$ coincides with the number of so-called *weak compositions* of the positive integer t^* into exactly mn parts, i.e. the number of ways of writing the positive integer t^* as the sum of an ordered sequence of mn non-negative integers (note that two sequences that differ in the order of their terms represent different configurations). This number is given by the *negative binomial* coefficient $\Omega_{t^*} = \binom{mn+t^*-1}{t^*}$ [68], whence

$$S^*_{\rm mic} = \ln \Omega_{t^*} = \ln \binom{mn+t^*-1}{t^*}. \tag{3.48}$$

In this case as well, one can confirm that the integration of the complex quantity $P^{-1}_{\rm can}(\mathbf{G}^*|\theta^*+i\psi)$ as specified in Eq. (3.20) produces precisely the same value of Ω_{t^*} used in Eq. (3.48) (see Appendix), thus retrieving the exact entropy

$$\begin{aligned} S^*_{\rm mic} &= \ln \int_{-\pi}^{+\pi} \frac{{\rm d}\psi}{2\pi} P^{-1}_{\rm can}(\mathbf{G}^*|\theta^*+{\rm i}\psi) \\ &= \ln \int_{-\pi}^{\pi} \frac{{\rm d}\psi}{2\pi} \frac{(1-e^{-\theta^*-{\rm i}\psi})^{-mn}}{e^{-(\theta^*+{\rm i}\psi)t^*}} \\ &= \ln \binom{mn+t^*-1}{t^*}. \end{aligned} \tag{3.49}$$

The relative entropy $S_n[P^*_{\rm mic}||P^*_{\rm can}]$, calculated using Eq. (3.15), equals

$$S_n[P^*_{\rm mic}||P^*_{\rm can}] = \ln \frac{(mn+t^*)^{mn+t^*}}{\binom{mn+t^*-1}{t^*}(t^*)^{t^*}(mn)^{mn}}. \tag{3.50}$$

Again, the origin of a non-zero relative entropy lies in the presence of dependencies among all the entries of $\mathbf{G}$ in the microcanonical ensemble, where they are coupled by the hard constraint $t(\mathbf{G}) = t^*$, and in the absence of such dependencies in the canonical ensemble, where each entry g_{ij} is independent and now *geometrically* (see Appendix) distributed with probability

$$p(g_{ij}|\theta^*) = e^{-\theta^* g_{ij}}(1 - e^{-\theta^*}), \quad g_{ij} \in \{0, 1, 2, \dots\}. \tag{3.51}$$

As a consequence, while in the microcanonical ensemble the constraint $t(\mathbf{G})$ is fixed to the constant value t^*, in the canonical ensemble it is a random variable with expected value t^* and variance

$$\Sigma^* = \mathrm{Var}_{\theta^*}[t] = nm\frac{e^{-\theta^*}}{(1 - e^{-\theta^*})^2} = t^*\left(1 + \frac{t^*}{mn}\right) \tag{3.52}$$

(see Appendix).

Ensemble equivalence for matrices under a global constraint

We can now study, in a combined fashion, the (non)equivalence of the canonical and microcanonical ensembles of both binary and weighted matrices with a global constraint t^*. To this end, we preliminarly notice that the reason why the quantity $\binom{k+l-1}{l}$ is called *negative* binomial is the fact that it can be formally rewritten as the following binomial coefficient with negative signs:

$$\binom{k+l-1}{l} = (-1)^l\binom{-k}{l}. \tag{3.53}$$

The above relation allows us to conveniently rewrite the relative entropy for the weighted case appearing in Eq. (3.50) as

$$S_n[P^*_{\mathrm{mic}}||P^*_{\mathrm{can}}] = \ln\frac{(-mn)^{-mn}}{\binom{-mn}{t^*}(t^*)^{t^*}(-mn - t^*)^{-mn-t^*}}. \tag{3.54}$$

Upon comparison with the corresponding Eq. (3.42) valid in the binary case, we can express the relative entropy in general as

$$S^{\pm}_n[P^*_{\mathrm{mic}}||P^*_{\mathrm{can}}] = \ln\frac{(\pm mn)^{\pm mn}}{\binom{\pm mn}{t^*}(t^*)^{t^*}(\pm mn - t^*)^{\pm mn-t^*}}, \tag{3.55}$$

where the superscript "+" applies to binary matrices (note that $t^* \leq mn$ in this case) and the superscript "−" applies to weighted matrices. Note that the expression for the weighted case can be formally retrieved by changing the sign of m in the expression valid for the binary case.

As we discussed in Subsection 3.2.6, checking for (non)equivalence requires studying the asymptotic behaviour of the relative entropy. In this case, we can calculate the asymptotic behaviour of $S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}]$ explicitly from the exact expression given by Eq. (3.55). Note that, in both the sparse and dense case (see Subsection 3.2.5), t^* and mn diverge in the thermodynamic limit. We can therefore apply Stirling's formula

$$k! = \sqrt{2\pi k}\left(\frac{k}{e}\right)^k [1+o(1)] \tag{3.56}$$

to Eq. (3.55), which yields

$$S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}] = \frac{1}{2}\ln\left[2\pi t^*\left(1-\frac{t^*}{\pm mn}\right)\right][1+o(1)]\,. \tag{3.57}$$

For purely pedagogical reasons, we check that the asymptotic behaviour found above is consistent with the one we would retrieve by using the expansion in Eq. (3.21), which leads to Eq. (3.26) and reduces the problem of the calculation of $S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}]$ to that of its leading order α_n. To this end, we note that in this case the matrix $\mathbf{\Sigma}^*$, being a 1×1 matrix, coincides with its only eigenvalue

$$(\lambda^*)^{\pm} = t^*\left(1-\frac{t^*}{\pm mn}\right), \tag{3.58}$$

where we have used Eq. (3.44) for binary $(+)$ and Eq. (3.52) for weighted $(-)$ matrices. Therefore

$$\begin{aligned}\alpha_n^{\pm} &= \ln\sqrt{2\pi(\lambda^*)^{\pm}} \\ &= \frac{1}{2}\ln\left[2\pi t^*\left(1-\frac{t^*}{\pm mn}\right)\right],\end{aligned} \tag{3.59}$$

which has indeed the same leading order as Eq. (3.57), thereby confirming the correctness of the saddle-point calculation. As an even stronger result, we are under the conditions for which Eq. (3.27) holds, a relationship that can be confirmed by comparing Eqs. (3.57) and (3.59). It should also be noted that, since t^* diverges in the thermodynamic limit, so does $(\lambda^*)^{\pm}$ and Eq. (3.21) leads to

$$\Omega_{t^*}^{\pm} = \frac{e^{S_n^{\pm}[P^*_{\rm can}]}}{\sqrt{2\pi t^*\left(1-\frac{t^*}{\pm mn}\right)}}[1+o(1)], \tag{3.60}$$

which is precisely what we get by applying Eq. (3.56) to the binomial and negative binomial coefficients appearing in the exact expression for $\Omega_{t^*}^{\pm}$ in the binary and weighted case respectively.

As stated in Eq. (3.28), checking whether the ensembles are equivalent boils down to checking whether $\alpha_n = o(n)$. Note that the only effect of the asymptotic scaling

of t^* is that the quantity $1 - t^*/(\pm mn)$ in Eq. (3.59) converges to 1 in the sparse case $t^*/mn = O(1/n)$ and to a different, but still finite and positive constant in the dense case $t^*/mn = O(1)$ (see Subsection 3.2.5). Therefore in both cases we have $\alpha_n = O(\ln t^*) = O(\ln mn)$. This implies $\alpha_n = o(n)$ independently of the asymptotic behaviour of m. This result shows that, *in presence of a global constraint, both binary and weighted matrices are under EE, irrespective of the scaling of t^* and m.* This finding confirms, in a generalized setting, the result obtained for networks with a given total number of links [5]. Since EE is preserved, we avoid the calculation of the limiting relative entropy ratio R_∞ defined in Eq. (3.32) in this case.

3.3.3 One-sided local constraints

We now consider ensembles of binary and weighted $n \times m$ matrices with one-sided local constraints, i.e. under the requirement that the n-dimensional vector $\vec{r}(\mathbf{G})$ with entries $r_i(\mathbf{G}) = \sum_{j=1}^{m} g_{ij}$ $(i = 1, n)$ takes a specific value $\vec{r}^* \equiv \vec{r}(\mathbf{G}^*)$. Note that, unlike the case of global constraints, here the number of constraints is extensive. As in the case with global constraints, it turns out that both $S^*_{\rm can}$ and $S^*_{\rm mic}$ can still be calculated exactly. Therefore we can again confirm the correctness of both the exact integral formula in Eq. (3.20) and the asymptotic expansion in Eq. (3.21). Despite these extensions are mathematically straightforward, we find a deep physical difference with respect to the case with global constraints: the presence of an extensive number of local constraints implies the breaking of the equivalence of canonical and microcanonical ensembles for both binary and weighted matrices. The calculation of the limiting relative entropy ratio R_∞ allows us to quantify the strength of nonequivalence and also to identify the conditions leading to the 'strong and unrestricted' form.

Binary matrices under one-sided local constraints

Let us first examine the case when the one-sided local constraints $\vec{r}^*$ are imposed on ensembles of binary matrices. As we show in the Appendix, in the canonical ensemble the likelihood is

$$P_{\rm can}(\mathbf{G}^*|\vec{\theta}) = \frac{e^{-\vec{\theta}\cdot\vec{r}^*}}{\prod_{i=1}^{n}(1+e^{-\theta_i})^m} \tag{3.61}$$

and reaches its maximum when the parameter $\vec{\theta}$ takes the value $\vec{\theta}^*$ with entries

$$\theta_i^* = \ln\frac{m - r_i^*}{r_i^*} \qquad i = 1, n, \tag{3.62}$$

corresponding to the soft constraint $\langle\vec{r}\rangle_{\vec{\theta}^*} = \vec{r}^*$. Substituting Eq. (3.62) into Eq. (3.61), we obtain the canonical entropy as

$$\begin{aligned} S^*_{\rm can} &= -\ln P_{\rm can}(\mathbf{G}^*|\vec{\theta}^*) \\ &= \sum_{i=1}^{n} \ln\frac{m^m}{(r_i^*)^{r_i^*}(m-r_i^*)^{m-r_i^*}}. \end{aligned} \tag{3.63}$$

Let us now turn to the microcanonical ensemble. Since the constraints are only one-sided, it is immediate to realize that the number $\Omega_{\vec{r}^*}$ of configurations realizing the hard constraint $\vec{r}(\mathbf{G}) = \vec{r}^*$ is a product of row-specific binomial coefficients, so that the microcanonical entropy S^*_{mic} can still be calculated exactly as the following simple generalization of Eq. (3.40):

$$S^*_{\text{mic}} = \ln \Omega_{\vec{r}^*} = \ln \prod_{i=1}^{n} \binom{m}{r_i^*} = \sum_{i=1}^{n} \ln \binom{m}{r_i^*}. \tag{3.64}$$

For the same reason, S^*_{mic} can also be exactly retrieved by explicitly integrating the complex-valued quantity $P_{\text{can}}(\mathbf{G}^*|\vec{\theta}^* + i\vec{\psi})$ as prescribed by Eq. (3.20) (see Appendix):

$$\begin{aligned} S^*_{\text{mic}} &= \ln \int_{-\vec{\pi}}^{+\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^n} P^{-1}_{\text{can}}(\mathbf{G}^*|\vec{\theta}^* + \mathrm{i}\vec{\psi}) \\ &= \ln \prod_{i=1}^{n} \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi_i}{2\pi} \frac{(1 + e^{-\theta_i^* - \mathrm{i}\psi_i})^m}{e^{-(\theta_i^* + \mathrm{i}\psi_i) r_i^*}} \\ &= \sum_{i=1}^{n} \ln \binom{m}{r_i^*}. \end{aligned} \tag{3.65}$$

Combining the above results, we can calculate the relative entropy from Eq. (3.15) as

$$S_n[P^*_{\text{mic}} || P^*_{\text{can}}] = \sum_{i=1}^{n} \ln \frac{m^m}{\binom{m}{r_i^*} (r_i^*)^{r_i^*} (m - r_i^*)^{m - r_i^*}}. \tag{3.66}$$

The above quantity encodes the following difference between the two ensembles: in the microcanonical ensemble, the hard constraint $\vec{r}(\mathbf{G}) = \vec{r}^*$ makes all the entries in each row of $\mathbf{G}$ mutually dependent, while leaving different rows independent of each other; on the other hand, in the canonical ensemble the soft constraint $\langle \vec{r} \rangle_{\vec{\theta}^*} = \vec{r}^*$ leaves *all* entries of the matrix independent. As in the case with a global constraint, each entry g_{ij} is still Bernoulli-distributed, but now with row-specific probability

$$p(g_{ij}|\vec{\theta}^*) = \frac{e^{-\theta_i^* g_{ij}}}{1 + e^{-\theta_i^*}}, \quad g_{ij} \in \{0, 1\}, \tag{3.67}$$

as we show in the Appendix. Correspondingly, in the microcanonical ensemble $\vec{r}$ is a deterministic vector fixed to the value $\vec{r}^*$, while in the canonical ensemble it is a random vector with expected value $\vec{r}^*$. The covariance matrix $\mathbf{\Sigma}^*$ between the entries of $\vec{r}$ (i.e. between the n constraints) in the canonical ensemble is a diagonal matrix with entries

$$\Sigma^*_{ij} = \delta_{ij} r_i^* \left(1 - \frac{r_i^*}{m}\right) \tag{3.68}$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$ (see Appendix). This implies that the eigenvalues $\{\lambda_i^*\}_{i=1}^n$ of $\mathbf{\Sigma}^*$ are

$$\lambda_i^* = r_i^* \left(1 - \frac{r_i^*}{m}\right) \qquad i = 1, n. \tag{3.69}$$

Again, we are going to discuss the (non)equivalence of the two ensembles together with the corresponding case of weighted matrices, after studying the latter below.

Weighted matrices under one-sided local constraints

We now move to the case when the one-sided local constraints $\vec{r}^*$ are imposed on ensembles of weighted matrices. The canonical ensemble under such constraints is characterized by the likelihood

$$P_{\text{can}}(\mathbf{G}^*|\vec{\theta}) = \frac{e^{-\vec{\theta}\cdot\vec{r}^*}}{\prod_{i=1}^n (1 - e^{-\theta_i})^{-m}}, \tag{3.70}$$

which is maximized by the parameter value $\vec{\theta}^*$ with entries

$$\theta_i^* = \ln \frac{m + r_i^*}{r_i^*} \qquad i = 1, n \tag{3.71}$$

realizing the soft constraint $\langle \vec{r} \rangle_{\vec{\theta}^*} = \vec{r}^*$ (see Appendix). If we insert Eq. (3.71) into Eq. (3.70), we get

$$\begin{aligned} S_{\text{can}}^* &= -\ln P_{\text{can}}(\mathbf{G}^*|\vec{\theta}^*) \\ &= \sum_{i=1}^n \ln \frac{(m + r_i^*)^{m+r_i^*}}{(r_i^*)^{r_i^*} m^m}. \end{aligned} \tag{3.72}$$

The microcanonical entropy S_{mic}^* is instead given by the following generalization of Eq. (3.48):

$$S_{\text{mic}}^* = \ln \Omega_{\vec{r}^*} = \sum_{i=1}^n \ln \binom{m + r_i^* - 1}{r_i^*}, \tag{3.73}$$

where we have expressed the number $\Omega_{\vec{r}^*}$ of configurations realizing the hard constraint $\vec{r}(\mathbf{G}) = \vec{r}^*$ as a product of row-specific negative binomial coefficients. Again, the microcanonical entropy can be obtained equivalently from Eq. (3.20) as follows (see Appendix):

$$\begin{aligned} S_{\text{mic}}^* &= \ln \int_{-\vec{\pi}}^{+\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^n} P_{\text{can}}^{-1}(\mathbf{G}^*|\vec{\theta}^* + \mathrm{i}\vec{\psi}) \\ &= \ln \prod_{i=1}^n \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi_i}{2\pi} \frac{(1 - e^{-\theta_i^* - \mathrm{i}\psi_i})^{-m}}{e^{-(\theta_i^* + \mathrm{i}\psi_i) r_i^*}} \\ &= \sum_{i=1}^n \ln \binom{m + r_i^* - 1}{r_i^*}. \end{aligned} \tag{3.74}$$

The relative entropy, which can be obtained from Eq. (3.15) as usual, equals

$$S_n[P^*_{\text{mic}}||P^*_{\text{can}}] = \sum_{i=1}^{n} \ln \frac{(m+r_i^*)^{m+r_i^*}}{\binom{m+r_i^*-1}{r_i^*}(r_i^*)^{r_i^*} m^m} \tag{3.75}$$

and encodes the difference between the microcanonical ensemble, where the entries of each row of $\mathbf{G}$ are mutually coupled by the hard constraint $\vec{r}(\mathbf{G}) = \vec{r}^*$ (while different rows are independent), and the canonical ensemble, where all entries of $\mathbf{G}$ are independent and *geometrically* (see Appendix) distributed with row-dependent probability

$$p(g_{ij}|\vec{\theta}^*) = e^{-\theta_i^* g_{ij}}(1-e^{-\theta_i^*}), \quad g_{ij} \in \{0,1,2,\dots\}. \tag{3.76}$$

As a consequence, while in the microcanonical ensemble the constraint $\vec{r}$ is fixed to the value $\vec{r}^*$, in the canonical ensemble it is a random vector fluctuating around $\vec{r}^*$ according to the diagonal covariance matrix $\mathbf{\Sigma}^*$ with entries

$$\Sigma^*_{ij} = \delta_{ij} r_i^* \left(1 + \frac{r_i^*}{m}\right) \tag{3.77}$$

(see Appendix) and eigenvalues

$$\lambda_i^* = r_i^* \left(1 + \frac{r_i^*}{m}\right) \qquad i = 1, n. \tag{3.78}$$

Ensemble nonequivalence for matrices under one-sided local constraints

We can now compactly discuss the (non)equivalence of canonical and microcanonical ensembles of both binary and weighted matrices under one-sided local constraints. As for the case of global constraints discussed in Subsection 3.3.2, we still have an exact knowledge of the canonical entropy, the microcanonical entropy, and the relative entropy. Moreover, these quantities can all be written, using Eq. (3.53), in compact expressions formally valid for both binary (+) and weighted (−) matrices. Indeed, the canonical entropy can be expressed by combining the expressions for $S_n[P^*_{\text{can}}] = S^*_{\text{can}}$ in Eqs. (3.63) and (3.72) into the unified formula

$$S_n^{\pm}[P^*_{\text{can}}] = \sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*}(m \mp r_i^*)^{\pm m - r_i^*}} \tag{3.79}$$

and, similarly, the microcanonical entropy can be obtained by formally combining Eqs. (3.64) and (3.73) into

$$S_n^{\pm}[P^*_{\text{mic}}] = \sum_{i=1}^{n} \ln \left[(\pm 1)^{r_i^*} \binom{\pm m}{r_i^*}\right]. \tag{3.80}$$

The above expressions can be used to calculate the relative entropy as

$$S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}] = \sum_{i=1}^{n} \ln \frac{(\pm m)^{\pm m}}{\binom{\pm m}{r_i^*} (r_i^*)^{r_i^*} (\pm m - r_i^*)^{\pm m - r_i^*}} \tag{3.81}$$

which indeed combines the expressions given in Eq. (3.66) for binary (+) matrices and Eq. (3.75) for weighted (−) matrices. Equation (3.81) extends Eq. (3.55) to the case of one-sided local constraints. We now consider different regimes.

- In the sparse case where $r_i^* = O(1)$ (for all i) and $m = O(n)$ (see Subsection 3.2.5), we can use Stirling's formula, given by Eq. (3.56), to expand $m!$ (but not $r_i^*!$) appearing in the (negative) binomial coefficient to get

$$\binom{\pm m}{r_i^*} \approx \frac{(\pm m)^{r_i^*}}{r_i^*!} \tag{3.82}$$

  and consequently

$$S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}] \approx \sum_{i=1}^{n} \ln \frac{e^{r_i^*} r_i^*!}{(r_i^*)^{r_i^*}} = O(n). \tag{3.83}$$

- In the dense case where both r_i^* (for all i) and m are $O(n)$, as discussed in Subsection 3.2.5 (so that r_i^*/m converges to a finite constant), we can use Stirling's formula to expand both $m!$ and $r_i^*!$ into Eq. (3.81) to obtain

$$\begin{aligned} S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}] &= \frac{1}{2} \sum_{i=1}^{n} \ln \left[2\pi r_i^* \left(1 - \frac{r_i^*}{\pm m} \right) \right] [1 + o(1)] \\ &= O(n \ln n) \end{aligned} \tag{3.84}$$

  for binary (+) (in which case $r_i^* \leq m$) and weighted (−) matrices.

- In the dense case where both m and r_i^* are finite, there is no asymptotic expansion that allows to simplify Eq. (3.81) in general, so $S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}]$ has to be evaluated explicitly (simple examples are provided below). The important general consideration is that, irrespective of the specific values of m and r_i^*,

$$S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}] = O(n). \tag{3.85}$$

Again, we can confirm that the above asymptotic expressions are consistent with what we would obtain from Eq. (3.26), which follows from the saddle-point approximation given by Eq. (3.21). To see this, noting that here $K = n$ and that Eqs. (3.68) and (3.77) indicate that $(\mathbf{\Sigma}^*)^{\pm}$ is a diagonal matrix with entries

$$(\Sigma^*_{ij})^{\pm} = \delta_{ij} r_i^* \left(1 - \frac{r_i^*}{\pm m} \right) \tag{3.86}$$

for binary (+) and weighted (−) one-sided constraints respectively, we can compactly express Eq. (3.25) as

$$\begin{aligned} \alpha_n^{\pm} &= \ln \sqrt{\det\left[2\pi(\mathbf{\Sigma}^*)^{\pm}\right]} \\ &= \frac{1}{2}\sum_{i=1}^{n} \ln\left[2\pi(\Sigma_{ii}^*)^{\pm}\right] \\ &= \frac{1}{2}\sum_{i=1}^{n} \ln\left[2\pi r_i^* \left(1 - \frac{r_i^*}{\pm m}\right)\right]. \end{aligned} \tag{3.87}$$

In the dense regime with diverging m, the conditions guaranteeing the strong result in Eq. (3.27) (all but a finite number of diverging eigenvalues of $\mathbf{\Sigma}^*$) hold, as can be confirmed by comparing Eqs. (3.84) and (3.87). Moreover, noticing from Eq. (3.56) that $e^k k!/k^k = \sqrt{2\pi k}[1 + o(1)]$, we see that Eq. (3.87) confirms the asymptotic behaviour of the relative entropy obtained also in Eqs. (3.83) and (3.85) for the other two regimes, under the respective assumptions on the scaling of m and k. Coincidentally, we also see that the stronger result in Eq. (3.27) turns out to be a very good approximation for the relative entropy even in these two regimes where, technically, the required conditions are not met. This means that, for the one-sided dense case with finite m, we can rewrite Eq. (3.26) asymptotically (i.e. for large n) as

$$\begin{aligned} S_n^{\pm}[P_{\text{mic}}^* || P_{\text{can}}^*] &= C_1(m)\alpha_n^{\pm} \\ &= C_1(m) \ln \sqrt{\det[2\pi(\mathbf{\Sigma}^*)^{\pm}]} \end{aligned} \tag{3.88}$$

where $C_1(m)$ is a finite and positive constant. Moreover, from the known inequality $e^k k!/k^k \geq \sqrt{2\pi k}$ for the factorial, we see that $C_1(m) \geq 1$ as implied by comparing Eqs. (3.81) and (3.87). Finally, we also know from Stirling's approximation that $C_1(m)$ is not much bigger than 1, i.e. $C_1(m) \gtrsim 1$, and that it rapidly approaches 1: indeed when m diverges Eq. (3.27) holds exactly, which implies

$$\lim_{m\to\infty} C_1(m) = 1. \tag{3.89}$$

The fact that, in all regimes, $S_n^{\pm}[P_{\text{mic}}^* || P_{\text{can}}^*]$ (or equivalently α_n) is at least of order $O(n)$ shows that Eq. (3.28) is violated and that *EE breaks down for both binary and weighted matrices under one-sided local constraints*, irrespective of the density and of the behaviour of m. This important finding generalizes the result, documented so far only for ensembles of binary graphs with given degree sequence [5, 7, 45] (and possibly modular structure [26]) and weighted graphs with given strength sequence [28], that EE breaks down in the presence of an extensive (i.e. growing like n) number of local constraints. Here, this result is extended to more general ensembles of matrices, i.e. asymmetric, rectangular matrices describing e.g. bipartite graphs, multivariate time series, multiplex social activity, multi-cast communication systems and multi-cell gene

expression profiles with variable m. More importantly, this generalized setting allows for a qualitatively new phenomenon to emerge, namely the onset of 'strong' EN, as we now show.

Indeed, we can investigate the 'strength' of nonequivalence by comparing the asymptotic behaviour of the relative entropy with that of the canonical entropy given by Eq. (3.79). This expression can be evaluated in the usual three regimes as follows.

In the sparse case with $r_i^* = O(1)$ and $m = O(n)$, noticing that asymptotically (for large n) we have $(m \mp r_i^*)^{\pm m - r_i^*} \approx m^{\pm m - r_i^*} e^{\mp r_i^*}$, Eq. (3.79) reduces to

$$S_n^{\pm}[P_{\rm can}^*] \approx \sum_{i=1}^{n} r_i^* \ln \frac{e^{\pm 1} m}{r_i^*} = O(n \ln n), \tag{3.90}$$

which dominates over the order $O(n)$ of the corresponding relative entropy $S_n^{\pm}[P_{\rm mic}^* || P_{\rm can}^*]$ calculated previously in Eq. (3.83) for the sparse case. This implies that the limiting relative entropy ratio defined in Eq. (3.32) is $R_\infty^{\pm} = 0$ for both binary $(+)$ and weighted $(-)$ constraints, meaning that in this case the breaking of EE is still 'weak' as in the case of graphs with local constraints.

In the dense case with $r_i^* = O(n)$ and $m = O(n)$, Eq. (3.79) can be evaluated as

$$\begin{aligned} S_n^{\pm}[P_{\rm can}^*] &= \sum_{i=1}^{n} \left[\pm m \ln \frac{m}{m \mp r_i^*} + r_i^* \ln \frac{m \mp r_i^*}{r_i^*} \right] \\ &= O(n^2) \end{aligned} \tag{3.91}$$

which, again, dominates over the order $O(n \ln n)$ of the corresponding relative entropy calculated in Eq. (3.84). Therefore we still have $R_\infty^{\pm} = 0$ (weak nonequivalence).

Finally, the dense case where both m and r_i^* remain finite as $n \to \infty$ is the subject of the rest of this Section. Equation (3.79) implies that

$$S_n^{\pm}[P_{\rm can}^*] = O(n) \tag{3.92}$$

which, upon comparison with Eq. (3.85), shows that now the relative entropy grows as fast as the canonical entropy, signalling the 'strong' form of EN. Using the combined expressions given in Eqs. (3.79) and (3.80), we can explicitly calculate the relative entropy ratio introduced in Eq. (3.31) as follows:

$$R_n^{\pm} = 1 - \frac{\sum_{i=1}^{n} \ln \left[(\pm 1)^{r_i^*} \binom{\pm m}{r_i^*} \right]}{\sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}}} > 0. \tag{3.93}$$

Using Eqs. (3.87) and (3.88), we obtain the alternative asymptotic (for large n) ex-

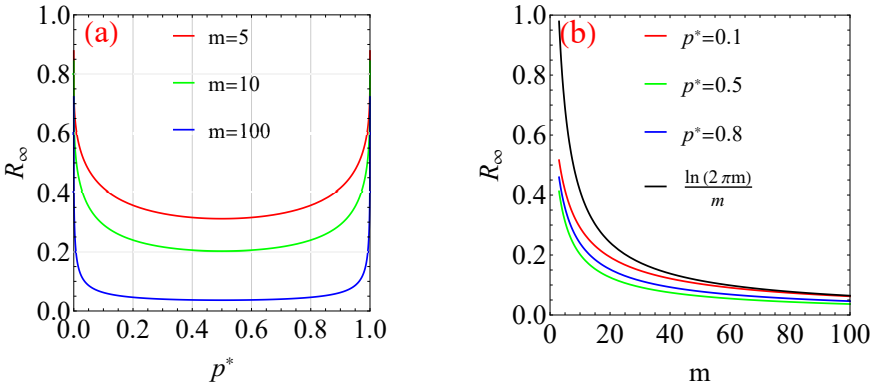

Figure 3.2. Strong ensemble nonequivalence, signalled by a positive limiting entropy ratio $R_\infty^+ > 0$, for binary matrices under homogeneous one-sided local constraints ($r_i^* = r^* \ \forall i$) in the dense case with finite m and r^*. (a) R_∞^+ as a function of $p^* = r^*/m$ for various values of m. Note that R_∞^+ is larger for smaller m and for values of p^* more distant from the uniform case ($p^* = 1/2$). (b) R_∞^+ as a function of m for various values of p^*. Note that, as m grows, R_∞^+ decays like $\ln(2\pi m)/m$.

pression

$$\begin{aligned} R_n^\pm &= \frac{S_n^\pm[P^*_{\rm mic}||P^*_{\rm can}]}{S_n^\pm[P^*_{\rm can}]} \\ &= C_1(m)\frac{\alpha_n^\pm}{\sum_{i=1}^n \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*}(m\mp r_i^*)^{\pm m - r_i^*}}} \\ &= \frac{C_1(m)}{2}\frac{\sum_{i=1}^n \ln\left[2\pi r_i^*\left(1-\frac{r_i^*}{\pm m}\right)\right]}{\sum_{i=1}^n \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*}(m\mp r_i^*)^{\pm m - r_i^*}}}. \end{aligned} \tag{3.94}$$

Comparing Eqs. (3.93) and (3.94) confirms that, as noticed above, $C_1(m) \approx 1$ also for finite m.

In general, taking the thermodynamic limit $n \to \infty$ in Eq. (3.93) or (3.94) requires the specification of the value of r_i^* for all i. For the sake of illustration, we can consider the simplest case where each constraint has the same value $r_i^* = r^*$ $(i = 1, n)$. Note that the resulting canonical entropy of matrices with constant one-sided constraint r^*, given by Eq. (3.79), coincides with the canonical entropy of matrices with the implied global constraint $t^* = nr^*$, given by Eqs. (3.39) and (3.47) in the binary and weighted case respectively. However, the microcanonical entropy in the one-sided case, given by Eq. (3.80), is strictly smaller than the corresponding one for matrices with the implied global constraint $t^* = nr^*$, given by Eqs. (3.40) and (3.48) in the binary and

weighted case respectively. From Eq. (3.93) we immediately find

$$R_\infty^\pm = \lim_{n\to\infty} R_n^\pm = 1 - \frac{\ln\left[(\pm 1)^{r^*}\binom{\pm m}{r^*}\right]}{\ln \frac{m^{\pm m}}{(r^*)^{r^*}(m\mp r^*)^{\pm m - r^*}}} > 0, \tag{3.95}$$

confirming strong nonequivalence as defined in Eq. (3.33). To gain numerical and visual insight about the behaviour of $R_\infty^\pm$ in Eq. (3.95), let us consider the binary and weighted cases separately.

In the binary case, Eq. (3.62) implies that, if $r_i^* = r^*$ for all i, then the Lagrange multipliers θ_i^* are all equal to

$$\theta_+^* \equiv \ln \frac{m - r^*}{r^*}. \tag{3.96}$$

Then, writing $p^* \equiv r^*/m = e^{-\theta_+^*}/(1 + e^{-\theta_+^*}) \in (0,1)$, from Eq. (3.95) we obtain

$$\begin{aligned} R_\infty^+ &= 1 - \frac{\ln \binom{m}{r^*}}{\ln \frac{m^m}{(r^*)^{r^*}(m-r^*)^{m-r^*}}} \\ &= 1 - \frac{\ln \binom{m}{p^* m}}{\ln \frac{m^m}{(p^* m)^{p^* m}(m - p^* m)^{m - p^* m}}} > 0. \end{aligned} \tag{3.97}$$

Using the above expression, in Fig. 3.2 we plot R_∞^+ as a function of either p^* (for fixed m) or m (for fixed p^*). We see that, for a wide range of values of p^*, R_∞^+ remains appreciably large for values of m up to one hundred. Moreover, values of p^* closer to 0 or 1 than to 1/2 make R_∞^+ larger. So, for empirical applications where the level of 'multiplexity' is moderate (i.e. small m), and especially away from the uniform case ($p^* = 1/2$), there is a significant entropy reduction from the canonical to the microcanonical ensemble. By contrast, as m increases while p^* remains fixed, R_∞^+ decreases like $\frac{\ln 2\pi m}{m}$, as can be easily realized by applying Stirling's formula to Eq. (3.97). This coincides with the system progressively moving to the different regime where both m and r_i^* grow as n grows, which results in weak EN and $R_\infty^+ = 0$ as previously noticed. Similarly, if r_i^* remains finite while m grows, we enter the sparse regime for which $R_\infty^+ = 0$ as previously noticed.

In the weighted case, Eq. (3.71) implies that if $r_i^* = r^*$ for all i, then $\theta_i^* = \theta_-^*$ for all i with

$$\theta_-^* \equiv \ln \frac{m + r^*}{r^*}. \tag{3.98}$$

Then, writing $q^* \equiv e^{-\theta_-^*} = \frac{r^*}{m + r^*} \in (0,1)$, from Eq. (3.95) we obtain

$$\begin{aligned} R_\infty^- &= 1 - \frac{\ln \binom{m + r^* - 1}{r^*}}{\ln \frac{(m+r^*)^{m+r^*}}{m^m (r^*)^{r^*}}} \\ &= 1 - \frac{\ln \binom{(m-1+q^*)/(1-q^*)}{m q^*/(1-q^*)}}{\frac{m}{1-q^*}\ln\frac{m}{1-q^*} - m\ln m - \frac{m q^*}{1-q^*}\ln\frac{m q^*}{1-q^*}} > 0. \end{aligned} \tag{3.99}$$

Figure 3.3 shows the behaviour of R_∞^- either as a function of q^* (with m fixed) or as a function of m (with q^* fixed). Considerations similar to the binary case apply. The main difference is that, while R_∞^+ has a symmetric behaviour around the value $p^* = 1/2$ (arising from the fundamental symmetry of exchanging p^* with $1 - p^*$ and $g_{ij} = 1$ with $g_{ij} = 0$ in the binary case), R_∞^- decreases monotonically as a function of q^* (due to the lack of any symmetry of that sort in the weighted case). So now R_∞^- is larger for smaller m and q^*.

The above results illustrate what we had anticipated previously, i.e. that if m is finite (and the matrices are necessarily dense) then the ensembles feature a strong form of EN. Here, this form of EN is also 'unrestricted', as it holds irrespective of the value of $\vec{C}^*$ or $\vec{\theta}^*$, i.e. throughout the parameter space. To the best of our knowledge, this is the first evidence of a situation for which EN occurs in a simultaneously 'strong and unrestricted' form, i.e. the most robust manifestation of the breaking of EE documented so far. Its ultimate origin is the presence of an extensive number of local constraints, and not of phase transitions.

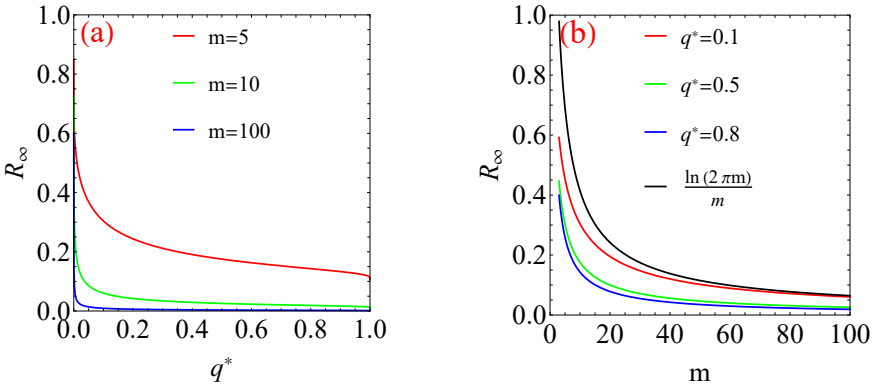

Figure 3.3. Strong ensemble nonequivalence, signalled by a positive limiting entropy ratio $R_\infty^- > 0$, for weighted matrices under homogeneous one-sided local constraints ($r_i^* = r^* \; \forall i$) in the dense case with finite m and r^*. (a) R_∞^- as a function of $q^* = \frac{r^*}{m+r^*}$ for various values of m. Note that R_∞^- is larger for smaller m and q^*. (b) R_∞^- as a function of m for various values of q^*. As in the binary case, R_∞^- decays like $\ln(2\pi m)/m$ as m grows.

3.3.4 Two-sided local constraints

We now discuss binary and weighted matrices under two-sided local constraints $(\vec{r}(\mathbf{G}), \vec{c}(\mathbf{G}))$, where $\vec{r}(\mathbf{G})$ is still the n-dimensional vector of row sums while $\vec{c}(\mathbf{G})$ is the m-dimensional vector of column sums, with entries $c_j(\mathbf{G}) = \sum_{i=1}^n g_{ij}$ $(j = 1, \ldots, m)$. We constrain both vectors to a given value $(\vec{r}^*, \vec{c}^*) \equiv (\vec{r}(\mathbf{G}^*), \vec{c}(\mathbf{G}^*))$. Note that the number of constraints is still extensive. Unlike the case of one-sided constraints, for two-sided constraints it is not possible to calculate the exact number $\Omega_{\vec{r}^*,\vec{c}^*}$ of configurations in the microcanonical ensemble. By contrast, all canonical calculations

can still be carried out analytically (although the value of the Lagrange multipliers can be determined only via implicit expressions). Therefore, as we show below, the matrix $\mathbf{\Sigma}^*$ of canonical covariances between the constraints becomes a crucial tool to calculate the asymptotic behaviour of the relevant microcanonical quantities.

Binary matrices under two-sided local constraints

As usual, we start from the binary case. As shown in the Appendix, in the canonical ensemble the likelihood is

$$P_{\rm can}(\mathbf{G}^*|\vec{\alpha},\vec{\beta}) = \frac{e^{-\vec{\alpha}\cdot\vec{r}^*-\vec{\beta}\cdot\vec{c}^*}}{\prod_{i=1}^n\prod_{j=1}^m[1+e^{-(\alpha_i+\beta_j)}]} \tag{3.100}$$

and is maximized by the parameter values $(\vec{\alpha}^*,\vec{\beta}^*)$ defined implicitly by the following set of $n+m$ coupled nonlinear equations:

$$r_i^* = \sum_{j=1}^m \frac{e^{-(\alpha_i^*+\beta_j^*)}}{1+e^{-(\alpha_i^*+\beta_j^*)}}, \quad i=1,n, \tag{3.101}$$

$$c_j^* = \sum_{i=1}^n \frac{e^{-(\alpha_i^*+\beta_j^*)}}{1+e^{-(\alpha_i^*+\beta_j^*)}}, \quad j=1,m. \tag{3.102}$$

Unfortunately, in general these equations cannot be solved analytically to express $(\vec{\alpha}^*,\vec{\beta}^*)$ as an explicit function of $(\vec{r}^*,\vec{c}^*)$. This is due to the fact that the presence of both row and column constraints couples all parameters. However, the equations can be solved numerically and the unique solution $(\vec{\alpha}^*,\vec{\beta}^*)$ can then be inserted into $P_{\rm can}(\mathbf{G}|\vec{\alpha},\vec{\beta})$. This gives complete analytical control over the canonical ensemble. In particular, the canonical entropy is

$$\begin{aligned} S_{\rm can}^* &= -\ln P_{\rm can}(\mathbf{G}^*|\vec{\alpha}^*,\vec{\beta}^*) \\ &= \vec{\alpha}^*\cdot\vec{r}^* + \vec{\beta}^*\cdot\vec{c}^* + \sum_{i=1}^n\sum_{j=1}^m \ln[1+e^{-(\alpha_i^*+\beta_j^*)}]. \end{aligned} \tag{3.103}$$

Note that, since this model has additional constraints with respect to the one-sided case with the same row sums $\vec{r}^*$, the canonical entropy above cannot be larger than the corresponding one-sided canonical entropy given by Eq. (3.63), i.e. we have the following upper bound:

$$S_{\rm can}^* \leq \sum_{i=1}^n \ln\frac{m^m}{(r_i^*)^{r_i^*}(m-r_i^*)^{m-r_i^*}}. \tag{3.104}$$

On the other hand, the microcanonical entropy $S_{\rm mic}^*$ cannot be computed analytically, although the asymptotic formulas based on Eqs. (3.24) and (3.25) can be used

to estimate it from the canonical covariance matrix $\mathbf{\Sigma}^*$. As we show later, this leads to an asymptotic estimate of the relative entropy based on Eq. (3.26). As for the canonical entropy, the microcanonical one cannot be larger than the corresponding one given by Eq. (3.64) in the one-sided case with the same row sums, with the only difference that now the resulting upper bound is tight:

$$S^*_{\text{mic}} < \sum_{i=1}^{n} \ln \binom{m}{r^*_i}. \tag{3.105}$$

Indeed, the configurations matching both the row and the column constraints in the two-sided case form a proper subset of the configurations matching only the row constraints in the one-sided case.

It should be noted that in the microcanonical ensemble both $\vec{r}$ and $\vec{c}$ are deterministic vectors fixed to the values $\vec{r}^*$ and $\vec{c}^*$ respectively, while in the canonical ensemble they are random vectors with expected values $\vec{r}^*$ and $\vec{c}^*$. In the microcanonical ensemble, the hard constraints $(\vec{r}(\mathbf{G}), \vec{c}(\mathbf{G})) = (\vec{r}^*, \vec{c}^*)$ create mutual dependencies among all the entries of $\mathbf{G}$. In the canonical ensemble, the soft constraint $(\langle\vec{r}\rangle_{\vec{\alpha}^*}, \langle\vec{c}\rangle_{\vec{\beta}^*}) = (\vec{r}^*, \vec{c}^*)$ leaves all entries of $\mathbf{G}$ independent. As in all other canonical binary ensembles considered above, each entry g_{ij} is Bernoulli-distributed, but now with its specific parameters:

$$p(g_{ij}|\vec{\alpha}^*, \vec{\beta}^*) = \frac{e^{-(\alpha^*_i+\beta^*_i)g_{ij}}}{1+e^{-(\alpha^*_i,\beta^*_i)}}, \quad g_{ij} \in \{0,1\}, \tag{3.106}$$

as shown in the Appendix.

For illustration, we consider the special case where the column sums are all equal to each other, i.e. $c^*_j = c^*$ for all j. In this case, since the corresponding Lagrange multipliers must also be all equal to each other ($\beta^*_j = \beta^*$ for all j), it is indeed possible to solve for the parameters explicitly. Indeed, Eqs. (3.101) and (3.102) reduce to the $n+1$ independent equations

$$r^*_i = m\frac{e^{-(\alpha^*_i+\beta^*)}}{1+e^{-(\alpha^*_i+\beta^*)}}, \quad i = 1, n, \tag{3.107}$$

$$c^* = \sum_{i=1}^{n} \frac{e^{-(\alpha^*_i+\beta^*)}}{1+e^{-(\alpha^*_i+\beta^*)}}, \tag{3.108}$$

where the second equation is simply the consistency condition $c^* = \sum_{i=1}^{n} r^*_i/m$ implied by the first one. This means that the parameter β^* is actually redundant, as it could in principle be reabsorbed into a shift of all the α^*_i's. In any case, the combination $\alpha^*_i + \beta^*$ is found explicitly by inverting Eq. (3.107):

$$\alpha^*_i + \beta^* = \ln \frac{m - r^*_i}{r^*_i}. \tag{3.109}$$

Note that, inserting this value into the expression for $S^*_{\rm can}$ in Eq. (3.103), we obtain exactly the canonical entropy found previously in Eq. (3.63) for the binary ensemble with one-sided (row) constraints specified by the same vector $\vec{r}^*$, i.e. the bound in Eq. (3.104) is fully saturated:

$$\begin{aligned} S^*_{\rm can} &= \vec{\alpha}^* \cdot \vec{r}^* + m\beta^* c^* + m \sum_{i=1}^{n} \ln[1 + e^{-(\alpha_i^* + \beta^*)}] \\ &= \sum_{i=1}^{n} \ln \frac{m^m}{(r_i^*)^{r_i^*} (m - r_i^*)^{m - r_i^*}}. \end{aligned} \tag{3.110}$$

Indeed, the two canonical ensembles are indistinguishable and all their properties are the same. However, the corresponding microcanonical ensembles remain very different, because in the two-sided case each of the m column sums has to match the exact value c^* separately, while in the one-sided case only the total sum mc^* of all the m column sums (which is necessarily implied by the row constraints) has to be matched exactly. Indeed Eq. (3.105) is a tight bound that cannot be saturated. Similarly, the covariance matrix $\mathbf{\Sigma}^*$ is now a $(n+m) \times (n+m)$ matrix (calculated later) and its determinant is different from the one obtained in the one-sided case, where the matrix is $n \times n$.

Again, we are going to discuss the (non)equivalence of the two ensembles together with the corresponding case of weighted matrices, after studying the latter below.

Weighted matrices under two-sided local constraints

We now discuss EN in weighted matrices with two-sided local constraints. The likelihood (see Appendix) is now

$$P_{\rm can}(\mathbf{G}^*|\vec{\alpha}, \vec{\beta}) = \frac{e^{-\vec{\alpha}\cdot\vec{r}^* - \vec{\beta}\cdot\vec{c}^*}}{\prod_{i=1}^{n} \prod_{j=1}^{m} [1 - e^{-(\alpha_i + \beta_j)}]^{-1}} \tag{3.111}$$

and is maximized by the unique parameter values $(\vec{\alpha}^*, \vec{\beta}^*)$ defined implicitly through the $n+m$ coupled nonlinear equations

$$r_i^* = \sum_{j=1}^{m} \frac{e^{-(\alpha_i^* + \beta_j^*)}}{1 - e^{-(\alpha_i^* + \beta_j^*)}}, \quad i = 1, n, \tag{3.112}$$

$$c_j^* = \sum_{i=1}^{n} \frac{e^{-(\alpha_i^* + \beta_j^*)}}{1 - e^{-(\alpha_i^* + \beta_j^*)}}, \quad j = 1, m. \tag{3.113}$$

that can be solved numerically. The solution $(\vec{\alpha}^*, \vec{\beta}^*)$, when inserted into $P_{\rm can}(\mathbf{G}|\vec{\alpha}, \vec{\beta})$, completely characterizes the canonical ensemble. The resulting canonical entropy is

$$\begin{aligned} S^*_{\rm can} &= -\ln P_{\rm can}(\mathbf{G}^*|\vec{\alpha}^*, \vec{\beta}^*) \\ &= \vec{\alpha}^* \cdot \vec{r}^* + \vec{\beta}^* \cdot \vec{c}^* - \sum_{i=1}^{n}\sum_{j=1}^{m} \ln[1 - e^{-(\alpha_i^* + \beta_j^*)}] \end{aligned} \tag{3.114}$$

and an upper bound is provided by the canonical entropy given in Eq. (3.72) for the one-sided case with the same row constraints $\vec{r}^*$:

$$S^*_{\rm can} \leq \sum_{i=1}^{n} \ln \frac{(m + r_i^*)^{m+r_i^*}}{(r_i^*)^{r_i^*} m^m}. \tag{3.115}$$

As in the binary two-sided case, the microcanonical entropy $S^*_{\rm mic}$ cannot be computed explicitly, but it can still be evaluated asymptotically from the determinant of the canonical covariance matrix $\mathbf{\Sigma}^*$ using Eqs. (3.24) and (3.25). Correspondingly, the relative entropy can be computed using Eq. (3.26). The microcanonical entropy given by Eq. (3.73) for the corresponding one-sided case is still a strict upper bound for the two-sided entropy:

$$S^*_{\rm mic} < \sum_{i=1}^{n} \ln \binom{m + r_i^* - 1}{r_i^*}. \tag{3.116}$$

As in the corresponding binary case, in the microcanonical ensemble both $\vec{r}$ and $\vec{c}$ are deterministic and fixed to the values $\vec{r}^*$ and $\vec{c}^*$, while in the canonical ensemble they are random with expected values $\vec{r}^*$ and $\vec{c}^*$. The coupled hard constraints $(\vec{r}(\mathbf{G}), \vec{c}(\mathbf{G})) = (\vec{r}^*, \vec{c}^*)$ create mutual dependencies among all the entries of $\mathbf{G}$ in the microcanonical ensemble. By contrast, the soft constraint $(\langle\vec{r}\rangle_{\vec{\alpha}^*}, \langle\vec{c}\rangle_{\vec{\beta}^*}) = (\vec{r}^*, \vec{c}^*)$ leaves all entries of $\mathbf{G}$ independent in the canonical ensemble. In the latter, as for all weighted matrices discussed so far, each entry g_{ij} is geometrically distributed, but now with its specific parameters:

$$p(g_{ij}|\vec{\alpha}^*, \vec{\beta}^*) = e^{-(\alpha_i^* + \beta_i^*)g_{ij}} \left[1 - e^{-(\alpha_i^* + \beta_i^*)}\right] \tag{3.117}$$

for $g_{ij} \in \{0, 1, 2, \dots\}$, as we show in the Appendix.

Here as well, the special case where the column sums are all equal to each other ($c_j^* = c^*$ for all j) provides a nice example. The corresponding Lagrange multipliers are in this case all equal to each other ($\beta_j^* = \beta^*$ for all j) and this allows us to solve for all parameters explicitly. In particular, Eqs. (3.112) and (3.113) reduce to the $n + 1$ independent equations

$$r_i^* = m \frac{e^{-(\alpha_i^* + \beta^*)}}{1 - e^{-(\alpha_i^* + \beta^*)}}, \quad i = 1, n, \tag{3.118}$$

$$c^* = \sum_{i=1}^{n} \frac{e^{-(\alpha_i^* + \beta^*)}}{1 - e^{-(\alpha_i^* + \beta^*)}}, \tag{3.119}$$

where, again, the second equation is equivalent to the consistency condition $c^* = \sum_{i=1}^n r_i^*/m$. Inverting Eq. (3.118), we obtain explicitly

$$\alpha_i^* + \beta^* = \ln \frac{m + r_i^*}{r_i^*} \tag{3.120}$$

which, if inserted into the expression for $S_{\rm can}^*$ in Eq. (3.114), produces exactly the canonical entropy found previously in Eq. (3.72) for the weighted ensemble with one-sided (row) constraints specified by the same vector $\vec{r}^*$:

$$\begin{aligned} S_{\rm can}^* &= \vec{\alpha}^* \cdot \vec{r}^* + m\beta^* c^* - m \sum_{i=1}^n \ln[1 - e^{-(\alpha_i^* + \beta^*)}] \\ &= \sum_{i=1}^n \ln \frac{(m + r_i^*)^{m+r_i^*}}{(r_i^*)^{r_i^*} m^m}. \end{aligned} \tag{3.121}$$

The upper bound in Eq. (3.115) is therefore fully saturated. Again, while the canonical ensembles are identical for the two cases, the microcanonical ensembles remain very different and the microcanonical entropy under two-sided constraints is strictly smaller than the one under one-sided constraints: the upper bound in Eq. (3.116) cannot be saturated. Similarly, the determinant of the covariance matrix $\mathbf{\Sigma}^*$, which here is a $(n+m) \times (n+m)$ matrix (that we calculate later on), is different from the one obtained in the one-sided case.

The (non)equivalence of the two ensembles is discussed below, in conjunction with the case of two-sided binary matrices.

Ensemble nonequivalence for matrices under two-sided local constraints

To investigate EN in the two-sided case, it is convenient to preliminary combine the results obtained so far in the binary (+) and weighted (−) cases as follows.

The canonical entropy $S_n^{\pm}[P_{\rm can}^*]$ can be evaluated by combining Eqs. (3.103) and (3.114), as well as the corresponding upper bounds given by Eqs. (3.104) and (3.115), into

$$\begin{aligned} S_n^{\pm}[P_{\rm can}^*] &= \vec{\alpha}^* \cdot \vec{r}^* + \vec{\beta}^* \cdot \vec{c}^* \pm \sum_{i=1}^n \sum_{j=1}^m \ln[1 \pm e^{-(\alpha_i^* + \beta_j^*)}] \\ &\leq \sum_{i=1}^n \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}} \end{aligned} \tag{3.122}$$

(see Eq. (3.79) for a comparison). It is easy to check that, in all the three regimes considered (sparse, dense with diverging m, dense with finite m), the above canonical entropy has the same qualitative behaviour as the corresponding quantity obtained previously in Eq. (3.79) for the one-sided case.

Unlike the one-sided case, the microcanonical entropy cannot be evaluated exactly, neither through a direct combinatorial formula nor via the complex integral approach, and we only have strict upper bounds given by Eqs. (3.105) and (3.116) in the binary and weighted case respectively, which we can combine as follows:

$$S_n^{\pm}[P^*_{\rm mic}] < \sum_{i=1}^{n} \ln \left[(\pm 1)^{r_i^*} \binom{\pm m}{r_i^*} \right] \tag{3.123}$$

(see Eq. (3.80) for a comparison).

We can now discuss EN in a combined fashion for binary and weighted matrices. While we cannot calculate the relative entropy exactly, we can correctly evaluate its asymptotic scaling via Eq. (3.26), because the canonical covariance matrix $(\mathbf{\Sigma}^*)^{\pm}$ between the constraints can still be calculated analytically as a function of the parameters $(\vec{\alpha}^*, \vec{\beta}^*)$, in both the binary and weighted cases. In particular, it is easy to see that the entries $(\mathbf{\Sigma}^*_{ij})^{\pm}$ are arranged into a block structure, with a square $n \times n$ diagonal block ($i, j \in [1, n]$) representing the covariance matrix between pairs of row sums, a square $m \times m$ diagonal block ($i, j \in [n+1, n+m]$) representing the covariance matrix between pairs of column sums, and two rectangular ($n \times m$ and $m \times n$) off-diagonal blocks representing the covariances between row and column sums ($i \in [1, n]$, $j \in [n+1, n+m]$ and $i \in [n+1, n+m]$, $j \in [1, n]$). As we show in the Appendix, these entries are

$$(\mathbf{\Sigma}^*_{ij})^{\pm} = \begin{cases} \delta_{ij} \sum_{k=1}^{m} \dfrac{e^{-(\alpha_i^* + \beta_k^*)}}{\left[1 \pm e^{-(\alpha_i^* + \beta_k^*)}\right]^2} & i, j \in [1, n], \\ \dfrac{e^{-(\alpha_i^* + \beta_{j-n}^*)}}{\left[1 \pm e^{-(\alpha_i^* + \beta_{j-n}^*)}\right]^2} & i \in [1, n],\ j \in [n+1, n+m] \\ \dfrac{e^{-(\alpha_j^* + \beta_{i-n}^*)}}{\left[1 \pm e^{-(\alpha_j^* + \beta_{i-n}^*)}\right]^2} & i \in [n+1, n+m],\ j \in [1, n] \\ \delta_{ij} \sum_{k=1}^{n} \dfrac{e^{-(\alpha_k^* + \beta_{j-n}^*)}}{\left[1 \pm e^{-(\alpha_k^* + \beta_{j-n}^*)}\right]^2} & i, j \in [n+1, n+m] \end{cases} . \tag{3.124}$$

The above expression is the generalization of Eq. (3.86) to the case of two-sided constraints. Once the values of $\vec{r}^*$ and $\vec{c}^*$ are specified, one can calculate the determinant of the above matrix and, through Eq. (3.26), the leading order of the relative entropy $S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}]$. As we show in the Appendix, the order of $\alpha_n^{\pm}$ confirms the same scalings for the relative entropy found previously in Eqs. (3.83), (3.84) and (3.85) for the one-sided case: namely, $\alpha_n^{\pm} = O(n)$ in the sparse regime, $\alpha_n^{\pm} = O(n \ln n)$ in the dense regime with $m = O(n)$, and $\alpha_n^{\pm} = O(n)$ in the dense regime with finite m.

In practice, unlike the one-sided case, calculating the values of $S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}]$ and $R_{\infty}^{\pm}$ (or bounds for them) as explicit functions of the constraints is not easy in general. It is however possible, and instructive, to consider a special case where

$S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}]$ and $R_\infty^{\pm}$ in this two-sided case (given the vectors $\vec{r}^*$ and $\vec{c}^*$) can be related to the corresponding values obtained in the one-sided case with the same vector $\vec{r}^*$ (but without a constraint on $\vec{c}^*$). Indeed, if we consider again the special case with constant column constraints ($c_j^* = c^*$, $j = 1, m$) then from our previous results in Eqs. (3.110) and (3.121) we recall that, for any given value of n, the two-sided canonical entropy $S_n^{\pm}[P^*_{\rm can}]$ is exactly equal to the one-sided canonical entropy given in Eq. (3.79) corresponding to the same vector $\vec{r}^*$, while of course the two-sided microcanonical entropy $S_n^{\pm}[P^*_{\rm mic}]$ is strictly smaller than the one-sided one given in Eq. (3.80). This automatically implies that $S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}]$ in the two-sided case is strictly larger than the corresponding one-sided relative entropy given in Eq. (3.81). This proves that the scaling of the relative entropy is always at least $O(n)$, irrespective of the density and of the value of m: in all regimes, EE breaks down for binary and weighted matrices under two-sided local constraints, as found in the one-sided case. The presence of the extra column constraints is not changing the qualitative behaviour of the relative entropy, but only its numerical value. Since the assumption of constant column sums only changes the values, but not the order, of the relative entropy, we expect that the scalings remain unchanged in the general case as well.

Moreover, EN has again the strong form ($R_\infty^{\pm} > 0$) in the sparse regime with finite m, because the value of $R_n^{\pm} = 1 - S_n^{\pm}[P^*_{\rm mic}]/S_n^{\pm}[P^*_{\rm can}]$ in the two-sided case is strictly larger than the corresponding one calculated previously for the one-sided case. In particular, we can use Eq. (3.93) to establish the following lower bound in the two-sided case with constant column constraints and finite m:

$$R_n^{\pm} > 1 - \frac{\sum_{i=1}^n \ln\left[(\pm 1)^{r_i^*}\binom{\pm m}{r_i^*}\right]}{\sum_{i=1}^n \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*}(m \mp r_i^*)^{\pm m - r_i^*}}} > 0. \tag{3.125}$$

The above inequality proves strong EN in this case as well. Again, we expect that relaxing the assumption of constant column sums will change only the value of $R_n^{\pm}$, but not its strict positivity.

Finally, we can also establish an upper bound for $R_n^{\pm}$ by rewriting Eq. (3.26) asymptotically for large n, in analogy with Eq. (3.88), as

$$S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}] = C_2(m)\alpha_n^{\pm} \tag{3.126}$$

where $C_2(m)$ is a finite positive constant and noticing that, since the covariance matrix $(\mathbf{\Sigma}^*)^{\pm}$ is positive-definite, we can use Hadamard's inequality stating that the determinant of a positive-definite matrix is less than or equal to the product of the diagonal entries of the matrix. This means

$$\alpha_n^{\pm} = \ln \sqrt{\det\left[2\pi(\mathbf{\Sigma}^*)^{\pm}\right]} \leq \tilde{\alpha}_n^{\pm} \tag{3.127}$$

where, using Eq. (3.124), we have introduced

$$\tilde{\alpha}_n^{\pm} = \frac{1}{2}\sum_{i=1}^n \ln\left[2\pi(\Sigma^*_{ii})^{\pm}\right] + \frac{1}{2}\sum_{j=1}^m \ln\left[2\pi(\Sigma^*_{jj})^{\pm}\right]. \tag{3.128}$$

Now, using Eqs. (3.109) and (3.120) in the binary and weighted case respectively, it is easy to show that, in the two-sided case with constant column sums, the first n diagonal entries of $(\mathbf{\Sigma}^*)^{\pm}$ are identical to the n diagonal entries of the covariance matrix in the corresponding one-sided case given by Eq. (3.86), i.e.

$$(\Sigma^*_{ii})^{\pm} = r^*_i \left(1 - \frac{r^*_i}{\pm m}\right), \quad i = 1, n. \tag{3.129}$$

Inserting the above expression into Eq. (3.128), and noticing that the last sum in the latter is strictly positive, we can write

$$\tilde{\alpha}^{\pm}_n < \frac{1}{2} \sum_{i=1}^{n} \ln \left[2\pi r^*_i \left(1 - \frac{r^*_i}{\pm m}\right)\right]. \tag{3.130}$$

Combining Eqs. (3.126), (3.127) and (3.130) we obtain the upper bound (for large n)

$$\begin{aligned} R^{\pm}_n &= C_2(m) \frac{\alpha^{\pm}_n}{S^{\pm}_n[P^*_{\text{can}}]} \\ &\leq C_2(m) \frac{\tilde{\alpha}^{\pm}_n}{S^{\pm}_n[P^*_{\text{can}}]} \\ &< \frac{C_2(m)}{2} \frac{\sum_{i=1}^{n} \ln \left[2\pi r^*_i \left(1 - \frac{r^*_i}{\pm m}\right)\right]}{\sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r^*_i)^{r^*_i} (m \mp r^*_i)^{\pm m - r^*_i}}} \\ &= \frac{C_2(m)}{C_1(m)} \left[1 - \frac{\sum_{i=1}^{n} \ln \left[(\pm 1)^{r^*_i} \binom{\pm m}{r^*_i}\right]}{\sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r^*_i)^{r^*_i} (m \mp r^*_i)^{\pm m - r^*_i}}}\right], \end{aligned} \tag{3.131}$$

where we have used Eqs. (3.93) and (3.94) established in the one-sided case. Comparing Eq. (3.131) with Eq. (3.125) we see that we must have $C_2(m)/C_1(m) > 1$. We conjecture that, in analogy with $C_1(m)$ in the one-sided case, $C_2(m) \gtrsim 1$. Moreover, here as well we know that $\lim_{m\to\infty} C_2(m) = 1$ as in Eq. (3.89). This means that we expect that, for n large, $C_2(m)/C_1(m) \approx 1$ so that the upper bound in Eq. (3.131) approaches the lower bound in Eq. (3.125), which is therefore a very good estimate of the actual value of $R^{\pm}_n$ in the two-sided case with constant column constraints:

$$R^{\pm}_n \approx 1 - \frac{\sum_{i=1}^{n} \ln \left[(\pm 1)^{r^*_i} \binom{\pm m}{r^*_i}\right]}{\sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r^*_i)^{r^*_i} (m \mp r^*_i)^{\pm m - r^*_i}}}. \tag{3.132}$$

Upon comparison with Eq. (3.93), we see that $R^{\pm}_n$ remains practically unchanged with respect to the one-sided case with the same value of $\vec{r}^*$.

Note that the above result means that the decrease Δ_{mic} in microcanonical entropy introduced by the extra column constraints is subleading with respect to the canonical

entropy. Indeed, denoting with $\{\cdot\}_h$ a quantity evaluated in the h-sided case (where $h = 1, 2$), and exploiting again the identity of the canonical entropies $\{S_n^{\pm}[P^*_{\rm can}]\}_1 = \{S_n^{\pm}[P^*_{\rm can}]\}_2$ and our conjecture $C_2(m)/C_1(m) \approx 1$, we can use Eqs. (3.15) and the results obtained so far to express the decrease in microcanonical entropy as

$$\begin{aligned}
\Delta_{\rm mic} &= \{S_n^{\pm}[P^*_{\rm mic}]\}_1 - \{S_n^{\pm}[P^*_{\rm mic}]\}_2 \\
&= \{S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}]\}_2 - \{S_n^{\pm}[P^*_{\rm mic}||P^*_{\rm can}]\}_1 \\
&= C_2(m)\{\alpha_n^{\pm}\}_2 - C_1(m)\{\alpha_n^{\pm}\}_1 \\
&< C_2(m)\{\tilde{\alpha}_n^{\pm}\}_2 - C_1(m)\{\alpha_n^{\pm}\}_1 \\
&\approx \frac{C_1(m)}{2}\sum_{j=1}^{m} \ln\left[2\pi(\Sigma^*_{jj})^{\pm}\right] \\
&= C_1(m)\frac{m}{2}\ln\sum_{k=1}^{n}\frac{2\pi\, e^{-(\alpha_k^*+\beta^*)}}{\left[1 \pm e^{-(\alpha_k^*+\beta^*)}\right]^2} \\
&= C_1(m)\frac{m}{2}\ln\sum_{k=1}^{n}\left[2\pi\frac{r_i^*}{m}\left(1-\frac{r_i^*}{\pm m}\right)\right],
\end{aligned} \tag{3.133}$$

which is of order $O(\ln n)$, while the canonical entropy is of order $O(n)$ in the dense case with finite m considered here.

Clearly, if we additionally consider constant row constraints, i.e. $r_i^* = r^*$ for $i = 1, n$ (where necessarily $r^* = c^* m/n$), then in analogy with Eq. (3.95) we can establish the following explicit lower bound for the value of $R_\infty^{\pm}$ in the two-sided case with constant row and column constraints:

$$R_\infty^{\pm} > 1 - \frac{\ln\left[(\pm 1)^{r^*}\binom{\pm m}{r^*}\right]}{\ln\frac{m^{\pm m}}{(r^*)^{r^*}(m \mp r^*)^{\pm m - r^*}}} > 0. \tag{3.134}$$

Our expectation in Eq. (3.132) suggests that the above lower bound is a very good approximation for the actual value of $R_\infty^{\pm}$:

$$R_\infty^{\pm} \approx 1 - \frac{\ln\left[(\pm 1)^{r^*}\binom{\pm m}{r^*}\right]}{\ln\frac{m^{\pm m}}{(r^*)^{r^*}(m \mp r^*)^{\pm m - r^*}}}, \tag{3.135}$$

leading to the same result as in Eq. (3.95) for the one-sided case.

The above results generalize the finding of strong EN to the two-sided case, again in the dense regime with finite m. The results do not change qualitatively, and apparently only slightly quantitatively, with respect to the one-sided case. This result points again at the fact that it is the extensivity of the constraints that plays the key role for EN: adding a finite number m of (column) constraints does not relevantly change the picture already obtained in the one-sided case.

3.4 Discussion and conclusions

We have studied the problem of EN in the general context of $n \times m$ matrices with given constraints. Such matrices can represent high-dimensional data such as multivariate time series, expression profiles, multiplex social activity, and other relational or structured data encountered in many settings. Their entries can either be binary (Boolean) or weighted (non-negative integers). The constraints imposed on these matrices represent sums over either all the entries of the matrix (single global constraint) or over individual rows (local one-sided constraints) and possibly also columns (local two-sided constraints). These constraints take the form of linear terms into the Hamiltonian at the exponent of the maximum-entropy probability distribution characterizing the matrix ensemble.

Global constraints do not account for the heterogeneity (either spatial or temporal, i.e. nonstationarity) in the physical data-generating process, as they lead to probability distributions with identical parameters for all the entries of the matrix. By contrast, local constraints produce probability distributions with different local (row- and possibly column-specific) parameters. Most modern data structures are heterogeneous and/or nonstationary, and are therefore characterized by (at least) the type of local constraints considered here. Indeed, maximum-entropy ensembles with local constraints are being increasingly used, either as null models for pattern detection or even as generative models and inference methods whenever there is only partial, local information available about the system [44, 14].

We have shown that local constraints break the asymptotic (i.e. for large n) equivalence of canonical and microcanonical ensembles, where the constraints are enforced in a soft and hard manner respectively. By contrast, global constraints preserve EE. Mathematically, EE is encountered when the relative entropy between the canonical and microcanonical probability distributions is $o(n)$. Importantly, the breakdown of EE observed here under local constraints occurs without phase transitions, which would require nonlinear constraints in the Hamiltonian and are therefore deliberately excluded from the cases we considered. The form of EN we observe under local constraints is also 'unrestricted', i.e. it holds for any value of the model parameters (here, for any graphical value of the constraints), while the mechanism for EN based on phase transitions requires specific parameters or phases. Our results hold in all regimes of density and for all values of m, and therefore generalize a recently discovered, alternative mechanism for the breakdown of EE observed so far in ensembles of binary graphs with given degree sequence [5, 26, 7, 45] and weighted graphs with given strength sequence [28].

At the same time, our results highlight a qualitatively new finding. While the systems with local constraints studied in the past exhibited a 'weak' degree of EN (where the relative entropy is of smaller order compared with the canonical entropy, while still growing at least linearly in n), here we identified a regime for which EN is as 'strong' as in presence of phase transitions (i.e. with the relative entropy being of the same order as the canonical entropy). This regime is obtained when both m and the expected value of each entry of the matrix are finite, i.e. $O(1)$. In practice, this means

that the data structure is one where n grows as the size of the system grows, while m remains finite. This circumstance is naturally encountered e.g. when n represents a large number of timesteps during which a small number m of synchronous time series are observed (e.g. for EEG signals), or when n represents a large number of genes for which expression levels are observed in a small number m of cells at the same time, or when n represents a large number of users whose activities or preferences are recorded for a small number m of platforms, items, or other dimensions.

The simultaneously 'strong' and 'unrestricted' form of EN discussed here has never been documented so far, to the best of our knowledge. Indeed, in all the settings that had been studied previously, m was necessarily equal to n since the matrices represented the special case of square adjacency matrices of graphs, therefore the regime leading to strong EN could not be observed.

EN has important practical consequences. A traditional expectation in statistical physics is that, in absence of phase transitions or long-range interactions, ensembles are equivalent and it is therefore legitimate to freely choose the ensemble to work with, e.g. based purely on mathematical or computational convenience. For instance, if the ensemble is used as a null model for a real system, one may either want to randomize the data numerically by keeping certain quantities fixed (in which case the microcanonical ensemble is the most efficient choice) or prefer an exact mathematical characterization of the probability of each configuration in the ensemble (in which case the canonical ensemble is the easiest to work with). This view has been challenged by the recent discovery of EN under local constraints. Nonequivalence imposes a principled choice of the ensemble, that can no longer be based on practical convenience. For instance, if one has reasons to believe that the hypothesis underlying the null model, or the partial information available about the system, should be treated as a hard constraint, then one is forced to choose the microcanonical ensemble. By contrast, if one believes the constraints should be treated as soft (for instance to account for possible measurement errors leading to noisy values of the constraints in the data), then one should take the canonical route.

Our observation of strong EN shows that the quantitative differences between the two descriptions of the same system are much bigger than previously encountered in the case of weak EN. These big quantitative differences are exemplified by Eqs. (3.34), (3.35) and (3.36). A 'wrong' choice of the ensemble can therefore lead to major errors in the estimation of the probability distribution characterizing the ensemble, of the resulting entropy, of the expected values of higher-order properties that are nonlinear functions of the constraints, etc. Conclusions of statistical analyses can therefore be highly biased.

However, besides this warning, the findings presented here are intended to offer also a constructive solution. The fact that it is possible to rigorously quantify the differences between the two ensembles via the explicit calculation of the relative entropy ratio R_n and its limiting value R_∞ implies that one can still make a convenient choice of the ensemble, while at the same time being able to retrieve the desired results for the other ensemble via the calculated value of R_∞. In other words, besides being a warning signal for strong EN, R_∞ is also a concrete tool allowing researchers to switch

more easily between alternative descriptions of the same system, by compensating for their irreducible differences. The calculations carried out here can hopefully serve as useful references for future quantitative research in a variety of domains.

Appendix 3.A Global constraints

Here we derive the main mathematical expressions for the case of global constraints on our ensembles of binary and weighted matrices. This corresponds to the case where the constraint $\vec{C}(\mathbf{G})$ is a simple scalar quantity $C(\mathbf{G})$ defined as the total value $C(\mathbf{G}) \equiv t(\mathbf{G}) \equiv \sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}$. There is only one scalar Lagrange multiplier θ entering the definition of the Hamiltonian

$$H(\mathbf{G},\theta) = \theta\, t(\mathbf{G}) = \theta \sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}. \tag{3.136}$$

The above Hamiltonian is the same for both binary and weighted matrices under a global constraint. However, the calculation of the partition function (hence of all the other properties) is different in the two cases.

3.A.1 Binary matrices under a global constraint

Let us consider binary matrices first ($g_{ij} = 0, 1$). The partition function can be calculated as follows:

$$\begin{aligned}
Z(\theta) &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-H(\mathbf{G},\theta)} \\
&= \sum_{\mathbf{G}\in\mathcal{G}} e^{-\theta\sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}} \\
&= \sum_{\mathbf{G}\in\mathcal{G}} \prod_{i=1}^{n}\prod_{j=1}^{m} e^{-\theta g_{ij}} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{g_{ij}=0,1} e^{-\theta g_{ij}} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m}(1+e^{-\theta}) \\
&= (1+e^{-\theta})^{mn}.
\end{aligned} \tag{3.137}$$

This leads to

$$\begin{aligned}
P_{\rm can}(\mathbf{G}|\theta) &= \frac{e^{-H(\mathbf{G},\theta)}}{Z(\theta)} \\
&= \frac{e^{-\theta\, t(\mathbf{G})}}{(1+e^{-\theta})^{mn}} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{e^{-\theta g_{ij}}}{1+e^{-\theta}}
\end{aligned} \tag{3.138}$$

and to Eq. (3.37) in the main text. Notice that Eq. (3.138) reveals that each entry g_{ij} of the matrix $\mathbf{G}$ is a Bernoulli-distributed random variable taking the value $g_{ij} = 1$ with probability $p(1|\theta) = e^{-\theta}/(1+e^{-\theta})$ and the value $g_{ij} = 0$ with probability $p(0|\theta) = 1/(1+e^{-\theta})$, i.e.

$$p(g_{ij}|\theta) = \frac{e^{-\theta g_{ij}}}{1+e^{-\theta}}. \tag{3.139}$$

(the entries of $\mathbf{G}$ are therefore *i.i.d.*). The expected value of g_{ij} is

$$\langle g_{ij}\rangle_\theta \equiv \sum_{g_{ij}=0,1} g_{ij} p(g_{ij}|\theta) = \frac{e^{-\theta}}{1+e^{-\theta}}, \tag{3.140}$$

while its variance is

$$\mathrm{Var}_\theta[g_{ij}] \equiv \langle g_{ij}^2\rangle_\theta - \langle g_{ij}\rangle_\theta^2 = \frac{e^{-\theta}}{(1+e^{-\theta})^2}. \tag{3.141}$$

The resulting expected value and variance of the constraint $t(\mathbf{G})$ are

$$\langle t\rangle_\theta = \sum_{i=1}^{n}\sum_{j=1}^{m}\langle g_{ij}\rangle_\theta = nm\frac{e^{-\theta}}{1+e^{-\theta}}, \tag{3.142}$$

$$\mathrm{Var}_\theta[t] = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{Var}_\theta[g_{ij}] = nm\frac{e^{-\theta}}{(1+e^{-\theta})^2}, \tag{3.143}$$

the latter identity following from the fact that, since all the entries of $\mathbf{G}$ are mutually independent, the variance of the constraint $t(\mathbf{G})$ is the sum of all variances.

Now, we have to find the parameter value θ^* that solves Eq. (3.5) or equivalently maximizes the log-likelihood $\ln P_{\mathrm{can}}(\mathbf{G}^*|\theta)$. This can be done by setting the expected value $\langle t\rangle_{\theta^*}$ equal to the desired value t^*, which leads to

$$e^{-\theta^*} = \frac{t^*}{mn - t^*}, \tag{3.144}$$

and to Eq. (3.38) in the main text. Inserting Eq. (3.144) into the expressions for $P_{\mathrm{can}}(\mathbf{G}^*|\theta)$ and $\mathrm{Var}_\theta[t]$ leads to the values of S^*_{can} and $\mathrm{Var}_{\theta^*}[t]$ given in Eqs. (3.39) and (3.44) in the main text. One can easily confirm that $\mathrm{Var}_{\theta^*}[t]$ coincides with the only ($K = 1$) entry of the 1×1 covariance matrix $\mathbf{\Sigma}^*$ obtained through Eq. (3.23), i.e.

$$\begin{aligned}\Sigma^* &= \left.\frac{\partial^2 \ln Z(\theta)}{\partial\theta^2}\right|_{\theta=\theta^*} \\ &= t^*\left(1 - \frac{t^*}{mn}\right) \\ &= \mathrm{Var}_{\theta^*}[t]\end{aligned} \tag{3.145}$$

and, trivially,

$$\det(\mathbf{\Sigma}^*) = t^* \left(1 - \frac{t^*}{mn}\right). \tag{3.146}$$

The calculation of the microcanonical entropy $S^*_{\rm mic}$ is in this case trivial and given by Eq. (3.40) in the main text. Given the simplicity of this example, it is instructive to show explicitly that the integral formula in Eq. (3.20), which can be calculated exactly in this case, gives the correct value of Ω_{t^*}. To do this, we use Eq. (3.138) to obtain the complex-valued quantity

$$P_{\rm can}(\mathbf{G}^*|\theta^* + \mathrm{i}\psi) = \frac{e^{-(\theta^*+\mathrm{i}\psi)t^*}}{[1 + e^{-(\theta^*+\mathrm{i}\psi)}]^{mn}} \tag{3.147}$$

and use Eq. (3.20) to calculate Ω_{t^*} as

$$\Omega_{t^*} = \frac{1}{2\pi} \int_{-\pi}^{\pi} [1 + e^{-(\theta^*+\mathrm{i}\psi)}]^{mn} e^{(\theta^*+\mathrm{i}\psi)t^*} \mathrm{d}\psi. \tag{3.148}$$

To calculate the above integral, we change variable from ψ to $z \equiv e^{-(\theta^*+\mathrm{i}\psi)}$, so that $\mathrm{d}z = \mathrm{d}e^{-(\theta^*+\mathrm{i}\psi)} = -\mathrm{i}z\mathrm{d}\psi$ and $\mathrm{d}\psi = \mathrm{i}\,\mathrm{d}z/z$. Then the integral becomes

$$\Omega_{t^*} = \frac{\mathrm{i}}{2\pi} \int_{e^{-(\theta^*-\mathrm{i}\pi)}}^{e^{-(\theta^*+\mathrm{i}\pi)}} (1+z)^{mn} z^{-(t^*+1)} \mathrm{d}z. \tag{3.149}$$

and, using the binomial formula

$$(1+x)^l = \sum_{k=0}^{l} \binom{l}{k} x^{l-k}, \tag{3.150}$$

we obtain

$$\Omega_{t^*} = \frac{\mathrm{i}}{2\pi} \sum_{k=1}^{mn} \binom{mn}{k} \int_{e^{-(\theta^*-\mathrm{i}\pi)}}^{e^{-(\theta^*+\mathrm{i}\pi)}} z^{mn-k-t^*-1} \mathrm{d}z. \tag{3.151}$$

Now, each integral in the above sum can be calculated using Cauchy's residue theorem, from which we know that the integral is non-zero only when the exponent of z is -1, in which case it equals $-2\pi\mathrm{i}$. This selects the only value $k = mn - t^*$ in the sum, so that

$$\Omega_{t^*} = \frac{\mathrm{i}}{2\pi} \binom{mn}{mn - t^*} (-2\pi\mathrm{i}) = \binom{mn}{t^*}, \tag{3.152}$$

which coincides with the binomial coefficient used in Eq. (3.40).

3.A.2 Weighted matrices under a global constraint

We now consider the case of weighted matrices ($g_{ij} = 0, 1, 2, \ldots, +\infty$) with a global constraint t^*. The Hamiltonian is still given by Eq. (3.136), while the partition

function is now calculated differently as follows:

$$\begin{aligned} Z(\theta) &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-H(\mathbf{G},\theta)} \\ &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-\theta\sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}} \\ &= \sum_{\mathbf{G}\in\mathcal{G}} \prod_{i=1}^{n}\prod_{j=1}^{m} e^{-\theta g_{ij}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{g_{ij}=0}^{+\infty} e^{-\theta g_{ij}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{1}{1-e^{-\theta}} \\ &= \frac{1}{(1-e^{-\theta})^{mn}}. \end{aligned} \tag{3.153}$$

The canonical probability is therefore

$$\begin{aligned} P_{\mathrm{can}}(\mathbf{G}|\theta) &= \frac{e^{-H(\mathbf{G},\theta)}}{Z(\theta)} \\ &= e^{-\theta\, t(\mathbf{G})}(1-e^{-\theta})^{mn} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m} e^{-\theta g_{ij}}(1-e^{-\theta}), \end{aligned} \tag{3.154}$$

which leads to Eq. (3.45) in the main text. Equation (3.154) shows that all the entries of $\mathbf{G}$ are *i.i.d.* random variables, in this case distributed according to a geometric distribution with success probability $e^{-\theta}$:

$$p(g_{ij}|\theta) = e^{-\theta g_{ij}}(1-e^{-\theta}). \tag{3.155}$$

The expected value of g_{ij} is now

$$\langle g_{ij}\rangle_\theta = \sum_{g_{ij}=0}^{+\infty} g_{ij}\, p(g_{ij}|\theta) = \frac{e^{-\theta}}{1-e^{-\theta}} \tag{3.156}$$

and its variance is

$$\mathrm{Var}_\theta[g_{ij}] \equiv \langle g_{ij}^2\rangle_\theta - \langle g_{ij}\rangle_\theta^2 = \frac{e^{-\theta}}{(1-e^{-\theta})^2} \tag{3.157}$$

(note the change of sign at the denominator with respect to Eq. (3.141)), from which we calculate the expected value and variance of the constraint $t(\mathbf{G})$ as

$$\langle t \rangle_\theta = \sum_{i=1}^{n}\sum_{j=1}^{m} \langle g_{ij} \rangle_\theta = nm \frac{e^{-\theta}}{1-e^{-\theta}}, \tag{3.158}$$

$$\mathrm{Var}_\theta[t] = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{Var}_\theta[g_{ij}] = nm \frac{e^{-\theta}}{(1-e^{-\theta})^2}. \tag{3.159}$$

The maximum-likelihood parameter value θ^* is found by setting the expected value $\langle t \rangle_{\theta^*}$ equal to t^*, resulting in

$$e^{-\theta^*} = \frac{t^*}{mn + t^*} \tag{3.160}$$

(notice again the change of sign with respect to the binary case) and to Eq. (3.46) in the main text. Substituting Eq. (3.160) into Eqs. (3.154) and (3.159) produces the expressions for S^*_{can} and $\mathrm{Var}_{\theta^*}[t]$ shown in Eqs. (3.47) and (3.52) in the main text. As for the binary case, one can easily confirm that $\mathrm{Var}_{\theta^*}[t]$ coincides with

$$\begin{aligned} \boldsymbol{\Sigma}^* &= \left. \frac{\partial^2 \ln Z(\theta)}{\partial \theta^2} \right|_{\theta=\theta^*} \\ &= t^* \left(1 + \frac{t^*}{mn} \right) \\ &= \mathrm{Var}_{\theta^*}[t] \end{aligned} \tag{3.161}$$

so that

$$\det(\boldsymbol{\Sigma}^*) = t^* \left(1 + \frac{t^*}{mn} \right). \tag{3.162}$$

Again, it is instructive to show that the complex integral in Eq. (3.20) gives the exact result corresponding to the microcanonical entropy reported in Eq. (3.48). Calculating the quantity

$$P_{\mathrm{can}}(\mathbf{G}^* | \theta^* + \mathrm{i}\psi) = \frac{e^{-(\theta^* + \mathrm{i}\psi)t^*}}{\left[1 - e^{-(\theta^* + \mathrm{i}\psi)}\right]^{-mn}} \tag{3.163}$$

and inserting it into Eq. (3.20) yields

$$\Omega_{t^*} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{(\theta^* + \mathrm{i}\psi)t^*}}{\left[1 - e^{-(\theta^* + \mathrm{i}\psi)}\right]^{mn}} \mathrm{d}\psi. \tag{3.164}$$

We first perform the change of variable $y \equiv e^{-(\theta^* + \mathrm{i}\psi)}$, $\mathrm{d}\psi = \mathrm{i}\mathrm{d}y/y$ and rearrange the integral as

$$\begin{aligned} \Omega_{t^*} &= \frac{\mathrm{i}}{2\pi} \int_{e^{-(\theta^* - \mathrm{i}\pi)}}^{e^{-(\theta^* + \mathrm{i}\pi)}} y^{-(t^*+1)} \left(\frac{1}{1-y} \right)^{mn} \mathrm{d}y \\ &= \frac{\mathrm{i}}{2\pi} \int_{e^{-(\theta^* - \mathrm{i}\pi)}}^{e^{-(\theta^* + \mathrm{i}\pi)}} y^{-(t^*+1)} \left(1 + \frac{y}{1-y} \right)^{mn} \mathrm{d}y. \end{aligned} \tag{3.165}$$

Then we perform a second change of variable $z \equiv y/(1-y)$, $\mathrm{d}y = \mathrm{d}z/(z+1)^2$ and apply the binomial formula in Eq. (3.150) twice to obtain

$$\begin{aligned}
\Omega_{t^*} &= \frac{\mathrm{i}}{2\pi} \sum_{k=0}^{mn} \binom{mn}{k} \int_{z_-}^{z_+} \left(\frac{z}{z+1}\right)^{-(t^*+1)} \frac{z^k}{(z+1)^2} \mathrm{d}z \\
&= \frac{\mathrm{i}}{2\pi} \sum_{k=0}^{mn} \binom{mn}{k} \int_{z_-}^{z_+} z^{k-t^*-1} (z+1)^{t^*-1} \mathrm{d}z \\
&= \frac{\mathrm{i}}{2\pi} \sum_{k=0}^{mn} \binom{mn}{k} \sum_{h=0}^{t^*-1} \binom{t^*-1}{h} \int_{z_-}^{z_+} z^{k+h-t^*-1} \mathrm{d}z
\end{aligned}$$

where we have defined

$$z_\pm \equiv \frac{e^{-(\theta^* \pm \mathrm{i}\pi)}}{1 - e^{-(\theta^* \pm \mathrm{i}\pi)}}. \tag{3.166}$$

Using again the residue theorem, the only non-zero integral is obtained for $h = t^* - k$, which selects the value

$$\begin{aligned}
\Omega_{t^*} &= \frac{1}{2\pi\mathrm{i}} \sum_{k=0}^{mn} \binom{mn}{k} \binom{t^*-1}{k-1} (-2\pi\mathrm{i}) \\
&= \sum_{k=0}^{mn} \binom{mn}{k} \binom{t^*-1}{k-1} \\
&= \sum_{k=0}^{mn} \binom{mn}{k} \binom{t^*-1}{t^*-k} \\
&= \binom{t^*+mn-1}{t^*}
\end{aligned} \tag{3.167}$$

(where we have used the generalized Vandermonde's identity). The above calculation retrieves exactly the negative binomial coefficient used in Eq. (3.48).

Appendix 3.B One-sided local constraints

We now consider the case of one-sided local constraints on ensembles of binary and weighted $n \times m$ matrices. The constraint $\vec{C}(\mathbf{G})$ is now an n-dimensional ($K = n$) vector $\vec{r}(\mathbf{G})$ where the entry $r_i(\mathbf{G}) = \sum_{j=1}^{m} g_{ij}$ is the i-th row sum of the matrix $\mathbf{G}$. Correspondingly, there is an n-dimensional vector $\vec{\theta}$ of Lagrange multipliers and the Hamiltonian is

$$H(\mathbf{G}, \vec{\theta}) = \vec{\theta} \cdot \vec{r}(\mathbf{G}) = \sum_{i=1}^{n} \theta_i r_i(\mathbf{G}) = \sum_{i=1}^{n} \theta_i \sum_{j=1}^{m} g_{ij} \tag{3.168}$$

for both binary and weighted matrices. The calculation of the resulting properties of binary and weighted ensembles is discussed separately below.

3.B.1 Binary matrices under one-sided local constraints

In the binary case, the partition function $Z(\vec{\theta})$ can be calculated from Eq. (3.168) according to the following generalization of Eq. (3.137):

$$\begin{aligned} Z(\vec{\theta}) &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-H(\mathbf{G},\vec{\theta})} \\ &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-\sum_{i=1}^{n}\theta_i \sum_{j=1}^{m} g_{ij}} \\ &= \sum_{\mathbf{G}\in\mathcal{G}} \prod_{i=1}^{n}\prod_{j=1}^{m} e^{-\theta_i g_{ij}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{g_{ij}=0,1} e^{-\theta_i g_{ij}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}(1+e^{-\theta_i}) \\ &= \prod_{i=1}^{n}(1+e^{-\theta_i})^m. \end{aligned} \tag{3.169}$$

The resulting canonical probability is

$$\begin{aligned} P_{\text{can}}(\mathbf{G}|\vec{\theta}) &= \frac{e^{-H(\mathbf{G},\vec{\theta})}}{Z(\vec{\theta})} \\ &= \frac{e^{-\vec{\theta}\cdot\vec{r}(\mathbf{G})}}{\prod_{i=1}^{n}(1+e^{-\theta_i})^m} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{e^{-\theta_i g_{ij}}}{1+e^{-\theta_i}}, \end{aligned} \tag{3.170}$$

which leads to Eq. (3.61) in the main text. As in the case of binary matrices under a global constraint, each entry g_{ij} of the matrix $\mathbf{G}$ is a Bernoulli-distributed random variable. However, while all these entries are still independent, the parameter of the distribution depends on the row being considered:

$$p(g_{ij}|\vec{\theta}) = \frac{e^{-\theta_i g_{ij}}}{1+e^{-\theta_i}}. \tag{3.171}$$

Consequently, Eqs. (3.140) and (3.141) generalize to

$$\langle g_{ij}\rangle_{\vec{\theta}} \equiv \sum_{g_{ij}=0,1} g_{ij}\, p(g_{ij}|\vec{\theta}) = \frac{e^{-\theta_i}}{1+e^{-\theta_i}}, \tag{3.172}$$

$$\text{Var}_{\vec{\theta}}[g_{ij}] \equiv \langle g_{ij}^2\rangle_{\vec{\theta}} - \langle g_{ij}\rangle_{\vec{\theta}}^2 = \frac{e^{-\theta_i}}{(1+e^{-\theta_i})^2}. \tag{3.173}$$

We can therefore calculate the expected value of each constraint $r_i(\mathbf{G})$ as

$$\langle r_i \rangle_{\vec{\theta}} = \sum_{j=1}^{m} \langle g_{ij} \rangle_{\vec{\theta}} = m \frac{e^{-\theta_i}}{1 + e^{-\theta_i}}, \quad i = 1, n. \tag{3.174}$$

Similarly, the variance of r_i is

$$\mathrm{Var}_{\vec{\theta}}[r_i] = \sum_{j=1}^{m} \mathrm{Var}_{\vec{\theta}}[g_{ij}] = m \frac{e^{-\theta_i}}{(1 + e^{-\theta_i})^2}, \quad i = 1, n \tag{3.175}$$

while all covariances between different constraints are zero, because of the independence of distinct entries of $\mathbf{G}$:

$$\mathrm{Cov}_{\vec{\theta}}[r_i, r_j] = \sum_{k=1}^{m} \sum_{l=1}^{m} \mathrm{Cov}_{\vec{\theta}}[g_{ik}, g_{jl}] = 0, \quad i \neq j. \tag{3.176}$$

We can combine Eqs. (3.175) and (3.176) as follows:

$$\mathrm{Cov}_{\vec{\theta}}[r_i, r_j] = \delta_{ij} m \frac{e^{-\theta_i}}{(1 + e^{-\theta_i})^2}, \tag{3.177}$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

Now, the parameter value $\vec{\theta}^*$ that maximizes the log-likelihood is found by equating the expected value $\langle \vec{r} \rangle_{\vec{\theta}^*}$ with the desired value $\vec{r}^*$. Inverting Eq. (3.174), this leads to

$$e^{-\theta_i^*} = \frac{r_i^*}{m - r_i^*} \qquad i = 1, n \tag{3.178}$$

or equivalently to Eq. (3.62) in the main text. The expression for the canonical entropy S^*_{can} given in Eq. (3.63) in the main text follows from substituting Eqs.(3.178) into Eq. (3.170). Similarly, the expression for the entries of the $n \times n$ covariance matrix $\mathbf{\Sigma}^*$ given in Eq. (3.68) in the main text follows from combining Eqs. (3.177) and (3.178). Note that Eq. (3.68) can also be obtained by differentiating the logarithm of Eq. (3.169) as prescribed by Eq. (3.23):

$$\begin{aligned} \Sigma^*_{ij} &= \left. \frac{\partial^2 \ln Z(\vec{\theta})}{\partial \theta_i \partial \theta_j} \right|_{\vec{\theta} = \vec{\theta}^*} \\ &= \delta_{ij} r_i^* \left(1 - \frac{r_i^*}{m} \right) \\ &= \mathrm{Cov}_{\vec{\theta}^*}[r_i, r_j]. \end{aligned} \tag{3.179}$$

These results imply

$$\det(\mathbf{\Sigma}^*) = \prod_{i=1}^{n} r_i^* \left(1 - \frac{r_i^*}{m} \right). \tag{3.180}$$

The microcanonical entropy $S^*_{\rm mic}$ can be directly calculated as Eq. (3.64) in the main text. We can still confirm that its value is exactly retrieved by using the integral formula in Eq. (3.20). From Eq. (3.170) we obtain

$$P_{\rm can}(\mathbf{G}^*|\vec{\theta^*}+\mathrm{i}\vec{\psi}) = \prod_{i=1}^{n} \frac{e^{-(\theta_i^*+\mathrm{i}\psi_i)r_i^*}}{[1+e^{-(\theta_i^*+\mathrm{i}\psi_i)}]^m}. \tag{3.181}$$

Using Eq. (3.20), we can calculate $\Omega_{\vec{r}^*}$ by exploiting again the binomial theorem, a change of variables ($z_i \equiv e^{-(\theta_i^*+\mathrm{i}\psi_i)}$, $\mathrm{d}z_i = -\mathrm{i}z_i\mathrm{d}\psi_i$) and the residue theorem as in Eqs. (3.149), (3.151) and (3.152):

$$\begin{aligned}
\Omega_{\vec{r}^*} &= \int_{-\vec{\pi}}^{+\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^n} \prod_{i=1}^{n} \frac{[1+e^{-(\theta_i^*+\mathrm{i}\psi_i)}]^m}{e^{-(\theta_i^*+\mathrm{i}\psi_i)r_i^*}} \\
&= \prod_{i=1}^{n} \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi_i}{2\pi} \frac{[1+e^{-(\theta_i^*+\mathrm{i}\psi_i)}]^m}{e^{-(\theta_i^*+\mathrm{i}\psi_i)r_i^*}} \\
&= \prod_{i=1}^{n} \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi_i}{2\pi} \sum_{k=0}^{m} \binom{m}{k} e^{-(\theta_i^*+\mathrm{i}\psi_i)(k-r_i^*)} \\
&= \prod_{i=1}^{n} \int_{\theta_i^*-\mathrm{i}\pi}^{\theta_i^*+\mathrm{i}\pi} \frac{\mathrm{d}z_i}{(-2\pi\mathrm{i})} \sum_{k=0}^{m} \binom{m}{k} {z_i}^{k-r_i^*-1} \\
&= \prod_{i=1}^{n} \frac{(-2\pi\mathrm{i})}{(-2\pi\mathrm{i})} \binom{m}{r_i^*} \\
&= \prod_{i=1}^{n} \binom{m}{r_i^*}
\end{aligned} \tag{3.182}$$

which coincides with Eq. (3.64).

3.B.2 Weighted matrices under one-sided local constraints

In the weighted case, the partition function is given by the following generalization of Eq. (3.153):

$$\begin{aligned} Z(\vec{\theta}) &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-H(\mathbf{G},\vec{\theta})} \\ &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-\sum_{i=1}^{n}\theta_i \sum_{j=1}^{m} g_{ij}} \\ &= \sum_{\mathbf{G}\in\mathcal{G}} \prod_{i=1}^{n}\prod_{j=1}^{m} e^{-\theta_i g_{ij}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{g_{ij}=0}^{+\infty} e^{-\theta_i g_{ij}} \\ &= \prod_{i=1}^{n} \frac{1}{(1-e^{-\theta_i})^m}. \end{aligned} \tag{3.183}$$

The resulting canonical probability is

$$\begin{aligned} P_{\text{can}}(\mathbf{G}|\vec{\theta}) &= \frac{e^{-H(\mathbf{G},\vec{\theta})}}{Z(\vec{\theta})} \\ &= \frac{e^{-\vec{\theta}\cdot\vec{r}(\mathbf{G})}}{\prod_{i=1}^{n}(1-e^{-\theta_i})^{-m}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m} \frac{e^{-\theta_i g_{ij}}}{(1-e^{-\theta_i})^{-1}}, \end{aligned} \tag{3.184}$$

leading to Eq. (3.70) in the main text. As in the case of weighted matrices under a global constraint, each entry g_{ij} of the matrix $\mathbf{G}$ is an independent and geometrically distributed random variable. On the other hand, as in the case of binary matrices under local constraints, the parameter of the distribution depends on the row being considered:

$$p(g_{ij}|\vec{\theta}) = e^{-\theta_i g_{ij}}(1-e^{-\theta_i}). \tag{3.185}$$

The resulting expected value and variance of g_{ij} are given by the following generalizations of Eqs. (3.156) and (3.157):

$$\langle g_{ij}\rangle_{\vec{\theta}} \equiv \sum_{g_{ij}=0,1} g_{ij}\, p(g_{ij}|\vec{\theta}) = \frac{e^{-\theta_i}}{1-e^{-\theta_i}}, \tag{3.186}$$

$$\text{Var}_{\vec{\theta}}[g_{ij}] \equiv \langle g_{ij}^2\rangle_{\vec{\theta}} - \langle g_{ij}\rangle_{\vec{\theta}}^2 = \frac{e^{-\theta_i}}{(1-e^{-\theta_i})^2}. \tag{3.187}$$

The expected value of each constraint $r_i(\mathbf{G})$ is therefore

$$\langle r_i \rangle_{\vec{\theta}} = \sum_{j=1}^{m} \langle g_{ij} \rangle_{\vec{\theta}} = m \frac{e^{-\theta_i}}{1 - e^{-\theta_i}}, \quad i = 1, n, \tag{3.188}$$

while the covariances between different constraints are

$$\mathrm{Cov}_{\vec{\theta}}[r_i, r_j] = \delta_{ij} m \frac{e^{-\theta_i}}{(1 - e^{-\theta_i})^2}. \tag{3.189}$$

As usual, note the change of sign at the denominator of Eqs. (3.188) and (3.189) with respect to the corresponding Eqs. (3.174) and (3.177) valid in the binary case.

Using Eq. (3.188), we set $\langle \vec{r} \rangle_{\vec{\theta}^*} = \vec{r}^*$ and solve for $\vec{\theta}^*$, finding

$$e^{-\theta_i^*} = \frac{r_i^*}{m + r_i^*} \qquad i = 1, n \tag{3.190}$$

as the parameter value that maximizes the log-likelihood. From the above expression, we get Eq. (3.71) and, using Eq. (3.184), Eq. (3.72) in the main text.

Similarly, the expression for the entries of the $n \times n$ covariance matrix $\mathbf{\Sigma}^*$ given in Eq. (3.68) in the main text follows from combining Eqs. (3.175), (3.176) and (3.178). Note that Eq. (3.68) can also be obtained by differentiating the logarithm of Eq. (3.169) as prescribed by Eq. (3.23):

$$\begin{aligned} \Sigma_{ij}^* &= \left. \frac{\partial^2 \ln Z(\vec{\theta})}{\partial \theta_i \partial \theta_j} \right|_{\vec{\theta} = \vec{\theta}^*} \\ &= \delta_{ij} r_i^* \left(1 + \frac{r_i^*}{m} \right) \\ &= \mathrm{Cov}_{\vec{\theta}^*}[r_i, r_j]. \end{aligned} \tag{3.191}$$

So, in analogy with the binary case,

$$\det(\mathbf{\Sigma}^*) = \prod_{i=1}^{n} r_i^* \left(1 + \frac{r_i^*}{m} \right). \tag{3.192}$$

The microcanonical entropy S_{mic}^* can be directly calculated as Eq. (3.73) in the main text. We can still confirm that its value is correctly retrieved by using the integral formula in Eq. (3.20). From Eq. (3.184) we obtain

$$P_{\mathrm{can}}(\mathbf{G}^* | \vec{\theta}^* + \mathrm{i}\vec{\psi}) = \prod_{i=1}^{n} \frac{e^{-(\theta_i^* + \mathrm{i}\psi_i) r_i^*}}{[1 - e^{-(\theta_i^* + \mathrm{i}\psi_i)}]^{-m}}. \tag{3.193}$$

Using Eq. (3.20), we can calculate $\Omega_{\vec{r}^*}$ by exploiting again the binomial theorem as

$$\Omega_{\vec{r}^*} = \int_{-\vec{\pi}}^{\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^n} \prod_{i=1}^{n} \frac{e^{(\beta_i^* + \mathrm{i}\psi_i)(r_i^*)}}{[1 - e^{-(\beta_i^* + \mathrm{i}\psi_i)}]^m}. \tag{3.194}$$

We can use the change of variables $y_i \equiv e^{-(\beta_i^*+\mathrm{i}\psi_i)}$, $\mathrm{d}\psi_i = -\mathrm{i}\mathrm{d}y_i/y_i$ and the relation $(1-y_i)^{-m} = (1+\frac{y_i}{1-y_i})^m$ to calculate $\Omega_{\vec{r}^*}$ as

$$\begin{aligned}\Omega_{\vec{r}^*} &= \prod_{i=1}^{n}\int_{\beta_i^*-\mathrm{i}\pi}^{\beta_i^*+\mathrm{i}\pi}\frac{\mathrm{d}y_i}{2\pi\mathrm{i}}\left(1+\frac{y_i}{1-y_i}\right)^m y_i^{-r_i^*-1} \\ &= \prod_{i=1}^{n}\int_{\beta_i^*-\mathrm{i}\pi}^{\beta_i^*+\mathrm{i}\pi}\frac{\mathrm{d}y_i}{2\pi\mathrm{i}}\sum_{k=0}^{m}\binom{m}{k}y_i^{-r_i^*-1}\left(\frac{y_i}{1-y_i}\right)^k.\end{aligned} \tag{3.195}$$

Using another change of variables $z_i = y_i/(1-y_i)$, $y_i = z_i/(z_i+1)$, $\mathrm{d}y_i = \mathrm{d}z_i/(z_i+1)^2$, and denoting $u_i^* = \frac{e^{-(\beta_i^*+\mathrm{i}\pi)}}{1-e^{-(\beta_i^*+\mathrm{i}\pi)}}$, we find

$$\begin{aligned}\Omega_{\vec{r}^*} &= \prod_{i=1}^{n}\int_{-u_i^*}^{u_i^*}\frac{\mathrm{d}z_i}{2\pi\mathrm{i}}\sum_{k=0}^{m}\binom{m}{k}\left(\frac{z_i}{z_i+1}\right)^{-r_i^*-1}\frac{(z_i)^k}{(z_i+1)^2} \\ &= \prod_{i=1}^{n}\int_{-u_i^*}^{u_i^*}\frac{\mathrm{d}z_i}{2\pi\mathrm{i}}\sum_{k=0}^{m}\binom{m}{k}(z_i+1)^{r_i^*-1}(z_i)^{k-r_i^*-1} \\ &= \prod_{i=1}^{n}\int_{-u_i^*}^{u_i^*}\frac{\mathrm{d}z_i}{2\pi\mathrm{i}}\sum_{k=0}^{m}\binom{m}{k}\sum_{l=0}^{r_i^*-1}\binom{r_i^*-1}{l}z_i^{l+k-r_i^*-1}.\end{aligned}$$

Now, according to Cauchy's residue theorem, only when $l+k=r_i^*$ we get a non-zero value. This allows us to further write

$$\Omega_{\vec{r}^*} = \prod_{i=1}^{n}\sum_{k=1}^{m}\binom{m}{k}\binom{r_i^*-1}{k-1} = \prod_{i=1}^{n}\binom{m+r_i^*-1}{r_i^*},$$

which coincides with Eq. (3.74) in the main text.

Appendix 3.C Two-sided local constraints

We now discuss ensembles of binary and weighted $n\times m$ matrices with two-sided local constraints. In this case $\vec{C}(\mathbf{G})$ is $(n+m)$-dimensional ($K=n+m$) and specified by the two vectors $(\vec{r}(\mathbf{G}),\vec{c}(\mathbf{G}))$, where $\vec{r}(\mathbf{G})$ is still the n-dimensional vector of row sums of the matrix $\mathbf{G}$ (as in the one-sided case) and, additionally, $\vec{c}(\mathbf{G})$ is the m-dimensional vector of column sums of $\mathbf{G}$, with entries $c_j(\mathbf{G}) = \sum_{i=1}^{n} g_{ij}$ ($j = 1,m$). The corresponding Lagrange multipliers take the form $(\vec{\alpha},\vec{\beta})$ where $\vec{\alpha}$ is n-dimensional and coupled to $\vec{r}(\mathbf{G})$, while $\vec{\beta}$ is m-dimensional and coupled to $\vec{c}(\mathbf{G})$.

The corresponding Hamiltonian is

$$\begin{aligned} H(\mathbf{G},\vec{\alpha},\vec{\beta}) &= \sum_{i=1}^{n}\alpha_i r_i(\mathbf{G}) + \sum_{j=1}^{m}\beta_j c_j(\mathbf{G}) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i+\beta_j)g_{ij}. \end{aligned} \tag{3.196}$$

As usual, the binary and weighted cases are discussed separately below.

3.C.1 Binary matrices under two-sided local constraints

Starting from the Hamiltonian in Eq. (3.196), the partition function of the canonical binary matrix ensemble can be still calculated exactly as a simple generalization of Eq. (3.169):

$$\begin{aligned} Z(\vec{\alpha},\vec{\beta}) &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-H(\mathbf{G},\vec{\alpha},\vec{\beta})} \\ &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i+\beta_j)g_{ij}} \\ &= \sum_{\mathbf{G}\in\mathcal{G}}\prod_{i=1}^{n}\prod_{j=1}^{m} e^{-(\alpha_i+\beta_j)g_{ij}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{g_{ij}=0,1} e^{-(\alpha_i+\beta_j)g_{ij}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}[1+e^{-(\alpha_i+\beta_j)}]. \end{aligned} \tag{3.197}$$

The resulting probability is

$$\begin{aligned} P_{\text{can}}(\mathbf{G}|\vec{\alpha},\vec{\beta}) &= \frac{e^{-H(\mathbf{G},\vec{\alpha},\vec{\beta})}}{Z(\vec{\alpha},\vec{\beta})} \\ &= \frac{e^{-\vec{\alpha}\cdot\vec{r}(\mathbf{G})-\vec{\beta}\cdot\vec{c}(\mathbf{G})}}{\prod_{i=1}^{n}\prod_{j=1}^{m}[1+e^{-(\alpha_i+\beta_j)}]} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{e^{-(\alpha_i+\beta_j)g_{ij}}}{1+e^{-(\alpha_i+\beta_j)}}. \end{aligned} \tag{3.198}$$

As in the case of binary matrices under global and one-sided local constraints, each entry g_{ij} of the matrix $\mathbf{G}$ is still an independent and Bernoulli-distributed random

variable, now controlled by the two parameters α_i and β_j. We can write the probability of g_{ij} as

$$p(g_{ij}|\vec{\alpha},\vec{\beta}) = \frac{e^{-(\alpha_i+\beta_j)g_{ij}}}{1+e^{-(\alpha_i+\beta_j)}}. \tag{3.199}$$

The expected value of g_{ij} is now

$$\langle g_{ij}\rangle_{\vec{\alpha},\vec{\beta}} \equiv \sum_{g_{ij}=0,1} g_{ij}\, p(g_{ij}|\vec{\alpha},\vec{\beta}) = \frac{e^{-(\alpha_i+\beta_j)}}{1+e^{-(\alpha_i+\beta_j)}} \tag{3.200}$$

and the variance is

$$\begin{aligned} \mathrm{Var}_{\vec{\alpha},\vec{\beta}}[g_{ij}] &\equiv \langle g_{ij}^2\rangle_{\vec{\alpha},\vec{\beta}} - \langle g_{ij}\rangle^2_{\vec{\alpha},\vec{\beta}} \\ &= \frac{e^{-(\alpha_i+\beta_j)}}{[1+e^{-(\alpha_i+\beta_j)}]^2}. \end{aligned} \tag{3.201}$$

The resulting expected values of the constraints are

$$\langle r_i\rangle_{\vec{\alpha},\vec{\beta}} = \sum_{j=1}^{m} \frac{e^{-(\alpha_i+\beta_j)}}{1+e^{-(\alpha_i+\beta_j)}}, \quad i=1,n, \tag{3.202}$$

$$\langle c_j\rangle_{\vec{\alpha},\vec{\beta}} = \sum_{i=1}^{n} \frac{e^{-(\alpha_i+\beta_j)}}{1+e^{-(\alpha_i+\beta_j)}}, \quad j=1,m. \tag{3.203}$$

The unique parameter values $(\vec{\alpha}^*,\vec{\beta}^*)$ that maximize the likelihood are found as usual by imposing that the expected values $(\langle\vec{r}\rangle_{\vec{\alpha}^*,\vec{\beta}^*}, \langle\vec{c}\rangle_{\vec{\alpha}^*,\vec{\beta}^*})$ match the desired values $(\vec{r}^*,\vec{c}^*)$. Unfortunately, in this case the values $(\vec{\alpha}^*,\vec{\beta}^*)$ cannot be determined analytically as a function of $(\vec{r}^*,\vec{c}^*)$, but they are defined implicitly by imposing

$$(\vec{r}^*,\vec{c}^*) = (\langle\vec{r}\rangle_{\vec{\alpha}^*,\vec{\beta}^*}, \langle\vec{c}\rangle_{\vec{\alpha}^*,\vec{\beta}^*}), \tag{3.204}$$

which leads to Eqs. (3.101) and (3.102) in the main text.

3.C.2 Weighted matrices under two-sided local constraints

In the canonical ensemble of weighted matrices under two-sided local constraints, the partition function is the following generalization of Eq. (3.183):

$$\begin{aligned} Z(\vec{\alpha},\vec{\beta}) &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-H(\mathbf{G},\vec{\alpha},\vec{\beta})} \\ &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i+\beta_j)g_{ij}} \\ &= \sum_{\mathbf{G}\in\mathcal{G}}\prod_{i=1}^{n}\prod_{j=1}^{m} e^{-\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i+\beta_j)g_{ij}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{g_{ij}=0}^{+\infty} e^{-(\alpha_i+\beta_j)g_{ij}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{1}{1-e^{-(\alpha_i+\beta_j)}}. \end{aligned} \tag{3.205}$$

The resulting canonical probability is

$$\begin{aligned} P_{\rm can}(\mathbf{G}|\vec{\alpha},\vec{\beta}) &= \frac{e^{-H(\mathbf{G},\vec{\alpha},\vec{\beta})}}{Z(\vec{\alpha},\vec{\beta})} \\ &= \frac{e^{-\vec{\alpha}\cdot\vec{r}(\mathbf{G})-\vec{\beta}\cdot\vec{c}(\mathbf{G})}}{\prod_{i=1}^{n}\prod_{j=1}^{m}\left[1-e^{-(\alpha_i+\beta_j)}\right]^{-1}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{e^{-(\alpha_i+\beta_j)g_{ij}}}{\left[1-e^{-(\alpha_i+\beta_j)}\right]^{-1}}. \end{aligned} \tag{3.206}$$

As in the case of weighted matrices under global and one-sided local constraints, each entry g_{ij} of the matrix $\mathbf{G}$ is an independent geometrically distributed random variable defined by the probability

$$p(g_{ij}|\vec{\alpha},\vec{\beta}) = e^{-(\alpha_i+\beta_j)g_{ij}}[1-e^{-(\alpha_i+\beta_j)}], \tag{3.207}$$

which is now controlled by the entry-specific pair of parameters α_i,β_j. The expected value of g_{ij} is

$$\langle g_{ij}\rangle_{\vec{\alpha},\vec{\beta}} \equiv \sum_{g_{ij}=0}^{+\infty} g_{ij}p(g_{ij}|\vec{\alpha},\vec{\beta}) = \frac{e^{-(\alpha_i+\beta_j)}}{1-e^{-(\alpha_i+\beta_j)}} \tag{3.208}$$

and the variance is

$$\begin{aligned} {\rm Var}_{\vec{\alpha},\vec{\beta}}[g_{ij}] &\equiv \langle g_{ij}^2\rangle_{\vec{\alpha},\vec{\beta}} - \langle g_{ij}\rangle_{\vec{\alpha},\vec{\beta}}^2 \\ &= \frac{e^{-(\alpha_i+\beta_j)}}{[1-e^{-(\alpha_i+\beta_j)}]^2}. \end{aligned} \tag{3.209}$$

The expected values of the constraints are

$$\langle r_i \rangle_{\vec{\alpha},\vec{\beta}} = \sum_{j=1}^{m} \frac{e^{-(\alpha_i+\beta_j)}}{1-e^{-(\alpha_i+\beta_j)}}, \quad i=1,n, \tag{3.210}$$

$$\langle c_j \rangle_{\vec{\alpha},\vec{\beta}} = \sum_{i=1}^{n} \frac{e^{-(\alpha_i+\beta_j)}}{1-e^{-(\alpha_i+\beta_j)}}, \quad j=1,m. \tag{3.211}$$

As in the two-sided binary case, the values $(\vec{\alpha}^*, \vec{\beta}^*)$ maximizing the likelihood cannot be determined analytically as a function of the empirical values $(\vec{r}^*, \vec{c}^*)$, but they are defined implicitly by imposing the equality

$$(\vec{r}^*, \vec{c}^*) = (\langle \vec{r} \rangle_{\vec{\alpha}^*,\vec{\beta}^*}, \langle \vec{c} \rangle_{\vec{\alpha}^*,\vec{\beta}^*}) \tag{3.212}$$

between the empirical and the expected values of the constraints. This equality leads to Eqs. (3.112) and (3.113) in the main text.

3.C.3 Determinant of the covariance matrix for two-sided local constraints

The covariance matrix $(\mathbf{\Sigma}^*)^{\pm}$ in binary $(+)$ and weighted $(-)$ ensembles of matrices under two-sided local constraints is an $(n+m)\times(n+m)$ matrix. It contains all covariances among the n row sums, all covariances among the m column sums, and all the covariances between row and column sums. If we order the constraints by considering first the n row sums $\vec{r}^*$ and then the m column sums $\vec{c}^*$ into the $(n+m)$-dimensional vector $\vec{C}^* = (\vec{r}^*, \vec{c}^*)$, and combine Eqs. (3.197) and (3.205) into the general partition function

$$Z^{\pm}(\vec{\alpha},\vec{\beta}) = \prod_{i=1}^{n}\prod_{j=1}^{m}[1 \pm e^{-(\alpha_i+\beta_j)}]^{\pm 1} \tag{3.213}$$

valid for binary (+) and weighted (−) matrices, we can determine the entries of $\mathbf{\Sigma}^{\pm}$ by applying the definition in Eq. (3.23). This yields

$$\Sigma_{ij}^{\pm} = \begin{cases} \dfrac{\partial^2 \ln Z^{\pm}(\vec{\alpha},\vec{\beta})}{\partial\alpha_i \partial\alpha_j} & i,j \in [1,n], \\ \dfrac{\partial^2 \ln Z^{\pm}(\vec{\alpha},\vec{\beta})}{\partial\alpha_i \partial\beta_{j-n}} & i \in [1,n],\ j \in [n+1,n+m] \\ \dfrac{\partial^2 \ln Z^{\pm}(\vec{\alpha},\vec{\beta})}{\partial\alpha_{i-n} \partial\beta_j} & i \in [n+1,n+m],\ j \in [1,n] \\ \dfrac{\partial^2 \ln Z^{\pm}(\vec{\alpha},\vec{\beta})}{\partial\beta_{i-n} \partial\beta_{j-n}} & i,j \in [n+1,n+m] \end{cases}$$

$$= \begin{cases} \mathrm{Cov}^{\pm}_{\vec{\alpha},\vec{\beta}}[r_i, r_j] & i,j \in [1,n], \\ \mathrm{Cov}^{\pm}_{\vec{\alpha},\vec{\beta}}[r_i, c_{j-n}] & i \in [1,n],\ j \in [n+1,n+m] \\ \mathrm{Cov}^{\pm}_{\vec{\alpha},\vec{\beta}}[c_{i-n}, r_j] & i \in [n+1,n+m],\ j \in [1,n] \\ \mathrm{Cov}^{\pm}_{\vec{\alpha},\vec{\beta}}[c_{i-n}, c_{j-n}] & i,j \in [n+1,n+m] \end{cases}$$

and, following Eq. (3.22),

$$(\Sigma^*_{ij})^{\pm} = (\Sigma_{ij})^{\pm}\big|_{(\vec{\alpha},\vec{\beta})=(\vec{\alpha}^*,\vec{\beta}^*)}\,. \tag{3.214}$$

It is easy to see that $(\mathbf{\Sigma}^*)^{\pm}$ is a combination of four blocks

$$(\mathbf{\Sigma}^*)^{\pm} = \begin{bmatrix} (\mathbf{A}^*)^{\pm} & (\mathbf{B}^*)^{\pm} \\ (\mathbf{C}^*)^{\pm} & (\mathbf{D}^*)^{\pm} \end{bmatrix}, \tag{3.215}$$

where each block has entries as described below. What determines these entries is the elements g_{ij} of the (binary or weighted) adjacency matrix $\mathbf{G}$ that different constraints have in common. The covariance between contraints that have no g_{ij} in common is zero (as such constraints are independent), while the covariance between constraints that share a term g_{ij} receives from that term a contribution equal to

$$\mathrm{Var}^{\pm}_{\vec{\alpha}^*,\vec{\beta}^*}[g_{ij}] = \frac{e^{-(\alpha_i^*+\beta_j^*)}}{[1 \pm e^{-(\alpha_i^*+\beta_j^*)}]^2}, \tag{3.216}$$

obtained combining Eqs. (3.201) and (3.209).

- Block $(\mathbf{A}^*)^{\pm}$ is the $n \times n$ covariance matrix between the row sums, with entries

$$\begin{aligned} (A^*_{ij})^{\pm} &= \mathrm{Cov}^{\pm}_{\vec{\alpha}^*,\vec{\beta}^*}[r_i, r_j] \\ &= \left.\frac{\partial^2 \ln Z^{\pm}(\vec{\alpha},\vec{\beta})}{\partial\alpha_i \partial\alpha_j}\right|_{(\vec{\alpha},\vec{\beta})=(\vec{\alpha}^*,\vec{\beta}^*)} \\ &= \delta_{ij} \sum_{k=1}^{m} \frac{e^{-(\alpha_i^*+\beta_k^*)}}{\left[1 \pm e^{-(\alpha_i^*+\beta_k^*)}\right]^2}. \end{aligned} \tag{3.217}$$

Note that $(\mathbf{A}^*)^{\pm}$ is a diagonal matrix, since different row sums are all independent.

- Block $(\mathbf{B}^*)^{\pm}$ is the $n \times m$ matrix of covariances between row sums and column sums, with entries

$$\begin{aligned}(B^*_{ij})^{\pm} &= \mathrm{Cov}^{\pm}_{\vec{\alpha}^*,\vec{\beta}^*}[r_i, c_j] \\ &= \left.\frac{\partial^2 \ln Z^{\pm}(\vec{\alpha},\vec{\beta})}{\partial\alpha_i \partial\beta_j}\right|_{(\vec{\alpha},\vec{\beta})=(\vec{\alpha}^*,\vec{\beta}^*)} \\ &= \frac{e^{-(\alpha^*_i+\beta^*_j)}}{\left[1 \pm e^{-(\alpha^*_i+\beta^*_j)}\right]^2}, \end{aligned} \tag{3.218}$$

where we now see that the matrix is not diagonal, as reach row sum r_i shares the entry g_{ij} with the column sum c_j.

- Similarly, block $(\mathbf{C}^*)^{\pm}$ is the $m \times n$ matrix of covariances between column sums and row sums, and is therefore the transpose of $(\mathbf{B}^*)^{\pm}$, as follows also from the fact that $(\mathbf{\Sigma}^*)^{\pm}$ must be symmetric. Indeed its entries are

$$\begin{aligned}(C^*_{ij})^{\pm} &= \mathrm{Cov}^{\pm}_{\vec{\alpha}^*,\vec{\beta}^*}[c_i, r_j] \\ &= \left.\frac{\partial^2 \ln Z^{\pm}(\vec{\alpha},\vec{\beta})}{\partial\alpha_j \partial\beta_i}\right|_{(\vec{\alpha},\vec{\beta})=(\vec{\alpha}^*,\vec{\beta}^*)} \\ &= \frac{e^{-(\alpha^*_j+\beta^*_i)}}{\left[1 \pm e^{-(\alpha^*_j+\beta^*_i)}\right]^2}. \end{aligned} \tag{3.219}$$

- Finally, block $(\mathbf{D}^*)^{\pm}$ is the $m \times m$ matrix of covariances among the column sums, with entries

$$\begin{aligned}(D^*_{ij})^{\pm} &= \mathrm{Cov}^{\pm}_{\vec{\alpha}^*,\vec{\beta}^*}[c_i, c_j] \\ &= \left.\frac{\partial^2 \ln Z^{\pm}(\vec{\alpha},\vec{\beta})}{\partial\beta_i \partial\beta_j}\right|_{(\vec{\alpha},\vec{\beta})=(\vec{\alpha}^*,\vec{\beta}^*)} \\ &= \delta_{ij} \sum_{k=1}^{n} \frac{e^{-(\alpha^*_k+\beta^*_j)}}{\left[1 \pm e^{-(\alpha^*_k+\beta^*_j)}\right]^2}. \end{aligned} \tag{3.220}$$

Like $(\mathbf{A}^*)^{\pm}$, $(\mathbf{D}^*)^{\pm}$ is a diagonal matrix, since different column sums are all independent.

Combining Eqs. (3.215), (3.217), (3.218), (3.219) and (3.220) proves Eq. (3.124) in the main text.

Now, in order to calculate the scaling of the determinant of $(\mathbf{\Sigma}^*)^{\pm}$, we follow the definition by Leibniz as

$$\det[(\mathbf{\Sigma}^*)^{\pm}] = \sum_{\sigma \in \mathbf{Z}_{n+m}} \mathrm{sgn}(\sigma) \prod_{l=1}^{n+m} (\Sigma^*_{l,\sigma_l})^{\pm}, \tag{3.221}$$

where σ is a permutation of the first $n+m$ integers that exchanges (without replacement) each of these integers i with another such integer $j = \sigma_i$, $\mathbf{Z}_{n+m}$ is the set of all such $(n+m)!$ permutations, and the symbol $\mathrm{sgn}(\sigma)$ represents the *parity* of σ: $\mathrm{sgn}(\sigma) = +1$ when σ is an even permutation (i.e. obtained by combining an even number of pairwise exchanges of the type $j = \sigma_i$ and $i = \sigma_j$) and $\mathrm{sgn}(\sigma) = -1$ when σ is an odd permutation (i.e. obtained by combining an odd number of pairwise exchanges). Let us call σ^0 the identity permutation, i.e. the one such that $\sigma^0_i = i$ for all i, and $\mathbf{Z}^0_{n+m} \equiv \mathbf{Z}_{n+m} \backslash \sigma^0$ the set of all other permutations. Clearly, $\mathrm{sgn}(\sigma^0) = +1$ because σ^0 involves an even number (zero) of exchanges. We can therefore rewrite Eq. (3.221) as

$$\det[(\mathbf{\Sigma}^*)^{\pm}] = \Delta^0 + \Delta' \tag{3.222}$$

where

$$\Delta^0 = \prod_{l=1}^{n+m} (\Sigma^*_{l,\sigma^0_l})^{\pm} = \prod_{i=1}^{n} (A^*_{ii})^{\pm} \prod_{j=1}^{m} (D^*_{jj})^{\pm} \tag{3.223}$$

is the product of the diagonal entries of $(\mathbf{\Sigma}^*)^{\pm}$ and

$$\Delta' = \sum_{\sigma \in \mathbf{Z}^0_{n+m}} \mathrm{sgn}(\sigma) \prod_{l=1}^{n+m} (\Sigma^*_{l,\sigma_l})^{\pm}. \tag{3.224}$$

We are going to show that Δ' is at most of the same order of Δ^0. Setting $c_n = 1/n$ in the sparse regime (for which we recall that $m = O(n)$ necessarily) and $c_n = 1$ in the dense regime (for which m can be either finite or $O(n)$), we note that each entry of the blocks $(\mathbf{B}^*)^{\pm}$ and $(\mathbf{C}^*)^{\pm}$ is of order $O(c_n)$, while each of the diagonal entries of block $(\mathbf{A}^*)^{\pm}$ is of order $O(c_n m)$ and each of the diagonal entries of block $(\mathbf{D}^*)^{\pm}$ is of order $O(c_n n)$. In general, the order of Δ^0 is therefore

$$\Delta^0 = O\left((c_n n)^m (c_n m)^n\right) \tag{3.225}$$

as clear from Eq. (3.223).

To this end, we note that each permutation σ appearing in Eq. (3.224) can be expressed as a combination of a certain number (say $a > 0$) of exchanges of pairs of the first $n+m$ integers. It is easy to see that all the $\binom{n}{2}$ exchanges of pairs of the first n integers give a zero contribution to Δ', because they lead to terms of the type $(\Sigma^*_{i,j})^{\pm} = 0$ where $i, j \in [0, n]$ with $i \neq j$ (combination of exchanges that lead again

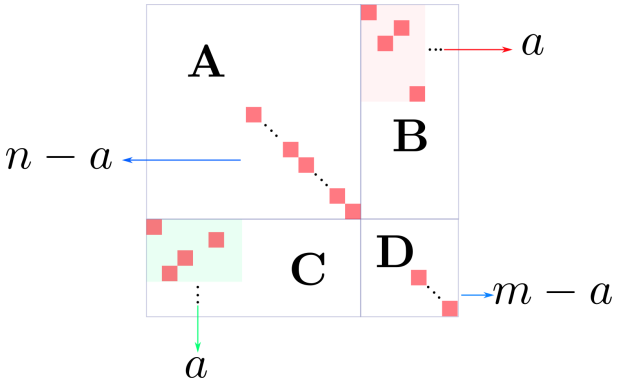

Figure 3.4. Illustration of the permutations producing the non-zero contributions to the determinant of the covariance matrix, as described in the text.

to $i = j$ are such that $j = \sigma_i = i$ and therefore do not lead to new permutations: they are already accounted for in permutations with lower a). Similarly, all the $\binom{m}{2}$ exchanges of pairs of the next m integers give a zero contribution to Δ', because they lead to terms of the type $(\Sigma^*_{i,j})^{\pm} = 0$ where $i, j \in [n+1, n+m]$ with $i \neq j$. Therefore the only exchanges leading to nonzero contributions to Δ' are the nm exchanges across the first n integers and the next m integers, i.e. those that lead to terms $(\Sigma^*_{i,j})^{\pm} > 0$ where $i \in [0, n]$ and $j \in [n+1, n+m]$ or $j \in [0, n]$ and $i \in [n+1, n+m]$ in Eq. (3.224). Each of these nontrivial contributing permutations involves a (unrepeated) exchanges of integers, where $a \in [1, nm]$. Compared with the identity σ_0, each of these permutations replaces a of the first n diagonal entries and a of the next m diagonal entries of $(\mathbf{\Sigma}^*)^{\pm}$ appearing in Eq. (3.223) with a number $2a$ of non-zero off-diagonal entries in blocks $(\mathbf{B}^*)^{\pm}$ and $(\mathbf{C}^*)^{\pm}$ (see Fig. 3.4).

Each such permutation therefore gives a contribution of order $(c_n n)^{m-a}(c_n m)^{n-a}$ to the summation in Eq. (3.224). Individually, each such contribution is subleading with respect to the term Δ^0. However, collectively all the contributions involving the same number a of exchanges contribute a term of order $E_a(c_n n)^{m-a}(c_n m)^{n-a}$ where E_a is the number of unrepeated exchanges of a pairs. The order of E_a is given by the number of distinct choices of a exchanges out of the nm possible ones, which is $\binom{nm}{a}$. This estimate does not control for the fact that, for each a, some of the exchanges reduce to simpler permutations already accounted for by smaller values of a, however the leading order is correct. Since nm is large, we can apply Stirling's approximation to $(nm)!$ and estimate the order of E_a as

$$E_a = O\left(\binom{nm}{a}\right) = O\left(\frac{(nm)^a}{a!}\right) \tag{3.226}$$

as in Eq. (3.82). Therefore all the permutations realized by a exchanges collectively

give a contribution of order

$$E_a(c_n n)^{m-a}(c_n m)^{n-a} = O\left(\frac{(c_n n)^m (c_n m)^n}{a!}\right) \tag{3.227}$$

and sign $(-1)^a$ to the sum in Eq. (3.224), so Δ' can be rewritten as a sum over a (with $a = 1, nm$) of terms of alternating sign. Qualitatively, and with an abuse of notation, the order of the entire sum defining $\det[(\mathbf{\Sigma}^*)^{\pm}]$ in Eq. (3.221) is (except for accidental cancellations due to particular combinations of values of the entries of $(\mathbf{\Sigma}^*)^{\pm}$)

$$\begin{aligned} O(\det[(\mathbf{\Sigma}^*)^{\pm}]) &= \sum_{a=1}^{nm} O\left(\frac{(-1)^a (c_n n)^m (c_n m)^n}{a!}\right) \\ &= O\left((c_n n)^m (c_n m)^n e^{-1}\right) \\ &= O\left((c_n n)^m (c_n m)^n\right). \end{aligned} \tag{3.228}$$

We therefore see that the order of Δ' does not exceed that of Δ^0, so the leading order of $\det[(\mathbf{\Sigma}^*)^{\pm}]$ is

$$\det[(\mathbf{\Sigma}^*)^{\pm}] = O(\Delta^0) = O\left((c_n n)^m (c_n m)^n\right). \tag{3.229}$$

In other words, the off-diagonal terms of $(\mathbf{\Sigma}^*)^{\pm}$ do not alter the order obtained by multiplying the diagonal terms. For finite m, we therefore have

$$\begin{aligned} \alpha_n^{\pm} &= \ln\sqrt{\det\left[2\pi(\mathbf{\Sigma}^*)^{\pm}\right]} \\ &= O\left(m\ln(c_n n) + n\ln(c_n m)\right). \end{aligned} \tag{3.230}$$

In the sparse case where $c_n = 1/n$ and $m = O(n)$, we have

$$\alpha_n^{\pm} = O\left(n\right), \tag{3.231}$$

while in the dense case with $c_n = 1$ and $m = O(n)$ we have

$$\alpha_n^{\pm} = O\left(n\ln n\right), \tag{3.232}$$

and finally in the dense case with $c_n = 1$ and finite m we have

$$\alpha_n^{\pm} = O\left(n\right), \tag{3.233}$$

confirming the same scalings for the relative entropy obtained in Eqs. (3.83), (3.84) and (3.85) for the one-sided case.

Chapter 4

New information-theoretic bounds for systems with local constraints

Abstract

The information-theoretic bounds are the limit of space to store the information generated by the information source and the limit of speed to reliable transmit information through a channel. In classical information theory, those bounds are determined by the Shannon entropy of the information sources. However, recent research shows that information sources in non-physical systems such as social networks or nervous systems are not a single variable with finite outcomes but a composition of numerous interacting units. Furthermore, these heterogeneous dependencies imply local constraints in those information sources. Thus, to find the new information-theoretical bounds of them, statistical ensembles with local constraints are used to describe those new information sources in this work. We find that under ensemble equivalence, information-theoretical bounds of information sources described by different statistical ensembles are equivalent. When heterogeneous dependencies implied local constraints break the ensemble equivalence, the information storage space of the information source described by the microcanonical ensemble with hard constraints is smaller than that of the canonical ensemble one with soft constraints. The extra sequences in the typical set of the canonical ensemble described sources have the same sum of *Hamiltonian* with the microcanonical ensemble one. But the constraints of each state in the sequence are not equal to hard constraints. Therefore, there is a tradeoff between the choosing of different ensembles to describe those information sources. Using the microcanonical one with hard constraints costs more calculation to obtain the probability distribution but requires less information storage space. Choosing the canonical one needs more space to store the information but requires less calculation to hold the 'soft' constraints[1].

[1]This chapter is based on the coming paper:
Qi Zhang, Diego Garlaschelli, "New information-theoretic bounds for systems with local constraints" (2021)

4.1 Introduction

The birth of modern information theory can be traced to 1948 when Shannon gave the first quantifiable definition of information in his fundamental paper [33]. It is believed to be promoted by the rapid progress of electronic communication systems in the first half of the 20th century. When all the engineers in the communication industry are desire to know what the smallest space needs to store the information that is generated by different information sources, and what is the maximum speed of reliable information transmission through a channel, i.e., the information-theoretical bounds of the communication systems [29].

To find the information-theoretical bounds of the communication systems, Shannon creatively divides the communication systems into three parts: the information source, the channel and the receiver. And all of them are described by the probability theory, e.g., information sources and the receivers are described by random variables; the channel used to transport the information is modelled by conditional probability. The smallest space needs to store the information generated by information sources is decided by the information entropy, which is the probabilistic uncertainty of the information source [33]. The maximum speed of reliable information transmission (channel capacity) is equal to the mutual information between the information source and receiver, which is determined by the conditional probability that is used to describe the channel [29].

Compared with 1948, the information needs to store and transmit in natural and artificial systems right now is much more complex, e.g., the activity of the neurons in the nervous system [9], the appearing of retweets and comments in a social network [36]. Information sources in these systems are not a single variable in the traditional information theory. Instead, they have numerous units, and almost all units are entangled with each other by different interactions. Thus, using the random variables with finite outcomes to model those new information sources is impossible, as the single variable can not describe the heterogeneous interactions among units.

Actually, signal generation by these information sources is not like the sampling of a random variable. It is closer to the change of the particles' status in the thermodynamic system under localized macroscopic properties [27]. Thus, the space to store the information that is generated by the billions of users in Twitter when there is break news in the real world, or the limit of information storage in the nervous system like the brain is equivalent to the quantify of the macroscopic property in thermodynamic systems with numerous particles. Fortunately, the signal generation at different times by those new information sources are still independent. Thus, we can not use the random variable with finite outcomes to describe those information sources. But we can use statistical ensembles from statistical physics to describe the status change of those heterogeneous interacted units and find the new information-theoretical bounds [1, 37, 27, 28].

In statistical physics, statistical ensembles are introduced by Gibbs to model the macroscopic properties of the numerous particles in the thermodynamic system from microscopic behaviour of them based on the probability theory [1]. In traditional

statistical physics, the particles are identical, and this is why all the ensembles are under global constraints, such as the fixed total energy or temperature. However, in the new information sources, the heterogeneous interacted units are not identical. The heterogeneous interactions among all units will imply local constraints. Thus, to describe new information sources, we need the statistical ensembles with localized macroscopic constraints [27].

In ensemble theory, systems with different constraints will be described by different statistical ensembles [1]. The *microcanonical ensemble* is used to describe systems with fixed total energy E^*, and the *canonical ensemble* is used to describe the system with fixed temperature $\beta = 1/KT$, where K is the Boltzmann constant, T is the absolute temperature. Obviously, from the energy isolation, the restriction in the microcanonical ensemble is harder than that in the canonical ensemble [15]. Thus, when the local constraints implied by the heterogeneous interactions among the units in the information sources have a different macroscopic property (hard or soft), the information sources also need to be described by different statistical ensembles. When the local constraints are hard, those information sources need to be described by the microcanonical ensemble. When the local constraints are soft, those information sources need to be described by the canonical ensemble [26, 45]. The two ensembles will conjugate with each other by setting the parameter $\beta = \beta^*$ to make the average total energy in the canonical ensemble equal to the fixed total energy in the microcanonical ensemble,$\langle E\rangle = E^*$.

When the system has finite sizes, the two ensembles are certainly different. But in the thermodynamic limit (number of particles goes to infinite), the fluctuation of constraints in the canonical ensemble will vanish. The microcanonical ensemble can be replaced by the canonical ensemble, which is mathematically easy to calculate [8]. This phenomenon is called ensemble equivalence (EE). The existence of EE also shows that the information carried by different ensemble descriptions of the thermodynamic system is the same. However, recent research on networks and system with long-range interactions also show that in the boundary of phase transitions or when the system is under extensive local constraints, this ensemble equivalence will be broken [8, 5, 27]. This ensemble nonequivalence (EN) will directly influence the information-theoretical bounds of the new information sources with different statistical ensembles descriptions under local constraints. As detecting the limits needs the length of the sequences used to record the status changing in the information sources goes to infinite.

In Shannon's setting, the information generated by the random variable described information source is carried by the sequences use to record the status changing of the information source, and most of the information is carried by equiprobable sequences that belong to the typical set of it [29]. Thus, the smallest space of information storage is determined by the size of the typical set. And the influence of ensemble nonequivalence in information sources with heterogeneous interacted units will be manifested by the difference between the typical sets of different ensembles.

In this chapter, the statistical ensembles with local constraints are used to describe the information sources with heterogeneous interacting units to find the new information-theoretical bounds of them [27]. As the extensive local constraints in the

new information source will lead the EN, we also need to check the influence of EN on the new information-theoretical bounds. We find that the information storage space of the microcanonical ensemble described information sources is smaller than the conjugate canonical ensemble one. As the typical set of the microcanonical ensemble is smaller than the conjugate canonical ensemble. The extra sequences in the typical set of the canonical ensemble have the same sum of *Hamilitonian* as sequences in the typical set of the microcanonical ensemble. But the local constraints of information sources are not the same. This result shows that using the microcanonical ensemble with hard constraints to describe information sources with heterogeneous interacting units needs less space to store the information generated by it than the conjugate canonical ensembles. But it needs cost more power on the calculations to obtain the probability distributions. Therefore, there is a tradeoff in the choosing of ensembles. Using the microcanonical ensemble needs less information storage space but cost more energy in the initial calculation. Using the canonical ensemble will reduce the energy cost in the initial calculation but requires more information storage space.

4.2 Ensemble described information sources

To describe those new information sources, we need a reasonable mathematical model to quantify the heterogeneous interactions. Actually, the quantification of interactions among different units is not a new problem in scientific research. Networks model as a specific case of the random matrix has already been widely used in different research fields to describe those interactions [12, 14]. Thus, in the following discussion, information sources with heterogeneous interacted units will be described by the random matrix with marginal sums as local constraints [27]. According to the traditional information theory, the limit of space to store the information generated by the random variable described information source is determined by the uncertainty of the information source. Therefore, to find new information-theoretical bounds, we also need to find the probability distribution of possible states of the random matrix with local constraints based on statistical ensembles.

According to statistical physics, the appearance of each state in the thermodynamic system with numerous particles is random. But the different macroscopic properties will determine the probability of each state's presence. For example, in the microcanonical ensemble, the total energy E^* of each state is equal to each other, so each state has the same probability. In the canonical ensemble, the total energy of each state is different, but the temperature β^* is fixed. Therefore, the probability of each state in the canonical ensemble is determined by β^* and the total energy of each state [1]. Thus, the different macroscopic properties of local constraints will affect the choice of ensembles to describe it.

Here, we use the $n \times m$ matrix $\mathbf{G}$ (m can equal to n) to represent the possible configurations of the interacted units in the information sources [27]. Each unit g_{ij} in the matrix $\mathbf{G}$ represents the degree of interaction in this system, and it will have different physical means when there is a different definition of i and j. The con-

straint of the matrix is $\vec{C}(\mathbf{G})$, and it is determined by the macroscopic property of the interactions among units in the information sources If the interactions are homogeneous, the constraints will be global, which is is the sum of all units in the matrix, $C(\mathbf{G}) = \sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}$. When interactions are heterogeneous, the constraints will be localized as the column and row sum of the matrix $\vec{C}(\mathbf{G}) = [\vec{c}, \vec{r}]$. The column local constraints is a vector with m units in it, $\vec{c} = [c_1, c_2, \cdots, c_j, \cdots, c_m]$, each unit is the sum of all the elements in column j of matrix $\mathbf{G}$ as $c_j = \sum_{i=1}^{n} g_{ij}$. This represents the sum of one kind of property of all the particles in the information source. The row constraints is $\vec{r} = [r_1, r_2, \cdots, r_i, \cdots, r_n]$, each unit is equal to $r_i = \sum_{j=1}^{m} g_{ij}$ [27]. It represents the total influence of the particle i in the information source. Two examples of information sources that need to be modelled by the matrix with local constraints are shown in FIG.4.1

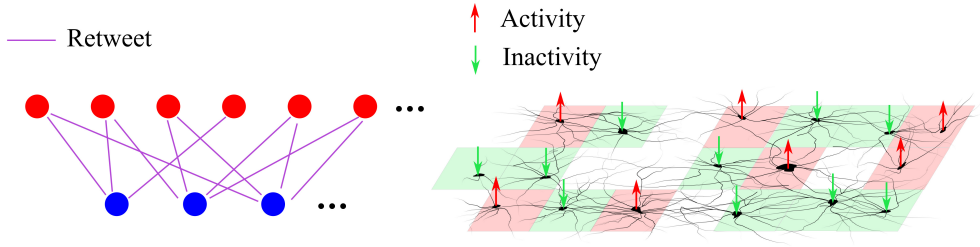

Figure 4.1. The left figure shows the retweets in social media. It can be treated as a bipartite network, where the m users retweet the n users' tweets, each g_{ji} here represents retweet from user j to user i. The right figure shows the activity of the neurons in the nervous system, i and j represents the spatial position of each neuron, g_{ji} here can be 0 or 1, to represents the activity of neurons in the position j, i. The local constraints in the two information sources can be the sum of all the activated neurons in the specific region in the nervous system or the fixed total number of retweets for each user in the social networks. The changing of all the units' state in the two systems is determined by the interactions with each other.

To analytically get the details of the interactions among the numerous units is difficult. Normally, we have the local constraints $\vec{C}(\mathbf{G})$ and the size of the information sources. Thus, to obtain the probability distribution of the states of the information source with the partial information, we need based on the maximum entropy principle introduced by Jaynes [15].

As the macroscopic property of the constraints will decide the ensemble used to describe it [7] so when the constraints are hard, each state of the information source have the same value of constraints as $\vec{C}^*$, the information source needs to be described by the microcanonical ensemble. The probability of each state in the microcanonical ensemble described information source $\mathcal{G}_{\text{mic}}$ is

$$P_{\text{mic}}(\mathbf{G}) = \Omega_{\vec{C}^*}^{-1}, \tag{4.1}$$

where $\Omega_{\vec{C}^*} = |\{\mathbf{G} \in \mathcal{G}_{\text{mic}} : \vec{C}(\mathbf{G}) = \vec{C}^*\}|$ is the number of states in microcanonical

ensemble with hard constraints $\vec{C}^*$. The Shannon entropy of the microcanonical ensemble $S_{\rm mic} = \ln \Omega_{\vec{C}^*}$ is also decided by the number of configurations in it.

If constraints are soft. Only the average value of the constraints in each state of the ensemble is required to equal to the hard constraints in the microcanonical ensemble as $\langle \vec{C}(\mathbf{G}) \rangle = \vec{C}^*$, then the information source needs to be described by the conjugate canonical ensemble. The probability of each state of the canonical ensemble described information source $\mathcal{G}_{\rm can}$ is equal to

$$P_{\rm can}(\mathbf{G}) = e^{-H(\mathbf{G},\vec{\beta}^*)} / Z(\vec{\theta}^*), \tag{4.2}$$

where $\vec{\beta}^*$ represnt parameters, which realize $\langle C(\mathbf{G}) \rangle_{\vec{\theta}^*}$ equal to C^* [15]. The partition function $Z(\vec{\beta}^*)$ is a normalization constant equal to $Z(\vec{\beta}^*) = \sum e^{-H(\mathbf{G},\vec{\beta}^*)}$, which is the sum of $e^{-H(\mathbf{G},\vec{\beta}^*)}$ of all the possible configuration of $\mathbf{G}$ in canonical ensemble. The *Hamiltonian* $H = \vec{C}(\mathbf{G}) \cdot \vec{\beta}^*$ is a liner combination of the constraints and parameter $\vec{\beta}^*$.

According to the definition of the partition function and the *Hamiltonian*, we can rewrite the probability $P_{\rm can}(\mathbf{G})$ as a product of all the interactions' probability in the information source as

$$P_{\rm can}(\mathbf{G}) = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{e^{-\beta_{ij}^* g_{ij}}}{\sum_{g_{ij} \in \mathbf{g}} e^{-\beta_{ij}^* g_{ij}}} \tag{4.3}$$

where $\mathbf{g}$ is the collection of all possible configuration of g_{ij} [27]. It means the localized interactions are independent. Thus, when the interaction is homogeneous, the parameter will be set equal to each other as $\beta_{ij}^* = \beta^*$. And then, the canonical ensemble is a $n \times m$ extension of the random variable. It means the classical description is a special case of the canonical ensemble descriptions.

In traditional information theory, the information generated by random variable described information sources are carried by sequences use to record the status changing of random variables [29]. In ensemble described information sources with local constraints, the information generated by it is carried by ensemble sequences. When using different ensembles to describe information sources, we will have different ensemble sequences. The structure of the microcanonical ensemble sequences $\mathbf{A}_{\rm mic}^{(l)}$ and canonical ensemble sequence $\mathbf{A}_{\rm can}^{(l)}$ with length l are shown in Fig.4.2.

The limit of information storage space is determined by the size of the corresponding typical set.

Let $\mathbf{G}^*$ denotes the state in the microcanonical ensemble with constraints $\vec{C}^*$, then probability of microcanonical ensemble sequence $\mathbf{A}_{\rm mic}^{(l)}$ is equal to

$$P(\mathbf{A}_{\rm mic}^{(l)}) = \prod_{k=1}^{l} P_{\rm mic}(\mathbf{G}^*) = \Omega_{\vec{C}^*}^{-l} = e^{-lS_{\rm mic}}. \tag{4.4}$$

As each state have the same probability, so all the microcanonical ensemble sequences have the same probability. According to the asymptotic equipartition property (AEP),

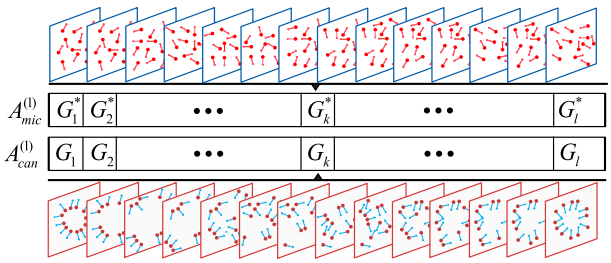

Figure 4.2. As we mentioned before, the states of the new information source are analogue to the recording of the behaviours of numerous particles. Thus, each ensemble sequence is the recording of the interacting units' behaviours in l times. In microcanonical ensemble sequences $\mathbf{A}^{(l)}_{\text{mic}}$, each state $\mathbf{G}^*_k$ have the same constraints $\vec{C}(\mathbf{G}^*_k) = \vec{C}^*$. But in canonical ensemble sequences $\mathbf{A}^{(l)}_{\text{can}}$, the total energy of each state is different. The average value of the constraints of all states should equal the hard constraints in the conjugate microcanonical ensemble $\langle \vec{C}(\mathbf{G}) \rangle = \vec{C}^*$.

microcanonical ensemble sequences have the same probability. Thus, all of them belong to the typical set of it [29].

The number of ensemble sequences in typical set of the microcanonical ensemble $|T^{\epsilon}_{\text{mic}}| = 1/P(\mathbf{A}^{(l)}_{\text{mic}})$ is equal to $e^{lS_{\text{mic}}}$. The smallest space need to store the information generated by the microcanonical ensemble described information source is equal to

$$\ln |T^{\epsilon}_{\text{mic}}| = l \times S_{\text{mic}}. \tag{4.5}$$

It is connected with the possible number of configurations in the microcanonical ensemble with hard constraints $\vec{C}^*$.

The canonical ensemble sequence $\mathbf{A}^{(l)}_{\text{can}}$ is generated by the canonical ensemble $\mathcal{G}_{\text{can}}$. According to Jaynes's work [15], the probability of each state in the canonical ensemble should realize the average value of constraints equal to hard constraints in the microcanonical ensemble as $\langle \vec{C}(\mathbf{G}) \rangle = \vec{C}^*$, and maximization the Shannon entropy S_{can} of it. Thus, probability of the canonical ensemble sequences $\mathbf{A}^{(l)}_{\text{can}}$ still equal to the production of all the l states as

$$P(\mathbf{A}^{(l)}_{\text{can}}) = \prod_{k=1}^{l} P_{\text{can}}(\mathbf{G}_k). \tag{4.6}$$

The smallest space to store the information carried by the canonical ensemble se-

quences $\mathbf{A}_{\text{can}}^{(l)}$ still can be estimated by the AEP as

$$\begin{aligned}\lim_{l\to\infty}\frac{1}{l}\ln P(\mathbf{A}_{\text{can}}^{(l)}) &= \lim_{l\to\infty}\frac{1}{l}\sum_{k=1}^{l}\ln P_{\text{can}}(\mathbf{G}_k)\\ &\to E[\ln P_{\text{can}}(\mathbf{G})]\\ &= S_{\text{can}}.\end{aligned} \tag{4.7}$$

The S_{can} is the Shannon entropy of the canonical ensemble. It is defined as $S_{\text{can}} = -\sum_{\mathbf{G}\in\mathcal{G}_{\text{can}}} P_{\text{can}}(\mathbf{G})\ln P_{\text{can}}(\mathbf{G})$. When symbol $\langle\cdot\rangle$ represents the average value, canonical entropy $S_{\text{can}} = \langle H + \ln Z(\vec{\theta}^*)\rangle$ will equal to $S_{\text{can}} = \langle\vec{C}\rangle\cdot\vec{\theta}^* + \ln Z(\vec{\theta}^*)$. In the setting of ensemble conjugation, the hard constraints equal to the average value of soft constraints as $\langle\vec{C}\rangle = \vec{C}^*$. It makes the value of S_{can} equals to $\vec{C}^*\cdot\vec{\theta}^* + \ln Z(\vec{\theta}^*)$, which is only based on the probability of state $\mathbf{G}^*$ and equal to logarithm of $P_{\text{can}}(\mathbf{G}^*)$.

As not all states in the canonical ensemble have the same probability, thus the probability of each canonical ensemble sequence may also be different. Therefore, to find the limit of information storage for the canonical ensemble described information source, we need to find the typical set of it. When use the ϵ to represent the bias between the canonical entropy function and the limit of the average value of $\ln P_{\text{can}}(\mathbf{G})$, probability of canonical ensemble sequence in typical set $\mathbf{A}_{T_{\text{can}}^{\epsilon}}^{(l)}$ have the property

$$e^{-l(S_{\text{can}}+\epsilon)} \le P(\mathbf{A}_{T_{\text{can}}^{\epsilon}}^{(l)}) \le e^{-l(S_{\text{can}}-\epsilon)}. \tag{4.8}$$

If the value of ϵ is equal to 0, then the ensemble sequences belong to the typical set of canonical ensembles can be identified by the sum of *Hamiltonian* in the ensemble sequences as

$$T_{\text{can}}^{\epsilon=0} = \{A_{\text{can}}^{(l)} | \sum_{k=1}^{l} H(\mathbf{G}_k,\vec{\theta}^*) = l\times H(\mathbf{G}^*,\vec{\theta}^*)\}. \tag{4.9}$$

This result shows that all sequences in the typical set of the conjugate microcanonical ensemble belong to the typical set of the canonical ensemble.

The number of ensemble sequences in the typical set is equal to $|T_{\text{can}}^{\epsilon=0}|$ and the smallest space needs to store the information is also equal to $\ln|T_{\text{can}}^{\epsilon=0}|$ as

$$\ln|T_{\text{can}}^{\epsilon=0}| = l\times S_{\text{can}} = l\times[\vec{C}^*\cdot\vec{\theta}^* + \ln Z(\vec{\theta}^*)]. \tag{4.10}$$

The result shows that states in the canonical ensemble with constraints $\vec{C}(\mathbf{G}) = \vec{C}^*$.determine the limit of information storage.

The generating of ensemble sequences by statistical ensembles with local constraints is independent. The space needs to store the information carried by the different kinds of ensemble sequences still decided by the Shannon entropy of the ensemble, which is used to describe the information source.

4.3 Information storage under ensemble nonequivalence

The measure-level ensemble equivalence is that the canonical probability distribution converges to the conjugate microcanonical one in the thermodynamic limit [8]. Under ensemble nonequivalence, there is always a difference between the two probability distributions, even in the thermodynamic limit. Specifically, probability of states with the constraints C^* in the canonical ensemble is smaller than that in the microcanonical ensemble $P_{\rm mic}(\mathbf{G}^*) > P_{\rm can}(\mathbf{G}^*)$. It is why the conjugate canonical ensemble always has a bigger Shannon entropy than the microcanonical ensemble [45].

The measure-level ensemble nonequivalence is easy to be shown in the probability distribution of the states as in FIG.4.3. And this difference can be quantified by the

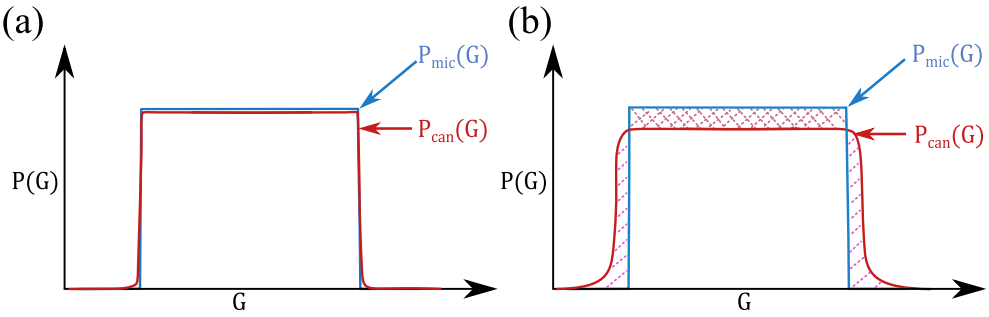

Figure 4.3. Probability distribution of states in the microcanonical and conjugate canonical ensemble under EE (a) and EN (b).

relative entropy between probability distributions of the microcanonical and canonical ensemble as

$$S(P_{\rm mic}||P_{\rm can}) = \sum_{\mathbf{G}\in\mathcal{G}_{\rm can}} P_{\rm mic}(\mathbf{G}) \ln \frac{P_{\rm mic}(\mathbf{G})}{P_{\rm can}(\mathbf{G})}. \tag{4.11}$$

The probability of states in microcanonical with constraints $\vec{C} \neq \vec{C}^*$ is equal to 0, so the relative entropy is decided by states in the two ensembles with constraints $\vec{C}^*$. And the value of it is equal to $S(P_{\rm mic}||P_{\rm can}) = \ln P_{\rm mic}(\mathbf{G}^*) - \ln P_{\rm can}(\mathbf{G}^*)$, which is the difference between the Shannon entropy of the two ensembles, $S(P_{\rm mic}||P_{\rm can}) = S_{\rm can} - S_{\rm mic}$. It directly connects with the difference of the typical set's size.

The relative entropy is difficult to obtain, as the value of $\Omega_{\vec{C}^*}$ is hard to calculate. However, according to the assumption that all the microscopic configurations in the microcanonical ensemble are the subset of the conjugate canonical ensemble, the number of configurations in the microcanonical ensemble can be estimated by the δ-function as $\Omega_{\vec{C}^*} = \sum_{\mathbf{G}\in\mathcal{G}} \int_{-\vec{\pi}}^{\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} e^{\mathrm{i}\vec{\psi}[\vec{C}^* - \vec{C}(\mathbf{G})]}$, which can be simplied as the function of the canonical probability $\Omega_{\vec{C}^*} = \int_{-\vec{\pi}}^{+\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} P_{\rm can}^{-1}(\mathbf{G}^*|\vec{\beta}^* + \mathrm{i}\vec{\psi})$ [37].

When the integration is hard to calculate we can still use the saddle-point technique to approach number of configurations in the microcanonical ensemble as

$$\Omega_{\vec{C}^*} = \frac{e^{S^*_{\text{can}}}}{\sqrt{\det(2\pi\boldsymbol{\Sigma}^*)}} \prod_{k=1}^{K} [1 + O(1/\lambda_k^*)], \tag{4.12}$$

which is based on the covariance matrix of constraints $\boldsymbol{\Sigma}^*$ [37]. Therefore, the relative entropy between the microcanonical and canonical ensemble is equal to

$$S(P_{\text{mic}}||P_{\text{can}}) = \frac{1}{2} \sum_{w=1}^{W} \ln \frac{2\pi\lambda_w^*}{[1 + O(1/\lambda_w^*)]^2}, \tag{4.13}$$

where λ_w^* is the wth no-zero eigenvalue of the covariance matrix of constraints in the canonical ensemble $\boldsymbol{\Sigma}^*$, W is the total number of constraints in the systems. When the matrix is under two-sided local constraints, the value of $W = n + m$ [27].

Each entry $\boldsymbol{\Sigma}^*_{kl}$ in the covariance matrix reprensents the covariance between local constraints C_k and C_l, $\boldsymbol{\Sigma}^*_{kl} = \text{Cov}[C_k, C_l]_{\vec{\beta}^*}$. The constraint C_k or C_l here can be the column local constraint r_i^* or c_j^* in the canonical ensemble. The value of $\boldsymbol{\Sigma}^*_{kl}$ equal to $\boldsymbol{\Sigma}^*_{kl} = \frac{\partial^2 \ln Z(\vec{\beta}^*)}{\partial\beta_k^* \partial\beta_l^*}$. It can be obtained from the partial differential of the logarithm of partition function of the canonical ensemble $\ln Z(\vec{\theta}^*)$. More details of the proofs can be found in [37].

According to Eq.(4.5) and Eq.(4.10), we can find the difference between the size of typical sets of different ensembles described information sources are connected with the relative entropy between the ensembles. This relative entropy is the indicator that is used to detect the measure-level ensemble nonequivalence. Thus, the EN in the information sources will directly affect the information-theoretical bounds of the new information sources.

As the Shannon entropy of the microcanonical ensemble is smaller than the conjugate canonical ensemble, the typical set of the microcanonical ensemble described information sources is also smaller than the canonical ensemble one. It means using the canonical ensemble to describe the information source with heterogeneous interacting units needs extra space to store the sequences that belong to the typical set of the canonical ensemble but not include in the typical set of the microcanonical ensemble. Moreover, this extra space is determined by the relative entropy between the two ensembles, or in other words, it will be affected by the degree of ensemble nonequivalence. Especially when each unit in the information source has a finite degree of freedom, there is a strong ensemble nonequivalence [27], the gap between the limit of space to store the information generated by the different ensemble described information source has the same order as the limit under the canonical ensemble descriptions.

The space requires to store the set of the extra sequences $T^{(l)}_{\tilde{\alpha}_n}$ in the typical set of the canonical ensemble but not belong to the typical set of the microcanonical

ensemble is equal to

$$\ln |T_{\tilde{\alpha}_n}^{(l)}| = l \times \frac{1}{2} \sum_{w=1}^{W} \ln \frac{2\pi\lambda_w^*}{[1+O(1/\lambda_w^*)]^2}. \tag{4.14}$$

Since sequences belong to the typical set of the canonical ensemble should have the sum of each state's *Hamiltonian* as $l \times H(\mathbf{G}^*, \vec{\theta}^*)$ and ensemble sequences generate by the microcanonical ensemble are all belong to the typical set of the canonical ensemble. Thus the extra sequences $T_{\tilde{\alpha}_n}^{\epsilon}$ should satisfy the following condition:

$$T_{\tilde{\alpha}_n} = \{T_{\text{can}}^{(l)} | \sum_{k=1}^{l} \vec{C}(\mathbf{G}_k) = l \times \vec{C}^*, \vec{C}(\mathbf{G}_k) \neq \vec{C}^*\}. \tag{4.15}$$

If χ_{can} is a collection of all the ensemble sequences generated by the canonical ensemble described information source, and χ_{mic} represents all the ensemble sequences generated by the microcanonical ensemble, then the relationship between the typical set of them under ensemble equivalence or nonequivalence is shown as in FIG.4.4.

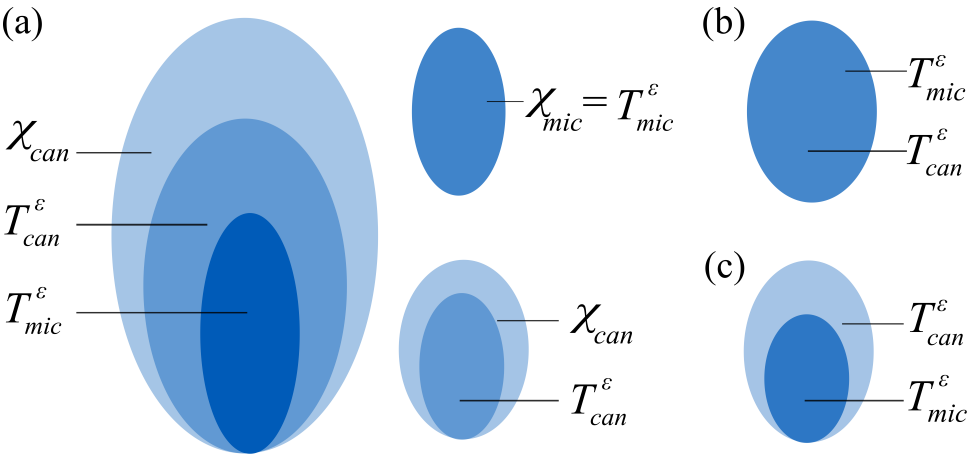

Figure 4.4. Every ensemble sequence generated by the microcanonical ensemble belongs to the typical set of it. But not all the ensemble sequences of the canonical ensemble are included in $T_{\text{can}}^{\epsilon}$. When the information source is under EE, the typical set of the microcanonical ensemble $T_{\text{mic}}^{\epsilon}$ is close to the typical set of the canonical ensemble as shown in (b). When the information sources are under EN, there is a non-vanished difference between the typical set of the canonical ensemble and the microcanonical ensemble like in (c).

Then we want to know if the difference between the information-theoretical bounds that is quantified by $\ln |T_{\tilde{\alpha}_n}^{\epsilon}|$ is bigger enough to affect the choice of the ensembles to describe information source? To solve this problem, we need to compare the limit of the extra space to store the set $T_{\tilde{\alpha}_n}$ with the space needs to store the information

that is generated by the canonical ensemble as

$$\lim_{l\to\infty} r = \frac{\ln|T^{\epsilon}_{\tilde{\alpha}_n}|}{\ln|T^{\epsilon}_{\rm can}|} = \frac{\frac{1}{2}\sum_{k=1}^{K}\frac{\ln(2\pi\lambda_k^*)}{[1+O(1/\lambda_w^*))]^2}}{-\vec{\theta}^*\cdot\vec{C}^* - \ln Z(\vec{\theta}^*)}. \tag{4.16}$$

If the limit value r is bigger than 0, then the space saved by choosing the microcanonical ensemble is bigger enough. This will happens when the information source with heterogeneous interacted units are under strong ensemble nonequivalence [27]. When the information source is under EE, the information-theoretical bounds of the different ensemble described information source is the same. Such as the information sources described by the ER model and the matrix with global constraints.

The proof details are in the appendix. When information sources need to be described by the matrix with local constraints, the ratio r is bigger than 0, and close to $r \sim \frac{\ln(2\pi m)}{m}$, where m is the degree of freedom for each unit in the information source [27]. Proof also can be found in the appendix.

4.4 Ensemble nonequivalence and channel capacity

As we already mentioned before, the information-theoretical bounds include the limit of information storage and the maximum speed of reliable information transmission through a channel. Here, to find the possible influence of EN on the channel capacity, we suppose there is a zero-less ensemble channel, which can transmit all the information generated by the ensemble described information sources correct. Then the information R should be transmitted through the channel is equal to

$$R = \begin{cases} l\times S_{\rm mic}, & \text{microcanonical ensemble} \\ l\times S_{\rm can}, & \text{canonical ensemble} \end{cases} \tag{4.17}$$

When the number of codes that can be used to carry the information is fixed, using the microcanonical ensemble to describe the information source will leave more redundancy for the information transmission. Furthermore, this redundancy can be quantified by the covariance of the matrix in the conjugate canonical ensemble.

4.5 Conclusions

In this chapter, we show that the information-theoretical bounds of the information sources with numerous heterogeneous interacting units are still decided by the uncertainty of the information sources, which are quantified by the Shannon entropy. As the random variable with finite outcomes is not enough to describe the new information sources, so we have introduced the statistical ensembles with local constraints to model those heterogeneous interactions. Thus, the new information-theoretical bounds are decided by the entropy of those statistical ensembles. As these new information sources have a huge number of units, the entropy of the statistical ensemble

described information sources would be affected by the possible appearance of ensemble nonequivalence. Under this case, different ensemble descriptions have different Shannon entropy. Using the microcanonical ensemble to describe the information source with interacting units needs a smaller information storage space, but the probability distribution of the microcanonical ensemble is hard to calculate. Using the canonical ensemble to describe the new information sources is easy to get the probability distribution but requires more information storage space. The extra sequences that need to store in the canonical ensemble descriptions are those sequences with the same sum of *Hamiltonian* as the sequences in the typical set of the microcanonical ensemble, but the constraints of each state are not equal to the hard constraints of the microcanonical ensemble. The size of the extra typical sets can be approached by the covariance matrix of constraints in the canonical ensemble. It means the difference between the information-theoretical bounds of different ensemble descriptions of the ensemble nonequivalent information sources is affected by the fluctuation in the local constraints. It has reinforced the conclusions in the traditional information theory that the uncertainty of the information sources will affect the information-theoretical bounds.

Appendix 4.A ER model described information sources

ER model $G(n,p)$ represents the probability of the system with n units, and each pair of units have a probability of p to connect with each other. The constraint in the ER model is the total number of interactions among the units in the information source. It still can be described by the microcanonical and the canonical ensemble when constraints in it have different properties.

As the number of the possible links in the ER model is equal to $n(n-1)/2$, the hard constraint C^* in the microcanonical ensemble is equal to the expectation value of the total number of links as $p \times n(n-1)/2$. Thus, the probability of each state in the microcanonical ensemble with these hard constraints is

$$P_{\rm mic}(\mathbf{G}^*) = 1/\binom{n(n-1)/2}{C^*}. \tag{4.18}$$

Each state has the same probability to appears in the process of signal generation. The value of the probability is decided by the total number of configurations with the same constraints.

When the ER model is under 'soft' constraints, each state of the information source does not need to have the same value of total interactions; only the average value of total interactions is equal to the C^*. Then the value of p is equal to $\frac{2C^*}{n(n-1)}$, probability of the state of the information source described by the canonical ensemble is

$$P_{\rm can}(\mathbf{G}) = p^{C(\mathbf{G})}(1-p)^{\frac{1}{2}n(n-1)-C(\mathbf{G})}. \tag{4.19}$$

The canonical entropy is equal to $-\ln P_{\rm can}(\mathbf{G}^*)$. When the value of soft constraints

is equal to C^*, the canonical entropy is

$$\begin{aligned} S_{\rm can} =& - C^* \ln 2C^* + \frac{n(n-1)}{2} \ln[n(n-1)] \\ & - (\frac{n(n-1)}{2} - C^*) \ln(n(n-1) - 2C^*). \end{aligned} \tag{4.20}$$

If the information source described by the microcanonical ER model ensembles is used to generate the information sequences with length l, the space to store the information that is generated by the microcanonical description is equal to

$$\ln |T^{\epsilon=0}_{\rm mic}| = l \times \ln \binom{n(n-1)/2}{C^*}, \tag{4.21}$$

where $T^{\epsilon=0}_{\rm mic}$ is the typical set of the microcanonical ensemble information sequence.

When information sources are described by the canonical ER model, the space to store the information that is generated by it is equal to

$$\ln |T^{\epsilon=0}_{\rm can}| = l \times S_{\rm can} = -l \times \ln P_{\rm can}(\mathbf{G}^*). \tag{4.22}$$

It is decided by the canonical entropy of the ER model.

The difference between the limit of information storage of the information source described by different ensembles is related to the relative entropy between the microcanonical and canonical ER model, which is equal to $S_{\rm can} - S_{\rm mic}$ when the two ensembles are conjugate with each other. We can get the result of relative entropy based on Stirling's formula as

$$S(P_{\rm mic}||P_{\rm can}) \approx \ln \sqrt{2\pi C^*(1 - 2C^*/[n(n-1)])}. \tag{4.23}$$

In the thermodynamic limit, the two ER model ensembles are equivalent to each other, as the limit value of the relative entropy density is equal to 0

$$\lim_{n\to\infty} \frac{1}{n} S(P_{\rm mic}||P_{\rm can}) = 0. \tag{4.24}$$

The space can be saved from canonical ensemble description to microcanonical ensemble description is equal to $l \times S(P_{\rm mic}||P_{\rm can})$, but compared with the total space needs in the canonical ensemble description, it is not so important, as the space ratio r will equal to 0 in the thermodynamic limit

$$r = \frac{S(P_{\rm mic}||P_{\rm can})}{S_{\rm can}} = 0. \tag{4.25}$$

Therefore, using the microcanonical ensemble only saves finite space of the information storage. The ensemble equivalence allows us to choose the canonical ensemble, which is mathematically easy to obtain.

Appendix 4.B Matrix described information sources

Matrix is a general model widely used to describe natural systems with heterogeneous interactions. [27]. The heterogeneous interactions among the units imply the local constraints in it. It means this information sources described by the matrix ensembles are under ensemble nonequivalence.

Thus, checking if the limit of information storage of the matrix ensemble described information source will be affected by the ensemble nonequivalence as we predicted is significant for our theory. We will start with matrice under global constraint, then extend the discussion to the local constrained one.

4.B.1 Matrix with global constraint

Global constraint is the sum of all the elements in each matrix is fixed as $C(\mathbf{G}) = \sum_{i=1,j=1}^{n,m} g_{ij}$. When the global constraint is hard, the constraint of each state in the matrix ensemble is equal to each other as C^*. The system can be described by the microcanonical ensemble. If the global constraint is soft, the average value of each matrix's constraints is equal to the hard constraints, $\langle C(\mathbf{G})\rangle = C^*$, the system can be described by the canonical ensemble. States both in the microcanonical and canonical ensemble with constraints equal to C^* is represented by $\mathbf{G}^*$.

The probability of each state in the matrix described by the microcanonical ensemble is equal to

$$P_{\text{mic}}(\mathbf{G}^*) = \frac{1}{\Omega_{C^*}}, \tag{4.26}$$

where Ω_{C^*} is the total number of states with global constraint equals to C^*.

In the matrix described by the canonical ensemble, the *Hamiltonian* of each matrix

$$H = \theta^* \cdot C(\mathbf{G}) \tag{4.27}$$

decides the probability of it. The θ^* is the maximum likelihood parameter realized $\langle C(\mathbf{G})\rangle = C^*$ and maximum the Shannon entropy. Thus, the probability of each state in the canonical ensemble is equal to

$$P_{\text{can}}(\mathbf{G}) = \frac{e^{-H(\mathbf{G},\theta^*)}}{Z(\theta^*)}, \tag{4.28}$$

$Z(\theta^*)$ is the partition function, and it is a normalization constant equal to $Z(\theta^*) = \sum_{\mathbf{G}\in\mathcal{G}} e^{-H}$.

If the matrix with global constraints is used as an information source, then the information generated from it is carried by a set of matrix ensemble sequences with length l.

When the global constrained matrix is described by the microcanonical ensemble, the space to store the information generated by it is equal to

$$\ln |T^{\epsilon}_{\text{mic}}| = l \times S_{\text{mic}} = l \times \ln \Omega_{C^*}. \tag{4.29}$$

For the canonical ensemble described matrix information sources, the space to store the information generated by it is equal to

$$\ln |T^{\epsilon}_{\text{can}}| = l \times S_{\text{can}} = -l \times \ln P_{\text{can}}(\mathbf{G}^*). \tag{4.30}$$

Relative entropy between the two ensembles' probability distribution is equal to $S_{\text{can}} - S_{\text{mic}}$. It can be estimated by the determinant of the covariance matrix of constraints in the canonical ensemble as

$$S(P_{\text{mic}}||P_{\text{can}}) \approx \ln \sqrt{2\pi \mathbf{\Sigma}^*} = \frac{1}{2} \ln 2\pi \frac{\partial^2 \ln Z(\theta^*)}{\partial {\theta^*}^2}. \tag{4.31}$$

When the elements in the matrix is chosen from different set, partition function $Z(\theta^*)$ is also different. For example when the element in the matrix is equal to 1 or 0, this matrix is binary matrix, the partition function of it is equal to $Z(\theta^*) = (e^{-\theta^*} + 1)^{mn}$. If the element in the matrix is chosen from the whole natural number set, the matrix is a weighted matrix, the partition function of it is equal to $Z(\theta^*) = (1 - e^{-\theta^*})^{-mn}$. According to the relationship between θ^* and C^* in the two different matrices, we can find the value of relative entropy of binary matrix is equal to $S(P_{\text{mic}}||P_{\text{can}}) = \frac{1}{2} \ln[2\pi C^*(1 - C^*/(mn))]$, and the relative entropy of weighted matrix is equal to $S(P_{\text{mic}}||P_{\text{can}}) = \frac{1}{2} \ln[2\pi C^*(1 + C^*/(mn))]$.

Because the value of $S(P_{\text{mic}}||P_{\text{can}})$ for the two different kinds of matrices are both grows like $o(n)$, the two different ensemble descriptions are equivalent to each other in the thermodynamic limit.

The ratio of space that can be saved from canonical ensemble description to microcanonical ensemble description is

$$r = [\frac{1}{2} \ln 2\pi \frac{\partial^2 \ln Z(\theta^*)}{\partial {\theta^*}^2}] / [-\theta^* \cdot C^* - \ln Z(\theta^*)]. \tag{4.32}$$

The value r of the binary matrix under global constraint is equal to

$$r = \frac{1}{2} \frac{\ln[2\pi C^*(1 - C^*/(mn))]}{mn \ln(mn) - C^* \ln C^* - (mn - C^*) \ln(mn - C^*)}. \tag{4.33}$$

The value of r for the weighted matrix ensemble under global constraints is equal to

$$r = \frac{1}{2} \frac{\ln[2\pi C^*(1 + C^*/(mn))]}{(mn + C^*) \ln(mn + C^*) - mn \ln(mn) - C^* \ln C^*}, \tag{4.34}$$

When n goes to infinite, the r is equal to 0 both in the two matrices. Therefore, when the system with global constraint is used as the information source, it is under ensemble equivalence. The space saved from canonical to microcanonical ensemble description can be neglected.

4.B.2 Matrix with local constraints

The local constraints of the information sources are implied by the heterogeneous interactions among the units in it. The local constraint is the sum of all the elements in each row or column in the matrix ensemble. According to the research in [27], matrix ensemble with local constraints is under ensemble nonequivalence. When the value of rows in the matrix under local constraints is finite $m \ll \infty$, the ensemble nonequivalence is as strong as the one in the boundary of phase transition [8].

In this section, we will introduce how the heterogeneous interaction will affect the limit of information storage. As the coupled constraints only can be analytically solved in two particular cases, we will put all the calculations on the matrix with one-sided local constraints [27].

The $m \times n$ matrix ensemble under local column constraints $\vec{C}^* = [c_1^*, c_2^*, \cdots, c_i^*, \cdots, c_n^*]$ has n constraints in it. Each c_i^* is the sum of all the elements in the column i as $c_i^* = \sum_{j=1}^{m} g_{ij}$. The property of constraints decides which ensembles will be used to describe this local constrained matrix.

In the microcanonical ensemble description, each state still have the same value of constraints as $\vec{C}^*$, and the probability of it is equal to

$$P_{\text{mic}}(\mathbf{G}^*) = 1/\Omega_{\vec{C}^*}, \tag{4.35}$$

where $\Omega_{\vec{C}^*}$ is the number of states in the matrix described by the microcanonical ensemble. In binary matrix, $\Omega_{\vec{C}^*} = \prod_{i=1}^{n} \binom{m}{r_i^*}$. In weighted matrix, $\Omega_{\vec{C}^*} = \prod_{i=1}^{n} \binom{m+r_i^*-1}{r_i^*}$. The space to store the information generated by it is equal to

$$\ln |T_{\text{mic}}^{\epsilon=0}| = l \times S_{\text{mic}} = l \times \ln \Omega_{\vec{C}^*}, \tag{4.36}$$

both in the binary and weighted matrix.

When the local column constraints are soft, the matrix needs to be described by the canonical ensemble. The probability of states in the canonical ensemble is also based on the *Hamiltonian* of it, which is defined as $H = \sum_{i=1}^{n} \beta_i^* c_i^*$. Where β_i^* is the correspond parameter which maximum the Shannon entropy and realized the $\langle \vec{C}(\mathbf{G}) \rangle = \vec{C}^*$. Therefore, the probability of states in the canonical matrix ensemble is

$$P_{\text{can}}(\mathbf{G}) = \frac{e^{-H}}{Z(\vec{\beta}^*)}. \tag{4.37}$$

The information generated by it is also carried in the canonical matrix ensemble sequences. The space to store the information is equal to

$$\ln |T_{\text{can}}^{\epsilon=0}| = l \times S_{\text{can}} = -l \times \ln P_{\text{can}}(\mathbf{G}^*). \tag{4.38}$$

The space saved from the canonical description to the microcanonical ensemble description can be estimated by the function of the determinant of the covariance

matrix of constraints in the canonical ensemble as

$$\begin{aligned} S(P_{\text{mic}}||P_{\text{can}}) &= \frac{1}{2}\sum_{k=1}^{n}\ln\frac{2\pi\lambda_k^*}{[1+O(1/\lambda_k^*)]^2} \\ &\approx \frac{1}{2}\sum_{k=1}^{n}\ln[2\pi\frac{\partial^2\ln Z(\vec{\beta}^*)}{\partial\beta_k^{*2}}]. \end{aligned} \tag{4.39}$$

As different matrices have different difference of partition function, so the relative entropy is also different. Partition function of binary matrix under is equal to $Z(\vec{\beta}^*) = \prod_{i=1}^{n}(e^{-\beta_i^*}+1)^m$. In weighted matrix, partition function is equal to $Z(\vec{\beta}^*) = \prod_{i=1}^{n}(1-e^{-\beta^*})^{-m}$. Then the value of relative entropy for binary matrix is equal to $S(P_{\text{mic}}||P_{\text{can}}) = \frac{1}{2}\sum_{i=1}^{n}\ln[2\pi\frac{r_i^*(m-r_i^*)}{m}]$. In weighted matrix, the value of relative entropy is equal to $S(P_{\text{mic}}||P_{\text{can}}) = \frac{1}{2}\sum_{i=1}^{n}\ln[2\pi\frac{r_i^*(m+r_i^*)}{m}]$. Both of those two matrix are under ensemble nonequivalence. The space can be saved from the canonical ensemble description to the microcanonical ensemble description has the same order as the increase of the ensemble sequences' length. The ratio r is still defined as

$$r = [\frac{1}{2}\sum_{k=1}^{n}\ln[2\pi\frac{\partial^2\ln Z(\vec{\beta}^*)}{\partial\beta_k^{*2}}]]/[-\vec{\beta}^*\cdot\vec{C}^* - \ln Z(\vec{\beta}^*)]. \tag{4.40}$$

In the binary matrix ensemble under local column constraints, the value of r is equal to

$$r = 1 - [\sum_{i=1}^{n}\ln\binom{m}{r_i^*}]/[\sum_{i=1}^{n}\ln[m^m/r_i^{*r_i^*}(m-r_i^*)^{m-r_i^*}]]. \tag{4.41}$$

In the weighted matrix ensemble, the value of r is equal to

$$r = 1 - [\sum_{i=1}^{n}\ln\binom{m+r_i^*-1}{r_i^*}]/[\sum_{i=1}^{n}\ln[\frac{(m-r_i^*)^{m-r_i^*}}{m^m r_i^{*r_i^*}}]. \tag{4.42}$$

When the matrix is in the thermodynamic limit, the limit value of r grows like $\frac{\ln(2\pi m)}{m}$. Thus, when the freedom of each element is finite $m \ll \infty$, the ratio is fixed. When the value of m is growing like $O(n)$, the ratio is close to 0. It means compare with canonical ensemble description, using the microcanonical ensemble will save $r\%$ of the space.

Under two-sided local constraints, it is impossible to calculate the number of states in the microcanonical ensemble. The increased number of constraints will decrease the possible configurations in the microcanonical ensemble, so the space to store the information generated by it is even smaller than the one with local column constraints [27]. The ensemble nonequivalence will affect the information-theoretical bounds.

Chapter 5

Information theory with coupled sources under ensemble nonequivalence

Abstract

Information theory is build to describe information transmission and storage in different systems. Restricted by the initial setting of the electronic communication system, the traditional information theory is based on a fundamental assumption that the signal generation by those information sources has an identical probability distribution and is independent of time. However, recent research on nervous systems and social networks shows that the information flows in those systems are generated by the numerous interacting units, and the signal generation in those systems is under temporal dependencies. It means the classical information theory based on the i.i.d assumption cannot deal with the coupled sources with temporal and spatial dependencies in non-artificial communication systems.

Motivated by the recent works on systems with local constraints in statistical physics, a generalization of information theory with coupled sources is built to find the limits of information transmission and storage based on the descriptions of information sequences with statistical ensembles under local constraints in this work. We find that the microcanonical ensemble description or the *Boltzmann entropy* is closer to the real limit of information storage than the canonical ensemble description with soft constraints or *Shannon entropy*. We also find that the classical information theory is a particular case of the canonical ensemble description when the dependencies are homogeneous. Moreover, the effectivity of classical information theory only holds when the microcanonical ensemble description and the canonical ensemble description of the signal generation are under the ensemble equivalence. Our result also shows that the finite temporal dependences of units in the information source are not enough to break the ensemble equivalence. The ensemble nonequivalence is formed by the extensive spatial interactions among all the units [1].

[1] This chapter is based on the coming paper:
Qi Zhang, Diego Garlaschelli, "Information theory with coupled source under ensemble nonequivalence" (2021)

5.1 Introduction

The classical information theory established in 1948 by Shannon is build to estimate the information-theoretical bounds of the information storage and the information transmission in communication systems [33]. This theory is based on an essential assumption that the information source works independently with an identical probability distribution (i.i.d) in the process of signal generation. According to this assumption, Shannon found that almost all the information generated by the i.i.d. information source $\mathbf{x}$ is carried by a set of equiprobable sequences $\{x_1, x_2, \cdots, x_n\}$, which is named as typical set $T_\epsilon^{(n)}$. Therefore, the smallest space needs to store the information generated by the information source is equal to the logarithm of the cardinality of the typical set $\ln |T_\epsilon^{(n)}|$. This limit of information storage will converge to the $n \times s(\mathbf{x})$ as a function of Shannon entropy $s(\mathbf{x})$ of the information sources when $(n \to \infty)$ the length of sequences goes to infinite [29]. This typicality is the asymptotic equipartition property (AEP). It also can be found in statistical physics, which shows that the equilibrium behaviour of a system in the thermodynamic limit is determined by typical microstates [8].

The signal generation under the i.i.d. assumption as a stationary process ignores the possible temporal and spatial dependencies between the units in information sources. However, in natural systems, especially when there are multivariates in the information source, these dependencies generally exist. For example, in the vertebrate retina, the activity of neurons is determined by pairwise correlations among neurons, and the limited energy can be used for each neuron simultaneously [69]. The pairwise correlations are spatial interactions among all the neurons. The finite energy that can be used by each neuron in the whole process of signal generation is the temporal constraint. The spatial and temporal dependencies also exist in the changes of cars' flow in the urban traffic networks [70] and fluctuations of the stock market [3]. These heterogeneous dependencies among the units in the information source make it impossible to find the limit of information storage by the classical AEP [71, 30]. Also, it may affect the symbol rate used to reliable transport the information through different channels. Thus, we need a new theory to describe the signal generation with spatial and temporal dependencies and find the information-theoretical bounds.

According to the classical information theory, the information generated by the source is carried by information sequences, which are used to record the behaviour of units in information sources. So even when the signal generation of the multivariate source is under spatial and temporal dependencies, the information generated by it is still carried by the multivariate information sequences $\{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n\}$. Thus, to find the information-theoretical bounds of those non-stationary signal generations, we should focus on the information sequences, not the information sources.

However, those sequences with heterogeneous dependencies and the increased length are impossible to be described by random variables with finite outcomes, but those macroscopic properties are analogue with the that in states of thermodynamic systems. Both of them have numerous interacting units, and the numbers of units

go to infinite in the thermodynamic limit. It means statistical ensembles in physics can be used to describe the multivariate and heterogeneous dependent information sequences, to find its information-theoretical bounds [1, 27].

In statistical physics, systems with different macroscopic properties need to be described by different ensembles. The microcanonical ensemble is used to describe systems with fixed total energy E^*. The canonical ensemble is used to describe systems with fixed temperature $\beta = 1/kT$ [1, 49, 27]. The two ensembles will conjugate with each other by setting the parameter β in the canonical ensemble equal to β^*, which makes the average value of the total energy in the canonical ensemble equal to the fixed total energy in the microcanonical ensemble ($\langle E \rangle = E^*$) [15].

Normally, in the thermodynamic limit, the two conjugate ensemble descriptions are believed to be equivalent. The microcanonical ensemble can be replaced by the canonical ensemble, which is mathematically easy to calculate. This one phenomenon is called ensemble equivalence (EE) [8]. However, in the past decades, the breakdown of EE also has been observed in various physical systems [39, 25, 21]. Especially when there are extensive local constraints in the system, the EE breaks in the whole parameter space of this system [5, 27]. Therefore, when statistical ensembles are used to describe the information sequences with different macroscopic properties, their information-theoretical bounds are affected by the possible appearance of the ensemble nonequivalence (EN).

In this work, a matrix ensemble with local constraints is used to describe the information sequences that are generated in the signal generation with heterogeneous dependencies. These heterogeneous dependencies are quantified by the total correlation (multi-information) among units in sequences [72]. We find that the classical information theory is a particular case of the canonical ensemble description with soft constraints. We also prove that the effectivity of the classical AEP in information theory is based on the EE of the signal generation. Most importantly, we find that the EN in the non-stationary process is led by the variable spatial interactions among units in the source, not the finite temporal dependence.

5.2 Ensemble described information sequences

In traditional statistical physics, statistical ensembles are under global constraints such as the fixed total energy and fixed temperature. The interactions among all units are homogeneous. However, this assumption breaks when the statistical ensembles are used to describe the information sequences generated by the information source with heterogeneous dependent units since the interactions among those units are homogeneous and time variant. It means the description of the information sequences needs the statistical ensemble with local constraints [27].

The information sequences are generated by the information source with heterogeneous spatial interactions and temporal dependencies. If the information source has m finite units, then the information generated by it in n times sampling should be recorded by the $m \times n$ matrix $\mathbf{X}$. Each unit x_{ji} represents the state of unit j

in information source at time i, and matrix $\mathbf{X}$ represents a particular state of the information sequences.

According to the definition of matrix should above, the spatial dependence among m units in the information source at different time can be modelled by the *column local constraints* $\vec{c} = [c_1, c_2, \cdots, c_i, \cdots, c_n]$, where c_i is the sum of all the units in column i of matrix $\mathbf{X}$ as $c_i = \sum_{j=1}^{m} x_{ij}$. The temporal dependence of each variable in the information sources is modelled by the *row local constraints* $\vec{r} = [r_1, r_2, \cdots, r_j, \cdots, r_m]$, where each $r_j = \sum_{i=1}^{n} x_{ji}$ is the sum of all the element in the jth row of matrix $\mathbf{X}$. In nervous systems, r_j represents the total energy theta can be used by the neuron X_j in the whole signal generation, c_i represents the total energy that can be used by m interacting units in time i.

The relationship between local constraints in the matrix ensemble and the dependencies in ensemble sequences is shown in FIG.5.1.

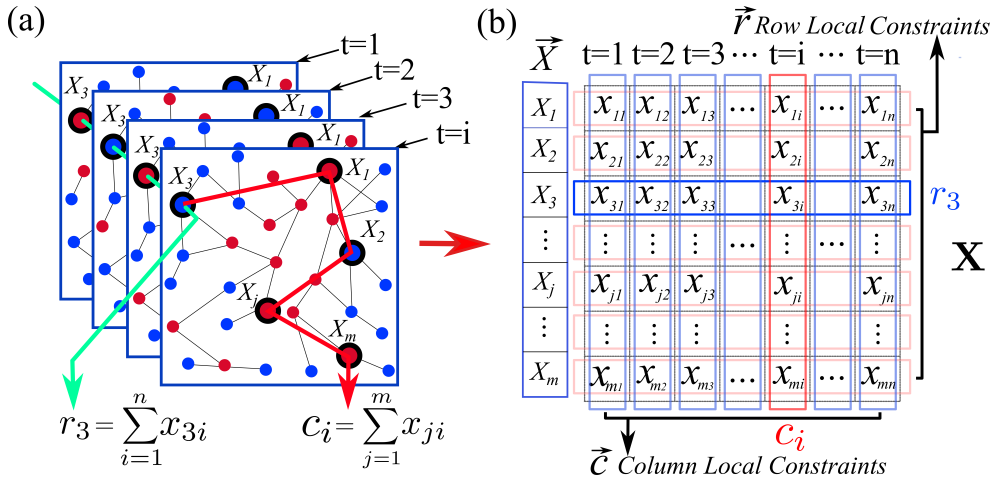

Figure 5.1. The status of m units in different times show in (a). The red and blue colour of each node represents it is active or non-active. The selected nodes (the big nodes with the black circle as margin) are used as units in the information source. The green line across different layers represents temporal dependence. The red line among nodes in each layer represents the spatial interaction among them. The localization of the two dependencies in the matrix shows in (b). The spatial interactions among all the units in the information source at time i are quantified by the column local constraints c_i. The temporal dependence of unit j is represented by the row local constraints r_j.

When the dependence is homogeneous, the information sequences should be modelled by the matrix ensemble with global constraints as $t^* = \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ji}$, which is the sum of all elements in the sequence [27].

The statistical ensemble is a probability distribution of all possible states in a specific thermodynamic system. The macroscopic property of constraints determines the value of probability for each state. For example, when the constraints are hard,

states of it are equiprobable. When the constraints are soft, the probability of each state is different. The two different kinds of constraints represent two different ways to describe the signal generation under dependencies. When the constraints are hard, we need the microcanonical ensemble. Otherwise, when the constraints are soft, we need the canonical ensemble [15].

5.2.1 Canonical ensemble description

When heterogeneous dependencies in sequences are *soft*, the constraints of each microscopic configuration are different. The average value of the constraints for all the sequences is equal to the hard constraints in the conjugate microcanonical ensemble. Then the sequences will be described by the canonical ensemble.

The probability of each state in the canonical ensemble is a parameter solution under the realization of the average constraints and the maximization of Shannon entropy

$$P_{\text{can}}(\mathbf{X}|\vec{\beta}) = e^{-H(\mathbf{X},\vec{\beta})}/Z(\vec{\beta}). \tag{5.1}$$

Vector $\vec{\beta}$ is an extension of the maximum likelihood parameter β base on extensive local constraints in it [27].

The partition function $Z(\vec{\beta})$ is a normalization constant, which collect the exponential function of all the possible configurations of the sequences $\mathbf{X}$ as $Z(\vec{\beta}) = \sum_{\mathbf{X}\in\mathcal{X}} e^{-H(\mathbf{X},\vec{\beta})}$. The symbol $H(\mathbf{X},\vec{\beta})$ represents the *Hamiltonian* of a sequence $\mathbf{X}$. It is a liner combination of constraint and maximum likelihood parameter, $H(\mathbf{X},\vec{\beta}) = \vec{C}(\mathbf{X})\cdot\vec{\beta}$ [15].

The microcanonical ensemble and canonical ensemble are conjugate with each other by setting the parameter $\vec{\beta} = \vec{\beta}^*$, to make the average value of constraints $\langle\vec{C}(\mathbf{X})\rangle$ in the canonical ensemble is equal to hard constraints $\vec{C}^*$ in the microcanonical ensemble as

$$\langle\vec{C}(\mathbf{X})\rangle = \sum_{\mathbf{X}\in\mathcal{X}} P_{\text{can}}(\mathbf{X}|\vec{\beta}^*)\vec{C}(\mathbf{X}) = \vec{C}^*, \tag{5.2}$$

where $\mathcal{X}$ represents the collection of all the possible information sequences.

As constraints $\vec{C}(\mathbf{X})$ is the sum of elements in each column and row, the probability of a sequence in the conjugate canonical ensemble is equal to the product of the probability of each unit in each sequence as

$$P_{\text{can}}(\mathbf{X}|\vec{\beta}^*) = \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{e^{-x_{ij}\beta^*_{ij}}}{\sum_{x_{ij}\in\S_{ij}} e^{-x_{ij}\beta^*_{ij}}}. \tag{5.3}$$

The value of $P_{\text{can}}(\mathbf{X}|\vec{\beta}^*)$ is governed by the parameter $\vec{\beta}^*$ and the value of each units x_{ij} in sequence $\mathbf{X}$. The symbol $\S_{ij}$ represents the collection of all the possible values of x_{ij}.

Then, according to the classical information theory, it is easy to find that the smallest space to store the information carried by canonical ensemble sequences is equal to the Shannon entropy of it as

$$S_{\text{can}} = \sum_{\mathbf{X}\in\mathcal{X}} P_{\text{can}}(\mathbf{X}|\vec{\beta}^*) \ln P_{\text{can}}(\mathbf{X}|\vec{\beta}^*). \tag{5.4}$$

This result coincides with the consequence in Shannon's information theory that the smallest space needs to store the information is determined by its uncertainty. However, the quantification of the uncertainty is affected by the appearance of heterogeneous interactions.

The Eq.5.3 shows that each unit in the sequence $\mathbf{X}$ under soft constraints is independent. Probability of each unit to gets value x_{ij} is governed by the localized parameter β^*_{ij}. Thus, the canonical ensemble is equal to the production of the marginal probability of each unit as

$$P_{\text{can}}(\mathbf{X}|\vec{\beta}^*) = \prod_{i=1}^{n}\prod_{j=1}^{m} P(x_{ji}). \tag{5.5}$$

Comparing with the microcanonical ensemble that realization all the constraints exactly, the canonical one is more like the localization of the dependencies on each unit in the information sequence.

When the two local constraints are working simultaneously, sequences are under coupled local constraints. The *Hamiltonian* is equal to $H = \sum_{i=1}^{n} \alpha_i^* c_i + \sum_{j=1^m} \beta_j^* r_j$. Then, we can get the probability distribution and the Shannon entropy of this canonical ensemble description. Details of the calculation are shown in the Appendix 5.C.

When only column-local constraints work on the signal generation or the signal generation is only constrained by soft spatial interactions among all the units in information sources, the canonical ensemble description of this signal generation still can be described by the coupled local constraints, but with the row local constraints equal to each other as $r_j^* = r^*$ and $\beta_j^* = \beta^*$. Then the *Hamiltonian* should equal to $H = \sum_{j=1}^{m}\sum_{i=1}^{n}(\beta^* + \alpha_i^*)x_{ji}$. The partition function and the Shannon entropy of this canonical ensemble can be found in the Appendix 5.B. It also can be described by the one-sided local constraints matrix in [27].

When the signal generation is only limited by the soft temporal dependence or when the units in the multivariate information source are independent, the signal generation also can be described by the coupled local constraints canonical ensembles with the unit in column local constraints equal to each other as $c_i^* = c^*$. The corresponding maximum likelihood parameter is also equal to each other as $\alpha_i^* = \alpha^*$ The *Hamiltonian* of this canonical ensemble description is $H = \sum_{j=1}^{m}\sum_{i=1}^{n}(\beta_j^* + \alpha^*)x_{ji}$. The Shannon entropy can be found in the Appendix 5.A.

5.2.2 Microcanonical ensemble description

In the microcanonical ensemble description, constraints of each state have the same value $\vec{C}^*$. The column and row local constraints are fixed exactly in each matrix.

Thus, in the signal generation, the possible state of each unit in the information source will be limited by the spatial dependencies and temporal interactions exactly. For example, the activity of neurons in the nervous system is decided by the energy that the neuron can use in signal generation. The fixed column local constraints $\vec{c}^*$ means the total energy that can be used by all neurons each time is finite. The fixed row local constraints $\vec{r}^*$ mean the total energy that can be used by each neuron in the whole process of signal generation is finite. Then the more energy cost by other units in the information source, the less energy will be left for the specific one to have different states. For each unit, the more energy cost in the past, the less will be left for the future.

The *hard* constraints of information sequences require the probability of each state with constraints $\vec{C}^*$ equal to each other as

$$P_{\text{mic}}(\mathbf{X}|\vec{C}^*) = 1/\Omega_{\vec{C}^*}, \tag{5.6}$$

where $\Omega_{\vec{C}^*} = |\mathcal{X}_{\text{mic}}|$ represents the total number of sequences with constraint $\vec{C}^*$ in the microcanonical ensemble.

According to the AEP, we can find that all sequences in the microcanonical ensemble belong to the typical set T_{mic} of it. Space needs to store the information carried by the typical set is equal to the *Boltzmann entropy* of information sequence $\mathbf{X}$ [10]

$$S_{\text{mic}} = \ln |T_{\text{mic}}| = \ln \Omega_{\text{mic}}. \tag{5.7}$$

The value of it is determined by the total number of configurations of information sequence with hard constraints $\vec{C}^*$.

The $\Omega_{\vec{C}^*}$ is hard to calculate analytical, especially when information sequences are under coupled constraints. However, according to the mechanism in [37], we can estimate the value of it by the covariance matrix of constraints in the canonical ensemble.

In the two conjugate ensembles, the total number of states in the microcanonical ensemble is equal to the number of states in the canonical ensemble with constraints equal to $\vec{C}^*$ [37]. Thus, the number of states in the microcanonical ensemble can be calculated from the canonical probability distribution with Dirac delta function (δ function) as

$$\begin{aligned} \Omega_{\vec{C}^*} &= \sum_{\mathbf{X}\in\mathcal{X}} \int_{-\vec{\pi}}^{\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} e^{i\vec{\psi}[\vec{C}^* - \vec{C}(\mathbf{X})]} \\ &= \int_{-\vec{\pi}}^{\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} P_{\text{can}}^{-1}(\mathbf{X}^*|\vec{\theta}^* + i\vec{\psi}). \end{aligned} \tag{5.8}$$

The integral is difficult to calculate when it is under coupled constraints [37]. But it is still possible to use saddle point technology to approach the value of $\Omega_{\vec{C}^*}$ as

$$\Omega_{\vec{C}^*} = \frac{e^{S_{\text{can}}}}{\sqrt{\det(2\pi\mathbf{\Sigma}^*)}} \prod_{k=1}^{K} [1 + O(1/\lambda_k^*)], \tag{5.9}$$

where $\mathbf{\Sigma}^*$ is the covariance matrix of constraints in the canonical ensemble whose entries are defined as

$$\begin{aligned}\Sigma^*_{ij} &\equiv \frac{\partial^2 \ln Z(\vec{\theta})}{\partial\theta_i \partial\theta_j} \mid_{\vec{\theta}=\vec{\theta}^*} \\ &= \mathrm{Cov}[C_i, C_j]_{\vec{\theta}^*} \\ &= \langle C_i C_j\rangle_{\vec{\theta}^*} - \langle C_i\rangle_{\vec{\theta}^*}\langle C_j\rangle_{\vec{\theta}^*},\end{aligned} \tag{5.10}$$

The $\{\lambda^*_k\}$ is the eigenvalue of covariance matrix $\mathbf{\Sigma}^*$ [37]. K is the number of constraints in the matrix $\mathbf{X}$

Then we can have the *Boltzmann* entropy of the microcanonical ensemble $S_{\mathrm{mic}} = \ln \Omega_{\vec{C}^*}$ is equal to the Shannon entropy of the conjugate canonical ensemble minus the correction part based on the covariance matrix of constraints in the canonical ensemble as

$$S_{\mathrm{mic}} = S_{\mathrm{can}} - \ln\sqrt{\det(2\pi\mathbf{\Sigma}^*)} + \sum_{k=1}^{K} \ln[1 + O(1/\lambda^*_k)]. \tag{5.11}$$

The correction part $\sum_{k=1}^{K} \ln[1 + O(1/\lambda^*_k)$ is negligible when the eigenvalue value of the covariance matrix λ^*_k is big enough. Thus, the space to store the information carried by the microcanonical ensemble sequences is smaller than the canonical one. Because the hard constraint in the microcanonical ensemble strictly modelled the influence of heterogeneous dependencies in information sequences. Therefore, the microcanonical ensemble description is closer to the natural process of signal generation under heterogeneous dependence than the canonical ensemble one, which is the maximum entropy approximation.

5.3 Total correlations of information sequences

The nonnegligible difference between the microcanonical and canonical ensemble description of the information sequences shows that the two different descriptions of the heterogeneous dependencies in the information sequences contains different information about the signal generation with temporal and spatial dependencies. However, we do not know, if this difference is related with the correaltions among units in the information sequences. Therefore, it is important to check if the nonnegligible ensemble difference is a manifestation of the correaltions of the units in the information sequences.

The matrix ensemble $\mathbf{X}$ gives a possible model for us to quantify the dependence among all units in information sequences as the total correlations $\mathbb{C}$ [72]

$$\mathbb{C} = \sum_{\mathbf{X}\in\mathcal{X}} P(\mathbf{X}) \ln \frac{P(\mathbf{X})}{\prod_{i=1}^{n}\prod_{j=1}^{m} P(x_{ji})}. \tag{5.12}$$

The symbol $\mathcal{X}$ represents the collection of all possible configurations of sequences $\mathbf{X}$ under heterogeneous dependencies.

However, with heterogeneous dependencies, both the actual probability $P(\mathbf{X})$ of the sequence $\mathbf{X}$ and the production of the marginal probability of each unit $\prod_{i=1}^{n}\prod_{j=1}^{m}P(x_{ij})$ are difficult to calculate. Therefore, the two different ensemble descriptions show above that are based on the maximum entropy principle proposed by Jaynes [15] give a way to approach the probability $P(\mathbf{X})$ and the production of marginal probability $\prod_{i=1}^{n}\prod_{j=1}^{m}P(x_{ij})$ from the biased information we know.

The probability of the matrix $\mathbf{X}$ with constraints $\vec{C}(\mathbf{X})$ to appear in the signal generation is decided by the number of configurations $\Omega_{\vec{C}(\mathbf{X})}$ in it. The microcanonical ensemble description strictly satisfies the requirement of signal generation. Thus, when the constraints are fixed, the probability $P(\mathbf{X})$ in the definition of total correlations can be replaced by the probability of states in the microcanonical ensemble description as

$$P(\mathbf{X}) = P_{\text{mic}}(\mathbf{X}|\vec{C}^{*}). \tag{5.13}$$

The production of each unit's marginal probability in Eq.5.12 is based on the assumption that the probability of each unit $P(x_{ji})$ can be calculated independently. However, when there is heterogeneous dependence, we can only use the canonical ensemble to approach the production of marginal probabilities, as the canonical ensemble has localized the dependencies of all the units in the sequence by the parameter β_{ji}. Thus, we can have the production of the marginal probability of the matrix $\mathbf{X}$ as

$$\prod_{i=1}^{n}\prod_{j=1}^{m}P(x_{ji}) = P_{\text{can}}(\mathbf{X}|\vec{\beta}^{*}). \tag{5.14}$$

It is determined by the parameter $\vec{\beta}^{*}$ and the corresponding constraints $\vec{C}(\mathbf{X})$ simultaneously.

Therefore, the total correlation in sequences with constraints $\vec{C}^{*}$ (the hard constraints in microcanonical ensemble, and the average value of constraints in canonical ensemble) can be approached by the relative entropy $S(P_{\text{mic}}||P_{\text{can}})$ between the two ensembles as

$$\mathbb{C} = S(P_{\text{mic}}||P_{\text{can}}) = \sum_{\mathbf{X}\in\mathcal{X}} P_{\text{mic}}(\mathbf{X}|\vec{C}^{*}) \ln \frac{P_{\text{mic}}(\mathbf{X}|\vec{C}^{*})}{P_{\text{can}}(\mathbf{X}|\vec{\beta}^{*})} \tag{5.15}$$

The probability of states in microcanonical ensemble with constraints not equal to $\vec{C}^{*}$ is 0, so the total correlation also can be calculated as

$$\mathbb{C} = \ln P_{\text{mic}}(\mathbf{X}^{*}) - \ln P_{\text{can}}(\mathbf{X}^{*}), \tag{5.16}$$

where $\mathbf{X}^{*}$ represents sequence with constraints $\vec{C}^{*}$ [27].

The Shanon entopy of the canonical ensemble is equal to $S_{\text{can}} = \ln P_{\text{can}}(\mathbf{X}^{*})$. The Boltzmann entropy of the microcanonical ensembl equal to $S_{\text{mic}} = \ln P_{\text{mic}}(\mathbf{X}^{*})$.Thus, the total correaltions $\mathbb{C}$ is the difference between the shannon entropy and Boltzmann entropy of the sequences with heterogeneous dependencies.

As the relative entropy density is the indicator of the measure ensemble nonequivalence [8], the total correlation $\mathbb{C}$ of the matrices $\mathbf{X}$ also has a close relationship with the EN in it.

Because the microcanonical entropy can be obtained by the covariance matrix of constraints in the canonical ensemble, we can find that the total correlation also can be calcualted by the covariance matrix of constraints in the canonical ensemble as

$$\mathbb{C} = \ln \sqrt{\det(2\pi \mathbf{\Sigma}^*)} - \ln \prod_{k=1}^{K} [1 + O(1/\lambda_k^*)]. \tag{5.17}$$

This result shows that the extra information needs to describe states in the canonical ensemble is determined by the relative fluctuation between the constraints in the canonical ensemble and the conjugate microcanonical ensemble. The bigger the fluctuation, the more difference between the two ensembles. The more information we need to store under the canonical ensemble description.

5.4 Classical information theory with ensemble description

As we already mentioned before, the ensemble description is an extension of the classical information theory with heterogeneous dependencies. Thus, in this part, we will show that the classical information theory is a particular case of the ensemble description of information sequences.

First, we will introduce the typical description of signal generating in classical information theory.

In classical information theory, there is a basic assumption that each unit in the sequence is independent. Moreover, the probability distribution of each unit to get different values is the same (independent-identical-distribution). For example, in the binary information source x, the probability of the sequence $A = [a_1, a_2, \cdots, a_i, \cdots, a_n]$ with t units equal to 1 is

$$P(A|t) = p^t (1-p)^{n-t}, \tag{5.18}$$

where p is the probability of the information source x to have the value of 1, and t is the total number of units with value 1 in the sequence.

When the length n of the sequence A goes to infinite, and the probability p of each unit to get value 1 fixed, we can find the average number of units in the sequence A to has value 1 is equal to $\langle t \rangle = t^* = n \times p$. It is a manifestation of the large number law.

The information that is generated by the information source x is carried by information sequences A in the typical set T_ϵ. This typical set can be detected by the AEP as

$$\frac{1}{n} \ln P(A|t) = \frac{t}{n} \ln p + \frac{n-t}{n} \ln(1-p). \tag{5.19}$$

When the length $n \to \infty$, the value of $\frac{t}{n}$ in the rescaled proability $\frac{1}{n} \ln P(A|t)$ will equal to $\frac{t^*}{n} = p$. Thus, the limit of the rescaled probability is equal to

$$\lim_{n\to\infty} \frac{1}{n} \ln P(A|p) = p \ln p + (1-p) \ln(1-p), \tag{5.20}$$

which is the minus of the Shannon entropy of the information source x. The Shannon entropy $s(x)$ of the information source is equal to

$$s(x) = -\frac{t^*}{n} \ln \frac{t^*}{n} - \frac{n-t^*}{n} \ln \frac{n-t^*}{n}. \tag{5.21}$$

Therefore, the typical set T_ϵ is the collection of all the sequences that satisfy the following condition

$$T_\epsilon = \{A | e^{-n(s(x)-\epsilon)} \leq P(A) \leq e^{-n(s(x)-\epsilon)}\}. \tag{5.22}$$

The space to store sequences in the typical set is equal to $\ln |T_{\epsilon=0}| = n \times s(x)$. The result shows one of the main results in Shannon's information theory that the space needs to store the information generated by the information source is decided by the uncertainty of the information source, which is the Shannon entropy of the information source x [33].

Next we will prove that the classical information theory is a particular case of the canonical ensemble described information sequences with coupled constraints when the parameter β^*_{ij} is equal to each other as β^*, and $m = 1$. It is also can be described by the canonical ensemble under global constraints $t = \sum_{i=1}^{n} a_i$.

When use the canonical ensemble undee global constraints to describe the information sequence in the classical information theory, the *Hamiltonian* of classical case is equal to $H(A) = t \cdot \beta^*$. The partition function is equal to $Z(\beta^*) = (1 + e^{-\beta^*})^n$. Probability of the sequence with global constraint equal to t is

$$P_{\text{can}}(A|(\beta^*, t)) = \frac{e^{-t\beta^*}}{(1+e^{-\beta^*})^n}. \tag{5.23}$$

The soft constraints require the average value of global constraints in the canonical ensemble equals to t^* as

$$\langle t \rangle_{\text{can}} = \sum_{A \in \mathcal{X}_{\text{can}}} t(A) P_{\text{can}}(A|\beta^*, t) = t^*. \tag{5.24}$$

Thus, the probability of unit a_i to have the value of 1 in the canonical ensemble is equal to

$$p = \frac{e^{-\beta^*}}{1+e^{-\beta^*}} = \frac{t^*}{n}. \tag{5.25}$$

The Shannon entropy of the information source under canonical ensemble description is equal to

$$s_{\text{can}}(x) = -[\frac{t^*}{n} \ln \frac{t^*}{n} + \frac{n-t^*}{n} \ln \frac{n-t^*}{n}]. \tag{5.26}$$

Hence, the typical set of sequences in the classical information theory can be obtained from the AEP as

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\ln P_{\text{can}}(A|\beta^*,t) &= \frac{1}{n}\sum_{i=1}^{n}\ln p(a_i)\\ &\to E\ln p(x)\\ &= -s_{\text{can}}(x),\end{aligned} \tag{5.27}$$

The sequences in the typical set should satisfy the following condition

$$T_{\text{can}} = \{A|e^{-ns_{\text{can}}(x)-\epsilon} \le P(A) \le e^{-ns_{\text{can}}(x)+\epsilon}\}. \tag{5.28}$$

According to the relationship show in Eq.5.23, we find sequences that belong to the typical set also can be represented by the Shannon entropy of the canonical ensemble. As the sum of n rescaled entropy $s_{\text{can}}(x)$ is equal to

$$\begin{aligned}n\times s_{\text{can}}(x) &= \frac{t^{*t^*}(n-t^*)^{n-t^*}}{n^n}\\ &= \ln P_{\text{can}}(A|(\beta^*,t^*))\\ &= S_{\text{can}}\end{aligned} \tag{5.29}$$

Thus, the space to store the information carried by those sequences is equal to

$$\ln|T_\epsilon^{\text{can}}| = S_{\text{can}} = n\times s(x). \tag{5.30}$$

The result of the canonical ensemble description is equivalent to the classical description of the information sources. It proves that when the interactions in the information sequences are homogeneous, the classical description in information theory is a special case of the canonical ensemble description.

Then we need to check what happens when we use the microcanonical ensemble to describe the classical signal generating of the binary information source x. The microcanonical ensemble description is different from the canonical one. The probability of these sequences with t^* units have a value 1 can be obtained by the microcanonical ensemble description as

$$P_{\text{mic}}(A|t^*) = 1/\binom{n}{t^*}. \tag{5.31}$$

All the sequences in the microcanonical ensemble belong to the typical set T_{mic} of it, so the smallest space needs to store the information carried by those sequences $\ln|T_{\text{mic}}|$ is equal to the Boltzmann entropy of the microcanonical ensemble

$$\ln|T_{\text{mic}}| = \ln\binom{n}{t^*} = \ln\Omega_{\text{mic}} = S_{\text{mic}}, \tag{5.32}$$

where Ω_{mic} is the total number of sequences under this hard constraints t^*. We can find that the canonical ensemble description showed above is equivalent to the classical information theory when the i.i.d. assumption hold.

The difference between the space to store the information carried by the sequences in the two ensembles is equal to

$$\ln |T_{\rm can}| - \ln |T_{\rm mic}| = S_{\rm can} - S_{\rm mic}. \tag{5.33}$$

It is the total correlations $\mathbb{C}$ in the sequence A

$$\mathbb{C} = \sum_{A\in\mathcal{A}} P_{\rm mic}(A|t^*) \ln \frac{P_{\rm mic}(A|t^*)}{P_{\rm can}(A|\theta^*)}. \tag{5.34}$$

When it is rescaled n,

$$\frac{1}{n}\mathbb{C} = \frac{1}{n}[\ln \frac{n^n}{t^{*t^*}(n-t^*)^{n-t^*}} - \ln \binom{n}{t^*}], \tag{5.35}$$

It is equal to the relative entropy density between the two ensembles.

According to the Stirling approximation, the limit of the rescaled difference is equal to 0

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{C} = \lim_{n\to\infty} \frac{1}{n}[\frac{1}{2}\ln 2\pi t^*(1-\frac{t^*}{n})] = 0. \tag{5.36}$$

It means the canonical ensemble will converge to the microcanonical one in the thermodynamic limit. It also shows that the limit of the information storage is the same in the two ensemble descriptions.

In statistical physics, this is the measure-level ensemble equivlaence [8]. The logarithm difference $\ln |T_{\rm can}| - \ln |T_{\rm mic}|$ is the relative entropy

$$\ln |T_{\rm can}| - \ln |T_{\rm mic}| = S_{\rm can} - S_{\rm mic} = S(P_{\rm mic}||P_{\rm can}), \tag{5.37}$$

which grows like $o(n)$.

The microcanonical ensemble description has realized the constraints in the signal generation exactly. The classical information theory under this case is a particular example of the canonical ensemble. It means the effectiveness of the classical information theory is based on the EE between the microcanonical ensemble and the canonical ensemble with the global constraints t^*.

5.5 Extensive number of constraints and ensemble nonequivalence

As already strictly prov proved in Chapter 3, that the ensemble nonequivalence is generally exist in the systems with extensive local constraints. When we use K to represents the numbers of constraints in a systems. In the matrix under global constraints, there is only one constraint, $K = 1$. When the $m \times n$ matrix is under row local constraits, there is m constraints in it, $K = m$. If the matrix is under column local constraints, then there is $K = n$ constraints in it. The matrix under

coupled local constraints has $K = m + n$ constraints. The extension of constraints in the matrix $\mathbf{X}$ is a localization of dependencies in information sequences. According to the canonical ensemble description, the units in the matrix under global constraints have a homogeneous interactions. When the matrix are under row local constraints, the units can be divided into m parts and each part have the same interactions. Obviously, when the $m \times n$ matrix under coupled constraints, the interactions have be localized to each unit.

The classical information theory is a special case of the global constrained canonical ensemble description, when the matrix $\mathbf{X}$ only has one row $m = 1$ and n columns, $t = \sum_{i=1}^{n} x_{1i}$ [27]. It is under EE, and the limit of information storage in the two ensemble descriptions is equivalent.

The single generation by the m independent variables with different probability distribution in the network information theory is an m extension of the global constrained classical information theory. We can use the matrix with the finite row local constraints $\vec{r}$ to describe the sequences generated by the i.i.d multivariate information source in the network information theory [27, 30]. There are $K = m$ constraints in it. It is also under EE. We can find the description of it in Appendix 5.A.

If we only focus on the spatial dependences, there will be n constraints in the sequence, $K = n$. The sequences generated by this non-stationary process can be represented by the states of the matrices with local column constraint $\vec{c}$. According to Stirling's approximation, the limit of rescaled total correlations is bigger than 0 as

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} [\frac{1}{2} \ln 2\pi c_i^* (1 - \frac{c_i^*}{m})] > 0. \tag{5.38}$$

Sequences with this spatial dependence are under EN. Details of proof are in Appendix 5.B.

The coupled local constraints are the two kinds of dependences work simultaneously, $\vec{C}(\mathbf{X}) = [\vec{c}, \vec{r}]$. There are $K = m + n$ constraints in it. The probability of the microcanonical ensemble is difficult to calculate, but we can still get the conjugate canonical probability as

$$P_{\text{can}}(\mathbf{X}|\vec{\theta^*}) = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{e^{-(\alpha_i^* + \beta_j^*) x_{ji}}}{e^{-(\alpha_i^* + \beta_j^*)} + 1}. \tag{5.39}$$

According to the results in [27], those sequences are also under EN.

The constraints' extension has two paths: the first one is from global to row-local constraints, then coupled local constraints, the second one is from global to column-local constraints, then to the coupled local constraints. As we already know, both of the two paths will break the EE, but the difference in the two paths will show which kind of dependence subleading the breaking of EE.

In order to check how is the total correlation will change when the constraint is extended in the system, we set a series of models with homogeneous dependencies. The homogeneous spatial interaction means the c_i in the column local constraint is

equal to each other as c^*. The homogeneous temporal dependencies need each element in the row local constraints equal to each other as $r_j = r^*$. And the global constraint is equal to $t^* = n \times c^* = m \times r^*$. All the three special cases can be described by coupled constraints matrix ensemble with different definitions of *Hamilotonian*. Thus, we have a series of models where the spatial interactions and temporal dependencies are homogeneous but still under extensive local constraints.

The signal generation by multivariate information source under the same temporal dependencies should have the same value with row local constraint $r_j^* = r^*$. The canonical entropy of this matrix ensemble is equal to

$$S_{\text{can}}^{(K=m)} = m \times \ln \frac{n^n}{r^{*r^*}(n-r^*)^{n-r^*}}. \tag{5.40}$$

As we already mentioned before, it is a m times linear extension of the single independent identical signal generating in the classical information theory, and it is under EE.

In information sequences with homogeneous spatial interactions in the information sources, the canonical entropy should equal to

$$S_{\text{can}}^{K=n} = n \times \ln \frac{n^n}{r^{*r^*}(n-r^*)^{n-r^*}}. \tag{5.41}$$

It is equivalent to the signal generated by the multivariate information source with an identical probability distribution, but variables in the information source are not independent. It has nonneglected dependencies among them.

If the homogeneous spatial interactions and temporal dependencies work simultaneously, information sequences with these coupled constraints have the same column and row local constraint as $c_i^* = c^*$, $r_j^* = r^*$. The canonical entropy of this matrix ensemble should equal to

$$S_{\text{can}}^{(K=m+n)} = n \times \ln \frac{m^m}{c^{*c^*}(m-c^*)^{m-c^*}} \tag{5.42}$$

when we focus on the spatial interactions.

The canonical entropy is equal to

$$S_{\text{can}}^{(K=m+n)} = m \times \ln \frac{n^n}{r^{*r^*}(n-r^*)^{n-r^*}} \tag{5.43}$$

when the calculation is focused on the temporal dependencies. It is also equivalent to the matrix ensemble with one side column local constraints,

Canonical entropies show above are all equivalent to the one with global constraints t^*, when the r^* is replaced by $r^* = t^*/m$, or the c^* is replaced by $c^* = t^*/n$. It shows that the canonical ensemble description of information Sequences under four different local constraints that are implied by the homogeneous dependencies are equivalent to each other,

$$S_{\text{can}}^{(K=1)} = S_{\text{can}}^{(K=m)} = S_{\text{can}}^{(K=n)} = S_{\text{can}}^{(K=m+n)}. \tag{5.44}$$

Thus, under soft constraints, the four homogeneous signal generations have the same information-theoretical bounds. The space to store the information generated by it is the same.

However, the Boltzmann entropy of information sequences with the four different dependences is different. Under the global constraint, it is equal to

$$S_{\text{mic}}^{(K=1)} = \ln \binom{mn}{t^*}. \tag{5.45}$$

Under row-local constraints, the Boltzmann entropy is equal to

$$S_{\text{mic}}^{(K=m)} = m \times \ln \binom{n}{r^*}. \tag{5.46}$$

When it is under column-local constraints, the Boltzmann entropy is equal to

$$S_{\text{mic}}^{(K=n)} = n \times \ln \binom{m}{c^*}. \tag{5.47}$$

When the two dependence is working Simultaneously, the Boltzmann entropy of the hard constrained sequences is difficult to calculate, but we can still get the approximation by the Eq (5.11), and the value of it is smaller than $S_{\text{mic}}^{(K=n)}$. Thus, the relationship between the four Boltzmann entropy is

$$S_{\text{mic}}^{(K=1)} > S_{\text{mic}}^{(K=m)} > S_{\text{mic}}^{(K=n)} > S_{\text{mic}}^{(K=m+n)}. \tag{5.48}$$

The relationship between the breaking of ensemble equivalence and the extension of constraints is shown in Fig.5.2.

These results show above proved that the temporal dependence of each independent variable in information sources is not enough to break the EE between the hard and soft constraint's description of information sequences. This EE allows classical information theory can be applied in the independent multivariate information sources system to estimate the limit of information storage [30].

5.6 Information transmission with coupled sources

In classical information theory, the maximum speed of reliable information transmission through a channel is the channel capacity. It is the other vital bounds of the information theory, and it is equal to the mutual information between the information source $\mathbf{x}$ and the received signal $\mathbf{y}$ [33].

To transport the information carried by matrix $\mathbf{X}$ with local constraints, we still need to code it by codes G with length L. In the receiver, we will receive a matrix $\mathbf{Y}$. The information we will transport through this channel is equal to $L \times R$, and it is decided by the uncertainty from the channel $\mathbf{H}(\mathbf{X}|\mathbf{Y})$ and the information carried

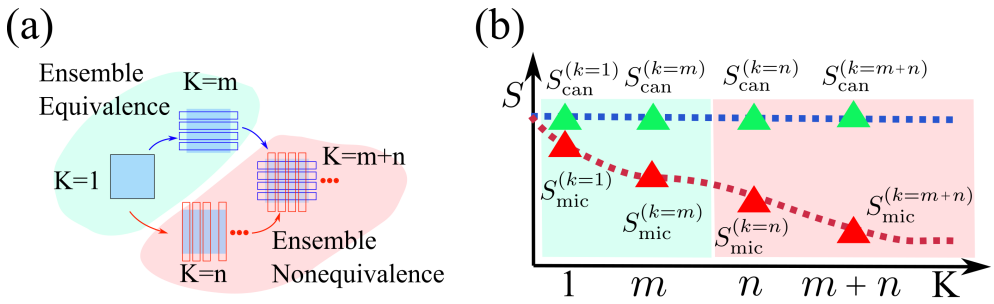

Figure 5.2. The relationship between the constraints' extension and the breaking of EE is shown in (a). The extension of constraints in matrix described systems have two paths. First, the constraints is extend from global constraint to row local constraints then to coupled constraints. The other path is the extendtion from global constraints to column local constraints then coupled constraints. Systems are still under EE when the constraints extend from global to row local constraints ($K = 1 \to K = m$), but when the constraints extend to the coupled local constraints from row local constraints $K = m \to K = m + n$, the EE is broken. On the other path, the extension of global constraint to column local constraints already breaks the EE. Obviously, the extension from column local constraints to coupled local constraints is already under EN. This result shows that the extension of global constraints to finite numbers local constraints is not enough to break the EE. Only when the extension of local constraints has the same order as the increasing of system's size, or even faster, EE will break by the extension of constraints. On the other hand, the relationship between the canonical entropy and microcanonical entropy of the systems with extensive constraints under homogeneous dependencies in (b) shows that the non-vanished fluctuation of the constraints in the canonical ensemble does not lead to the breaking of EE. The EN is formed by the reduction of possible configurations of sequences in the microcanonical ensemble. Because the canonical ensemble with soft constraints has the same Shannon entropy when it is under homogeneous local and global constraints. But the Boltzmann entropy of the microcanonical ensemble with hard constraints is declined.

by the sequences $\mathbf{X}$ sent to the receiver $\mathbf{I}(\mathbf{X};\mathbf{Y})$. Thus, we will have the relationship as

$$L \times R = \mathbf{H}(\mathbf{X}|\mathbf{Y}) + \mathbf{I}(\mathbf{X};\mathbf{Y}) \geq S(\mathbf{X}). \tag{5.49}$$

In the zero-error channel, the matrix we received $\mathbf{Y}$ is the matrix we sent $\mathbf{X}$, the value of $\mathbf{H}(\mathbf{X}|\mathbf{Y}) = 0$, the mutual information $\mathbf{I}(\mathbf{X};\mathbf{Y}) = S(\mathbf{X})$, then the effective information we need transport is equal to the Shannon entropy of the sequence $S(\mathbf{X})$. When the message we received is a random matrix, the mutual information $\mathbf{I}(\mathbf{X};\mathbf{Y}) = 0$, but the channel uncertainty is equal to $\mathbf{H}(\mathbf{X}|\mathbf{Y}) = S(\mathbf{X})$.

Both in the two cases, the smallest information needs to transport are all decided by the entropy of the sequences $S(\mathbf{X})$. Thus, the symbol rate $R = S(\mathbf{G})/L$ or the channel capacity is also affected by the hard or soft constraints in this signal generation process. When using the microcanonical to describe the sequences, the Boltzmann entropy S_{mic} is smaller than the Shannon entropy in the conjugate canonical ensemble S_{can}. Thus, the symbol rate R_{mic} needs for the microcanonical ensemble is smaller than the symbol rate R_{can},

$$R_{\text{mic}} \leq R_{\text{can}}. \tag{5.50}$$

It means the microcanonical ensemble with local constraints is reliable than the canonical ensemble one when we use the same channel to transport the information carried by the sequences ensemble. Because it needs less information carried for each symbol in the code, the system will have more redundancy.

5.7 Conclusions

The information is not only the bits that flow in the electronic communication systems. It also generally exists in the natural systems, e.g., the activity of neurons in the nervous system and the fluctuation of the financial system. The birth of classical information theory ignores the interactions and dependencies among the information sources, as it is limited by the initial structure of artificial communication systems. However, this imperfection has been amplified when the classical information theory is attempted to describe the signal generation in natural systems with heterogeneous dependencies and interactions.

The existence of heterogeneous dependencies in the signal generation has broken the i.i.d assumption in the classical information theory, which is the cornerstone of the finding that the uncertainty of information sources decides the limit of the information storage. Thus, finding the information-theoretical bounds (i.e., the limit of information storage and channel capacity) of the natural signal generation should focus on the information sequences, not the interacting and temporal dependent information source.

The information sequence with heterogeneous interactions has an extensive length. It is analogous to the state in the thermodynamic system. That is why statistical ensembles can be used to describe the information sequences in this work.

In ensemble descriptions of the information sequence, the heterogeneous dependencies imply local constraints in the statistical ensemble. The microcanonical ensemble description requires all the local constraints fixed in the information sequences generated by the heterogeneous interacting information sources. It is closer to the real process of signal generation in natural systems. On the contrary, the canonical ensemble description is a maximum entropy approximation of the signal generation. The total correlation is equal to the relative entropy between the microcanonical and canonical ensemble, which is the indicator of the measure level EN. It means the EN appearance in the heterogeneous interacted signal generation also has a connection with the degree of dependences among the units in it.

The sequences described by the microcanonical ensemble with hard local constraints need less space to store the information carried by them compared with the canonical one. In the information transmission, the microcanonical ensemble also needs a lower symbol rate to transport the information than the canonical ensemble. Both results show that the microcanonical ensemble description is better than the canonical one from the limit of information storage. However, the probability distribution of the microcanonical ensemble is mathematically difficult to calculate than the canonical ensemble. Thus, when we want to build a communication system with a multi-coupled source, we still need to consider the trade-off between the cost of calculation to maintain the hard constraints in the microcanonical ensemble and the waste of space and channel capacity to use the canonical ensemble with soft constraints.

We also find that the classical information theory under the i.i.d. assumption is a special case of the canonical ensemble description when the column-local constraints have the same value or when the sequences have homogeneous spatial dependence. The effectiveness of the classical information theory is based on the EE between the canonical ensemble and the natural signal generation.

This non-stationary process also gives a chance to learn how is the extension of constraints causes EN in a system. From the canonical ensemble description, the global constraints and two different local constraints are special cases of the coupled constraints matrix ensemble. The extension of constraints in it has two paths. The first one is from global constraints to row local constraints then coupled constraints. The second one is from global constraints to column local constraints, then coupled constraints. The matrix under global constraints and finite row local constraints is in the EE. However, this equivalence will break when there are column-local constraints and coupled constraints. The column-local constraints are used to describe the spatial interaction among the units in the information source. Thus, the EN in sequences is caused by the spatial interactions among the units in the signal generation, not the temporal dependence of finite independent variables in the information source. Even the row local constraints also have influences on all the units in sequences.

The same Shannon entropy and the decreased Boltzmann entropy in the process of constraints' extention illustrate that the breaking of EE is caused by the reduction of the possible configurations in the microcanonical ensemble when there are homogeneous dependencies. This mechanism is different from the traditional one, which shows that the EN is caused by the non-vanished fluctuation of constraints in the

canonical ensemble. This finding extended our understanding of the EN in statistical physics.

Appendix 5.A Row local constraints and multivariate independence source

There is m rows in the matrices $\mathbf{X}$, which means there are m units in the information source. The signal generation of the multivariate independent information source is under ensemble equivalence, as there is a finite number of local constraints in it, and there is no phase transition in the information sequence [27]. This signal generation still can be described by the classical information theory. Then according to the AEP, we can find the limit of information storage.

The sequence generated by the information source $\vec{x} = [b_1, b_2, \cdots, b_j, \cdots, b_m]$ with m independent variables is an $m \times n$ matrix B. As the m units in the information source are independent with each other, the sequence B can be divided into m row vectors $\mathbf{B} = \{\vec{R}_1; \vec{R}_2; \cdots; \vec{R}_m\}$. Each row vectors $\vec{R}_j$ of the matrix B has n elements in it. According to the classical information theory, the m i.i.d. random variables may have different probability to get different value, so the probability of sequence $\mathbf{B}$ to appear in the signal generation is equal to

$$P(\mathbf{B}) = \prod_{j=1}^{m} P(\vec{R}_j) = \prod_{j=1}^{m} p_j^{r_j}(1-p_j)^{n-r_j}. \tag{5.51}$$

Here we still focus on the binary information sequence. Thus, p_j is the probability of each unit in row j to have value 1. The r_j is the number of units in row j to have a value of 1, and it will affect the process of signal generating when different constraints model it.

The value of p_j can be obtained by the average value of total units with value 1 in each row as $p_j = r_j^*/n$. When the jth variable is represented by b_j, then the AEP will be generalized as

$$\begin{aligned} \frac{1}{n}\ln P(\mathbf{B}) &= \sum_{j=1}^{m}[\frac{r_j^*}{n}\ln\frac{r_i^*}{n} + \frac{n-r_j^*}{n}\ln\frac{n-r_j^*}{n}] \\ &\to -\sum_{j=1}^{m} s(b_j) \end{aligned} \tag{5.52}$$

Sequences belonging to the typical set of this system with multivariate independence information source should satisfy the following condition

$$T_\epsilon = \{\mathbf{B}|e^{-n\sum_{j=1}^{m} s(b_j)-\epsilon} \leq P(\mathbf{B}) \leq e^{-n\sum_{j=1}^{m} s(b_j)+\epsilon}\}. \tag{5.53}$$

The space to store the information carried by it is equal to

$$\ln|T_\epsilon| = n \times \sum_{j=1}^{m} s(b_j). \tag{5.54}$$

This result shows that the uncertainty of the information source still decides the limit of information storage. Even there are m independent variables in it.

Next, we will introduce the ensemble description of the information sequence generated by the multivariate independent information sources.

Canonical ensemble description- The information sequence $\mathbf{B}$ generated by the independent variables with different probability distribution can be modeled by the matrix $\mathbf{X}$ with the row local constraints $\vec{r}(\mathbf{X}) = [r_1^*, r_2^*, \cdots, r_m^*]$, where $r_j^* = \sum_{i=1}^{n} x_{ji}$. The maximum likelihood parameter $\vec{\beta} = [\beta_1^*, \beta_2^*, \cdots, \beta_m^*]$ has m elements. The *Hamiltonian* is still the linear combination of the constraints and parameters $H(\mathbf{X}) = \sum_{j=1}^{m} \beta_j^* r_j^*$. In the binary case, the partition function of this matrix ensemble is

$$Z(\vec{\beta}) = \prod_{j=1}^{m} (1 + e^{-\beta_j^*})^n. \tag{5.55}$$

Then we can have the canonical probability of the state $\mathbf{X}$ with row local constraints $\vec{r}(\mathbf{X})$ as

$$P_{\text{can}}(\mathbf{X}) = \prod_{j=1}^{m} \frac{e^{-\beta_j^* r_j^*}}{(1 + e^{-\beta_j^*})^n}. \tag{5.56}$$

The parameter β_j^* is decided by the corresponding average value of row local constraints $\langle r_j \rangle$.

Space to store the information carried by the information sequences $\mathbf{B}$ still can be quantified by the AEP. The rescaled logarithm of the probability is equal to

$$\frac{1}{n} \ln P_{\text{can}}(\mathbf{X}) = \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n} \ln \frac{e^{-\beta_j^* x_{ji}}}{1 + e^{-\beta_j^*}}. \tag{5.57}$$

When the length n goes to infinite, the probability of the units in jth row to have value 1 is equal to average value of x_{ji} as

$$\langle x_{ji} \rangle = \frac{e^{-\beta_j^*}}{1 + e^{-\beta_j^*}} = \frac{r_j^*}{n}. \tag{5.58}$$

The sum of the logarithms of the probability for the $m \times n$ units will equal the sum of the m variables Shannon entropy

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} \ln P_{\text{can}}(\mathbf{X}) &= \lim_{n \to \infty} \sum_{j=1}^{m} [\frac{r_j^*}{n} \ln \frac{r_j^*}{n} + \frac{n - r_j^*}{n} \ln \frac{n - r_j^*}{n}] \\ &\to \sum_{j=1}^{m} s(b_j). \end{aligned} \tag{5.59}$$

Then space to store the information carried by the sequences in the typical set $T_{\text{can}}^{\epsilon=0}$, which satisfy the following condition

$$T_{\text{can}}^{\epsilon=0} = \{\mathbf{X} | P_{\text{can}}(\mathbf{X}) = e^{n \sum_{j=1}^{m} s(b_j)}\} \tag{5.60}$$

The logarithm of the number of sequence in the typical set also equal to $\ln|T_{\text{can}}| = n\sum_{j=1}^{m} s(b_j)$, which is the Shannon entropy of the sequences $S_{\text{can}}(\mathbf{X})$

$$\ln|T_{\text{can}}^{\epsilon=0}| = S_{\text{can}}(\mathbf{X}) = \sum_{j=1}^{m} \ln[\frac{n^n}{r_j^{*r_j^*}(n-r_j^*)^{n-r_j^*}}]. \tag{5.61}$$

It is equal to the $\ln|T_\epsilon| = n \times \sum_{j=1}^{m} s(b_j)$ in the classical information theory. This result shows that the method in the classical information theory is a particular case of the canonical ensemble description.

Microcanonical ensemble description- The microcanonical ensemble under the hard constraints needs the total number of units with the value 1 in each row of all the sequences $\mathbf{X}$ are fixed the same as $\vec{r^*} = [r_1^*, r_2^*, \cdots, r_j^*, \cdots, r_m^*]$. The probability of each sequence decided by the total number of configurations of this local row constrained sequences with constraints $\vec{r^*}$ as

$$P_{\text{mic}}(\mathbf{X}) = 1/\prod_{j=1}^{m} \binom{n}{r_j^*}. \tag{5.62}$$

All the sequences in the microcanonical ensemble belong to the typical set of it, the space to store the information carried by it is equal to the *Boltzmann entropy* of the sequences as

$$\ln|T_{\text{mic}}^{(\vec{r^*})}| = \sum_{j=1}^{m} \ln\binom{n}{r_j^*} = \ln\Omega_{\vec{r^*}}. \tag{5.63}$$

The rescaled difference between the logarithm of the two ensemble's probability is close to the rescaled total correaltions $\frac{1}{n}\mathbb{C}$ as

$$\frac{1}{n}\mathbb{C} = \frac{1}{n}\sum_{j=1}^{m}[\ln\frac{n^n}{r_j^{*r_j^*}(n-r_j^*)^{n-r_j^*}} - \ln\binom{n}{r_j^*}] \tag{5.64}$$

The asymptotic behaviour of $\frac{1}{n}\mathbb{C}$ decided by m, in this work, $m = o(n)$, thus the limit of $\frac{1}{n}\mathbb{C}$ is equal to 0 as

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{C} = \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{m}[\frac{1}{2}2\pi r_j^*(1-\frac{r_j^*}{n})] = 0. \tag{5.65}$$

The signal generalization by the classical independent multivariate information sources is under ensemble equivalence. That is why we can still use the AEP in the estimation of information-theoretical bounds.

Appendix 5.B Column local constraints and non-stationary process

In the binary matrix $\mathbf{X}$, when the constraints are the time-variation total energy can be used by all the m units in different times, the process recorded by the matrices

ensemble $\mathcal{X}$ is under column local constraints $\vec{c^*}(\mathbf{X}) = [c_1^*, c_2^*, \cdots c_n^*]$. At time i, the constraint is the sum of all the units in column i of matrice $\mathbf{X}$ as $c_i^* = \sum_{j=1}^{m} x_{ji}$. The probability of the variable getting different values in the signal generation will change with time, so applying the classical information theory in this process is impossible, but we can still use the ensemble descriptions.

Microcanonical ensemble description- In the microcanonical ensemble description of sequences, the probability of each state is based on the total number of configurations in it as

$$P_{\text{mic}}(\mathbf{X}|\vec{c^*}) = 1/\prod_{i=1}^{n} \binom{m}{c_i^*}. \tag{5.66}$$

Obviously, all the sequences with hard constraints still belong to the typical set of it, so the space to store the information carried by it is equal to the *Boltzmann entropy* of it as

$$\ln |T_{\text{mic}}^{(\vec{c^*})}| = -\ln P_{\text{mic}}(\mathbf{X}|\vec{c^*}) = \sum_{i=1}^{n} \ln \binom{m}{c_i^*}. \tag{5.67}$$

This is the smallest space need to store the information generated under the constraints $\vec{c^*}$.

Canonical ensemble description- We need the canonical ensemble to describe the sequences $\mathbf{X}$ with soft constraints. The correspond parameter $\vec{\alpha}^* = [\alpha_1^*, \cdots, \alpha_n^*]$ has n elements in it. The *Hamiltonian* of the binary matrix still is a linear combination of the parameter and constraints, $H = \sum_{i=1}^{n} c_i \alpha_i^*$. The partition function is equal to

$$Z(\vec{\alpha}^*) = \prod_{i=1}^{n} (e^{-\alpha_i^*} + 1)^m, \tag{5.68}$$

and the probability of each sequence in the system is equal to

$$P_{\text{can}}(\mathbf{X}|\vec{\alpha}^*) = \prod_{i=1}^{n} \frac{e^{-\alpha_i^* c_i}}{(e^{-\alpha_i^*} + 1)^m}. \tag{5.69}$$

The average value of x_{ji} equal to

$$\langle x_{ji} \rangle = \frac{e^{-\alpha_i^*}}{e^{-\alpha_i^*} + 1} = \frac{c_i^*}{m}. \tag{5.70}$$

This condition imply the relationship between the parameter α_i^* and the corresponding constraints c_i^* as

$$e^{-\alpha_i^*} = \frac{c_i^*}{m - c_i^*}. \tag{5.71}$$

The space to store the information carried by it can not be approached by the AEP,

but we can find the Shannon entropy of the whole sequences as

$$\begin{aligned} S_{\text{can}} &= \sum_{i=1}^{n} \alpha_i^* c_i^* + \ln Z(\vec{\alpha}^*) \\ &= \sum_{i=1}^{n} [-c_i^* \ln \frac{c_i^*}{m - c_i^*} + m \ln \frac{m}{m - c_i^*}] \\ &= \sum_{i=1}^{n} [\ln \frac{m^m}{c_i^{*c_i^*} (m - c_i^*)^{m - c_i^*}}]. \end{aligned} \tag{5.72}$$

Under this constraint, it is difficult to find the typical set of the sequences by the classical information theory, but we can still find the difference of the rescaled logarithm of the probability between microcanonical and canonical ensemble is equal to the rescaled total correlations in the sequence as $\frac{1}{n} \ln \frac{P_{\text{mic}}(D)}{P_{\text{can}}(D)} = \frac{1}{n} \mathbb{C}$, which is equal to

$$\frac{1}{n} \mathbb{C} = \frac{1}{n} \sum_{i=1}^{n} [\ln \frac{m^m}{c_i^{*c_i^*} (m - c_i^*)^{m - c_i^*}} - \ln \binom{m}{c_i^*}]. \tag{5.73}$$

As we already know the total correlations is equal to the relative entropy, thus the difference between the rescaled logarithm of the probability is equal to the relative entropy density as $\frac{1}{n} \mathbb{C} = \frac{1}{n} S(P_{\text{mic}} || P_{\text{can}})$. When the value of n goes to infinite, the limit

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{C} = O(n) > 0, \tag{5.74}$$

so this signal generation is under ensemble nonequivalence.

The total correlation between the microcanonical ensemble and canonical ensemble of the matrix $\mathbf{X}$ with column local constraints $\vec{c}^*$ grows like $O(n)$ in the thermodynamic limit. It means the information carried by the typical set of the two ensemble descriptions is different under the spatial interactions. This difference also manifested in the appearance of measure level ensemble nonequivalence in it.

Appendix 5.C Coupled local constraints and multi-coupled process

In the former two subsections, we have introduced the ensemble description of the information generating with independent spatial or temporal dependencies. They are modelled by the matrix ensemble with local column or row constraints. In this subsection, we will study what will happens when the two constraints work simultaneously $\vec{C}(\mathbf{X}) = [\vec{c}, \vec{r}]$.

Canonical ensemble description- As we know, under the coupled local constraints, the maximum likelihood parameter will be $\vec{\theta} = [\vec{\alpha}, \vec{\beta}]$. The constraint $\vec{c}$ is the local column constraints. The $\vec{r}$ is the local row constraints. The corresponding

parameters $\vec{\theta}$ also comes from the maximum likelihood parameter of row and column local constraints. The *Hamiltonian* is still the linear combination of constraints and parameters as

$$H = \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) x_{ji}. \tag{5.75}$$

We still focus on the binary matrix, so the partition function of these sequences with coupled local constraints is

$$Z(\vec{\theta}) = \prod_{i=1}^{n} \prod_{j=1}^{m} [e^{-(\alpha_i + \beta_j)} + 1]. \tag{5.76}$$

Probability of the sequence $\mathbf{X}$ to appears in the signal generating process is equal to

$$P_{\text{can}}(\mathbf{X}) = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{e^{-(\alpha_i^* + \beta_j^*) x_{ji}}}{e^{-(\alpha_i^* + \beta_j^*)} + 1}. \tag{5.77}$$

The probability of each unit in matrix $\mathbf{X}$ is decided by the value of x_{ji} and the corresponding parameter α_i and β_j. The smallest space needs to store the information carried by them is still equal to the Shannon entropy of it as

$$S_{\text{can}} = \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) \langle x_{ji} \rangle + \ln Z(\vec{\theta}). \tag{5.78}$$

The average value of x_{ji} is difficult to get exactly, but we can still find the exact value of it under the special setting of the constraints.

Appendix 5.D Homogeneous dependencies under different constraints

As a special case of heterogeneous dependencies, homogeneous dependencies gives a chance for us to check what will happens when the canonical ensemble descriptions are equivalent under different constraints, but the microcanonical ensemble descriptions are different. Obviously, the global constraints and one-sided local constraints (both the column and row local constraints) are all the special cases of the coupled constraints under homogeneous dependencies. Thus, in this part, we will introduce the coupled constrained ensemble descriptions of the signal generation with homogeneous dependencies under different constraints.

5.D.1 Global constraint t^*

The canonical ensemble description of the information sequence under global constraints is a special case of the coupled local constraints when there is only one constraint, and the corresponding maximum likelihood parameter is equal to $\alpha^* + \beta^*$.

The *Hamiltonian* of it is equal to

$$H = \sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha^* + \beta^*)x_{ji}. \tag{5.79}$$

The partition function will be

$$Z(\theta^*) = \prod_{i=1}^{n}\prod_{j=1}^{m}[e^{-(\alpha^*+\beta^*)} + 1] = [e^{-(\alpha^*+\beta^*)} + 1]^{mn}. \tag{5.80}$$

The probability of the states under this special case is equal to

$$P_{\text{can}}(\mathbf{X}) = \frac{e^{-\sum_{j=1}^{m}\sum_{i=1}^{n}(\alpha^*+\beta^*)x_{ji}}}{[e^{-(\alpha^*+\beta^*)} + 1]^{mn}}. \tag{5.81}$$

When the sum of all elements in sequence is equal to t, the probability of $\mathbf{X}$ to appears in the signal generation is equal to that in the canonical ensemble with global constraints t^* as

$$P_{\text{can}}(\mathbf{X}) = \frac{e^{-(\alpha^*+\beta^*)t}}{[e^{-(\alpha^*+\beta^*)} + 1]^{mn}}. \tag{5.82}$$

When the average value of the total number of units in the information sequence with value 1 $\langle t\rangle = t^*$ as

$$\langle t\rangle = \sum_{\mathbf{X}\in\mathcal{X}} t(\mathbf{X})P_{\text{can}}(\mathbf{X}) = t^*. \tag{5.83}$$

We can have the parameter $\alpha^* + \beta^*$ is equal to

$$e^{-(\alpha^*+\beta^*)} = \frac{t^*}{mn - t^*}. \tag{5.84}$$

Then we can find the Shannon entropy of the information sequence with the number of constraints equal to 1 is equal to $\ln P_{\text{can}}(\mathbf{X}|t^*)$ as

$$S_{\text{can}}^{(K=1)} = \ln \frac{mn^{mn}}{t^{*t^*}(mn - t^*)^{mn-t^*}}. \tag{5.85}$$

Then microcanonical ensemble description of the information sequence with global constraints t^* have Ω_{t^*} states in it. Thus, the probability of each state in it is equal to

$$P_{\text{mic}}(\mathbf{X}|t^*) = \frac{1}{\Omega_{t^*}} = 1/\binom{mn}{t^*}. \tag{5.86}$$

Therefore, the *Boltzmann* entropy of the microcanonical ensemble with global constraints is equal to

$$S_{\text{mic}}^{(K=1)} = \ln \binom{mn}{t^*} \tag{5.87}$$

According to Stirling's approximation and the results in [27], the difference between the Shannon entropy of canonical and the microcanonical ensemble is equal to

$$S_{\text{can}}^{(K=1)} - S_{\text{mic}}^{(K=1)} = \frac{1}{2}\ln\left[2\pi t^*\left(1 - \frac{t^*}{\pm mn}\right)\right][1 + o(1)]\,. \tag{5.88}$$

Thus, it is under ensemble equivalence.

5.D.2 Row local constraints

When we use $r^* = t^*/m$ as the row local constraints, there are m constraints in the information sequence. It also can be model by the matrix with coupled constraints, when the elements in the column local constraints is equal to each other as $c_i^* = c^*$ but the elements in the row local constraints is $\vec{r}^*$. We will have another special case of the coupled constrained ensemble description, which is that the interactions among all the units in the information sources are identified in the whole process of signal generating, but the temporal dependence of each unit is different. Then the corresponding parameter will change as $\alpha_i^* = \alpha^*$. Thus, the *Hamiltonian* of this special case will be

$$H = \sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha^* + \beta_j^*)x_{ji}. \tag{5.89}$$

The partition function of this coupled local constraints sequences will be

$$Z(\vec{\theta}^*) = \prod_{i=1}^{n}\prod_{j=1}^{m}[e^{-(\alpha^*+\beta_j^*)} + 1]. \tag{5.90}$$

We can get the probability of states in the sequences under this special constraints as

$$P_{\text{cam}}(\mathbf{X}|\vec{\theta}^*) = \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{e^{-(\alpha^*+\beta_j^*)x_{ji}}}{e^{-(\alpha^*+\beta_j^*)}+1}. \tag{5.91}$$

Because each element in the column local constraints are equal to each other as $c_i^* = c^*$, the average value of x_{ji} in this sequence is equal to

$$\langle x_{ji}\rangle = \frac{e^{-(\alpha^*+\beta_j^*)}}{e^{-(\alpha^*+\beta_j^*)}+1} = \frac{r_j^*}{n}. \tag{5.92}$$

It implies the relationship between $\alpha^* + \beta_j^*$ and r_j^* as

$$e^{-(\alpha^*+\beta_j^*)} = \frac{r_j^*}{n - r_j^*}. \tag{5.93}$$

The Shannon entropy of this sequences under this local constraints is equal to

$$\begin{aligned} S_{\text{can}}^{(K=m)} &= \ln P_{\text{can}}(\mathbf{X}^*|\vec{\theta}^*) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}[(\alpha^* + \beta_j^*)\langle x_{ji}\rangle] + \ln Z(\vec{\theta}^*) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}[-\frac{r_j^*}{n}\ln\frac{r_j^*}{n-r_j^*}] + \ln Z(\vec{\theta}^*) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}[-\frac{r_j^*}{n}\ln\frac{r_j^*}{n-r_j^*} + \ln\frac{n}{n-r_j^*}] \\ &= \sum_{j=1}^{m}[-r_j^*\ln\frac{r_j^*}{n-r_j^*} + \ln\frac{n^n}{(n-r_j^*)^n}] \\ &= \sum_{j=1}^{m}[\ln\frac{n^n}{{r_j^*}^{r_j^*}(n-r_j^*)^{n-r_j^*}}], \end{aligned} \tag{5.94}$$

It also equals the Shannon entropy of the sequences under one-sided row local constraints.

The microcanonical ensemble description of this special case is equal to the one where there are only row local constraints in the information sequence. Thus, we can have the *Boltzmann* entropy of this information source as

$$S_{\text{mic}}^{(K=m)} = \sum_{j=1}^{m}\ln\binom{n}{r_j^*} \tag{5.95}$$

It is also under ensemble equivalence [27].

5.D.3 Column local constraints

When all the elements in the row constraints is equal to each other as $r_j^* = r^*$, the column local constraints still remain as $\vec{c}^*$, the corresponding maximum likelihood parameter will be $\vec{\theta}^* = [\vec{\alpha}^*, \vec{\beta}^*]$, but all the elements in $\vec{\beta}^*$ is equal to each other as $\beta_j^* = \beta^*$. Then the *Hamiltonian* of this coupled constrained canonical ensemble will be

$$H = \sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i^* + \beta^*)x_{ji}. \tag{5.96}$$

The partition function need to consider all the possible configurations as

$$Z(\vec{\theta}^*) = \prod_{i=1}^{n}\prod_{j=1}^{m}[e^{-(\alpha_i^*+\beta^*)} + 1]. \tag{5.97}$$

Probability of states under this constraint is equal to

$$P_{\text{cam}}(\mathbf{X}|\vec{\theta}^*) = \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{e^{-(\alpha_i^*+\beta^*)x_{ji}}}{e^{-(\alpha_i^*+\beta^*)}+1}. \tag{5.98}$$

The average value of each element $\langle x_{ij}\rangle$ is

$$\langle x_{ij}\rangle = \frac{e^{-(\alpha_i^*+\beta^*)}}{e^{-(\alpha_i^*+\beta^*)}+1}. \tag{5.99}$$

Each element in the row local constraints is equal to each other as $r_j^* = r^*$, so the average value of element x_{ji} should also equal to the $\frac{c_i^*}{m}$. Thus, we will have the relationship follows

$$\frac{e^{-(\alpha_i^*+\beta^*)}}{e^{-(\alpha_i^*+\beta^*)}+1} = \frac{c_i^*}{m}, e^{-(\alpha_i^*+\beta^*)} = \frac{c_i^*}{m-c_i^*}. \tag{5.100}$$

The smallest space to store the information carried by the canonical ensemble described sequences is equal to the Shannon entropy of it as

$$\begin{aligned}
S_{\text{can}}^{(K=n)} =& \ln P_{\text{cam}}(\mathbf{X}^*|\vec{\theta}^*) \\
=& \vec{C}^* \cdot \vec{\theta}^* + \ln Z(\vec{\theta}^*) \\
=& \sum_{i=1}^{n}\sum_{j=1}^{m}[(\alpha_i^*+\beta^*)\frac{e^{-(\alpha_i^*+\beta^*)}}{e^{-(\alpha_i^*+\beta^*)}+1} \\
& + \ln[e^{-(\alpha_i^*+\beta^*)}+1]] \\
=& \sum_{i=1}^{n}\sum_{j=1}^{m}[-\frac{c_i^*}{m}\ln\frac{c_i^*}{m-c_i^*} + \ln\frac{m}{m-c_i^*}] \\
=& \sum_{i=1}^{n}[-\ln\frac{{c_i^*}^{c_i^*}}{(m-c_i^*)^{c_i^*}} + \ln\frac{m^m}{(m-c_i^*)^m}] \\
=& \sum_{i=1}^{n}\ln\frac{m^m}{{c_i^*}^{c_i^*}(m-c_i^*)^{m-c_i^*}},
\end{aligned} \tag{5.101}$$

It is the same as the Shannon entropy of the sequences under soft constraints when there is only column local constraints $\vec{c}^*$.

The microcanonical ensemble description is also equal to the one when there only has column-local constraints. Therefore, the entropy of this microcanonical ensemble description is equal to

$$S_{\text{mic}}^{(K=n)} = \sum_{i=1}^{n}\ln\binom{m}{c_i^*}. \tag{5.102}$$

According to the results in section 5.B, this coupled constrianted information sequence is under ensemble nonequivlaence.

Chapter 6

Conclusions

Statistical physics deals with the description of systems with many interacting microscopic constituents and develops tools to characterize the macroscopic properties emerging in the so-called thermodynamic limit, where the number of constituents goes to infinity. Formally, this is achieved by introducing the concept of statistical ensemble, i.e. a collection of (unobservable) microscopic configurations of the system, each of which is assigned a certain probability that is calculated from certain (observable) macroscopic constraints. Traditionally, the constraints considered in physics are global conserved quantities, such as the total energy and the total number of particles. There are different ways in which the constraints can be enforced, and different resulting ensembles. The two most important possibilities are the microcanonical ensemble, where the constraint is enforced in a hard fashion (i.e. every configuration in the ensemble matches the constraint exactly, as in an isolated system with a fixed total energy), and the canonical ensemble, where the constraint is enforced in a soft fashion (i.e. the constraint is met as an ensemble average, as in a system with fluctuating energy in a thermal bath at fixed temperature). Traditionally, these two ensembles are believed to become asymptotically the same in the thermodynamic limit, a notion that goes under the name of ensemble equivalence. Ensemble equivalence implies that, as the system becomes larger, it does not matter which of the two descriptions is adopted; in particular, the canonical and microcanonical entropies per particle become the same.

However, evidence has accumulated that ensemble equivalence can break down in certain circumstances. The most studied scenario where this can happen is the presence of phase transitions. In this case there are certain phases, or regions in parameter space, where ensemble equivalence breaks down (the canonical and microcanonical entropies per particle are different) and other phases where it is restored. A much more recent scenario for the breakdown of ensemble equivalence is the presence of local constraints in the system, i.e. constraints attached to each of the fundamental units, or at least a finite fraction of them: in other words, the number of constraints is extensive in the size of the system. An example is that of networks with a given number of connections (degree) for each node separately. In this case, it turns out that ensemble equivalence is broken throughout the entire parameter space, as in fact the simplest such models do not even have phase transitions.

In this thesis, we studied novel aspects of the breakdown of ensemble equivalence under local constraints and identified for the first time, to the best of our knowledge, their impact on certain key results in information theory.

In Chapter 1 we set the stage by introducing the main notions and establishing an informal analogy between ensembles in statistical physics and typical sequences in

information theory. Systems with given macroscopic properties in statistical physics have a direct analogue in typical sets in information theory, which are in turn at the basis of several key results, including the maximum compressibility of data and the minimum rate of information transmission to preserve lossless communication.

In Chapter 2 we introduced the first model that studies ensemble nonequivalence in presence of both an extensive number of local constraints and a phase transition. The model is also the first one wherein ensemble nonequivalence is studied on weighted networks, i.e. networks where links can carry different weights. By making the constraints of each node of the network depend on a global temperature-like parameter, we showed that it is possible to induce a form of Bose-Einstein condensation whereby a finite fraction of the total link weights concentrates among a finite number of nodes. We also showed that the traditional criterion for ensemble equivalence, i.e. the vanishing of the relative fluctuations for the constraints in the canonical ensemble, becomes incorrect in the case of local constraints: nonvanishing relative fluctuations capture the onset of the condensation transition, but they do not capture the breakdown of ensemble equivalence. This is quite different from what people have been intuitively reasoning so far about relative fluctuations, and shows that the novel mechanism based on local constraints is quite subtle.

In Chapter 3 we significantly extended the framework of systems with local constraints, from the case of networks with constraints on each node to that of generic systems that can be described by rectangular matrices with constraints on the row and/or column sums. Such constrained matrices can represent a wide range of systems with spatial heterogeneity and/or temporal non-stationarity. We found that, in this more general setting, ensemble nonequivalence can occur in an even stronger way, with the difference between canonical and microcanonical entropies being of the same order as the entropies themselves. Such form of ensemble nonequivalence is as strong as the traditional one encountered in presence of phase transitions, while at the same time maintaining the property of being unrestricted in parameter space, as happens for the other known systems with local constraints. Therefore, it is the most robust form of nonequivalence documented so far. For many specific settings, we calculated explicitly the mathematical quantities distinguishing the two ensembles.

In Chapter 4 me made a major leap from statistical physics to information theory in order to investigate the consequences of ensemble nonequivalence for the compression of modern big data structures such as large networks or long multivariate time series. We first established a rigorous analogy between typical sets, i.e. the collectively most probable outcomes of an information source, and microcanonical ensembles as subsets of the canonical configurations of a physical system. We then showed that, when ensemble equivalence holds, the analogy is actually an identity: the microcanonical configurations coincide with the typical set of the canonical ensemble, implying no difference between sources subject to hard and soft constraints. However, when ensemble equivalence breaks down, the typical set of the canonical ensemble is irreducible to the microcanonical ensemble. In this case, we showed that, for hard constraints, standard information-theoretic results have to be generalized using the calculations of Chapter 3. We revised in particular the traditional

information-theoretic bounds for data compression based on Shannon entropy. We found that microcanonical sources require less storage space but more computational costs, while the opposite is true for canonical sources. This highlights a novel trade-off between memory and computation.

Finally, in Chapter 5 we considered an even more general setting with multiple information sources (such as those generating multivariate time series in finance or neuroscience) where constraints can couple both the outputs of different sources at the same time (spatial constraints) and the outputs of the same source at different times (temporal constraints). We found that temporal constraints never break ensemble equivalence, while spatial constraints do break it if the number of sources is finite. Again, while for canonical sources the standard Shannon theory remains valid, for microcanonical sources various information-theoretic quantities have to be corrected using the calculations of Chapter 3. Moreover, we find that the normalized (per output) total correlation between all sources subject to hard constraints coincides with the normalized (per output) difference between canonical and microcanonical entropies, which in turn only requires the knowledge of the covariance matrix between the canonical constraints and can therefore be calculated explicitly. If spatial constraints are deactivated, ensemble equivalence is restored and the normalized total correlation vanishes, so the microcanonical outputs become asymptotically mutually independent just like the canonical outputs.

Bibliography

[1] JW Gibbs. Elementary principles in statistical physics. *The Collected Works of JW Gibbs (Yale University, New Haven, CT, 1957)*, 2, 1902.

[2] Sergey V Buldyrev, Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.

[3] Stefano Battiston, J Doyne Farmer, Andreas Flache, Diego Garlaschelli, Andrew G Haldane, Hans Heesterbeek, Cars Hommes, Carlo Jaeger, Robert May, and Marten Scheffer. Complexity theory and financial regulation. *Science*, 351(6275):818–819, 2016.

[4] Aniello Lampo, Javier Borge-Holthoefer, Sergio Gómez, and Albert Solé-Ribalta. Multiple abrupt phase transitions in urban transport congestion. *Physical Review Research*, 3(1):013267, 2021.

[5] Tiziano Squartini, Joey de Mol, Frank den Hollander, and Diego Garlaschelli. Breaking of ensemble equivalence in networks. *Physical review letters*, 115(26):268701, 2015.

[6] F den Hollander, M Mandjes, A Roccaverde, and NJ Starreveld. Breaking of ensemble equivalence for perturbed erd\h {o} sr\'enyi random graphs. *arXiv preprint arXiv:1807.07750*, 2018.

[7] Diego Garlaschelli, Frank den Hollander, and Andrea Roccaverde. Covariance structure behind breaking of ensemble equivalence in random graphs. *Journal of Statistical Physics*, 173(3-4):644–662, 2018.

[8] Hugo Touchette. Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels. *Journal of Statistical Physics*, 159(5):987–1016, 2015.

[9] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[10] Ludwig Boltzmann. *Lectures on gas theory*. Courier Corporation, 2012.

[11] James Clerk Maxwell. Ii. illustrations of the dynamical theory of gases. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 20(130):21–37, 1860.

[12] Mark Newman. *Networks*. Oxford university press, 2018.

[13] Alessandro Campa, Thierry Dauxois, and Stefano Ruffo. Statistical mechanics and dynamics of solvable models with long_range interactions. *Physics Reports*, 480(3):57–159, 2009.

[14] Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and Guido Caldarelli. The statistical physics of real-world networks. *Nature Reviews Physics*, 1(1):58, 2019.

[15] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[16] Mark EJ Newman. Analysis of weighted networks. *Physical review E*, 70(5):056131, 2004.

[17] Mark Ed Newman, Albert-László Ed Barabási, and Duncan J Watts. *The structure and dynamics of networks.* Princeton university press, 2006.

[18] Diego Garlaschelli and Maria I Loffredo. Maximum likelihood: Extracting unbiased information from complex networks. *Physical Review E*, 78(1):015101, 2008.

[19] Diego Garlaschelli and Maria I Loffredo. Generalized bose-fermi statistics and structural correlations in weighted networks. *Physical review letters*, 102(3):038701, 2009.

[20] Diego Garlaschelli, Sebastian E Ahnert, Thomas Fink, and Guido Caldarelli. Low-temperature behaviour of social and economic networks. *Entropy*, 15(8):3148–3169, 2013.

[21] Richard S Ellis, Kyle Haven, and Bruce Turkington. Nonequivalent statistical equilibrium ensembles and refined stability theorems for most probable flows. *Nonlinearity*, 15(2):239, 2002.

[22] Richard S Ellis, Kyle Haven, and Bruce Turkington. Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles. *Journal of Statistical Physics*, 101(5-6):999–1064, 2000.

[23] Richard S Ellis, Hugo Touchette, and Bruce Turkington. Thermodynamic versus statistical nonequivalence of ensembles for the mean-field blume–emery–griffiths model. *Physica A: Statistical Mechanics and its Applications*, 335(3-4):518–538, 2004.

[24] Donald Lynden-Bell. Negative specific heat in astronomy, physics and chemistry. *Physica A: Statistical Mechanics and its Applications*, 263(1-4):293–304, 1999.

[25] Pierre-Henri Chavanis. Gravitational instability of isothermal and polytropic spheres. *Astronomy & Astrophysics*, 401(1):15–42, 2003.

[26] Diego Garlaschelli, Frank Den Hollander, and Andrea Roccaverde. Ensemble nonequivalence in random graphs with modular structure. *Journal of Physics A: Mathematical and Theoretical*, 50(1):015001, 2016.

[27] Qi Zhang and Diego Garlaschelli. Strong ensemble nonequivalence in systems with local constraints. *arXiv preprint arXiv:2107.04920*, 2021.

[28] Qi Zhang and Diego Garlaschelli. Ensemble nonequivalence and bose-einstein condensation in weighted networks. *arXiv preprint arXiv:2012.09998*, 2020.

[29] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[30] Abbas El Gamal and Young-Han Kim. *Network information theory*. Cambridge university press, 2011.

[31] Harry Nyquist. Certain factors affecting telegraph speed. *Transactions of the American Institute of Electrical Engineers*, 43:412–422, 1924.

[32] Ralph VL Hartley. Transmission of information. *Bell System technical journal*, 7(3):535–563, 1928.

[33] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[34] Stephen F Altschul, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, 1990.

[35] Ricard Solé. Using information theory to decode network coevolution. *Science*, 368(6497):1315–1316, 2020.

[36] Gema Bello-Orgaz, Jason J Jung, and David Camacho. Social big data: Recent achievements and new challenges. *Information Fusion*, 28:45–59, 2016.

[37] Tiziano Squartini and Diego Garlaschelli. Reconnecting statistical physics and combinatorics beyond ensemble equivalence. *arXiv preprint arXiv:1710.11422*, 2017.

[38] M Blume, VJ Emery, and Robert B Griffiths. Ising model for the λ transition and phase separation in he 3-he 4 mixtures. *Physical review A*, 4(3):1071, 1971.

[39] Julien Barré, David Mukamel, and Stefano Ruffo. Inequivalence of ensembles in a system with long-range interactions. *Physical Review Letters*, 87(3):030601, 2001.

[40] M d'Agostino, F Gulminelli, Ph Chomaz, M Bruno, F Cannata, R Bougault, F Gramegna, I Iori, N Le Neindre, GV Margagliotti, et al. Negative heat capacity in the critical region of nuclear fragmentation: an experimental evidence of the liquid-gas phase transition. *Physics Letters B*, 473(3-4):219–225, 2000.

[41] Julien Barré and Bruno Gonçalves. Ensemble inequivalence in random graphs. *Physica A: Statistical Mechanics and its Applications*, 386(1):212–218, 2007.

[42] Charles Radin and Lorenzo Sadun. Phase transitions in a complex network. *Journal of Physics A: Mathematical and Theoretical*, 46(30):305002, 2013.

[43] Michael Kastner. Nonequivalence of ensembles for long-range quantum spin systems in optical lattices. *Physical review letters*, 104(24):240403, 2010.

[44] Tiziano Squartini and Diego Garlaschelli. *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics*. Springer, 2017.

[45] Andrea Roccaverde. Is breaking of ensemble equivalence monotone in the number of constraints? *Indagationes Mathematicae*, 30(1):7–25, 2019.

[46] Juyong Park and Mark EJ Newman. Statistical mechanics of networks. *Physical Review E*, 70(6):066117, 2004.

[47] Ginestra Bianconi and Albert-László Barabási. Bose-einstein condensation in complex networks. *Physical review letters*, 86(24):5632, 2001.

[48] M Ángeles Serrano, Marián Boguñá, and Romualdo Pastor-Satorras. Correlations in weighted networks. *Physical Review E*, 74(5):055101, 2006.

[49] Tiziano Squartini, Rossana Mastrandrea, and Diego Garlaschelli. Unbiased sampling of network ensembles. *New Journal of Physics*, 17(2):023052, 2015.

[50] https://it.mathworks.com/matlabcentral/fileexchange/46912-max-sam-package-zip.

[51] https://meh.imtlucca.it.

[52] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[53] Albert Einstein. Quantentheorie des einatomigen idealen gases. *Königliche Preußische Akademie der Wissenschaften.*, 261-267, 1924.

[54] Patrick Navez, Dmitri Bitouk, Mariusz Gajda, Zbigniew Idziaszek, and Kazimierz Rzaz ewski. Fourth statistical ensemble for the bose-einstein condensate. *Physical review letters*, 79(10):1789, 1997.

[55] Martin Holthaus, Eva Kalinowski, and Klaus Kirsten. Condensate fluctuations in trapped bose gases: Canonical vs. microcanonical ensemble. *Annals of Physics*, 270(1):198–230, 1998.

[56] WJ Mullin and JP Fernandez. Bose-einstein condensation, fluctuations, and recurrence relations in statistical mechanics. *American Journal of Physics*, 71(7):661–669, 2003.

[57] Sourav Chatterjee and Persi Diaconis. Fluctuations of the bose–einstein condensate. *Journal of Physics A: Mathematical and Theoretical*, 47(8):085201, 2014.

[58] SV Tarasov, Vl V Kocharovsky, and VV Kocharovsky. Grand canonical versus canonical ensemble: Universal structure of statistics and thermodynamics in a critical region of bose-einstein condensation of an ideal gas in arbitrary trap. *Journal of Statistical Physics*, 161(4):942–964, 2015.

[59] A Crisanti, A Sarracino, and M Zannetti. Condensation versus ordering: From the spherical models to bose-einstein condensation in the canonical and grand canonical ensemble. *Physical Review Research*, 1(2):023022, 2019.

[60] C. Kittel and H. Kroemer. *Thermal Physics*. W.H. Freeman, San Francisco, 1980.

[61] Pierfrancesco Dionigi, Diego Garlaschelli, Frank den Hollander, and Michel Mandjes. A spectral signature of breaking of ensemble equivalence for constrained random graphs. https://arxiv.org/abs/2009.05155, 2020.

[62] Assaf Almog and Diego Garlaschelli. Binary versus non-binary information in real time series: empirical results and maximum-entropy matrix models. *New journal of physics*, 16(9):093015, 2014.

[63] Andrea Roccaverde. Is breaking of ensemble equivalence monotone in the number of constraints? *Indagationes Mathematicae*, 30(1):7–25, 2019.

[64] Stefan Semrau, Johanna E Goldmann, Magali Soumillon, Tarjei S Mikkelsen, Rudolf Jaenisch, and Alexander Van Oudenaarden. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nature communications*, 8(1):1096, 2017.

[65] P. Wang, W. He, and J. Zhao. A tale of three social networks: User activity comparisons across facebook, twitter, and foursquare. *IEEE*, 18(2):10–15, 2013.

[66] Assaf Almog, M Renate Buijink, Ori Roethler, Stephan Michel, Johanna H Meijer, Jos HT Rohling, and Diego Garlaschelli. Uncovering functional signature in neural systems via random matrix theory. *PLoS computational biology*, 15(5):e1006934, 2019.

[67] F den Hollander, M Mandjes, A Roccaverde, NJ Starreveld, et al. Ensemble equivalence for dense graphs. *Electronic Journal of Probability*, 23, 2018.

[68] Silvia Heubach and Toufik Mansour. *Combinatorics of compositions and words*. CRC Press, 2009.

[69] A Moujahid, A d'Anjou, FJ Torrealdea, and Francisco Torrealdea. Energy and information in hodgkin-huxley neurons. *Physical Review E*, 83(3):031912, 2011.

[70] Luis E Olmos, Serdar Çolak, Sajjad Shafiei, Meead Saberi, and Marta C González. Macroscopic dynamics and the collapse of urban traffic. *Proceedings of the National Academy of Sciences*, 115(50):12654–12661, 2018.

[71] René Marois and Jason Ivanoff. Capacity limits of information processing in the brain. *Trends in cognitive sciences*, 9(6):296–305, 2005.

[72] Milan Studenỳ and Jirina Vejnarová. The multiinformation function as a tool for measuring stochastic dependence. In *Learning in graphical models*, pages 261–297. Springer, 1998.

Samenvatting

Statistische fysica houdt zich bezig met de beschrijving van systemen die zijn opgebouwd uit heel veel microscopische bestanddelen die met elkaar wisselwerken. Zij ontwikkelt methoden om de macroscopische eigenschappen te karakteriseren die gevonden worden men het aantal bestanddelen naar het oneindige laat gaan, de zogenaamde thermodynamische limiet. De introductie van het z.g. statistisch ensemble, een verzameling van (niet-waarneembare) microscopische configuraties van het systeem, die elk een bepaalde waarschijnlijkheid krijgen toegewezen en die wordt berekend op basis van bepaalde (waarneembare) macroscopische randvoorwaarden, spelt hierbij een grote rol. In de natuurkunde zijn de randvoorwaarden doorgaans grootheden die op globale schaal behouden zijn, zoals de totale energie en het totale aantal deeltjes. Er zijn verschillende manieren om deze randvoorwaarden op te leggen, elk leidend tot een ander ensembles. De twee belangrijkste zijn het microcanonieke en het canonieke ensemble. Bij het microcanonike ensemble wordt de randvoorwaarde op een harde manier afgedwongen (namelijk elke configuratie in het ensemble moet aan de randvoorwaarde voldoen, zoals b.v. in een geïsoleerd systeem met een vaste totale energie). Bij het canonieke ensemble daarentegen, wordt, de randvoorwaarde op een zachte manier afgedwongen (aan de randvoorwaarde wordt gemiddeld voldaan, zoals in een systeem met fluctuerende energie die in contact staat met een warmtebad met vaste temperatuur). In het algemeen wordt aangenomen dat deze twee ensembles in de thermodynamische limiet asymptotisch identiek worden, een begrip dat bekend staat als ensemble equivalentie. Ensemble equivalentie houdt in dat, naarmate het systeem groter wordt, het niet uitmaakt welke van de twee beschrijvingen wordt genomen; in het bijzonder, de canonieke en microcanonieke entropieën per deeltje worden hetzelfde.

Er zijn echter aanwijzingen dat ensemble equivalentie onder bepaalde omstandigheden verdwijnt, bijvoorbeeld in de buurt van faseovergangen. In bepaalde fasen, gebieden in de parameterruimte, is geen ensemble equivalentie (de canonieke en microcanonieke entropieën per deeltje zijn verschillend) en in andere fasen is het wel. Onlangs heeft men zich gerealiseerd dat ensemble equivalentie ook afwezig kan zijn als het systeem lokale randvoorwaarden heeft, d.w.z. randvoorwaarden die aan elk van de fundamentele eenheden zijn verbonden, of ten minste met een eindige fractie ervan. Het aantal randvoorwaarden schaalt dan met de grootte van het systeem. Netwerken met een gegeven aantal verbindingen (graad) voor elk knooppunt afzonderlijk zijn hiervan een voorbeeld. In dat geval blijkt dat ensemble equivalentie nergens in de parameterruimte van toepassing is, omdat zelfs het meest eenvoudige model niet eens een faseovergang vertoont.

In dit proefschrift bestuderen wij nieuwe aspecten van de afwezigheid van ensemble equivalentie bij lokale randvoorwaarden en identificeren wij de impact daarvan op bepaalde sleutelbegrippen in de informatietheorie. In Hoofdstuk 1 leggen wij de basis door de belangrijkste begrippen te introduceren en een informele analogie vast

te stellen tussen ensembles in de statistische fysica en de "typische set" in de informatietheorie. Systemen met gegeven macroscopische eigenschappen in de statistische fysica krijgen hierdoor een analogon in zo'n typische set. Laatstgenoemden zijn van groot belang voor zaken als de maximale compressibiliteit van data en de minimale snelheid van informatieoverdracht om verliesvrije communicatie te behouden.

In Hoofdstuk 2 introduceren wij het eerste model waarin wij ensemble niet-equivalentie bestuderen, met zowel een extensief aantal lokale randvoorwaarden als een faseovergang. Het model is het eerste waarbij ensemble niet-equivalentie wordt bestudeerd op gewogen netwerken, d.w.z. netwerken waar lijnen (verbindingen) verschillende gewichten kunnen dragen. Door de randvoorwaarden van elk knooppunt van het netwerk afhankelijk te maken van een globale, op een temperatuur lijkende parameter, tonen wij aan dat het mogelijk is om een vorm van Bose-Einstein condensatie tot stand te brengen waarbij een eindige fractie van de totale lijngewichten zich concentreert tussen een eindig aantal knooppunten. Ook laten wij zien dat het traditionele criterium voor ensemble equivalentie, n.l. het verdwijnen van de relatieve fluctuaties van de randvoorwaarden in het canonieke ensemble, in het geval van lokale randvoorwaarden niet meer van toepassing is. De relatieve fluctuaties zijn karakteristiek voor het begin van de condensatieovergang, maar niet voor het verbreken van ensemble equivalentie. Dit is echt nieuw, en laat zien dat het nieuwe mechanisme op basis van lokale randvoorwaarden vrij subtiel is.

In Hoofdstuk 3 breiden wij het raamwerk van systemen met lokale randvoorwaarden aanzienlijk uit, van het geval van netwerken met randvoorwaarden bij elk knooppunt tot dat van generieke systemen die kunnen worden beschreven door rechthoekige matrices met randvoorwaarden die gelden voor de rij- en/of kolomsommen. Dergelijke matrices kunnen een breed scala aan systemen vertegenwoordigen met ruimtelijke heterogeniteit en/of temporele niet-stationariteit. In deze meer algemene setting, blijkt ensemble niet-equivalentie op een nog sterkere manier op te kunnen treden, waarbij het verschil tussen canonieke en microcanonieke entropieën van dezelfde orde is als de entropieën zelf. Deze vorm van ensemble niet-equivalentie is even sterk als die wordt aangetroffen bij de aanwezigheid van faseovergangen, maar behoudt tegelijkertijd de eigenschap dat zij onbeperkt is in de parameterruimte, zoals het geval is bij de andere bekende systemen met lokale, randvoorwaarden. Het is dan ook de meest robuuste vorm van niet-equivalentie die tot dusver is gevonden. Voor een groot aantal specifieke instellingen hebben wij expliciet de wiskundige grootheden berekend die de twee ensembles van elkaar onderscheiden.

In Hoofdstuk 4 maak ik een grote sprong van de statistische natuurkunde naar de informatietheorie om de gevolgen van ensemble-ongelijkwaardigheid te onderzoeken voor de compressie van moderne grote gegevensstructuren zoals grote netwerken of lange multivariate tijdreeksen. Eerst stellen wij een rigoureuze analogie vast tussen "typische sets", d.w.z. het collectief meest waarschijnlijke uitkomsten van een informatiebron, en microcanonieke ensembles als deelverzamelingen van de canonieke configuraties van een fysisch systeem. Vervolgens tonen wij aan dat, wanneer ensemble-equivalentie geldt, de analogie in feite een identiteit is: de microcanonieke configuraties vallen samen met de "typische sets" van het canonieke ensemble, wat in-

houdt dat er geen verschil is tussen bronnen die onderworpen zijn aan harde en zachte randvoorwaarden. Wanneer de ensemble-equivalentie echter wegvalt, is de "typische set" van het canonieke ensemble onherleidbaar tot het microcanonieke ensemble. In dit geval laten wij zien dat, voor harde randvoorwaarden, de standaard informatietheoretische resultaten met behulp van de methoden van hoofdstuk 3 moeten worden veralgemeniseerd. Wij hebben in het bijzonder de traditionele informatietheoretische grenzen voor gegevenscompressie op basis van Shannon-entropie herzien. Wij zien dat microcanonieke bronnen minder opslagruimte vergen maar meer rekenkracht, terwijl het omgekeerde geldt voor canonieke bronnen. Dit wijst op een nieuwe wisselwerking en afweging tussen geheugen en rekenkracht.

Ten slotte bestuderen wij in Hoofdstuk 5 een nog algemenere setting. Wij kijken naar systemen met meerdere informatiebronnen, zoals die welke multivariate tijdreeksen genereren (vaak gezien in de financiële wereld of in de neurowetenschappen) waar randvoorwaarden zowel de outputs van verschillende bronnen tegelijkertijd kunnen koppelen (ruimtelijke randvoorwaarden) als de outputs van dezelfde bron op verschillende tijdstippen (temporele randvoorwaarden). Wij vinden dat temporele randvoorwaarden nooit ensemble equivalentie doorbreken, terwijl ruimtelijke randvoorwaarden wel doen als het aantal bronnen eindig is. Nogmaals, terwijl voor canonieke bronnen de standaard Shannon theorie geldig blijft, moeten voor microcanonieke bronnen verschillende informatietheoretische grootheden worden gecorrigeerd met behulp van de berekeningen van Hoofdstuk 3. Bovendien vinden wij dat de (per output) genormaliseerde totale correlatie tussen alle bronnen die onderhevig zijn aan harde randvoorwaarden samenvalt met het (per output) genormaliseerde verschil tussen canonieke en microcanonieke entropieën. Dit vereist, uitsluitend de kennis van de covariantiematrix tussen de canonieke randvoorwaarden en kan daarom expliciet worden berekend. Als ruimtelijke randvoorwaarden worden opgeheven, wordt ensemble equivalentie hersteld en verdwijnt de genormaliseerde totale correlatie, dus de microcanonieke outputs wordt asymptotisch wederzijds onafhankelijk, net als de canonieke outputs.

List of Publications

1. Qi Zhang, Diego Garlaschelli, "Information theory with coupled source under ensemble nonequivalence", *In preparation* (2021).

2. Qi Zhang, Diego Garlaschelli, "New information-theoretic bounds for systems with local constraints", *In preparation* (2021).

3. Qi Zhang, Diego Garlaschelli, "Strong ensemble nonequivalence in systems with local constraints", arXiv:2107.04920 (2021).

4. Qi Zhang, Diego Garlaschelli, "Ensemble nonequivalence and BEC in weighted networks", arXiv:2012.09998 (2020).

5. Qi Zhang, Meizhu Li, Yong Deng, "Measuring the structural similarity of nodes in complex networks based on relative entropy", *Physica A*, 491 (2017): 749-763 .

6. Qi Zhang, Meizhu Li, Yong Deng, "A new structural entropy of complex networks based on Tsallis nonextensive statistical mechanics", *International Journal of Modern Physics C*, 10 (2016) 1650118.

7. Qi Zhang, Chuanhai Luo, Meizhu Li, Yong Deng, Sankaran Mahadevan, "Tsallis information dimension of complex networks", *Physica A*, 419 (2015): 707-717.

8. Liguo Fei, Qi Zhang, Yong Deng, "Identifying influential nodes in complex networks based on the inverse-square law", *Physica A*, 512 (2018): 1044-1059.

9. Meizhu Li, Qi Zhang, Yong Deng, "Evidential identification of influential nodes in network of networks", *Chaos, Solitons & Fractals*, 117 (2018): 283-296.

10. Xinyang Deng, Qi Zhang, Yong Deng, Zhen Wang, "A novel framework of classical and quantum prisoner's dilemma games on coupled networks", *Scientific reports*, 6 (2016): 23024.

11. Meizhu Li, Yong Hu, Qi Zhang, Yong Deng, "A novel distance function of D numbers and its application in product engineering", *Engineering Applications of Artificial Intelligence*, 47 (2016): 61-67.

12. Meizhu Li, Xi Lu, Qi Zhang, Yong Deng, "Multiscale probability transformation of basic probability assignment", *Mathematical Problems in Engineering*, (2014), 319264.

Curriculum Vitae

I was born in 1989 in Sichuan, a province in Southwest China. I received my early education in this province, including primary, middle and high schools.

In 2009, I left my hometown and went to Chongqing for my undergraduate studies at Southwest University. At that time, I was trained as an automation engineer. I continued my master study in the Department of Computer Science at the same university but with a focus on complex networks under the supervision of Prof Yong Deng in 2014.

In 2016, I went to Leiden, the Netherlands, to start my PhD work under the supervision of Dr Diego Garlaschelli at the Lorentz Institute for Theoretical Physics at Leiden University. I spent five years researching systems with local constraints from the viewpoint of statistical physics and information theory.

Acknowledgments

My Phd in Leiden is coming to an end, and I can not imagine what my life and research would be like without the support of so many people.

First, I would like to express my sincere gratitude to my supervisor Diego Garlaschelli. To me, what he has done in the past years is much more than mentoring a PhD student on specific projects. His patience and sensitivity showed me how to figure out physically important problems and develop exciting ideas with scientific rigour. Through his guidance, motivation, and training, I was transformed from being an engineer to become a physicist. Psychologically, his help always gave me the power to face the problems in my life and research. He always told me "Qi, don't worry, we will figure it out". The knowledge I have been learned from him taught me how to become a good physicist and a better person. It great honour to have him as my supervisor. I would also like to thank Prof. Eric Eliel for being my PhD promotor and kindly helping me go through all the necessary administrative procedures.

I also want to thank Tiziano for the discussions in Lucca when I tried to extend statistical ensembles with local constraints into information theory. I want to thank Alexandru and Hamed for their suggestions and comments on this thesis. I also want to thank Eric, Joseph and Peter for their help with my Dutch summary.

I would also like to thank other members of Diego's group: Assaf, Valerio, Elena, Andrea, Janusz, Maria, Pierfrancesco and Federica. I felt so lucky to be part of this wonderful team. I enjoyed all the discussions, conferences, dinners and outings we had together. I am indebted to the secretaries at the institute–Fran, Manon, and Marianne. Their assistance has made my work and life much easier. I would like to thank my friends and colleagues in the Physics and Astronomy institutes of Leiden: Ireth, Bahareh, Zhihong, Yujie, Weichun, Donggang, Xuxing, Zujia, Jiangnan, Junxiang, Xuechen, and Yuhan. Thanks for your help both in research and life.

In addition, I would like to thank my Chinese friends in Leiden: Enchen Zhou, Xiaoyu Wei, Tian Qin, Yang Deng, Hui Wang, Wei Yang, Jing Hu, Jing Wei, Ying Wang, Zhengzheng Zhang, Mengjie Shen and many others whose names are not listed here.

I want to thank my parents. Their unconditional support is always the solid backing in my life and work.

At last, I want to thank my dearest Meizhu. Her presence and love in the past ten years are the major source of encouragement for me.