# Breaking of ensemble equivalence for complex networks

Roccaverde, A.

# Breaking of Ensemble Equivalence for Complex Networks

Andrea Roccaverde

# Breaking of Ensemble Equivalence for Complex Networks

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 5 december 2018
klokke 15.00 uur

door

Andrea Roccaverde
geboren te Modena
in 1990

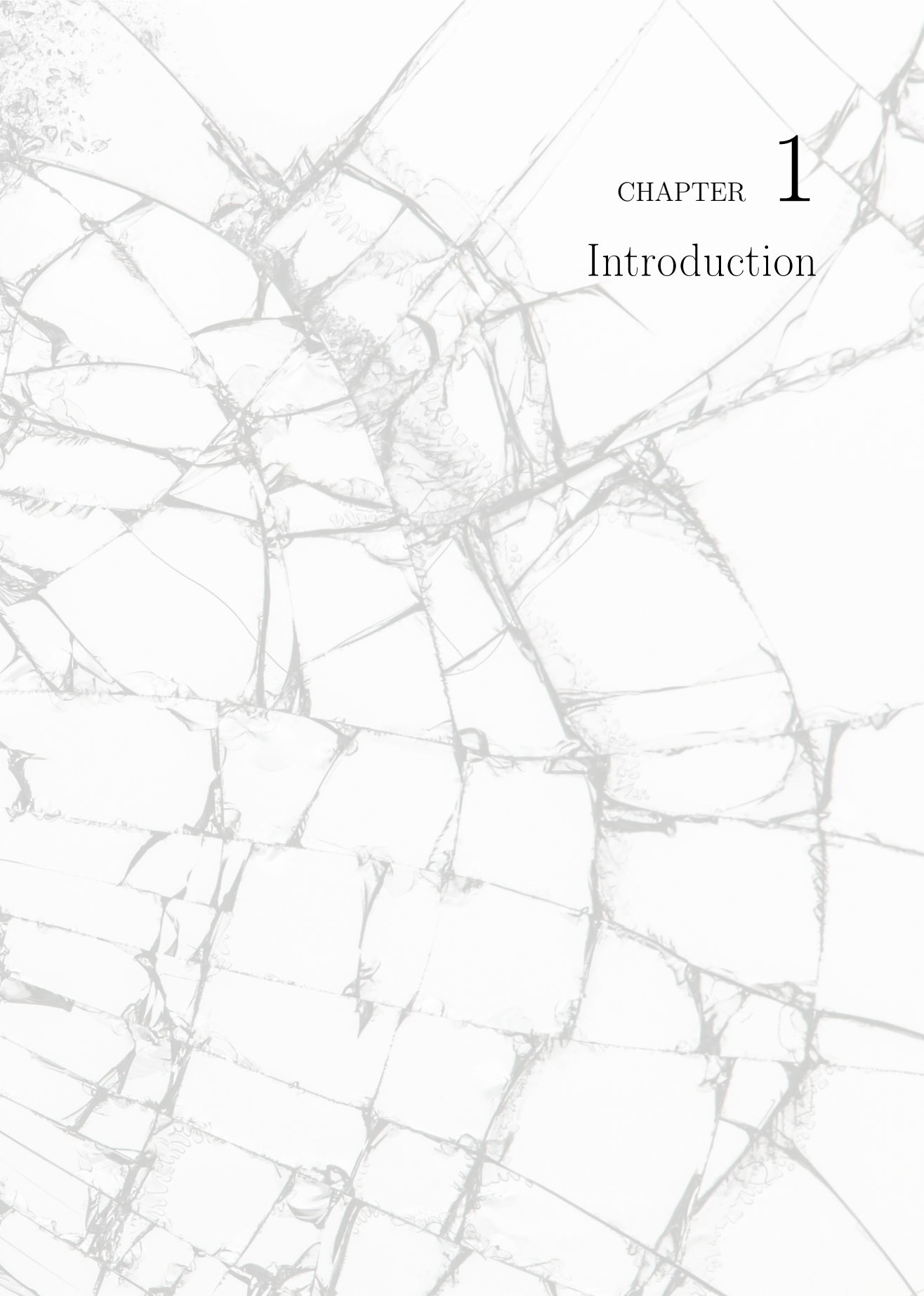
Samenstelling van de promotiecommissie:

**1$^e$ Promotor:**
Prof. dr. W. Th. F. den Hollander (Universiteit Leiden)

**2$^e$ Promotor:**
Dr. D. Garlaschelli (Universiteit Leiden)

**Overige Leden:**
Prof. dr. A. Doelman (Universiteit Leiden, secretary)
Prof. dr. R.W. van der Hofstad (Universiteit Eindhoven)
Dr. T. Squartini (IMT Institute for Advanced Studies in Lucca)
Prof. dr. H. Touchette (Stellenbosch University)

# Contents



CHAPTER 1

# Introduction

## §1.1 Gibbs Ensembles

In order to provide an introduction to Gibbs Ensembles we borrow from Garlaschelli, den Hollander, Roccaverde [49].

Statistical physics aims at describing collective behavior in systems consisting of a very large number of interacting particles (= atoms or molecules). This is a daunting task: a glass of water or a piece of iron can easily contain $10^{23}$ particles. Still, the hope is that the macroscopic properties of these particles as a whole can be explained from the microscopic interactions between them. For instance, we want to explain *why* water turns into ice (or vapor) at an appropriate temperature and *how* this transition exactly takes place. Similarly, we want to explain why a piece of iron at low temperature becomes magnetized when it is moved close to a magnet and remains magnetized after it is moved away from the magnet. We also want to understand why this does not happen when the temperature is high. For most physical systems the trajectories of the particles are so chaotic that they cannot be captured by explicit formulas. A full description would require knowledge of the positions and the speeds of all the particles at all times, which clearly is hopeless. Yet, we need large numbers of particles to explain collective phenomena: a single water molecule cannot transit from water to ice (or vapor).

The way out of this dilemma is offered by statistical physics: for most purposes a full microscopic description is not necessary: it suffices to have a macroscopic description in terms of a small number of relevant quantities, such as pressure, density and temperature. In statistical physics, the system is assumed to be a random sample, drawn from a set of allowed microscopic configurations that are consistent with a set of given macroscopic constraints. These constraints determine the pressure, density and temperature of the system. Collective phenomena, such as whether the system is a solid (ice), a liquid (water) or a gas (vapor), should follow from a combination of the microscopic interactions and the macroscopic constraints.

A physical system is rarely isolated. Typically, it is part of a larger system, that in turn is part of an even larger system, etc. For example, the molecules in a glass of water depend on what is outside the glass. The molecules at the top of the glass interact with the air above it. This air also contains water molecules, and a lively exchange takes place close to the surface of the water. We are thus tempted to believe that, in order to understand what happens inside the glass of water, we need to model all the molecules around the glass as well, and perhaps even all the molecules in the room the glass finds itself in. Fortunately, this is not the case, since the water molecules can only interact over short distances.

Statistical physics deals with the definition of the appropriate probability distribution over the set of allowed microscopic configuration, such as the locations and the speeds of the particles and how they bounce off each other and the wall of the container. These distributions need to take the macroscopic constraints into account. For instance, when the temperature is high the particles move quickly, which should be reflected in the choice of the probability distribution for the positions and the speeds. These probability distributions are called, in statistical physics, *ensembles*. They were introduced for the first time by Boltzmann [19] and then reformulated

in their modern probabilistic form by Gibbs [54]. Each ensemble describes how the system interacts with its surroundings and therefore represents a particular physical situation.

1 The *microcanonical ensemble*, where hard constraints are placed on both the energy and the number of particles: both are set to fixed values and are not allowed to vary.

2 The *canonical ensemble*, where a soft constraint is placed on the energy of the particles (in the sense that it may vary but with a fixed average), while a hard constraint is placed on the number of particles.

3 The *grandcanonical ensemble*, where both the energy and the number of particles are soft.

For systems of finite size, the three ensembles lead to different behavior. Therefore, in practical situations, the choice of ensemble is important and must be based on the physical situation that is described. In particular, an experimental physicist would use the microcanonical ensemble to model an isolated system (= a system that exchanges neither heat nor particles with its surroundings), the canonical ensemble to model a closed system (= a system that exchanges heat with an "external reservoir", with which it is in thermal equilibrium, but no particles), and the grandcanonical ensemble to model an open system (= a system that exchanges both heat and particles with the external reservoir, with which it is in thermal and chemical equilibrium). Choosing the wrong ensemble amounts to choosing the wrong microscopic probability distribution on which the computation of macroscopic quantities is based. For instance, if the experimental physicist is certain that the system under study does not exchange particles with its surroundings, then the grandcanonical ensemble is clearly not the right choice, and it would make the microscopic description of the system more noisy than is necessary.

## §1.2 Equivalence of Ensembles

Statistical physics also deals with the problem of determining whether the ensembles give the same predictions when the system is very large. Traditionally, in physics books the three ensembles are assumed to be *thermodynamically equivalent*: for large systems fluctuations of macroscopic quantities around their average value are expected to be small and to be asymptotically vanishing as the number of particles tends to infinity. In the latter limit, called the thermodynamic limit, the soft constraints effectively become hard constraints. The assumption of ensemble equivalence dates back to Gibbs [54] and has been verified for traditional models of physical systems with short-range microscopic interactions and subject to a small number of macroscopic constraints. However, ensemble equivalence is not a simple concept. It has to be defined and studied carefully: depending on the level of description considered it can take different forms. We will talk about this in more detail in Section 1.3.

The general idea is that ensemble equivalence is convenient because it allows us to choose any of the three ensembles to work with. Soft constraints often are computationally easier to work with than hard constraints, which makes the choice of the canonical ensemble and the grandcanonical ensemble more convenient than that of the microcanonical ensemble. If ensemble equivalence holds and the system is large enough, then all three ensembles lead to the same macroscopic outcome for most of the relevant quantities. However, ensemble equivalence does not hold in general. This fact is important because, in such a situation, an experimental physicist must make a careful choice what ensemble to use for modeling the system, even when the system is large. A wrong choice means a wrong answer to macroscopic questions. Despite the fact that many textbooks still convey the message that ensemble equivalence holds for all physical systems, over the last decades various examples of physical systems have been found for which it breaks down ([85], [86], [87] and [29]).

Thus, breaking of ensemble equivalence means that different choices of ensemble lead to asymptotically different behavior. Consequently, while for applications based on ensemble equivalent models the choice of the working ensemble can be arbitrary and can be based on mathematical convenience, for those based on nonequivalent models the choice must be dictated by a criterion indicating which ensemble is the appropriate one to use. This criterion must be based on the *a priori* knowledge that is available about the network, i.e., which form of constraint (hard or soft) applies in practice.

## §1.3 Definition of Ensemble Equivalence

In this section we give a brief introduction to the problem of ensemble equivalence. In his treatise [54], Gibbs argued that, in the so-called thermodynamic limit (when the number of particles goes to infinity), the microcanonical and the canonical ensemble become equivalent. Gibbs's argument was that, when the system is large, the fluctuations of the energy, in the canonical ensemble, become negligible with respect to the total energy. The canonical ensemble therefore essentially chooses a unique value of the energy, equal to the energy used to define the microcanonical ensemble. In this sense the use of the canonical ensemble, instead of the more complicated microcanonical ensemble, is justified and the ensembles are said to be equivalent. In other words, the equilibrium properties of the system can be described by using either the energy or the temperature as parameters. This equivalence can be proved in simple cases, for example, an ideal gas or non-interacting systems.

Many other complex systems have been studied with the help of statistical ensembles, which brought physicists to assume that the two ensembles are always equivalent ([89, 59, 6, 68, 91]). However more recently systems have been found where breaking of ensemble equivalence occurs. These examples include models of fluid turbulence [41], star formation [72] and networks [7, 85, 95].

The problem of ensemble (non)equivalence has been formulated in a more rigorous manner. Ellis, Haven and Turkington [40] studied two types of (non)equivalence. The first type is equivalence at the *thermodynamic level*, that has been studied the most so far. The second type is equivalence at the *macrostate level*, introduced in [40].

Another type of equivalence is the equivalence the *measure level*. Following [97], we present the problem of ensemble equivalence in these three different forms:

- *Thermodynamic equivalence*: the microcanonical and the canonical ensemble are said to be thermodynamically equivalent when the entropy (as a function of the energy) and the free energy (as a function of the temperature) are one-to-one related by a Legendre transform.
- *Macrostate equivalence*: the microcanonical and the canonical ensemble are said to be macrostate equivalent when the equilibrium values of the macrostate predicted by the microcanonical ensemble and the equilibrium values of the macrostate predicted by the canonical ensemble are the same.
- *Measure equivalence*: the microcanonical and the canonical ensemble are said to be measure equivalent when the Gibbs distribution defining the canonical ensemble at the microstate level converges to the distribution defined by Boltzmann's equiprobability postulate defining the microcanonical ensemble.

Touchette [97] proves that thermodynamic nonequivalence occurs whenever the microcanonical entropy function has one or more points of non-concavity and that macrostate and the thermodynamic equivalence are essentially equivalent. Measure equivalence is also proved to be equivalent to the other two types of equivalence. The main conclusion is that the three 'different' levels of ensemble equivalence are equivalent, whenever the setting is a general particle system, under the assumption that thermodynamic functions and equilibrium macrostates exists and are defined through large deviation principles (see [97] for more details).

The physical reason behind ensemble nonequivalence still remains to be clarified and part of this thesis is to understand some of the hidden mechanisms behind it. The type of equivalence considered throughout the thesis is that at the measure level. In the following we give the precise definition of the latter.

## §1.3.1 Measure equivalence

Equivalence at the measure level concerns convergence of the canonical ensemble to the microcanonical ensemble at the microscopic level. In this section we give the mathematical definition of this type of equivalence and explain the idea behind it.

### Relative entropy

$P$ and $Q$ are two discrete probability measures defined on the same space $\mathcal{X}$, with $P$ absolutely continuous with respect to $Q$ ($P \ll Q$). The *relative entropy* of $P$ with respect to $Q$ is defined as

$$S(P|Q) = \sum_{i \in \mathcal{X}} P(i) \ln \frac{P(i)}{Q(i)}. \tag{1.1}$$

The relative entropy $S(P|Q)$ is not a distance (it is not symmetric and does not satisfy the triangle inequality). However, $S(P|Q)$ is non-negative and equals zero if and only

if $P = Q$ almost everywhere. Moreover, Pinsker's inequality shows that $S(P|Q)$ is an upper bound on the total variation distance, namely,

$$d_{TV}(P,Q) = \frac{1}{2}\sum_{i\in\mathcal{X}} |P(i) - Q(i)| \leq \sqrt{S(P|Q)}. \tag{1.2}$$

In the case of the microcanonical and the canonical ensemble, defined for an $N$-particle system (i.e., see equations (1.6) and (1.9)), we get $P^N_{\text{mic}} \ll P^N_{\text{can}}$ (but not vice versa). Therefore, the relative entropy of the microcanonical ensemble with respect to the canonical ensemble can be computed and takes the form

$$S(P^N_{\text{mic}}|P^N_{\text{can}}) = \sum_{i\in\mathcal{X}} P^N_{\text{mic}}(i) \ln \frac{P^N_{\text{mic}}(i)}{P^N_{\text{can}}(i)}. \tag{1.3}$$

The *specific relative entropy* is defined as the limit

$$s_\infty = \lim_{N\to\infty} \frac{1}{N} S(P^N_{\text{mic}}|P^N_{\text{can}}). \tag{1.4}$$

**1.3.1 Definition (Measure equivalence).** The microcanonical and the canonical ensemble are said to be equivalent at the measure level if

$$s_\infty = 0.$$

The immediate implication of ensemble equivalence (at the measure level) combined with Pinsker's inequality in (1.2) is that the total variation $d_{TV}(P^N_{\text{mic}}, P^N_{\text{can}})$ grows slower than $\sqrt{N}$ as $N \to \infty$.

## §1.4 Statistical Ensembles for Complex Networks

Ensemble (non)equivalence is usually studied for systems in which the Boltzmann distribution describes a certain physical interaction that is encapsulated in the energy. However, as already shown by Jaynes [61], the Boltzmann distribution describes much more general ensembles of systems with given constraints, namely, *all solutions to the maximum-entropy problem of inference from partial information.* In what follows we argue that, for any discrete enumeration problem where we need to count microcanonical configurations compatible with a given constraint, there exists a 'dual' problem involving canonical configurations induced by the same constraint. We define microcanonical and canonical ensembles for *complex networks*, and provide examples of networks that exhibit equivalence and nonequivalence of the ensembles at the measure level, introduced in Definition 1.3.1. The statistical mechanics approach turns out to be very powerful in the study of real-world networks, for which a detailed knowledge of the architecture is typically not available. The way proposed here is to study the complex networks through a probabilistic description, i.e., statistical ensembles. To that end the network is assumed to be a random sample drawn from a set of allowed configurations that are consistent with a set of known *topological constraints* [95]. In the following we give a rigorous definition of these statistical ensembles for complex networks.

### §1.4.1 Microcanonical and Canonical Ensemble for Complex Networks

In Section 1.1 we explained why statistical physics deals with the definition of the appropriate probability distribution and what are the possible effects this has on the experiments. Here we consider two of the key choices of probability distribution, namely:

(1) The *microcanonical ensemble*, where the constraints are *hard* (i.e., are satisfied by each individual configuration).

(2) The *canonical ensemble*, where the constraints are *soft* (i.e., hold as ensemble averages, while individual configurations may violate the constraints).

(In both ensembles, the entropy is *maximal* subject to the given constraints.) We start by giving the rigorous definitions of the microcanonical and the canonical ensemble for complex networks.

For $n \in \mathbb{N}$, let $\mathcal{G}_n$ denote the set of all simple undirected graphs with $n$ nodes. Any graph $G \in \mathcal{G}_n$ can be represented as an $n \times n$ matrix with elements

$$g_{ij}(G) = \begin{cases} 1 & \text{if there is a link between node } i \text{ and node } j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.5}$$

Let $\vec{C}$ denote a vector-valued function on $\mathcal{G}_n$. Given is a specific value $\vec{C}^*$, which we assume to be *graphical*, i.e., realisable by at least one graph in $\mathcal{G}_n$.
The *microcanonical probability distribution* on $\mathcal{G}_n$ with *hard constraint* $\vec{C}^*$ is defined as

$$P_{\text{mic}}(G) = \begin{cases} \Omega_{\vec{C}^*}^{-1}, & \text{if } \vec{C}(G) = \vec{C}^*, \\ 0, & \text{else,} \end{cases} \tag{1.6}$$

where

$$\Omega_{\vec{C}^*} = |\{G \in \mathcal{G}_n \colon \vec{C}(G) = \vec{C}^*\}| \tag{1.7}$$

is the number of graphs that realise $\vec{C}^*$. The *canonical probability distribution* $P_{\text{can}}(G)$ on $\mathcal{G}_n$ is defined as the solution of the maximisation of the *entropy*

$$S_n(P_{\text{can}}) = -\sum_{G \in \mathcal{G}_n} P_{\text{can}}(G) \ln P_{\text{can}}(G) \tag{1.8}$$

subject to the normalisation condition $\sum_{G \in \mathcal{G}_n} P_{\text{can}}(G) = 1$ and to the *soft constraint* $\langle \vec{C} \rangle = \vec{C}^*$, where $\langle \cdot \rangle$ denotes the average w.r.t. $P_{\text{can}}$. This gives the formula (see [61])

$$P_{\text{can}}(G) = \frac{\exp[-H(G, \vec{\theta}^*)]}{Z(\vec{\theta}^*)}, \tag{1.9}$$

where

$$H(G, \vec{\theta}) = \vec{\theta} \cdot \vec{C}(G) \tag{1.10}$$

is the *Hamiltonian* and

$$Z(\vec{\theta}) = \sum_{G \in \mathcal{G}_n} \exp[-H(G, \vec{\theta})] \tag{1.11}$$

is the *partition function*. In (1.9) the parameter $\vec{\theta}$ must be set equal to the particular value $\vec{\theta}^*$ that realises $\langle \vec{C} \rangle = \vec{C}^*$. This value is unique and maximises the likelihood of the model given the data (see [51]). We next proceed with the equivalence definition of the ensembles.

## §1.4.2 $\alpha_n$-Equivalence of Ensembles

In order to define the equivalence at the measure level for complex networks, we follow [92, 48, 50] and define the *relative entropy* of $P_{\text{mic}}$ w.r.t. $P_{\text{can}}$ as

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) = \sum_{G \in \mathcal{G}_n} P_{\text{mic}}(G) \log \frac{P_{\text{mic}}(G)}{P_{\text{can}}(G)}, \tag{1.12}$$

and the $\alpha_n$-*relative entropy* as [50]

$$s_{\alpha_n} = {\alpha_n}^{-1} S_n(P_{\text{mic}} \mid P_{\text{can}}), \tag{1.13}$$

where $\alpha_n$ is a *scale parameter*. The limit of the relative entropy $\alpha_n$-density is defined as

$$s_{\alpha_\infty} \equiv \lim_{n \to \infty} s_{\alpha_n} = \lim_{n \to \infty} {\alpha_n}^{-1} S_n(P_{\text{mic}} \mid P_{\text{can}}) \in [0, \infty]. \tag{1.14}$$

We say that the microcanonical and the canonical ensemble are equivalent on scale $\alpha_n$ if and only if

$$s_{\alpha_\infty} = 0. \tag{1.15}$$

This is a generalization of the standard measure equivalence definition given in Section 1.3.1. In fact, for complex networks, a specific parameter corresponding to the number of particles (the volume of the system) does not exist. For example, we could decide to use the number of nodes $n$ as the parameter representing the volume of the system. On the other hand, we could also use the number of edges $\binom{n}{2}$, or even other $n$ dependent factors instead. It becomes clear that computing the *specific relative entropy*, i.e., dividing by the number $n$ of particles, does not have a precise meaning in the context of complex networks. This is the main reason why we choose to be general and use a parameter $\alpha_n$. The choice of the scale $\alpha_n$ at which we check for (non)equivalence in this thesis is flexible and depends on the number of nodes $n$, on the constraint at hand and on its value as well. Indeed, we consider different choices of $\alpha_n$ for different models. In certain cases, we in fact prefer to reverse the point of view and look for the 'natural' or 'critical' scale $\alpha_n$ at which $s_{\alpha_\infty}$ is positive and finite. This second approach allows us to immediately conclude that the ensembles are $\beta_n$-equivalent for all $\beta_n = \omega(\alpha_n)$ and nonequivalent when $\beta_n = \Omega(\alpha_n)$. For instance, if the constraint is on the *degree sequence*, then in the *sparse regime* the critical scale turns out to be $\alpha_n = n$ [92], [48] (in which case $s_{\alpha_\infty}$ is the specific relative entropy 'per vertex'), while

in the *dense regime* it turns out to be $\alpha_n = n \log n$ [50]. For more details, see Section 1.5 for the sparse regime and Section 1.6 for the dense regime. On the other hand, if the constraint is on the *total numbers of edges and triangles*, with values different from what is typical for the Erdős-Renyi random graph in the dense regime, then the critical scale turns out to be $\alpha_n = n^2$ [38] (in which case $s_{\alpha_\infty}$ is the specific relative entropy 'per edge'). This is discussed in more detail in Section 1.8.

Before considering specific cases, we recall an important observation made in [92]. The definition of $H(G, \vec{\theta})$ ensures that, for any $G_1, G_2 \in \mathcal{G}_n$, $P_{\rm can}(G_1) = P_{\rm can}(G_2)$ whenever $\vec{C}(G_1) = \vec{C}(G_2)$ (i.e., the canonical probability is the same for all graphs having the same value of the constraint). We may therefore rewrite (1.12) as

$$S_n(P_{\rm mic} \mid P_{\rm can}) = \log \frac{P_{\rm mic}(G^*)}{P_{\rm can}(G^*)}, \tag{1.16}$$

where $G^*$ is *any* graph in $\mathcal{G}_n$ such that $\vec{C}(G^*) = \vec{C}^*$ (recall that we have assumed that $\vec{C}^*$ is realisable by at least one graph in $\mathcal{G}_n$). The definition in (1.14) then becomes

$$s_{\alpha_\infty} = \lim_{n\to\infty} {\alpha_n}^{-1} \big[ \log P_{\rm mic}(G^*) - \log P_{\rm can}(G^*) \big], \tag{1.17}$$

which shows that breaking of ensemble equivalence coincides with $P_{\rm mic}(G^*)$ and $P_{\rm can}(G^*)$ having different large deviation behavior on scale $\alpha_n$. (This is perfectly in line with what was discussed in Section 1.3). Note that (1.17) involves the microcanonical and canonical probabilities of a *single* configuration $G^*$ realising the hard constraint. Apart from its theoretical importance, this fact greatly simplifies mathematical calculations.

To analyse breaking of ensemble equivalence, ideally we would like to be able to identify an underlying *large deviation principle* on a natural scale $\alpha_n$. This is generally difficult, and so far has only been achieved in the dense regime with the help of *graphons*. See [38] and Section 1.8 to understand why.

## §1.5 Summary of Chapter 2

While there is consensus that nonequivalence occurs when the microcanonical specific entropy is non-concave as a function of the energy density in the thermodynamic limit, the classification of the physical mechanisms at the *origin* of nonequivalence is still open. A possible and natural mechanism is the presence of *long-range interactions*. Similarly, phase transitions are naturally associated with long-range order. These "standard mechanisms" for ensemble nonequivalence have been documented also in the study of random graphs.

In Chapter 2 we study certain classes of unipartite networks [92], and show that ensemble nonequivalence can manifest itself via an additional, novel mechanism, unrelated to non-additivity or phase transitions: namely, the presence of an *extensive* number of local topological constraints, i.e., the degrees and/or the strengths (for weighted graphs) of all nodes.[1] This finding explains previously documented signatures of nonequivalence in random graphs with local constraints, such as a finite

[1] While in binary (i.e., simple) graphs the *degree* of a node is defined as the number of edges

difference between the microcanonical and the canonical entropy densities [1] and the non-vanishing of the relative fluctuations of the constraints [95]. How generally this result holds beyond the specific uni-partite and bi-partite cases considered so far remains an open question. By considering a much more general class of random graphs with a variable number of constraints, we confirm that the presence of an extensive number of local topological constraints breaks ensemble equivalence, even in the absence of phase transitions or non-additivity.

We start from the characterization of nonequivalence in the simple cases of uni-partite and bi-partite graphs already explored in [92], and subsequently move on to a very general class of graphs with an arbitrary multilayer structure and tunable intra-layer and inter-layer connectivity. The main theorems proved, which (mostly) concern the *sparse regime*, not only characterize nonequivalence *qualitatively*, they also provide a *quantitative* formula for the specific relative entropy. We discuss various important implications of our results, describing properties that are fully general, but also focusing on several special cases of empirical relevance. In addition, we provide an interpretation of the specific relative entropy formula in terms of Poissonisation of the degrees. We also discuss the implications of our results for the study of several empirically relevant classes of "modular" networks that have recently attracted interest in the literature, such as networks with a so-called multi-partite, multiplex [16], time-varying [58], block-model [57], [62] or community structure [43], [84].

## §1.6 Summary of Chapter 3

In Chapter 3 we take a fresh look at breaking of ensemble equivalence by analyzing a formula for the relative entropy, based on the *covariance structure* of the canonical ensemble, recently put forward by Squartini and Garlaschelli [93]. We consider the case of a random graph with a given degree sequence (configuration model) and show that the formula correctly predicts that the specific relative entropy is determined by the scaling of the determinant of the covariance matrix of the constraints in the so called $\delta$-tame dense regime, while it requires an extra correction term in the sparse regime and the ultra-dense regime. We also show that the different behaviors found in the different regimes correspond to the degrees being asymptotically Gauss in the dense regime and asymptotically Poisson in the sparse regime, and the dual degrees being asymptotically Poisson in the ultra-dense regime. We also show that, in general, in the canonical ensemble the degrees are distributed according to a multivariate version of the *Poisson-Binomial* distribution [100], which admits the Gauss distribution and the Poisson distribution as limits in appropriate regimes.

incident to that node, in weighted graphs (i.e., graphs where edges can carry weights) the *strength* of a node is defined as the total weight of all the edges incident to that node. In Chapter 2, we focus on binary graphs only.

## §1.7 Summary of Chapter 4

In Chapters 2 and 3 breaking of ensemble equivalence between the microcanonical ensemble and the canonical ensemble is shown to occur when the constraint is put on the degree sequence (*configuration model*). In this case the constraint becomes a function of the number $n$ of nodes and we can therefore ask an interesting question: How is the relative entropy affected when the number of constraints is reduced, possibly in a way that depends on $n$?

In Chapter 4 we answer this question by analyzing the effect on the relative entropy when the number of constraints is reduced, i.e., when only *part* of the nodes are constrained degree (and the remaining nodes are left unconstrained). Intuitively, the relative entropy is expected to decrease as the number of constraints decreases. However, this is not a trivial issue, because when the number of constraints is reduced both the microcanonical ensemble and the canonical ensemble change. We consider random graphs with a prescribed partial degree sequence (reduced constraint). The breaking of ensemble equivalence is studied by analyzing how the relative entropy changes as a function of the number of constraints. In particular it is shown that the relative entropy is a monotone function in the number of constraints at the *macroscopic level*, i.e., when a positive fraction of the constraints is removed. More precisely, when only $m$ nodes are constrained and the remaining $n-m$ nodes are unconstrained, the relative entropy turns out to grow like $m \log n$ as $n \to \infty$.

Our analysis is based on a recent formula put forward by Squartini and Garlaschelli [93]. This formula predicts that the relative entropy is determined by the covariance matrix of the constraints in the canonical ensemble, in the regime where the graph is dense. Our result implies that ensemble equivalence breaks down whenever the dense regime is $\delta$-tame, irrespective of the number of degrees $m$ that are constrained, provided $m$ is not too close to $n$. It is further shown that the expression of the relative entropy corresponds, in the dense regime, to the degrees in the microcanonical ensemble being asymptotically *multivariate Dirac* and in the canonical ensemble being asymptotically *Gauss*.

## §1.8 Summary of Chapter 5

In Chapter 5 we analyze breaking of ensemble equivalence for the case in which topological constraints are imposed not only on the total number of edges but also on the total number of wedges, triangles, etc. We work in the dense regime, in which the number of edges per vertex scales proportionally with the number of vertices $n$. We compute the relative entropy of the two ensembles in the limit as $n$ goes to $\infty$, where the two ensembles are said to be *equivalent* if this relative entropy divided by $n^2$ tends to zero (which, up to a constant, can be interpreted as the relative entropy per edge). In particular, we show that the relative entropy divided by $n^2$ tends to $s_\infty > 0$ when the constraints are *frustrated*. We base our analysis on a *large deviation principle for graphons* and we provide results for three different choices of constraints.

## §1.9 Summary of Chapter 6

In Chapter 5 we considered a random graph subject to constraints on the total number of edges and the total number of triangles, in the dense regime. With the help of *large deviation theory for graphons*, we derived a variational formula for $s_\infty = \lim_{n\to\infty} n^{-2} s_n$, where $n$ is the number of vertices and $s_n$ is the *relative entropy* of the microcanonical ensemble with respect to the canonical ensemble. In Chapter 6 we analyze the behavior of $s_\infty$ when the constraints are close to but different from those of the Erdős-Rényi random graph. It turns out that the behavior changes when the total number of triangles is larger, respectively, smaller than that of the Erdős-Rényi random graph with a given total number of edges. In particular, we find that $s_\infty > 0$ when the constraints are *frustrated*, i.e., $T_2^* \neq T_1^{*3}$ with $T_1^*$ the edge density and $T_2^*$ the triangle density. The Erdős-Rényi random graph corresponds to $T_2^* = T_1^{*3}$, for which $s_\infty = 0$. We identify the scaling behavior of $s_\infty$ for fixed $T_1^*$ and $T_2^* \downarrow T_1^{*3}$, respectively, $T_2^* \uparrow T_1^{*3}$, and prove that the way in which $s_\infty$ tends to zero is different for the two limits. We also identify what the constrained random graph asymptotically looks like in the microcanonical ensemble.

## §1.10 Development of the chapters

This thesis presents new results about breaking of ensemble equivalence for complex networks. Chapter 2 investigates the role of the number of constraints in the breaking of ensemble equivalence phenomenon. Chapter 2 continues and generalizes the work in [92] and shows that nonequivalence occurs in the presence of an extensive number of topological constraints. Chapter 2 first considers the class of unipartite graphs with the constraint on the degree sequence, in the sparse regime. After that, results are extended to the class of bipartite graphs, and to more complicated classes of graphs with a modular structure. The dense regime is investigated in Chapter 3, where a formula of the relative entropy based on the covariance structure of the canonical ensemble, recently put forward by Garlaschelli and Squartini [93], is confirmed. The study of the configuration model is continued in Chapter 4, where a different question is answered. While extensivity of the number of constraints in the number of nodes was shown to play a crucial role in the phenomenon of breaking of ensemble equivalence, it remains an open question how reduction of the number of constraints affects this phenomenon. Chapter 4 analyzes the effect on the relative entropy when the number of constraint is reduced. It shows that, under certain hypothesis, breaking of ensemble equivalence is monotone in the number of constraints.

Chapters 5 and 6 conclude this thesis with a study of dense graphs with constraints on subgraph structures. In Chapter 5 breaking of ensemble equivalence is analyzed for the case of topological constraints on the number of edges and different subgraphs (wedges, triangles, etc.) at the same time. Here a large deviation principle for graphons is used to prove that breaking of ensemble equivalence occurs whenever the constraints are *frustrated*. Chapter 6 is a continuation of Chapter 5, for the case where the constraints are on the number of edges and triangles at the same time. In particular, constraints are chosen to be close to, but different from, the

so called Erdős-Rényi line. It turns out that when the total number of triangles is larger or smaller then the total number of edges, the behavior of the relative entropy is completely different.

## §1.11 Conclusions and Open Problems

In this thesis we analyze breaking of ensemble equivalence for Complex Networks with different types of constraints and in different regimes. The main conclusion of Chapter 2 and 3 is that the physical mechanism behind breaking of ensemble equivalence seems to be the *extensivity* of the number of constraints. In fact, both in the *sparse* and in the *dense* regime, the ensembles are shown to be non-equivalent whenever the the number of constraints grows extensively with the number of nodes. Moreover, Chapter 4 shows how breaking of ensemble equivalence reduces as the number of constrained nodes is reduced. On the other hand, Chapter 5 and 6 show a completely different mechanism behind breaking of ensemble equivalence, namely, *frustration* of the constraints. In the specific case where the constraint is on the number of edges and the number of triangles, the canonical ensemble scales like an Erdős-Rényi random graph with an appropriate edge density, but the microcanonical ensemble does not.

We conclude this introductory chapter with a number of open problems that can serve as a starting point for a future study of breaking of ensemble equivalence phenomenon in complex networks.

1 **Meaning of (non)equivalence**

In this thesis we analyze breaking of ensemble equivalence at the measure level, i.e., we study the limit of the $\alpha_n$-relative entropy (1.17) for different constraints, for different regimes and for different values of $\alpha_n$. One consequence of $\alpha_n$-equivalence can be derived through (1.2), also known as Pinsker's inequality. This relates a pseudo distance (the relative entropy) to a distance (the total variation distance) and implies that, whenever the ensembles are $\alpha_n$-equivalent, the total variation distance between the microcanonical and the canonical ensemble does not grow faster than $\sqrt{\alpha_n}$. On the other hand, Pinsker's inequality does not provide full information about what nonequivalence means for typical quantities characterizing the network. It would be interesting to understand what nonequivalence translates into for simulations of real-world networks.

2 **Monotonicity of the relative entropy in the number of constraints**

In Chapter 4 we analyze the effect on the breaking of ensemble equivalence when the number of constraints is reduced, i.e., when only part of the nodes are constrained in their degree (and the remaining nodes are left unconstrained). We find that the relative entropy is a monotone function in the number of constraints when a *positive fraction* of the constraints is removed.

a The result of Chapter 4 is based on a formula recently put forward by Squartini and Garlaschelli (see [93], which provides compelling evidence but not a rigorous proof). It would be interesting to prove the monotonicity

property for the relative entropy in a way that does not depend on this formula and possibly for different regimes and other types of constraints as well.

b Chapter 4 analyzes the relative entropy at a macroscopic level, but nothing is said about the microscopic level. More precisely, it would be interesting to understand how the relative entropy changes when a *single* constraint is removed, rather than a positive fraction of constraints. For example, what is the effect when the longest degree is removed? Is the effect the same or not when we decide to remove the smallest degree, or any other degree for that matter?

3 **Functions of the constraints**

In this thesis we analyze breaking of ensemble equivalence for a few specific types of constraint. The constraint is put on the number of edges in Chapter 2 and on the degree sequence in Chapters 3 and 4. In Chapter 5 and 6 the constraint is put on the number of edges and the number of triangles. It would be interesting to have a theorem proving the (non)equivalence of ensembles for general types of constraint, and possibly for general functions of the constraints as well.

CHAPTER 2

# Ensemble Nonequivalence in Random Graphs with Modular Structure

This chapter is based on:
D. Garlaschelli, F. den Hollander, and A. Roccaverde. Ensemble nonequivalence in random graphs with modular structure. *J. Phys. A*, 50(1):015001, 35, 2017

**Abstract**

Breaking of equivalence between the microcanonical ensemble and the canonical ensemble, describing a large system subject to hard and soft constraints, respectively, was recently shown to occur in large random graphs. Hard constraints must be met by every graph, soft constraints must be met only on average, subject to maximal entropy. In Squartini, de Mol, den Hollander and Garlaschelli (2015) it was shown that ensembles of random graphs are nonequivalent when the degrees of the nodes are constrained, in the sense of a non-zero limiting specific relative entropy as the number of nodes diverges. In that paper, the nodes were placed either on a single layer (uni-partite graphs) or on two layers (bi-partite graphs). In the present paper we consider an arbitrary number of intra-connected and inter-connected layers, thus allowing for modular graphs with a multi-partite, multiplex, time-varying, block-model or community structure. We give a full classification of ensemble equivalence in the sparse regime, proving that breakdown occurs as soon as the number of local constraints (i.e., the number of constrained degrees) is extensive in the number of nodes, irrespective of the layer structure. In addition, we derive an explicit formula for the specific relative entropy and provide an interpretation of this formula in terms of Poissonisation of the degrees.

# §2.1 Introduction and main results

## §2.1.1 Background and outline

For systems with many interacting components a detailed microscopic description is infeasible and must be replaced by a probabilistic description, where the system is assumed to be a random sample drawn from a set of allowed microscopic configurations that are consistent with a set of known macroscopic properties, referred to as *constraints*. Statistical physics deals with the definition of the appropriate probability distribution over the set of microscopic configurations and with the calculation of the resulting macroscopic properties of the system. The three main choices of probability distribution are: (1) the *microcanonical ensemble*, where the constraints are *hard* (i.e., are satisfied by each individual configuration); (2) the *canonical ensemble*, where the constraints are *soft* (i.e., hold as ensemble averages, while individual configurations may violate the constraints); (3) the *grandcanonical ensemble*, where also the number of components is considered as a soft constraint.

For systems that are large but finite, the three ensembles are obviously different and, in fact, represent different physical situations: (1) the microcanonical ensemble models completely isolated systems (where both the energy and the number of particles are "hard"); (2) the canonical ensemble models closed systems in thermal equilibrium with a heat bath (where the energy is "soft" and the number of particles is "hard"); (3) the grandcanonical ensemble models open systems in thermal and chemical equilibrium (where both the energy and the number of particles are "soft"). However, in the limit as the number of particles diverges, the three ensembles are traditionally *assumed* to become equivalent as a result of the expected vanishing of the fluctuations of the soft constraints, i.e., the soft constraints are expected to become asymptotically hard. This assumption of *ensemble equivalence*, which dates back to Gibbs [53], has been verified in traditional models of physical systems with short-range interactions and a finite number of constraints, but it does *not* hold in general. Nonetheless, equivalence is considered to be one of the pillars of statistical physics and underlies many of the results that contribute to our current understanding of large real-world systems.

Despite the fact that many textbooks still convey the message that ensemble equivalence holds for all systems, as some sort of universal asymptotic property, over the last decades various examples have been found for which it breaks down. These examples range from astrophysical processes [73], [96], [56], [72], [32], quantum phase separation [15], [8], [98], nuclear fragmentation [35], and fluid turbulence [40], [41]. Across these examples, the signatures of ensemble nonequivalence differ, which calls for a rigorous mathematical definition of ensemble (non)equivalence: (i) *thermodynamic equivalence* refers to the existence of an invertible Legendre transform between the microcanonical entropy and canonical free energy [98]; (ii) *macrostate equivalence* refers to the equivalence of the canonical and microcanonical sets of equilibrium values of macroscopic properties [98]; (iii) *measure equivalence* refers to the asymptotic equivalence of the microcanonical and canonical probability distributions in the thermodynamic limit, i.e., the vanishing of their specific relative entropy [97]. The latter

reference reviews the three definitions and shows that, under certain hypotheses, they are identical.

In the present paper we focus on the equivalence between microcanonical and canonical ensembles, although nonequivalence can in general involve the grandcanonical ensemble as well [106]. While there is consensus that nonequivalence occurs when the microcanonical specific entropy is non-concave as a function of the energy density in the thermodynamic limit, the classification of the physical mechanisms at the *origin* of nonequivalence is still open. In most of the models studied in the literature, nonequivalence appears to be associated with the non-additivity of the energy of the subparts of the system or with phase transitions [23], [24], [97]. A possible and natural mechanism for non-additivity is the presence of *long-range interactions*. Similarly, phase transitions are naturally associated with long-range order. These "standard mechanisms" for ensemble nonequivalence have been documented also in the study of random graphs. In [7], a Potts model on a random regular graph is studied in both the microcanonical and canonical ensemble, where the microscopic configurations are the spin configurations (not the configurations of the network itself). It is found that the long-range nature of random connections, which makes the model non-additive and the microcanonical entropy non-concave, ultimately results in ensemble nonequivalence. In [85], [86], [87] and [29], random networks with given densities of edges and triangles are considered, and phase transitions characterised by jumps in these densities are found, with an associated breaking of ensemble equivalence (where the microscopic configurations are network configurations).

Recently, the study of certain classes of uni-partite and bi-partite random graphs [92], [47] has shown that ensemble nonequivalence can manifest itself via an additional, novel mechanism, unrelated to non-additivity or phase transitions: namely, the presence of an *extensive* number of local topological constraints, i.e., the degrees and/or the strengths (for weighted graphs) of all nodes.[1] This finding explains previously documented signatures of nonequivalence in random graphs with local constraints, such as a finite difference between the microcanonical and canonical entropy densities [1] and the non-vanishing of the relative fluctuations of the constraints [95]. How generally this result holds beyond the specific uni-partite and bi-partite cases considered so far remains an open question, on which we focus in the present paper. By considering a much more general class of random graphs with a variable number of constraints, we confirm that the presence of an extensive number of local topological constraints breaks ensemble equivalence, even in the absence of phase transitions or non-additivity.

The remainder of our paper is organised as follows. In Section 2.1.2 we give the definition of measure equivalence and, following [92], show that it translates into a simple pointwise criterion for the large deviation properties of the microcanonical and canonical probabilities. In Section 2.1.3 we introduce our main theorems in pedagogical order, starting from the characterisation of nonequivalence in the simple

[1] While in binary (i.e., simple) graphs the *degree* of a node is defined as the number of edges incident to that node, in weighted graphs (i.e., graphs where edges can carry weights) the *strength* of a node is defined as the total weight of all edges incident to that node. In this paper, we focus on binary graphs only.

cases of uni-partite and bi-partite graphs already explored in [92], and subsequently moving on to a very general class of graphs with arbitrary multilayer structure and tunable intra-layer and inter-layer connectivity. Our main theorems, which (mostly) concern the *sparse regime*, not only characterise nonequivalence *qualitatively*, they also provide a *quantitative* formula for the specific relative entropy. In Section 2.2 we discuss various important implications of our results, describing properties that are fully general but also focussing on several special cases of empirical relevance. In addition, we provide an interpretation of the specific relative entropy formula in terms of Poissonisation of the degrees. We also discuss the implications of our results for the study of several empirically relevant classes of "modular" networks that have recently attracted interest in the literature, such as networks with a so-called multi-partite, multiplex [16], time-varying [58], block-model [57], [62] or community structure [43], [84]. In Section 2.3, finally, we provide the proofs of our theorems.

In future work we will address the *dense regime*, which requires the use of *graphons*. In that regime we expect nonequivalence to persist, and in some cases become even more pronounced.

### §2.1.2 Microcanonical ensemble, canonical ensemble, relative entropy

For $n \in \mathbb{N}$, let $\mathcal{G}_n$ denote the set of all simple undirected graphs with $n$ nodes. Let $\mathcal{G}_n^\sharp \subseteq \mathcal{G}_n$ be some non-empty subset of $\mathcal{G}_n$, to be specified later. Informally, the restriction from $\mathcal{G}_n$ to $\mathcal{G}_n^\sharp$ allows us to forbid the presence of certain links, in such a way that the $n$ nodes are effectively partitioned into $M \in \mathbb{N}$ groups of nodes (or "layers") of sizes $n_1, \dots, n_M$ with $\sum_{i=1}^M n_i = n$. This restriction can be made explicit and rigorous through the definition of a superstructure, which we call the *master graph*, that will be introduced later. A given choice of $\mathcal{G}_n^\sharp$ corresponds to the selection of a specific class of *multilayer* graphs with desired intra-layer and inter-layer connectivity, such as graphs with a multipartite, multiplex, time-varying, block-model or community structure. In the simplest case, $\mathcal{G}_n^\sharp = \mathcal{G}_n$, which reduces to the ordinary choice of uni-partite (single-layer) graphs. This example, along with various more complicated examples, is considered explicitly later on.

In general, any graph $\mathbf{G} \in \mathcal{G}_n^\sharp$ can be represented as an $n \times n$ matrix with elements

$$g_{i,j}(\mathbf{G}) = \begin{cases} 1 & \text{if there is a link between node } i \text{ and node } j, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Let $\vec{C}$ denote a vector-valued function on $\mathcal{G}_n^\sharp$. Given a specific value $\vec{C}^*$, which we assume to be *graphic*, i.e., realisable by at least one graph in $\mathcal{G}_n^\sharp$, the *microcanonical probability distribution* on $\mathcal{G}_n^\sharp$ with *hard constraint* $\vec{C}^*$ is defined as

$$P_{\text{mic}}(\mathbf{G}) = \begin{cases} 1/\Omega_{\vec{C}^*}, & \text{if } \vec{C}(\mathbf{G}) = \vec{C}^*, \\ 0, & \text{else,} \end{cases} \tag{2.2}$$

where

$$\Omega_{\vec{C}^*} = |\{\mathbf{G} \in \mathcal{G}_n^\sharp \colon \vec{C}(\mathbf{G}) = \vec{C}^*\}| > 0 \tag{2.3}$$

is the number of graphs that realise $\vec{C}^*$. The *canonical probability distribution* $P_{\rm can}(\mathbf{G})$ on $\mathcal{G}_n^\sharp$ is defined as the solution of the maximisation of the *entropy*

$$S_n(P_{\rm can}) = -\sum_{\mathbf{G}\in\mathcal{G}_n^\sharp} P_{\rm can}(\mathbf{G}) \ln P_{\rm can}(\mathbf{G}) \tag{2.4}$$

subject to the *soft constraint* $\langle\vec{C}\rangle = \vec{C}^*$, where $\langle\cdot\rangle$ denotes the average w.r.t. $P_{\rm can}$, and subject to the normalisation condition $\sum_{\mathbf{G}\in\mathcal{G}_n^\sharp} P_{\rm can}(\mathbf{G}) = 1$. This gives

$$P_{\rm can}(\mathbf{G}) = \frac{\exp[-H(\mathbf{G},\vec{\theta}^*)]}{Z(\vec{\theta}^*)}, \tag{2.5}$$

where

$$H(\mathbf{G},\vec{\theta}) = \vec{\theta}\cdot\vec{C}(\mathbf{G}) \tag{2.6}$$

is the *Hamiltonian* (or *energy*) and

$$Z(\vec{\theta}) = \sum_{\mathbf{G}\in\mathcal{G}_n^\sharp} \exp[-H(\mathbf{G},\vec{\theta})] \tag{2.7}$$

is the *partition function*. Note that in (2.5) the parameter $\vec{\theta}$ must be set to the particular value $\vec{\theta}^*$ that realises $\langle\vec{C}\rangle = \vec{C}^*$. This value also maximises the likelihood of the model, given the data [51].

It is worth mentioning that, in the social network analysis literature [25], maximum-entropy canonical ensembles of graphs are traditionally known under the name of Exponential Random Graphs (ERGs). Indeed, many of the examples of canonical graph ensembles that we will consider in this paper, or variants thereof, have been studied previously as ERG models of social networks. Recently, ERGs have also entered the physics literature [1], [2], [14], [81], [94], [95], [74], [45] ,[46], [62], [44], [82], [13] because of the wide applicability of techniques from statistical physics for the calculation of canonical partition functions. We will refer more extensively to these models, and to the empirical situations for which they have been proposed, in Section 2.2.2. Apart for a few exceptions [1], [82], [92], these previous studies have not addressed the problem of ensemble (non)equivalence of ERGs. The aim of the present paper is to do so exhaustively, and in a mathematically rigorous way, via the following definitions.

The *relative entropy* of $P_{\rm mic}$ w.r.t. $P_{\rm can}$ is

$$S_n(P_{\rm mic} \mid P_{\rm can}) = \sum_{\mathbf{G}\in\mathcal{G}_n^\sharp} P_{\rm mic}(\mathbf{G}) \ln \frac{P_{\rm mic}(\mathbf{G})}{P_{\rm can}(\mathbf{G})}, \tag{2.8}$$

and the *specific relative entropy* is

$$s_n = n^{-1}\, S_n(P_{\rm mic} \mid P_{\rm can}). \tag{2.9}$$

Following [97], [92], we say that the two ensembles are measure equivalent if and only if their specific relative entropy vanishes in the *thermodynamic limit* $n\to\infty$, i.e.,

$$s_\infty = \lim_{n\to\infty} n^{-1}\, S_n(P_{\rm mic} \mid P_{\rm can}) = 0. \tag{2.10}$$

It should be noted that, for a given choice of $\mathcal{G}_n^\sharp$ and $\vec{C}$, there may be different ways to realise the thermodynamic limit, corresponding to different ways in which the numbers $\{n_i\}_{i=1}^M$ of nodes inside the $M$ layers grow relatively to each other. So, (2.10) implicitly requires an underlying *specific definition of the thermodynamic limit*. Explicit examples will be considered in each case separately, and certain different realisations of the thermodynamic limit will indeed be seen to lead to different results. With this in mind, we suppress the $n$-dependence from our notation of quantities like $\mathbf{G}$, $\vec{C}$, $\vec{C}^*$, $P_{\rm mic}$, $P_{\rm can}$, $H$, $Z$. When letting $n \to \infty$ it will be understood that $\mathbf{G} \in \mathcal{G}_n^\sharp$ always.

Before considering specific cases, we recall an important observation made in [92]. The definition of $H(\mathbf{G}, \vec{\theta})$ ensures that, for any $\mathbf{G}_1, \mathbf{G}_2 \in \mathcal{G}_n^\sharp$, $P_{\rm can}(\mathbf{G}_1) = P_{\rm can}(\mathbf{G}_2)$ whenever $\vec{C}(\mathbf{G}_1) = \vec{C}(\mathbf{G}_2)$ (i.e., the canonical probability is the same for all graphs having the same value of the constraint). We may therefore rewrite (2.8) as

$$S_n(P_{\rm mic} \mid P_{\rm can}) = \ln \frac{P_{\rm mic}(\mathbf{G}^*)}{P_{\rm can}(\mathbf{G}^*)}, \tag{2.11}$$

where $\mathbf{G}^*$ is *any* graph in $\mathcal{G}_n^\sharp$ such that $\vec{C}(\mathbf{G}^*) = \vec{C}^*$ (recall that we have assumed that $\vec{C}^*$ is realisable by at least one graph in $\mathcal{G}_n^\sharp$). The condition for equivalence in (2.10) then becomes

$$\lim_{n\to\infty} n^{-1} \left[ \ln P_{\rm mic}(\mathbf{G}^*) - \ln P_{\rm can}(\mathbf{G}^*) \right] = 0, \tag{2.12}$$

which shows that the breaking of ensemble equivalence coincides with $P_{\rm mic}(\mathbf{G}^*)$ and $P_{\rm can}(\mathbf{G}^*)$ having different large deviation behaviour. Importantly, this condition is entirely local, i.e., it involves the microcanonical and canonical probabilities of a *single* configuration $\mathbf{G}^*$ realising the hard constraint. Apart from its theoretical importance, this fact greatly simplifies mathematical calculations. Note that (2.12), like (2.10), implicitly requires a specific definition of the thermodynamic limit. For a given choice of $\mathcal{G}_n^\sharp$ and $\vec{C}$, different definitions of the thermodynamic limit may result either in ensemble equivalence or in ensemble nonequivalence.

## §2.1.3 Main Theorems (Theorems 2.1.1-2.1.10)

Most of the constraints that will be considered below are *extensive* in the number of nodes.

#### Single layer: uni-partite graphs

The first class of random graphs we consider is specified by $M = 1$ and $\mathcal{G}_n^\sharp = \mathcal{G}_n$. This choice corresponds to the class of (simple and undirected) *uni-partite graphs*, where links are allowed between each pair of nodes. We can think of these graphs as consisting of a single layer of nodes, inside which all links are allowed. Note that in this simple case the thermodynamic limit $n \to \infty$ can be realised in a unique way, which makes (2.10) and (2.12) already well-defined.

**Constraints on the degree sequence.** For a uni-partite graph $\mathbf{G} \in \mathcal{G}_n$, the degree sequence is defined as $\vec{k}(\mathbf{G}) = (k_i(\mathbf{G}))_{i=1}^n$ with $k_i(\mathbf{G}) = \sum_{j \neq i} g_{i,j}(\mathbf{G})$. In what follows we constrain the degree sequence to a *specific value* $\vec{k}^*$, which (in accordance with our aforementioned general prescription for $\vec{C}^*$) we assume to be *graphical*, i.e., there is at least one graph with degree sequence $\vec{k}^*$. The constraints are therefore

$$\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^n \in \mathbb{N}_0^n, \tag{2.13}$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ with $\mathbb{N} = \{1, 2, \ldots\}$. This class is also known as the *configuration model* ([11], [18], [77], [78], [33], [95]; see also [99, Chapter 7]). In [92] the breaking of ensemble equivalence was studied in the *sparse regime* defined by the condition

$$m^* = \max_{1 \leq i \leq n} k_i^* = o(\sqrt{n}). \tag{2.14}$$

Let $\mathcal{P}(\mathbb{N}_0)$ denote the set of probability distributions on $\mathbb{N}_0$. Let

$$f_n = n^{-1} \sum_{i=1}^n \delta_{k_i^*} \in \mathcal{P}(\mathbb{N}_0), \tag{2.15}$$

be the *empirical degree distribution*, where $\delta_k$ denotes the point measure at $k$. Suppose that there exists a degree distribution $f \in \mathcal{P}(\mathbb{N}_0)$ such that

$$\lim_{n \to \infty} \|f_n - f\|_{\ell^1(g)} = 0, \tag{2.16}$$

where $g \colon \mathbb{N}_0 \to [0, \infty)$ is given by

$$g(k) = \log\left(\frac{k!}{k^k e^{-k}}\right), \qquad k \in \mathbb{N}_0, \tag{2.17}$$

and $\ell^1(g)$ is the vector space of functions $h \colon \mathbb{Z} \to \mathbb{R}$ with $\|h\|_{\ell^1(g)} = \sum_{k \in \mathbb{N}_0} |h(k)| g(k) < \infty$. For later use we note that

$$g(0) = 0, \qquad k \mapsto g(k) \text{ is strictly increasing}, \qquad g(k) = \tfrac{1}{2}\log(2\pi k) + O(k^{-1}), \quad k \to \infty. \tag{2.18}$$

**2.1.1 Theorem.** *Subject to* (2.13)–(2.14) *and* (2.16)*, the specific relative entropy equals*

$$s_\infty = \|f\|_{\ell^1(g)} > 0. \tag{2.19}$$

Thus, when we constrain the degrees we break the ensemble equivalence.

**2.1.2 Remark.** It is known that $\vec{k}^*$ is graphical if and only if $\sum_{i=1}^n k_i^*$ is even and

$$\sum_{i=1}^j k_i^* \leq j(j-1) + \sum_{i=j+1}^n \min(j, k_i^*), \qquad j = 1, \ldots, n-1. \tag{2.20}$$

In [5], the case where $k_i^*$, $i \in \mathbb{N}$, are i.i.d. with probability distribution $f$ is considered, and it is shown that

$$\lim_{n \to \infty} f^{\otimes n}\Big((k_1^*, \ldots, k_n^*) \text{ is graphical} \,\Big|\, \sum_{i=1}^n k_i^* \text{ is even}\Big) = 1 \tag{2.21}$$

as soon as $f$ satisfies $0 < \sum_{k \text{ even}} f(k) < 1$ and $\lim_{n\to\infty} n \sum_{k\geq n} f(k) = 0$. (The latter condition is slightly weaker than the condition $\sum_{k\in\mathbb{N}_0} kf(k) < \infty$.) In what follows we do *not* require the degrees to be drawn in this manner, but when we let $n \to \infty$ we always implicitly assume that the limit is taken *within the class of graphical degree sequences*.

**2.1.3 Remark.** A different yet similar definition of sparse regime, replacing (2.14), is given in van der Hofstad [99, Chapter 7]. This condition is formulated in terms of bounded second moment of the empirical degree distribution $f_n$ in the limit as $n \to \infty$. Theorem 2.1.1 carries over.

**Constraints on the total number of links only.** We now relax the constraints, and fix only the total number of links $L(\mathbf{G}) = \frac{1}{2}\sum_{i=1}^{n} k_i(\mathbf{G})$. The constraint therefore becomes

$$\vec{C}^* = L^*. \tag{2.22}$$

It should be note that in this case, the canonical ensemble coincides with the Erdős-Rényi random graph model, where each pair of nodes is independently connected with the same probability. As shown in [1], [92], in this case the usual result that the ensembles are asymptotically equivalent holds.

**2.1.4 Theorem.** *Subject to* (2.22)*, the specific relative entropy equals* $s_\infty = 0$.

**Two layers: bi-partite graphs**

The second class of random graphs we consider are *bi-partite graphs*. Here $M = 2$ and nodes are placed on two (non-overlapping) layers (say, top and bottom), and only links *across* layers are allowed. Let $\Lambda_1$ and $\Lambda_2$ denote the sets of nodes in the top and bottom layer, respectively. The set of all bi-partite graphs consisting of $n_1 = |\Lambda_1|$ nodes in the top layer and $n_2 = |\Lambda_2|$ nodes in the bottom layer is denoted by $\mathcal{G}_n^\sharp = \mathcal{G}_{n_1,n_2} \subset \mathcal{G}_n$. Bi-partiteness means that, for all $\mathbf{G} \in \mathcal{G}_{n_1,n_2}$, we have $g_{i,j}(\mathbf{G}) = 0$ if $i, j \in \Lambda_1$ or $i, j \in \Lambda_2$.

In a bipartite graph $\mathbf{G} \in \mathcal{G}_{n_1,n_2}$, we define the degree sequence of the top layer as $\vec{k}_{1\to2}(\mathbf{G}) = (k_i(\mathbf{G}))_{i\in\Lambda_1}$, where $k_i(\mathbf{G}) = \sum_{j\in\Lambda_2} g_{i,j}(\mathbf{G})$. Similarly, we define the degree sequence of the bottom layer as $\vec{k}_{2\to1}(\mathbf{G}) = (k'_i(\mathbf{G}))_{i\in\Lambda_2}$, where $k'_i(\mathbf{G}) = \sum_{j\in\Lambda_1} g_{i,j}(\mathbf{G})$. The symbol $s \to t$ highlights the fact that the degree sequence of layer $s$ is built from links pointing from $\Lambda_s$ to $\Lambda_t$ $(s, t = 1, 2)$. The degree sequences $\vec{k}_{1\to2}(\mathbf{G})$ and $\vec{k}_{2\to1}(\mathbf{G})$ are related by the condition that they both add up to the total number of links $L(\mathbf{G})$:

$$L(\mathbf{G}) = \sum_{i\in\Lambda_1} k_i(\mathbf{G}) = \sum_{j\in\Lambda_2} k'_j(\mathbf{G}). \tag{2.23}$$

**Constraints on the top and the bottom layer.** We first fix the degree sequence on both layers, i.e., we constrain $\vec{k}_{1\to2}(\mathbf{G})$ and $\vec{k}_{2\to1}(\mathbf{G})$ to the values $\vec{k}^*_{1\to2} = (k^*_i)_{i\in\Lambda_1}$ and $\vec{k}^*_{2\to1} = (k'^*_i)_{i\in\Lambda_2}$ respectively. The constraints are therefore

$$\vec{C}^* = \{\vec{k}^*_{1\to2}, \vec{k}^*_{2\to1}\}. \tag{2.24}$$

As mentioned before, we allow $n_1$ and $n_2$ to depend on $n$, i.e., $n_1 = n_1(n)$ and $n_2 = n_2(n)$. In order not to overburden the notation, we suppress the dependence on $n$ from the notation.

We abbreviate

$$\begin{aligned} m^* &= \max_{i\in\Lambda_1} k_i^*, \quad m'^* = \max_{j\in\Lambda_2} k_j'^*, \\ f_{1\to2}^{(n_1)} &= {n_1}^{-1} \sum_{i\in\Lambda_1} \delta_{k_i^*}, \quad f_{2\to1}^{(n_2)} = {n_2}^{-1} \sum_{j\in\Lambda_2} \delta_{k_j'^*}, \end{aligned} \tag{2.25}$$

and assume the existence of

$$A_1 = \lim_{n\to\infty} \frac{n_1}{n_1+n_2}, \quad A_2 = \lim_{n\to\infty} \frac{n_2}{n_1+n_2}. \tag{2.26}$$

(This assumption is to be read as follows: choose $n_1 = n_1(n)$ and $n_2 = n_2(n)$ in such a way that the limiting fractions $A_1$ and $A_2$ exist.) The *sparse regime* corresponds to

$$m^* m'^* = o({L^*}^{2/3}), \qquad n \to \infty. \tag{2.27}$$

We further assume that there exist $f_{1\to2}, f_{2\to1} \in \mathcal{P}(\mathbb{N}_0)$ such that

$$\lim_{n\to\infty} \|f_{1\to2}^{(n_1)} - f_{1\to2}\|_{\ell^1(g)} = 0, \quad \lim_{n\to\infty} \|f_{2\to1}^{(n_2)} - f_{2\to1}\|_{\ell^1(g)} = 0. \tag{2.28}$$

The specific relative entropy is

$$s_{n_1+n_2} = \frac{S_{n_1+n_2}(P_{\mathrm{mic}} \mid P_{\mathrm{can}})}{n_1+n_2}. \tag{2.29}$$

**2.1.5 Theorem.** *Subject to* (2.24) *and* (2.26)–(2.28)*,*

$$s_\infty = \lim_{n\to\infty} \frac{S_{n_1+n_2}(P_{\mathrm{mic}} \mid P_{\mathrm{can}})}{n_1+n_2} = A_1 \|f_{1\to2}\|_{\ell^1(g)} + A_2 \|f_{2\to1}\|_{\ell^1(g)}. \tag{2.30}$$

Since $A_1 + A_2 = 1$, it follows that $s_\infty > 0$, so in this case ensemble equivalence never holds.

**Constraints on the top layer only.** We now partly relax the constraints and only fix the degree sequence $\vec{k}_{1\to2}(\mathbf{G})$ to the value

$$\vec{C}^* = \vec{k}_{1\to2}^* = \big(k_i^*\big)_{i\in\Lambda_1}, \tag{2.31}$$

while leaving $\vec{k}_{2\to1}(\mathbf{G})$ unspecified (apart for the condition (2.23)). The microcanonical number of graphs satisfying the constraint is

$$\Omega_{\vec{k}_{1\to2}^*} = \prod_{i\in\Lambda_1} \binom{n_2}{k_i^*}. \tag{2.32}$$

The canonical ensemble can be obtained from (2.5) by setting

$$H(\mathbf{G}, \vec{\theta}) = \vec{\theta} \cdot \vec{k}_{1\to2}(\mathbf{G}). \tag{2.33}$$

Setting $\vec{\theta} = \vec{\theta^*}$ in order that equation (2.5) is satisfied, we can write the canonical probability as

$$P_{\text{can}}(\mathbf{G}) = \prod_{i \in \Lambda_1} (p_i^*)^{k_i(\mathbf{G})} (1 - p_i^*)^{n_2 - k_i(\mathbf{G})} \tag{2.34}$$

with $p_i^* = \frac{k_i^*}{n_2}$. Let

$$f_{n_1} = {n_2}^{-1} \sum_{i \in \Lambda_2} \delta_{k_i^*} \in \mathcal{P}(\mathbb{N}_0). \tag{2.35}$$

Suppose that there exists an $f \in \mathcal{P}(\mathbb{N}_0)$ such that

$$\lim_{n \to \infty} \|f_{n_1} - f\|_{\ell^1(g)} = 0. \tag{2.36}$$

The relative entropy per node can be written as

$$s_{n_1+n_2} = \frac{S_{n_1+n_2}(P_{\text{mic}} \mid P_{\text{can}})}{n_1 + n_2} = \frac{n_1}{n_1 + n_2} \|f_{n_1}\|_{\ell^1(g_{n_2})}, \tag{2.37}$$

with

$$g_{n_2}(k) = -\log \left[ \text{Bin}\left(n_2, \tfrac{k}{n_2}\right)(k) \right] \mathbb{I}_{0 \le k \le n_2}, \qquad k \in \mathbb{N}_0, \tag{2.38}$$

and $\text{Bin}(n_2, \frac{k}{n_2})(k) = \binom{n_2}{k} (\frac{k}{n_2})^k (\frac{n_2 - k}{k})^{n_2 - k}$ for $k = 0, \ldots, n_2$ and equals to 0 for $k > n_2$. We follow the convention $0 \log(0) = 0$.

In this partly relaxed case, different scenarios are possible depending on the specific realisation of the thermodynamic limit, i.e., on how $n_1, n_2$ tend to infinity. The ratio between the sizes of the two layers $c = \lim_{n \to \infty} \frac{n_2}{n_1} = \frac{A_2}{A_1}$ plays an important role.

**2.1.6 Theorem.** *Subject to* (2.31) *and* (2.36)*:*
(1) *If* $n_2 \to^{n \to \infty} \infty$ *with* $n_1$ *fixed* ($c = \infty$)*, then* $s_\infty = \lim_{n \to \infty} s_{n_1+n_2} = 0$.
(2) *If* $n_1, n_2 \to^{n \to \infty} \infty$ *with* $c = \infty$*, then* $s_\infty = \lim_{n \to \infty} s_{n_1+n_2} = 0$.
(3) *If* $n_1 \to^{n \to \infty} \infty$ *with* $n_2$ *fixed* ($c = 0$)*, then*

$$s_\infty = \lim_{n \to \infty} s_{n_1+n_2} = \|f\|_{\ell^1(g_{n_2})}. \tag{2.39}$$

(4) *If* $n_1, n_2 \to^{n \to \infty} \infty$ *with* $c \in [0, \infty)$*, then*

$$s_\infty = \frac{1}{1 + c} \|f\|_{\ell^1(g)}. \tag{2.40}$$

**Constraints on the total number of links only.** We now fully relax the constraints and only fix the total number of links, i.e.,

$$\vec{C^*} = L^*. \tag{2.41}$$

In analogy with the corresponding result for the uni-partite case (Theorem 2.1.4), in this case ensemble equivalence is restored.

**2.1.7 Theorem.** *Subject to* (2.41)*, the specific relative entropy equals* $s_\infty = 0$.

**Multiple layers**

We now come to our most general setting where we fix a finite number $M \in \mathbb{N}$ of layers. Each layer $s$ has $n_s$ nodes, with $\sum_{s=1}^{M} n_s = n$. Let $v_i^{(s)}$ denote the $i$-th node of layer $s$, and $\Lambda_s = \{v_1^{(s)}, \dots, v_{n_s}^{(s)}\}$ denote the set of nodes in layer $s$. We may allow links *both within and across* layers, while constraining the numbers of links among different layers separately. But we may as well switch off links inside or between (some of the) layers. The actual choice can be specified by a superstructure, which we denote as the *master graph* $\mathbf{\Gamma}$, in which self-loops are allowed but multi-links are not. The nodes set of $\mathbf{\Gamma}$ is $\{1, \dots, M\}$ and the associated adjacency matrix has entries

$$\gamma_{s,t}(\mathbf{\Gamma}) = \begin{cases} 1 & \text{if a link between layers } s \text{ and } t \text{ exists} \\ 0 & \text{otherwise.} \end{cases} \tag{2.42}$$

The chosen set of all multi-layer graphs with given numbers of nodes, layers, and admissible edges (we admit edges only between layers connected in the *master graph*) is $\mathcal{G}_n^\sharp = \mathcal{G}_{n_1,\dots,n_M}(\mathbf{\Gamma}) \subseteq \mathcal{G}_n$. In 2.2.2 we discuss various empirically relevant choices of $\mathbf{\Gamma}$ explicitly, while here we keep our discussion entirely general.

Given a graph $\mathbf{G}$, for each pair of layers $s$ and $t$ (including $s = t$) we define the *$t$-targeted degree sequence* of layer $s$ as $\vec{k}_{s\to t}(\mathbf{G}) = \big(k_i^t(\mathbf{G})\big)_{i\in\Lambda_s}$, where $k_i^t(\mathbf{G}) = \sum_{j\in\Lambda_t} g_{i,j}(\mathbf{G})$ is the number of links connecting node $i$ to all other nodes in layer $t$. For each pair of layers $s$ and $t$ such that $\gamma_{s,t}(\mathbf{\Gamma}) = 1$, we enforce the value $\vec{k}^*_{s\to t} = \big(k_i^{*\,t}\big)_{i\in\Lambda_s}$ as a constraint for the $t$-targeted degree sequence of layer $s$. For $\gamma_{s,t}(\mathbf{\Gamma}) = 0$ we have $\vec{k}^*_{s\to t} = \vec{0}$, but this constraint is automatically enforced by the master graph. Thus, the relevant constraints are

$$\vec{C}^* = \Big\{ \vec{k}^*_{s\to t} \colon\ s,t = 1,\dots,M \ \ \gamma_{s,t}(\mathbf{\Gamma}) = 1 \Big\}. \tag{2.43}$$

We abbreviate

$$L^*_{s,t} = \sum_{i\in\Lambda_s} k_i^{*\,t} = \sum_{j\in\Lambda_t} k_j^{*\,s}, \quad m^*_{s\to t} = \max_{i\in\Lambda_s} k_i^{*\,t}, \quad f^{(n_s)}_{s\to t} = n_s^{-1} \sum_{i\in\Lambda_s} \delta_{k_i^{*\,t}}, \tag{2.44}$$

where $L^*_{s,t}$ is the number of links between layers $s$ and $t$ (note that $L^*_{s,s}$ is *twice* the number of links inside layer $s$), and assume the existence of

$$A_s = \lim_{n_1,\dots,n_M\to\infty} \frac{n_s}{n} \quad \forall\, s, \tag{2.45}$$

where $\sum_{s=1}^{M} A_s = 1$. (As before, this assumption is to be read as follows: choose $n_s = n_s(n)$, $1 \le s \le M$, in such a way that the limiting fractions $A_s$,$1 \le s \le M$, exist.) The *sparse regime* corresponds to

$$\begin{aligned} &m^*_{s\to t} m^*_{t\to s} = o({L^*_{s,t}}^{2/3}), && n_s, n_t \to \infty \ \forall\, s \neq t, \\ &m^*_{s\to s} = o(n_s^{1/2}), && n_s \to \infty \ \forall\, s. \end{aligned} \tag{2.46}$$

We further assume that there exists $f_{s\to t} \in \mathcal{P}(\mathbb{N}_0)$ such that

$$\lim_{n_s\to\infty} \|f^{(n_s)}_{s\to t} - f_{s\to t}\|_{\ell^1(g)}, \qquad \lim_{n_s\to\infty} \|f^{(n_s)}_{s\to s} - f_{s\to s}\|_{\ell^1(g)} = 0. \tag{2.47}$$

**2.1.8 Theorem.** *Subject to* (2.43) *and* (2.45)–(2.47),

$$s_\infty = \sum_{\substack{s,t=1 \\ \gamma_{s,t}(\mathbf{\Gamma})=1}}^{M} A_s \, \|f_{s\to t}\|_{\ell^1(g)}. \tag{2.48}$$

The above result shows that, unless $A_s = 0$ whenever $\gamma_{s,t}(\mathbf{\Gamma}) = 1$ (i.e., unless only the nodes of the master graph that have no links or self-loops contribute a finite fraction of nodes in the corresponding layers), ensemble equivalence does not hold.

**Relaxing constraints in the multilayer case**

We next study the effects of relaxing constraints. This deserves a separate discussion, since in the multi-partite setting there are more possible ways of relaxing the constraints than in the uni-partite and bi-partite settings.

**One class of layers.** We first fix *two kinds of constraints*: (1) the total number of links between some pairs of layers; (2) the degree sequence between some other pairs of layers. We define the set of the edges of the *master graph* as $\mathcal{E} = \{(s,t) \in (M \times M) \colon \gamma_{s,t}(\mathbf{\Gamma}) = 1\}$. Then, we partition $\mathcal{E}$ into two parts, namely $\mathcal{D}, \mathcal{L} \subseteq \mathcal{E}$, with $\mathcal{D} \cap \mathcal{L} = \emptyset$, $\mathcal{D}$ and $\mathcal{L}$ symmetric, by requiring that $(s,t) \in \mathcal{D}$ $(\in \mathcal{L})$ when $(t,s) \in \mathcal{D}$ $(\in \mathcal{L})$. For each pair of layers $(s,t) \in \mathcal{D}$ we fix the degree sequence $\vec{k}^*_{s\to t}$ of every node of $\Lambda_s$ linking to $\Lambda_t$. As before, we impose that $\sum_{i\in\Lambda_s} k_i^{*\,t} = \sum_{j\in\Lambda_t} k_j^{*\,s}$. For each pair of layers $(s,t) \in \mathcal{L}$ we fix the total number of links $L^*_{s,t}$ $(L^*_{s,t} = L^*_{t,s})$.

The effect of relaxing some constraints affects the specific relative entropy: this will decrease because the pairs of layers with relaxed constraints (i.e., the pairs in $\mathcal{L}$) no longer contribute.

**2.1.9 Theorem.** *Subject to the above relaxation,*

$$s_\infty = \sum_{(s,t)\in\mathcal{D}} A_s \, \|f_{s\to t}\|_{\ell^1(g)}. \tag{2.49}$$

In particular, equivalence holds if and only if $\mathcal{D} = \emptyset$ or $A_s = 0$ for all $s$ endpoints of elements in $\mathcal{E}$. Note that, if $\mathcal{D} = \emptyset$, then we have a finite number of constraints (at most $M^2$), and this implies equivalence of the ensembles.

**Two classes of layers.** We may further generalise Theorem 2.1.8 as follows. Suppose that we have two classes of layers, $\mathcal{M}_1$ and $\mathcal{M}_2$. For every pair of layers $s,t \in \mathcal{M}_1$ such that $\gamma_{s,t}(\mathbf{\Gamma}) = 1$, we fix the degree sequences $\vec{k}^*_{s\to t}$ and $\vec{k}^*_{t\to s}$. For every pair of layers $s \in \mathcal{M}_1$, $t \in \mathcal{M}_2$, $\gamma_{s,t}(\mathbf{\Gamma}) = 1$ we fix the degree sequence $\vec{k}^*_{s\to t}$ from the layer in $\mathcal{M}_1$ to the layer in $\mathcal{M}_2$ (but not vice versa). We show that the resulting specific relative entropy is a mixture of the one in Theorem 2.1.8 and the one in Theorem 2.1.6. For $s = 1, \dots, M$ we set $A_s = \lim_{n_1,n_2,\dots,n_M\to\infty} \frac{n_s}{n}$.

**2.1.10 Theorem.** *Subject to the above relaxation,*

$$s_\infty = \sum_{\substack{s\in\mathcal{M}_1, \ t\in\mathcal{M}_1\cup\mathcal{M}_2 \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} A_s \, \|f_{s\to t}\|_{\ell^1(g)}. \tag{2.50}$$

In particular,

$$s_\infty = 0 \quad \Longleftrightarrow \quad A_s = 0 \; \forall\, s \in \big\{ u \in \mathcal{M}_1 \colon \exists\, t \in \mathcal{M}_1 \cup \mathcal{M}_2 \text{ with } \gamma_{u,t}(\mathbf{\Gamma}) = 1 \big\}. \tag{2.51}$$

**Another way for relaxing constraints.** We may think about another way for relaxing the constraints. We assume that $\gamma_{s,t}(\mathbf{\Gamma}) = 1$ for all $s, t = 1, 2, \dots, M$ and we fix $\vec{k}^*_s = \sum_{t=1}^M \vec{k}^*_{s\to t}$ for each $s = 1, 2, \dots, M$. This means that for each node we fix its degree sequence (no matter to which target layer, possibly its own layer). In this case we lose the multi-layer structure: constraints are no longer involving pairs of layers and the graphs are effectively uni-partite. This is the same case described in the configuration model of Theorem 2.1.1. There are still an extensive number of local constraints, and the ensembles are nonequivalent.



## §2.2 Discussion

In this section we discuss various important implications of our results. We first consider properties that are fully general, and afterwards focus on several special cases of empirical relevance.

### §2.2.1 General considerations

**Poissonisation.** The function $g$ in (2.17) has an interesting interpretation, namely,

$$g(k) = S\big(\, \delta[k] \mid \text{Poisson}[k] \,\big) \tag{2.52}$$

is the relative entropy of the Poisson distribution with average $k$ w.r.t. the Dirac distribution with average $k$. The specific relative entropy in (2.1.1) for the uni-partite setting can therefore be seen as a sum over $k$ of contributions coming from the nodes with fixed, respectively, average degree $k$. The microcanonical ensemble forces the degree of these nodes to be exactly $k$ (which corresponds to $\delta[k]$), while the canonical ensemble, under the sparseness condition in (2.14), forces their degree to be Poisson distributed with average $k$. The same condition ensures that in the limit as $n \to \infty$ the constraints act on the nodes essentially independently.

The same interpretation applies to Theorems 2.1.5–2.1.6 and 2.1.8–2.1.10. The result in Theorem 2.1.6(3) shows that in the bi-partite setting, when one of the layers tends to infinity while the other layer does not, Poissonisation does not set in fully. Namely, we have

$$s_n = \sum_{k=1}^n f(k) g_n(k), \qquad g_n(k) = S\big(\, \delta[k] \mid \text{Bin}(n, \tfrac{k}{n}) \,\big). \tag{2.53}$$

In words, the canonical ensemble forces the nodes in the infinite layer with average degree $k$ to draw their degrees towards the $n$ nodes in the finite layer essentially independently, giving rise to a binomial distribution. Only in the limit as $n \to \infty$ does this distribution converge to the Poisson distribution with average $k$.

**Additivity vs. non-additivity.** In all the other examples known so far in the literature, the generally accepted explanation for the breaking of ensemble equivalence is the presence of a non-additive energy, induced e.g. by long-range interactions [23], [24]. However, in the examples considered in the present paper, nonequivalence has a different origin, namely, the presence of an extensive number of local constraints. As we now show, this mechanism is completely unrelated to non-additivity and is therefore a novel mechanism for ensemble nonequivalence.

Intuitively, the energy of a system is additive when, upon partitioning the units of the system into non-overlapping subunits, the 'interaction' energy between these subunits is negligible with respect to the internal energy of the subunits themselves. The 'physical' size of the systems considered in this paper is given by the number $n$ of nodes, i.e., we are defining the network to become 'twice as large' when the number of nodes is doubled. Think, for instance, of a population of $n$ individuals and the corresponding social network connecting these individuals: we say that the size of the network doubles when the population doubles. Consistently, in (2.9) we have defined the specific relative entropy $s_n$ by diving $S_n$ by $n$. In accordance with this reasoning, in order to establish whether in our systems ensemble equivalence has anything to do with energy additivity, we need to define the latter *node-wise*, i.e., with respect to partitioning the set of nodes into nonoverlapping subsets. Note that, in the presence of more than one layer, we have allowed for the number of nodes in some layer(s) to remain finite (in general, to grow subextensively) as the total number of nodes goes to infinity (see for instance Theorem 2.1.6). In such a situation it makes sense to study additivity only with respect to the nodes in those layers that are allowed to grow extensively in the thermodynamic limit.

Formally, if we let $\mathcal{I}$ denote the union of all layers for which $A_s > 0$ (see (2.45)), then we say that the energy is *node-additive* if the Hamiltonian (2.6) can be written as

$$H(\mathbf{G},\vec{\theta}) = \sum_{i\in\mathcal{I}} H_i(\mathbf{G},\vec{\theta}) \qquad \forall\,\mathbf{G}\in\mathcal{G}_n^\sharp, \tag{2.54}$$

where the $\{H_i\}_{i\in\mathcal{I}}$ do not depend on common subgraphs of $\mathbf{G}$ (i.e., each of them can be restricted to a distinct subgraph of $\mathbf{G}$), and are therefore independent random variables.

The case of uni-partite graphs with fixed degree sequence (Theorem 2.1.1) is an example of ensemble *nonequivalence* with *non-additive* Hamiltonian, because the latter is defined as $H(\mathbf{G},\vec{\theta}) = \sum_{i=1}^n \theta_i k_i(\mathbf{G})$ and cannot be rewritten in the form of (2.54) with independent $\{H_i(\mathbf{G},\vec{\theta})\}$: the degrees $k_i(\mathbf{G})$ and $k_j(\mathbf{G})$ of any two distinct nodes $i$ and $j$ depend on a common subgraph of $\mathbf{G}$, i.e., the dyad $g_{i,j}(\mathbf{G})$. In the example of uni-partite graphs with a fixed total number of links (see (2.22)), the energy has the form $H(\mathbf{G},\vec{\theta}) = \theta L(\mathbf{G}) = \frac{1}{2}\theta\sum_{i=1}^n k_i(\mathbf{G})$, which is still *non-additive*. However, the ensembles are in this case *equivalent* (see Theorem 2.1.4).

By contrast, the case of bi-partite graphs with fixed degree sequence on the top layer and the nodes in the other layer growing subextensively (case (3) of Theorem 2.1.6) is an example of ensemble *nonequivalence* with an *additive* Hamiltonian. Indeed, from (2.33) we see that $H(\mathbf{G},\vec{\theta})$ is now a linear combination of the $n_1$ degrees of the nodes in layer $\Lambda_1$, each of which depends only on the (bi-partite) subgraph ob-

tained from the corresponding node of the top layer and all the nodes of the bottom layer. Here, unlike the uni-partite case, all these subgraphs are disjoint. Despite being node-additive, when $A_1 = 1$ ($c = 0$) this Hamiltonian leads to nonequivalence, as established in (2.39). Similar examples can be engineered using some of the relaxations in Section 2.1.3. Finally, the case of bi-partite graphs with fixed total number of links (Theorem 2.1.7) is an example of ensemble *equivalence* with an *additive* Hamiltonian.

The four examples above show that additivity or non-additivity of the Hamiltonian does *not* influence the breaking of ensembles equivalence in the examples considered here. What matters is the *extensiveness* of the number of constraints. This observation was already made in [92], and is confirmed in full generality for the multi-layer setting treated in the present paper. Indeed, our results indicate that, whenever the number $\kappa$ of constraints on the degrees is *subextensive*, i.e., $\kappa = o(n)$ where $n$ is the number of nodes, ensemble equivalence is restored.

Note that the above notion of *node additivity* should not be confused with that of *edge additivity*, i.e., the fact that the Hamiltonian can be written as a sum over independent pairs of nodes. Due to the linearity of the chosen (local) constraints on the entries $\{g_{i,j}\}_{i,j=1}^{n}$ of the adjacency matrix of the graph $\mathbf{G}$, our examples are always edge-additive (irrespective of whether they are ensemble-equivalent), while they may or may not be node-equivalent, as we have seen. In either case, there is no relation between additivity and equivalence.

We stress again that the extensivity of the (local) constraints is, with respect to the mechanisms for nonequivalence already explored in the literature so far, an additional (and previously unrecognised) *sufficient* mechanism. It is obviously not the only one, and definitely *not a necessary one*, as exemplified by the fact that, in dense networks, nonequivalence has been found even in the presence of only two constraints, such as the total numbers of edges and triangles [85, 86, 87, 29]. However, while in the previous examples the breaking of equivalence arises from the nonlinearity (with respect to $\{g_{i,j}\}$) of some constraint and is typically found in a specific (usually critical) region of the parameter space separating phases where ensemble equivalence still applies, in our setting ensemble nonequivalence arises from the extensiveness of the number of (linear) constraints and extends to the entire space of parameters of the models. In this sense it is a stronger form of nonequivalence. Moreover, while the nonequivalence of network ensembles with a finite number of constraints was previously reported only for dense graphs, we are documenting it for the unexplored regime of sparse graphs.

**A principled choice of ensembles.** Ensembles of random graphs with constraints are used for many practical purposes. Two important examples are *pattern detection* and *network reconstruction*. For concreteness, we briefly illustrate these examples before we emphasize the implications that our results have for these and other applications.

Pattern detection is the identification of nontrivial structural properties in a real-world network, through the comparison of such network with a suitable null model [94]. For instance, *community detection* is the identification of groups of nodes that are more densely connected with each other than expected under a null model [43], [84] (in Section 2.2.2 we discuss the relation between our models and community detection

in more detail). A null model is a random graph model that preserves some simple topological properties of the real network (typically local, like the degree sequence) and is otherwise completely random. So, maximum-entropy ensembles of graphs with given degrees are a key tool for pattern detection.

Network reconstruction employs purely local topological information to infer the higher-order structural properties of a real-world network [74]. This problem arises whenever the complete structure of a network is not known (for instance, due to confidentiality or privacy issues), but local properties are. An example relevant for the epidemiology of sexually transmitted diseases is the network of sexual contacts among people, for which only aggregate information (the total number of contacts with different partners) can be typically surveyed in a population. In such cases, optimal inference about the network can be achieved by maximising the entropy subject to the known (local) constraints, which again leads to the ensembles with fixed degrees considered here.

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

Among the possible criteria, we suggest one that we believe appropriate whenever the available data are subject to (even small) errors, i.e., when the measured value $\vec{C}^*$ entering as input in the construction of the random graph ensemble is, strictly speaking, the best available estimate for some unknown 'true' (error-free) value $\vec{C}^\times$. In this situation, we want that possible small deviations of $\vec{C}^*$ from $\vec{C}^\times$ result in small devations of $P^*_{\rm mic}$ and $P^*_{\rm can}$ from the corresponding $P^\times_{\rm mic}$ and $P^\times_{\rm can}$. Now, if $\vec{C}^* \neq \vec{C}^\times$ (no matter how "small" and in which norm this difference is taken), then $P^*_{\rm mic}$ will attach zero probability to any graph $\mathbf{G}^\times$ that realises the 'true' constraint $\vec{C}^\times$: $P^*_{\rm mic}(\mathbf{G}^\times) = 0$, while $P^\times_{\rm mic}(\mathbf{G}^\times) \neq 0$. Indeed, $P^*_{\rm mic}$ and $P^\times_{\rm mic}$ will have non-overlapping supports, so they will sample distinct sets of graphs. This means that even small initial errors in the knowledge of the constraints will be severely propagated to the entire microcanonical ensemble, and inference based on the latter will be highly biased. In particular, the 'true' network will never be sampled by $P^*_{\rm mic}$. On the other hand, if the difference between $\vec{C}^*$ and $\vec{C}^\times$ is small, then the difference between $P^*_{\rm can}$ and $P^\times_{\rm can}$ will also be small. So, even though $\vec{C}^\times$ is unknown, any graph $\mathbf{G}^\times$ that realises this value will be given a probability $P^*_{\rm can}(\mathbf{G}^\times)$ that is nonzero and not very different from the probability $P^\times_{\rm can}(\mathbf{G}^\times)$ that would be obtained by knowing the true value $\vec{C}^\times$. In general, small deviations of $\vec{C}^*$ from $\vec{C}^\times$ imply that $P^*_{\rm can}(\mathbf{G})$ is not very different from $P^\times_{\rm can}(\mathbf{G})$ for any graph $\mathbf{G}$, as desired. This implies that *even if $\vec{C}^*$ is affected by small errors, then a principled choice of ensembles is the canonical one.* So, besides being the mathematically simpler option, we argue that canonical ensembles

are also the most appropriate choice in the presence of 'noise'. A similar claim was already made in [95], and is here strengthened by our proof of nonequivalence.

## §2.2.2 Special cases of empirical relevance

Different choices of the master graph $\mathbf{\Gamma}$ induce different structural features in the graphs of the ensemble $\mathcal{G}_n^\sharp$. Convenient choices allow us to consider certain classes of graphs that have been introduced recently to study appropriate types of real-world networks of empirical relevance. We discuss some of these choices below. The full generality of our results in Section 2.1.3 allows us to immediately draw conclusions about the (non)equivalence of the corresponding ensembles in each case of interest. As an important outcome of this discussion, all the empirically relevant ensembles of graphs turn out to be nonequivalent. In line with our general observation at the end of the previous section, this implies that a principled choice of ensembles is needed in all practical applications.



**Scale-free uni-partite networks.** Clearly, the trivial case when the master graph has a single node ($M = 1$) with a self-loop, i.e., $\gamma_{11}(\mathbf{\Gamma}) = 1$, corresponds to the class of uni-partite graphs we considered in Section 2.1.3. Many real-world networks, at least at a certain level of aggregation, admit such uni-partite representation. Examples include the Internet, the World Wide Web and many biological, social and economic networks. A common property displayed by most of these real-world networks is the presence of a "broad" empirical degree distribution, often consistent with a power-law distribution with an upper cut-off [17]. Networks with a power-law degree distribution are said to be *scale-free* [22]. This empirical observation implies that real-world networks are very different from Erdős-Rényi random graphs (which have a much narrower degree distribution) and are more closely reproduced by a configuration model with a truncated power-law degree distribution $f_n$ (see (2.15)) of the form $f_n(k) = A_{\gamma,n} k^{-\gamma} \mathbb{I}_{1 \leq k \leq k_c(n)}$ with $\gamma > 1$, $A_{\gamma,n}$ the normalisation constant, and $\lim_{n\to\infty} k_c(n) = \infty$ and $k_c(n) = o(\sqrt{n})$. The so-called *structural cut-off* $k_c(n)$ makes the networks sparse, as in condition (2.14) [17]. Since $\lim_{n\to\infty} \|f_n - f\|_{\ell^1(g)} = 0$ with $f(k) = k^{-\gamma}/\zeta(\gamma)$ for $k \geq 1$ and 0 elsewhere, where $\zeta$ is the Riemann zeta-function, our result in (2.19) tells us that

$$s_\infty = \sum_{k\in\mathbb{N}} g(k)\, f(k) = \frac{1}{\zeta(\gamma)} \sum_{k\in\mathbb{N}} g(k)\, k^{-\gamma}. \tag{2.55}$$

Since $g(k) = \frac{1}{2}\log(2\pi k) + O(k^{-1})$ as $k \to \infty$, we find that $s_\infty$ tends to 1 as $\gamma \to \infty$ and diverges like $\sim 1/2(\gamma - 1)$ as $\gamma \downarrow 1$. This result shows that the simplest random graph ensemble consistent with the scale-free character of real-world networks is nonequivalent. Interestingly, as the tail exponent $\gamma$ decreases, the degree distribution becomes broader and the degree of nonequivalence increases. A similar conclusion was drawn in [92].

**2.2.1 Remark.** Suppose that for each $n \in \mathbb{N}$ the degrees are drawn in an i.i.d. manner from the truncated degree distribution $f_n$. Suppose further that $\sum_{k\in\mathbb{N}_0} k f(k) <$

$\infty$, i.e., $\gamma > 2$. Then, because $\sup_{n\in\mathbb{N}} \sum_{k\in\mathbb{N}_0} k f_n(k) = \sum_{k\in\mathbb{N}_0} k f(k) < \infty$, conditional on the sum of the degrees being even, the degree sequence is graphical with a probability tending to one as $n \to \infty$. This fact is the analogue of the result in [5] mentioned in Remark 2.1.2, and its proof is a straightforward extension of the argument in [5]. Truncation improves the chance of being graphical.

**Multipartite networks.** The case when the master graph has only $M = 2$ interconnected nodes *and no self-loops*, i.e., $\gamma_{1,2}(\mathbf{\Gamma}) = \gamma_{2,1}(\mathbf{\Gamma}) = 1$ and $\gamma_{1,1}(\mathbf{\Gamma}) = \gamma_{2,2}(\mathbf{\Gamma}) = 0$, coincides with the class of bi-partite graphs discussed in Section 2.1.3. Popular real-world examples relevant to economics, ecology and scientometrics are bank-firm, plant-pollinator and author-paper networks, respectively. In this case as well, empirical evidence shows that real-world bi-partite networks have broad degree distributions (at least on one of the two layers, and typically on both). Random graph models with only a global constraint on the total number of links (as in Theorem 2.1.7) are therefore unrealistic. The minimal ensemble that is consistent with the properties of most real-world bi-partite networks requires the specification of the degree sequence(s) as constraint(s) and is therefore nonequivalent.

A direct generalisation of the bi-partite case is when $\mathbf{\Gamma}$ is an $M$-dimensional matrix with zeroes along the diagonal and ones off the diagonal: $\gamma_{s,s}(\mathbf{\Gamma}) = 0\ \forall s$ and $\gamma_{s,t}(\mathbf{\Gamma}) = 1$ for all $s \neq t$. The induced graphs in $\mathcal{G}_n^\sharp$ have an "all-to-all" multipartite structure (i.e., links are allowed between all pairs of distinct layers, but not inside layers). From our Theorem 2.1.8 it follows that if the $t$-targeted degree sequences are specified as a constraint, then the relative entropy in the all-to-all multipartite case is

$$s_\infty = \sum_{\substack{s,t=1 \\ s\neq t}}^{M} A_s \, \|f_{s\to t}\|_{\ell^1(g)} > 0, \tag{2.56}$$

which proves again ensemble nonequivalence.

**Stochastic block-models.** Another important example is when the master graph is a complete graph with all self-loops realised, i.e., $\gamma_{s,t}(\mathbf{\Gamma}) = 1$ for all $s, t$. This prescription generates the class of so-called *stochastic block-models*, which are very popular in the social network analysis literature [57], [62], [44]. The earliest and simplest stochastic block-model [57] is one where only the total numbers of links between all pairs of blocks (including within each block) are specified. When we identify blocks with layers, this model coincides with our relaxed model considered in Theorem 2.1.9, with $\mathcal{D} = \emptyset$. It follows as a corollary that this model is ensemble equivalent:

$$s_\infty = 0. \tag{2.57}$$

However, this model predicts that, within each block, the expected topological properties of the network are those of an Erdős-Rényi random graph, a property that is contradicted by empirical evidence. So, unless the number of blocks is chosen to be comparable with the number of nodes (which in our case is contradicted by the

requirement that $M$ is finite), the traditional block-model is not a good model of real-world networks.

More recently, emphasis has been put on the more realistic *degree-corrected stochastic block-model* [62], where an additional constraint is put on the degree of all nodes. An even more constrained variant of this model has been proposed in [44], where the constraints coincide with the $t$-targeted degree sequences $\{\vec{k}_{s\to t}\}_{s,t}$ among all pairs of blocks. To distinguish this model from the "generic" degree-corrected block-model, we call it the *targeted degree-corrected block-model*. This coincides with our model in Section 2.1.3, with the block structure given by the (complete) master graph. From Theorem 2.1.8 we calculate the relative entropy as

$$s_\infty = \sum_{s,t=1}^{M} A_s \, \|f_{s\to t}\|_{\ell^1(g)} > 0. \tag{2.58}$$

We can therefore conclude that, unlike the traditional block-model considered above, the targeted degree-corrected model is ensemble nonequivalent. We also note that, unlike stated in [44], the targeted degree-corrected block-model is not just a re-parametrisation of the untargeted degree-corrected model. While fixing the targeted degree sequences automatically realises the constraints of the untargeted model, the converse is not true. Being a relaxation of the targeted model, we expect the untargeted model to have a relative entropy smaller than in (2.58), further illustrating the difference between the two models. Yet, we expect the relative entropy in the untargeted model to be strictly positive for, every choice of the degree sequence, since there is still an extensive number of active constraints. This would support the claim made in [82] that, *for small values of the degrees*, the degree-corrected block-models with soft and hard constraints are not equivalent in the thermodynamic limit. At the same time, it would contradict the claim made in the same reference that, *if all degrees become large* (but still in the sparse regime), the two ensembles become equivalent. Indeed, from the behaviour of $g(k)$ for large $k$ (see (2.18)) and the normalisation by $n$ in (2.9), we expect a finite specific relative entropy in that case as well.

**Networks with community structure.** Another very important class of graphs that are studied intensively in the literature are graphs with community structure [43], [84]. This class is related to the block-models described above, but is in general different. Community structure is loosely defined as the presence of groups of nodes that are more densely interconnected internally than with each other. One of the possible ways to quantitatively define the presence of communities in a real-world network is in terms of a positive difference between the realised number of intra-community links and the corresponding expected number calculated under a certain null hypothesis. This definition can be made more explicit by introducing the concept of *modularity* [43], [84]. For a graph with $n$ nodes, a non-overlapping partition of nodes into $M$ communities can be specified by the $n$-dimensional vector $\vec{\sigma}$, where the $i$-th entry $\sigma_i \in \{1, \dots, M\}$ is an integer number labelling the community to which node $i$ is assigned by that particular partition. For a given real-world graph $\mathbf{G}^*$, the

modularity is a function on the space of possible partitions, defined as

$$Q_{\mathbf{G}^*}(\vec{\sigma}) = K_{\mathbf{G}^*} \sum_{1 \leq i < j \leq n} (g_{ij}(\mathbf{G}^*) - \langle g_{ij} \rangle)\, \delta_{\sigma_i, \sigma_j}, \tag{2.59}$$

where $K_{\mathbf{G}^*}$ is an (inessential) normalisation constant (independent of the partition $\vec{\sigma}$) intended to have the property $Q_{\mathbf{G}^*} \in [-1, +1]$, and $\langle g_{ij} \rangle$ is the expected value of $g_{ij}(\mathbf{G})$ under the null hypothesis. The null hypothesis leads to a *null model* for the real-world network $\mathbf{G}^*$. The most popular choice for this null model is the canonical configuration model in the sparse regime, which gives $\langle g_{ij} \rangle = k_i^* k_j^* / 2L^*$ for $i \neq j$ and $\langle g_{ii} \rangle = 0$, where $k_i^*$, $k_j^*$ and $L^*$ are all calculated on $\mathbf{G}^*$ (see (2.72) in the proof of Theorem 2.1.1).

Now, if the real-world network $\mathbf{G}^*$ is indeed composed of communities, then the partition $\vec{\sigma}^\dagger$ that encodes these communities will be such that $Q_{\mathbf{G}^*}(\vec{\sigma}^\dagger) > 0$, i.e., the total number of links inside communities will be larger than the expected number under the null model. More stringently, the 'optimal' partition into communities can be defined as the one that maximises $Q_{\mathbf{G}^*}(\vec{\sigma})$, provided that the corresponding value $\max_{\vec{\sigma}} Q_{\mathbf{G}^*}(\vec{\sigma})$ is positive. Indeed, one of the most popular ways in which communities are looked for in real-world networks is through the process of modularity maximisation. The higher the value of the maximised modularity, the sharper the community structure. In practice, the problem of community detection is complicated by the possible existence of many local minima of $Q_{\mathbf{G}^*}(\vec{\sigma})$ and by the fact that $Q_{\mathbf{G}^*}(\vec{\sigma}^\dagger)$ may be positive even for "noisy communities", i.e., communities induced by chance only out of randomness in the data.

In our setting, community structure can be easily induced in the multilayer graph ensemble $\mathcal{G}_n^\sharp = \mathcal{G}_{n_1,\ldots,n_M}(\mathbf{\Gamma})$ through a convenient choice of the master graph $\mathbf{\Gamma}$ and of the constrained $t$-targeted degree sequences $\{\vec{k}^*_{s \to t}\}$. First, we identify the $M$ layers $\{\Lambda_s\}$ with the desired communities and define the corresponding partition $\vec{\sigma}^\dagger$ through $\sigma_i^\dagger = \Lambda_s$ if $i \in \Lambda_s$. Next, we require that the master graph $\mathbf{\Gamma}$ has all possible self-loops, plus a desired number of additional edges that need not be maximal (pairs of distinct communities are not necessarily connected in real-world networks). Finally, we need to require that the $t$-targeted degree sequences induce an excess of intra-community links with respect to the null model, so that the modularity is at least positive, i.e., $Q_{\mathbf{G}^*}(\vec{\sigma}^\dagger) > 0$, and at best maximised by the desired partition, i.e.,

$\vec{\sigma}^\dagger = \text{argmax}_{\vec{\sigma}}\, Q_{\mathbf{G}^*}(\vec{\sigma})$. To this end, we rewrite

$$\begin{aligned} Q_{\mathbf{G}^*}(\vec{\sigma}^\dagger) &= K_{\mathbf{G}^*} \sum_{1\le i<j\le n} \left(g_{ij}(\mathbf{G}^*) - \langle g_{ij}\rangle\right) \delta_{\sigma_i^\dagger,\sigma_j^\dagger} \\ &= \frac{K_{\mathbf{G}^*}}{2} \sum_{1\le i,j\le n} \left(g_{ij}(\mathbf{G}^*) - \langle g_{ij}\rangle\right) \delta_{\sigma_i^\dagger,\sigma_j^\dagger} \\ &= \frac{K_{\mathbf{G}^*}}{2} \sum_{s=1}^{M} \sum_{i,j\in\Lambda_s} \left(g_{ij}(\mathbf{G}^*) - \frac{k_i^* k_j^*}{2L^*}\right) \\ &= \frac{K_{\mathbf{G}^*}}{2} \sum_{s=1}^{M} \left(L_{s,s}^* - \frac{1}{2L^*}\Big(\sum_{i\in\Lambda_s} k_i^*\Big)^2\right) \\ &= \frac{K_{\mathbf{G}^*}}{2} \sum_{s=1}^{M} \left(L_{s,s}^* - \frac{1}{\sum_{s,t=1}^{M} L_{s,t}^*}\Big(\sum_{t=1}^{M} L_{s,t}^*\Big)^2\right), \end{aligned} \tag{2.60}$$

where we use $g_{ii}(\mathbf{G}^*) = \langle g_{ii}\rangle = 0$, $k_i^* = \sum_{t=1}^{M} k_i^{*t}$ and $2L^* = \sum_{s,t=1}^{M} L_{s,t}^*$. So, the weaker condition $Q_{\mathbf{G}^*}(\vec{\sigma}^\dagger) > 0$ is realised by requiring that $\{\vec{k}_{s\to t}^*\}$ satisfies the inequality

$$\sum_{s=1}^{M} L_{s,s}^* > \frac{\sum_{s=1}^{M}\Big(\sum_{t=1}^{M} L_{s,t}^*\Big)^2}{\sum_{s,t=1}^{M} L_{s,t}^*}, \tag{2.61}$$

where $L_{s,t}^* = \sum_{i\in\Lambda_s} k_i^{*t}$. The above inequality explicitly states that the number of realised intra-community edges counted in the left-hand side should be larger than the expected number calculated in the right-hand side. The stronger condition $\vec{\sigma}^\dagger = \text{argmax}_{\vec{\sigma}}\, Q_{\mathbf{G}^*}(\vec{\sigma})$ should instead be enforced by looking for the specific $\{\vec{k}_{s\to t}^*\}$ that maximises (2.60).

Independently of how communities are induced in our framework, our results show that *ensembles of random graphs with community structure* (according to the definition above) *are nonequivalent*, with a relative entropy given by (2.48) where the degree distributions $\{f_{s\to t}\}$ are induced by suitable $t$-targeted degree sequences that realise (2.61) and possibly also $\vec{\sigma}^\dagger = \text{argmax}_{\vec{\sigma}}\, Q_{\mathbf{G}^*}(\vec{\sigma})$.

**Multiplex networks and time-varying graphs.** Two other important classes of graphs that have recently gained attention are those of *multiplex* networks [16] and *time-varying graphs* [58].

Multiplex networks are networks where the same set of nodes can be connected by $M$ different types of links [16]. Two examples, both studies in [52], are the multiplex of international trade in different products (where nodes are world countries and links of different type represent international trade in different products) and the multiplex of flights by different airlines (where nodes are airports and links of different type represent flights operated by different companies). An equivalent and widely used representation for a multiplex is one where a number $M$ of layers is introduced, the same nodes are replicated in each layer, and inside each layer an ordinary graph is constructed, specified by all links of a single type. Links only exist within layers,

and not across layers. Indeed, what 'couples' the different layers and makes a real-world multiplex different from a collection of independent layers is the empirical fact that the topological properties of the layer-specific networks are typically strongly (either positively or negatively) correlated. For instance, networks of trade in different products have a similar structure, and most notably countries that are 'hubs' in one layer are likely to be hubs in other layers as well. By contrast, airports that are hubs for a domestic airline are likely not to be hubs for other domestic airlines [52]. This means that, for each node $i$ in real-world networks, the $M$ numbers of intra-layer links (i.e., the *intra-layer degrees*) are in general (anti)correlated.

Time-varying graphs are collections of temporal snapshots of the same network [58]. If the set of nodes in the network does not change with time, then a time-varying graph can be represented as a multiplex where each temporal snapshot is a single layer. (Note that multiplex networks themselves can vary over time [79].) Again, while not interacting directly via links, the different layers are mutually dependent because of empirical correlations between the properties of the same physical network across its temporal snapshots. Therefore this type of time-varying graphs can be treated in a way formally similar to that used for multiplex networks, the only difference being that a natural temporal ordering can be defined for the snapshots of time-varying graphs, while this is in general not true for the layers of a multiplex.

In our framework, a multiplex or time-varying network can be introduced by identifying each link type with a layer $\Lambda_s$ and by requiring that the only edges of the master graph $\mathbf{\Gamma}$ are self-loops, i.e., $\gamma_{s,s}(\mathbf{\Gamma}) = 1$ for $s = 1, M$ and $\gamma_{s,t}(\mathbf{\Gamma}) = 0$ for $s \neq t$. Note that this specification, which implies $\vec{k}^*_{s\to t} = \vec{0}$ for $s \neq t$, is somehow 'dual' to the one defining all-to-all multipartite networks (see above). The fact that nodes in different layers are replicas of the same set of $n$ nodes implies that $|\Lambda_s|$ is the same for all $s$, i.e., $n_s = n/M$. Finally, the 'coupling' between the topological properties of different layers can be introduced by assigning (anti)correlated $t$-targeted degree sequences, i.e., by choosing (anti)correlated entries for every pair of vectors $\vec{k}^*_{s\to s}$ and $\vec{k}^*_{t\to t}$, $s \neq t$. Real-world multiplexes, including the two examples made above, are well reproduced by such a model [52]. Our results imply that the relevant ensembles are nonequivalent. In particular, as a corollary of Theorem 2.1.8 we have

$$s_\infty = \frac{1}{M}\sum_{s=1}^{M} \|f_{s\to s}\|_{\ell^1(g)}. \tag{2.62}$$

So, the relative entropy between the microcanonical and canonical distributions is the average of the relative entropy of all layers, where for each layer $s$ the relative entropy is the same as that obtained for a uni-partite network with $n/M$ nodes and limiting degree distribution $f_{s\to s}$ (see Theorem 2.1.1). Moreover, the presence of correlations between $\vec{k}^*_{s\to s}$ and $\vec{k}^*_{t\to t}$ translate into dependencies between $\|f_{s\to s}\|_{\ell^1(g)}$ and $\|f_{t\to t}\|_{\ell^1(g)}$. In particular, in case of perfect correlation ($\vec{k}^*_{s\to s} = \vec{k}^*_{t\to t}$ for all $s, t$), all the degree distributions are equal to a common one $f_{s\to s} = f\ \forall s$, and we get

$$s_\infty = \|f\|_{\ell^1(g)}. \tag{2.63}$$

In this case, the degree of nonequivalence is the same as that obtained for a single

uni-partite network with $n/M$ nodes and limiting degree distribution $f$ (see Theorem 2.1.1).

**Interdependent multilayer networks.** Finally, we discuss the class of *interdependent multilayer networks*, which are multiplex networks with the addition of inter-layer links [16]. Nodes in different layers are still replicas of the same set of nodes, so we still have $n_s = n/M$ for all $s$. Similarly, the topological properties of different intra-layer networks are still (anti)correlated, which can be again realised by choosing (anti)correlated entries for every pair of vectors $\vec{k}^*_{s\to s}$ and $\vec{k}^*_{t\to t}$, $s \neq t$. However, while we still require $\gamma_{s,s}(\mathbf{\Gamma}) = 1$ for $s = 1, M$, now we no longer require $\gamma_{s,t}(\mathbf{\Gamma}) = 0$ for $s \neq t$. Therefore the degree of nonequivalence can only increase with respect to (2.62). Indeed, Theorem 2.1.8 now leads to

$$s_\infty = \frac{1}{M} \sum_{\substack{s,t=1 \\ \gamma_{s,t}(\mathbf{\Gamma})=1}}^{M} \|f_{s\to t}\|_{\ell^1(g)}, \tag{2.64}$$

which shows that the relative entropy is no longer only an average over the layer-specific relative entropies, since inter-layer relative entropies give additional contributions.

**Networks of networks.** A final class of graphs worth mentioning is the so-called *networks of networks*, sometimes constructed by different 'micro-networks' that are coupled together into a 'macro-network' where each node is a micro-network itself [34]. This class is similar to the interdependent multilayer networks considered above, but here there is no identification of the nodes in different layers to the same physical entity. An example is provided by multi-scale transport networks, where different cities are internally characterised by their local urban transport networks and at the same time are coupled through a long-distance inter-city transport network (like highways or flights). In our framework, this class of network can be induced by identifying the master graph $\mathbf{\Gamma}$ with the macro-network, and the $M$ intra-layer subgraphs with the micro-networks. To have all micro-networks non-empty, the master graph must have all self-loops realised. This case is similar to the block-model mentioned above, but now the master graph itself can be chosen to have nontrivial structural properties, such as community structure, to resemble the specific properties of real-world networks of networks.

If the $t$-targeted degree sequences $\{\vec{k}^*_{s\to t}\}$ $(s, t = 1, M)$ are all enforced as constraints, then the relative entropy is given by (2.48) with $\gamma_{s,s}(\mathbf{\Gamma}) = 1$ for all $s$. However, in this class of models it is often more natural to assume that the internal degree sequence $\vec{k}^*_{s\to s}$ of each micro-network (layer) $s$ is enforced (in order to get realistic micro-network topologies), while between every pair $s, t$ $(s \neq t)$ of micro-networks only the number of links $L^*_{s,t}$ is fixed (because the topology of the master graph is already chosen in order to replicate the empirical macro-network). This leads to the relaxed model in Theorem 2.1.9 with $\mathcal{D} = \{(s, s)\colon\ s = 1, M\}$. The relative entropy is

therefore

$$s_\infty = \sum_{s=1}^{M} A_s \, \|f_{s\to s}\|_{\ell^1(g)} \tag{2.65}$$

and is still positive, even though the links among micro-networks do not contribute to it.



## §2.3 Proofs of Theorems 2.1.1-2.1.10

### §2.3.1 Proof of Theorem 2.1.1

*Proof.* The microcanonical number $\Omega_{\vec{k}^\star}$ is not known in general, but asymptotic results exist in the *sparse regime* defined by the condition (2.14). For this regime it was shown in [10], [76] that

$$\Omega_{\vec{k}^*} = \frac{\sqrt{2}\,(\frac{2L^*}{e})^{L^*}}{\prod_{i=1}^n k_i^*!}\, e^{-(\overline{k^{*2}}/2\overline{k^*})^2 + \frac{1}{4} + o(n^{-1}\overline{k^*}^3)}, \tag{2.66}$$

where $\overline{k^*} = n^{-1}\sum_{i=1}^n k_i^*$ (average degree), $L^* = n\overline{k^*}/2$ (number of links), $\overline{k^{*2}} = n^{-1}\sum_{i=1}^n k_i^{*2}$ (average square degree). The canonical ensemble has Hamiltonian $H(\mathbf{G},\vec{\theta}) = \sum_{i=1}^n \theta_i k_i(\mathbf{G})$, where $\mathbf{G}$ is a graph belonging to $\mathcal{G}_n$, and $k_i(\mathbf{G}) = \sum_{j\neq i} g_{i,j}(\mathbf{G})$ is the degree of the node $i$. The partition function equals

$$\begin{aligned} Z(\theta) &= \sum_{\mathbf{G}\in\mathcal{G}_n} e^{-H(\mathbf{G},\vec{\theta})} = \sum_{\mathbf{G}\in\mathcal{G}_n} \prod_{1\le i<j\le n} e^{-\theta_i g_{i,j}(\mathbf{G})} \\ &= \sum_{\mathbf{G}\in\mathcal{G}_n} \prod_{1\le i<j\le n} e^{-(\theta_i+\theta_j) g_{i,j}(\mathbf{G})} = \prod_{1\le i<j\le n} (1+e^{-(\theta_i+\theta_j)}). \end{aligned} \tag{2.67}$$

The canonical probability equals

$$P_{\rm can}(\mathbf{G}\mid\vec{\theta}) = \frac{\prod_{1\le i<j\le n} e^{-(\theta_i+\theta_j)g_{i,j}(\mathbf{G})}}{Z(\vec{\theta})} = \prod_{1\le i<j\le n} \frac{e^{-(\theta_i+\theta_j)g_{i,j}(\mathbf{G})}}{1+e^{-(\theta_i+\theta_j)}}. \tag{2.68}$$

Setting $p_{ij}^* \equiv e^{-\theta_i^*-\theta_j^*}/(1+e^{-\theta_i^*-\theta_j^*})$, and $\vec{\theta}^*$ such that

$$\sum_{j\neq i} \frac{e^{-\theta_i^*-\theta_j^*}}{1+e^{-\theta_i^*-\theta_j^*}} = k_i^* \quad \forall\, i \tag{2.69}$$

we have

$$P_{\rm can}(\mathbf{G}) = \prod_{1\le i<j\le n} (p_{ij}^*)^{g_{ij}} (1-p_{ij}^*)^{1-g_{ij}}. \tag{2.70}$$

It is ensured by (2.14) that $\lim_{n\to\infty} \frac{1}{n}\sum_{1\le i<j\le n} \widehat{p}_{ij}^{*\,2} = 0$, a condition under which we can show that (2.70) has the same asymptotic behaviour as

$$\widehat{P}_{\rm can}(\mathbf{G}) = \prod_{1\le i<j\le n} (\widehat{p}_{ij}^*)^{g_{ij}} (1-\widehat{p}_{ij}^*)^{1-g_{ij}}, \tag{2.71}$$

with

$$\widehat{p}^{\,*}_{ij} = e^{-\theta^*_i-\theta^*_j} = \frac{k^*_i k^*_j}{2L^*}. \tag{2.72}$$

Indeed,

$$\frac{1}{n}\log\left(\frac{\widehat{P}_{\rm can}(\mathbf{G})}{P_{\rm can}(\mathbf{G})}\right) = \frac{1}{n}\sum_{1\le i<j\le n} g_{i,j}\log(1-\widehat{p}^{\,*}_{ij}) - \frac{1}{n}\sum_{1\le i<j\le n}\log(1-\widehat{p}^{\,*\,2}_{ij}) \to 0, \quad n\to\infty, \tag{2.73}$$

because

$$\sum_{1\le i<j\le n} g_{i,j}\log(1-\widehat{p}^{\,*}_{ij}) \le (m^*)^2 + O(\widehat{p}^{\,*\,2}_{ij}) \tag{2.74}$$

and

$$0 \le \frac{1}{n}\sum_{1\le i<j\le n}\widehat{p}^{\,2}_{ij} = \frac{1}{2}\left[\frac{\sum_{i=1}^n k_i^2}{\sqrt{n}\sum_{i=1}^n k_i}\right]^2 \le \frac{1}{2}\frac{(m^*)^2}{n} \to 0, \qquad n\to\infty. \tag{2.75}$$

This implies $\sum_{1\le i<j\le n}\ln(1-p^*_{ij}) = -\sum_{1\le i<j\le n} k^*_i k^*_j/2L^* + o(n)$. Thus,

$$\ln P_{\rm can}(\mathbf{G}^*) \;\; = \;\; \sum_{i=1}^n k^*_i \ln k^*_i - L^*\ln(2L^*) - L^* + o(n). \tag{2.76}$$

Combining (2.66) and (2.76), we obtain (recall (2.17))

$$S_n(P_{\rm mic} \mid P_{\rm can}) = \sum_{i=1}^n g(k^*_i) + o(n), \qquad n\to\infty, \tag{2.77}$$

where $g(k) = \log\left(\frac{k!}{k^k e^{-k}}\right)$, as defined in (2.17). With the help of (2.15) this reads

$$n^{-1}\, S_n(P_{\rm mic} \mid P_{\rm can}) = \sum_{k\in\mathbb{N}_0} f_n(k)g(k) + o(1) = \|f_n\|_{\ell^1(g)} + o(1), \tag{2.78}$$

which together with (2.16) yields the claim.

## §2.3.2 Proof of Theorem 2.1.4

*Proof.* The microcanonical ensemble is easy: the number of graphs with a fixed fraction $\lambda \in (0,1)$ of links is

$$\Omega_{L^*} = \binom{\binom{n}{2}}{L^*} = \binom{K}{\lambda K}, \qquad K = \binom{n}{2}. \tag{2.79}$$

The canonical ensemble has the Hamiltonian $H(\mathbf{G},\theta) = \theta L(\mathbf{G})$, where $\mathbf{G}$ is a graph belonging to $\mathcal{G}_n$, and $L(\mathbf{G}) = \sum_{1\le i<j\le n} g_{i,j}(\mathbf{G})$ is the number of links in $\mathbf{G}$. The partition function equals

$$Z(\theta) = \sum_{\mathbf{G}\in\mathcal{G}_n} e^{-H(\mathbf{G},\theta)} = \sum_{\mathbf{G}\in\mathcal{G}_n}\prod_{1\le i<j\le n} e^{-\theta g_{i,j}(\mathcal{G})} = \prod_{1\le i<j\le n}(1+e^{-\theta}). \tag{2.80}$$

The canonical probability equals

$$\begin{aligned} P_{\text{can}}(\mathbf{G} \mid \theta) &= \frac{e^{-\sum_{1\le i<j\le n}\theta g_{i,j}(\mathbf{G})}}{Z(\theta)} = \prod_{1\le i<j\le n}\frac{e^{-\theta g_{i,j}(\mathbf{G})}}{1+e^{-\theta}} \\ &= \prod_{1\le i<j\le n} p^{g_{i,j}(\mathbf{G})}(1-p)^{1-g_{i,j}(\mathbf{G})} \end{aligned} \tag{2.81}$$

with $p = \frac{e^{-\theta}}{1+e^{-\theta}}$. We search for $\theta^*$ such that

$$L^* = \sum_{1\le i<j\le n} p^*, \qquad p^* = \frac{e^{-\theta^*}}{1+e^{-\theta^*}}. \tag{2.82}$$

It follows that $p^* = \lambda$. Thus,

$$\begin{aligned} \log P_{\text{mic}}(\mathbf{G}^*) &= -\log(K)! + \log(\lambda K)! + \log((1-\lambda)K)! \\ &= -K[\log K - 1] + \lambda K[\log \lambda K - 1] \\ &\quad + [(1-\lambda)K][\log((1-\lambda)K) - 1] + o(n) \\ &= K\log(1-\lambda) + \lambda K \log\left(\frac{\lambda}{1-\lambda}\right) + o(n), \\ \log P_{\text{can}}(\mathbf{G}^*) &= \lambda K \log(\lambda) + (1-\lambda)K\log(1-\lambda). \end{aligned} \tag{2.83}$$

This in turn implies that

$$\lim_{n\to\infty}\frac{S_n(P_{\text{mic}} \mid P_{\text{can}})}{n} = 0. \tag{2.84}$$

## §2.3.3 Proof of Theorem 2.1.5

*Proof.* We start by describing the canonical ensemble. The Hamiltonian is

$$\begin{aligned} H(\mathbf{G}|\vec{\theta},\vec{\phi}) &= \sum_{i\in\Lambda_1} k_i(\mathbf{G})\theta_i + \sum_{j\in\Lambda_2} k'_j(\mathbf{G})\phi_j \\ &= \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\theta_i g_{i,j}(\mathbf{G}) + \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\phi_j g_{i,j}(\mathbf{G}) = \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}(\theta_i+\phi_j)g_{i,j}(\mathbf{G}). \end{aligned} \tag{2.85}$$

The partition function is

$$Z(\vec{\theta},\vec{\phi}) = \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}} e^{-\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}(\theta_i+\phi_j)g_{i,j}(\mathbf{G})} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2}\left(1+e^{-(\theta_i+\phi_j)}\right). \tag{2.86}$$

The canonical probability becomes

$$\begin{aligned} P_{\text{can}}(\mathbf{G} \mid \vec{\theta},\vec{\phi}) &= \frac{e^{-\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}(\theta_i+\phi_j)g_{i,j}(\mathbf{G})}}{Z(\vec{\theta},\vec{\phi})} \\ &= \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2}\frac{e^{-(\theta_i+\phi_j)g_{i,j}(\mathbf{G})}}{1+e^{-(\theta_i+\phi_j)}} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} p_{i,j}^{g_{i,j}(\mathbf{G})}(1-p_{i,j})^{1-g_{i,j}(\mathbf{G})}, \end{aligned} \tag{2.87}$$

where $p_{i,j} = \frac{e^{-(\theta_i+\phi_j)}}{1+e^{-(\theta_i+\phi_j)}}$. We search for $(\vec{\theta}^*, \vec{\phi}^*)$ that solves the system of equations

$$\begin{cases} \sum_{j\in\Lambda_2} p^*_{i,j} = k^*_i, \\ \sum_{i\in\Lambda_1} p^*_{i,j} = k'^*_j, \end{cases} \tag{2.88}$$

where $p^*_{i,j} = \frac{e^{-(\theta^*_i+\phi^*_j)}}{1+e^{-(\theta^*_i+\phi^*_j)}}$. If $\mathbf{G}^*$ is any graph in $\mathcal{G}_{n_1,n_2}$ such that $k_i(\mathbf{G}^*) = k^*_i$ and $k'_j(\mathbf{G}^*) = k'^*_j$, then

$$P_{\rm can}(\mathbf{G}) = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} p^{*\,g_{i,j}(\mathbf{G})}_{i,j}(1-p^*_{i,j})^{1-g_{i,j}(\mathbf{G})}. \tag{2.89}$$

Under the sparseness condition (2.27), we can replace $p^*_{i,j}$ with the following quantity. Define $\widehat{p}^*_{i,j} = e^{-(\theta^*_i+\phi^*_j)}$ and consider the system of equations

$$\begin{cases} \sum_{j\in\Lambda_2} \widehat{p}^*_{i,j} = k^*_i, \\ \sum_{i\in\Lambda_1} \widehat{p}^*_{i,j} = k'^*_j. \end{cases} \tag{2.90}$$

This has solution

$$\widehat{p}^*_{i,j} = \frac{k^*_i k'^*_j}{L^*}, \qquad L^* = \sum_{i\in\Lambda_1} k^*_i = \sum_{j\in\Lambda_2} k'^*_j. \tag{2.91}$$

We define

$$\widehat{P}_{\rm can}(\mathbf{G}) = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} (\widehat{p}^*_{i,j})^{g_{ij}(\mathbf{G})}(1-\widehat{p}^*_{i,j})^{1-g_{ij}(\mathbf{G})}, \tag{2.92}$$

and note that

$$\frac{1}{n_1+n_2}\log\left(\frac{\widehat{P}_{\rm can}(\mathbf{G})}{P_{\rm can}(\mathbf{G})}\right) \to 0, \qquad n_1, n_2 \to \infty. \tag{2.93}$$

The crucial point is to prove that $\frac{1}{n_1+n_2}\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\widehat{p}^{*\,2}_{i,j} \to 0$. This allows us to write

$$\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\log(1-p^*_{i,j}) = -\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\frac{k^*_i k'^*_j}{L^*} + o(n_1+n_2), \qquad n_1, n_2 \to \infty. \tag{2.94}$$

Indeed,

$$0 \le \frac{1}{n_1+n_2}\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\widehat{p}^{*\,2}_{i,j} = \frac{1}{n_1+n_2}\frac{\sum_{i\in\Lambda_1}k^{*\,2}_i\sum_{j\in\Lambda_2}k'^{*\,2}_j}{\sum_{i\in\Lambda_1}k^*_i\sum_{j\in\Lambda_2}k'^*_j} \le \frac{m^*m'^*}{\sqrt{n_1n_2}}\frac{\sqrt{n_1n_2}}{n_1+n_2} \to 0, \tag{2.95}$$

because $m^*m'^* = o(L^{*\,2/3})$ implies $m^*m'^* = o(\sqrt{n_1n_2})$.

Combining (2.89) and (2.94), we have

$$\begin{aligned} \log P_{\rm can}(\mathbf{G}^*) &= \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2} g_{i,j}(\mathbf{G}^*)\log\left(\frac{k^*_i k'^*_j}{L^*}\right) - \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\frac{k^*_i k'^*_j}{L^*} + o(n_1+n_2) \\ &= \sum_{i\in\Lambda_1} k^*_i\log\left(k^*_i\right) + \sum_{j\in\Lambda_2} k'^*_j\log\left(k'^*_j\right) - L^*\log L^* - L^* + o(n_1+n_2), \end{aligned} \tag{2.96}$$

which concludes our computation for the canonical ensemble.

Microcanonical probabilities come from the results in [55], where it is shown that, as $n \to \infty$, the number of bi-partite graphs with degree sequences $\vec{k^*}, \vec{k'^*}$ on the two layers is given by

$$\Omega_{\vec{k^*},\vec{k'^*}} = \frac{L^*!}{\prod_{i\in\Lambda_1} k_i^*! \prod_{j\in\Lambda_2} k_j'^*!}\, e^{o(n_1+n_2)}. \tag{2.97}$$

Hence

$$\log P_{\mathrm{mic}}(\mathbf{G}^*) = -\log \Omega_{\vec{s^*},\vec{t^*}} = \sum_{i\in\Lambda_1} k_i^*! + \sum_{j\in\Lambda_2} k'^*_j! - \log(L^*!) + o(n_1+n_2). \tag{2.98}$$

From (2.96) and (2.98) we get

$$\begin{aligned} S_{n_1+n_2}(P_{\mathrm{can}} \mid P_{\mathrm{mic}}) &= \log P_{\mathrm{mic}}(\mathbf{G}^*) - \log P_{\mathrm{can}}(\mathbf{G}^*) \\ &= \sum_{i\in\Lambda_1} \log\left(\frac{k_i^*!}{k_i^{*k_i^*}}\right) + \sum_{j\in\Lambda_2} \log\left(\frac{k_j'^*!}{k_j'^{*k_j'^*}}\right) \\ &\quad + L^*\log L^* + L^* - \log(L^*!) + o(n_1+n_2) \\ &= \sum_{i\in\Lambda_1} g(k_i^*) + \sum_{j\in\Lambda_2} g(k_j'^*) + o(n_1+n_2), \end{aligned} \tag{2.99}$$

where in the last line we use $L^* = \sum_{i\in\Lambda_1} k_i^* = \sum_{j\in\Lambda_2} k'^*_j$ and Stirling's approximation for $\log(L^*!)$. Since

$$\begin{aligned} {n_1}^{-1} \sum_{i\in\Lambda_1} g(k_i^*) &= \sum_{k\in\mathbb{N}_0}^{n_2} f_{1\to2}^{(n_1)}(k) g(k) = \|f_{1\to2}^{(n_1)}\|_{\ell^1(g)}, \\ {n_2}^{-1} \sum_{j\in\Lambda_2} g(k_j'^*) &= \sum_{k\in\mathbb{N}_0}^{n_1} f_{2\to1}^{(n_2)}(k) g(k) = \|f_{2\to1}^{(n_2)}\|_{\ell^1(g)}, \end{aligned} \tag{2.100}$$

we get, with the help of (2.28),

$$\lim_{n\to\infty} \frac{S_{n_1+n_2}(P_{\mathrm{can}} \mid P_{\mathrm{mic}})}{n_1+n_2} = A_1 \|f_{1\to2}\|_{\ell^1(g)} + A_2 \|f_{2\to1}\|_{\ell^1(g)}, \tag{2.101}$$

which proves the claim.

## §2.3.4 Proof of Theorem 2.1.6

*Proof.* The number of bi-partite graphs with constraint $\vec{k}^*$ on the top layer is

$$\Omega_{\vec{k}^*} = \prod_{i\in\Lambda_1} \binom{n_2}{k_i^*}. \tag{2.102}$$

In order to calculate the canonical probability, we calculate the partition function:

$$\begin{aligned} Z(\vec{\theta}) &= \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}} e^{-\sum_{i\in\Lambda_1}\theta_i \sum_{j\in\Lambda_2} g_{i,j}(\mathbf{G})} \\ &= \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}} \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} e^{-\theta_i g_{i,j}(\mathbf{G})} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} [1+e^{-\theta_i}]. \end{aligned} \tag{2.103}$$

The canonical probability becomes

$$\begin{aligned} P_{\mathrm{can}}(\mathbf{G}|\vec{\theta}) &= \frac{e^{-\sum_{i\in\Lambda_1}\theta_i\sum_{j\in\Lambda_2}g_{i,j}(\mathbf{G})}}{Z(\vec{\theta})} \\ &= \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2}\frac{e^{-\theta_i g_{i,j}(\mathbf{G})}}{1+e^{-\theta_i}} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} p_i^{g_{i,j}(\mathbf{G})}(1-p_i)^{1-g_{i,j}(\mathbf{G})} \end{aligned} \tag{2.104}$$

with $p_i = \frac{e^{-\theta_i}}{1+e^{-\theta_i}}$. We search for $\theta_i^*$ such that

$$k_i^* = \sum_{j\in\Lambda_2} p_i^* = n_2 p_i^*, \qquad p_i^* = \frac{e^{-\theta_i^*}}{1+e^{-\theta_i^*}}. \tag{2.105}$$

It follows that $p_i = \frac{k_i^*}{n_2}$ (recall (2.34)). According to (2.11) we have

$$\begin{aligned} S_{n_1+n_2}(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) &= \ln\frac{P_{\mathrm{mic}}(\mathbf{G}^*)}{P_{\mathrm{can}}(\mathbf{G}^*)} \\ &= -\sum_{i\in\Lambda_1}\log\binom{n_2}{k_i^*} - \sum_{i\in\Lambda_1} k_i^*\log\left(\frac{k_i^*}{n_2}\right) \\ &\quad - \sum_{i\in\Lambda_1}(n_2-k_i^*)\log\left(1-\frac{k_i^*}{n_2}\right) \\ &= n_1 n_2\log n_2 - \sum_{i\in\Lambda_1}\log\left[\binom{n_2}{k_i^*}{k_i^*}^{k_i^*}(n_2-k_i^*)^{(n_2-k_i^*)}\right]. \end{aligned} \tag{2.106}$$

Abbreviate $U_a(x) \equiv \log\left[\binom{a}{x}x^x(a-x)^{a-x}\right]$ and write

$$S_{n_1+n_2}(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = n_1 n_2\log n_2 - \sum_{i\in\Lambda_1} U_{n_2}(k_i^*) = n_1 n_2\log n_2 - n_1\sum_{k=0}^{n_2} f_{n_1}(k)U_{n_2}(k). \tag{2.107}$$

For the relative entropy per node this gives

$$\begin{aligned} s_{n_1+n_2} &= \frac{n_1}{n_1+n_2}\sum_{k=0}^{n_2} f_{n_1}(k)\, n_2\log n_2 - \frac{n_1}{n_1+n_2}\sum_{k=0}^{n_2} f_{n_1}(k)U_{n_2}(k) \\ &= -\frac{n_1}{n_1+n_2}\sum_{k=0}^{n_2} f_{n_1}(k)\log\mathrm{Bin}\big(n_2,\tfrac{k}{n_2}\big)(k) = \frac{n_1}{n_1+n_2}\|f_{n_1}\|_{\ell^1(g_{n_2})}. \end{aligned} \tag{2.108}$$

**Case (1).** Recall (2.17). Note that $x \mapsto z(x) = e^{g(x)}$ is non-decreasing:

$$\frac{z(x-1)}{z(x)} = \left(\frac{x}{x-1}\right)^{x-1}\frac{1}{e} \le 1. \tag{2.109}$$

It therefore follows that

$$\begin{aligned}\|f_{n_1}\|_{\ell^1(g_{n_2})} &= -\sum_{k=0}^{n_2} f_{n_1}(k)\log \mathrm{Bin}\big(n_2,\tfrac{k}{n_2}\big)(k) = \sum_{k=0}^{n_2} f_{n_1}(k)\log\left(\frac{z(k)z(n_2-k)}{z(n_2)}\right)\\ &= \sum_{k\in\mathbb{N}_0} f_{n_1}(k)\log\left(\frac{z(k)z(n_2-k)}{z(n_2)}\right)\mathbb{I}_{k\le n_2} \le \sum_{k\in\mathbb{N}_0}\mathbb{I}_{0\le k\le n_2} f_{n_1}(k)\log z(k)\\ &\le \sum_{k\in\mathbb{N}_0} f_{n_1}(k)\log z(k) = \|f_{n_1}\|_{\ell^1(g)} < \infty.\end{aligned} \tag{2.110}$$

By (2.36) and dominated convergence, we may exchange limit and sum to obtain

$$\lim_{n\to\infty} s_{n_1,n_2} = \lim_{n_2\to\infty}\frac{n_1}{n_1+n_2}\sum_{k\in\mathbb{N}_0} f_{n_1}(k)\lim_{n_2\to\infty}\log\left(\frac{z(k)z(n_2-k)}{z(n_2)}\right)\mathbb{I}_{0\le k\le n_2} = 0, \tag{2.111}$$

where we use that $\lim_{n\to\infty}\frac{n_1}{n_1+n_2}=0$ and $\lim_{n\to\infty}\frac{z(n_2-k)}{z(n_2)}=1$ for all $k\in\mathbb{N}_0$.

**Case (2).** Using (2.110) and (2.36), we get

$$0 \le s_{n_1,n_2} = \frac{n_1}{n_1+n_2}\|f_{n_1}\|_{\ell^1(g)} \to^{n\to\infty} \frac{1}{1+c}\|f\|_{\ell^1(g)} = 0. \tag{2.112}$$

**Case (3).** Estimate

$$0 \le |\|f_{n_1}\|_{\ell^1(g_{n_2})} - \|f_{n_1}\|_{\ell^1(g_{n_2})}| \le \|f_{n_1}-f\|_{\ell^1(g_{n_2})} \le \|f_{n_1}-f\|_{\ell^1(g)} \to^{n\to\infty} 0. \tag{2.113}$$

**Case (4).**

$$0 \le |\|f_{n_1}\|_{\ell^1(g_{n_2})} - \|f\|_{\ell^1(g)}| \le \sum_{k\in\mathbb{N}_0}|f_{n_1}(k)-f(k)||g_{n_2}(k)\mathbb{I}_{k\le n_2}-g(k)| \le 2\|f_{n_1}-f\|_{\ell^1(g)}. \tag{2.114}$$

Since $\frac{n_1}{n_1+n_2} = \frac{1}{1+\frac{n_2}{n_1}} \to \frac{1}{1+c}$, the claim follows.

## §2.3.5 Proof of Theorem 2.1.7

*Proof.* The microcanonical ensemble is easy: the number of bi-partite graphs with a fixed fraction $\lambda\in(0,1)$ of links is

$$\Omega_{L^*} = \binom{n_1n_2}{L^*} = \binom{n_1n_2}{\lambda n_1n_2}. \tag{2.115}$$

The canonical ensemble has the Hamiltonian $H(\mathbf{G},\theta)=\theta L(\mathbf{G})$, where $\mathbf{G}$ is a bi-partite graph belonging to $\mathcal{G}_{n_1,n_2}$, and $L(\mathbf{G})=\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2} g_{i,j}(\mathbf{G})$ is the number of links in $\mathbf{G}$. The partition function equals

$$Z(\theta) = \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}} e^{-H(\mathbf{G},\theta)} = \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}}\prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} e^{-\theta g_{i,j}(\mathcal{G})} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2}(1+e^{-\theta}). \tag{2.116}$$

The canonical probability equals

$$\begin{aligned} P_{\mathrm{can}}(\mathbf{G} \mid \vec{\theta}) &= \frac{e^{-\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\theta g_{i,j}(\mathbf{G})}}{Z(\vec{\theta})} \\ &= \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2}\frac{e^{-\theta g_{i,j}(\mathbf{G})}}{1+e^{-\theta}} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} p^{g_{i,j}(\mathbf{G})}(1-p)^{1-g_{i,j}(\mathbf{G})} \end{aligned} \tag{2.117}$$

with $p = \frac{e^{-\theta}}{1+e^{-\theta}}$. We search for $\theta^*$ such that

$$L^* = \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2} p^*, \qquad p^* = \frac{e^{-\theta^*}}{1+e^{-\theta^*}}. \tag{2.118}$$

It follows that $p^* = \lambda$. Thus,

$$\begin{aligned} \log P_{\mathrm{mic}}(\mathbf{G}^*) &= -\log(n_1n_2)! + \log(\lambda n_1 n_2)! + \log((1-\lambda)n_1n_2)! \\ &= -n_1n_2[\log n_1n_2 - 1] + \lambda n_1n_2[\log \lambda n_1 n_2 - 1] \\ &\quad + [1-\lambda n_1 n_2][\log((1-\lambda)n_1n_2) - 1] + o(n_1+n_2) \\ &= n_1n_2\log(1-\lambda) + \lambda n_1n_2 \log\left(\frac{\lambda}{1-\lambda}\right) + o(n_1+n_2), \\ \log P_{\mathrm{can}}(\mathbf{G}^*) &= n_1n_2\log(1-\lambda) + \lambda n_1n_2\log\left(\frac{\lambda}{1-\lambda}\right). \end{aligned} \tag{2.119}$$

This in turn implies that

$$\lim_{n_1,n_2\to\infty}\frac{S_{n_1+n_2}(P_{\mathrm{mic}} \mid P_{\mathrm{can}})}{n_1+n_2} = 0. \tag{2.120}$$

## §2.3.6 Proof of Theorem 2.1.8

*Proof.* The proof is based on the previous theorems. We start by looking at the Hamiltonian of the system. For each admitted pair of layers ($\gamma_{s,t}(\mathbf{\Gamma}) = 1$) we define Lagrange multipliers $\vec{\theta}_{s\to t} = (\theta_1^{(t)}, \dots, \theta_{n_s}^{(t)})$. The Hamiltonian equals

$$\begin{aligned} &H\big(\mathbf{G} \mid \vec{\theta}_{s\to t};\ s,t = 1,\dots,M,\ \gamma_{s,t}(\mathbf{\Gamma}) = 1\big) \\ &= \sum_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}}\sum_{\substack{i\in\Lambda_s\\ j\in\Lambda_t}}(\theta_i^t + \theta_j^s)g_{i,j}(\mathbf{G}) + \sum_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M}\sum_{\substack{i,j\in\Lambda_s\\ i<j}}(\theta_i^s+\theta_j^s)g_{i,j}(\mathbf{G}) \\ &= \sum_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}}\sum_{\substack{i\in\Lambda_s\\ j\in\Lambda_t}} H_{s,t}(\mathbf{G}^{(st)} \mid \vec{\theta}_{s\to t}, \vec{\theta}_{t\to s}) + \sum_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M}\sum_{\substack{i,j\in\Lambda_s\\ i<j}} H_{s,s}(\mathbf{G}^{(ss)} \mid \vec{\theta}_{s\to s}), \end{aligned} \tag{2.121}$$

where

$$\begin{aligned} H_{s,t}(\mathbf{G}^{(st)} \mid \vec{\theta}_{s\to t}, \vec{\theta}_{t\to s}) &= \sum_{\substack{i\in\Lambda_s \\ j\in\Lambda_t}} (\theta_i^t + \theta_j^s) g_{i,j}(\mathbf{G}^{(st)}), \\ H_{s,s}(\mathbf{G}^{(ss)} \mid \vec{\theta}_{s\to s}) &= \sum_{\substack{i,j\in\Lambda_s \\ i<j}} (\theta_i^s + \theta_j^s) g_{i,j}(\mathbf{G}^{(ss)}), \end{aligned} \tag{2.122}$$

and $\mathbf{G}^{(st)}$ ($\mathbf{G}^{(ss)}$) is the bi-partite (uni-partite) graph between layers $s$ and $t$ (inside layer $s$) obtained from the multi-partite graph $\mathbf{G}$. The $n_s \times n_t$ matrix representing the bi-partite graph has, for each $i \in \Lambda_s$ and $j \in \Lambda_t$, elements $g_{i,j}(\mathbf{G}^{(st)}) = g_{i,j}(\mathbf{G})$. Note that $H_{s,t}(\mathbf{G}^{(st)} \mid \vec{\theta}_{s\to t}, \vec{\theta}_{t\to s})$ is the Hamiltonian of the bi-partite graph $\mathbf{G}^{(st)}$ between layers $s$ and $t$ with constraints $\vec{k}^*_{s\to t}$, and $H_{s,s}(\mathbf{G}^{(ss)} \mid \vec{\theta}_{s\to s})$ is the Hamiltonian of the uni-partite graph $\mathbf{G}^{(ss)}$ of the layer $s$ with constraints $\vec{k}^*_{s\to s}$.

The partition function of the canonical ensemble equals

$$\begin{aligned} &Z\big(\vec{\theta}_{s\to t};\ s,t=1,\dots,M,\ \gamma_{s,t}(\mathbf{\Gamma})=1\big) \\ &= \sum_{\mathbf{G}\in\mathcal{G}_{n_1,\dots,n_M}(\mathbf{\Gamma})} e^{-H(\mathbf{G}\,|\,\vec{\theta}_{s\to t};\ s,t=1,2,\dots,M:\ \gamma_{s,t}(\mathbf{\Gamma})=1)} \\ &= \prod_{\substack{1\le s<t\le M \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \sum_{\mathbf{G}^{(st)}\in\mathcal{G}_{n_s,n_t}} e^{-H_{s,t}(\mathbf{G}^{(st)}\,|\,\vec{\theta}_{s\to t},\vec{\theta}_{t\to s})} \prod_{\substack{s=1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \sum_{\mathbf{G}^{(ss)}\in\mathcal{G}_{n_s,n_s}} e^{-H_{s,s}(\mathbf{G}^{(ss)}\,|\,\vec{\theta}_{s\to s})} \\ &= \prod_{\substack{1\le s<t\le M \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} Z^{(st)}(\vec{\theta}_{s\to t}, \vec{\theta}_{t\to s}) \prod_{\substack{s=1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} Z^{(ss)}(\vec{\theta}_{s\to s}), \end{aligned} \tag{2.123}$$

where $Z^{(st)}(\vec{\theta}_{s\to t}, \vec{\theta}_{t\to s})$ is the partition function of the set of bi-partite graphs $\mathcal{G}_{n_s,n_t}$ with constraints $\vec{k}^*_{s\to t}$ on the top layer and $\vec{k}^*_{t\to s}$ on the bottom layer, and $Z^{(ss)}(\vec{\theta}_{s\to s})$ is the partition function of the set of graph $\mathcal{G}_{n_s}$ with constraint $\vec{k}^*_{s\to s}$. The canonical ensemble is

$$P_{\text{can}}(\mathbf{G}) = \prod_{\substack{1\le s<t\le M \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} P^{(st)}_{\text{can}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} P^{(ss)}_{\text{can}}(\mathbf{G}^{(ss)}), \tag{2.124}$$

where $P^{(st)}_{\text{can}}(\mathbf{G}^{(st)})$ is the canonical probability of the bi-partite graph $\mathbf{G}^{(st)}$ with constraints $\vec{k}^*_{s\to t}$ on the top layer and $\vec{k}^*_{t\to s}$ on the bottom layer, and $P^{(ss)}_{\text{can}}(\mathbf{G}^{(ss)})$ is the canonical probability of the uni-partite graph $\mathbf{G}^{(ss)}$ with constraint $\vec{k}^*_{s\to s}$.

We can split the microcanonical probability as products of microcanonical prob-

abilities for simpler cases. The number of graphs with constraints $\vec{C}^*$ is

$$\begin{aligned}
&\Omega_{\vec{k}^*_{s\to t};\, s,t\in\{1,\dots,M\},\ \gamma_{s,t}(\mathbf{\Gamma})=1} \\
&= \left| \left\{ \mathbf{G} \in \mathcal{G}_{n_1,\dots,n_M}(\mathbf{\Gamma}) : \sum_{j\in\Lambda_t} g_{i,j}(\mathbf{G}) = k_i^{*\,t}\ \forall\, i \in \Lambda_s\, \forall\, \gamma_{s,t} = 1 \right\} \right| \\
&= \prod_{\substack{1\le s<t\le M \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} |A_{st} \cap B_{st}| \prod_{\substack{s=1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \left| \left\{ \mathbf{G}^{(ss)} \in \mathcal{G}_{n_s} : \sum_{j\in\Lambda_s} g_{i,j}(\mathbf{G}^{(ss)}) = s_i^{*\,h}\ \forall\, i \in \Lambda_s \right\} \right| \\
&= \prod_{\substack{1\le s<t\le M \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \Omega_{\vec{k}^*_{s\to t},\vec{k}^*_{t\to s}} \prod_{\substack{s=1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \Omega_{\vec{k}^*_{s\to s}},
\end{aligned} \tag{2.125}$$

where $A_{st} = \left\{ \mathbf{G}^{(st)} \in \mathcal{G}_{n_s,n_k} : \sum_{j\in\Lambda_t} g_{i,j}(\mathbf{G}^{(st)}) = k_i^{*\,t}\ \forall\, i \in \Lambda_s \right\}$ and $B_{st} = \left\{ \mathbf{G}^{(st)} \in \mathcal{G}_{n_s,n_k} : \sum_{i\in\Lambda_s} g_{i,j}(\mathbf{G}^{(st)}) = k_j^{*\,s}\ \forall\, j \in \Lambda_t \right\}$.

This means the microcanonical probability can be factorised as

$$P_{\mathrm{mic}}(\mathbf{G}) = \prod_{\substack{1\le s<t\le M \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} P^{(st)}_{\mathrm{mic}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} P^{(ss)}_{\mathrm{mic}}(\mathbf{G}^{(ss)}), \tag{2.126}$$

where $P^{(st)}_{\mathrm{mic}}(\mathbf{G}^{(st)})$ is the microcanonical probability of the bi-partite graph $\mathbf{G}^{(st)}$ with constraints $\vec{k}^*_{s\to t}$ on the top layer and $\vec{k}^*_{t\to s}$ on the bottom layer, and $P^{(ss)}_{\mathrm{mic}}(\mathbf{G}^{(ss)})$ is the microcanonical probability of the uni-partite graph $\mathbf{G}^{(ss)}$ with constraint $\vec{k}^*_{s\to s}$.

Equations (2.124) and (2.126) imply that the relative entropy equals the sum

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = \sum_{\substack{1\le s<t\le M \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} S_n(P^{(st)}_{\mathrm{mic}} \mid P^{(st)}_{\mathrm{can}}) + \sum_{\substack{s=1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} S_n(P^{(ss)}_{\mathrm{mic}} \mid P^{(ss)}_{\mathrm{can}}). \tag{2.127}$$

We can now apply Theorems 2.1.1 and 2.1.5 to get the asymptotic relative entropy

per nodes as

$$\begin{aligned}
&\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\text{mic}} \mid P_{\text{can}})}{n} \\
&= \sum_{\substack{1\le s<t\le M \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{(st)}_{\text{mic}} \mid P^{(st)}_{\text{can}})}{n} + \sum_{\substack{s=1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{(ss)}_{\text{mic}} \mid P^{(ss)}_{\text{can}})}{n} \\
&= \sum_{\substack{1\le s<t\le M \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \left\{A_s \left\|f_{s\to t}\right\|_{\ell^1(g)} + A_t \left\|f_{t\to s}\right\|_{\ell^1(g)}\right\} + \sum_{\substack{s=1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \left\{A_s \left\|f_{s\to s}\right\|_{\ell^1(g)}\right\} \\
&= \sum_{\substack{s,t=1 \\ \gamma_{s,t}(\mathbf{\Gamma})}}^{M} A_s \left\|f_{s\to t}\right\|_{\ell^1(g)}.
\end{aligned} \tag{2.128}$$

## §2.3.7 Proof of Theorem 2.1.9

*Proof.* We start by studying the Hamiltonian. For each pair $(s,t)$ of layers in $\mathcal{D}$, we define Lagrange multipliers $\vec{\theta}_{s\to t} = (\theta^t_1, \dots, \theta^t_{n_s})$. For each pair $(s,t)$ of layers in $\mathcal{L}$, we define a Lagrange multiplier $\theta_{s,t}$. The Hamiltonian is

$$\begin{aligned}
&H\big(\mathbf{G} \mid \vec{\theta}_{s\to t}, \theta_{l,m};\, (s,t)\in\mathcal{D}, (l,m)\in\mathcal{L}\big) \\
&= H_{\mathcal{D}}(\mathbf{G} \mid \vec{\theta}_{s\to t};\, (s,t)\in\mathcal{D}) + H_{\mathcal{L}}(\mathbf{G} \mid \theta_{l,m};\, (l,m)\in\mathcal{L})
\end{aligned} \tag{2.129}$$

with

$$\begin{aligned}
H_{\mathcal{D}}(\mathbf{G} \mid \vec{\theta}_{s\to t};\, (s,t)\in\mathcal{D}) &= \sum_{\substack{1\le s<t\le M \\ (s,t)\in\mathcal{D}}} \sum_{\substack{i\in\Lambda_s \\ j\in\Lambda_t}} (\theta^t_i + \theta^s_j) g_{i,j}(\mathbf{G}) \\
&\quad + \sum_{\substack{s=1 \\ (s,s)\in\mathcal{D}}}^{M} \sum_{\substack{i,j\in\Lambda_s \\ i<j}} (\theta^s_i + \theta^s_j) g_{i,j}(\mathbf{G}), \\
H_{\mathcal{L}}(\mathbf{G} \mid \theta_{s,t};\, (s,t)\in\mathcal{L}) &= \sum_{\substack{1\le s<t\le M \\ (s,t)\in\mathcal{L}}} \sum_{\substack{i\in\Lambda_s \\ j\in\Lambda_t}} (\theta_{s,t}) g_{i,j}(\mathbf{G}) + \sum_{\substack{s=1 \\ (s,s)\in\mathcal{L}}}^{M} \sum_{\substack{i,j\in\Lambda_s \\ i<j}} (\theta_{s,s}) g_{i,j}(\mathbf{G}).
\end{aligned} \tag{2.130}$$

Consequently, the canonical ensemble is

$$P_{\text{can}}(\mathbf{G}) = P^{\mathcal{D}}_{\text{can}}(\mathbf{G}) P^{\mathcal{L}}_{\text{can}}(\mathbf{G}) \tag{2.131}$$

with

$$\begin{aligned} P^{\mathcal{D}}_{\rm can}(\mathbf{G}) &= \prod_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{D}}} {P^{(st)}_{\rm can}}^{\mathcal{D}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\(s,s)\in\mathcal{D}}}^{M} {P^{(ss)}_{\rm can}}^{\mathcal{D}}(\mathbf{G}^{(ss)}),\\ P^{\mathcal{L}}_{\rm can}(\mathbf{G}) &= \prod_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{L}}} {P^{(st)}_{\rm can}}^{\mathcal{L}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\(s,s)\in\mathcal{L}}}^{M} {P^{(ss)}_{\rm can}}^{\mathcal{L}}(\mathbf{G}^{(ss)}). \end{aligned} \tag{2.132}$$

Here,

- $\mathbf{G}^{(st)}$ ($\mathbf{G}^{(ss)}$ ) is the bi-partite (uni-partite) graph between layers $s$ and $t$ (and itself) obtained from the multi-partite graph $\mathbf{G}$. The $n_s \times n_t$ ($n_s \times n_s$) matrix representing this bi-partite (uni-partite) graph has, for each $i \in \Lambda_s$ and $j \in \Lambda_t$ (for each $i, j \in \Lambda_s$), elements $g_{i,j}(\mathbf{G}^{(st)}) = g_{i,j}(\mathbf{G})$ ($g_{i,j}(\mathbf{G}^{(ss)}) = g_{i,j}(\mathbf{G})$).
- ${P^{(st)}_{\rm can}}^{\mathcal{D}}(\mathbf{G}^{(st)})$ (${P^{(ss)}_{\rm can}}^{\mathcal{D}}(\mathbf{G}^{(ss)})$) is the canonical probability of the bi-partite (uni-partite) graph $\mathbf{G}^{(st)}$ ($\mathbf{G}^{(ss)}$) with constraints $\vec{k}^{*}_{s\to t}$ on the top layer and $\vec{k}^{*}_{t\to s}$ on the bottom layer (with constraint $\vec{k}^{*}_{s\to s}$).
- ${P^{(st)}_{\rm can}}^{\mathcal{L}}(\mathbf{G}^{(st)})$ (${P^{(ss)}_{\rm can}}^{\mathcal{L}}(\mathbf{G}^{(ss)})$) is the canonical probability of the bi-partite (uni-partite) graph $\mathbf{G}^{(st)}$ ($\mathbf{G}^{(ss)}$) with constraint $L^{*}_{s,t}$ ($L^{*}_{s,s}$).

We can split the microcanonical probability as products of microcanonical probabilities of simpler cases. The number of graphs with such a type of constraints is

$$\Omega_{\vec{k}^{*}_{s\to t},L^{*}_{l,m};\,(s,t)\in\mathcal{D},(l,m)\in\mathcal{L}} = \Omega_{\vec{k}^{*}_{s\to t};\,(s,t)\in\mathcal{D}}\,\Omega_{L_{l,m};(l,m)\in\mathcal{L}}. \tag{2.133}$$

This means that the microcanonical probability can be factorised as

$$P_{\rm mic}(\mathbf{G}) = P^{\mathcal{D}}_{\rm mic}(\mathbf{G})P^{\mathcal{L}}_{\rm mic}(\mathbf{G}) \tag{2.134}$$

with

$$\begin{aligned} P^{\mathcal{D}}_{\rm mic}(\mathbf{G}) &= \prod_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{D}}} {P^{(st)}_{\rm mic}}^{\mathcal{D}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\(s,s)\in\mathcal{D}}}^{M} {P^{(ss)}_{\rm mic}}^{\mathcal{D}}(\mathbf{G}^{(ss)}),\\ P^{\mathcal{L}}_{\rm mic}(\mathbf{G}) &= \prod_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{L}}} {P^{(st)}_{\rm mic}}^{\mathcal{L}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\(s,s)\in\mathcal{L}}}^{M} {P^{(ss)}_{\rm mic}}^{\mathcal{L}}(\mathbf{G}^{(ss)}). \end{aligned} \tag{2.135}$$

Here,

- ${P^{(st)}_{\rm mic}}^{\mathcal{D}}(\mathbf{G}^{(st)})$ (${P^{(ss)}_{\rm mic}}^{\mathcal{D}}(\mathbf{G}^{(ss)})$) is the microcanonical probability of the bi-partite (uni-partite) graph $\mathbf{G}^{(st)}$ ($\mathbf{G}^{(ss)}$) with constraints $\vec{k}^{*}_{s\to t}$ on the top layer and $\vec{k}^{*}_{t\to s}$ on the bottom layer (with constraint $\vec{k}^{*}_{s\to s}$).
- ${P^{(st)}_{\rm mic}}^{\mathcal{L}}(\mathbf{G}^{(st)})$ (${P^{(ss)}_{\rm mic}}^{\mathcal{L}}(\mathbf{G}^{(ss)})$) is the microcanonical probability of the bi-partite (uni-partite) graph $\mathbf{G}^{(st)}$ ($\mathbf{G}^{(ss)}$) with constraint $L^{*}_{s,t}$ ($L^{*}_{s,s}$).

The relative entropy becomes

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) = S_n(P^{\mathcal{D}}_{\text{mic}} \mid P^{\mathcal{D}}_{\text{can}}) + S_n(P^{\mathcal{L}}_{\text{mic}} \mid P^{\mathcal{L}}_{\text{can}}). \tag{2.136}$$

It follows that

$$\begin{aligned} &\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\text{mic}} \mid P_{\text{can}})}{n} \\ &\quad= \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{\mathcal{D}}_{\text{mic}} \mid P^{\mathcal{D}}_{\text{can}})}{n} + \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{\mathcal{L}}_{\text{mic}} \mid P^{\mathcal{L}}_{\text{can}})}{n}. \end{aligned} \tag{2.137}$$

Using Theorem 2.1.8 we get

$$\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{\mathcal{D}}_{\text{mic}} \mid P^{\mathcal{D}}_{\text{can}})}{n} = \sum_{(s,t)\in\mathcal{D}} A_s \, \|f_{s\to t}\|_{\ell^1(g)}. \tag{2.138}$$

Moreover,

$$\begin{aligned} &\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{\mathcal{L}}_{\text{mic}} \mid P^{\mathcal{L}}_{\text{can}})}{n} \\ &= \lim_{n_1,\dots,n_M\to\infty} \sum_{\substack{1\le s<t\le M \\ (s,t)\in\mathcal{L}}} \frac{S_n(P^{(st)^{\mathcal{L}}}_{\text{mic}} \mid P^{(st)^{\mathcal{L}}}_{\text{can}})}{n} + \lim_{n_1,\dots,n_M\to\infty} \sum_{\substack{s=1 \\ (s,s)\in\mathcal{L}}}^{M} \frac{S_n(P^{(ss)^{\mathcal{L}}}_{\text{mic}} \mid P^{(ss)^{\mathcal{L}}}_{\text{can}})}{n}. \end{aligned} \tag{2.139}$$

Using Theorems 2.1.4 and 2.1.7, we get

$$\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{(st)^{\mathcal{L}}}_{\text{mic}} \mid P^{(st)^{\mathcal{L}}}_{\text{can}})}{n} = \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{(ss)^{\mathcal{L}}}_{\text{mic}} \mid P^{(ss)^{\mathcal{L}}}_{\text{can}})}{n} = 0, \tag{2.140}$$

which proves the claim.

## §2.3.8 Proof of Theorem 2.1.10

*Proof.* The proof is based on the previous theorems. For each pair of layers $s, t \in \mathcal{M}_1$ we define Lagrange multipliers $\vec{\theta}_{s\to t} = (\theta^t_1, \dots, \theta^t_{n_s})$ and $\vec{\theta}_{t\to s} = (\theta^s_1, \dots, \theta^s_{n_t})$. For each pair of layers $s, \in \mathcal{M}_1,\ t \in \mathcal{M}_2$ we define $\vec{\theta}_{s\to t} = (\theta^t_1, \dots, \theta^t_{n_s})$. The Hamiltonian is

$$\begin{aligned} &H\big(\mathbf{G} \mid \vec{\theta}_{s\to t};\ s \in \mathcal{M}_1,\ t \in \mathcal{M}_1 \cup \mathcal{M}_2,\ \gamma_{s,t}(\mathbf{\Gamma}) = 1\big) \\ &= \sum_{\substack{s,t\in\mathcal{M}_1 \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to t}\vec{s}_{s\to t}(\mathbf{G}) + \sum_{\substack{s\in\mathcal{M}_1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to s}\vec{s}_{s\to s}(\mathbf{G}) + \sum_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2 \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to t}\vec{s}_{s\to t}(\mathbf{G}) \\ &= H_{\mathcal{M}_1\to\mathcal{M}_1} + H_{\mathcal{M}_1\to\mathcal{M}_2}, \end{aligned} \tag{2.141}$$

with

$$\begin{aligned} H_{\mathcal{M}_1\to\mathcal{M}_1} &= \sum_{\substack{s,t\in\mathcal{M}_1 \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to t}\vec{s}_{s\to t}(\mathbf{G}) + \sum_{\substack{s\in\mathcal{M}_1 \\ \gamma_{s,s}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to s}\vec{s}_{s\to s}(\mathbf{G}), \\ H_{\mathcal{M}_1\to\mathcal{M}_2} &= \sum_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2 \\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to t}\vec{s}_{s\to t}(\mathbf{G}). \end{aligned} \tag{2.142}$$

Consequently, the canonical ensemble is

$$P_{\rm can}(\mathbf{G}) = P_{\rm can}^{\mathcal{M}_1\to\mathcal{M}_1}(\mathbf{G}) P_{\rm can}^{\mathcal{M}_1\to\mathcal{M}_2}(\mathbf{G}) \tag{2.143}$$

with

$$\begin{aligned} P_{\rm can}^{\mathcal{M}_1\to\mathcal{M}_1}(\mathbf{G}) &= \prod_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} {P_{\rm can}^{(st)}}^{top,bot}(\mathbf{G}^{(st)}) \prod_{\substack{s\in\mathcal{M}_1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}} P_{\rm can}^{(ss)}(\mathbf{G}^{(ss)}),\\ P_{\rm can}^{\mathcal{M}_1\to\mathcal{M}_2}(\mathbf{G}) &= \prod_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} {P_{\rm can}^{(st)}}^{top}(\mathbf{G}^{(st)}). \end{aligned} \tag{2.144}$$

Here,

- $\mathbf{G}^{(st)}$ ($\mathbf{G}^{(ss)}$) is the bi-partite (uni-partite) graph between layers $s$ and $t$ (itself) obtained from the multi-partite graph $\mathbf{G}$. The $n_s \times n_t$ ($n_s \times n_s$) matrix representing this bi-partite (uni-partite) graph has, for each $i \in \Lambda_s$ and $j \in \Lambda_t$ (for each $i, j \in s$), elements $g_{i,j}(\mathbf{G}^{(st)}) = g_{i,j}(\mathbf{G})$ ($g_{i,j}(\mathbf{G}^{(ss)}) = g_{i,j}(\mathbf{G})$).
- ${P_{\rm can}^{(st)}}^{top,bot}(\mathbf{G}^{(st)})$ is the canonical probability of the bi-partite graph $\mathbf{G}^{(st)}$ with constraints $\vec{k}^*_{s\to t}$ on the top layer and $\vec{k}^*_{t\to s}$ on the bottom layer.
- $P_{\rm can}^{(ss)}(\mathbf{G}^{(ss)})$ is the canonical probability of the uni-partite graph $\mathbf{G}^{(ss)}$ with constraint $\vec{k}^*_{s\to s}$.
- ${P_{\rm can}^{(st)}}^{top}(\mathbf{G}^{(st)})$ is the canonical probability of the bi-partite graph $\mathbf{G}^{(st)}$ with constraint $\vec{k}^*_{s\to t}$ on the top layer.

We can split the microcanonical probability as products of microcanonical probabilities for simpler cases. The number of graphs with such a type of constraints is

$$\Omega_{\vec{k}^*_{s\to t};\ s\in\mathcal{M}_1,\ t\in\mathcal{M}_1\cup\mathcal{M}_2,\ \gamma_{s,t}(\mathbf{\Gamma})=1} \tag{2.145}$$

$$= \Omega_{\vec{k}^*_{s\to t};\ s,t\in\mathcal{M}_1,\ \gamma_{s,t}(\mathbf{\Gamma})=1} \Omega_{\vec{k}^*_{s\to t};\ s\in\mathcal{M}_1,\ t\in\mathcal{M}_2,\ \gamma_{s,t}(\mathbf{\Gamma})=1}. \tag{2.146}$$

This means that the microcanonical probability can be factorised as

$$P_{\rm mic}(\mathbf{G}) = P_{\rm mic}^{\mathcal{M}_1\to\mathcal{M}_1}(\mathbf{G}) P_{\rm mic}^{\mathcal{M}_1\to\mathcal{M}_2}(\mathbf{G}) \tag{2.147}$$

with

$$\begin{aligned} P_{\rm mic}^{\mathcal{M}_1\to\mathcal{M}_1}(\mathbf{G}) &= \prod_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} {P_{\rm mic}^{(st)}}^{top,bot}(\mathbf{G}^{(st)}) \prod_{\substack{s\in\mathcal{M}_1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}} P_{\rm mic}^{(ss)}(\mathbf{G}^{(ss)}),\\ P_{\rm mic}^{\mathcal{M}_1\to\mathcal{M}_2}(\mathbf{G}) &= \prod_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} {P_{\rm mic}^{(st)}}^{top}(\mathbf{G}^{(st)}). \end{aligned} \tag{2.148}$$

Here,

- $P_{\text{mic}}^{(st)^{top,bot}}(\mathbf{G}^{(st)})$ is the microcanonical probability of the bi-partite graph $\mathbf{G}^{(st)}$ with constraints $\vec{k}^*_{s\to t}$ on the top layer and $\vec{k}^*_{t\to s}$ on the bottom layer.
- $P_{\text{mic}}^{(ss)}(\mathbf{G}^{(ss)})$ is the microcanonical probability of the uni-partite graph $\mathbf{G}^{(ss)}$ with constraint $\vec{k}^*_{s\to s}$.
- $P_{\text{mic}}^{(st)^{top}}(\mathbf{G}^{(st)})$ is the microcanonical probability of the bi-partite graph $\mathbf{G}^{(st)}$ with constraint $\vec{k}^*_{s\to t}$ on the top layer.

The relative entropy becomes

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) = S_n(P_{\text{mic}}^{\mathcal{M}_1\to\mathcal{M}_1} \mid P_{\text{can}}^{\mathcal{M}_1\to\mathcal{M}_1}) + S_n(P_{\text{mic}}^{\mathcal{M}_1\to\mathcal{M}_2} \mid P_{\text{can}}^{\mathcal{M}_1\to\mathcal{M}_2}). \tag{2.149}$$

It follows that

$$\begin{aligned}&\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\text{can}} \mid P_{\text{can}})}{n}\\ &= \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\text{mic}}^{\mathcal{M}_1\to\mathcal{M}_1} \mid P_{\text{can}}^{\mathcal{M}_1\to\mathcal{M}_1})}{n} + \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\text{mic}}^{\mathcal{M}_1\to\mathcal{M}_2} \mid P_{\text{can}}^{\mathcal{M}_1\to\mathcal{M}_2})}{n}.\end{aligned} \tag{2.150}$$

Using again Theorem 2.1.8 we get

$$\begin{aligned}&\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\text{mic}}^{\mathcal{M}_1\to\mathcal{M}_1} \mid P_{\text{can}}^{\mathcal{M}_1\to\mathcal{M}_1})}{n}\\ &= \sum_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \left\{A_s \left\|f_{s\to t}\right\|_{\ell^1(g)} + A_t \left\|f_{t\to s}\right\|_{\ell^1(g)}\right\} + \sum_{\substack{s\in\mathcal{M}_1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}} A_s \left\|f_{s\to s}\right\|_{\ell^1(g)}\\ &= \sum_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} A_s \left\|f_{s\to t}\right\|_{\ell^1(g)}.\end{aligned} \tag{2.151}$$

From Theorem 2.1.6 we get

$$\begin{aligned}&\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\text{mic}}^{\mathcal{M}_1\to\mathcal{M}_2} \mid P_{\text{can}}^{\mathcal{M}_1\to\mathcal{M}_2})}{n}\\ &= \lim_{n_1,\dots,n_M\to\infty} \sum_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \frac{S_n(P_{\text{mic}}^{(st)^{\mathcal{M}_1\to\mathcal{M}_2}} \mid P_{\text{can}}^{(st)^{\mathcal{M}_1\to\mathcal{M}_2}})}{n}\\ &= \sum_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} A_s \left\|f_{s\to l}\right\|_{\ell^1(g)},\end{aligned} \tag{2.152}$$

which concludes the proof.

CHAPTER 3

# Covariance structure behind breaking of ensemble equivalence in random graphs

This chapter is based on:
D. Garlaschelli, F. den Hollander, and A. Roccaverde. Covariance structure behind breaking of ensemble equivalence in random graphs. *J. Stat. Phys.*, Jul 2018


**Abstract**

For a random graph subject to a topological constraint, the *microcanonical ensemble* requires the constraint to be met by every realisation of the graph ('hard constraint'), while the *canonical ensemble* requires the constraint to be met only on average ('soft constraint'). It is known that *breaking of ensemble equivalence* may occur when the size of the graph tends to infinity, signalled by a non-zero specific relative entropy of the two ensembles. In this paper we analyse a formula for the relative entropy of generic discrete random structures recently put forward by Squartini and Garlaschelli. We consider the case of a random graph with a given degree sequence (configuration model), and show that in the dense regime this formula correctly predicts that the specific relative entropy is determined by the scaling of the determinant of the matrix of canonical covariances of the constraints. The formula also correctly predicts that an extra correction term is required in the sparse regime and in the ultra-dense regime. We further show that the different expressions correspond to the degrees in the canonical ensemble being asymptotically *Gaussian* in the dense regime and asymptotically *Poisson* in the sparse regime (the latter confirms what we found in earlier work), and the dual degrees in the canonical ensemble being asymptotically *Poisson* in the ultra-dense regime. In general, we show that the degrees follow a multivariate version of the *Poisson-Binomial* distribution in the canonical ensemble.

# §3.1 Introduction and main results

## §3.1.1 Background and outline

For most real-world networks, a detailed knowledge of the architecture of the network is not available and one must work with a probabilistic description, where the network is assumed to be a random sample drawn from a set of allowed configurations that are consistent with a set of known *topological constraints* [95]. Statistical physics deals with the definition of the appropriate probability distribution over the set of configurations and with the calculation of the resulting properties of the system. Two key choices of probability distribution are:

(1) the *microcanonical ensemble*, where the constraints are *hard* (i.e., are satisfied by each individual configuration);

(2) the *canonical ensemble*, where the constraints are *soft* (i.e., hold as ensemble averages, while individual configurations may violate the constraints).

(In both ensembles, the entropy is *maximal* subject to the given constraints.)

In the limit as the size of the network diverges, the two ensembles are traditionally *assumed* to become equivalent, as a result of the expected vanishing of the fluctuations of the soft constraints (i.e., the soft constraints are expected to become asymptotically hard). However, it is known that this equivalence may be broken, as signalled by a non-zero specific relative entropy of the two ensembles (= on an appropriate scale). In earlier work various scenarios were identified for this phenomenon (see [92], [48], [38] and references therein). In the present paper we take a fresh look at breaking of ensemble equivalence by analysing a formula for the relative entropy, based on the *covariance structure* of the canonical ensemble, recently put forward by Squartini and Garlaschelli [93]. We consider the case of a random graph with a given degree sequence (configuration model) and show that this formula correctly predicts that the specific relative entropy is determined by the scaling of the determinant of the covariance matrix of the constraints in the dense regime, while it requires an extra correction term in the sparse regime and the ultra-dense regime. We also show that the different behaviours found in the different regimes correspond to the degrees being asymptotically Gaussian in the dense regime and asymptotically Poisson in the sparse regime, and the dual degrees being asymptotically Poisson in the ultra-dense regime. We further note that, in general, in the canonical ensemble the degrees are distributed according to a multivariate version of the *Poisson-Binomial* distribution [100], which admits the Gaussian distribution and the Poisson distribution as limits in appropriate regimes.

Our results imply that, in all three regimes, ensemble equivalence breaks down in the presence of an extensive number of constraints. This confirms the need for a *principled choice* of the ensemble used in practical applications. Three examples serve as an illustration:

(a) *Pattern detection* is the identification of nontrivial structural properties in a real-world network through comparison with a suitable *null model*, i.e., a random

graph model that preserves certain local topological properties of the network (like the degree sequence) but is otherwise completely random.

(b) *Community detection* is the identification of groups of nodes that are more densely connected with each other than expected under a null model, which is a popular special case of pattern detection.

(c) *Network reconstruction* employs purely local topological information to infer higher-order structural properties of a real-world network. This problem arises whenever the global properties of the network are not known, for instance, due to confidentiality or privacy issues, but local properties are. In such cases, optimal inference about the network can be achieved by maximising the entropy subject to the known local constraints, which again leads to the two ensembles considered here.

Breaking of ensemble equivalence means that different choices of the ensemble lead to asymptotically different behaviours. Consequently, while for applications based on ensemble-equivalent models the choice of the working ensemble can be arbitrary and can be based on mathematical convenience, for those based on ensemble-nonequivalent models the choice should be dictated by a criterion indicating which ensemble is the appropriate one to use. This criterion must be based on the *a priori* knowledge that is available about the network, i.e., which form of the constraint (hard or soft) applies in practice.

The remainder of this section is organised as follows. In Section 3.1.2 we introduce the constraints to be considered, which are on the *degree sequence*. In Section 3.1.3 we introduce the various regimes we will be interested in and state a formula for the relative entropy when the constraint is on the degree sequence. In Section 3.1.4 we state the formula for the relative entropy proposed in [93] and present our main theorem. In Section 3.1.5 we close with a discussion of the interpretation of this theorem and an outline of the remainder of the paper.

The microcanonical and the canonical ensemble, as well as the relative entropy density have been defined in Section 1.4.1 and 1.4.2.

## §3.1.2 Constraint on the degree sequence

The degree sequence of a graph $G \in \mathcal{G}_n$ is defined as $\vec{k}(G) = (k_i(G))_{i=1}^n$ with $k_i(G) = \sum_{j \neq i} g_{ij}(G)$. In what follows we constrain the degree sequence to a *specific value* $\vec{k}^*$, which we assume to be *graphical*, i.e., there is at least one graph with degree sequence $\vec{k}^*$. The constraint is therefore

$$\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^n \in \{1, 2, \dots, n-2\}^n, \tag{3.1}$$

The microcanonical ensemble, when the constraint is on the degree sequence, is known as the *configuration model* and has been studied intensively (see [95, 92, 99]). For later use we recall the form of the canonical probability in the configuration model, namely,

$$P_{\mathrm{can}}(G) = \prod_{1 \leq i < j \leq n} \left(p_{ij}^*\right)^{g_{ij}(G)} \left(1 - p_{ij}^*\right)^{1 - g_{ij}(G)} \tag{3.2}$$

with

$$p^*_{ij} = \frac{e^{-\theta^*_i-\theta^*_j}}{1+e^{-\theta^*_i-\theta^*_j}} \tag{3.3}$$

and with the vector of Lagrange multipliers tuned to the value $\vec{\theta}^* = (\theta^*_i)_{i=1}^n$ such that

$$\langle k_i\rangle = \sum_{j\neq i} p^*_{ij} = k^*_i, \qquad 1\leq i\leq n. \tag{3.4}$$

Using (1.16), we can write

$$S_n(P_{\rm mic}\mid P_{\rm can}) = \log\frac{P_{\rm mic}(G^*)}{P_{\rm can}(G^*)} = -\log[\Omega_{\vec{k}^*}P_{\rm can}(G^*)] = -\log Q[\vec{k^*}](\vec{k^*}), \tag{3.5}$$

where $\Omega_{\vec{k}}$ is the number of graphs with degree sequence $\vec{k}$,

$$Q[\vec{k^*}](\vec{k}\,) = \Omega_{\vec{k}}\,P_{\rm can}\big(G^{\vec{k}}\big) \tag{3.6}$$

is the probability that the degree sequence is equal to $\vec{k}$ under the canonical ensemble with constraint $\vec{k^*}$, $G^{\vec{k}}$ denotes an arbitrary graph with degree sequence $\vec{k}$, and $P_{\rm can}\big(G^{\vec{k}}\big)$ is the canonical probability in (3.2) rewritten for one such graph:

$$P_{\rm can}\big(G^{\vec{k}}\big) = \prod_{1\leq i<j\leq n}\big(p^*_{ij}\big)^{g_{ij}(G^{\vec{k}})}\big(1-p^*_{ij}\big)^{1-g_{ij}(G^{\vec{k}})} = \prod_{i=1}^n (x^*_i)^{k_i}\prod_{1\leq i<j\leq n}(1+x^*_ix^*_j)^{-1}. \tag{3.7}$$

In the last expression, $x^*_i = e^{-\theta^*_i}$, and $\vec{\theta} = (\theta^*_i)_{i=1}^n$ is the vector of Lagrange multipliers coming from (3.3).

## §3.1.3 Relevant regimes

The breaking of ensemble equivalence was analysed in [48] in the so-called *sparse regime*, defined by the condition

$$\max_{1\leq i\leq n} k^*_i = o(\sqrt{n}\,). \tag{3.8}$$

It is natural to consider the opposite setting, namely, the *ultra-dense regime* in which the degrees are close to $n-1$,

$$\max_{1\leq i\leq n}(n-1-k^*_i) = o(\sqrt{n}\,). \tag{3.9}$$

This can be seen as the *dual* of the sparse regime. We will see in Appendix B that under the map $k^*_i\mapsto n-1-k^*_i$ the microcanonical ensemble and the canonical ensemble preserve their relationship, in particular, their relative entropy is invariant.

It is a challenge to study breaking of ensemble equivalence *in between* the sparse regime and the ultra-dense regime, called the *dense regime*. In what follows we consider a subclass of the dense regime, called the *$\delta$-tame regime*, in which the graphs are subject to a certain uniformity condition.

**3.1.1 Definition.** A degree sequence $\vec{k}^* = (k_i^*)_{i=1}^n$ is called $\delta$-tame if and only if there exists a $\delta \in \left(0, \frac{1}{2}\right]$ such that

$$\delta \leq p_{ij}^* \leq 1 - \delta, \qquad 1 \leq i \neq j \leq n, \tag{3.10}$$

where $p_{ij}^*$ are the canonical probabilities in (3.2)–(3.4).

**3.1.2 Remark.** The name $\delta$-tame is taken from [9], which studies the number of graphs with a $\delta$-tame degree sequence. Definition 3.1.1 is actually a reformulation of the definition given in [9]. See Appendix A for details.

The condition in (3.10) implies that

$$(n-1)\delta \leq k_i^* \leq (n-1)(1-\delta), \qquad 1 \leq i \leq n, \tag{3.11}$$

i.e., $\delta$-tame graphs are nowhere too thin (sparse regime) nor too dense (ultra-dense regime).

It is natural to ask whether, conversely, condition (3.11) implies that the degree sequence is $\delta'$-tame for some $\delta' = \delta'(\delta)$. Unfortunately, this question is not easy to settle, but the following lemma provides a partial answer.

**3.1.3 Lemma.** *Suppose that $\vec{k}^* = (k_i^*)_{i=1}^n$ satisfies*

$$(n-1)\alpha \leq k_i^* \leq (n-1)(1-\alpha), \qquad 1 \leq i \leq n, \tag{3.12}$$

*for some $\alpha \in (\frac{1}{4}, \frac{1}{2}]$. Then there exist $\delta = \delta(\alpha) > 0$ and $n_0 = n_0(\alpha) \in \mathbb{N}$ such that $\vec{k}^* = (k_i^*)_{i=1}^n$ is $\delta$-tame for all $n \geq n_0$.*

*Proof.* The proof follows from [9, Theorem 2.1]. In fact, by picking $\beta = 1 - \alpha$ in that theorem, we find that we need $\alpha > \frac{1}{4}$. The theorem also gives information about the values of $\delta = \delta(\alpha)$ and $n_0 = n_0(\alpha)$.

## §3.1.4 Linking ensemble nonequivalence to the canonical covariances

In this section we investigate an important formula, recently put forward in [93], for the scaling of the relative entropy under a general constraint. The analysis in [93] allows for the possibility that not all the constraints (i.e., not all the components of the vector $\vec{C}$) are linearly independent. For instance, $\vec{C}$ may contain redundant replicas of the same constraint(s), or linear combinations of them. Since in the present paper we only consider the case where $\vec{C}$ is the degree sequence, the different components of $\vec{C}$ (i.e., the different degrees) are linearly independent.

When a $K$-dimensional constraint $\vec{C}^* = (C_i^*)_{i=1}^K$ with independent components is imposed, then a key result in [93] is the formula

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) \sim \log \frac{\sqrt{\det(2\pi Q)}}{T}, \qquad n \to \infty, \tag{3.13}$$

where

$$Q = (q_{ij})_{1\le i,j\le K} \tag{3.14}$$

is the $K\times K$ covariance matrix of the constraints under the canonical ensemble, whose entries are defined as

$$q_{ij} = \mathrm{Cov}_{P_{\mathrm{can}}}(C_i, C_j) = \langle C_i\, C_j\rangle - \langle C_i\rangle\langle C_j\rangle, \tag{3.15}$$

and

$$T = \prod_{i=1}^{K}\left[1 + O\left(1/\lambda_i^{(K)}(Q)\right)\right], \tag{3.16}$$

with $\lambda_i^{(K)}(Q) > 0$ the $i$-th eigenvalue of the $K\times K$ covariance matrix $Q$. This result can be formulated rigorously as



**3.1.1 Formula ([93]).** If all the constraints are linearly independent, then the limiting relative entropy $\alpha_n$-density equals

$$s_{\alpha_\infty} = \lim_{n\to\infty}\frac{\log\sqrt{\det(2\pi Q)}}{\alpha_n} + \tau_{\alpha_\infty} \tag{3.17}$$

with $\alpha_n$ the 'natural' speed and

$$\tau_{\alpha_\infty} = -\lim_{n\to\infty}\frac{\log T}{\alpha_n}. \tag{3.18}$$

The latter is zero when

$$\lim_{n\to\infty}\frac{|I_{K_n,R}|}{\alpha_n} = 0 \quad \forall\, R < \infty, \tag{3.19}$$

where $I_{K,R} = \{i = 1,\dots,K\colon\ \lambda_i^{(K)}(Q) \le R\}$ with $\lambda_i^{(K)}(Q)$ the $i$-th eigenvalue of the $K$-dimensional covariance matrix $Q$ (the notation $K_n$ indicates that $K$ may depend on $n$). Note that $0 \le I_{K,R} \le K$. Consequently, (3.19) is satisfied (and hence $\tau_{\alpha_\infty} = 0$) when $\lim_{n\to\infty} K_n/\alpha_n = 0$, i.e., when the number $K_n$ of constraints grows slower than $\alpha_n$.

**3.1.4 Remark ([93]).** Formula 3.1.1, for which [93] offers compelling evidence but not a mathematical proof, can be rephrased by saying that the natural choice of $\alpha_n$ is

$$\tilde{\alpha}_n = \log\sqrt{\det(2\pi Q)}. \tag{3.20}$$

Indeed, if all the constraints are linearly independent and (3.19) holds, then $\tau_{\tilde{\alpha}_n} = 0$ and

$$s_{\tilde{\alpha}_\infty} = 1, \tag{3.21}$$

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = [1 + o(1)]\,\tilde{\alpha}_n. \tag{3.22}$$

We now present our main theorem, which considers the case where the constraint is on the degree sequence: $K_n = n$ and $\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^n$. This case was studied in [48], for which $\alpha_n = n$ in the *sparse regime with finite degrees*. Our results here focus on three new regimes, for which we need to increase $\alpha_n$: the *sparse regime with growing degrees*, the *δ-tame regime*, and the *ultra-dense regime with growing dual degrees*. In all these cases, since $\lim_{n\to\infty} K_n/\alpha_n = \lim_{n\to\infty} n/\alpha_n = 0$, Formula 3.1.1 states that (3.17) holds with $\tau_{\tilde{\alpha}_n} = 0$. Our theorem provides a rigorous and independent mathematical proof of this result.

**3.1.5 Theorem.** *Formula* 3.1.1 *is true with* $\tau_{\alpha_\infty} = 0$ *when the constraint is on the degree sequence* $\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^n$*, the scale parameter is* $\alpha_n = n\,\overline{f_n}$ *with*

$$\overline{f_n} = n^{-1}\sum_{i=1}^{n} f_n(k_i^*) \qquad \textit{with} \qquad f_n(k) = \frac{1}{2}\log\left[\frac{k(n-1-k)}{n}\right], \tag{3.23}$$

*and the degree sequence belongs to one of the following three regimes:*

- *The sparse regime with growing degrees:*

$$\max_{1\le i\le n} k_i^* = o(\sqrt{n}\,), \qquad \lim_{n\to\infty}\min_{1\le i\le n} k_i^* = \infty. \tag{3.24}$$

- *The δ-tame regime (see* (3.2) *and Lemma* 3.1.3*):*

$$\delta \le p_{ij}^* \le 1-\delta, \quad 1 \le i \ne j \le n. \tag{3.25}$$

- *The ultra-dense regime with growing dual degrees:*

$$\max_{1\le i\le n} (n-1-k_i^*) = o(\sqrt{n}\,), \qquad \lim_{n\to\infty}\min_{1\le i\le n} (n-1-k_i^*) = \infty. \tag{3.26}$$

*In all three regimes there is breaking of ensemble equivalence, and*

$$s_{\alpha_\infty} = \lim_{n\to\infty} s_{\alpha_n} = 1. \tag{3.27}$$

## §3.1.5 Discussion and outline

Comparing (3.21) and (3.27), and using (3.20), we see that Theorem 3.1.5 shows that if the constraint is on the degree sequence, then

$$S_n(P_{\rm mic} \mid P_{\rm can}) \sim n\overline{f_n} \sim \log\sqrt{\det(2\pi Q)} \tag{3.28}$$

in each of the three regimes considered. Below we provide a heuristic explanation for this result (as well as for our previous results in [48]) that links back to (3.5). In Section 3.2 we prove Theorem 3.1.5.

**Poisson-Binomial degrees in the general case.** Note that (3.5) can be rewritten as

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) = S\big(\,\delta[\vec{k^*}] \mid Q[\vec{k^*}]\,\big), \tag{3.29}$$

where $\delta[\vec{k^*}] = \prod_{i=1}^n \delta[k_i^*]$ is the *multivariate Dirac distribution* with average $\vec{k^*}$. This has the interesting interpretation that the relative entropy between the distributions $P_{\text{mic}}$ and $P_{\text{can}}$ *on the set of graphs* coincides with the relative entropy between $\delta[\vec{k^*}]$ and $Q[\vec{k^*}]$ *on the set of degree sequences.*

To be explicit, using (3.6) and (3.7), we can rewrite $Q[\vec{k^*}](\vec{k})$ as

$$Q[\vec{k^*}](\vec{k}) = \Omega_{\vec{k}} \prod_{i=1}^n (x_i^*)^{k_i} \prod_{1\le i<j\le n} (1 + x_i^* x_j^*)^{-1}. \tag{3.30}$$

We note that the above distribution is a multivariate version of the *Poisson-Binomial distribution* (or Poisson's Binomial distribution; see Wang [100]). In the univariate case, the Poisson-Binomial distribution describes the probability of a certain number of successes out of a total number of independent and (in general) not identical Bernoulli trials [100]. In our case, the marginal probability that node $i$ has degree $k_i$ in the canonical ensemble, irrespectively of the degree of any other node, is indeed a univariate Poisson-Binomial given by $n-1$ independent Bernoulli trials with success probabilities $\{p_{ij}^*\}_{j\neq i}$. The relation in (3.29) can therefore be restated as

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) = S\big(\,\delta[\vec{k^*}] \mid \text{PoissonBinomial}[\vec{k^*}]\,\big), \tag{3.31}$$

where PoissonBinomial$[\vec{k^*}]$ is the multivariate Poisson-Binomial distribution given by (3.30), i.e.,

$$Q[\vec{k^*}] = \text{PoissonBinomial}[\vec{k^*}]. \tag{3.32}$$

The relative entropy can therefore be seen as coming from a situation in which the microcanonical ensemble forces the degree sequence to be exactly $\vec{k^*}$, while the canonical ensemble forces the degree sequence to be Poisson-Binomial distributed with average $\vec{k^*}$.

It is known that the univariate Poisson-Binomial distribution admits two asymptotic limits: (1) a Poisson limit (if and only if, in our notation, $\sum_{j\neq i} p_{ij}^* \to \lambda > 0$ and $\sum_{j\neq i}(p_{ij}^*)^2 \to 0$ as $n\to\infty$ [100]); (2) a Gaussian limit (if and only if $p_{ij}^* \to \lambda_j > 0$ for all $j \neq i$ as $n\to\infty$, as follows from a central limit theorem type of argument). If all the Bernoulli trials are identical, i.e., if all the probabilities $\{p_{ij}^*\}_{j\neq i}$ are equal, then the univariate Poisson-Binomial distribution reduces to the ordinary Binomial distribution, which also exhibits the well-known Poisson and Gaussian limits. These results imply that also the general multivariate Poisson-Binomial distribution in (3.30) admits limiting behaviours that should be consistent with the Poisson and Gaussian limits discussed above for its marginals. This is precisely what we confirm below.

**Poisson degrees in the sparse regime.** In [48] it was shown that, for a sparse degree sequence,

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) \sim \sum_{i=1}^n S\big(\,\delta[k_i^*] \mid \text{Poisson}[k_i^*]\,\big). \tag{3.33}$$

The right-hand side is the sum over all nodes $i$ of the relative entropy of the *Dirac distribution* with average $k_i^*$ w.r.t. the *Poisson distribution* with average $k_i^*$. We see that, under the sparseness condition, the constraints act on the nodes essentially independently. We can therefore reinterpret (3.33) as the statement

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) \sim S\big(\delta[\vec{k^*}] \mid \text{Poisson}[\vec{k^*}]\big), \tag{3.34}$$

where $\text{Poisson}[\vec{k^*}] = \prod_{i=1}^n \text{Poisson}[k_i^*]$ is the *multivariate Poisson distribution* with average $\vec{k^*}$. In other words, in this regime

$$Q[\vec{k^*}] \sim \text{Poisson}[\vec{k^*}], \tag{3.35}$$

i.e. the joint multivariate Poisson-Binomial distribution (3.30) essentially decouples into the product of marginal univariate Poisson-Binomial distributions describing the degrees of all nodes, and each of these Poisson-Binomial distributions is asymptotically a Poisson distribution.

Note that the Poisson regime was obtained in [48] under the condition in (3.8), which is less restrictive than the aforementioned condition $k_i^* = \sum_{j\neq i} p_{ij}^* \to \lambda > 0$, $\sum_{j\neq i}(p_{ij}^*)^2 \to 0$ under which the Poisson distribution is retrieved from the Poisson-Binomial distribution [100]. In particular, the condition in (3.8) includes both the case with growing degrees included in Theorem 3.1.5 (and consistent with Formula 3.1.1 with $\tau_{\alpha_\infty} = 0$) and the case with finite degrees, which *cannot* be retrieved from Formula 3.1.1 with $\tau_{\alpha_\infty} = 0$, because it corresponds to the case where all the $n = \alpha_n$ eigenvalues of $Q$ remain finite as $n$ diverges (as the entries of $Q$ themselves do not diverge), and indeed (3.19) does not hold.

**Poisson degrees in the ultra-dense regime.** Since the ultra-dense regime is the dual of the sparse regime, we immediately get the heuristic interpretation of the relative entropy when the constraint is on an ultra-dense degree sequence $\vec{k}^*$. Using (3.34) and the observations in Appendix B (see, in particular (B.2)), we get

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) \sim S\big(\delta[\vec{\ell^*}] \mid \text{Poisson}[\vec{\ell^*}]\big), \tag{3.36}$$

where $\vec{\ell^*} = (\ell_i^*)_{i=1}^n$ is the dual degree sequence given by $\ell_i^* = n - 1 - k_i^*$. In other words, under the microcanonical ensemble the dual degrees follow the distribution $\delta[\vec{\ell^*}]$, while under the canonical ensemble the dual degrees follow the distribution $Q[\vec{\ell^*}]$, where in analogy with (3.35),

$$Q[\vec{\ell^*}] \sim \text{Poisson}[\vec{\ell^*}]. \tag{3.37}$$

Similar to the sparse case, the multivariate Poisson-Binomial distribution (3.30) reduces to a product of marginal, and asymptotically Poisson, distributions governing the different degrees.

Again, the case with finite dual degrees cannot be retrieved from Formula 3.1.1 with $\tau_{\alpha_\infty} = 0$, because it corresponds to the case where $Q$ has a diverging (like $n = \alpha_n$) number of eigenvalues whose value remains finite as $n \to \infty$, and (3.19) does not hold. By contrast, the case with growing dual degrees can be retrieved from Formula 3.1.1 with $\tau_{\alpha_\infty} = 0$ because (3.19) holds, as confirmed in Theorem 3.1.5.

**Gaussian degrees in the dense regime.** We can reinterpet (3.28) as the statement

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) \sim S\big(\,\delta[\vec{k^*}] \mid \mathrm{Normal}[\vec{k^*}, Q]\,\big), \tag{3.38}$$

where $\mathrm{Normal}[\vec{k^*}, Q]$ is the *multivariate Normal distribution* with mean $\vec{k^*}$ and covariance matrix $Q$. In other words, in this regime

$$Q[\vec{k^*}] \sim \mathrm{Normal}[\vec{k^*}, Q], \tag{3.39}$$

i.e., the multivariate Poisson-Binomial distribution (3.30) is asymptotically a multivariate Gaussian distribution whose covariance matrix is in general not diagonal, i.e., the dependencies between degrees of different nodes do *not* vanish, unlike in the other two regimes. Since all the degrees are growing in this regime, so are all the eigenvalues of $Q$, implying (3.19) and consistently with Formula 3.1.1 with $\tau_{\alpha_\infty} = 0$, as proven in Theorem 3.1.5.

Note that the right-hand side of (3.38), being the relative entropy of a discrete distribution with respect to a continuous distribution, needs to be properly interpreted: the Dirac distribution $\delta[\vec{k^*}]$ needs to be smoothened to a continuous distribution with support in a small ball around $\vec{k^*}$. Since the degrees are large, this does not affect the asymptotics.

**Crossover between the regimes.** An easy computation gives

$$S\big(\,\delta[k_i^*] \mid \mathrm{Poisson}[k_i^*]\,\big) = g(k_i^*) \quad \text{with} \quad g(k) = \log\left(\frac{k!}{e^{-k}k^k}\right). \tag{3.40}$$

Since $g(k) = [1 + o(1)]\frac{1}{2}\log(2\pi k)$, $k \to \infty$, we see that, as we move from the sparse regime with finite degrees to the sparse regime with growing degrees, the scaling of the relative entropy in (3.33) nicely links up with that of the dense regime in (3.38) via the common expression in (3.28). Note, however, that since the sparse regime with growing degrees is in general incompatible with the dense $\delta$-tame regime, in Theorem 3.1.5 we have to obtain the two scalings of the relative entropy under disjoint assumptions. By contrast, Formula 3.1.1 with $\tau_{\alpha_\infty} = 0$, and hence (3.22), unifies the two cases under the simpler and more general requirement that all the eigenvalues of $Q$, and hence all the degrees, diverge. Actually, (3.22) is expected to hold in the even more general hybrid case where there are both finite and growing degrees, provided the number of finite-valued eigenvalues of $Q$ grows slower than $\alpha_n$ [93].

**Other constraints.** It would be interesting to investigate Formula 3.1.1 for constraints other than on the degrees. Such constraints are typically much harder to analyse. In [38] constraints are considered on the total number of edges and the total number of triangles *simultaneously* ($K = 2$) in the dense regime. It was found that, with $\alpha_n = n^2$, breaking of ensemble equivalence occurs for some 'frustrated' choices of these numbers. Clearly, this type of breaking of ensemble equivalence does not arise from the recently proposed [93] mechanism associated with a diverging number of constraints as in the cases considered in this paper, but from the more traditional [97] mechanism of a phase transition associated with the frustration phenomenon.

**Outline.** Theorem 3.1.5 is proved in Section 3.2. In Appendix A we show that the canonical probabilities in (3.2) are the same as the probabilities used in [9] to define a $\delta$-tame degree sequence. In Appendix B we explain the duality between the sparse regime and the ultra-dense regime.

## §3.2 Proof of the Main Theorem

In Section 3.2.2 we prove Theorem 3.1.5. The proof is based on two lemmas, which we state and prove in Section 3.2.1.

### §3.2.1 Preparatory lemmas

The following lemma gives an expression for the relative entropy.

**3.2.1 Lemma.** *If the constraint is a $\delta$-tame degree sequence, then the relative entropy in* (1.16) *scales as*

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = [1 + o(1)]\, \tfrac{1}{2} \log[\det(2\pi Q)], \tag{3.41}$$

*where $Q$ is the covariance matrix in* (3.14)*. This matrix $Q = (q_{ij})$ takes the form*

$$\begin{cases} q_{ii} = k_i^* - \sum_{j \neq i} (p_{ij}^*)^2 = \sum_{j \neq i} p_{ij}^* (1 - p_{ij}^*), & 1 \leq i \leq n, \\ q_{ij} = p_{ij}^* (1 - p_{ij}^*), & 1 \leq i \neq j \leq n. \end{cases} \tag{3.42}$$

*Proof.* To compute $q_{ij} = \mathrm{Cov}_{P_{\mathrm{can}}}(k_i, k_j)$ we take the second order derivatives of the log-likelihood function

$$\begin{aligned} &\mathcal{L}(\vec{\theta}) = \log P_{\mathrm{can}}(G^* \mid \vec{\theta}) \\ &= \log \left[ \prod_{1 \leq i < j \leq n} p_{ij}^{g_{ij}(G^*)} (1 - p_{ij})^{(1 - g_{ij}(G^*))} \right], \quad p_{ij} = \frac{e^{-\theta_i - \theta_j}}{1 + e^{-\theta_i - \theta_j}} \end{aligned} \tag{3.43}$$

in the point $\vec{\theta} = \vec{\theta}^*$ [93]. Indeed, it is easy to show that the first-order derivatives are [51]

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\vec{\theta}) = \langle k_i \rangle - k_i^*, \quad \frac{\partial}{\partial \theta_i} \mathcal{L}(\vec{\theta}) \bigg|_{\vec{\theta} = \vec{\theta}^*} = k_i^* - k_i^* = 0 \tag{3.44}$$

and the second-order derivatives are

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathcal{L}(\vec{\theta}) \bigg|_{\vec{\theta} = \vec{\theta}^*} = \langle k_i\, k_j \rangle - \langle k_i \rangle \langle k_j \rangle = \mathrm{Cov}_{P_{\mathrm{can}}}(k_i, k_j). \tag{3.45}$$

This readily gives (3.42).

The proof of (3.41) uses [9, Eq. (1.4.1)], which says that if a $\delta$-tame degree sequence is used as constraint, then

$$P_{\mathrm{mic}}^{-1}(G^*) = \Omega_{\vec{C}^*} = \frac{e^{H(p^*)}}{(2\pi)^{n/2} \sqrt{\det(Q)}}\, e^C, \tag{3.46}$$

where $Q$ and $p^*$ are defined in (3.42) and (3.70) below, while $e^C$ is sandwiched between two constants that depend on $\delta$:

$$\gamma_1(\delta) \leq e^C \leq \gamma_2(\delta). \tag{3.47}$$

From (3.46) and the relation $H(p^*) = -\log P_{\rm can}(G^*)$, proved in Lemma A.1 below, we get the claim.

The following lemma shows that the diagonal approximation of $\log(\det Q)/n\overline{f}_n$ is good when the degree sequence is $\delta$-tame.

**3.2.2 Lemma.** *Under the $\delta$-tame condition,*

$$\log(\det Q_D) + o(n\,\overline{f}_n) \leq \log(\det Q) \leq \log(\det Q_D) \tag{3.48}$$

*with $Q_D = \mathrm{diag}(Q)$ the matrix that coincides with $Q$ on the diagonal and is zero off the diagonal.*



*Proof.* We use [60, Theorem 2.3], which says that if

(1) $\det(Q)$ is real,

(2) $Q_D$ is non-singular with $\det(Q_D)$ real,

(3) $\lambda_i(A) > -1$, $1 \leq i \leq n$,

then

$$e^{-\frac{n\rho^2(A)}{1+\lambda_{\min}(A)}} \det Q_D \leq \det Q \leq \det Q_D. \tag{3.49}$$

Here, $A = Q_D^{-1}Q_{\rm off}$, with $Q_{\rm off}$ the matrix that coincides with $Q$ off the diagonal and is zero on the diagonal, $\lambda_i(A)$ is the $i$-th eigenvalue of $A$ (arranged in decreasing order), $\lambda_{\min}(A) = \min_{1\leq i\leq n}\lambda_i(A)$, and $\rho(A) = \max_{1\leq i\leq n}|\lambda_i(A)|$.

We begin by verifying (1)–(3).

(1) Since $Q$ is a symmetric matrix with real entries, $\det Q$ exists and is real.

(2) This property holds thanks to the $\delta$-tame condition. Indeed, since $q_{ij} = p^*_{i,j}(1 - p^*_{i,j})$, we have

$$0 < \delta^2 \leq q_{ij} \leq (1-\delta)^2 < 1, \tag{3.50}$$

which implies that

$$0 < (n-1)\delta^2 \leq q_{ii} = \sum_{j\neq i} q_{ij} \leq (n-1)(1-\delta)^2. \tag{3.51}$$

(3) It is easy to show that $A = (a_{ij})$ is given by

$$a_{ij} = \begin{cases} \frac{q_{ij}}{q_{ii}}, & 1 \leq i \neq j \leq n, \\ 0, & 1 \leq i \leq n, \end{cases} \tag{3.52}$$

where $q_{ij}$ is given by (3.42). Since $q_{ij} = q_{ji}$, the matrix $A$ is symmetric. Moreover, since $q_{ii} = \sum_{j\neq i} q_{ij}$, the matrix $A$ is also Markov. We therefore have

$$1 = \lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A) \geq -1. \tag{3.53}$$

From (3.50) and (3.52) we get

$$0 < \frac{1}{n-1}\left(\frac{\delta}{1-\delta}\right)^2 \leq a_{ij} \leq \frac{1}{n-1}\left(\frac{1-\delta}{\delta}\right)^2. \tag{3.54}$$

This implies that the Markov chain on $\{1,\ldots,n\}$ with transition matrix $A$ starting from $i$ can return to $i$ with a positive probability after an arbitrary number of steps $\geq 2$. Consequently, the last inequality in (3.53) is strict.

We next show that

$$\frac{n\rho^2(A)}{1+\lambda_{\min}(A)} = o(n\,\overline{f}_n). \tag{3.55}$$

Together with (3.49) this will settle the claim in (3.48). From (3.53) it follows $\rho(A) = 1$, so we must show that

$$\lim_{n\to\infty}[1+\lambda_{\min}(A)]\,\overline{f}_n = \infty. \tag{3.56}$$

Using [104, Theorem 4.3], we get

$$\lambda_{\min}(A) \geq -1 + \frac{\min_{1\leq i\neq j\leq n}\pi_i a_{ij}}{\min_{1\leq i\leq n}\pi_i}\,\mu_{\min}(L) + 2\gamma. \tag{3.57}$$

Here, $\pi = (\pi_i)_{i=1}^n$ is the invariant distribution of the reversible Markov chain with transition matrix $A$, while $\mu_{\min}(L) = \min_{1\leq i\leq n}\lambda_i(L)$ and $\gamma = \min_{1\leq i\leq n} a_{ii}$, with $L = (L_{ij})$ the matrix such that, for $i \neq j$, $L_{ij} = 1$ if and only if $a_{ij} > 0$, while $L_{ii} = \sum_{j\neq i} L_{ij}$.

We find that $\pi_i = \frac{1}{n}$ for $1 \leq i \leq n$, $L_{ij} = 1$ for $1 \leq i \neq j \leq n$, $L_{ii} = n-1$ for $1 \leq i \leq n$, and $\gamma = 0$. Hence (3.57) becomes

$$\lambda_{\min}(A) \geq -1 + (n-2)\min_{1\leq i\neq j\leq n} a_{ij} \geq -1 + \frac{n-2}{n-1}\left(\frac{\delta}{1-\delta}\right)^2, \tag{3.58}$$

where the last inequality comes from (3.54). To get (3.56) it therefore suffices to show that $\overline{f}_\infty = \lim_{n\to\infty}\overline{f}_n = \infty$. But, using the $\delta$-tame condition, we can estimate

$$\begin{aligned}\frac{1}{2}\log\left[\frac{(n-1)\delta(1-\delta+n\delta)}{n}\right] \leq \overline{f}_n &= \frac{1}{2n}\sum_{i=1}^n \log\left[\frac{k_i^*(n-1-k_i^*)}{n}\right] \\ &\leq \frac{1}{2}\log\left[\frac{(n-1)(1-\delta)(\delta+n(1-\delta))}{n}\right],\end{aligned} \tag{3.59}$$

and both bounds scale like $\frac{1}{2}\log n$ as $n\to\infty$.

## §3.2.2 Proof (Theorem 3.1.5)

*Proof.* We deal with each of the three regimes in Theorem 3.1.5 separatetely.

**The sparse regime with growing degrees.** Since $\vec{k}^* = (k_i^*)_{i=1}^n$ is a sparse degree sequence, we can use [48, Eq. (3.12)], which says that

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = \sum_{i=1}^n g(k_i^*) + o(n), \qquad n \to \infty, \tag{3.60}$$

where $g(k) = \log\left(\frac{k!}{k^k e^{-k}}\right)$ is defined in (3.40). Since the degrees are growing, we can use Stirling's approximation $g(k) = \frac{1}{2}\log(2\pi k) + o(1)$, $k \to \infty$, to obtain

$$\sum_{i=1}^n g(k_i^*) = \tfrac{1}{2}\sum_{i=1}^n \log(2\pi k_i^*) + o(n) = \tfrac{1}{2}\left[n\log 2\pi + \sum_{i=1}^n \log k_i^*\right] + o(n). \tag{3.61}$$

Combining (3.60)–(3.61), we get

$$\frac{S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}})}{n\,\overline{f}_n} = \tfrac{1}{2}\left[\frac{\log 2\pi}{\overline{f}_n} + \frac{\sum_{i=1}^n \log k_i^*}{n\overline{f}_n}\right] + o(1). \tag{3.62}$$

Recall (3.23). Because the degrees are sparse, we have

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n \log k_i^*}{n\overline{f}_n} = 2. \tag{3.63}$$

Because the degrees are growing, we also have

$$\overline{f}_\infty = \lim_{n\to\infty} \overline{f}_n = \infty. \tag{3.64}$$

Combining (3.62)–(3.64) we find that $\lim_{n\to\infty} S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}})/n\,\overline{f}_n = 1$.

**The ultra-dense regime with growing dual degrees.** If $\vec{k}^* = (k_i^*)_{i=1}^n$ is an ultra-dense degree sequence, then the dual $\vec{\ell}^* = (\ell_i^*)_{i=1}^n = (n-1-k_i^*)_{i=1}^n$ is a sparse degree sequence. By Lemma B.2, the relative entropy is invariant under the map $k_i^* \to \ell_i^* = n-1-k_i^*$. So is $\bar{f}_n$, and hence the claim follows from the proof in the sparse regime.

**The $\delta$-tame regime.** It follows from Lemma 3.2.1 that

$$\lim_{n\to\infty} \frac{S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}})}{n\,\overline{f}_n} = \tfrac{1}{2}\left[\lim_{n\to\infty} \frac{\log 2\pi}{\overline{f}_n} + \lim_{n\to\infty} \frac{\log(\det Q)}{n\,\overline{f}_n}\right]. \tag{3.65}$$

From (3.59) we know that $\overline{f}_\infty = \lim_{n\to\infty} \overline{f}_n = \infty$ in the *$\delta$-tame regime*. It follows from Lemma 3.2.2 that

$$\lim_{n\to\infty} \frac{\log(\det Q)}{n\,\overline{f}_n} = \lim_{n\to\infty} \frac{\log(\det Q_D)}{n\,\overline{f}_n}. \tag{3.66}$$

To conclude the proof it therefore suffices to show that

$$\lim_{n\to\infty} \frac{\log(\det Q_D)}{n\,\overline{f}_n} = 2. \tag{3.67}$$

Using (3.51) and (3.59), we may estimate

$$\frac{2\log[(n-1)\delta^2]}{\log\frac{(n-1)(1-\delta)(\delta+n(1-\delta))}{n}} \le \frac{\sum_{i=1}^n \log(q_{ii})}{n\,\overline{f}_n} = \frac{\log(\det Q_D)}{n\,\overline{f}_n} \le \frac{2\log[(n-1)(1-\delta)^2]}{\log\frac{(n-1)\delta(1-\delta+n\delta)}{n}}. \tag{3.68}$$

Both sides tend to 2 as $n \to \infty$, and so (3.67) follows.

# §A Appendix

Here we show that the canonical probabilities in (3.2) are the same as the probabilities used in [9] to define a $\delta$-tame degree sequence.

For $q = (q_{ij})_{1\le i,j\le n}$, let

$$E(q) = -\sum_{1\le i\ne j\le n} q_{ij}\log q_{ij} + (1-q_{ij})\log(1-q_{ij}). \tag{3.69}$$

be the entropy of $q$. For a given degree sequence $(k_i^*)_{i=1}^n$, consider the following maximisation problem:

$$\begin{cases} \max E(q), \\ \sum_{j\ne i} q_{ij} = k_i^*,\ 1\le i\le n, \\ 0\le q_{ij}\le 1,\ 1\le i\ne j\le n. \end{cases} \tag{3.70}$$

Since $q \mapsto E(q)$ is strictly concave, it attains its maximum at a unique point.

**A.1 Lemma.** *The canonical probability takes the form*

$$P_{\mathrm{can}}(G) = \prod_{1\le i<j\le n} \left(p_{ij}^*\right)^{g_{ij}(G)} \left(1-p_{ij}^*\right)^{1-g_{ij}(G)}, \tag{3.71}$$

*where $p^* = (p_{ij}^*)$ solves* (3.70)*. In addition,*

$$\log P_{\mathrm{can}}(G^*) = -H(p^*). \tag{3.72}$$

*Proof.* It was shown in [48] that, for a degree sequence constraint,

$$P_{\mathrm{can}}(G) = \prod_{1\le i<j\le n} \left(p_{ij}^*\right)^{g_{ij}(G)} \left(1-p_{ij}^*\right)^{1-g_{ij}(G)} \tag{3.73}$$

with $p_{ij}^* = \frac{e^{-\theta_i^*-\theta_j^*}}{1+e^{-\theta_i^*-\theta_j^*}}$, where $\vec{\theta}^*$ has to be tuned such that

$$\sum_{j\ne i} p_{ij}^* = k_i^*, \qquad 1\le i\le n. \tag{3.74}$$

On the other hand, the solution of (3.70) via the Lagrange multiplier method gives that

$$q_{ij}^* = \frac{e^{-\phi_i^*-\phi_j^*}}{1+e^{-\phi_i^*-\phi_j^*}}, \tag{3.75}$$

where $\vec{\phi}^*$ has to be tuned such that

$$\sum_{j\neq i} q_{ij}^* = k_i^*, \qquad 1 \leq i \leq n. \tag{3.76}$$

This implies that $q_{ij}^* = p_{ij}^*$ for all $1 \leq i \neq j \leq n$. Moreover,

$$\begin{aligned}\log P_{\rm can}(G^*) + H(p^*) &= \sum_{1\leq i<j\leq n} g_{ij}(G^*) \log\left(\frac{p_{ij}^*}{1-p_{ij}^*}\right) - \sum_{1\leq i<j\leq n} p_{ij}^* \log\left(\frac{p_{ij}^*}{1-p_{ij}^*}\right) \\ &= -\sum_{1\leq i<j\leq n} g_{ij}(G^*)(\theta_i^*+\theta_j^*) + \sum_{1\leq i<j\leq n} p_{ij}^*(\theta_i^*+\theta_j^*) = \sum_{i=1}^n \theta_i^* \sum_{j\neq i}(p_{ij}^* - g_{ij}(G^*)) = 0,\end{aligned} \tag{3.77}$$

where the last equation follows from the fact that

$$\sum_{j\neq i} g_{ij}(G^*) = \sum_{j\neq i} p_{ij}^* = k_i^*, \qquad 1 \leq i \leq n. \tag{3.78}$$

# §B Appendix

We explain the duality between the sparse regime and the ultra-dense regime.

Let $\vec{k}^* = (k_i^*)_{i=1}^n$ be an ultra-dense degree sequence,

$$\max_{1\leq i\leq n}(n-1-k_i^*) = o(\sqrt{n}), \tag{3.79}$$

and let $\vec{\ell}^* = (\ell_i^*)_{i=1}^n$ be the *dual* degree sequence defined by $\ell_i^* = n-1-k_i^*$. Clearly, $\vec{\ell}^* = (\ell_i^*)_{i=1}^n$ is a sparse degree sequence,

$$\max_{1\leq i\leq n} \ell_i^* = o(\sqrt{n}). \tag{3.80}$$

**B.1 Lemma.** *Let $P_{\rm can}$ and $\widehat{P_{\rm can}}$ denote the canonical ensembles in* (1.9) *when $\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^n$, respectively, $\vec{C}^* = \vec{\ell}^* = (\ell_i^*)_{i=1}^n$. Then*

$$P_{\rm can}(G) = \widehat{P_{\rm can}}(G_c), \qquad G \in \mathcal{G}_n, \tag{3.81}$$

*where $G$ and $G_c$ are complementary graphs, i.e.,*

$$g_{ij}(G_c) = 1 - g_{ij}(G), \qquad 1 \leq i \neq j \leq n. \tag{3.82}$$

*Proof.* From the definition of the canonical probabilities we have

$$P_{\rm can}(G) = P_{\rm can}(G \mid \vec{\theta}^*), \qquad \widehat{P_{\rm can}}(G) = P_{\rm can}(G \mid \vec{\phi}^*), \tag{3.83}$$

where

$$P_{\rm can}(G \mid \vec{\theta}) = \frac{\exp[-\vec{\theta}\cdot\vec{k}(G)]}{Z(\vec{\theta})}, \qquad \vec{k}(G) = \sum_{j\neq i} g_{ij}(G), \tag{3.84}$$

and the values $\vec{\theta}^*$ and $\vec{\phi}^*$ are such that

$$\left.\frac{\partial F(\vec{\theta})}{\partial \theta_i}\right|_{\vec{\theta}=\vec{\theta}^*} = -\langle k_i \rangle_{P_{\mathrm{can}}(\cdot \mid \vec{\theta}^*)} = -k_i^*, \tag{3.85}$$

$$\left.\frac{\partial F(\vec{\theta})}{\partial \theta_i}\right|_{\vec{\theta}=\vec{\phi}^*} = -\langle k_i \rangle_{P_{\mathrm{can}}(\cdot \mid \vec{\phi}^*)} = -\ell_i^*. \tag{3.86}$$

The free energy is $F(\vec{\theta}) = \log Z(\vec{\theta})$, and its $i$-th partial derivative in the Lagrange multiplier that fixes the average of the $i$-th constraint. We show that $\vec{\theta}^* = -\vec{\phi}^*$.

Write

$$Z(\vec{\theta}) = \sum_{G \in \mathcal{G}_n} e^{-\vec{\theta}\cdot\vec{k}(G)} = \sum_{G \in \mathcal{G}_n} e^{-\sum_{i=1}^n \theta_i (n-1-k(G_c))} = e^{-(n-1)\sum_{i=1}^n \theta_i}\, Z(-\vec{\theta}). \tag{3.87}$$

Using (3.85) and (3.87), we get

$$-k_i^* = \left.\frac{\partial F(\vec{\theta})}{\partial \theta_i}\right|_{\vec{\theta}=\vec{\theta}^*} = -(n-1) + \langle k_i \rangle_{P_{\mathrm{can}}(\cdot \mid -\vec{\theta}^*)}. \tag{3.88}$$

Since $k_i^* = n - 1 - \ell_i^*$, we obtain

$$\ell_i^* = \langle k_i \rangle_{P_{\mathrm{can}}(\cdot \mid -\vec{\theta}^*)}. \tag{3.89}$$

From (3.86), (3.89) and the uniqueness of the Lagrange multipliers, we get

$$\vec{\theta}^* = -\vec{\phi}^*. \tag{3.90}$$

Using (3.87) and (3.90), we obtain

$$\begin{aligned} \widehat{P_{\mathrm{can}}}(G_c) = P_{\mathrm{can}}(G_c \mid \vec{\phi}^*) = P_{\mathrm{can}}(G_c \mid -\vec{\theta}^*) = \frac{\exp[\vec{\theta}^* \cdot \vec{k}(G_c)]}{Z(-\vec{\theta}^*)} \\ = \frac{\exp[-\vec{\theta}^* \cdot \vec{k}(G)]}{Z(-\vec{\theta}^*)\, e^{-(n-1)\sum_{i=1}^n \theta_i^*}} = \frac{\exp[-\vec{\theta}^* \cdot \vec{k}(G)]}{Z(\vec{\theta}^*)} = P_{\mathrm{can}}(G), \end{aligned} \tag{3.91}$$

which settles (3.81).

**B.2 Lemma.** *Let*

- $P_{\mathrm{mic}}$ *and* $P_{\mathrm{can}}$ *denote the microcanonical ensemble in* (1.6)*, respectively, the canonical ensemble in* (1.9)*, when* $\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^n$ *with* $k_i^*$ *satisfying* (3.79)*.*
- $\widehat{P_{\mathrm{mic}}}$ *and* $\widehat{P_{\mathrm{can}}}$ *denote the microcanonical ensemble in* (1.6)*, respectively, the canonical ensemble in* (1.9)*, when* $\vec{C}^* = \vec{\ell}^* = (\ell_i^*)_{i=1}^n$ *with* $\ell_i^* = n - 1 - k_i^*$ *the dual degree satisfying* (3.80)*.*

*Then the relative entropy in* (1.16) *satisfies*

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = S_n(\widehat{P_{\mathrm{mic}}} \mid \widehat{P_{\mathrm{can}}}). \tag{3.92}$$

*Proof.* Consider a graph $G^*$ with degree sequence $\vec{k}(G^*) = \vec{k}^*$. Then

$$P_{\mathrm{mic}}(G^*) = |\{G \in \mathcal{G}_n \colon \vec{k}(G) = \vec{k}^*\}|^{-1} = |\{G \in \mathcal{G}_n \colon \vec{k}(G) = \vec{\ell}^*\}|^{-1} = \widehat{P_{\mathrm{mic}}}(G_c^*), \tag{3.93}$$

where $G_c^*$ and $G^*$ are complementary graphs, so that $\vec{k}(G_c^*) = \vec{\ell}^*$. Using Lemma B.1, we have

$$P_{\mathrm{can}}(G^*) = \widehat{P_{\mathrm{can}}}(G_c^*). \tag{3.94}$$

Combine (1.16), (3.93) and (3.94), to get

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = \log \frac{P_{\mathrm{mic}}(G^*)}{P_{\mathrm{can}}(G^*)} = \log \frac{\widehat{P_{\mathrm{mic}}}(G_c^*)}{\widehat{P_{\mathrm{can}}}(G_c^*)} = S_n(\widehat{P_{\mathrm{mic}}} \mid \widehat{P_{\mathrm{can}}}). \tag{3.95}$$

CHAPTER 4

# Is Breaking of Ensemble Equivalence Monotone in the Number of Constraints?

This chapter is based on:
A. Roccaverde. Is breaking of ensemble equivalence monotone in the number of constraints? *Indagationes Mathematicae*, 2018

**Abstract**

Breaking of ensemble equivalence between the microcanonical ensemble and the canonical ensemble may occur for random graphs whose size tends to infinity, and is signaled by a non-zero specific relative entropy of the two ensembles. In [48] and [50] it was shown that breaking occurs when the constraint is put on the degree sequence (configuration model). It is not known what is the effect on the relative entropy when the number of constraints is reduced, i.e., when only part of the nodes are constrained in their degree (and the remaining nodes are left unconstrained). Intuitively, the relative entropy is expected to decrease. However, this is not a trivial issue because when constraints are removed both the microcanonical ensemble and the canonical ensemble change. In this paper a formula for the relative entropy valid for generic discrete random structures, recently formulated by Squartini and Garlaschelli, is used to prove that the relative entropy is monotone in the number of constraints when the constraint is on the degrees of the nodes. It is further shown that the expression for the relative entropy corresponds, in the dense regime, to the degrees in the microcanonical ensemble being asymptotically *multivariate Dirac* and in the canonical ensemble being asymptotically *Gaussian*.

# §4.1 Introduction and main results

## §4.1.1 Background

For most real-world networks, a detailed knowledge of the architecture of the network is not available and one must work with a probabilistic description, where the network is assumed to be a random sample drawn from a set of allowed configurations that are consistent with a set of known *topological constraints* [95]. Statistical physics deals with the definition of the appropriate probability distribution over the set of configurations and with the calculation of the resulting properties of the system. Two key choices of probability distribution are:

(1) the *microcanonical ensemble*, where the constraints are *hard* (i.e., are satisfied by each individual configuration);

(2) the *canonical ensemble*, where the constraints are *soft* (i.e., hold as ensemble averages, while individual configurations may violate the constraints).

(In both ensembles, the entropy is *maximal* subject to the given constraints.)

Breaking of ensemble equivalence means that different choices of the ensemble lead to asymptotically different behaviors. Consequently, while for applications based on ensemble-equivalent models the choice of ensemble can be based on mathematical convenience, for those based on ensemble-nonequivalent models the choice should be determined by the system one wants to apply to, i.e., dictated by a theoretical criterion that indicates *a priori* which ensemble is the appropriate one to be used. It is known that ensemble equivalence may be broken, signaled by a non-zero specific relative entropy between the two ensembles. It is expected that when the number of constraints grows extensively in the number of nodes, then typically there is breaking of ensemble equivalence. This has been shown to be the case when the setting is simple or bipartite graphs and the constraint is on the number of links (1 constraint and ensemble equivalence) or on the full degree sequence ($n$ constraints and non-equivalence) [48]. Later, in [50] and [38], also the dense regime was investigated and it was shown that the relative entropy between the two ensembles grows even faster. In general, the constraint is a multidimensional vector and its components represent the single quantities that are constrained. From now on, with the word 'constraint' we mean the 'vector constraint' and with the plural 'constraints' we mean the 'components' of the vector. This means when we talk about the number of constraints we actually mean the dimension of the vector constraint. In some cases this number can be very large, for example, when the constraint is on the degree sequence (a large number of nodes which need all to have the right degree).

Once the constraint becomes a function of the number $n$ of nodes (for example, the degree sequence), we can ask an interesting question: How is the relative entropy affected when the number of constraints is reduced, possibly in a way that depends on $n$? Intuitively, the relative entropy should decrease, but this is not a trivial issue because both the microcanonical and the canonical ensemble change when the constraints are changed. Of particular interest for the present paper is the main

result of [50]. There it was proven that, when a $\delta$-tame degree sequence is put as a constraint on the set of simple graphs, than the relative entropy between the two ensembles grows as $n \log n$. We consider random graphs with a prescribed partial degree sequence (reduced constraint). The breaking of ensemble equivalence is studied by analyzing how the relative entropy changes as a function of the number of constraints, in particular, it is shown that the relative entropy is a monotone function of the number of constraints. More precisely, when only $m$ nodes are constrained and the remaining $n - m$ nodes are left unconstrained, the relative entropy is shown to grow like $m \log n$. Our analysis is based on a recent formula put forward by Squartini and Garlaschelli [93]. This formula predicts that the relative entropy is determined by the covariance matrix of the constraints under the canonical ensemble, in the regime where the graph is dense. Our result implies that ensemble equivalence breaks down whenever the regime is $\delta$-tame, irrespective of the number of degrees $m$ that are constrained, provided $m$ is not of order $n$.

## Outline

Our paper is organized as follows. In Section 4.1 the background, the model and the main theorem are discussed. In Section 4.2 the main theorem is proved, together with a few basic lemmas that are needed along the way. Appendix A derives an expression for the canonical ensemble when a partial degree sequence is put as constraint. Appendix B discusses the $\delta$-tame condition for a partial degree sequence.

The remainder of Section 4.1 is organized as follows. In Section 4.1.1 we discussed the background of the problem. In Section 4.1.2 we describe the model when the constraint is put on the full degree sequence, in Section 4.1.3 when the constraint is put on the partial degree sequence. Here we also define the $\delta$-tame regime when the constraint is on the partial degree sequence. In Section 4.1.4 we state a formula for the relative entropy presented in [93] and state the main theorem. In section 4.1.5 we interpret the main theorem by stating how the degrees are distributed in the two ensembles.

The microcanonical and the canonical ensemble, as well as the relative entropy density have been defined in Section 1.4.1 and 1.4.2.

## §4.1.2 Constraint on the full degree sequence

The model of this section comes from [48] and [50]. The full degree sequence of a graph $G \in \mathcal{G}_n$ is defined as the vector $\vec{k}(G) = (k_i(G))_{i=1}^n$ with $k_i(G) = \sum_{j \neq i} g_{ij}(G)$. The degree sequence is set to a specific value $\vec{k}^*$, which we assume to be graphical, i.e., there is at least one graph with degree sequence $\vec{k}^*$. The constraint is therefore

$$\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^n \in \{1, 2, \ldots, n-2\}^n. \tag{4.1}$$

This constraint was studied in various regimes: in [48] in the sparse regime, and in [50] in the ultra-dense and the $\delta$-tame regime. The microcanonical ensemble, when the constraint is put on the degree sequence, is known as the *configuration model* and

has been studied in detail (see [95, 92, 99]). In the sparse (and in the ultra-dense) regime, the microcanonical ensemble cannot be computed exactly, but there are good approximations with an error that is vanishing when the relative entropy is computed in the limit as $n \to \infty$ [48], [9]. In the $\delta$-tame regime, this approximation does not hold, but the relative entropy can still be investigated with other tools [50]. The canonical ensemble can be computed in every regime and takes the form

$$P_{\text{can}}(G) = \prod_{1 \le i < j \le n} \left(p_{ij}^*\right)^{g_{ij}(G)} \left(1 - p_{ij}^*\right)^{1 - g_{ij}(G)}, \tag{4.2}$$

with

$$p_{ij}^* = \frac{e^{-\theta_i^* - \theta_j^*}}{1 + e^{-\theta_i^* - \theta_j^*}}, \tag{4.3}$$

and with the vector of Lagrange multipliers $\vec{\theta}^* = (\theta_i^*)_{i=1}^n$ tuned such that

$$\langle k_i \rangle = \sum_{\substack{1 \le j \le n \\ j \ne i}} p_{ij}^* = k_i^*, \qquad \forall\ 1 \le i \le n. \tag{4.4}$$

The results in [48] show that there is breaking of ensemble equivalence with $\alpha_n = n$ when the regime is sparse and ultra-dense. The results in [50] show that the relative entropy grows like $\alpha_n = \frac{1}{2} n \log n$. The purpose of the paper is to investigate what happens when part of the $n$ constraints degrees are removed and how the relative entropy is affected by this. In the next section the partial constraint is presented and the main theorem is stated.



## §4.1.3 Constraint on the partial degree sequence

In this section we look at a different model. The constraint is put on the partial degree sequence instead of on the full degree sequence, more precisely, only the first $m < n$ nodes are constrained while the remaining nodes are left unconstrained. The partial degree sequence of a graph $G \in \mathcal{G}_n$ is defined as the vector $\vec{k}(G) = (k_i(G))_{i=1}^m$ where $k_i(G) = \sum_{1 \le j \ne j \le n} g_{ij}(G)$. The constraint is set to be a *specific m-dimensional* vector $\vec{k}^*$, which we assume to be *graphical*, i.e., there exist at least one graph $G^* \in \mathcal{G}_n$ with partial degree sequence $\vec{k}^*$. The constraint is therefore

$$\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^m \in \{1, 2, \ldots, n-2\}^m, \tag{4.5}$$

As mentioned above, the microcanonical ensemble can be computed approximately when the constraint is put on the full degree sequence. However, when the constraint is put on the partial degree sequence, no good approximation is available. The situation is different for the canonical ensemble, which can still be computed. Appendix A is dedicated to the study of the canonical ensemble when a partial degree sequence is put as a constraint. This leads to

$$P_{\text{can}}(G) = 2^{-\binom{n-m}{2}} \prod_{1 \le i < j \le m} \left(p_{ij}^*\right)^{g_{ij}(G)} \left(1 - p_{ij}^*\right)^{1 - g_{ij}(G)} \prod_{i=1}^{m} \left(p_i^*\right)^{s_i(G)} \left(1 - p_i^*\right)^{n - m - s_i(G)} \tag{4.6}$$

with

$$p^*_{ij} = \frac{e^{-\theta^*_i-\theta^*_j}}{1+e^{-\theta^*_i-\theta^*_j}}, \qquad p^*_i = \frac{e^{-\theta^*_i}}{1+e^{-\theta^*_i}}, \qquad s_i(G) = \sum_{j=m+1}^{n} g_{ij}(G), \tag{4.7}$$

and with the vector of Lagrange multipliers $\vec{\theta}^* = (\theta^*_i)_{i=1}^m$ tuned such that

$$\langle k_i \rangle = \sum_{\substack{1\le j\le m\\ j\neq i}} p^*_{ij} + (n-m)p^*_i = k^*_i, \qquad 1 \le i \le m. \tag{4.8}$$

The canonical ensemble has an interesting dual structure, consisting of the product of two canonical probabilities, which we call unipartite probability and bipartite probability, and an overall factor $2^{-\binom{n-m}{2}}$. The unipartite probability,

$$\prod_{1\le i<j\le m} \left(p^*_{ij}\right)^{g_{ij}(G)} \left(1-p^*_{ij}\right)^{1-g_{ij}(G)},$$

is precisely the canonical ensemble obtained when the constraint is put on the full degree sequence $\vec{u}^* = (u^*_i)_{i=1}^m$, with $u^*_i = \sum_{\substack{1\le j\le m\\ j\neq i}} p^*_{ij}$, on the subset of graphs with $m$ nodes $\mathcal{G}_m$. The bipartite probability,

$$\prod_{i=1}^{m} \left(p^*_i\right)^{s_i(G)} \left(1-p^*_i\right)^{n-m-s_i(G)},$$

is precisely the canonical bipartite probability obtained when the constraint is put on the top layer of a bipartite graph. More precisely, the configuration space is the set of bipartite graphs $\mathcal{G}_{m,n-m}$ with $m$ nodes in the top layer and $n-m$ nodes in the bottom layer. The constraint is put on the degree sequence in the top layer only and corresponds to the vector $\vec{b}^* = (b^*_i)_{i=1}^m$ with $b^*_i = (n-m)p^*_i$. Moreover, the average $i$-th degree $\langle k_i \rangle$ with respect to the canonical ensemble (4.6) equals $k^*_i$ and is given by the balance equation (4.8). This equation shows that the $i$-th unipartite constraint $u^*_i$ and the $i$-th bipartite constraint $b^*_i$ sum up to the $i$-th original constraint $k^*_i$.

**4.1.1 Definition.** A partial degree sequence $\vec{k}^* = (k^*_i)_{i=1}^m$, put as a constraint on the set of configurations $\mathcal{G}_n$ with $m < n$, is said to be $\delta$-tame if and only if there exists a $\delta \in \left(0, \frac{1}{2}\right]$ such that

$$\delta \le p^*_{ij} \le 1-\delta, \qquad 1 \le i \neq j \le m, \tag{4.9}$$

where $p^*_{ij}$ are the canonical probabilities in (4.6)–(4.8).

It is easy to prove that, given a $\delta$-tame partial degree sequence $\vec{k}^* = (k^*_i)_{i=1}^m$, the bipartite probabilities $(p^*_i)_{i=1}^m$ are also $\delta$-tame, namely, satisfy

$$\delta' \le p^*_i \le 1-\delta', \qquad \forall\ 1 \le i \le m, \tag{4.10}$$

for some $\delta' \in \left(0, \frac{1}{2}\right]$. This is discussed in more detail in Appendix B. Condition (4.9) has a trivial implication for the degree sequence:

$$(m-1)\delta \le u^*_i \le (m-1)(1-\delta), \qquad 1 \le i \le m, \tag{4.11}$$

$$(n-m)\delta' \le b_i^* \le (n-m)(1-\delta'), \qquad 1 \le i \le m. \tag{4.12}$$

Since $\delta' = \frac{1}{1+(\frac{1-\delta}{\delta})^{3/2}} < \delta$ for all $\delta \in [0, 1/2)$ and $u_i^* + b_i^* = k_i^*$, it follows that

$$(n-1)\delta' \le k_i^* \le (n-1)(1-\delta'), \qquad 1 \le i \le m. \tag{4.13}$$

This means that $\delta$-tame graphs are neither too thin (sparse regime) nor too dense (ultra-dense regime). It is natural to ask whether, conversely, condition (4.13), or a similar condition involving only the original degrees $\vec{k}^* = (k_i^*)_{i=1}^m$, is sufficient to prove that the partial degree sequence is $\delta$-tame for some $\delta = \delta(\delta')$, in the sense of Definition 4.1.1. Unfortunately, this question is not easy to settle, but the following lemma provides a partial answer.

**4.1.2 Lemma.** *Suppose that $\vec{k}^* = (k_i^*)_{i=1}^m$ satisfies*

$$(n-1)\delta' + (n-m) \le k_i^* \le (n-1)(1-\delta'), \qquad 1 \le i \le m, \tag{4.14}$$

*for some $\delta' \in (\frac{1}{4}, \frac{1}{2}]$. Then there exist $\delta = \delta(\delta') > 0$ and $n_0 = n_0(\delta') \in \mathbb{N}$ such that $\vec{k}^* = (k_i^*)_{i=1}^m$ is a $\delta$-tame partial degree sequence, in the sense of Definition 4.1.1, for all $n \ge n_0$.*

*Proof.* Condition (4.14), with $u_i^* = k_i^* - b_i^*$ and $b_i^* \in [0, n-m]$, gives

$$(n-1)\delta' \le u_i^* \le (n-1)(1-\delta'), \qquad 1 \le i \le m. \tag{4.15}$$

The proof follows from (4.15) and [9, Theorem 2.1]. In fact, applying that theorem with $\alpha = \delta'$, $\beta = 1 - \delta'$ and with $\delta' > \frac{1}{4}$, we get

$$\delta \le p_{ij}^* \le 1 - \delta, \qquad 1 \le i \ne j \le m. \tag{4.16}$$

Moreover, [9, Theorem 2.1] also gives information about the values of $\delta = \delta(\delta')$ and $n_0 = n_0(\delta')$.

## §4.1.4 Linking ensemble nonequivalence to the canonical covariances

In this section we describe an important formula, recently put forward in [93], for the scaling of the relative entropy under a general constraint. The analysis in [93] allows for the possibility that not all the constraints (i.e., not all the components of the vector $\vec{C}$) are linearly independent. For instance, $\vec{C}$ may contain redundant replicas of the same constraint(s), or linear combinations of them. Since in the present paper we only consider the case where $\vec{C}$ is the degree sequence, the different components of $\vec{C}$ (i.e., the different degrees) are linearly independent.

When a $K$-dimensional constraint $\vec{C}^* = (C_i^*)_{i=1}^K$ with independent components is imposed, then a key result in [93] is the formula

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) \sim \log \frac{\sqrt{\det(2\pi Q)}}{T}, \qquad n \to \infty, \tag{4.17}$$

where

$$Q = (q_{ij})_{1 \le i,j \le K} \tag{4.18}$$

is the $K \times K$ covariance matrix of the constraints under the canonical ensemble, whose entries are defined as

$$q_{ij} = \mathrm{Cov}_{P_{\mathrm{can}}}(C_i, C_j) = \langle C_i\, C_j \rangle - \langle C_i \rangle \langle C_j \rangle, \tag{4.19}$$

and

$$T = \prod_{i=1}^{K} \left[ 1 + O\left( 1/\lambda_i^{(K)}(Q) \right) \right], \tag{4.20}$$

with $\lambda_i^{(K)}(Q) > 0$ the $i$-th eigenvalue of the $K \times K$ covariance matrix $Q$. This result can be formulated more rigorously as follows.

**4.1.1 Formula ([93]).** If all the constraints are linearly independent, then the limiting relative entropy $\alpha_n$-density equals

$$s_{\alpha_\infty} = \lim_{n \to \infty} \frac{\log \sqrt{\det(2\pi Q)}}{\alpha_n} + \tau_{\alpha_\infty} \tag{4.21}$$

with

$$\tau_{\alpha_\infty} = - \lim_{n \to \infty} \frac{\log T}{\alpha_n}. \tag{4.22}$$

The latter is zero when

$$\lim_{n \to \infty} \frac{|I_{K_n,R}|}{\alpha_n} = 0 \quad \forall\, R < \infty, \tag{4.23}$$

where $I_{K,R} = \{ i = 1, \ldots, K \colon \lambda_i^{(K)}(Q) \le R \}$ with $\lambda_i^{(K)}(Q)$ the $i$-th eigenvalue of the $K$-dimensional covariance matrix $Q$ (the notation $K_n$ indicates that $K$ may depend on $n$). Note that $0 \le I_{K,R} \le K$. Consequently, (4.23) is satisfied (and hence $\tau_{\alpha_\infty} = 0$) when $\lim_{n \to \infty} K_n/\alpha_n = 0$, i.e., when the number $K_n$ of constraints grows slower than $\alpha_n$.

**4.1.3 Remark ([93]).** Formula 4.1.1, for which [93] offers compelling evidence but not a mathematical proof, can be rephrased by saying that the natural choice of $\alpha_n$ is

$$\tilde{\alpha}_n = \log \sqrt{\det(2\pi Q)}. \tag{4.24}$$

Indeed, if all the constraints are linearly independent and (4.23) holds, then $\tau_{\tilde{\alpha}_n} = 0$ and

$$s_{\tilde{\alpha}_\infty} = 1, \tag{4.25}$$

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = [1 + o(1)]\, \tilde{\alpha}_n. \tag{4.26}$$

Formula 4.1.1 has been verified in several examples, namely, all the models in [48] and [50].

Next we present our main theorem, which considers the case where the constraint is on the partial degree sequence $\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^m$ in the *$\delta$-tame regime* defined in Definition 4.1.1.

**4.1.4 Theorem.** *Suppose that:*

- *The constraint is put on the partial degree sequence* $\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^m$ *on the space of simple graphs* $\mathcal{G}_n$ *with* $0 \leq m \leq n$.
- $\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^m$ *is a* $\delta$*-tame partial degree sequence, namely, the canonical probabilities* $(p_{ij}^*)_{1 \leq i \neq j \leq m}$ *satisfy*

$$\delta \leq p_{ij}^* \leq 1 - \delta, \quad 1 \leq i \neq j \leq m. \tag{4.27}$$

- *Formula 4.1.1 is valid in the above framework.*
- *The scale parameter is* $\alpha_n = \dfrac{m \log n}{2}$.
- $m = m(n)$ *satisfies*

$$\lim_{n \to \infty} \frac{n - m}{m} \log n = \infty. \tag{4.28}$$

*Then there is breaking of ensemble equivalence, and*

$$s_{\alpha_\infty} = \lim_{n \to \infty} s_{\alpha_n} = 1. \tag{4.29}$$

Condition (4.28) fails when $n - m = O(\frac{m}{\log n})$, i.e., when the number of unconstrained nodes is sufficiently small. We expect that (4.29) continues to hold even in this case, but our proof breaks down.



## §4.1.5 Discussion

Theorem 4.1.4 analyses the relative entropy at a macroscopic level, but says nothing about what happens at the microscopic level. More precisely, it does not identify how the relative entropy changes when a single constraint is removed, rather than a positive fraction of constraints. A microscopic analysis could reveal what is the effect when e.g. the longest degree is removed, or the smallest degree, or any other degree. The result in Theorem 4.1.4 is far from trivial. In fact, when the number of constraints is reduced, it can become either easier or more difficult to compute microcanonical and canonical ensembles. The case when the constraint is put on the degree sequence provides a clear example. If the constraint is put on the full degree sequence, then the microcanonical ensemble can be *asymptotically computed* [9]. As soon as one or more degrees are removed (meaning that some nodes are left unconstrained), the structure of the problem changes completely. The symmetry of the constraints is broken by the removal, and this makes it more difficult to compute the number of graphs with a prescribed partial degree sequence. On the other hand, the canonical problem can still be solved and has an interesting structure (Appendix A). This makes it possible to use the formula proposed by Garlaschelli and Squartini [93], which only makes use of the canonical ensemble to analyze the relative entropy between the two ensembles. Theorem 4.1.4 clearly exhibits the monotonicity property of the relative entropy in the case where the constraint is put on the degrees. Indeed, under the hypotheses

written above, the relative entropy $S_n(P_{\rm mic} \mid P_{\rm can})$ grows like $m \log n$, where $m$ is the number of constrained nodes and $n$ is the total number of nodes. This shows that the relative entropy is monotone in the number of constraints on scale $n$.

We next provide a heuristic explanation for Theorem 4.1.4 (in analogy with what was done in [48] and [50]).

**Heuristic explanation of Theorem 4.1.4.** Using (1.16), we can write the relative entropy between the ensembles as

$$S_n(P_{\rm mic} \mid P_{\rm can}) = \log \frac{P_{\rm mic}(G^*)}{P_{\rm can}(G^*)} = -\log[\Omega^n_{\vec{k}^*} P_{\rm can}(G^*)] = -\log Q^n[\vec{k^*}](\vec{k^*}), \quad (4.30)$$

where $\Omega^n_{\vec{k}^*}$ is the number of graphs with $n$ nodes and partial degree sequence $\vec{k}^* = (k_i^*)_{i=1}^m$,

$$Q^n[\vec{k^*}](\vec{k}\,) = \Omega^n_{\vec{k}}\, P_{\rm can}\big(G^{\vec{k}}\big) \quad (4.31)$$

is the probability that the partial degree sequence is equal to $\vec{k}$ under the canonical ensemble with constraint $\vec{k}^*$, $G^{\vec{k}}$ denotes an arbitrary graph with partial degree sequence $\vec{k}$, and $P_{\rm can}\big(G^{\vec{k}}\big)$ is the canonical probability rewritten for one such graph. Indeed, (1.9) shows that the canonical probability is constant for all graphs with the same constraint, in our case, for all graphs with the same partial degree sequence. Using (1.9) and (4.65) we can rewrite the canonical probability in the form

$$P_{\rm can}\big(G^{\vec{k}}\big) = 2^{-\binom{n-m}{2}} \prod_{i=1}^m \frac{{x_i^*}^{k_i}}{(1+x_i^*)^{n-m}} \prod_{1\le i<j\le m} (1+x_i^* x_j^*)^{-1}, \quad (4.32)$$

where $x_i^* = e^{-\theta_i^*}$, and $\vec{\theta}^* = (\theta_i^*)_{i=1}^m$ is the vector of Lagrange multipliers coming from (4.7). Equation (4.30) can be rewritten as

$$S_n(P_{\rm mic} \mid P_{\rm can}) = S\big(\,\delta[\vec{k^*}] \mid Q^n[\vec{k^*}]\,\big), \quad (4.33)$$

where $\delta[\vec{k^*}] = \prod_{i=1}^m \delta[k_i^*]$ is the *multivariate Dirac distribution* with average $\vec{k^*}$. We can interpret the relative entropy between $P_{\rm mic}$ and $P_{\rm can}$ *on the set of graphs* $\mathcal{G}_n$ as the relative entropy between $\delta[\vec{k^*}]$ and $Q^n[\vec{k^*}]$ *on the set of degree sequences*. More precisely, combining (4.31) and (4.32), we can rewrite $Q^n[\vec{k^*}](\vec{k})$ as

$$Q^n[\vec{k^*}](\vec{k}) = \Omega^n_{\vec{k}}\; 2^{-\binom{n-m}{2}} \prod_{i=1}^m \frac{{x_i^*}^{k_i}}{(1+x_i^*)^{n-m}} \prod_{1\le i<j\le m} (1+x_i^* x_j^*)^{-1}. \quad (4.34)$$

The distribution in (4.34) is a multivariate version of the *Poisson-Binomial distribution* [100]. The univariate Poisson-Binomial distribution describes the probability of a certain number of successes out of a total number of independent, possibly non-identical, Bernoulli trials. In our case, the marginal probability that node $i$ has degree $k_i$, under the canonical ensemble, irrespectively of the degree of any other node, is a univariate Poisson-Binomial distribution with a total number of $n-1$ Bernoulli trials: $m-1$ independent Bernoulli trials with success probabilities $\{p_{ij}^*\}_{1\le j\ne i\le m}$ and $n-m$

independent Bernoulli trials with the same success probability $p_i^*$. The relation in (4.33) therefore becomes

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) = S\big(\,\delta[\vec{k^*}] \mid \text{PoissonBinomial}[\vec{k^*}]\,\big), \tag{4.35}$$

where PoissonBinomial$[\vec{k^*}]$ is the multivariate Poisson-Binomial distribution given by (4.34), i.e.,

$$Q^n[\vec{k^*}] = \text{PoissonBinomial}[\vec{k^*}]. \tag{4.36}$$

The relative entropy between the microcanonical and the canonical ensemble can be seen as coming from the limiting situation in which the microcanonical ensemble forces the degree sequence to be exactly $\vec{k^*}$, while the canonical ensemble forces the degree sequence to be distributed as a multivariate Poisson-Binomial with average $\vec{k^*}$.

**Two different regimes for the Poisson-Binomial distribution.** As has already been said in Chapter 3, the univariate Poisson-Binomial distribution admits two asymptotic limits: Poisson or Gaussian [100]. A Poisson limit occurs whenever $\sum_{j\neq i} p_{ij}^* \to \lambda > 0$ and $\sum_{j\neq i}(p_{ij}^*)^2 \to 0$ as $n \to \infty$, while a Gaussian limit occurs whenever $p_{ij}^* \to \lambda_j > 0$ for all $j \neq i$ as $n \to \infty$. In the simple case of identical Bernoulli trials, i.e., all the probabilities $\{p_{ij}^*\}_{j\neq i}$ are equal, the univariate Poisson-Binomial distribution reduces to a Binomial distribution, which is known to admit Poisson and Gaussian limits. This implies that also the multivariate Poisson-Binomial distribution in (4.34) admits limits that should be consistent with the Poisson and Gaussian ones for its marginals. Below we present two different situations.



**Gaussian constrained degrees in the $\delta$-tame regime.** Comparing (4.25) and (4.29), and using (4.24), we see that Theorem 4.1.4 shows that if the constraint is on the partial degree sequence, then

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) \sim m \log n \sim \log \sqrt{\det(2\pi Q)} \tag{4.37}$$

in the $\delta$-tame regime and under the condition in (4.28). Equation (4.37) can be reinterpreted as

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) \sim S\big(\,\delta[\vec{k^*}] \mid \text{Normal}[\vec{k^*}, Q]\,\big), \tag{4.38}$$

where Normal$[\vec{k^*}, Q]$ is the *multivariate Normal distribution* with mean $\vec{k^*}$ and covariance matrix $Q$. Basically, in the $\delta$-tame regime,

$$Q^n[\vec{k^*}] \sim \text{Normal}[\vec{k^*}, Q]. \tag{4.39}$$

The multivariate Poisson-Binomial distribution in (4.34) becomes asymptotically a multivariate Gaussian distribution which, in general, has a non-diagonal covariance matrix, i.e., there can be dependence between the degrees of the different nodes.

The right-hand side of (4.38) deserves clarification, because it has to be properly interpreted. In fact, it describes the relative entropy of a discrete distribution with respect to a continuous distribution. Technically, the Dirac distribution $\delta[\vec{k^*}]$ must be smoothed to a continuous distribution with support on a small ball around $\vec{k^*}$. Since the degrees are large, this does not affect the asymptotics.

**Poisson-Binomial unconstrained degrees in the $\delta$-tame regime.** It is interesting to study the distribution of the degrees of the unconstrained nodes in the canonical ensemble. The canonical probability of the $(m+1)$-th degree (the first unconstrained node) can be computed and the same steps can be used to compute the canonical probabilities of the other unconstrained nodes, which follow the same probability law. The canonical probability that the $(m+1)$-th node is equal to some value $x \in \{0, 1, \dots, n-1\}$ can be written as:

$$\begin{aligned}
&\sum_{\substack{G\in\mathcal{G}_n\\ k_{m+1}(G)=x}} P_{\text{can}}(G)\\
&= 2^{-\binom{n-m}{2}} \sum_{\substack{G\in\mathcal{G}_n\\ k_{m+1}(G)=x}} \prod_{1\le i<j\le m} \left(p^*_{ij}\right)^{g_{ij}(G)} \left(1-p^*_{ij}\right)^{1-g_{ij}(G)}\\
&\prod_{i=1}^{m} \left(p^*_i\right)^{s_i(G)} \left(1-p^*_i\right)^{n-m-s_i(G)}\\
&= 2^{-\binom{n-m}{2}} \sum_{G\in A} \prod_{i=1}^{m} \left(p^*_i\right)^{g_{im+1}(G)} \left(1-p^*_i\right)^{1-g_{im+1}(G)}\\
&= 2^{-\binom{n-m}{2}} 2^{\binom{n-m-1}{2}} \sum_{G\in A\cap B} \prod_{i=1}^{m} {p^*_i}^{g_{im+1}(G)} \left(1-p^*_i\right)^{1-g_{im+1}(G)}\\
&= \sum_{G\in A\cap B} \left(\tfrac{1}{2}\right)^{(n-m-1)} \prod_{i=1}^{m} {p^*_i}^{g_{im+1}(G)} \left(1-p^*_i\right)^{1-g_{im+1}(G)}\\
&= P(Po-Bi[p^*_1,\dots,p^*_m,\tfrac{1}{2},\dots,\tfrac{1}{2}]=x),
\end{aligned} \tag{4.40}$$

where

$$\begin{aligned}
A &= \{G\in\mathcal{G}_n : k_{m+1}(G)=x,\ g_{ij}(G)=0\ \forall i\in[1,m],\ j=\in[1,n]\setminus\{m+1\},\ i\neq j\},\\
B &= \{G\in\mathcal{G}_n : g_{ij}(G)=0\ \forall i=m+2,\dots,n,\ j=m+2,\dots,n,\ i\neq j\},
\end{aligned} \tag{4.41}$$

and $Po-Bi[p^*_1,\dots,p^*_m,\frac{1}{2},\dots,\frac{1}{2}]$ is the Poisson-Binomial distribution given by the $m$ independent trials $p^*_i$, $i=1,\dots,m$, and the $n-m-1$ independent Bernoulli trials with the same success probability $\frac{1}{2}$. This means that, for each $j=m+1,\dots,n$, the canonical probability of the degree of the $j$-th node is distributed as a Poisson-Binomial random variable with $n-1$ entries: $p^*_1,\dots,p^*_1,\frac{1}{2},\dots,\frac{1}{2}$.

## §4.2 Proof of the Main Theorem

The proof is based on two lemmas, which are stated and proved in Section 4.2.1. In Section 4.2.2 Theorem 4.1.4 is proved.

### §4.2.1 Preparatory lemmas

The following lemma gives an expression for the relative entropy.

**4.2.1 Lemma.** *If the constraint is on the partial degree sequence* $(k_i^*)_{i=1}^m$*, then the relative entropy in* (1.16) *equals*

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) = \tfrac{1}{2}\log[\det(2\pi Q)] - \log T^*, \tag{4.42}$$

*where* $Q$ *is the covariance matrix in* (4.18) *and* $T^*$ *is the error in* (4.20)*. The matrix* $Q = (q_{ij})$ *takes the form*

$$\begin{cases} q_{ii} = \sum_{1\le j\le m, j\neq i} p_{ij}^*(1-p_{ij}^*) + (n-m)p_i^*(1-p_i^*), & 1 \le i \le m, \\ q_{ij} = p_{ij}^*(1-p_{ij}^*), & 1 \le i \neq j \le m. \end{cases} \tag{4.43}$$

*Proof.* To compute $q_{ij} = \text{Cov}_{P_{\text{can}}}(k_i, k_j)$, take the second order derivatives of the log-likelihood function

$$\begin{aligned} &\mathcal{L}(\vec{\theta}) = \log P_{\text{can}}(G^* \mid \vec{\theta}) \\ &= \log\left[2^{-\binom{n-m}{2}} \prod_{1\le i<j\le n} p_{ij}^{g_{ij}(G^*)}(1-p_{ij})^{(1-g_{ij}(G^*))} \prod_{i=1}^m p_i^{s_i(G^*)}(1-p_i)^{n-m-s_i(G^*)}\right], \end{aligned} \tag{4.44}$$

with

$$p_{ij} = \frac{e^{-\theta_i-\theta_j}}{1+e^{-\theta_i-\theta_j}}, \quad p_i = \frac{e^{-\theta_i}}{1+e^{-\theta_i}}, \tag{4.45}$$

in the point $\vec{\theta} = \vec{\theta}^*$ [93]. It is easy to show that the first-order derivatives are [51]

$$\frac{\partial}{\partial\theta_i}\mathcal{L}(\vec{\theta}) = \langle k_i\rangle - k_i^*, \quad \frac{\partial}{\partial\theta_i}\mathcal{L}(\vec{\theta})\Big|_{\vec{\theta}=\vec{\theta}^*} = k_i^* - k_i^* = 0 \tag{4.46}$$

and the second-order derivatives are

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\mathcal{L}(\vec{\theta})\Big|_{\vec{\theta}=\vec{\theta}^*} = \langle k_i\rangle\langle k_j\rangle - \langle k_i\, k_j\rangle = -\text{Cov}_{P_{\text{can}}}(k_i, k_j). \tag{4.47}$$

Taking the second-order derivatives of the log-likelihood function, we get (4.43). The proof of (4.42) uses [93, Formula 25].

The following lemma shows that a diagonal approximation of the matrix $Q$ is good for a $\delta$-tame partial degree sequence and $\alpha_n = m\log n$.

**4.2.2 Lemma.** *Under the* $\delta$*-tame condition,*

$$\log(\det Q_D) + o(m\log n) \le \log(\det Q) \le \log(\det Q_D) \tag{4.48}$$

*with* $Q_D = \text{diag}(Q)$ *the matrix that coincides with* $Q$ *on the diagonal and is zero off the diagonal.*

*Proof.* Use [60, Theorem 2.3], which says that if

(1) $\det(Q)$ is real,

(2) $Q_D$ is non-singular with $\det(Q_D)$ real,

(3) $\lambda_i(A) > -1$, $1 \le i \le m$,

then

$$e^{-\frac{m\rho^2(A)}{1+\lambda_{\min}(A)}} \det Q_D \le \det Q \le \det Q_D. \tag{4.49}$$

Here, $A = Q_D^{-1} Q_{\text{off}}$, with $Q_{\text{off}}$ the matrix that coincides with $Q$ off the diagonal and is zero on the diagonal, $\lambda_i(A)$ is the $i$-th eigenvalue of $A$ (arranged in decreasing order), $\lambda_{\min}(A) = \min_{1\le i\le n} \lambda_i(A)$, and $\rho(A) = \max_{1\le i\le n} |\lambda_i(A)|$.
We verify (1)–(3).
(1) Since $Q$ is a symmetric matrix with real entries, $\det Q$ exists and is real.

(2) This property holds thanks to the $\delta$-tame condition and Lemma B.1. In fact

$$0 < \delta^2 \le q_{ij} \le (1-\delta)^2 < 1, \tag{4.50}$$

and

$$(m-1)\delta^2 + (n-m)\delta'^2 \le q_{ii} =\le (m-1)(1-\delta)^2 + (n-m)(1-\delta')^2. \tag{4.51}$$

(3) It is easy to show that $A = (a_{ij})$ is given by

$$a_{ij} = \begin{cases} \frac{q_{ij}}{q_{ii}} = \frac{p^*_{ij}(1-p^*_{ij})}{\sum_{1\le k\le m, k\ne i} p^*_{ik}(1-p^*_{ik}) + (n-m)p^*_i(1-p^*_i)}, & 1 \le i \ne j \le m \\ 0 & 1 \le i = j \le m, \end{cases} \tag{4.52}$$

where $q_{ij}$ is given by (4.43). The Gershgorin circle theorem says the eigenvalues of the matrix $A$ satisfy

$$|\, \lambda_i(A) \,| \le R_i = \sum_{j\ne i} a_{ij} = \frac{\sum_{1\le k\le m, k\ne i} p^*_{ik}(1-p^*_{ik})}{\sum_{1\le k\le m, k\ne i} p^*_{ik}(1-p^*_{ik}) + (n-m)p^*_i(1-p^*_i)}, \quad 1 \le i \le m. \tag{4.53}$$

Using the $\delta$-tame condition, we find the bound

$$|\, \lambda_i(A) \,| \le \max_{1\le i\le m} R_i < 1 - A(\delta), \tag{4.54}$$

with $A(\delta) = \frac{(n-m)\delta'}{(m-1)(1-\delta)^2 + (n-m)(1-\delta')}$. In principle, $A(\delta)$ also depends on $\delta'$, but $\delta'$ is itself function of $\delta$. Equation (4.54) immediately gives $\rho(A) < 1$, namely

$$-\frac{m\rho^2(A)}{1+\lambda_{\min}(A)} > -\frac{m}{1+\lambda_{\min}(A)}. \tag{4.55}$$

Next we show that

$$-\frac{m}{1+\lambda_{\min}(A)} = o(m \log n). \tag{4.56}$$

Together with (4.49) this will settle the claim in (4.48). We must show that

$$\lim_{n\to\infty} (1 + \lambda_{\min}(A)) \log n = \infty. \tag{4.57}$$

Using equation (4.54) again, it follows $1 + \lambda_{\min}(A) > A(\delta)$. Therefore it suffices to prove that

$$\lim_{n\to\infty} A(\delta) \log n = \infty. \tag{4.58}$$

The result is trivial when $A(\delta)$ is constant ($\frac{n-m}{m} \to$ constant) or $A(\delta) \to \infty$ ($\frac{n-m}{m} \to \infty$) . On the other hand, when $A(\delta) \to 0$ ($\frac{n-m}{m} \to 0$), the condition $\frac{n-m}{m} \log n \to \infty$ is needed to conclude the proof.

## §4.2.2 Proof (Theorem 4.1.4)

*Proof.* When $\alpha_n = \frac{m \log n}{2}$, Lemma 4.2.1 says

$$\lim_{n\to\infty} \frac{S_n(P_{\text{mic}} \mid P_{\text{can}})}{\alpha_n} = \lim_{n\to\infty} \frac{\log 2\pi}{m \log n} + \lim_{n\to\infty} \frac{\log(\det Q)}{m \log n} - \lim_{n\to\infty} \frac{\log T^*}{2m \log n}. \tag{4.59}$$

The last term (the error) tends to zero. In fact, in [93] it is proved that $\lim_{n\to\infty} \frac{\log T^*}{m \log n} = 0$ unless the number of eigenvalues of $Q$ that have a finite limit as $n \to \infty$ which is indeed the case when a partial $\delta$-tame degree sequence is put as a constraint and $\alpha_n = \frac{m \log n}{2}$.
Using the $\delta$-tame condition, we get from Lemma 4.2.2 that

$$\lim_{n\to\infty} \frac{\log(\det Q)}{m \log n} = \lim_{n\to\infty} \frac{\log(\det Q_D)}{m \log n}. \tag{4.60}$$

To conclude the proof it therefore suffices to show that

$$\lim_{n\to\infty} \frac{\log(\det Q_D)}{m \log n} = 1. \tag{4.61}$$

Using (4.51), we have

$$\begin{aligned} &\frac{\log[(m-1)\delta^2 + (n-m)\delta']}{\log n} \le \frac{\sum_{i=1}^m \log q_{ii}}{m \log n} = \frac{\log(\det Q_D)}{m \log n} \\ &\le \frac{\log[(m-1)(1-\delta)^2 + (n-m)(1-\delta')]}{\log n}. \end{aligned} \tag{4.62}$$

Both sides tend to 1 as $n \to \infty$, and so (4.61) follows.

# §A Appendix

In this appendix we identify the structure of the canonical ensemble when the constraint is put on the partial degree sequence for the first $m < n$ nodes. The partial degree sequence $\vec{k}(G) = (k_i(G))_{i=1}^m$ is set to a *specific m-dimensional* vector $\vec{k}^*$, which is assumed to be *graphical*, i.e., there is at least one graph $G^* \in \mathcal{G}_n$ with partial degree sequence $\vec{k}^*$. The constraint is therefore

$$\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^m \in \{1, 2, \ldots, n-2\}^m. \tag{4.63}$$

The canonical ensemble has Hamiltonian $H(G,\vec{\theta}) = \sum_{i=1}^{m} \theta_i k_i(G)$, where $G$ is a graph belonging to $\mathcal{G}_n$, and $k_i(G) = \sum_{j\neq i} g_{ij}(G)$ is the degree of node $i$. It is easy to transform the Hamiltonian into

$$H(G,\vec{\theta}) = \sum_{1\leq i<j\leq m} (\theta_i+\theta_j) g_{ij}(G) + \sum_{i=1}^{m} \theta_i \sum_{j=m+1}^{n} g_{ij}(G) \tag{4.64}$$

Using this form, we see that the partition function equals

$$\begin{aligned} Z(\vec{\theta}) &= \sum_{G\in\mathcal{G}_n} e^{-H(G,\vec{\theta})} = \sum_{G\in\mathcal{G}_n} \prod_{1\leq i<j\leq m} e^{-(\theta_i+\theta_j)g_{ij}(G)} \prod_{i=1}^{m}\prod_{j=m+1}^{n} e^{-\theta_i g_{ij}(G)} \\ &= 2^{\binom{n-m}{2}} \prod_{1\leq i<j\leq m} (1+e^{-(\theta_i+\theta_j)}) \prod_{i=1}^{m}\prod_{j=m+1}^{n} (1+e^{-\theta_i}) \\ &= 2^{\binom{n-m}{2}} \prod_{1\leq i<j\leq m} (1+e^{-(\theta_i+\theta_j)}) \prod_{i=1}^{m} (1+e^{-\theta_i})^{(n-m)}. \end{aligned} \tag{4.65}$$

Inserting the partition function into the canonical expression, we get

$$P_{\text{can}}(G \mid \theta) = 2^{-\binom{n-m}{2}} \prod_{1\leq i<j\leq m} p_{ij}^{g_{ij}(G)} (1-p_{ij})^{1-g_{ij}(G)} \prod_{i=1}^{m} p_i^{s_i(G)} (1-p_i)^{n-m-s_i(G)} \tag{4.66}$$

with

$$p_{ij} = \frac{e^{-\theta_i-\theta_j}}{1+e^{-\theta_i-\theta_j}}, \qquad p_i = \frac{e^{-\theta_i}}{1+e^{-\theta_i}}, \qquad s_i(G) = \sum_{j=m+1}^{n} g_{ij}(G). \tag{4.67}$$

It remains to tune the Lagrange multipliers to the values such that the average constraint equals the vector $\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^m \in \{1,2,\ldots,n-2\}^m$. The average energy of the $i$-th degree with respect to the probability distribution $P_{\text{can}}(\cdot \mid \theta)$ corresponds to the derivative with respect to $\theta_i$ of the logarithm of the partition function (free energy). This means that the values $(\theta_i^*)_{i=1}^m$ must satisfy

$$\langle k_i \rangle = \sum_{\substack{1\leq j\leq m \\ j\neq i}} p_{ij}^* + (n-m)p_i^* = k_i^*, \qquad 1\leq i\leq m \tag{4.68}$$

with

$$p_{ij}^* = \frac{e^{-\theta_i^*-\theta_j^*}}{1+e^{-\theta_i^*-\theta_j^*}}, \qquad p_i^* = \frac{e^{-\theta_i^*}}{1+e^{-\theta_i^*}}. \tag{4.69}$$

The canonical ensemble therefore takes the form

$$P_{\text{can}}(G) = 2^{-\binom{n-m}{2}} \prod_{1\leq i<j\leq m} {p_{ij}^*}^{g_{ij}(G)} \left(1-p_{ij}^*\right)^{1-g_{ij}(G)} \prod_{i=1}^{m} {p_i^*}^{s_i(G)} (1-p_i^*)^{n-m-s_i(G)}. \tag{4.70}$$

The expression in (4.70) has an interpretation. Indeed, the canonical formula che be split into two parts:

$$P_{\text{can}}(G) = P_{\text{can}}^U(G) P_{\text{can}}^B(G)\ 2^{-\binom{n-m}{2}}, \tag{4.71}$$

with

$$P^{U}_{\text{can}}(G) = \prod_{1\leq i<j\leq m} p^{*\,g_{ij}(G)}_{ij} \left(1-p^{*}_{ij}\right)^{1-g_{ij}(G)} \tag{4.72}$$

and

$$P^{B}_{\text{can}}(G) = \prod_{i=1}^{m} p^{*\,s_i(G)}_{i} \left(1-p^{*}_{i}\right)^{n-m-s_i(G)} . \tag{4.73}$$

The unipartite probability, $P^{U}_{\text{can}}(G)$, is the canonical probability obtained when the constraint is put on the full degree sequence $\vec{u}^* = (u^*_i)_{i=1}^m$ on the set $\mathcal{G}_m$. The constrained degree sequence is precisely $u^*_i = \sum_{\substack{1\leq j\leq m \\ j\neq i}} p^*_{ij}$. The bipartite probability $P^{B}_{\text{can}}(G)$ is the canonical bipartite probability obtained when the constraint is put only on the top layer of a bipartite graph. In this case the configuration space is the set of bipartite graphs $\mathcal{G}_{m,n-m}$ with $m$ nodes on the top layer and $n-m$ nodes on the bottom layer. The constrained top layer degree sequence is $\vec{b}^* = (b^*_i)_{i=1}^m$, where $b^*_i = (n-m)p^*_i$. The third factor $2^{-\binom{n-m}{2}}$ is the inverse of the number of possible (unconstrained) graphs with $n-m$ nodes. In conclusion, the canonical probability in (4.70) can be interpreted as the product of two canonical probabilities, $P^{U}_{\text{can}}(G)$ and $P^{B}_{\text{can}}(G)$, and the number $2^{-\binom{n-m}{2}}$. Both canonical probabilities have an $m$-dimensional degree sequence as a constraint $\vec{u}^* = (u^*_i)_{i=1}^m$ and $\vec{b}^* = (b^*_i)_{i=1}^m$, put on the respective configuration spaces. Furthermore, two degree sequences sum up to the original degree sequence, namely,

$$u^*_i + b^*_i = k^*_i \qquad \forall i = 1,\ldots,m. \tag{4.74}$$

For this reason $(p^*_{ij})_{i,j=1}^m$ are called the unipartite probabilities and $(p^*_i)_{i=1}^m$ the bipartite probabilities.



## §B Appendix

In this appendix we identify the structure of the $\delta$-tame condition when a partial degree sequence $(k^*_i)_{i=1}^m$ is put as a constraint on $\mathcal{G}_n$. The definition comes from the situation where a full degree sequence $(k^*_i)_{i=1}^n$ is fixed on $\mathcal{G}_n$ [9]. In the full degree sequence situation the canonical probability takes the form

$$P_{\text{can}}(G) = \prod_{1\leq i<j\leq n} \left(p^{*}_{ij}\right)^{g_{ij}(G)} \left(1-p^{*}_{ij}\right)^{1-g_{ij}(G)} \tag{4.75}$$

with

$$p^*_{ij} = \frac{e^{-\theta^*_i-\theta^*_j}}{1+e^{-\theta^*_i-\theta^*_j}} \quad \forall\ i\neq j, \tag{4.76}$$

and with the vector of Lagrange multipliers $\vec{\theta}^* = (\theta^*_i)_{i=1}^n$ tuned such that

$$\langle k_i\rangle = \sum_{\substack{1\leq j\leq n \\ j\neq i}} p^*_{ij} = k^*_i, \qquad 1\leq i\leq n. \tag{4.77}$$

The degree sequence $(k_i^*)_{i=1}^n$ is said to be $\delta$-tame when there exists a $\delta \in (0, \frac{1}{2}]$ such that, for each $1 \leq i \neq j \leq n$, the canonical probabilites satisfy

$$\delta < p_{ij}^* < 1 - \delta. \tag{4.78}$$

**B.1 Definition ($\delta$-tame partial degree sequence).** We say that such a sequence is $\delta$-tame when there exists a $\delta \in (0, \frac{1}{2}]$ such that, for each $1 \leq i \neq j \leq m$, the canonical probabilites defined in (4.6)–(4.8) satisfy

$$\delta < p_{ij}^* < 1 - \delta \qquad \forall\ 1 \leq i \neq j \leq m. \tag{4.79}$$

**B.2 Lemma.** *If $(k_i^*)_{i=1}^m$ is a partial degree sequence on $\mathcal{G}_n$ and it is $\delta$-tame in the sense of Definition B.1, then the canonical bipartite probabilities satisfy*

$$\delta' < p_i^* < 1 - \delta' \qquad \forall\ 1 \leq i \neq j \leq m, \tag{4.80}$$

*for some $\delta' \in (0, \frac{1}{2}]$.*

*Proof.* The canonical probabilities, tuned with the proper $(\theta_i^*)_{i=1}^m$, satisfy

$$p_{ij}^* = \frac{x_i x_j}{1 + x_i x_j}, \qquad p_i^* = \frac{x_i}{1 + x_i}, \quad x_i = e^{-\theta_i^*}. \tag{4.81}$$

Since $(k_i)_{i=1}^m$ is a partial $\delta$-tame degree sequence, Definition B.1 says that

$$\delta < p_{ij}^* < (1 - \delta). \tag{4.82}$$

From this it follows that

$$\frac{\delta}{1-\delta} < x_i x_j < \frac{1-\delta}{\delta}. \tag{4.83}$$

Using (4.83) for different indices $i, j, k$, we get

$$\left(\frac{\delta}{1-\delta}\right)^2 < x_i^2 x_j x_k < \left(\frac{1-\delta}{\delta}\right)^2. \tag{4.84}$$

Using again (4.83) for the indices $j$ and $k$, we get

$$\left(\frac{\delta}{1-\delta}\right)^{3/2} < x_i < \left(\frac{1-\delta}{\delta}\right)^{3/2}. \tag{4.85}$$

Using (4.85) and $p_i^* = \frac{x_i}{1+x_i} = \frac{1}{1+\frac{1}{x_i}}$, we obtain that

$$\delta' < p_i^* < 1 - \delta' \tag{4.86}$$

with $\delta' = \frac{1}{1+(\frac{1-\delta}{\delta})^{3/2}}$. Note that $0 < \delta \leq \frac{1}{2}$ implies $0 < \delta' \leq \frac{1}{2}$.

CHAPTER 5

# Ensemble Equivalence for dense graphs

This chapter is based on:
F. den Hollander, M. Mandjes, A. Roccaverde, and N. J. Starreveld. Ensemble equivalence for dense graphs. *Electron. J. Probab.*, 23:Paper No. 12, 26, 2018

**Abstract**

In this paper we consider a random graph on which topological restrictions are imposed, such as constraints on the total number of edges, wedges, and triangles. We work in the dense regime, in which the number of edges per vertex scales proportionally to the number of vertices $n$. Our goal is to compare the micro-canonical ensemble (in which the constraints are satisfied for every realization of the graph) with the canonical ensemble (in which the constraints are satisfied on average), both subject to maximal entropy. We compute the relative entropy of the two ensembles in the limit as $n$ grows large, where two ensembles are said to be *equivalent* in the dense regime if this relative entropy divided by $n^2$ tends to zero. Our main result, whose proof relies on large deviation theory for graphons, is that breaking of ensemble equivalence occurs when the constraints are *frustrated*. Examples are provided for three different choices of constraints.

# §5.1 Introduction

Section 5.1.1 gives background and motivation, Section 5.1.2 describes relevant literature, while Section 5.1.3 outlines the remainder of the paper.

## §5.1.1 Background and motivation

For large networks a detailed description of the architecture of the network is infeasible and must be replaced by a *probabilistic* description, where the network is assumed to be a random sample drawn from a set of allowed graphs that are consistent with a set of empirically observed features of the network, referred to as *constraints*. Statistical physics deals with the definition of the appropriate probability distribution over the set of graphs and with the calculation of its relevant properties (Gibbs [53]). The two main choices[1] of probability distribution are:

(1) The *microcanonical ensemble*, where the constraints are *hard* (i.e., are satisfied by each individual graph).

(2) The *canonical ensemble*, where the constraints are *soft* (i.e., hold as ensemble averages, while individual graphs may violate the constraints).

For networks that are large but finite, the two ensembles are obviously different and, in fact, represent different empirical situations: they serve as *null-models* for the network after incorporating what is known about the network *a priori* via the constraints. Each ensemble represents the unique probability distribution with *maximal entropy* respecting the constraints. In the limit as the size of the graph diverges, the two ensembles are traditionally *assumed* to become equivalent as a result of the expected vanishing of the fluctuations of the soft constraints, i.e., the soft constraints are expected to become asymptotically hard. This assumption of *ensemble equivalence*, which is one of the corner stones of statistical physics, does however *not* hold in general (we refer to Touchette [97] for more background).

In Squartini *et al.* [92] the question of the possible breaking of ensemble equivalence was investigated for two types of constraint:

(I) The total number of edges.

(II) The degree sequence.

In the *sparse regime*, where the empirical degree distribution converges to a limit as the number of vertices $n$ tends to infinity such that the maximal degree is $o(\sqrt{n})$, it was shown that the relative entropy of the micro-canonical ensemble w.r.t. the canonical ensemble divided by $n$ (which can be interpreted as the relative entropy per vertex) tends to $s_\infty$, with $s_\infty = 0$ in case the constraint concerns the total number of edges, and $s_\infty > 0$ in case the constraint concerns the degree sequence. For the latter

[1] The microcanonical ensemble and the canonical ensemble work with a fixed number of vertices. There is a third ensemble, the *grandcanonical ensemble*, where also the size of the graph is considered as a soft constraint.

case, an explicit formula was derived for $s_\infty$, which allows for a *quantitative analysis* of the breaking of ensemble equivalence.

In the present paper we analyse what happens in the *dense regime*, where the number of edges per vertex is of order $n$. We consider case (I), yet allow for constraints not only on the total number of edges but also on the total number of wedges, triangles, etc. We show that the relative entropy divided by $n^2$ (which, up to a constant, can be interpreted as the relative entropy per edge) tends to $s_\infty$, with $s_\infty > 0$ when the constraints are *frustrated*. Our analysis is based on a large deviation principle for graphons.

## §5.1.2 Relevant literature

In the past few years, several papers have studied the microcanonical ensemble and the canonical ensemble. Most papers focus on dense graphs, but there are some interesting advances for sparse graphs as well. Closely related to the canonical ensemble are the *exponential random graph model* (Bhamidi *et al.* [12], Chatterjee and Diaconis [29]) and the *constrained exponential random model* (Aristoff and Zhu [3], Kenyon and Yin [67], Yin [102], Zhu [105]).

In Aristoff and Zhu [3], Kenyon *et al.* [63], Radin and Sadun [86], the authors study the microcanonical ensemble, focusing on the constrained entropy density. In [3] directed graphs are considered with a *hard* constraint on the number of directed edges and $j$-stars, while in [63, 86] the focus is on undirected graphs with a *hard* constraint on the edge density, $j$-star density and triangle density, respectively. Following the work in Bhamidi *et al.* [12] and in Chatterjee and Diaconis [29], a deeper understanding has developed of how these models behave as the size of the graph tends to infinity. Most results concern the asymptotic behaviour of the partition function (Chatterjee and Diaconis [29], Kenyon, Radin, Ren and Sadun [63]) and the identification of regions where phase transitions occur (Aristoff and Zhu [4], Lubetsky and Zhao [70], Yin [101]). For more details we refer the reader to the recent monograph by Chatterjee [27], and references therein. Significant contributions for sparse graphs were made in Chatterjee and Dembo [28] and in subsequent work of Yin and Zhu [103].

For an overview on random graphs and their role as models of complex networks, we refer the reader to the recent monograph by van der Hofstad [99]. The most important distinction between our paper and the existing literature on exponential random graphs is that in the canonical ensemble we impose a *soft* constraint.

## §5.1.3 Outline

The remainder of this paper is organised as follows. Section 5.2 defines the two ensembles, gives the definition of equivalence of ensembles in the dense regime, recalls some basic facts about graphons, and states the large deviation principle for the Erdős-Rényi random graph. Section 5.3 states a key theorem in which we give a *variational representation* of $s_\infty$ when the constraint is on *subgraph counts*, properly normalised. Section 5.4 presents our main theorem for ensemble equivalence, which provides three

examples for which *breaking of ensemble equivalence* occurs when the constraints are *frustrated*. In particular, the constraints considered are on the number of edges, triangles and/or stars. Frustration corresponds to the situation where the canonical ensemble scales like an Erdős-Rényi random graph model with an appropriate edge density but the microcanonical ensemble does not. The proof of the main theorem is given in Sections 5.5–5.6, and relies on various papers in the literature dealing with exponential random graph models. Appendix A discusses convergence of Lagrange multipliers associated with the canonical ensemble.

## §5.2 Key notions

In Section 5.2.1 we introduce the model and give our definition of equivalence of ensembles in the dense regime (Definition 5.2.1 below). In Section 5.2.2 we recall some basic facts about graphons (Propositions 5.2.4–5.2.6 below). In Section 5.2.3 we recall the large deviation principle for the Erdős-Rényi random graph (Proposition 5.2.7 and Theorem 5.2.8 below), which is the key tool in our paper.

### §5.2.1 Microcanonical ensemble, canonical ensemble, relative entropy

For $n \in \mathbb{N}$, let $\mathcal{G}_n$ denote the set of all $2^{\binom{n}{2}}$ simple undirected graphs with $n$ vertices. Any graph $G \in \mathcal{G}_n$ can be represented by a symmetric $n \times n$ matrix with elements

$$h^G(i,j) := \begin{cases} 1 & \text{if there is an edge between vertex } i \text{ and vertex } j, \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

Let $\vec{C}$ denote a vector-valued function on $\mathcal{G}_n$. We choose a specific vector $\vec{C}^*$, which we assume to be *graphic*, i.e., realisable by at least one graph in $\mathcal{G}_n$. For this $\vec{C}^*$ the *microcanonical ensemble* is the probability distribution $P_{\text{mic}}$ on $\mathcal{G}_n$ with *hard constraint* $\vec{C}^*$ defined as

$$P_{\text{mic}}(G) := \begin{cases} 1/\Omega_{\vec{C}^*}, & \text{if } \vec{C}(G) = \vec{C}^*, \\ 0, & \text{otherwise,} \end{cases} \qquad G \in \mathcal{G}_n, \tag{5.2}$$

where

$$\Omega_{\vec{C}^*} := |\{G \in \mathcal{G}_n \colon \vec{C}(G) = \vec{C}^*\}| \tag{5.3}$$

is the number of graphs that realise $\vec{C}^*$. The *canonical ensemble* $P_{\text{can}}$ is the unique probability distribution on $\mathcal{G}_n$ that maximises the *entropy*

$$S_n(\mathrm{P}) := -\sum_{G \in \mathcal{G}_n} \mathrm{P}(G) \log \mathrm{P}(G) \tag{5.4}$$

subject to the *soft constraint* $\langle \vec{C} \rangle = \vec{C}^*$, where

$$\langle \vec{C} \rangle := \sum_{G \in \mathcal{G}_n} \vec{C}(G)\, \mathrm{P}(G). \tag{5.5}$$

This gives the formula (see Jaynes [61])

$$P_{\text{can}}(G) := \frac{1}{Z(\vec{\theta}^*)}\, \mathrm{e}^{H(\vec{\theta}^*, \vec{C}(G))}, \qquad G \in \mathcal{G}_n, \tag{5.6}$$

with

$$H(\vec{\theta}^*, \vec{C}(G)) := \vec{\theta}^* \cdot \vec{C}(G), \qquad Z(\vec{\theta}^*) := \sum_{G \in \mathcal{G}_n} \mathrm{e}^{\vec{\theta}^* \cdot \vec{C}(G)}, \tag{5.7}$$

denoting the *Hamiltonian* and the *partition function*, respectively. In (5.6)–(5.7) the parameter $\vec{\theta}^*$ (which is a real-valued vector the size of the constraint playing the role of a Langrange multiplier) must be set to the unique value that realises $\langle \vec{C} \rangle = \vec{C}^*$. The Lagrange multiplier $\vec{\theta}^*$ exists and is unique. Indeed, the gradients of the constraints in (5.5) are linearly independent vectors. Consequently, the Hessian matrix of the entropy of the canonical ensemble in (5.6) is a positive definite matrix, which implies uniqueness of the Lagrange multiplier.

The *relative entropy* of $P_{\text{mic}}$ with respect to $P_{\text{can}}$ is defined as

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) := \sum_{G \in \mathcal{G}_n} P_{\text{mic}}(G) \log \frac{P_{\text{mic}}(G)}{P_{\text{can}}(G)}. \tag{5.8}$$

**5.2.1 Definition.** In the dense regime, if [2]

$$s_\infty := \lim_{n \to \infty} \frac{1}{n^2}\, S_n(P_{\text{mic}} | P_{\text{can}}) = 0, \tag{5.9}$$

then $P_{\text{mic}}$ and $P_{\text{can}}$ are said to be *equivalent*.

Before proceeding, we recall an important observation made in Squartini *et al.* [92]. For any $G_1, G_2 \in \mathcal{G}_n$, $P_{\text{can}}(G_1) = P_{\text{can}}(G_2)$ whenever $\vec{C}(G_1) = \vec{C}(G_2)$, i.e., the canonical probability is the same for all graphs with the same value of the constraint. We may therefore rewrite (5.8) as

$$S_n(P_{\text{mic}} \mid P_{\text{can}}) = \log \frac{P_{\text{mic}}(G^*)}{P_{\text{can}}(G^*)}, \tag{5.10}$$

where $G^*$ is *any* graph in $\mathcal{G}_n$ such that $\vec{C}(G^*) = \vec{C}^*$ (recall that we assumed that $\vec{C}^*$ is realisable by at least one graph in $\mathcal{G}_n$). This fact greatly simplifies computations.

**5.2.2 Remark.** All the quantities above depend on $n$. In order not to burden the notation, we exhibit this $n$-dependence only in the symbols $\mathcal{G}_n$ and $S_n(P_{\text{mic}} \mid P_{\text{can}})$. When we pass to the limit $n \to \infty$, we need to specify how $\vec{C}(G)$, $\vec{C}^*$ and $\vec{\theta}^*$ are chosen to depend on $n$. This will be done in Section 5.3.1.

[2] In Squartini *et al.* [92], which was concerned with the *sparse regime*, the relative entropy was divided by $n$ (the number of vertices). In the *dense regime*, however, it is appropriate to divide by $n^2$ (the order of the number of edges).

## §5.2.2 Graphons

There is a natural way to embed a simple graph on $n$ vertices in a space of functions called *graphons*. Let $W$ be the space of functions $h\colon [0,1]^2 \to [0,1]$ such that $h(x,y) = h(y,x)$ for all $(x,y) \in [0,1]^2$. A finite simple graph $G$ on $n$ vertices can be represented as a graphon $h^G \in W$ in a natural way as (see Fig. 5.1)

$$h^G(x,y) := \begin{cases} 1 & \text{if there is an edge between vertex } \lceil nx \rceil \text{ and vertex } \lceil ny \rceil, \\ 0 & \text{otherwise.} \end{cases} \tag{5.11}$$

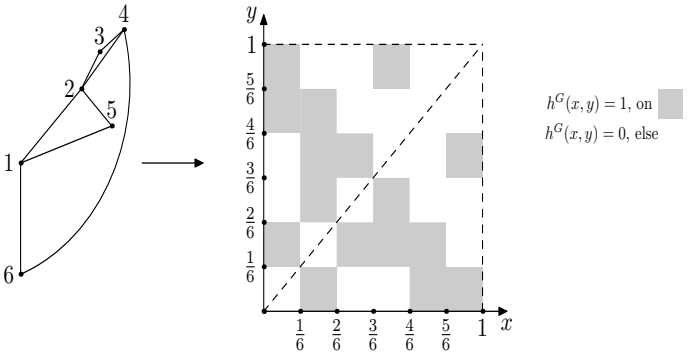



*Figure 5.1: An example of a graph $G$ and its graphon representation $h^G$.*

The space of graphons $W$ is endowed with the *cut distance*

$$d_\square(h_1,h_2) := \sup_{S,T\subset[0,1]} \left| \int_{S\times T} \mathrm{d}x\,\mathrm{d}y\,[h_1(x,y) - h_2(x,y)] \right|, \qquad h_1,h_2 \in W. \tag{5.12}$$

On $W$ there is a natural equivalence relation $\equiv$. Let $\Sigma$ be the space of measure-preserving bijections $\sigma\colon [0,1] \to [0,1]$. Then $h_1(x,y) \equiv h_2(x,y)$ if $h_1(x,y) = h_2(\sigma x, \sigma y)$ for some $\sigma \in \Sigma$. This equivalence relation yields the quotient space $(\tilde{W}, \delta_\square)$, where $\delta_\square$ is the metric defined by

$$\delta_\square(\tilde{h}_1,\tilde{h}_2) := \inf_{\sigma_1,\sigma_2} d_\square(h_1^{\sigma_1}, h_2^{\sigma_2}), \qquad \tilde{h}_1,\tilde{h}_2 \in \tilde{W}. \tag{5.13}$$

To avoid cumbersome notation, throughout the sequel we suppress the $n$-dependence. Thus, by $G$ we denote any simple graph on $n$ vertices, by $h^G$ its image in the graphon space $W$, and by $\tilde{h}^G$ its image in the quotient space $\tilde{W}$. Let $F$ and $G$ denote two simple graphs with vertex sets $V(F)$ and $V(G)$, respectively, and let $\hom(F,G)$ be the number of homomorphisms from $F$ to $G$. The *homomorphism density* is defined as

$$t(F,G) := \frac{1}{|V(G)|^{|V(F)|}} \hom(F,G). \tag{5.14}$$

Two graphs are said to be *similar* when they have similar homomorphism densities.

**5.2.3 Definition.** A sequence of labelled simple graphs $(G_n)_{n\in\mathbb{N}}$ is left-convergent when $(t(F,G_n))_{n\in\mathbb{N}}$ converges for any simple graph $F$.

Consider a simple graph $F$ on $k$ vertices with edge set $E(F)$, and let $h \in W$. Similarly as above, define the density

$$t(F,h) := \int_{[0,1]^k} \mathrm{d}x_1 \cdots \mathrm{d}x_k \prod_{(i,j)\in E(F)} h(x_i,x_j). \tag{5.15}$$

If $h^G$ is the image of a graph $G$ in the space $W$, then

$$t(F,h^G) = \int_{[0,1]^k} \mathrm{d}x_1 \cdots \mathrm{d}x_k \prod_{(i,j)\in E(F)} h^G(x_i,x_j) = \frac{1}{|V(G)|^{|V(F)|}} \hom(F,G) = t(F,G). \tag{5.16}$$

Hence a sequence of graphs $(G_n)_{n\in\mathbb{N}}$ is left-convergent to $h \in W$ when

$$\lim_{n\to\infty} t(F,G_n) = t(F,h). \tag{5.17}$$

We conclude this section with three basic facts that will be needed later on. The first gives the relation between left-convergence of sequences of graphs and convergence in the quotient space $(\tilde{W},\delta_\square)$, the second is a compactness property, while the third shows that the homomorphism density is Lipschitz continuous with respect to the $\delta_\square$-metric.

**5.2.4 Proposition (Borgs *et al.* [20]).** *For a sequence of labelled simple graphs $(G_n)_{n\in\mathbb{N}}$ the following properties are equivalent:*
*(i) $(G_n)_{n\in\mathbb{N}}$ is left-convergent.*
*(ii) $(\tilde{h}^{G_n})_{n\in\mathbb{N}}$ is a Cauchy sequence in the metric $\delta_\square$.*
*(iii) $(t(F,h^{G_n}))_{n\in\mathbb{N}}$ converges for all finite simple graphs $F$.*
*(iv) There exists an $h \in W$ such that $\lim_{n\to\infty} t(F,h^{G_n}) = t(F,h)$ for all finite simple graphs $F$.*

**5.2.5 Proposition (Lovász and Szegedy [69]).** *$(\tilde{W},\delta_\square)$ is compact.*

**5.2.6 Proposition (Borgs *et al.* [20]).** *Let $G_1, G_2$ be two labelled simple graphs, and let $F$ be a simple graph. Then*

$$|t(F,G_1) - t(F,G_2)| \le 4|E(F)|\delta_\square(G_1,G_2). \tag{5.18}$$

For a more detailed description of the structure of the space $(\tilde{W},\delta_\square)$ we refer the reader to Borgs *et al.* [20, 21] and Diao *et at.* [39].

## §5.2.3 Large deviation principle for the Erdős-Rényi random graph

In this section we recall a few key facts from the literature about rare events in Erdős-Rényi random graphs, formulated in terms of a large deviation principle. Importantly, the scale that is used is $n^2$, the order of the number of *edges* in the graph.

We start by introducing the large deviation rate function. For $p \in (0,1)$ and $u \in [0,1]$, let

$$\begin{aligned} I_p(u) &:= \frac{1}{2} u \log\left(\frac{u}{p}\right) + \frac{1}{2}(1-u)\log\left(\frac{1-u}{1-p}\right), \\ I(u) &:= \frac{1}{2} u \log u + \frac{1}{2}(1-u)\log(1-u) = I_{\frac{1}{2}}(u) - \tfrac{1}{2}\log 2, \end{aligned} \tag{5.19}$$

with the convention that $0 \log 0 = 0$. For $h \in W$ we write, with a mild abuse of notation,

$$I_p(h) := \int_{[0,1]^2} \mathrm{d}x\, \mathrm{d}y\; I_p(h(x,y)), \qquad I(h) := \int_{[0,1]^2} \mathrm{d}x\, \mathrm{d}y\; I(h(x,y)). \tag{5.20}$$

On the quotient space $(\tilde{W}, \delta_\square)$ we define $I_p(\tilde{h}) = I_p(h)$, where $h$ is any element of the equivalence class $\tilde{h}$.

**5.2.7 Proposition (Chatterjee and Varadhan [31]).** *The function $I_p$ is well-defined on $\tilde{W}$ and is lower semi-continuous under the $\delta_\square$-metric.*

Consider the set $\mathcal{G}_n$ of all graphs on $n$ vertices and the Erdős-Rényi probability distribution $\mathbb{P}_{n,p}$ on $\mathcal{G}_n$. Through the mappings $G \to h^G \to \tilde{h}^G$ we obtain a probability distribution on $W$ (with a slight abuse of notation again denoted by $\mathbb{P}_{n,p}$), and a probability distribution $\tilde{\mathbb{P}}_{n,p}$ on $\tilde{W}$.

**5.2.8 Theorem (Chatterjee and Varadhan [31]).** *For every $p \in (0,1)$, the sequence of probability distributions $(\tilde{\mathbb{P}}_{n,p})_{n\in\mathbb{N}}$ satisfies the large deviation principle on $(\tilde{W}, \delta_\square)$ with rate function $I_p$ defined by* (5.20)*, i.e.,*

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}(\tilde{C}) &\le -\inf_{\tilde{h}\in\tilde{W}} I_p(\tilde{h}) \qquad \forall\, \tilde{C} \subset \tilde{W} \text{ closed}, \\ \liminf_{n\to\infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}(\tilde{O}) &\ge -\inf_{\tilde{h}\in\tilde{O}} I_p(\tilde{h}) \qquad \forall\, \tilde{O} \subset \tilde{W} \text{ open}. \end{aligned} \tag{5.21}$$

Using the large deviation principle we can find asymptotic expressions for the number of simple graphs on $n$ vertices with a given property. In what follows a property of a graph is defined through an *operator* $T\colon W \to \mathbb{R}^m$ for some $m \in \mathbb{N}$. We assume that the operator $T$ is continuous with respect to the $\delta_\square$-metric, and for some $\vec{T}^* \in \mathbb{R}^m$ we consider the sets

$$\tilde{W}^* := \big\{\tilde{h} \in \tilde{W}\colon\, T(\tilde{h}) = \vec{T}^*\big\}, \quad \tilde{W}_n^* := \big\{\tilde{h} \in \tilde{W}^*\colon\, \tilde{h} = \tilde{h}^G \text{ for some } G \text{ on } n \text{ vertices}\big\}. \tag{5.22}$$

By the continuity of the operator $T$, the set $\tilde{W}^*$ is closed. Therefore, using Theorem 5.2.8, we obtain the following asymptotics for the cardinality of $\tilde{W}_n^*$.

**5.2.9 Corollary (Chatterjee [26]).** *For any measurable set $\tilde{W}^* \subset \tilde{W}$, with $\tilde{W}_n^*$ as defined in* (5.22)*,*

$$-\inf_{\tilde{h}\in\mathrm{int}(\tilde{W}^*)} I(\tilde{h}) \le \liminf_{n\to\infty} \frac{\log|\tilde{W}_n^*|}{n^2} \le \limsup_{n\to\infty} \frac{\log|\tilde{W}_n^*|}{n^2} \le -\inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}), \tag{5.23}$$

*where* $\mathrm{int}(\tilde{W}^*)$ *is the interior of* $\tilde{W}^*$.

## §5.3 Variational characterisation of ensemble equivalence

In this section we present a number of preparatory results we will need in Section 5.4 to state our theorem on the equivalence between $P_{\text{mic}}$ and $P_{\text{can}}$. Our main result is Theorem 5.3.4 below, which gives us a *variational characterisation of ensemble equivalence*. In Section 5.3.1 we introduce our constraints on the subgraph counts. In Section 5.3.2 we rephrase the canonical ensemble in terms of graphons. In Section 5.3.3 we state and prove Theorem 5.3.4.

### §5.3.1 Subgraph counts

First we introduce the concept of subgraph counts, and point out how the corresponding canonical distribution is defined. Label the simple graphs in any order, e.g., $F_1$ is an edge, $F_2$ is a wedge, $F_3$ is triangle, etc. Let $C_k(G)$ denote the number of subgraphs $F_k$ in $G$. In the dense regime, $C_k(G)$ grows like $n^{V_k}$, where $V_k = |V(F_k)|$ is the number of vertices in $F_k$. For $m \in \mathbb{N}$, consider the following *scaled vector-valued function* on $\mathcal{G}_n$:

$$\vec{C}(G) := \left( \frac{p(F_k)C_k(G)}{n^{V_k-2}} \right)_{k=1}^{m} = n^2 \left( \frac{p(F_k)C_k(G)}{n^{V_k}} \right)_{k=1}^{m}. \tag{5.24}$$

The term $p(F_k)$ counts the edge-preserving permutations of the vertices of $F_k$, i.e., $p(F_1) = 2$ for an edge, $p(F_2) = 2$ for a wedge, $p(F_3) = 6$ for a triangle, etc. The term $C_k(G)/n^{V_k}$ represents a subgraph density in the graph $G$. The additional $n^2$ guarantees that the full vector scales like $n^2$, the scaling of the large deviation principle in Theorem 5.2.8. For a simple graph $F_k$ we define the *homomorphism density* as

$$t(F_k, G) := \frac{\hom(F_k, G)}{n^{V_k}} = \frac{p(F_k)C_k(G)}{n^{V_k}}, \tag{5.25}$$

which does not distinguish between permutations of the vertices. Hence the Hamiltonian becomes

$$H(\vec{\theta}, \vec{T}(G)) = n^2 \sum_{k=1}^{m} \theta_k \, t(F_k, G) = n^2(\vec{\theta} \cdot \vec{T}(G)), \qquad G \in \mathcal{G}_n, \tag{5.26}$$

where

$$\vec{T}(G) := (t(F_k, G))_{k=1}^{m}. \tag{5.27}$$

The canonical ensemble with parameter $\vec{\theta}$ thus takes the form

$$P_{\text{can}}(G \mid \vec{\theta}) := \mathrm{e}^{n^2\left[\vec{\theta}\cdot\vec{T}(G) - \psi_n(\vec{\theta})\right]}, \qquad G \in \mathcal{G}_n, \tag{5.28}$$

where $\psi_n$ replaces the *partition function*:

$$\psi_n(\vec{\theta}) := \frac{1}{n^2} \log \sum_{G \in \mathcal{G}_n} \mathrm{e}^{n^2(\vec{\theta} \cdot \vec{T}(G))}. \tag{5.29}$$

In the sequel we take $\vec{\theta}$ equal to a specific value $\vec{\theta}^*$, so as to meet the soft constraint, i.e.,

$$\langle \vec{T} \rangle = \sum_{G \in \mathcal{G}_n} \vec{T}(G)\, P_{\mathrm{can}}(G) = \vec{T}^*. \tag{5.30}$$

The canonical probability then becomes

$$P_{\mathrm{can}}(G) = P_{\mathrm{can}}(G \mid \vec{\theta}^*) \tag{5.31}$$

In Section 5.5.1 we will discuss how to find $\vec{\theta}^*$.

**5.3.1 Remark.** (i) The constraint $\vec{T}^*$ and the Lagrange multiplier $\vec{\theta}^*$ in general depend on $n$, i.e., $\vec{T}^* = \vec{T}^*_n$ and $\vec{\theta}^* = \vec{\theta}^*_n$ (recall Remark 5.2.2). We consider constraints that converge when we pass to the limit $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \vec{T}^*_n = \vec{T}^*_\infty. \tag{5.32}$$

Consequently, we expect that

$$\lim_{n\to\infty} \vec{\theta}^*_n = \vec{\theta}^*_\infty. \tag{5.33}$$

Throughout the sequel we *assume* that (5.33) holds. If convergence fails, then we may still consider subsequential convergence. The subtleties concerning (5.33) are discussed in Appendix A.
(ii) In what follows, we suppress the dependence on $n$ and write $\vec{T}^*, \vec{\theta}^*$ instead of $\vec{T}^*_n, \vec{\theta}^*_n$, but we keep the notation $\vec{T}^*_\infty, \vec{\theta}^*_\infty$ for the limit. In addition, throughout the sequel we write $\vec{\theta}, \vec{\theta}_\infty$ instead of $\vec{\theta}^*, \vec{\theta}^*_\infty$ when we view these as parameters that do not depend on $n$. This distinction is crucial when we take the limit $n \to \infty$.

## §5.3.2 From graphs to graphons

In (5.16) we saw that if we map a finite simple graph $G$ to its graphon $h^G$, then for each finite simple graph $F$ the homomorphism densities $t(F, G)$ and $t(F, h^G)$ are identical. If $(G_n)_{n\in\mathbb{N}}$ is left-convergent, then

$$\lim_{n\to\infty} \vec{T}(G_n) = (t(F_k, h))_{k=1}^m \tag{5.34}$$

for some $h \in W$, as an immediate consequence of Theorem 5.2.4. We further see that the expression in (5.26) can be written in terms of graphons as

$$H(\vec{\theta}, \vec{T}(G)) = n^2 \sum_{k=1}^m \theta_k\, t(F_k, h^G). \tag{5.35}$$

With this scaling the *hard constraint* is denoted by $\vec{T}^*$, has the interpretation of the *density* of an observable quantity in $G$, and defines a subspace of the quotient space $\tilde{W}$, which we denote by $\tilde{W}^*$, and which consists of all graphons that meet the hard constraint, i.e.,

$$\tilde{W}^* := \{\tilde{h} \in \tilde{W} \colon\ \vec{T}(h) = \vec{T}^*\}. \tag{5.36}$$

The *soft constraint* in the canonical ensemble becomes $\langle \vec{T} \rangle = \vec{T}^*$ (recall (5.5)).

## §5.3.3 Variational formula for specific relative entropy

In what follows, the limit as $n \to \infty$ of the partition function $\psi_n(\vec{\theta})$ defined in (5.29) plays an important role. This limit has a variational representation that will be key to our analysis.

**5.3.2 Theorem (Chatterjee and Diaconis [29]).** *Let $\vec{T}\colon \tilde{W} \to \mathbb{R}^m$ be the operator defined in* (5.27)*. For any $\vec{\theta} \in \mathbb{R}^m$ (not depending on $n$),*

$$\lim_{n\to\infty} \psi_n(\vec{\theta}) = \sup_{\tilde{h}\in\tilde{W}} \Big(\vec{\theta}\cdot\vec{T}(\tilde{h}) - I(\tilde{h})\Big) \tag{5.37}$$

*with $I$ and $\psi_n$ as defined in* (5.20) *and* (5.29)*.*

**5.3.3 Theorem (Chatterjee and Diaconis [29]).** *Let $F_1, \dots, F_m$ be subgraphs as defined in Section* 5.3.1*. Suppose that $\theta_2, \dots, \theta_m \geq 0$. Then*

$$\lim_{n\to\infty} \psi_n(\vec{\theta}) = \sup_{0\leq u\leq 1} \left(\sum_{i=1}^{m} \theta_i\, u^{E(F_k)} - I(u)\right), \tag{5.38}$$

*where $E(F_k)$ denotes the number of edges in the subgraph $F_k$.*

The key result in this section is the following variational formula for $s_\infty$ defined in Definition 5.2.1. Recall that for $n \in \mathbb{N}$ we write $\vec{\theta}^*$ for $\vec{\theta}^*_n$.

**5.3.4 Theorem.** *Consider the microcanonical ensemble defined in* (5.2) *with constraint $\vec{T} = \vec{T}^*$ defined in* (5.27)*, and the canonical ensemble defined in* (5.28)–(5.29) *with parameter $\vec{\theta} = \vec{\theta}^*$ such that, for every $n \in \mathbb{N}$,* (5.30)*,* (5.32) *and* (5.33) *hold. Then*

$$s_\infty = \lim_{n\to\infty} \frac{1}{n^2}\, S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = \sup_{\tilde{h}\in\tilde{W}} \big[\vec{\theta}^*_\infty\cdot\vec{T}(\tilde{h}) - I(\tilde{h})\big] - \sup_{\tilde{h}\in\tilde{W}^*} \big[\vec{\theta}^*_\infty\cdot\vec{T}(\tilde{h}) - I(\tilde{h})\big], \tag{5.39}$$

*where $I$ is defined in* (5.19) *and $\tilde{W}^* = \{\tilde{h} \in \tilde{W}\colon\ \vec{T}(\tilde{h}) = \vec{T}^*_\infty\}$.*

*Proof.* From (5.10) we have

$$s_\infty = \lim_{n\to\infty} \frac{1}{n^2} \big[\log\, P_{\mathrm{mic}}(G^*) - \log\, P_{\mathrm{can}}(G^*)\big], \tag{5.40}$$

where $G^*$ is any graph in $\mathcal{G}_n$ such that $\vec{T}(G^*) = \vec{T}^*$. For the microcanonical ensemble we have

$$\log P_{\mathrm{mic}}(G^*) = -\log \Omega_{\vec{T}^*} = -\log \mathbb{P}_{\frac{1}{2},n}\left(\{G \in \mathcal{G}_n\colon\ \vec{T}(G) = \vec{T}^*\}\right) - \binom{n}{2}\log 2, \tag{5.41}$$

where

$$\Omega_{\vec{T}^*} = |\{G \in \mathcal{G}_n\colon\ \vec{T}(G) = \vec{T}^*\}| > 0. \tag{5.42}$$

Define the operator $\vec{T}\colon W \to \mathbb{R}^m$, $h \mapsto (t(F_k,h))_{k=1}^m$. This operator can be extended to an operator (with a slight abuse of notation again denoted by $\vec{T}$) on the quotient space $(\tilde{W}, \delta_\square)$ by defining $\vec{T}(\tilde{h}) = \vec{T}(h)$ with $h \in \tilde{h}$. Define the following sets

$$\tilde{W}^* := \big\{\tilde{h} \in \tilde{W}\colon\ T(\tilde{h}) = \vec{T}^*_\infty\big\}, \qquad \tilde{W}^*_n := \big\{\tilde{h} \in \tilde{W}^*\colon\ \tilde{h} = \tilde{h}^G \text{ for some } G \in \mathcal{G}_n\big\}. \tag{5.43}$$

From the continuity of the operator $\vec{T}$ on $\tilde{W}$, we see that $\tilde{W}^*$ is a compact subspace of $\tilde{W}$, and hence is also closed. From Theorem 5.2.6 we have that $\vec{T}$ is a Lipschitz continuous operator on the space $(\tilde{W}, \delta_\square)$. Since $\tilde{W}$ is a compact space, we have that

$$\lim_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}_{\frac{1}{2},n}\Big(\{G \in \mathcal{G}_n\colon\ \vec{T}(G) = \vec{T}^*\}\Big) = -\inf_{\tilde{h}\in\tilde{W}^*} I_{\frac{1}{2}}(\tilde{h}) = -\inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}) - \tfrac{1}{2}\log 2. \tag{5.44}$$

The large deviation principle applied to (5.41) yields

$$\lim_{n\to\infty} \frac{1}{n^2} \log P_{\mathrm{mic}}(G^*) = \inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}). \tag{5.45}$$

Consider the canonical ensemble and a graph $G_n^*$ on $n$ vertices such that $\vec{T}(G_n^*) = \vec{T}^*$. By Definition 5.2.3, Proposition 5.2.4, and (5.32) we may suppose that $(G_n^*)_{n\in\mathbb{N}}$ is left-convergent and converges to the graphon $h^*$. Since $\vec{T}$ is continuous, we have that $\vec{T}(G_n^*)$ converges to $\vec{T}(h^*) = \vec{T}^*_\infty$. From (5.28) we have that

$$\lim_{n\to\infty} \frac{1}{n^2} \log P_{\mathrm{can}}(G_n^*) = \vec{\theta}^*_\infty \cdot \vec{T}^*_\infty - \psi_\infty(\vec{\theta}^*_\infty). \tag{5.46}$$

By Theorem 5.3.2,

$$\psi_\infty(\vec{\theta}^*_\infty) = \sup_{\tilde{h}\in\tilde{W}} \big[\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\big]. \tag{5.47}$$

There is an additional subtlety in proving (5.47) in our setup because $\vec{\theta}^*$ depends on $n$. This dependence is treated in Appendix A. Combining (5.45) and (5.47), we get

$$s_\infty = \lim_{n\to\infty} \frac{1}{n^2} S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = \inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}) - \vec{\theta}^*_\infty \cdot \vec{T}^*_\infty + \sup_{\tilde{h}\in\tilde{W}} \big[\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\big]. \tag{5.48}$$

By definition all elements $\tilde{h} \in \tilde{W}^*$ satisfy $\vec{T}(\tilde{h}) = \vec{T}^*_\infty$. Hence the expression in the right-hand side of (5.48) can be written as

$$\sup_{\tilde{h}\in\tilde{W}} \big[\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\big] - \sup_{\tilde{h}\in\tilde{W}^*} \big[\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\big], \tag{5.49}$$

which settles the claim.

**5.3.5 Remark.** Theorem 5.3.4 and the compactness of $\tilde{W}^*$ give us a *variational characterisation* of ensemble equivalence: $s_\infty = 0$ if and only if at least one of the maximisers of $\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})$ in $\tilde{W}$ also lies in $\tilde{W}^* \subset \tilde{W}$. Equivalently, $s_\infty = 0$ when at least one the maximisers of $\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})$ satisfies the hard constraint.

## §5.4 Main theorem

The variational formula for the relative entropy $s_\infty$ in Theorem 5.3.4 allows us to identify examples where ensemble equivalence holds ($s_\infty = 0$) or is broken ($s_\infty > 0$). We already know that if the constraint is on the edge density alone, i.e., $T(G) = t(F_1, G) = T^*$, then $s_\infty = 0$ (see Garlaschelli *et al.* [48]). In what follows we will look at three models:

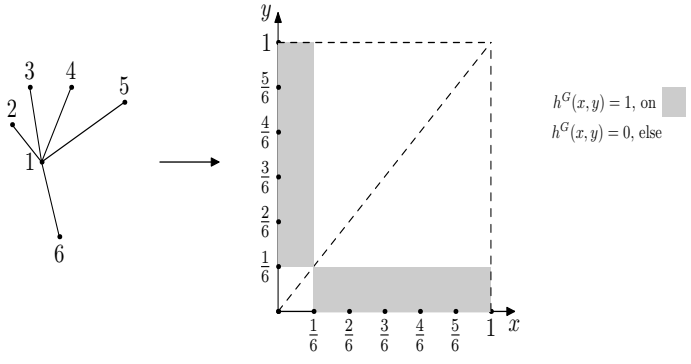



*Figure 5.2: A 5-star graph and its graphon representation.*

(I) The constraint is on the triangle density, i.e., $\vec{T}_2(G) = t(F_3, G) = T_2^*$ with $F_3$ the triangle. This will be referred to as the *Triangle Model*.

(II) The constraint is on the edge density and triangle density, i.e., $\vec{T}(G) = (t(F_1, G), t(F_3, G)) = (T_1^*, T_2^*)$ with $F_1$ the edge and $F_3$ the triangle. This will be referred to as the *Edge-Triangle Model*.

(III) The constraint is on the $j$-star density, i.e., $\vec{T}(G) = t(T[j], G) = T[j]^*$ with $T[j]$ the $j$-star graph, consisting of 1 root vertex and $j \in \mathbb{N} \setminus \{1\}$ vertices connected to the root but not connected to each other (see Fig. 5.2). This will be referred to as the *Star Model*.

For a graphon $h \in W$ (recall (5.15)), the edge density and the triangle density equal

$$\begin{aligned} T_1(h) &= \int_{[0,1]^2} \mathrm{d}x_1 \mathrm{d}x_2 \, h(x_1, x_2), \\ T_2(h) &= \int_{[0,1]^3} \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3 \, h(x_1, x_2) h(x_2, x_3) h(x_3, x_1), \end{aligned} \tag{5.50}$$

while the $j$-star density equals

$$T[j](h) = \int_{[0,1]} \mathrm{d}x \int_{[0,1]^j} \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_j \prod_{i=1}^{j} h(x, x_i). \tag{5.51}$$

**5.4.1 Theorem.** *For the above three types of constraint:*

(I) (a) *If* $T_2^* \geq \frac{1}{8}$ *, then* $s_\infty = 0$.

(b) *If* $T_2^* = 0$*, then* $s_\infty = 0$.

(II) (a) *If* $T_2^* = T_1^{*3}$*, then* $s_\infty = 0$.

(b) *If* $T_2^* \neq T_1^{*3}$ *and* $T_2^* \geq \frac{1}{8}$*, then* $s_\infty > 0$.

(c) *If* $T_2^* \neq T_1^{*3}$*,* $0 < T_1^* \leq \frac{1}{2}$ *and* $0 < T_2^* < \frac{1}{8}$*, then* $s_\infty > 0$.

(d) *If* $T_1^* = \frac{1}{2} + \epsilon$ *with* $\epsilon \in \left(\frac{\ell-2}{2\ell}, \frac{\ell-1}{2\ell+2}\right)$*,* $\ell \in \mathbb{N} \setminus \{1\}$*, and* $T_2^*$ *is such that* $(T_1^*, T_2^*)$ *lies on the scallopy curve in Fig.* 5.3*, then* $s_\infty > 0$.

(e) *If* $0 < T_1^* \leq \frac{1}{2}$ *and* $T_2^* = 0$*, then* $s_\infty = 0$.

(III) *For every* $j \in \mathbb{N} \setminus \{1\}$*, if* $T[j]^* \geq 0$*, then* $s_\infty = 0$.

Here, $T_1^*, T_2^*, T[j]^*$ are in fact the limits $T_{1,\infty}^*, T_{2,\infty}^*, T[j]_\infty^*$ in (5.32), but in order to keep the notation light we now also suppress the index $\infty$.

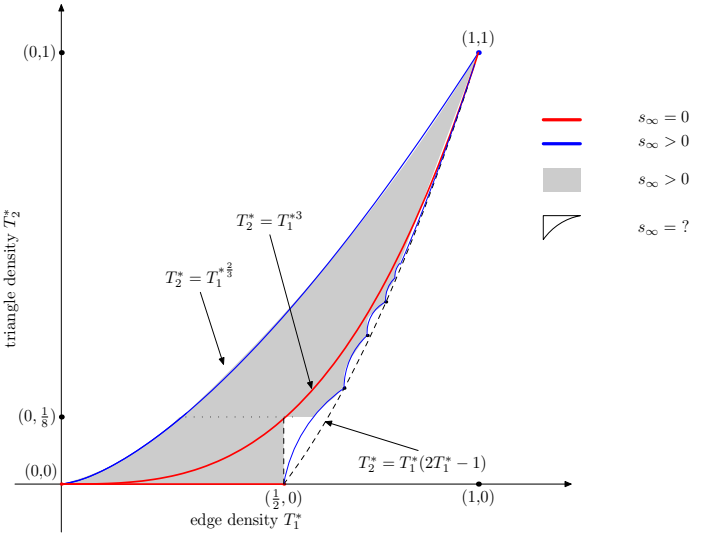



*Figure 5.3: The admissible edge-triangle density region is the region on and between the blue curves (cf. Radin and Sadun [86]).*

Theorem 5.4.1, which states our main results on ensemble equivalence and which is proven in Sections 5.5–5.6, is illustrated in Fig. 5.3. The region on and between the blue curves corresponds to the set of all realisable graphs: if the pair $(e, t)$ lies in this region, then there exists a graph with edge density $e$ and triangle density $t$. The red curves represent ensemble equivalence, the blue curves and the grey region represent breaking of ensemble equivalence, while in the white region between the red curve and the lower blue curve we do not know what happens. Breaking of ensemble equivalence arises from frustration between the edge and the triangle density.

Each of the cases in Theorem 5.4.1 corresponds to typical behavior of graphs drawn from the two ensembles:

- In cases (I)(a) and (II)(a), graphs drawn from both ensembles are asymptotically like Erdős-Rényi random graphs with parameter $p = T_2^{*1/3}$.
- In cases (I)(b) and (II)(e), almost all graphs drawn from both ensembles are asymptotically like bipartite graphs.
- In cases (II)(b), (II)(c) and (II)(d), we do not know what graphs drawn from the canonical ensemble look like. Graphs drawn from the microcanonical ensemble do not look like Erdős-Rényi random graphs. The structure of graphs drawn from the microcanonical ensemble when the constraint is as in (II)(d) has been determined in Pirkhurko and Razborov [83] and Radin and Sadun [86]. The vertex set of a graph drawn from the microcanonical ensemble can be partitioned into $\ell$ subsets: the first $\ell - 1$ have size $\lfloor cn \rfloor$ and the last has size between $\lfloor cn \rfloor$ and $2\lfloor cn \rfloor$, where $c$ is a known constant depending on $\ell$. The graph has the form of a complete $\ell-$partite graph on these pieces, plus some additional edges in the last piece that create no additional triangles.
- In case (III), graphs drawn from both ensembles are asymptotically like Erdős-Rényi random graphs with parameter $p = T[j]^{*1/j}$.

**5.4.2 Remark.** Similar results hold for the Edge-Wedge-Triangle Model and the Edge-Star Model.

Here are three open questions:

- Identify in which cases (5.32) implies (5.33).
- Is $s_\infty = 0$ as soon as the constraint involves a *single* subgraph count only?
- What happens for subgraphs other than edges, wedges, triangles and stars? Is again $s_\infty > 0$ under appropriate frustration?

## §5.5 Choice of the tuning parameter

The tuning parameter is to be chosen so as to satisfy the soft constraint (5.30), a procedure that in equilibrium statistical physics is referred to as the *averaging principle*. Depending on the choice of constraint, finding $\vec{\theta}^*$ may not be easy, neither analytically nor numerically. In Section 5.5.1 we investigate how $\vec{\theta}^*$ behaves as we vary $\vec{T}^*$ for fixed $n$. We focus on the Edge-Triangle Model (a slight adjustment yields the same results for the Triangle Model). In Section 5.5.2 we investigate how averages under the canonical ensemble, like (5.30), behave when $n \to \infty$. Here we can treat general constraints defined in (5.27).

For the behaviour of our constrained models, the *sign* of the coordinates of the tuning parameter $\vec{\theta}^*$ is of pivotal importance, both for a fixed $n \in \mathbb{N}$ and asymptotically (see Bhamidi *et al.* [12], Chatterjee and Diaconis [29], Radin and Yin [87], and references therein). We must therefore carefully keep track of this sign. The key results in this direction are Lemmas 5.5.1 and 5.5.2 below.

## §5.5.1 Tuning parameter for fixed $n$

**5.5.1 Lemma.** *Consider the Triangle Model with the constraint given by the triangle density $T_2^*$. For every $n$, $\theta^* \geq 0$ if and only if $T_2^* \geq \frac{1}{8}$.*

*Proof.* The proof is similar to that of Lemma 5.5.2 below.

**5.5.2 Lemma.** *Consider the Edge-Triangle Model. For every $n$, $\theta_2^* \geq 0$ if and only if $T_2^* \geq \frac{1}{8}$, irrespective of $T_1^*$. Furthermore, $\theta_1^* \geq 0$ if and only if $T_1^* \geq \frac{1}{2}$.*

*Proof.* Define, for $\theta_1, \theta_2 \in \mathbb{R}$, the function

$$g(\theta_1, \theta_2) := \sum_{G \in \mathcal{G}_n} \exp\left[n^2 \left(\theta_1 (T_1(G) - \tfrac{1}{2}) + \theta_2 (T_2(G) - \tfrac{1}{8})\right)\right]. \tag{5.52}$$

We first prove that $g$ attains a unique global minimum at $(\theta_1, \theta_2) = (0, 0)$. Consider the canonical ensemble $P_{\text{can}}$ as defined in (5.28) and (5.31), with $\vec{T}$ as defined above, and the probability distribution $\mathrm{P}_{\text{hom}}$ on $\mathcal{G}_n$ that assigns probability $2^{-\binom{n}{2}}$ to every graph $G \in \mathcal{G}_n$. Since $\mathrm{P}_{\text{hom}}$ is absolutely continuous with respect to $P_{\text{can}}$, the relative entropy $S_n(\mathrm{P}_{\text{hom}} | P_{\text{can}})$ is well defined:

$$S_n(\mathrm{P}_{\text{hom}} \mid P_{\text{can}}) = \sum_{G \in \mathcal{G}_n} \mathrm{P}_{\text{hom}}(G) \log \frac{\mathrm{P}_{\text{hom}}(G)}{P_{\text{can}}(G)} \geq 0. \tag{5.53}$$

Using the form of the canonical ensemble we get, after some straightforward calculations, that, for all $\theta_1, \theta_2 \in \mathbb{R}$,

$$\sum_{G \in \mathcal{G}_n} \exp\left[n^2 \big(\theta_1 T_1(G) + \theta_2 T_2(G)\big)\right] \geq 2^{\binom{n}{2}} \exp\left[n^2 \big(\theta_1 \tfrac{1}{2} + \theta_2 \tfrac{1}{8}\big)\right], \tag{5.54}$$

where the term in the right-hand side comes from the relation

$$\sum_{G \in \mathcal{G}_n} \frac{1}{2^{\binom{n}{2}}} \left(\theta_1 T_1(G) + \theta_2 T_2(G)\right) = \theta_1 \tfrac{1}{2} + \theta_2 \tfrac{1}{8}. \tag{5.55}$$

Observe that the left-hand side represents the average edge and triangle density, multiplied with $\theta_1, \theta_2$, in an Erdős-Rényi random graph with parameters $(n, \frac{1}{2})$. From (5.54) we find that $g(\theta_1, \theta_2) \geq 2^{\binom{n}{2}} = g(0, 0)$ for all $\theta_1, \theta_2 \in \mathbb{R}$, and so $g$ attains a global minimum at $(0, 0)$. In what follows we show that this global minimum is unique. A straightforward computation shows that $\partial_{\theta_1} g(\theta_1, \theta_2) = \partial_{\theta_2} g(\theta_1, \theta_2) = 0$ if and only if $\langle T_1 \rangle = \frac{1}{2}$ and $\langle T_2 \rangle = \frac{1}{8}$. Furthermore, the Hessian matrix is a covariance matrix and hence is positive semi-definite. For $\vec{\theta} = (\theta_1, \theta_2) = (0, 0)$ we know that $\langle T_1 \rangle = \frac{1}{2}$ and $\langle T_2 \rangle = \frac{1}{8}$. Hence, by uniqueness of the multiplier $\vec{\theta}^*$ for the constraint $T_1^* = \frac{1}{2}$, $T_2^* = \frac{1}{8}$, we obtain that $g$ has a unique global minimum at $(0, 0)$. Moreover, this shows that $g$ has no other stationary points. Consider the parameter $(\theta_1, \theta_2) = (\theta_1^*, \theta_2^*)$. We have

$$\begin{aligned} \partial_{\theta_2} g(\theta_1^*, \theta_2^*) &= \big(\langle T_2 \rangle - \tfrac{1}{8}\big) \exp[-n^2(\theta_1^* \tfrac{1}{2} + \theta_2^* \tfrac{1}{8})] \sum_{G \in \mathcal{G}_n} \exp\left[n^2 \left(\theta_1^* T_1(G) + \theta_2^* T_2(G)\right)\right] \\ &= \big(T_2^* - \tfrac{1}{8}\big) \exp[-n^2(\theta_1^* \tfrac{1}{2} + \theta_2^* \tfrac{1}{8})] \sum_{G \in \mathcal{G}_n} \exp\left[n^2 \left(\theta_1^* T_1(G) + \theta_2^* T_2(G)\right)\right]. \end{aligned} \tag{5.56}$$

If $T_2^* \geq \frac{1}{8}$, then $\partial_{\theta_2} g(\theta_1^*, \theta_2^*) \geq 0$. Because $g$ has a unique stationary point at $(0,0)$, which is a global minimum, we get $\theta_2^* \geq 0$. Similarly, we can show that if $T_2^* < \frac{1}{8}$, then $\theta_2^* < 0$. Suppose that $T_1^* \geq \frac{1}{2}$. For the parameter $(\theta_1, \theta_2) = (\theta_1^*, \theta_2^*)$ we have

$$\begin{aligned}\partial_{\theta_1} g(\theta_1^*, \theta_2^*) &= \left(\langle T_1 \rangle - \tfrac{1}{2}\right) \exp[-n^2(\theta_1^* \tfrac{1}{2} + \theta_2^* \tfrac{1}{8})] \sum_{G \in \mathcal{G}_n} \exp\left[n^2 \left(\theta_1^* T_1(G) + \theta_2^* T_2(G)\right)\right] \\ &= \left(T_1^* - \tfrac{1}{2}\right) \exp[-n^2(\theta_1^* \tfrac{1}{2} + \theta_2^* \tfrac{1}{8})] \sum_{G \in \mathcal{G}_n} \exp\left[n^2 \left(\theta_1^* T_1(G) + \theta_2^* T_2(G)\right)\right].\end{aligned} \tag{5.57}$$

Arguing in a similar way as before, we conclude that $\theta_1^* \geq 0$ if and only if $T_1^* \geq \frac{1}{2}$.

Consider the Edge-Triangle Model and suppose that the constraint $(T_1^*, T_2^*)$ is such that $T_2^* = T_1^{*3}$. Then $\theta_2^* = 0$ and $\theta_1^*$ matches the constraint on the edge density only. The following lemma shows that in this case the canonical ensemble behaves like the Erdős-Rényi model with parameter $T_1^*$, a fact that will be needed later to prove equivalence.

**5.5.3 Lemma.** *Consider the Edge-Triangle Model with the constraint given by the edge-triangle densities $\vec{T}^* = (T_1^*, T_2^*)$ with $T_2^* = T_1^{*3}$. Consider the canonical ensemble as defined in* (5.31)*. Then, for every $n \in \mathbb{N}$,*

$$\theta_1^* = \frac{1}{2} \log \frac{T_1^*}{1 - T_1^*}, \qquad \theta_2^* = 0. \tag{5.58}$$

*Proof.* From the definition of the canonical ensemble we have that, for $G \in \mathcal{G}_n$,

$$P_{\text{can}}(G) = P_{\text{can}}(G \mid \vec{\theta}^*) = \mathrm{e}^{n^2\left[\theta_1^* T_1(G) + \theta_2^* T_2(G) - \psi_n(\vec{\theta}^*)\right]}, \tag{5.59}$$

where $\psi_n(\vec{\theta}^*)$ is the partition function defined in (5.29). For the specific value $\vec{\theta} = \vec{\theta}^*$ we have that (recall (5.30))

$$\langle T_1 \rangle = T_1^*, \qquad \langle T_2 \rangle = T_2^* = T_1^{*3}. \tag{5.60}$$

We claim that the correct parameter is $\vec{\theta}^* = (\frac{1}{2} \log \frac{T_1^*}{1-T_1^*}, 0)$. The average fraction of edges is $T_1^*$ (see Park and Newman [81]). The average number of triangles is

$$\begin{aligned}\langle T_2 \rangle &= \frac{\sum_{G \in \mathcal{G}_n} T_2(G) \exp\left[n^2 \left(\frac{1}{2} \log \frac{T_1^*}{1-T_1^*} T_1(G)\right)\right]}{\sum_{G \in \mathcal{G}_n} \exp\left[n^2 \left(\frac{1}{2} \log \frac{T_1^*}{1-T_1^*} T_1(G)\right)\right]} \\ &= \frac{\sum_{G \in \mathcal{G}_n} T_2(G) (T_1^*)^{E(G)} (1 - T_1^*)^{\binom{n}{2} - E(G)}}{\sum_{G \in \mathcal{G}_n} (T_1^*)^{E(G)} (1 - T_1^*)^{\binom{n}{2} - E(G)}} = T_1^{*3},\end{aligned}$$

where the last equation comes from the fact we are calculating the average number of triangles in an Erdős-Rényi model with probability $T_1^*$. Since the multiplier $\vec{\theta}^*$ is unique, the proof is complete.

## §5.5.2 Tuning parameter for $n \to \infty$

In Lemma 5.5.4 below we show how averages under the canonical ensemble behave asymptotically when $\vec{\theta}$ does not depend on $n$. In Lemma A.2 we will look at what happens when $\vec{\theta}$ is a one-dimensional multiplier and depends on $n$.

**5.5.4 Lemma.** *Suppose that the operator $\vec{T}\colon W \to \mathbb{R}^m$ is bounded and continuous with respect to the $\delta_\square$-norm as defined in* (5.13)*. For $\vec{\theta} \in \mathbb{R}^m$ independent of $n$, consider the variational problem*

$$\sup_{\tilde{h}\in\tilde{W}} \big[\vec{\theta}\cdot\vec{T}(\tilde{h}) - I(\tilde{h})\big], \tag{5.61}$$

*where $I$ is defined in* (5.19)*. Suppose that the supremum is attained at a unique point, denoted by $\tilde{h}^*(\vec{\theta})$. Then*

$$\lim_{n\to\infty} \sum_{G\in\mathcal{G}_n} T_k(G)\, P_{\mathrm{can}}(G \mid \vec{\theta}\,) = T_k\big(\tilde{h}^*(\vec{\theta})\big), \qquad k = 1,\dots,m. \tag{5.62}$$

*Proof.* The average of $T_k$ under the canonical probability distribution is equal to

$$\sum_{G\in\mathcal{G}_n} T_k(G) P_{\mathrm{can}}(G \mid \vec{\theta}) = \sum_{G\in\mathcal{G}_n} T_k(G)\, \mathrm{e}^{n^2\left[\vec{\theta}\cdot\vec{T}(G) - \psi_n(\vec{\theta})\right]} =: \mathcal{J}_n. \tag{5.63}$$

Pick $\delta > 0$ and consider the $\delta$-ball $B_\delta(\tilde{h}^*)$ around the maximiser $\tilde{h}^*$ in the quotient space $(\tilde{W}, \delta_\square)$, i.e.,

$$B_\delta(\tilde{h}^*) := \big\{\tilde{h} \in \tilde{W}\colon \delta_\square(\tilde{h}, \tilde{h}^*) < \delta\big\}. \tag{5.64}$$

We denote by $G^\delta$ a graph on $n$ vertices whose graphon is a representative element of the class $\tilde{h}^G$. With a slight abuse of notation, we denote by $G^\delta$ both the graph and the corresponding graphon, and by $\tilde{h}^G$ the corresponding equivalence class in the quotient space $(\tilde{W}, \delta_\square)$. Since $(\tilde{W}, \delta_\square)$ is compact space (recall Proposition 5.2.5), and the graphons associated with finite graphs form a countable family that is dense in $(\tilde{W}, \delta_\square)$ (see Diao *et al.* [39], Lovász and Szegedy [69]), there exists a sequence $(\tilde{h}^{G_n})_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \delta_\square(\tilde{h}^{G_n}, \tilde{h}^*) = 0$. For $n$ large enough the neighbourhood $B_\delta(\tilde{h}^*)$ contains elements of the sequence $(\tilde{h}^{G_n})_{n\in\mathbb{N}}$ and, due to the Lipschitz property (recall Proposition 5.2.6), $\delta_\square(\tilde{h}^{G_n}, \tilde{h}^*) < \delta$ implies $|T_k(\tilde{h}^{G_n}) - T_k(\tilde{h}^*)| < C_k\delta$ for some constant $C_k > 0$ and $k = 1,\dots,m$.

**Upper bound for $\mathcal{J}_n$.** We decompose the sum over $G \in \mathcal{G}_n$ into two parts: the first over $G$ whose graphon lies in $B_\delta(\tilde{h}^*)$, the second over $G$ whose graphon lies in $B_\delta(\tilde{h}^*)^c =: \tilde{W}^{\delta,\#}$. We further denote by

$$\mathcal{G}_n^\delta := \big\{G \in \mathcal{G}_n\colon\ |T_k(\tilde{h}^G) - T_k(\tilde{h}^*)| < \delta,\ \ k = 1,\dots,m\big\}, \tag{5.65}$$

the set of all graphs whose subgraph densities $T_k(G)$ are $\delta$-close to $T_k(\tilde{h}^*)$. A graph from this set is denoted by $G^\delta$. We define the set

$$\mathcal{G}_n^{\delta,\#} := \big\{G \in \mathcal{G}_n\colon\ \tilde{h}^G \in \tilde{W}^{\delta,\#}\big\} \tag{5.66}$$

and, for $k = 1, \dots, m$, obtain the following upper bound:

$$\begin{aligned}
\mathcal{J}_n &= \sum_{G \in \mathcal{G}_n^\delta} T_k(G)\, \mathrm{e}^{n^2\left[\vec{\theta}\cdot\vec{T}(G) - \psi_n(\vec{\theta})\right]} + \sum_{G \in \mathcal{G}_n^{\delta,\#}} T_k(G)\, \mathrm{e}^{n^2\left[\vec{\theta}\cdot\vec{T}(G) - \psi_n(\vec{\theta})\right]} \\
&\leq \frac{(T_k(G^\delta) + \delta) \sum\limits_{G \in \mathcal{G}_n^\delta} \mathrm{e}^{n^2 \vec{\theta}\cdot\vec{T}(G)}}{\sum\limits_{G \in \mathcal{G}_n^\delta} \mathrm{e}^{n^2 \vec{\theta}\cdot\vec{T}(G)}} + \sum_{G \in \mathcal{G}_n^{\delta,\#}} T_k(G)\, \mathrm{e}^{n^2\left[\vec{\theta}\cdot\vec{T}(G) - \psi_n(\vec{\theta})\right]} \\
&= (T_k(G^\delta) + \delta) + \frac{\sum\limits_{G \in \mathcal{G}_n^{\delta,\#}} T_k(G)\, \mathrm{e}^{n^2 \vec{\theta}\cdot\vec{T}(G)}}{\sum\limits_{G \in \mathcal{G}_n} \mathrm{e}^{n^2 \vec{\theta}\cdot\vec{T}(G)}}.
\end{aligned} \tag{5.67}$$

Next, we further bound the second term in (5.67). By definition, for every $n \in \mathbb{N}$ the range of the operator $\vec{T}$ is a finite set

$$R_n := \left\{\vec{g} \in [0,\infty)^m \colon\ \vec{T}(G) = \vec{g},\ G \in \mathcal{G}_n\right\}. \tag{5.68}$$

For the set $R_n$ we observe that $|R_n| = o(n^{m^2})$. In addition, introduce the sets

$$\begin{aligned}
\mathcal{G}_n^{\vec{g}} &:= \{G \in \mathcal{G}_n \colon\ \vec{T}(G) = \vec{g}\}, \\
R_n^{\delta,\#} &:= \{\vec{g} \in [0,\infty)^m \colon\ \vec{T}(G) = \vec{g}, G \in \mathcal{G}_n^{\delta,\#}\} \subset R_n.
\end{aligned} \tag{5.69}$$

The operator $\vec{T}$ is bounded, and so there exists an $M > 0$ such that $\|\vec{T}(G)\| \leq M$ for all $G \in \mathcal{G}_n$. Hence, the second term in (5.67) can be bounded from above by

$$\frac{\sum_{G \in \mathcal{G}_n^{\delta,\#}} T_k(G)\, \mathrm{e}^{n^2 \vec{\theta}\cdot\vec{T}(G)}}{\sum_{G \in \mathcal{G}_n} \mathrm{e}^{n^2 \vec{\theta}\cdot\vec{T}(G)}} \leq \frac{|R_n^{\delta,\#}|\, M\, \exp\left[n^2 \sup_{\vec{g} \in R_n^{\delta,\#}} (\vec{\theta}\cdot\vec{g} + \frac{1}{n^2}\log|\mathcal{G}_n^{\vec{g}}|)\right]}{\exp\left[n^2 \sup_{\vec{g} \in R_n} (\vec{\theta}\cdot\vec{g} + \frac{1}{n^2}\log|\mathcal{G}_n^{\vec{g}}|)\right]}. \tag{5.70}$$

By the large deviation principle in Theorem 5.2.8, we have

$$\frac{1}{n^2}\log|\mathcal{G}_n^{\vec{g}}| = \inf_{\tilde{h} \in \tilde{W}^{\vec{g}}} I(h) + o(1), \tag{5.71}$$

where $\tilde{W}^g = \{\tilde{h} \in \tilde{W} \colon\ \vec{T}(\tilde{h}) = \vec{g}\}$. As a consequence, (5.70) is majorised by

$$\begin{aligned}
&M\,|R_n^*|\,\mathrm{e}^{o(n^2)} \exp\left[n^2\left(\sup_{\vec{g} \in R_n^{\delta,\#}}\left[\vec{\theta}\cdot\vec{g} - \inf_{\tilde{h} \in \tilde{W}^{\vec{g}}} I(\tilde{h})\right] - \sup_{\vec{g} \in R_n}\left[\vec{\theta}\cdot\vec{g} - \inf_{\tilde{h} \in \tilde{W}^{\vec{g}}} I(\tilde{h})\right]\right)\right] \\
&= M\,|R_n^*|\,\mathrm{e}^{o(n^2)} \\
&\exp\left[n^2\left(\sup_{\vec{g} \in R_n^{\delta,\#}} \sup_{\tilde{h} \in \tilde{W}^{\vec{g}}}\left[\vec{\theta}\cdot\vec{T}(\tilde{h}) - I(\tilde{h})\right] - \sup_{\vec{g} \in R_n} \sup_{\tilde{h} \in \tilde{W}^{\vec{g}}}\left[\vec{\theta}\cdot\vec{T}(\tilde{h}) - I(\tilde{h})\right]\right)\right] \\
&= M\,|R_n^*|\,\mathrm{e}^{o(n^2)} \exp\left[n^2\left(\sup_{\tilde{h} \in \tilde{W}^{\delta,\#}}\left[\vec{\theta}\cdot\vec{T}(\tilde{h}) - I(\tilde{h})\right] - \sup_{\tilde{h} \in \tilde{W}}\left[\vec{\theta}\cdot\vec{T}(\tilde{h}) - I(\tilde{h})\right]\right)\right].
\end{aligned} \tag{5.72}$$

The last equation can be justified as follows. Define the sets

$$\tilde{W}_n = \big\{\tilde{h} \in \tilde{W} \colon\ \tilde{h} = \tilde{h}^G \text{ for some graph } G \in \mathcal{G}_n\big\}, \qquad \tilde{W}_n^{\delta,\#} = \tilde{W}^{\delta,\#} \cap \tilde{W}_n. \quad (5.73)$$

Since the graphons associated with finite graphs form a countable set that is dense in $(\tilde{W}, \delta_\square)$, we have that

$$\tilde{W} = \mathrm{cl}\left(\bigcup_{n\in\mathbb{N}} \tilde{W}_n\right), \qquad \tilde{W}^{\delta,\#} = \mathrm{cl}\left(\bigcup_{n\in\mathbb{N}} \tilde{W}_n^{\delta,\#}\right), \quad (5.74)$$

where cl denotes closure. Using (5.74), and recalling that $\vec{T}$ is continuous and $I$ is lower-semicontinuous, we get

$$\lim_{n\to\infty} \sup_{\vec{g}\in R_n^{\delta,\#}} \sup_{\tilde{h}\in\tilde{W}^{\vec{g}}} \big[\vec{\theta}\cdot\vec{T}(\tilde{h}) - I(h)\big] = \sup_{\tilde{h}\in\tilde{W}^{\delta,\#}} \big[\vec{\theta}\cdot\vec{T}(\tilde{h}) - I(\tilde{h})\big], \quad (5.75)$$

and a similar result can be established for the second supremum in the exponent in (5.72). The exponent in (5.72) is negative for all $\delta > 0$ and is independent of $n$. Moreover, by the left-continuity of the graph sequence $(G_n^\delta)_{n\in\mathbb{N}}$, we have that $\lim_{n\to\infty} T_k(G_n^\delta) = T_k(\tilde{h}^*)$ for every $k = 1, \dots, m$ and every $\delta > 0$. Combined with the inequality in (5.67), we obtain, for $k = 1, \dots, m$,

$$\lim_{n\to\infty} \sum_{G\in\mathcal{G}_n} T_k(G)\, \mathrm{e}^{n^2\left[\vec{\theta}\cdot\vec{T}(G) - \psi_n(\vec{\theta})\right]} \le T_k(\tilde{h}^*). \quad (5.76)$$

**Lower bound for $\mathcal{J}_n$.** We distinguish two cases: $T_k(\tilde{h}^*) = 0$ and $T_k(\tilde{h}^*) > 0$. For the first case we trivially get the lower bound

$$\lim_{n\to\infty} \sum_{G\in\mathcal{G}_n} T_k(G)\, \mathrm{e}^{n^2\vec{\theta}\cdot\vec{T}(G)} \ge 0 = T_k(\tilde{h}^*). \quad (5.77)$$

For the second case we show the equivalent upper bound for the inverse, i.e.,

$$\lim_{n\to\infty} \frac{\sum_{G\in\mathcal{G}_n} \mathrm{e}^{n^2\vec{\theta}\cdot\vec{T}(G)}}{\sum_{G\in\mathcal{G}_n} T_k(G)\, \mathrm{e}^{n^2\vec{\theta}\cdot\vec{T}(G)}} \le \frac{1}{T_k(\tilde{h}^*)}. \quad (5.78)$$

Using the fact that $T_k(\tilde{h}^*) \neq 0$ is bounded, and using a similar reasoning as for the upper bound on $\mathcal{J}_n$, the latter is easily verified.

**5.5.5 Remark.** The convergence in (5.62) is not necessarily uniform in $\vec{\theta}$. Our results in Theorem (5.4.1) (II)(b)-(II)(d) indicate that breaking of ensemble equivalence manifests itself through non uniform convergence in (5.62). In Lemma (A.2) we show that uniform convergence holds when the constraint is on the triangle density only, which explains our result in Theorem (5.4.1) (I).

**5.5.6 Remark.** The analogue of Lemma 5.5.4 when the supremum in (5.61) has multiple maximisers in $\tilde{W}$ is considerably more involved.

As observed in Remark 5.2.2, in general the tuning parameter $\vec{\theta}^*$ depends on $n$. We discuss this dependence in Appendix A.

## §5.6 Proof of the Main Theorem 5.4.1

We proceed by computing the relative entropy $s_\infty$. In Sections 5.6.1, 5.6.3, 5.6.4, 5.6.5, 5.6.6 and 5.6.8 we treat the limiting regime where all constraints and parameters are the limiting parameters as in (5.32) and (5.33). In Sections 5.6.2 and 5.6.7 we write $T^*_{\infty,1}, T^*_{\infty,2}, \theta^*_{\infty,1}$ for the limiting regime.

### §5.6.1 Proof of (I)(a) (Triangle model $T_2^* \geq \frac{1}{8}$)

*Proof.* Theorem 5.3.4 says that

$$s_\infty = \sup_{\tilde{h}\in\tilde{W}} \big[\theta^* T_2(\tilde{h}) - I(\tilde{h})\big] - \sup_{\tilde{h}\in\tilde{W}^*} \big[\theta^* T_2(\tilde{h}) - I(\tilde{h})\big]. \tag{5.79}$$

Consider the first term in the right-hand side (5.79). From Lemma 5.5.1 we know that $\theta^* \geq 0$ if and only if $T_2^* \geq \frac{1}{8}$. From Theorem 5.3.3 it follows that if $\theta^* \geq 0$, then

$$\sup_{\tilde{h}\in\tilde{W}} \big[\theta^* T_2(\tilde{h}) - I(\tilde{h})\big] = \sup_{u\in[0,1]} \big[\theta^* u^3 - I(u)\big] = \sup_{u\in[0,1]} \ell_3(u;\theta^*). \tag{5.80}$$

From Radin and Yin [87, Proposition 3.2] we know that $\ell_3(u,\theta^*)$ attains a unique global maximum. Let $u^*(\theta^*) = \arg\sup_{u\in[0,1]} \ell_3(u;\theta^*)$ be the unique global maximiser. Using Lemma A.2, we obtain that $u^*(\theta^*) = {T_2^*}^{1/3}$, which leads to

$$\sup_{u\in[0,1]} \ell_3(u;\theta^*) = \theta^* u^*(\theta^*)^3 - I\big(u^*(\theta^*)\big) = \theta^* T_2^* - I\big({T_2^*}^{1/3}\big). \tag{5.81}$$

As to the second term in the right-hand side of (5.79), we use Chatterjee and Varadhan [31, Proposition 4.2], which states that, for $T_2^* \in (\frac{1}{8}, 1]$,

$$\inf_{\tilde{h}\in\tilde{W}} I(\tilde{h}) := \inf\big\{I(\tilde{h})\colon\ \tilde{h}\in\tilde{W}, T_2(\tilde{h}) = T_2^*\big\} = \inf\big\{I(\tilde{h})\colon\ \tilde{h}\in\tilde{W}, T_2(\tilde{h}) \geq T_2^*\big\}. \tag{5.82}$$

Moreover, $I$ is convex at the point $x = {T_2^*}^{1/3}$, and hence from Chatterjee and Varadhan [31, Theorem 4.3] we have that $\inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}) = I({T_2^*}^{1/3})$. Combining this with (5.81), we conclude that $s_\infty = 0$.

### §5.6.2 Proof of (I)(b) ($T_2^* = 0$)

Consider the Triangle Model with the constraint given by the triangle density $T^* = 0$. It was proven by Erdős *et al.* [42] that almost all triangle-free graph have a bipartite structure. For the case of dense graphs, the condition $T^* = 0$ means that the number of triangles in the graph is of order $o(n^2)$. In the proof we will see that the two ensembles are equivalent and that graphs drawn from the two ensembles have a bipartite structure.

*Proof.* From the construction of the canonical ensemble $P_{\text{can}}$ in Section 5.1.3, we observe that $P_{\text{can}}(G) = 0$ when $T(G) > 0$. This is a direct consequence of (5.5). We write

$$\mathcal{G}_n^0 := \{G \in \mathcal{G}_n : T(G) = 0\} \tag{5.83}$$

for the collection of all graphs with triangle density equal to zero. From (5.6) we obtain that $P_{\mathrm{can}}(G) = 0$ if $G \notin \mathcal{G}_n^0$ and $P_{\mathrm{can}}(G) = |\mathcal{G}_n^0|^{-1}$ if $G \in \mathcal{G}_n^0$. Hence $P_{\mathrm{can}}(G) = P_{\mathrm{mic}}(G)$ when the constraint is given by $T^* = 0$, which yields

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = 0 \qquad \forall n \in \mathbb{N} \tag{5.84}$$

and hence $s_\infty = 0$.

## §5.6.3 Proof of (II)(a) (Edge-Triangle model $T_2^* = T_1^{*3}$)

For the case $T_1^* = T_2^{*\frac{1}{3}}$ we have shown in Lemma 5.5.3 that the canonical ensemble essentially behaves like an Erdős-Rényi model with parameter $p = T_1^*$. Furthermore, the microcanonical ensemble also has an explicit expression, which is found by using the following lemma.

**5.6.1 Lemma.** *If $T_1^* = T_2^{*\frac{1}{3}}$, then*

$$\inf_{\tilde{h} \in \tilde{W}^*} I(\tilde{h}) = I\big(T_2^{*\frac{1}{3}}\big) = I\big(T_1^*\big). \tag{5.85}$$

*Proof.* Consider an element $\tilde{h} \in \tilde{W}^*$ with $\tilde{W}^* := \{\tilde{h} \in \tilde{W} \colon T_1(\tilde{h}) = T_1^* = T_2^{*\frac{1}{3}}, T_2(\tilde{h}) = T_2^*\}$. Using the convexity of $I$ on $\tilde{W}$ and Jensen's inequality, we get

$$I(\tilde{h}) = \int_{[0,1]^2} \mathrm{d}x\, \mathrm{d}y\, I(h(x,y)) \geq I\left(\int_{[0,1]^2} \mathrm{d}x\, \mathrm{d}y\, h(x,y)\right) = I\big(T_1(\tilde{h})\big) = I(T_1^*). \tag{5.86}$$

Hence $I(\tilde{h}) \geq I(T_2^{*\frac{1}{3}})$ for every $\tilde{h} \in \tilde{W}^*$, which proves the claim.

*Proof of (II)(a).* Consider the relative entropy $s_\infty$ as defined in (5.9) and (5.10). Using Lemma 5.5.3, we obtain the expression

$$s_\infty = -\frac{1}{2} T_1^* \log(T_1^*) - \frac{1}{2}(1 - T_1^*) \log(1 - T_1^*) + \inf_{\tilde{h} \in \tilde{W}^*} I(\tilde{h}). \tag{5.87}$$

From Lemma 5.6.1 we have that $\inf_{\tilde{h} \in \tilde{W}^*} I(\tilde{h}) = I(T_1^*)$, which yields $s_\infty = 0$.

## §5.6.4 Proof of (II)(b) ($T_2^* \neq T_1^{*3}$ and $T_2^* \geq \frac{1}{8}$)

*Proof.* From Lemma 5.5.2 we know that if $T_1^* \geq \frac{1}{2}$ and $T_2^* \geq \frac{1}{8}$, then $\theta_1^* \geq 0$ and $\theta_2^* \geq 0$ while if $T_1^* < \frac{1}{2}$ and $T_2^* \geq \frac{1}{8}$, then $\theta_1^* < 0$ and $\theta_2^* \geq 0$. An argument similar as above yields

$$\sup_{\tilde{h} \in \tilde{W}} \Big[\theta_1^* T_1(\tilde{h}) + \theta_2^* T_2(\tilde{h}) - I(\tilde{h})\Big] = \sup_{u \in [0,1]} \ell_3(u; \vec{\theta}^*), \tag{5.88}$$

where for $\theta_1^* \geq 0$ and $\theta_2^* \geq 0$ the last supremum has a unique solution (see Radin and Yin [87, Proposition 3.2]), while for $\theta_1^* < 0$ and $\theta_2^* \geq 0$ it either has a unique solution or two solutions. We treat these two cases separately.

**Unique solution.** Because of the uniqueness of the solution, not all realisable hard constraints can be met in the limit (see Lemma 5.5.4). We observe that, if $T_2^* \geq \frac{1}{8}$ and $T_2^* \neq T_1^{*3}$, in the limit as $n \to \infty$ the canonical ensemble becomes Erdős-Rényi with parameter $p$. This regime is known as the *high-temperature* regime (see Bhamidi *et al.* [12] and Chatterjee and Diaconis [29]). In what follows we determine the parameter $p$ of the canonical ensemble in the limit. From Bhamidi *et al.* [12, Theorem 7] we have that $p = u^*(\vec{\theta}^*)^{\frac{1}{3}}$ with $u^*(\vec{\theta}^*)^{\frac{1}{3}}$ the unique maximiser of (5.88). The expression in (5.88) thus takes the form

$$\begin{aligned} &\sup_{\tilde{h}\in\tilde{W}} \Big[\theta_1^* T_1(\tilde{h}) + \theta_2^* T_2(\tilde{h}) - I(\tilde{h})\Big] \\ &= \sup_{u\in[0,1]} \ell_3(u;\vec{\theta}^*) = \theta_1^* u^*(\vec{\theta}^*)^{\frac{1}{3}} + \theta_2^* u^*(\vec{\theta}^*) - I\big(u^*(\vec{\theta}^*)^{\frac{1}{3}}\big). \end{aligned} \tag{5.89}$$

Consider the second term in the right-hand side of (5.39). From the definition of $\tilde{W}^*$ it is straightforward to see that

$$\sup_{\tilde{h}\in\tilde{W}^*} \Big[\theta_1^* T_1(\tilde{h}) + \theta_2^* T_2(\tilde{h}) - I(\tilde{h})\Big] = \theta_1^* T_1^* + \theta_2^* T_2^* - \inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}), \tag{5.90}$$

where $\tilde{W}^* = \{\tilde{h} \in \tilde{W} \colon T_1(\tilde{h}) = T_1^*, T_2(\tilde{h}) = T_2^*\}$. We observe that, due to $T_2^* \neq T_1^{*3}$, the constant function $h \equiv u^*(\vec{\theta}^*)^{\frac{1}{3}}$ does not lie in $\tilde{W}^*$. This shows that $s_\infty > 0$.

**Two solutions.** The regime in which the right-hand side of (5.88) has two solutions is known as the *low-temperature* regime. In this case the hard constraints $(T_1^*, T_2^*)$, with $T_1^* \in [\frac{1}{4}, \frac{1}{2})$, $T_2^* \geq \frac{1}{8}$, lie on a curve on the $(T_1, T_2)$-plane in such a way such that the tuning parameters $(\theta_1^*, \theta_2^*)$ lie on the *phase transition* curve found in Chatterjee and Diaconis [29] and Radin and Yin [87]. Denote the two solutions of (5.88) by $u_1^*, u_2^*$. Because of the constraint we are considering, we have that neither of them lies in $\tilde{W}^*$. From the compactness of the latter space we see that $s_\infty > 0$.

## §5.6.5 Proof of (II)(c) ($T_2^* \neq T_1^{*3}$, $0 < T_1^* \leq \frac{1}{2}$ and $0 < T_2^{*3} < \frac{1}{8}$)

For the case $0 < T_1^* \leq \frac{1}{2}$, $T_2^* < \frac{1}{8}$ we know from Lemma 5.5.2 that $\theta_1^* \leq 0$ and $\theta_2^* < 0$ for every $n$. Hence, because of (5.33), we have that $\theta_1^* \leq 0$ and $\theta_2^* < 0$. This regime is significantly harder to analyse than the previous regimes. Consider the relative entropy $s_\infty$ and the variational representation given in (5.39). We consider two cases: $T_2^* > T_1^{*3}$ and $T_2^* \leq T_1^{*3}$.

**Case $T_2^* > T_1^{*3}$.** In this case we have the straightforward inequality

$$s_\infty \geq \theta_2^* \left(T_1^{*3} - T_2^*\right) - I(T_1^*) + \inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}). \tag{5.91}$$

Since $T_1^{*3} < T_2^*$, we have $\theta_2^* \left(T_1^{*3} - T_2^*\right) > 0$. We show that

$$\inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}) = \inf\{I(\tilde{h})\colon \tilde{h} \in \tilde{W},\, T_1(\tilde{h}) = T_1^*,\, T_2(\tilde{h}) = T_2^*\} > I(T_1^*). \tag{5.92}$$

Using the convexity of $I$ on $\tilde{W}$ and Jensen's inequality, we obtain that $I(\tilde{h}) \geq I(T_1^*)$ for all $\tilde{h} \in \tilde{W}^*$. Hence

$$\inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}) > \inf\{I(\tilde{h})\colon\ \tilde{h} \in \tilde{W},\, T_1(\tilde{h}) = T_1^*\} = I(T_1^*), \tag{5.93}$$

which settles (5.92). Hence $s_\infty > 0$.

**Case $T_2^* \leq T_1^{*3}$.** We argue similarly as above. We have the straightforward inequality

$$s_\infty \geq \theta_1^* \left(T_2^{*\frac{1}{3}} - T_1^*\right) - I(T_2^{*\frac{1}{3}}) + \inf_{\tilde{h}\in\tilde{W}} I(\tilde{h}). \tag{5.94}$$

We have seen above that $\inf_{\tilde{h}\in\tilde{W}} I(\tilde{h}) > I(T_1^*)$. We further now that $I$ is decreasing on $[0, \frac{1}{2}]$, and so $I(T_1^*) \geq I(T_2^{*1/3})$. Hence $s_\infty > 0$.

## §5.6.6 Proof of (II)(d) ($(T_1^*, T_2^*)$ on the scallopy curve)

We show that if $(T_1^*, T_2^*)$ lies on the lower blue curve in Fig. 5.3 (referred to as the scallopy curve), then $s_\infty > 0$. The case where $T_2^* \geq \frac{1}{8}$ can be dealt with directly via Theorem (II)(b). The proof below deals with the case $T_2^* < \frac{1}{8}$.

*Proof.* We give the proof for $\ell = 2$, the extension to $\ell > 2$ being similar.

Suppose that $T_1^* = \frac{1}{2} + \epsilon$ with $\epsilon \in (0, \frac{1}{6})$, and that $T_2^*$ is chosen as small as possible. It is known that graphs with a relatively high edge density and with a triangle density that is as small as possible have a $d$-partite structure with edges added in a suitable way so that the desired triangle density is obtained (see Radin and Sadun [86] and Pikhurko and Raborov [83]). Consider a graph on $n$ vertices, denoted by $G$, with edge density $T_1 \in (\frac{1}{2}, \frac{2}{3})$ and triangle density as small as possible. The structure of such graphs has been described above before Section 5.5. The graphon counterpart of such graphs is the optimiser of the second supremum in the right-hand side of the variational formula for $s_\infty$. Using Radin and Sadun [86, Theorem 4.2], we obtain

$$\sup_{\tilde{h}\in\tilde{W}^*} \left[\theta_1^* T_1(\tilde{h}) + \theta_2^* T_2(\tilde{h}) - I(\tilde{h})\right] = \theta_1^* T_1^* + \theta_2^* T_2^* - \frac{(1-c(\epsilon))^2}{2}\, I(p(\epsilon)), \tag{5.95}$$

where

$$c(\epsilon) = \frac{2+\sqrt{1-6\epsilon}}{6}, \qquad p(\epsilon) = \frac{4c(\epsilon)(1-2c(\epsilon))}{(1-c(\epsilon))^2}. \tag{5.96}$$

In order to lighten the notation, we drop the dependence of $c$ and $p$ on $\epsilon$. Furthermore, the optimising graphon has the form

$$h_\epsilon^*(x,y) = \begin{cases} 1 & \text{if } x < c < y \text{ or } y < c < x, \\ p & \text{if } c < x < \frac{1+c}{2} < y \text{ or } c < y < \frac{1+c}{2} < x, \\ 0 & \text{otherwise,} \end{cases} \qquad (x,y) \in [0,1]^2, \tag{5.97}$$

which has triangle density

$$T_2(h_\epsilon^*) = \frac{(2+\sqrt{1-6\epsilon})^2}{36}\, \frac{1-\sqrt{1-6\epsilon}}{3} = T(\epsilon). \tag{5.98}$$

Let $\tilde{\mathcal{F}}_\epsilon$ be the set of all maximisers of $\theta_1^* T_1(\tilde{h}) + \theta_2^* T_2(\tilde{h}) - I(\tilde{h})$ on $\tilde{W}$. We show that $h_\epsilon^* \notin \tilde{\mathcal{F}}_\epsilon$, which yields $s_\infty > 0$. From Chatterjee and Diaconis [29, Theorem 6.1] we know that if $\tilde{h} \in \tilde{W}$ maximises $\theta_1^* T_1(\tilde{h}) + \theta_2^* T_2(\tilde{h}) - I(\tilde{h})$ on $\tilde{W}$, then it must satisfy the Euler-Lagrange equations and it must be bounded away from 0 and 1. Hence we see that $\tilde{h}_\epsilon^*$ cannot be a stationary point of $\theta_1^* T_1(\tilde{h}) + \theta_2^* T_2(\tilde{h}) - I(\tilde{h})$ on $\tilde{W}$, and hence cannot be a maximiser.

## §5.6.7 Proof of (II)(e) ($0 < T_1^* \leq \frac{1}{2}$ and $T_2^* = 0$)

*Proof.* Consider the Edge-Triangle Model with constraint given by the edge and triangle densities $T_1^* \in (0, \frac{1}{2}]$ and $T_2^* = 0$. Working as in Section 5.6.2, we find that the canonical ensemble assigns positive probability only to graphs satisfying the constraint $T_2^* = 0$. Defining $\mathcal{G}_n^0$ as in (5.83) we obtain

$$P_{\text{can}}(G \mid \vec{\theta}) = \begin{cases} e^{n^2\left[\theta_1 T_1(G) - \psi_n(\vec{\theta})\right]} & \text{if } G \in \mathcal{G}_n^0, \\ 0 & \text{else,} \end{cases} \tag{5.99}$$

where $\psi_n(\vec{\theta}) = \sum_{G \in \mathcal{G}_n^0} e^{n^2 \theta_1 T_1(G)}$ is the partition function. From (5.99) we observe that the canonical probability distribution depends only on the edge parameter $\theta_1$. The parameter $\theta_1$ is chosen equal to $\theta_1^*$ that matches the soft constraint, i.e.,

$$\sum_{G \in \mathcal{G}_n^0} T_1(G)\, P_{\text{can}}(G \mid \vec{\theta}^*) = T_1^*. \tag{5.100}$$

Arguing as in the proof of Chatterjee and Diaconis [29, Theorem 3.1] we find that the relative entropy equals

$$s_\infty = \sup_{\tilde{h} \in \tilde{W}^0} \left[\theta_{\infty,1}^* T_1(\tilde{h}) - I(\tilde{h})\right] - \sup_{\tilde{h} \in \tilde{W}^*} \left[\theta_{\infty,1}^* T_1(\tilde{h}) - I(\tilde{h})\right], \tag{5.101}$$

where

$$\tilde{W}^0 := \{\tilde{h} \in \tilde{W} \colon\, T_2(\tilde{h}) = 0\}, \qquad \tilde{W}^* := \{\tilde{h} \in \tilde{W} \colon\, T_1(\tilde{h}) = T_{\infty,1}^*\,, T_2(\tilde{h}) = 0\}. \tag{5.102}$$

Using Chatterjee and Diaconis [29, Theorem 7.1 and Theorem 8.2], we obtain that $s_\infty = 0$.

## §5.6.8 Proof of (III) (Star model $T[j]^* \geq 0$)

*Proof.* From Chatterjee and Diaconis [29, Theorem 6.4] we have that, for all $\theta_\infty^* \in \mathbb{R}$,

$$\sup_{\tilde{h} \in \tilde{W}} \left[\theta^* W(\tilde{h}) - I(\tilde{h})\right] = \sup_{u \in [0,1]} \left[\theta^* u^2 - I(u)\right], \tag{5.103}$$

which by Radin and Yin [87, Proposition 3.1] has a unique solution, which we denote by $u^*(\theta^*)$. Using Theorem 5.3.4 we get that

$$s_\infty = \theta^* u^*(\theta^*)^2 - I(u^*(\theta^*)) - \theta^* T^* + \inf_{\tilde{h} \in \tilde{W}^*} I(\tilde{h}), \tag{5.104}$$

where, by Lemma A.2, we have that $u^*(\theta^*) = T^{*\frac{1}{2}}$. This yields

$$s_\infty = -I\big(T^{*\frac{1}{2}}\big) + \inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}). \tag{5.105}$$

We show that $\inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}) = I(T^{*\frac{1}{2}})$. This is done by slightly modifying the proof of Chatterjee and Diaconis [29, Theorem 6.4]. Indeed, observe that

$$T[j](h) = \int_{[0,1]} \mathrm{d}x\, M(x)^j, \qquad M(x) = \int_{[0,1]} \mathrm{d}y\, h(x,y). \tag{5.106}$$

Since $I$ is convex we have

$$\int_{[0,1]^2} \mathrm{d}x\, \mathrm{d}y\, I(h(x,y)) \geq \int_{[0,1]} \mathrm{d}x\, I(M(x)), \qquad h \in W, \tag{5.107}$$

with equality if and only if $h(x,y)$ is the same for almost all $y$. Since $h$ is a symmetric function, we get that equality holds if and only if $h$ is constant. For the constant function $h \equiv (T_j)^{1/j} \in W^* := \{h \in W : T_j(h) = T_j\}$, (5.107) is an equality. Hence, for any minimiser of $I$ on $\tilde{W}^*$ the inequality must be an equality, and thus any minimiser must be constant. This shows that $s_\infty = 0$.

# §A Appendix

In this appendix we elaborate on the assumption made in (5.33), i.e., the multiplier $\vec{\theta}^*_n$ converges to a limit $\vec{\theta}^*_\infty$ as $n \to \infty$. In order to get a meaningful limit, we consider constraints $\vec{T}^*_n$ such that

$$\lim_{n\to\infty} \vec{T}^*_n = \vec{T}^*_\infty. \tag{5.108}$$

It is straightforward to deduce from Corollary 5.2.9 and (5.26)–(5.30) that if $\{\vec{T}^*_n\}$ is bounded away from 0 and 1 component-wise, then $(\vec{\theta}^*_n)_{n\in\mathbb{N}}$ is bounded away from $-\infty$ and $+\infty$ component-wise. Such a sequence contains a converging subsequence, say, $(\vec{\theta}^*_{n_k})_{k\in\mathbb{N}}$, which in general need not be unique. Thus, as long as the constraint is component-wise bounded away from 0 and 1, the asymptotic expressions derived in this paper exist, but their values may depend on the subsequence we choose. The value of $s_\infty$ depends on the chosen subsequence, but whether it is positive or zero (i.e., whether there is equivalence) does *not*. A deeper investigation of the behaviour of $\{\vec{\theta}^*_n\}_{n\in\mathbb{N}}$ is interesting, but is beyond the scope of this paper.

We first extend Theorem 5.3.4 for the case when the tuning parameter $\vec{\theta}^*$ depends on $n$.

**A.1 Lemma.** *Consider the microcanonical ensemble defined in* (5.2) *with constraint* $\vec{T} = \vec{T}^*_n$ *defined as in* (5.27)*, and the canonical ensemble defined in* (5.28)–(5.29) *with parameter* $\vec{\theta} = \vec{\theta}^*_n$ *such that* (5.30) *holds. If the conditions in Remark* 5.3.1 *hold, then* (5.39) *holds too.*

*Proof.* The proof of Theorem 5.3.4 carries over to the setting in which the parameter $\vec{\theta}^*$ depends on $n$, i.e., $\vec{\theta}^* = \vec{\theta}^*_n$. The only non-trivial step is to show that

$$\lim_{n\to\infty} \psi_n(\vec{\theta}^*_n) = \psi_\infty(\vec{\theta}^*_\infty). \tag{5.109}$$

In the proof of Theorem 5.3.4 we have shown the pointwise convergence

$$\lim_{n\to\infty} \psi_n(\vec{\theta}) = \psi_\infty(\vec{\theta}), \tag{5.110}$$

for every $\vec{\theta} \in \mathbb{R}^m$, independently of $n$. A straightforward computation shows that $\nabla\psi_n(\vec{\theta}) = (\langle T_1\rangle, \dots, \langle T_m\rangle)$, recall (5.30) . Observe that for the specific choice of the parameter $\vec{\theta} = \vec{\theta}^*_n = \vec{\theta}^*$, we have that $\nabla\psi_n(\vec{\theta}^*_n) = (T^*_1, \dots, T^*_m)$, which yields $\|\nabla\psi_n(\vec{\theta})\| \le m$ for all $n \in \mathbb{N}$ and $\vec{\theta} \in \mathbb{R}^m$. We prove (5.109) under the assumptions made in Remark 5.3.1,

$$\begin{aligned} |\psi_n(\vec{\theta}^*_n) - \psi_\infty(\vec{\theta}^*_\infty)| &\le |\psi_n(\vec{\theta}^*_n) - \psi_n(\vec{\theta}^*_\infty)| + |\psi_n(\vec{\theta}^*_\infty) - \psi_\infty(\vec{\theta}^*_\infty)| \\ &\le \|\nabla\psi_n(\vec{\eta})\|\,\|\vec{\theta}^*_n - \vec{\theta}^*_\infty\| + |\psi_n(\vec{\theta}^*_\infty) - \psi_\infty(\vec{\theta}^*_\infty)| \\ &\le m\,\|\vec{\theta}^*_n - \vec{\theta}^*_\infty\| + |\psi_n(\vec{\theta}^*_\infty) - \psi_\infty(\vec{\theta}^*_\infty)| \to 0, \quad n\to\infty, \end{aligned} \tag{5.111}$$

where the second inequality follows from the mean-value theorem for some $\vec{\eta} = c\,\vec{\theta}^*_n + (1-c)\,\vec{\theta}^*_\infty$, $c \in (0,1)$. The rest of the proof of Theorem 5.3.4 carries over intact.

In the following lemma we extend the result of Lemma 5.5.4 for the case the operator $\vec{T}$ is the triangle density $T_2$ . This extension is needed in the proof of Theorem 5.4.1 (I).

**A.2 Lemma.** *Consider the operator $T_2\colon \tilde{W} \to \mathbb{R}$ which is bounded and continuous with respect to the $\delta_\square$-norm as defined in* (5.13)*. For $n \in \mathbb{N}$, consider the tuning parameter $\theta^*_n$ according to* (5.30)*, i.e.,*

$$\sum_{G\in\mathcal{G}_n} T_2(G)\,P_{\mathrm{can}}(G) = T^*_2. \tag{5.112}$$

*Suppose that $T^*_2 \ge \frac{1}{8}$ and that the limits $T^*_\infty, \theta^*_\infty$ in* (5.33) *exists. Then*

$$\lim_{n\to\infty} \sum_{G\in\mathcal{G}_n} T_2(G)\,P_{\mathrm{can}}(G) = \lim_{n\to\infty} \frac{\sum_{G\in\mathcal{G}_n} T_2(G)\,\mathrm{e}^{n^2\theta^*_n T_2(G)}}{\sum_{G\in\mathcal{G}_n} \mathrm{e}^{n^2\theta^*_n T_2(G)}} = u^*(\theta^*_\infty), \tag{5.113}$$

*where*

$$u^*(\theta) = \mathit{arg}\sup_{0\le u\le 1} [\theta u^3 - I(u)]. \tag{5.114}$$

*Proof.* From Lemma 5.5.2, since $T^*_2 \ge \frac{1}{8}$ we have that $\theta^*_n \ge 0$ for all $n$. Consequently, $\theta^* \ge 0$. Define, for $\theta \ge 0$, the function

$$f_n(\theta) := \sum_{G\in\mathcal{G}_n} T_2(G) P_{\mathrm{can}}(G \mid \vec{\theta}) = \frac{\sum_{G\in\mathcal{G}_n} T_2(G)\,\mathrm{e}^{n^2\theta T_2(G)}}{\sum_{G\in\mathcal{G}_n} \mathrm{e}^{n^2\theta T_2(G)}} \tag{5.115}$$

and consider the variational problem in (5.61). From Chatterjee and Diaconis [29] we have that, for $\theta \geq 0$,

$$\psi_\infty(\theta) := \sup_{\tilde{h} \in \tilde{W}} \big[\theta T(\tilde{h}) - I(\tilde{h})\big] = \sup_{0 \leq u \leq 1} \big[\theta u^3 - I(u)\big]. \tag{5.116}$$

From Radin and Sadun [86, Theorem 2.1] we have that the function $\theta \to u^*(\theta)$ is differentiable on $[0, \infty)$. We also observe that

$$u^*(0) = \tfrac{1}{2}, \qquad \lim_{\theta \to \infty} u^*(\theta) = 1. \tag{5.117}$$

Moreover, for very $n$, $\theta \mapsto f_n(\theta)$ is continuous on $[0, \infty)$. Hence, combining Lemma 5.5.4, the continuity of $f_n$ for every $n$, the analyticity of the limiting function $\theta \mapsto u^*(\theta)$ and (5.117), we obtain that if the limit $\theta_\infty$ in (5.33) exists, then

$$\lim_{n \to \infty} f_n(\theta_n^*) = u^*(\theta_\infty^*) = T_\infty^*, \tag{5.118}$$

which proves the claim.

CHAPTER 6

# Breaking of Ensemble Equivalence for Perturbed Erdős-Rényi Random Graphs

This chapter is based on:
F. den Hollander, M. Mandjes, A. Roccaverde, and N. Starreveld. Breaking of ensemble equivalence for perturbed erdős-rényi random graphs. *arXiv:1807.07750*

**Abstract**

In a previous paper we analysed a simple undirected random graph subject to constraints on the total number of edges and the total number of triangles. We considered the dense regime in which the number of edges per vertex is proportional to the number of vertices. We showed that, as soon as the constraints are *frustrated*, i.e., do not lie on the Erdős-Rényi line, there is breaking of ensemble equivalence, in the sense that the specific relative entropy per edge of the *microcanonical ensemble* with respect to the *canonical ensemble* is strictly positive in the limit as the number of vertices tends to infinity. In the present paper we analyse what happens near the Erdős-Rényi line. It turns out that the way in which the specific relative entropy tends to zero depends on whether the total number of triangles is slightly larger or slightly smaller than typical. We identify what the constrained random graph looks like asymptotically in the microcanonical ensemble.

## §6.1 Introduction

In this paper we analyse random graphs that are subject to *constraints*. Statistical physics prescribes what probability distribution on the set of graphs we should choose when we want to model a given type of constraint [53]. Two important choices are:

(1) The *microcanonical ensemble*, where the constraints are *hard* (i.e., are satisfied by each individual graph).

(2) The *canonical ensemble*, where the constraints are *soft* (i.e., hold as ensemble averages, while individual graphs may violate the constraints).

For random graphs that are large but finite, the two ensembles are obviously different and, in fact, represent different empirical situations. Each ensemble represents the unique probability distribution with *maximal entropy* respecting the constraints. In the limit as the size of the graph diverges, the two ensembles are traditionally *assumed* to become equivalent as a result of the expected vanishing of the fluctuations of the soft constraints, i.e., the soft constraints are expected to behave asymptotically like hard constraints. This assumption of *ensemble equivalence* is one of the corner stones of statistical physics, but it does *not* hold in general (see [97] for more background).

In a series of papers the question of possible breaking of ensemble equivalence was investigated for various choices of the constraints, including the degree sequence and the total number of edges, wedges and triangles. Both the *sparse regime* (where the number of edges per vertex remains bounded) and the *dense regime* (where the number of edges per vertex is of the order of the number of vertices) have been considered. The effect of *community structure* on ensemble equivalence has been investigated as well. Relevant references are [48], [50], [37], [93] and [92]. In [37] we considered a random graph subject to constraints on the total number of edges and the total number of triangles, in the dense regime. With the help of *large deviation theory for graphons*, see [31], we derived a variational formula for $s_\infty = \lim_{n\to\infty} n^{-2} s_n$, where $n$ is the number of vertices and $s_n$ is the *relative entropy* of the microcanonical ensemble with respect to the canonical ensemble. We found that $s_\infty > 0$ when the constraints are *frustrated*. In the present paper we analyse the behaviour of $s_\infty$ when the constraints are close to but different from those of the Erdős-Rényi random graph, and we identify what the constrained random graph looks like asymptotically in the microcanonical ensemble. It turns out that the behaviour changes when the total number of triangles is larger, respectively, smaller than that of the Erdős-Rényi random graph with a given total number of edges.

While breaking of ensemble equivalence is a relatively new concept in the theory of random graphs, there are many studies on the asymptotic structure of random graphs. In the pioneering work [31], followed by [70], the large deviation principle for dense Erdős-Rényi random graphs was proven and the asymptotic structure of constrained Erdős-Rényi random graphs was described as the solution of a variational problem. In the past few years significant progress has been made regarding sparse random graphs as well. We refer the reader to [30], [36], [71] and [105]. Two other random graph models that have been extensively studied are the exponential random graph model

and the constrained exponential random graph model. Exponential random graphs, which are related to the canonical ensemble we consider in this paper, were introduced rather early in the physics literature, see [80] and the references therein, and rigorously analysed in detail in [12] and [29]. In [12] the authors investigated the mixing time of the Glauber dymanics and they showed that, in some cases, graphs drawn from the exponential random graph model, behave asymptotically like Erdős-Rényi random graphs with a biased parameter. In [29] the authors verified and generalised this result using the machinery developed in [31]. Their main result was an asymptotic expression for the logarithm of the partition function in terms of a variational problem. Additionally, they showed that, in the edge-triangle model, a phase transition, which is defined as a discontinuity in the derivative of the logarithm of the partition function, occurs for specific values of the parameters. The existence of phase transitions in the exponential random graph model was investigated further in [87] and [101] and for directed graphs in [4]. An analysis of sparse exponential random graphs was carried out in [103]. A second random graph model, which is also related to the random graph models we study in this paper, and has received significant attention in the literature, is the constrained exponential random graph model, we refer the reader to [3], [63], [67] and [102] for a detailed description and analysis. A stream of research that is relevant to our work concerns the asymptotic description of the structure of graphs drawn from the microcanonical ensemble with a constraint on the edge and triangle density. In [86] the authors studied the behavior of random graphs with edge and triangle densities close to the Erdős-Rényi curve. They managed to identify the scaling behavior close to the curve by proving a bound on the entropy function. In one of the results in this paper we rigorously prove the results of [86] and we determine the exact structure of constrained random graphs close to the Erdős-Rényi curve. The same question was investigated in [75] for a constraint on the edge and triangle density close to the lower boundary curve of the admissibility region. In [64] the authors managed to determine, through extensive simulations, curves in the admissibility region where phase transitions occur in the structure of constrained random graphs.

The remainder of this paper is organised as follows. In Section 6.2 we define the two ensembles, give the definition of equivalence of ensembles in the dense regime, some basic facts about graphons and we the *variational representation* of $s_\infty$ derived in [37] when the constraints are on the total numbers of subgraphs drawn from a finite collection of subgraphs. We also recall the analysis of $s_\infty$ in [37] for the special case where the subgraphs are the edges and the triangles. In Section 6.3 we state our main theorems. Proofs are given in Sections 6.4 and 6.5.

## §6.2 Definitions and preliminaries

The microcanonical and the canonical ensemble, as well as the relative entropy density have been defined in Section 1.4.1 and 1.4.2. Graphons and their properties have been defined in Section 5.2.2 5.2.3 and 5.3 of Chapter 5. In this section we recall the definition of ensemble equivalence in the dense regime and the main two theorems of Chapter 5.

**6.2.1 Definition.** Following [37] $P_{\text{mic}}$ and $P_{\text{can}}$ are said to be equivalent in the dense regime when

$$s_\infty := \lim \frac{1}{n^2} S_n(P_{\text{mic}} \mid P_{\text{can}}) = 0. \tag{6.1}$$

The key result in [37] is the following variational formula for $s_\infty$.

**6.2.2 Theorem.** [37] *Subject to* (5.30), (5.32) *and* (5.33),

$$\lim_{n\to\infty} n^{-2} S_n(P_{\text{mic}} \mid P_{\text{can}}) =: s_\infty \tag{6.2}$$

*with*

$$s_\infty = \sup_{\tilde{h}\in\tilde{W}} \left[\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\right] - \sup_{\tilde{h}\in\tilde{W}^*} \left[\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\right]. \tag{6.3}$$

Theorem 6.2.2 and the compactness of $\tilde{W}^*$ give us a *variational characterisation* of ensemble equivalence: $s_\infty = 0$ if and only if at least one of the maximisers of $\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})$ in $\tilde{W}$ also lies in $\tilde{W}^* \subset \tilde{W}$. Equivalently, $s_\infty = 0$ when at least one the maximisers of $\vec{\theta}^*_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})$ satisfies the hard constraint. Theorem 6.2.2 allows us to identify cases where ensemble equivalence holds ($s_\infty = 0$) or is broken ($s_\infty > 0$). In [37] a detailed analysis was given for the special case where the constraint is on the total number of edges and the total number of triangles. The analysis in [37] relied on the large deviation principle for dense Erdős-Rényi random graphs established in [31]. The function defined in (5.19) plays a crucial role and is related to the rate function of the large deviation principle.

**6.2.3 Theorem.** [37] *For the edge-triangle model, $s_\infty = 0$ when*

- $T_2^* = T_1^{*3}$,
- $0 < T_1^* \leq \frac{1}{2}$ *and* $T_2^* = 0$,

*while $s_\infty > 0$ when*

- $T_2^* \neq T_1^{*3}$ *and* $T_2^* \geq \frac{1}{8}$,
- $T_2^* \neq T_1^{*3}$, $0 < T_1^* \leq \frac{1}{2}$ *and* $0 < T_2^* < \frac{1}{8}$,
- $(T_1^*, T_2^*)$ *lies on the scallopy curve in Figure* 6.1.

Here, $T_1^*, T_2^*$ are in fact the limits $T_{1,\infty}^*, T_{2,\infty}^*$ in (5.32), but in order to keep the notation light we now also suppress the index $\infty$.
Theorem 6.2.3 is illustrated in Fig. 6.1. The region on and between the blue curves corresponds to the choices of $(T_1^*, T_2^*)$ that are *graphical*, i.e., there exists a graph with edge density $T_1^*$ and triangle density $T_2^*$. The red curves represent ensemble equivalence, the blue curves and the grey region represent breaking of ensemble equivalence, while in the white region between the red curve and the lower blue curve we do not know what happens. Breaking of ensemble equivalence arises from *frustration* between the values of $T_1^*$ and $T_2^*$.

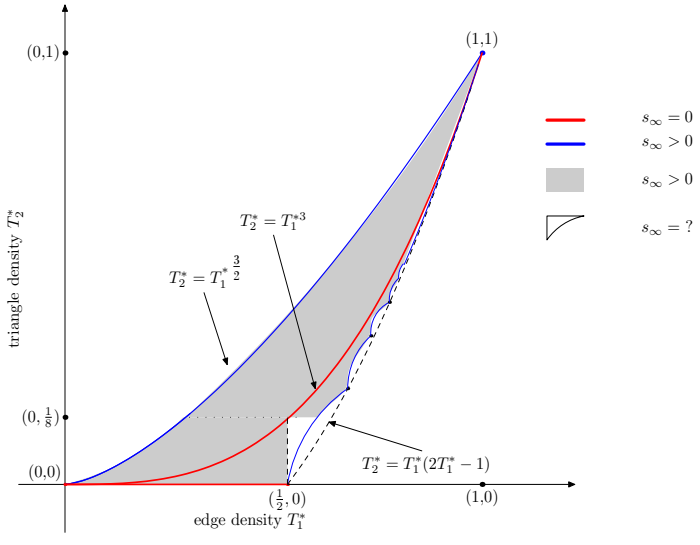



*Figure 6.1: The admissible edge-triangle density region is the region on and between the blue curves [86].*

The lower blue curve, called the *scallopy curve*, consist of infinitely many pieces labelled by $\ell \in \mathbb{N} \setminus \{1\}$. The $\ell$-th piece corresponds to $T_1^* \in (\frac{\ell-1}{\ell}, \frac{\ell}{\ell+1}]$ and a $T_2^*$ that is a function of $T_1^*$ given by

$$T_2^* = \frac{(\ell-1)\left(\ell - 2\sqrt{\ell(\ell - T_1^*(\ell+1))}\right)\left(\ell + \sqrt{\ell(\ell - T_1^*(\ell+1))}\right)^2}{\ell^2(\ell+1)^2}. \tag{6.4}$$

We refer the reader to [83], [85], [86] and [88] for more details.

## §6.3 Theorems

In this section we present our results which address the following two issues:

- In Theorems 6.3.1–6.3.3 we identify the scaling behaviour of $s_\infty$ for fixed $T_1^*$ and $T_2^* \downarrow T_1^{*3}$, respectively, $T_2^* \uparrow T_1^{*3}$. It turns out that the way in which $s_\infty$ tends to zero differs in the two cases.
- In Propositions 6.3.5–6.3.7 we characterise some possible asymptotic structures of random graphs drawn from the microcanonical ensemble when the hard constraint is on the edge and triangle density. Our results indicate that the structure of the graphs differs for $T_2^* \downarrow T_1^{*3}$, respectively, $T_2^* \uparrow T_1^{*3}$.

In the sequel we make the following two assumptions:

**Assumption 1.** Fix the edge density $T_1^* \in (0,1)$ and consider the triangle density $T_1^{*3} + \epsilon$, for some $\epsilon$ either positive or negative. For this pair of constraints we consider

the Lagrange multipliers $\vec{\theta}^*_\infty(\epsilon) := (\theta^*_1(\epsilon), \theta^*_2(\epsilon))$ as defined in Section 5.3.1. Then, for $\epsilon$ sufficiently small, we have the representation

$$\sup_{\tilde{h}\in\tilde{W}} \Big[\theta^*_1(\epsilon)T_1(\tilde{h}) + \theta^*_2(\epsilon)T_2(\tilde{h}) - I(\tilde{h})\Big] = \theta_1 T^*_1 - I(T^*_1) + (\gamma_1 T^*_1 + \gamma_2 T^{*3}_1)\epsilon + O(\epsilon^2), \tag{6.5}$$

where $\theta_1 := \theta_1(0), \gamma_1 = \theta'_1(0)$ and $\gamma_2 = \theta'_2(0)$.

In Section 6.4.1 we show that Assumption 1 is true when $T^*_1 \in [\frac{1}{2}, 1)$. For $T^*_1 \in (0, \frac{1}{2})$ we can prove (6.8) and (6.9) below, but with $\geq$ replacing the equality. If Assumption 1 is true, then we again obtain (6.8) and (6.9) with equality. If it fails, then we have strict inequality.

**Assumption 2.** Fix the edge density $T^*_1 \in (0, 1)$ and consider the triangle density $T^{*3}_1 + \epsilon$, for some $\epsilon$ either positive or negative. For this pair of constraints we consider the microcanonical entropy

$$-J(\epsilon) := \sup\{-I(\tilde{h}) :\ \tilde{h} \in \tilde{W},\ T_1(\tilde{h}) = T^*_1, T_2(\tilde{h}) = T^{*3}_1 + \epsilon\}. \tag{6.6}$$

Then for $\epsilon$ sufficiently small the solution of (6.6), denoted by $h^*_\epsilon$, has the following form

$$h^*_\epsilon = T^*_1 + g_\epsilon, \qquad \text{where} \qquad g_\epsilon = g_{11}1_{I\times I} + g_{12}1_{(I\times J)\cup(J\times I)} + g_{22}1_{J\times J}, \tag{6.7}$$

with $g_{11}, g_{12}, g_{22} \in [-T^*_1, 1 - T^*_1]$ and $I, J \subset [0, 1]$.

Assumption 2 is based on the following intuitive argument. Suppose we want to maximise the microcanonical entropy among all piecewise constant graphons. Then we expect the entropy to decrease when we add more structure, i.e., more steps, in the graphon. A piecewise constant graphon with $m$ steps corresponds to a random graph where the vertices are divided into $m$ groups, and within each group we make an ER random graph with some probability. We expect that the microcanonical entropy will decrease as $m$ increases. This statement is also supported by extensive numerical experiments performed in [65].

The methodology we rely on in order to analyse the variational problem in (6.6) does not always identify the exact optimal graphon. It identifies a candidate optimal graphon, which is sufficient, in some cases for the scaling behaviour of the relative entropy $s_\infty$. We call these graphons *balance optimal*. Roughly speaking, a balance optimal graphon is obtained when solving the optimisation problem in (6.6) in a smaller class of graphons than the whole class of graphons that satisfy the hard constraint. This is the class of graphons satisfying the conditions in Assumption 1 and such that the values $g_{11}, g_{12}, g_{22}$ all correspond to contributions of the same order. The precise definition of a balance optimal graphon is given in Section 6.5. We want to investigate in this chapter whether the global maximizer of (6.6) lies in this smaller class of graphons. We show that balance optimisers have specific structural properties. But, for the case of a perturbation upwards, the unique optimal graphon does not lie in this class, and this happens because $\lambda(I)$ gets very small as $\epsilon \downarrow 0$ while $g_{11}$ stays constant. We refer the reader to [66]. For the case of a perturbation downwards the

exact structure of the unique optimal graphon is still not known: the only results we are aware of come from an extensive numerical study, [63]. From this numerical study it seems that, at least for $T_1^* \in (0, \tilde{T}_1^*)$, with $\tilde{T}_1^* \approx 0.44$, the unique global optimiser is indeed a balance optimal graphon. In this chapter we investigate this question further by identifying the balance optimal graphons and comparing them with the results established numerically in [63].

Balance optimal graphons are candidate optimisers of $J(\epsilon)$. In what follows, because all the graphons we derive are balance optimal graphons, we simply speak of optimal graphons. When at some point a clear distinction is needed we say so. Another important feature is that balance optimal graphons are in general not unique. In the following sections we construct various balance optimal graphons, exhibiting the different structures that can emerge. The variational problem $J(\epsilon)$ in (6.6) has been solved in [66] for the case $T_2^* > T_1^{*3}$, while the case $T_2^* < T_1^{*3}$ still remains unsolved. In this chapter we consider only a small perturbation around the typical values, but the advantage of our method is that it is simpler and yields more intuition about the way the constraint is attained. Moreover, it also applies for the case $\epsilon < 0$, which has not been rigorously analysed before. In [63] the authors identify the maximizers of the microcanonical entropy numerically. The optimal graphons obtained numerically in [63] agree structurally with the balance optimal graphons that we find.

**6.3.1 Theorem.** *For $T_1^* \in (0,1)$ and $T_1^* \neq \frac{1}{2}$,*

$$\lim_{\epsilon \downarrow 0} \epsilon^{-1} s_\infty(T_1^*, T_1^{*3} + 3T_1^*\epsilon) = \frac{6}{1-2T_1^*} \log \frac{T_1^*}{1-T_1^*} \in (0,\infty). \tag{6.8}$$

**6.3.2 Theorem.** *For $T_1^* \in (0, \frac{1}{2}]$,*

$$\limsup_{\epsilon \downarrow 0} \epsilon^{-2/3} s_\infty(T_1^*, T_1^{*3} - T_1^{*3}\epsilon) \leq \frac{1}{4} \frac{T_1^*}{1-T_1^*} \in (0,\infty). \tag{6.9}$$

**6.3.3 Theorem.** *For $T_1^* \in (\frac{1}{2}, 1)$,*

$$\limsup_{\epsilon \downarrow 0} \epsilon^{-2/3} s_\infty(T_1^*, T_1^{*3} - T_1^{*3}\epsilon) \leq f(T_1^*, \bar{T}_1^*) \in (0,\infty), \tag{6.10}$$

*where $\bar{T}_1^* \in (-T_1^*, 0)$ is the unique point where the function $x \to f(T_1^*, x)$, defined by*

$$f(T_1^*, x) := T_1^{*2} \frac{I(T_1^* + x) - I(T_1^*) - I'(T_1^*)x}{x^2}, \quad x \in (-T_1^*, 0), \tag{6.11}$$

*attains its global minimum.*

We illustrate these results in Figure 6.2. In the left panel we plot the limits in the right-hand side of (6.9)–(6.10) as a function of $T_1^*$. In the right panel we plot $s_\infty(T_1^*, T_1^{*3} + \epsilon)$ as a function of $\epsilon$, for $\epsilon$ sufficiently small, and for four different values of $T_1^*$.

**6.3.4 Remark.** We believe, and there is numerical evidence in [63], that the results in (6.9) and (6.10) hold with equality and that the corresponding limits exist.

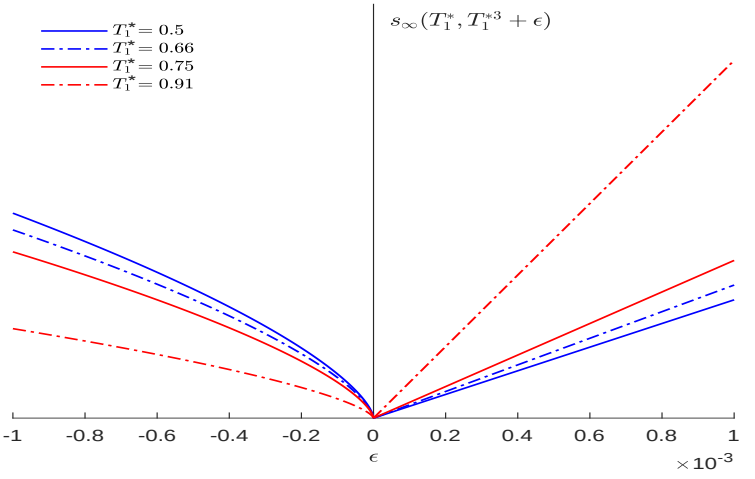



*Figure 6.2: Limit of scaled $s_\infty$ as a function of $\epsilon$ for $\epsilon$ sufficiently small.*

In Proposition 6.3.5–6.3.7 below we identify the structure of balance optimal graphons corresponding to the perturbed constraints in the microcanonical ensemble in the limit as $n \to \infty$.

**6.3.5 Proposition.** *When the ER-line is approached from above, a balance optimal graphon is given by*

$$h = T_1^* + \sqrt{\epsilon}\, g^* + O(\epsilon) \qquad \text{(global perturbation)} \tag{6.12}$$

*with $g^*$ given by*

$$g^*(x,y) = \begin{cases} 2, & (x,y) \in [0,\frac{1}{2}]^2, \\ 0, & (x,y) \in [0,\frac{1}{2}] \times (\frac{1}{2},1] \cup (\frac{1}{2},1] \times [0,\frac{1}{2}], \\ -2, & (x,y) \in (\frac{1}{2},1]^2. \end{cases} \tag{6.13}$$

It is important to mention that the balance optimal graphon determined in (6.13) is not unique, in the sense that there are multiple graphons that are balance optimal. From Proposition 6.3.5 we also see that it is possible that the class of balance optimisers does not contain the actual unique optimiser of $J(\epsilon)$. For this pair of constraints, and from [66], we have that the actual unique optimiser, denoted by $h_\epsilon^*$, is given by

$$h_\epsilon^*(x,y) = \begin{cases} h_{11}, & (x,y) \in [0,\lambda\epsilon]^2, \\ 1 - T_1^* + h_1\epsilon, & (x,y) \in [0,\lambda\epsilon] \times (\lambda\epsilon,1] \cup (\lambda\epsilon,1] \times [0,\lambda\epsilon], \\ T_1^* + h_2\epsilon, & (x,y) \in (\lambda\epsilon,1]^2, \end{cases} \tag{6.14}$$

where

$$\lambda := \frac{1}{(1-2T_1^*)^2}, \qquad h_1 := \frac{1}{2}h_2, \qquad h_2 := -\frac{2}{1-2T_1^*}. \tag{6.15}$$

The term $h_{11}$ solves the equation $I'(h_{11}) = 3I'(1 - T_1^*)$ and is constant as $\epsilon \downarrow 0$. For details on this issue we refer to [66]. As mentioned above, balance optimal graphons have the structural property that $g_{11}, g_{12}, g_{22}$ all contribute equally to the constraint. This is not the case for the graphon in (6.14) because only $g_{12}$ and $g_{22}$ contribute to the constraint to leading order. The exact computations are provided in Section 6.5.

From (6.13) and (6.14) we observe that balance optimal graphons can have structures very different from the optimal graphons.

**6.3.6 Proposition.** *When the ER-line is approached from below and $T_1^* \in (0, \frac{1}{2}]$, a balance optimal graphon is given by*

$$h = T_1^* + \epsilon^{1/3} g^* + O(\epsilon^{1/3}) \qquad \text{(global perturbation)} \tag{6.16}$$

*with $g^*$ given by*

$$g^*(x,y) = \begin{cases} -T_1^*, & (x,y) \in [0, \frac{1}{2}]^2, \\ T_1^*, & (x,y) \in [0, \frac{1}{2}] \times (\frac{1}{2}, 1] \cup (\frac{1}{2}, 1] \times [0, \frac{1}{2}], \\ -T_1^*, & (x,y) \in (\frac{1}{2}, 1]^2. \end{cases} \tag{6.17}$$

*This $g^*$ is not unique, in the sense that there are multiple graphons that are balance optimal.*

**6.3.7 Proposition.** *When the ER-line is approached from below and $T_1^* \in (\frac{1}{2}, 1)$, the unique balance optimal graphon is given by*

$$h = T_1^* + g_\epsilon^* \qquad \text{(local perturbation)} \tag{6.18}$$

*with $g_\epsilon^*$ defined by*

$$g_\epsilon^*(x,y) := \begin{cases} \frac{T_1^{*2}}{T_1^*} \epsilon^{2/3}, & (x,y) \in [0, 1 - \frac{T_1^*}{|T_1^*|} \epsilon^{1/3}]^2 \\ T_1^* \epsilon^{1/3}, & (x,y) \in [0, 1 - \frac{T_1^*}{|T_1^*|} \epsilon^{1/3}] \times [1 - \frac{T_1^*}{|T_1^*|} \epsilon^{1/3}, 1] \quad \text{or} \\ & (x,y) \in [1 - \frac{T_1^*}{|T_1^*|} \epsilon^{1/3}, 1] \times [0, 1 - \frac{T_1^*}{|T_1^*|} \epsilon^{1/3}], \\ \bar{T}_1^*, & (x,y) \in [1 - \frac{T_1^*}{|T_1^*|} \epsilon^{1/3}, 1]^2, \end{cases} \tag{6.19}$$

*with $\bar{T}_1^* \in (-T_1^*, 0)$ defined in Theorem 6.3.3.*

In conclusion, Theorems 6.3.1–6.3.3 say that at a fixed density of the edges it is less costly in terms of relative entropy to increase the density of triangles than to decrease it. The ER-line represents a crossover in the cost (see Figure 6.2, right panel). Above the ER-line the cost is linear in the distance, below the ER-line the cost is proportional to the $\frac{2}{3}$-power of the distance. Propositions 6.3.5–6.3.7 show that the optimal perturbation of the ER-graphon is global above the ER-line and below the ER-line when the edge density is less than $\frac{1}{2}$ and local below the ER-line when the edge density is larger than $\frac{1}{2}$.

## §6.4 Proofs of Theorems 6.3.1-6.3.3

In this section we prove Theorems 6.3.1–6.3.3. Along the way we use the results given in Propositions 6.3.5–6.3.7, which we prove in Section 6.5.

### §6.4.1 Proof of Theorem 6.3.1

For ease of notation we drop the superscript $*$ from the constraint on the edge density and write $T_1$ instead of $T_1^*$. Let

$$T_1(\epsilon) = T_1, \qquad T_2(\epsilon) = T_1^3 + 3T_1\epsilon. \tag{6.20}$$

The factor $3T_1$ appearing in front of the $\epsilon$ is put in for convenience. We know that for every pair of graphical constraints $(T_1(\epsilon), T_2(\epsilon))$ there exists a unique pair of Lagrange multipliers $(\theta_1(\epsilon), \theta_2(\epsilon))$ corresponding to these constraints. For an elaborate discussion on this issue we refer the reader to [37]. By considering the Taylor expansion of the Lagrange multipliers $(\theta_1(\epsilon), \theta_2(\epsilon))$ around $\epsilon = 0$, we obtain

$$\theta_1(\epsilon) = \theta_1 + \gamma_1\epsilon + \tfrac{1}{2}\Gamma_1\epsilon^2 + O(\epsilon^3), \qquad \theta_2(\epsilon) = \gamma_2\epsilon + \tfrac{1}{2}\Gamma_2\epsilon^2 + O(\epsilon^3), \tag{6.21}$$

where

$$\theta_1(0) = \theta_1 = I'(T_1),\ \gamma_1 = \theta_1'(0),\ \Gamma_1 = \theta_1''(0),\ \theta_2(0) = 0,\ \gamma_2 = \theta_2'(0),\ \Gamma_2 = \theta_2''(0). \tag{6.22}$$

We denote the two terms in the expression for $s_\infty$ in (6.3) by $I_1, I_2$, i.e.,

$$s_\infty = \sup_{\tilde{h}\in\tilde{W}} \left[\vec{\theta}_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\right] - \sup_{\tilde{h}\in\tilde{W}^*} \left[\vec{\theta}_\infty \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\right] = I_1 - I_2, \tag{6.23}$$

and we let $s_\infty(\epsilon)$ denote the relative entropy corresponding to the perturbed constraints. We distinguish between the cases $T_1 \in [\frac{1}{2}, 1)$ and $T_1 \in (0, \frac{1}{2})$.

**Case I** $T_1 \in [\frac{1}{2}, 1)$**:** From [37, Section 5], if $T_1 \in [\frac{1}{2}, 1)$ and $T_2 \in [\frac{1}{8}, 1)$, then the corresponding Lagrange multipliers $(\theta_1, \theta_2)$ are both non-negative. Hence from [29, Theorem 4.1] we have that

$$I_1 := \sup_{\tilde{h}\in\tilde{W}} \left[\theta_1(\epsilon)T_1(\tilde{h}) + \theta_2(\epsilon)T_2(\tilde{h}) - I(\tilde{h})\right] = \sup_{0\le u\le 1} \left[\theta_1(\epsilon)u + \theta_2(\epsilon)u^2 - I(u)\right], \tag{6.24}$$

and, consequently,

$$I_1 = \sup_{0<u<1} \left[\theta_1(\epsilon)u + \theta_2(\epsilon)u^3 - I(u)\right] = \theta_1(\epsilon)u^*(\epsilon) + \theta_2(\epsilon)u^*(\epsilon)^3 - I(u^*(\epsilon)). \tag{6.25}$$

The optimiser $u^*(\epsilon)$ corresponding to the perturbed multipliers $\theta_1^*(\epsilon)$ and $\theta_2^*(\epsilon)$ is analytic in $\epsilon$, as shown in [87]. Therefore, a Taylor expansion around $\epsilon = 0$ gives

$$u^*(\epsilon) = T_1 + \delta\epsilon + \tfrac{1}{2}\Delta\epsilon^2 + O(\epsilon^3), \tag{6.26}$$

where $\delta = u^{*\prime}(0)$ and $\Delta = u^{*\prime\prime}(0)$. Hence $I_1$ can be written as

$$I_1 = \theta_1 T_1 - I(T_1) + (\gamma_1 T_1 + \gamma_2 T_1^3)\epsilon + O(\epsilon^2). \tag{6.27}$$

Moreover,

$$\begin{aligned} I_2 &= \big[\theta_1 + \gamma_1\epsilon + \tfrac{1}{2}\Gamma_1\epsilon^2 + O(\epsilon^3)\big]T_1 + \big[\gamma_2\epsilon + \tfrac{1}{2}\Gamma_2\epsilon^2 + O(\epsilon^3)\big](T_1^3 + 3T_1\epsilon) - \inf_{\tilde{h}\in\tilde{W}_\epsilon^*} I(\tilde{h}) \\ &= \theta_1 T_1 + \gamma_1 T_1\epsilon + \tfrac{1}{2}\Gamma_1 T_1\epsilon^2 + T_1^3\gamma_2\epsilon + \tfrac{1}{2}\Gamma_2 T_1^3\epsilon^2 + 3T_1\gamma_2\epsilon^2 - J^{\downarrow}(\epsilon) + O(\epsilon^3), \end{aligned} \tag{6.28}$$

where

$$J^{\downarrow}(\epsilon) := \inf_{\tilde{h}\in\tilde{W}_\epsilon^*} I(\tilde{h}), \qquad \tilde{W}_\epsilon^* := \{\tilde{h}\in\tilde{W}\colon\, T_1(\tilde{h}) = T_1,\ T_2(\tilde{h}) = T_1^3 + 3T_1\epsilon\}. \tag{6.29}$$

Consequently,

$$s_\infty(T_1^*, T_1^{*3} + 3T_1^*\epsilon) = J^{\downarrow}(\epsilon) - I(T_1) + O(\epsilon^2). \tag{6.30}$$

Denote by $\tilde{h}_\epsilon^{(2)}$ one of the, possibly multiple, balance optimisers of the variational problem $J^{\downarrow}(\epsilon)$. From Proposition 6.3.5 we know that, for $\epsilon$ sufficiently small, any graphon in the equivalence class $\tilde{h}_\epsilon^{(2)}$, denoted by $h_\epsilon^{(2)}$, has the form $h_\epsilon^{(2)} = T_1 + \sqrt{\epsilon}g^* + O(\epsilon)$, where the graphon $g^*$ was defined in (6.13). By considering the Taylor expansion of the function $I$ around $\epsilon = 0$, we get

$$\begin{aligned} I(h_\epsilon^{(2)}) &= I(T_1) + I'(T_1)\sqrt{\epsilon}\int_{[0,1]^2} \mathrm{d}x\,\mathrm{d}y\, g^*(x,y) \\ &\quad + \tfrac{1}{2}I''(T_1)\,\epsilon\int_{[0,1]^2} \mathrm{d}x\,\mathrm{d}y\, g^*(x,y)^2 + o(\epsilon) \\ &= I(T_1) + \tfrac{1}{2}I''(T_1)\,\epsilon\int_{[0,1]^2} \mathrm{d}x\,\mathrm{d}y\, g^*(x,y)^2 + o(\epsilon) \\ &= I(T_1) + I''(T_1)\epsilon + o(\epsilon) \\ &= I(T_1) + \frac{1}{2}\frac{1}{T_1(1-T_1)}\epsilon + o(\epsilon). \end{aligned} \tag{6.31}$$

But, from (6.14), a straightforward computation of the entropy of $h_\epsilon^*$ shows that

$$J^{\downarrow}(\epsilon) = I(T_1) + \frac{6}{1-2T_1^*}\log\frac{T_1^*}{1-T_1^*}\epsilon + o(\epsilon). \tag{6.32}$$

Hence we obtain that the global optimiser is not a balance optimiser and that

$$s_\infty(T_1^*, T_1^{*3} + 3T_1^*\epsilon) = \frac{6}{1-2T_1^*}\log\frac{T_1^*}{1-T_1^*}\,\epsilon + o(\epsilon). \tag{6.33}$$

**Case II** $T_1 \in (0, \frac{1}{2})$**:** Consider the term

$$I_1 := \sup_{\tilde{h}\in\tilde{W}} \Big[\theta_1(\epsilon)T_1(\tilde{h}) + \theta_2(\epsilon)T_2(\tilde{h}) - I(\tilde{h})\Big],$$

as above. If Assumption 1 applies, then this case is proved in the same way as Case I. Otherwise, consider the straightforward lower bound

$$\sup_{\tilde{h}\in\tilde{W}}\Big[\theta_1(\epsilon)T_1(\tilde{h})+\theta_2(\epsilon)T_2(\tilde{h})-I(\tilde{h})\Big]\geq\sup_{0\leq u\leq 1}\big[\theta_1(\epsilon)u+\theta_2(\epsilon)u^3-I(u)\big]\,. \tag{6.34}$$

The arguments used in Case I after (6.25) apply, and the result in (6.30) is obtained with an inequality instead of an equality.

## §6.4.2 Proof of Theorem 6.3.2

In this section we omit the computations that are similar to those in the proof of Theorem 6.3.1. Let

$$T_1(\epsilon)=T_1,\qquad T_2(\epsilon)=T_1^3-T_1^3\;\epsilon. \tag{6.35}$$

The factor $T_1^3$ appearing in front of the $\epsilon$ is put in for convenience in the computations. The perturbed Lagrange multipliers are

$$\theta_1(\epsilon)=\theta_1+\gamma_1\epsilon+\tfrac{1}{2}\Gamma_1\epsilon^2+O(\epsilon^3),\qquad \theta_2(\epsilon)=\gamma_2\epsilon+\tfrac{1}{2}\Gamma_2\epsilon^2+O(\epsilon^3), \tag{6.36}$$

where

$$\theta_1=I'(T_1),\qquad \gamma_1=\theta_1'(0),\qquad \Gamma_1=\theta_1''(0)\qquad \gamma_2=\theta_2'(0),\qquad \Gamma_2=\theta_2''(0). \tag{6.37}$$

We denote the two terms in the expression for $s_\infty$ in (6.3) by $I_1, I_2$, i.e., $s_\infty = I_1 - I_2$, and let $s_\infty(\epsilon)$ denote the perturbed relative entropy. The computations for $I_1$ are similar as before, because the exact form of the constraint does not affect the expansions in (6.26) and (6.27). For $I_2$, on the other hand, we have

$$\begin{aligned} I_2&=\theta_1T_1+\gamma_1T_1\epsilon+\tfrac{1}{2}\Gamma_1T_1\epsilon^2+T_1^3\gamma_2\epsilon+\tfrac{1}{2}\Gamma_2T_1^3\epsilon^2-T_1^3\gamma_2\epsilon^2-J_1^{\uparrow}(\epsilon)\\ &=\theta_1T_1+\gamma_1T_1\epsilon+T_1^3\gamma_2\epsilon-J_1^{\uparrow}(\epsilon)+O(\epsilon^2),\end{aligned} \tag{6.38}$$

where

$$J_1^{\uparrow}(\epsilon):=\inf_{\tilde{h}\in\tilde{W}_\epsilon^*}I(\tilde{h}),\qquad \tilde{W}_\epsilon^*:=\{\tilde{h}\in\tilde{W}\colon\,T_1(\tilde{h})=T_1,\;T_2(\tilde{h})=T_1^3-T_1^3\;\epsilon\}. \tag{6.39}$$

Consequently,

$$s_\infty(T_1^*,T_1^*-T_1^{*3}\epsilon)=J_1^{\uparrow}(\epsilon)-I(T_1)+O(\epsilon^2). \tag{6.40}$$

Denote by $\tilde{h}_\epsilon^*$ one of the, possibly multiple, optimisers of the variational problem $J_1^{\uparrow}(\epsilon)$. From Proposition 6.3.6 we know that, for $T_1^*\in(0,\frac{1}{2}]$, a balance optimal graphon in the equivalence class $\tilde{h}_\epsilon^*$, denoted by $h_\epsilon^*$ for simplicity in the notation, has the form

$$h_\epsilon^*=T_1^*+\epsilon^{1/3}g^*+O(\epsilon^{1/3}) \tag{6.41}$$

with $g^*$ given by

$$g^*(x,y)=\begin{cases}-T_1^*, & (x,y)\in[0,\frac{1}{2}]^2,\\ \;\;T_1^*, & (x,y)\in[0,\frac{1}{2}]\times(\frac{1}{2},1]\cup(\frac{1}{2},1]\times[0,\frac{1}{2}],\\ -T_1^*, & (x,y)\in(\frac{1}{2},1]^2.\end{cases} \tag{6.42}$$

Hence

$$J_1^{\uparrow}(\epsilon) \leq I(T_1) + \frac{1}{2}T_1^{*2}I''(T_1)\epsilon^{2/3} \leq I(T_1) + \frac{1}{4}\frac{T_1^*}{1-T_1^*}\epsilon^{2/3}, \tag{6.43}$$

which gives

$$s_\infty(T_1^*, T_1^* - T_1^{*3}\epsilon) \leq \frac{1}{4}\frac{T_1^*}{1-T_1^*}\epsilon^{2/3} + o(\epsilon^{2/3}). \tag{6.44}$$

## §6.4.3 Proof of Theorem 6.3.3

The computations leading to the expression for the relative entropy in the right-hand side of (6.10) are similar as those in Section 6.4.2, and we omit them. Hence we have

$$s_\infty(T_1^*, T_1^* - T_1^{*3}\epsilon) = J_2^{\uparrow}(\epsilon) - I(T_1) + O(\epsilon^2), \tag{6.45}$$

where, for $T_1 \in (\frac{1}{2}, 1)$,

$$J_2^{\uparrow}(\epsilon) := \inf_{\tilde{h}\in\tilde{W}_\epsilon^*} I(\tilde{h}), \qquad \tilde{W}_\epsilon^* := \{\tilde{h} \in \tilde{W} : T_1(\tilde{h}) = T_1,\ T_2(\tilde{h}) = T_1^3 - T_1^3\ \epsilon\}. \tag{6.46}$$

Denote by $\tilde{h}_\epsilon^*$ one of the, possibly multiple, optimisers of the variational problem $J_2^{\uparrow}(\epsilon)$. From Proposition 6.3.7 we know that, for $T_1 \in (\frac{1}{2}, 1)$, a balance optimal graphon, in the equivalence class $\tilde{h}_\epsilon^*$, denoted by $h_\epsilon^*$ for simplicity in the notation, has the form

$$h_\epsilon^* = T_1^* + g_\epsilon^* \tag{6.47}$$

with $g_\epsilon^*$ given by

$$g_\epsilon^*(x,y) := \begin{cases} \frac{T_1^{*2}}{\bar{T}_1^*}\epsilon^{2/3}, & (x,y) \in [0, 1 - \frac{T_1^*}{|\bar{T}_1^*|}\epsilon^{1/3}]^2 \\ T_1^*\epsilon^{1/3}, & (x,y) \in [0, 1 - \frac{T_1^*}{|\bar{T}_1^*|}\epsilon^{1/3}] \times [1 - \frac{T_1^*}{|\bar{T}_1^*|}\epsilon^{1/3}, 1] \ \text{ or} \\ & (x,y) \in [1 - \frac{T_1^*}{|\bar{T}_1^*|}\epsilon^{1/3}, 1] \times [0, 1 - \frac{T_1^*}{|\bar{T}_1^*|}\epsilon^{1/3}], \\ \bar{T}_1^*, & (x,y) \in [1 - \frac{T_1^*}{|\bar{T}_1^*|}\epsilon^{1/3}, 1]^2. \end{cases} \tag{6.48}$$

The term $\bar{T}_1^* \in (-T_1^*, 0)$ is defined in Theorem 6.3.3. Hence we have

$$s_\infty(T_1^*, T_1^* - T_1^{*3}\epsilon) \leq f(T_1^*, \bar{T}_1^*)\epsilon^{2/3} + o(\epsilon^{2/3}), \tag{6.49}$$

where $\bar{T}_1^* \in (-T_1^*, 0)$ is the unique point where the global minimum of the function $x \to f(T_1^*, x)$ defined by

$$f(T_1^*, x) := T_1^{*2}\frac{I(T_1^* + x) - I(T_1^*) - I'(T_1^*)x}{x^2}, \quad x \in (-T_1, 0). \tag{6.50}$$

We need to show that, for every $T_1^* \in (0,1)$ and for every $x \in (-T_1, 0)$, $f(T_1, x) > 0$ or equivalently that

$$I(T_1^* + x) - I(T_1^*) - I'(T_1^*)x > 0. \tag{6.51}$$

From the mean-value theorem we have that there exists $\xi \in (T_1^* + x, T_1^*)$ such that $I'(T_1^* + x) - I(T_1^*) = I'(\xi)x$. Hence we have that

$$f(T_1^*, x) = (I'(\xi) - I'(T_1^*))x > 0, \tag{6.52}$$

which follows because $I'$ is an increasing function, $x \in (-T_1, 0)$ and $\xi \in (T_1^* + x, T_1^*)$. More detailed arguments are provided in the following section.

## §6.5 Proofs of Propositions 6.3.5–6.3.7

In this section we prove Propositions 6.3.5–6.3.7. In Section 6.5.1 we prove Proposition 6.3.5 and in Section 6.5.2 we prove Propositions 6.3.6 - 6.3.7. The proof of Proposition 6.3.7 is similar to the proof of Proposition 6.3.6, only the computations are different. In Section 6.4 the following variational problems were encountered:

(1) For $T_1 \in (0,1)$,

$$J^{\downarrow}(\epsilon) = \inf\big\{I(\tilde{h})\colon\ \tilde{h} \in \tilde{W},\, T_1(\tilde{h}) = T_1,\, T_2(\tilde{h}) = T_1^3 + 3T_1\epsilon\big\}. \tag{6.53}$$

(2) For $T_1 \in (0, \frac{1}{2}]$,

$$J_1^{\uparrow}(\epsilon) = \inf\big\{I(\tilde{h})\colon\ \tilde{h} \in \tilde{W},\, T_1(\tilde{h}) = T_1,\, T_2(\tilde{h}) = T_1^3 - T_1^3\epsilon\big\}. \tag{6.54}$$

(3) For $T_1 \in (\frac{1}{2}, 1)$,

$$J_2^{\uparrow}(\epsilon) = \inf\big\{I(\tilde{h})\colon\ \tilde{h} \in \tilde{W},\, T_1(\tilde{h}) = T_1,\, T_2(\tilde{h}) = T_1^3 - T_1^3\epsilon\big\}. \tag{6.55}$$

In order to prove Propositions 6.3.5–6.3.7, we need to analyse these three variational problems, for $\epsilon$ sufficiently small, which is the objective of this section. The variational formula in (6.53) has been rigorously analysed in [66], and hence we study the variational formulas in (6.54) and (6.55), under the assumption that the optimiser lies in the class of balance optimal graphons. We remind the reader that we suppose Assumption 2 to be true. We analyse the variational formulas with the help of a perturbation argument. In particular, we show that the balance optimal perturbations are those given in (6.12), (6.16) and (6.18), respectively. The results in Propositions 6.3.6–6.3.7 follow directly from the following two lemmas.

**6.5.1 Lemma.** *Let $T_1 \in (0, \frac{1}{2}]$. For $\epsilon > 0$ consider the variational formula for $J_1^{\uparrow}(\epsilon)$ given in* (6.54)*. Then, for $\epsilon$ sufficiently small,*

$$J_1^{\uparrow}(\epsilon) \leq I(T_1) + \frac{1}{4}\frac{T_1}{1-T_1}\epsilon^{2/3} + o(\epsilon^{2/3}). \tag{6.56}$$

**6.5.2 Lemma.** *Let $T_1 \in (\frac{1}{2}, 1)$. For $\epsilon > 0$ consider the variational formula for $J_2^{\uparrow}(\epsilon)$ given in* (6.55)*. Then, for $\epsilon$ sufficiently small,*

$$J_2^{\uparrow}(\epsilon) \leq I(T_1) + f(T_1, \bar{T}_1^*)\epsilon^{2/3} + o(\epsilon^{2/3}), \tag{6.57}$$

*where $f(T_1, x)$, $x \in (-T_1, 0)$, and $\bar{T}_1^*$ were defined in Theorem 6.3.3.*

**6.5.3 Remark.** As argued in Remark 6.3.4, we believe, and there is numerical evidence in [63], that the results in (6.56) and (6.57) hold with equality.

In what follows we use the notation $f(\epsilon) \asymp g(\epsilon)$, for two functions $f, g$, when $\frac{f(\epsilon)}{g(\epsilon)}$ converges to a *positive* constant, as $\epsilon \downarrow 0$.

## §6.5.1 Proof of Proposition 6.3.5

In this section we prove Proposition 6.3.5 given that Assumption 2 holds. In order to find the optimal perturbation when the ER-line is approached from above, we need to solve $J^{\downarrow}(\epsilon)$ in (6.53). The following construction shows intuitively why balance optimal perturbations have the form given in (6.12). Consider an inhomogeneous ER-random graph on $n$ vertices. We split the vertices of the graph into two parts of equal size, i.e., of size $n/2$. In one part we connect two vertices with probability $T_1+2\sqrt{\epsilon}$, in the other part we connect two vertices with probability $T_1-2\sqrt{\epsilon}$, and we connect vertices lying in different parts with probability $T_1$. This graph has expected edge density equal to

$$\frac{1}{\binom{n}{2}}\left(T_1\left(\frac{n}{2}\right)^2+(T_1+2\sqrt{\epsilon})\binom{\frac{n}{2}}{2}+(T_1-2\sqrt{\epsilon})\binom{\frac{n}{2}}{2}\right)=T_1. \tag{6.58}$$

Similarly, the expexted triangle density is equal to

$$\begin{aligned}&\frac{1}{\binom{n}{3}}\left(\binom{\frac{n}{2}}{3}(T_1+2\sqrt{\epsilon})^3+\frac{n}{2}\binom{\frac{n}{2}}{2}2T_1^2T_1+\binom{\frac{n}{2}}{3}(T_1-2\sqrt{\epsilon})^3\right)\\&=T_1^3+3T_1\frac{n-4}{n-1}\epsilon\\&\sim T_1^3+3T_1\epsilon,\end{aligned}$$

for $n$ large. Below when we speak of optimal perturbation we mean balance optimal. In the proof below we will see that the optimal perturbation is indeed given by the graphon counterpart of the inhomogeneous ER-random graph described above. We now proceed to the technical details of the proof.

With a slight abuse of notation we write $I(\cdot)$ for both cases of a graphon and a real number. We consider the variational formula $J^{\downarrow}(\epsilon)$, with $\epsilon>0$ given in (6.53). We denote by $\tilde{h}_\epsilon^{*\downarrow}$ one of the, possibly multiple, optimisers of $J^{\downarrow}(\epsilon)$. For simplicity in the notation, in what follows we work with a representative element, denoted by $h_\epsilon^{*\downarrow}$, of the equivalence class $\tilde{h}_\epsilon^{*\downarrow}$. We write the optimiser $h_\epsilon^{*\downarrow}$ in the form $h_\epsilon^{*\downarrow}=T_1+\Delta H_\epsilon$ for some bounded symmetric function $\Delta H_\epsilon$ defined on the unit square $[0,1]^2$ and taking values in $\mathbb{R}$. This term will be called the *perturbation term*. The optimiser $h_\epsilon^{*\downarrow}$ has to satisfy the conditions on the edge and triangle densities, i.e.,

$$T_1(h_\epsilon^{*\downarrow})=T_1,\qquad T_2(h_\epsilon^{*\downarrow})=T_1^3+3T_1\epsilon. \tag{6.59}$$

Hence the perturbation term $\Delta H_\epsilon$ needs to satisfy the constraints

$$(G_1)\colon\quad \int_{[0,1]^2}\mathrm{d}x\,\mathrm{d}y\ \Delta H_\epsilon(x,y)=0 \tag{6.60}$$

and

$$\begin{aligned}(G_2)\colon\ 3T_1&\int_{[0,1]^3}\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\ \ \Delta H_\epsilon(x,y)\Delta H_\epsilon(y,z)\\&+\int_{[0,1]^3}\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\ \ \Delta H_\epsilon(x,y)\Delta H_\epsilon(y,z)\Delta H_\epsilon(z,x)=3T_1\epsilon.\end{aligned} \tag{6.61}$$

In what follows we prove the result stated in Proposition 6.3.5, i.e., the optimal perturbation is a three-step function and is of order $\sqrt{\epsilon}$.

In Assumption 2 it is stated that it suffices to restrict to graphons that can be written in the form $T_1+\Delta H_\epsilon^{(2)}$, where $\Delta H_\epsilon^{(2)}$ is a bounded symmetric function defined on $[0,1]^2$, taking three non-zero values. In what follows, for simplicity in the computations and without loss of generality, we suppose that the optimal graphon has the form

$$\Delta H_\epsilon^{(2)} = g_{11}1_{I\times I} + g_{12}1_{(I\times J)\cup(J\times I)} + g_{22}1_{J\times J}. \tag{6.62}$$

Then $(G_1)$ above becomes

$$\lambda(I)^2 g_{11} + 2\lambda(I)(1-\lambda(I))g_{12} + (1-\lambda(I))^2 g_{22} = 0 \tag{6.63}$$

and the two integrals in $(G_2)$ become

$$\begin{aligned}\int_{[0,1]^3} \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\ \Delta H_\epsilon(x,y)\Delta H_\epsilon(y,z) = {}& \lambda(I)^3 g_{11}^2 + 2\lambda(I)^2(1-\lambda(I))g_{11}g_{12} \\ & + 2\lambda(I)(1-\lambda(I))^2 g_{12}g_{22} + \lambda(I)(1-\lambda(I))g_{12}^2 \\ & + (1-\lambda(I))^2 g_{22}^2,\end{aligned} \tag{6.64}$$

and

$$\begin{aligned}\int_{[0,1]^3} \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\ \Delta H_\epsilon(x,y)\Delta H_\epsilon(y,z) = {}& \lambda(I)^3 g_{11}^2 + 2\lambda(I)^2(1-\lambda(I))g_{11}g_{12} \\ & + 2\lambda(I)(1-\lambda(I))^2 g_{12}g_{22} + \lambda(I)(1-\lambda(I))g_{12}^2 \\ & + (1-\lambda(I))^2 g_{22}^2,\end{aligned} \tag{6.65}$$

and a similar expression can be computed for the second integral in $(G_2)$. We now give the formal definition of a balance optimal graphon:

**6.5.4 Definition.** For $T_1 \in (0,1)$, a graphon $T_1 + \tilde{h}_\epsilon$, $\epsilon > 0$, is called balanced if it has the structure given in (6.62) and the terms $\lambda(I)^2 g_{11}, \lambda(I)(1-\lambda(I))g_{12}$ and $(1-\lambda(I))^2 g_{22}$ are all of the same order when $\epsilon$ is sufficiently small.

**6.5.5 Definition.** For $\epsilon > 0$ a graphon $\tilde{h}_\epsilon$ is called balance optimal if it solves the following optimisation problem:

$$J_{bal}(\epsilon) := \inf\{I(\tilde{h}), \tilde{h} \in \tilde{W}, \tilde{h} \text{ is balanced, } T_1(\tilde{h}) = T_1, T_2(\tilde{h}) = T_1^3 + 3T_1\epsilon\}. \tag{6.66}$$

It is straightforward to observe that, for $\epsilon > 0$,

$$J_{bal}(\epsilon) \geq J(\epsilon). \tag{6.67}$$

In what follows we essentially determine $J_{bal}(\epsilon)$ for $\epsilon$ sufficiently small. We distinguish two cases, first $g_{12} = 0$ and then $g_{12} \neq 0$.

**Case $g_{12} = 0$:** The values of $g_+$ and $g_-$ are such so that $T_1 + \Delta H_\epsilon^{(2)}$ satisfies the conditions in (6.60) and (6.61). We proceed with the condition in (6.61). A standard computation yields

$$\int_{[0,1]^3} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\ \Delta H_\epsilon^{(2)}(x,y)\Delta H_\epsilon^{(2)}(y,z) = \lambda(I)^3\, g_+^2 + \lambda(J)^3\, g_-^2 \tag{6.68}$$

and

$$\int_{[0,1]^3} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\ \Delta H_\epsilon^{(2)}(x,y)\Delta H_\epsilon^{(2)}(y,z)\Delta H_\epsilon^{(2)}(z,x) = \lambda(I)^3\, g_+^3 + \lambda(J)^3\, g_-^3. \tag{6.69}$$

From (6.60) we obtain the first order condition

$$\lambda(I)^2 g_+ + \lambda(J)^2 g_- = 0. \tag{6.70}$$

Using the condition in (6.70), we get that (6.61) equals

$$g_-^2\, 3T_1 \frac{\lambda(J)^3}{\lambda(I)}(\lambda(J)+\lambda(I)) - g_-^3 \frac{\lambda(J)^3}{\lambda(I)^3}(\lambda(I)^3 - \lambda(J)^3) = 3T_1\epsilon + o(\epsilon). \tag{6.71}$$

There are multiple ways in which the condition in (6.71) can be met. We show that the lowest possible value of $I$ is attained when $g_+ \asymp \sqrt{\epsilon}, g_- \asymp -\sqrt{\epsilon}$ and $\lambda(I), \lambda(J)$ are constant. To that end we distinguish the following cases:

(I)

$$g_-^2\, 3T_1 \frac{\lambda(J)^3}{\lambda(I)}(\lambda(J)+\lambda(I)) \asymp \epsilon, \qquad g_-^3 \frac{\lambda(J)^3}{\lambda(I)^3}(\lambda(I)^3 - \lambda(J)^3) = o(\epsilon), \tag{6.72}$$

which splits into three sub-cases:

(Ia)

$$g_+ \asymp \epsilon^{1/2}, \quad g_- \asymp -\epsilon^{1/2}, \quad \frac{\lambda(J)}{\lambda(I)} \asymp 1. \tag{6.73}$$

(Ib)

$$g_+ \asymp \epsilon^{1/2+\delta/3}, \quad g_- \asymp -\epsilon^{1/2-\delta}, \quad \frac{\lambda(J)^3}{\lambda(I)} \asymp \epsilon^{2\delta}, \quad \delta \in (0, \tfrac{1}{2}). \tag{6.74}$$

(Ic)

$$g_+ \asymp \epsilon^{1/2-3\delta}, \quad g_- \asymp -\epsilon^{1/2+\delta}, \quad \frac{\lambda(J)^3}{\lambda(I)} \asymp \epsilon^{-2\delta}, \quad \delta \in (0, \tfrac{1}{6}). \tag{6.75}$$

(1d)

$$g_+ \asymp \epsilon^{2/3}, \quad g_- = \bar{g} \in (-T_1, 0), \quad \lambda(J) \asymp \epsilon^{1/3}. \tag{6.76}$$

(II)

$$g_-^2\, 3T_1 \frac{\lambda(J)^3}{\lambda(I)}(\lambda(J)+\lambda(I)) \asymp \epsilon^{1+\delta}, \quad -g_- \frac{1}{\lambda(I)^2} \asymp \epsilon^{-\delta}, \quad \delta > 0. \tag{6.77}$$

A simple calculation shows that, in all the cases above, $\lambda(I) + \lambda(J) \asymp 1$ and $\lambda(I)^3 - \lambda(J)^3 \asymp 1$, and hence we can omit these two factors from the analysis below. In what follows we exclude cases (Ib), (Ic) and (II) one by one by comparing them to graphons of the type given in case (Ia).

**Case (Ib):** We show that, for $\epsilon > 0$ sufficiently small, graphons having the structure indicated in (Ia) yield smaller values of the function $I$ than graphons with the structure in (Ib). We consider two graphons, denoted by $T_1 + g^*$ and $T_1 + \hat{g}^*$, where $g^*$ is as in Case (Ia) and $\hat{g}^*$ is as in Case (Ib). Before giving the technical details of the proof, we present a heuristic argument why $I(T_1 + g^*) < I(T_1 + \hat{g}^*)$. In what follows we will denote by $B(p)$ a Bernoulli random variable with parameter $p$. The function $-I(x)$, $x \in [0,1]$, defined in (5.19) represents the entropy of a $B(x)$ random variable with parameter $x$. On the graphon space the function $-I(h)$, $h \in W$, defined in (5.20), can be seen as the expectation of the entropy of a Bernoulli random variable with a random parameter (the expectation is with respect to the random parameter), i.e., $B(h(X,Y))$ with $(X,Y)$ a uniformly distributed random variable on $[0,1]^2$. For $h \in W$ we have

$$-I(h) = \int_{[0,1]^2} \mathrm{d}x\, \mathrm{d}y\, [-I(h(x,y))] = \mathbb{E}[-I(h(X,Y))]. \tag{6.78}$$

Hence we have the following equivalence

$$I(T_1 + g^*) < I(T_1 + \hat{g}^*) \Leftrightarrow \mathbb{E}[-I(T_1 + g^*(X,Y))] > \mathbb{E}[-I(T_1 + \hat{g}^*(X,Y))], \tag{6.79}$$

where $(X,Y)$ is a uniformly distributed random vector on $[0,1]^2$. Instead of working with entropy, it is intuitively simpler to work with the relative entropy with respect to the random variable $B(\frac{1}{2})$. The relative entropy is defined by

$$I_{\frac{1}{2}}(x) := x \log \frac{x}{\frac{1}{2}} + (1-x) \log \frac{1-x}{\frac{1}{2}}, \qquad x \in [0,1]. \tag{6.80}$$

Note that

$$\begin{aligned} &\mathbb{E}[-I(T_1 + g^*(X,Y))] > \mathbb{E}[-I(T_1 + \hat{g}^*(X,Y))] \Leftrightarrow \\ &\mathbb{E}[I_{\frac{1}{2}}(T_1 + g^*(X,Y))] < \mathbb{E}[I_{\frac{1}{2}}(T_1 + \hat{g}^*(X,Y))]. \end{aligned} \tag{6.81}$$

We first give an intuitive argument and afterwards prove that

$$\mathbb{E}[I_{\frac{1}{2}}(T_1 + g^*(X,Y))] < \mathbb{E}[I_{\frac{1}{2}}(T_1 + \hat{g}^*(X,Y))]. \tag{6.82}$$

We distinguish between the cases $T_1 \in (0, \frac{1}{2}]$ and $T_1 \in (\frac{1}{2}, 1)$. The case $T_1 \in (0, \frac{1}{2}]$ follows by using similar arguments as in case $T_1 \in (\frac{1}{2}, 1)$. We treat in detail only the case $T_1 \in (\frac{1}{2}, 1)$.

The relative entropy of a random variable with respect to $B(\frac{1}{2})$ is zero if and only if that random variable is equal to $B(\frac{1}{2})$. So, in order to compare the relative entropies in (6.82), we need to see how close the Bernoulli random variables with random parameters $T_1 + g^*(X,Y)$ and $T_1 + \hat{g}^*(X,Y)$ are to $B(\frac{1}{2})$. We are considering the case $T_1 > \frac{1}{2}$. Hence the random variables $B(T_1 + g^*(X,Y))$ and $B(T_1 + \hat{g}^*(X,Y))$ will be close to $B(\frac{1}{2})$ when the random parameters $T_1 + g^*(X,Y)$ and $T_1 + \hat{g}^*(X,Y)$ are close to $\frac{1}{2}$. This is the case when $g^*(X,Y)$ and $\hat{g}^*(X,Y)$ are negative. These events occur with probabilities

$$\mathbb{P}(T_1 + g^*(X,Y) < T_1) = \mathbb{P}(g^*(X,Y) < 0) = \mathbb{P}(g^*(X,Y) = g_-) = \lambda(J)^2 \asymp 1, \tag{6.83}$$

because of the properties of the graphon in Case (Ia). Similarly, we have that

$$\mathbb{P}(T_1 + \hat{g}^*(X,Y) < T_1) = \mathbb{P}(\hat{g}^*(X,Y) < 0) = \mathbb{P}(\hat{g}^*(X,Y) = g_-) = \lambda(\hat{J})^2 \asymp \epsilon^{4\delta/3}, \tag{6.84}$$

for some $\delta \in (0, \frac{1}{2}]$, because of the properties of the graphon in Case (Ib). Hence we see that the random variable $B(T_1 + g^*(X,Y))$ is closer to the random variable $B(\frac{1}{2})$ with much higher probability than the random variable $B(T_1 + \hat{g}^*(X,Y))$. We can see this by computing the corresponding expectations,

$$\mathbb{E}(g^*(X,Y) \mid g^*(X,Y) = g_-)\, \mathbb{P}(g^*(X,Y) = g_-) = g_-\, \mathbb{P}(g^*(X,Y) = g_-) \asymp \epsilon^{1/2}, \tag{6.85}$$

while

$$\begin{aligned}\mathbb{E}(\hat{g}^*(X,Y) \mid \hat{g}^*(X,Y) = \hat{g}_-)\, \mathbb{P}(\hat{g}^*(X,Y) = \hat{g}_-) &= \hat{g}_-\, \mathbb{P}(\hat{g}^*(X,Y) = \hat{g}_-) \\ &\asymp \epsilon^{1/2-\delta}\epsilon^{4\delta/3} = \epsilon^{1/2+\delta/3}.\end{aligned}$$

In what follows we complete this argument by adding the technical details. We work out the expressions in the left-hand and right-hand sides of (6.82). The expression in the right-hand side of (6.82) can be written as

$$\mathbb{E}[I_{\frac{1}{2}}(T_1 + g^*(X,Y))] = L I_{\frac{1}{2}}(T_1 + g_+) + K I_{\frac{1}{2}}(T_1 + g_-) + (1 - L - K) I_{\frac{1}{2}}(T_1), \tag{6.86}$$

for some constants $L := \mathbb{P}(g^*(X,Y) = g_+)$ and $K = \mathbb{P}(g^*(X,Y) = g_-)$ independent of $\epsilon$. Similarly,

$$\mathbb{E}[I_{\frac{1}{2}}(T_1 + \hat{g}^*(X,Y))] = \lambda(\hat{I})^2 I_{\frac{1}{2}}(T_1 + \hat{g}_+) + \epsilon^{4\delta/3} I_{\frac{1}{2}}(T_1 + \hat{g}_-) + (1 - \lambda(\hat{I})^2 - \epsilon^{4\delta/3}) I_{\frac{1}{2}}(T_1), \tag{6.87}$$

where $\lambda(\hat{I})^2 = \mathbb{P}(\hat{g}^*(X,Y) = \hat{g}_+) \asymp 1$ and $\mathbb{P}(\hat{g}^*(X,Y) = \hat{g}_-) \asymp \epsilon^{4\delta/3}$. Moreover, we recall that from the properties of the graphons in Case (Ia) and Case (Ib) we get

$$g_+ \asymp \sqrt{\epsilon}, \quad g_- \asymp -\sqrt{\epsilon}, \quad \hat{g}_+ \asymp \epsilon^{1/2+\delta/3}, \quad \hat{g}_- \asymp \epsilon^{1/2-\delta}, \quad \delta \in (0, \tfrac{1}{2}]. \tag{6.88}$$

Hence, for $T_1 \in (\frac{1}{2}, 1]$ and $\epsilon$ sufficiently small, because of (6.88), we obtain the following inequalities:

$$I_{\frac{1}{2}}(T_1 + g_+) > I_{\frac{1}{2}}(T_1 + \hat{g}_+) > I_{\frac{1}{2}}(T_1 + g_-) > I_{\frac{1}{2}}(T_1 + \hat{g}_-). \tag{6.89}$$

Using a Taylor expansion of the function $I$ around $T_1$ and the first order conditions

$$L g_+ + K g_- = 0 \quad \text{and} \quad \lambda(\hat{I})^2 \hat{g}_+ + \lambda(\hat{J})^2 \hat{g}_- = 0, \tag{6.90}$$

we observe that (6.86) and (6.87) are read

$$\mathbb{E}[I_{\frac{1}{2}}(T_1 + g^*(X,Y))] = I_{\frac{1}{2}}(T_1) + \tfrac{1}{2} I''_{\frac{1}{2}}(T_1)(L g_+^2 + K g_-^2) + o\left(g_+^2 + g_-^2\right) \tag{6.91}$$

and

$$\begin{aligned}\mathbb{E}[I_{\frac{1}{2}}(T_1 + \hat{g}^*(X,Y))] = I_{\frac{1}{2}}(T_1) &+ \tfrac{1}{2} I''_{\frac{1}{2}}(T_1)(\lambda(\hat{I}^2)\hat{g}_+^2 + \lambda(\hat{J})^2 \hat{g}_-^2) \\ &+ o\left(\lambda(\hat{I}^2)\hat{g}_+^2 + \lambda(\hat{J})^2 \hat{g}_-^2\right).\end{aligned} \tag{6.92}$$

Using (6.88), we observe that $Lg_+^2 + Kg_-^2 \asymp \epsilon$ and

$$\lambda(\hat{I}^2)\hat{g}_+^2 + \lambda(\hat{J})^2\hat{g}_-^2 \asymp \epsilon^{1+2\delta/3} + \epsilon^{4/3\delta}\epsilon^{1-2\delta} \asymp \epsilon^{1-2\delta/3}. \tag{6.93}$$

Hence, for $\epsilon$ sufficiently small,

$$\mathbb{E}\left[I_{\frac{1}{2}}(T_1 + g^*(X,Y))\right] < \mathbb{E}\left[I_{\frac{1}{2}}(T_1 + \hat{g}^*(X,Y))\right], \tag{6.94}$$

which proves (6.82).

Similar arguments can be used for the case $T_1 \in (0, \frac{1}{2})$ to show that graphons, as in Case (Ic), yield larger values of $I(\cdot)$ for $\epsilon$ sufficiently small. We omit the details.

**Case (1d):** In this case we have that the optimal graphon is constant on a subset of the unit square with a size tending to zero as $\epsilon \downarrow 0$. Such a graphon yields

$$\begin{aligned} I(T_1 + g^*) &= \lambda(I)^2 I(T_1 + g_+) + 2(1-\lambda(I))(1-\lambda(J))I(T_1) + \lambda(J)^2 I(T_1 + g_-) \\ &= \lambda(I)^2(I(T_1) + I'(T_1)g_+ + o(\epsilon^{2/3})) + 2(1-\lambda(I))(1-\lambda(J))I(T_1) \\ &\quad + \lambda(J)^2 I(T_1 + g_-) \\ &= I(T_1) - \lambda(J)^2 I(T_1) - \lambda(J)^2 \bar{g} I'(T_1) + \lambda(J)^2 I(T_1 + \bar{g}) \\ &= I(T_1) + \epsilon^{2/3}\left(I(T_1 + \bar{g}) - \bar{g}I'(T_1) - I(T_1)\right) + o(\epsilon^{2/3}). \end{aligned} \tag{6.95}$$

The second equality follows by considering a Taylor expansion around $\epsilon = 0$ in the terms that go to zero as $\epsilon \downarrow 0$, i.e, $g_+$. In the third equality we use (6.70). What remains is to show that

$$I(T_1 + \bar{g}) - I(T_1) - \bar{g}I'(T_1) > 0 \tag{6.96}$$

for $\bar{g} \in (-T_1, 0)$. From the mean-value theorem we have that $I(T_1+\bar{g}) - I(T_1) = I'(\xi)\bar{g}$ for some $\xi \in (T_1+\bar{g}, T_1)$. Since $\bar{g} < 0$ and $I$ is a convex function, i.e., $I'$ is an increasing function, we have that $I'(\xi) < I'(T_1)$. This proves the claim above. From (6.95) we observe that graphons having the form as in Case (1d) yield larger values of $I$, for $\epsilon$ sufficiently small, than graphons as in Case (1a).

**Case (II):** This case is simpler to exclude than the ones above. Indeed, suppose that (6.77) holds. Then either $\lambda(I)$ should become small or $-g_-$ should become large. But $g_- \asymp -\epsilon^{-\delta}$ is not possible because $g_-$ should stay bounded in $(-T_1, 0)$ as $\epsilon \downarrow 0$. Hence the only possibility is $\lambda(I) \asymp \epsilon^{\eta}$ and $g_- \asymp -\epsilon^{\zeta}$ for some $\eta, \zeta$ such that $\zeta - 2\eta = -\delta$, because of the second condition in (6.77). From the first condition in (6.77) we have that $2\zeta - \eta = 1 + \delta$. Solving these two equations we obtain that $\eta = \frac{1}{3} + \delta$ and $\zeta = \frac{2}{3} + \delta$. From (6.70) we then get that $g_+ \asymp \epsilon^{-\delta}$, which is not possible because $g_+$ should stay bounded in $(0, 1 - T_1)$ as $\epsilon \downarrow 0$.

At this point we summarise our findings. We considered the variational formula for $J^{\downarrow}(\epsilon)$ given in (6.53) and we assumed that we can restrict ourselves to piece-wise constant graphons (see Assumption 2) subject to the constraints in (6.60) and (6.61).

Afterwards, without loss of generality, we restricted ourselves to an even smaller class of graphons, those of the form

$$g = g_+ 1_{I\times I} + g_- 1_{J\times J} \tag{6.97}$$

for some $g_+ > 0$, $g_- < 0$ and $I, J \subset [0,1]$ with $\lambda(I)^2 + \lambda(J)^2 \leq 1$. At the end of this section we elaborate on the case $g_{12} \neq 0$. More specifically, we have shown that the optimal perturbation satisfies $g_+ \asymp \epsilon^{1/2}$, $g_- \asymp -\epsilon^{1/2}$ and $\lambda(I) \asymp 1$, $\lambda(J) \asymp 1$. Hence the solution to $J^{\downarrow}(\epsilon)$ has the form $T_1 + g^*\sqrt{\epsilon} + o(\epsilon)$, where $g^* = g_+ 1_{L\times L} + g_- 1_{K\times K}$ for some $g_+ > 0, g_- < 0, L, K \in (0,1)$ independent of $\epsilon$, is a symmetric function defined on $[0,1]^2$. From the constraints (6.60) and (6.61) we have that $g_+ L^2 = -g_- K^2$ and $L^3 g_+^2 + K^3 g_-^2 = 1$. A simple calculation shows that

$$\begin{aligned} I(T_1 + g\sqrt{\epsilon}) &= I(T_1) + I'(T_1)(L^2 g_+ + K^2 g_-)\sqrt{\epsilon} + \tfrac{1}{2} I''(T_1)(L^2 g_+^2 + K^2 g_-^2)\epsilon + o(\epsilon) \\ &= I(T_1) + \tfrac{1}{2} I''(T_1)(L^2 g_+^2 + K^2 g_-^2)\epsilon + o(\epsilon). \end{aligned}$$

Hence, in order to find the optimal graphon we need to solve the following optimisation problem:

$$\begin{aligned} &\min \left(L^2 g_+^2 + K^2 g_-^2\right) \\ &\text{such that } L + K \leq 1, \quad g_+ L^2 + g_- K^2 = 0, \quad L^3 g_+^2 + K^3 g_-^2 = 1. \end{aligned} \tag{6.98}$$

This is equivalent to

$$\begin{aligned} &\min \left(\frac{1}{K} + \frac{1}{L} - \frac{2}{K+L}\right) \\ &\text{such that } L + K \leq 1. \end{aligned} \tag{6.99}$$

From a standard computation we find that the optimal $K, L$ should satisfy $K + L = 1$. Hence we need to minimize $\frac{1-2L+L^2}{L(1-L)}$. This function is convex in $L \in (0,1)$ and attains a unique minimum at the point $L = \frac{1}{2}$. Having computed $L, K$ we find $g_+ = -g_- = 2$, and so the optimal solution to $J^{\downarrow}(\epsilon)$, for $\epsilon$ sufficiently small, is the graphon

$$h_\epsilon^{*\downarrow}(x,y) = \begin{cases} T_1 + 2\sqrt{\epsilon}, & \text{if } (x,y) \in [0,\tfrac{1}{2}]^2, \\ T_1, & \text{if } (x,y) \in [0,\tfrac{1}{2}] \times (\tfrac{1}{2},1] \text{ or } (\tfrac{1}{2},1] \times [0,\tfrac{1}{2}], \\ T_1 - 2\sqrt{\epsilon}, & \text{if } (x,y) \in (\tfrac{1}{2},1]^2. \end{cases} \tag{6.100}$$

A standard computation shows that $T_1(h_\epsilon^{*\downarrow}) = T_1$ and $T_2(h_\epsilon^{*\downarrow}) = T_1^3 + 3T_1\epsilon$.

**Case** $g_{12} \neq 0$**:** By following similar arguments as for the case $g_{12} = 0$, we can show that the optimal values of $g_{11}, g_{12}, g_{22}$, $K$ and $L$ can be retrieved by solving the following optimisation problem:

$$\begin{aligned} &\min \left(L^2 g_{11}^2 + K^2 g_{22}^2 + 2LK g_{12}^2\right) \\ &\text{such that} \\ &\qquad L + K = 1, \\ &\qquad L^2 g_{11} + K^2 g_{22} + 2LK g_{12} = 0, \\ &\qquad L^3 g_{11}^2 + K^3 g_{22}^2 + 2L^2 K g_{12} g_{11} + 2LK^2 g_{12} g_{22} + LK g_{12}^2 = 1. \end{aligned} \tag{6.101}$$

Suppose first that $L = K = \frac{1}{2}$. Then we have the following optimisation problem:

$$\begin{aligned}&\min \tfrac{1}{4}\left(g_{11}^2 + g_{22}^2 + 2g_{12}^2\right)\\&\text{such that}\\&\qquad g_{11} + g_{22} + 2g_{12} = 0,\\&\qquad g_{11}^2 + g_{22}^2 + 2g_{12}g_{11} + 2g_{12}g_{22} + 2g_{12}^2 = 8.\end{aligned}$$

Introducing Lagrange multipliers, we obtain the solution $g_{12} = 0$ and $g_{11} = -g_{22} = 2$. For arbitrary $L, K$, substituting

$$g_{12} = -\frac{1}{2}\left(\frac{L}{1-L}g_{11} + \frac{1-L}{L}g_{22}\right)$$

into (6.101) and differentiating the Lagrangian with respect to $g_{12}$, we obtain $g_{12} = 0$. We observe at this point that this argument holds only for the case where $g_{11}, g_{12}$ and $g_{22}$ go to zero as $\epsilon \downarrow 0$. This is not the case for the actual optimal graphon in (6.14).

**Case $g_{12} \neq 0$ and $g_{22} = 0$:** From (6.96) we observe that $g_{22} = 0$ yields an equality. Hence in this case the microcanonical entropy will of order $\epsilon$ instead of $\epsilon^{2/3}$. From the first-order constraint in (6.60) we obtain

$$g_{12} = -\frac{1}{2}\frac{\lambda}{(1-\lambda)}g_{11}, \tag{6.102}$$

where $\lambda := \lambda(I)$. Then the second order constraint reads

$$g_{11}^2\frac{1}{4}\frac{\lambda^2}{(1-\lambda)^2}\lambda(1-\lambda) = \epsilon. \tag{6.103}$$

Following similar arguments as before, we can show that the case $g_{11} \asymp \epsilon^{\delta}$, $\lambda \asymp \epsilon^{1/3-\delta/3}$ $g_{12} \asymp -\epsilon^{2/3+\delta/2}$ is not optimal. The case where $g_{11}$ or $g_{12}$ are constant, independently of $\epsilon$, is also not optimal, since if one of them is constant then the entropy cost will be $\epsilon^{2/3}$ instead of $\epsilon$. A standard computation yields

$$I(T_1 + g^*) = I(T_1) + \frac{1}{2}I''(T_1)\left(2 + 4\frac{1-\lambda}{\lambda}\right)\epsilon + o(\epsilon), \tag{6.104}$$

while for the graphon defined in (6.100) we have

$$I(h_\epsilon^{*\downarrow}) = I(T_1) + I''(T_1)\epsilon + o(\epsilon). \tag{6.105}$$

Hence we see that $I(T_1 + g^*) > I(T_1 + h_\epsilon^{*\downarrow})$ if and only if $1 - \lambda$ is constant and independent of $\epsilon$. If $1 - \lambda \asymp \epsilon^{\delta}$, then further analysis is needed in order to establish the optimal graphon. In any case, the graphon $h_\epsilon^{*\downarrow}$ is balance optimal, as desired.

## §6.5.2 Proof of Lemma 6.5.1 and Lemma 6.5.2

In this section we provide the technical details leading to the optimal perturbation of the variational formula in (6.54). We denote one of the, possibly multiple, optimizers

of (6.54) by $\tilde{h}_\epsilon^{*\uparrow}$. In the proof, in order to keep the notation light, we denote a representative element of this class by $h_\epsilon^{*\uparrow}$. We start by writing the optimizer in the form $h_\epsilon^{*\uparrow} = T_1 + \Delta H_\epsilon$ for some perturbation term $\Delta H_\epsilon$. The perturbation term has to be a bounded symmetric function defined on the unit square $[0,1]^2$ taking values in $\mathbb{R}$. The optimizer $h_\epsilon^{*\uparrow}$ has to satisfy the constraints

$$T_1(h_\epsilon^{*\uparrow}) = T_1, \qquad T_2(h_\epsilon^{*\uparrow}) = T_1^3 - T_1^3\epsilon, \tag{6.106}$$

and so the perturbation $\Delta H_\epsilon$ needs to satisfy the two constraints

$$(K_1): \int_{[0,1]^2} \mathrm{d}x\,\mathrm{d}y\ \Delta H_\epsilon(x,y) = 0 \tag{6.107}$$

and

$$\begin{aligned}(K_2):\ 3T_1 &\int_{[0,1]^3} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\ \ \Delta H_\epsilon(x,y)\Delta H_\epsilon(y,z) \\ &+ \int_{[0,1]^3} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\ \ \Delta H_\epsilon(x,y)\Delta H_\epsilon(y,z)\Delta H_\epsilon(z,x) = -T_1^3\epsilon.\end{aligned} \tag{6.108}$$

Again, from Assumption 2, we restrict to graphons having the form $T_1 + \Delta H_\epsilon$ where

$$\Delta H_\epsilon = g_{11}1_{I\times I} + g_{12}1_{(I\times J)\cup(J\times I)} + g_{22}1_{J\times J}, \tag{6.109}$$

$g_{11}, g_{12}, g_{22} \in (-T_1, 1-T_1)$ and $I \subset [0,1]$, $J = I^c$. From (6.107) we get

$$\lambda(I)^2 g_{11} + 2\lambda(I)(1-\lambda(I))g_{12} + (1-\lambda(I))^2 g_{22} = 0, \tag{6.110}$$

which yields

$$g_{12} = -\frac{1}{2}\left(\frac{\lambda(I)}{1-\lambda(I)}g_{11} + \frac{1-\lambda(I)}{\lambda(I)}g_{22}\right). \tag{6.111}$$

A standard computation shows that the second order integral in (6.108) is equal to

$$\lambda(I)^3 g_{11}^2 + (1-\lambda(I))^3 g_{22}^2 + 2\lambda(I)(1-\lambda(I))g_{12}(\lambda(I)g_{11} + (1-\lambda(I))g_{22} + \frac{1}{2}g_{12}. \tag{6.112}$$

By (6.111) this is equal to

$$\frac{1}{4}\lambda(I)(1-\lambda(I))\left(\frac{\lambda(I)}{(1-\lambda)}g_{11} - \frac{1-\lambda(I)}{\lambda(I)}g_{22}\right)^2. \tag{6.113}$$

From (6.108) we observe that, for $\epsilon$ sufficiently small, the first integral will dominate the second integral when $g_{11}, g_{12}$ and $g_{22}$ depend on $\epsilon$. Hence, in order to obtain the condition in (6.108), it must be that

$$\int_{[0,1]^3} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\ \Delta H_\epsilon(x,y)\Delta H_\epsilon(y,z) = 0. \tag{6.114}$$

Then (6.113) yields

$$g_{11} = \frac{(1-\lambda(I))^2}{\lambda(I)^2}g_{22} \tag{6.115}$$

and from (6.111) also

$$g_{12} = -\frac{1-\lambda(I)}{\lambda(I)} g_{22}. \tag{6.116}$$

The third order integral in (6.108) then yields

$$g_{22}\frac{1-\lambda(I)}{\lambda(I)} = -T_1\epsilon^{1/3}. \tag{6.117}$$

We distinguish three cases,

(1)

$$g_{11} \asymp -\epsilon^{1/3}, \quad g_{12} \asymp \epsilon^{1/3} \quad g_{22} \asymp -\epsilon^{1/3}, \quad \frac{1-\lambda}{\lambda} \asymp 1, \tag{6.118}$$

(2)

$$g_{11} \asymp -\epsilon^{2/3-\delta} \ , g_{12} \asymp \epsilon^{1/3}, \ g_{22} \asymp -\epsilon^{\delta}, \ \frac{1-\lambda(I)}{\lambda(I)} \asymp \epsilon^{1/3-\delta}, \ \delta \in (0, \tfrac{1}{3}), \tag{6.119}$$

(3)

$$g_{11} \asymp -\epsilon^{2/3} \ , g_{12} \asymp \epsilon^{1/3}, \ g_{22} = \bar{g} \in (-T_1, 0), \ \frac{1-\lambda(I)}{\lambda(I)} \asymp \epsilon^{1/3}. \tag{6.120}$$

For each of the cases above we compute the value of the function $I$.

**Case (1):** For graphons as in Case 1, we have

$$\begin{aligned}
I(T_1 + \Delta H_\epsilon) =& \lambda(I)^2 I(T_1 + g_{11}) + 2\lambda(I)(1-\lambda(I))I(T_1 + g_{12}) \\
&+ (1-\lambda(I))^2 I(T_1 + g_{22}) \\
=& I(T_1) + \frac{1}{2}I''(T_1)\big(\lambda(I)^2 g_{11}^2 + 2\lambda(I)(1-\lambda(I))g_{12}^2 \\
&+ (1-\lambda(I))^2 g_{22}^2\big)\epsilon^{2/3} + o(\epsilon^{2/3}) \\
=& I(T_1) + \frac{1}{2}I''(T_1)\left(\frac{(1-\lambda(I))^4}{\lambda(I)^2} + 2\frac{(1-\lambda(I))^3}{\lambda(I)} + (1-\lambda(I))^2\right) g_{22}^2 \\
&+ o(\epsilon^{2/3}) \\
=& I(T_1) + \frac{1}{2}I''(T_1)\frac{(1-\lambda(I))^2}{\lambda(I)^2} g_{22}^2 + o(\epsilon^{2/3}) \\
=& I(T_1) + \frac{1}{2}I''(T_1)T_1^2\epsilon^{2/3} + o(\epsilon^{2/3}) \\
=& I(T_1) + \frac{1}{4}\frac{T_1}{1-T_1}\epsilon^{2/3} + o(\epsilon^{2/3}).
\end{aligned}$$

We observe that there exist multiple graphons that can yield this result. The only constraint we impose is $g_{22}\frac{1-\lambda(I)}{\lambda(I)} = -T_1\epsilon^{1/3}$. For example, the graphon $T_1 + g^*\epsilon^{1/3}$

with

$$g^*(x,y) = \begin{cases} -T_1, & (x,y) \in [0,\frac{1}{2}]^2, \\ T_1, & (x,y) \in [0,\frac{1}{2}] \times (\frac{1}{2},1] \cup (\frac{1}{2},1] \times [0,\frac{1}{2}], \\ -T_1, & (x,y) \in (\frac{1}{2},1]^2, \end{cases} \quad , \tag{6.121}$$

as in (6.17) is balance optimal. In Case (2) below we construct more graphons that are balance optimal.

**Case (2):** A similar computation as above shows that

$$\begin{aligned} I(T_1 + \Delta H_\epsilon) &= I(T_1) + \frac{1}{2} I''(T_1) T_1^2 \epsilon^{2/3} + o(\epsilon^{2/3}) \\ &= I(T_1) + \frac{1}{4} \frac{T_1}{1-T_1} \epsilon^{2/3} + o(\epsilon^{2/3}). \end{aligned} \tag{6.122}$$

From cases (1) and (2) we observe that various graphons are balance optimal. Hence we need to investigate the higher-order terms in order to determine the optimal graphon.

**Case (3):** For this case we have

$$\begin{aligned} I(T_1 + \Delta H_\epsilon) &= \lambda(I)^2 \big(I(T_1) + I'(T_1) g_{11}\big) + 2\lambda(I)(1-\lambda(I)) \big(I(T_1) + I'(T_1) g_{12}\big) \\ &\quad + (1-\lambda(I))^2 I(T_1 + \bar{g}) + o(\epsilon^{2/3}) \\ &= I(T_1) + (1-\lambda(I))^2 \left(-I'(T_1)\bar{g} - I(T_1) + I(T_1+\bar{g})\right) + o(\epsilon^{2/3}) \\ &= I(T_1) + T_1^2 \left( \frac{I(T_1+\bar{g}) - I(T_1) - I'(T_1)\bar{g}}{\bar{g}^2} \right) \epsilon^{2/3} + o(\epsilon^{2/3}). \end{aligned} \tag{6.123}$$

Therefore, in order to determine the optimal graphon, we need to find, for a given $T_1 \in (0,1)$, the minimum of the function

$$f(T_1, x) := T_1^2 \frac{I(T_1+x) - I(T_1) - I'(T_1)x}{x^2} \tag{6.124}$$

in $(-T_1, 0)$. We analyze this function for every $T_1 \in (0,1)$ as $x$ varies from $-T_1$ to 0. For $x = -T_1$ we have

$$f(T_1, -T_1) = -I(T_1) + T_1 I'(T_1) = -\frac{1}{2}\log(1-T_1), \tag{6.125}$$

while for $x \uparrow 0$ we have

$$\lim_{x\uparrow 0} f(T_1, x) = T_1^2 \lim_{x\uparrow 0} \frac{I'(T_1+x) - I'(T_1)}{2x} = \frac{1}{2} T_1^2 I''(T_1) = \frac{1}{4}\frac{T_1}{1-T_1}. \tag{6.126}$$

The first derivative is equal to

$$f'(T_1, x) = T_1^2 \frac{\left(I'(T_1+x) - I'(T_1)\right) x^2 - 2x \left(I(T_1+x) - I(T_1) - I'(T_1)x\right)}{x^4} \tag{6.127}$$

and at the boundary points we have

$$\lim_{x\downarrow -T_1} f'(T_1,x) = -\infty, \quad \lim_{x\uparrow 0} f'(T_1,x) = \frac{1}{6} I^{(3)}(T_1) = -\frac{1}{12}\frac{1-2T_1}{(T_1(1-T_1))^2}. \tag{6.128}$$

We observe that $I^{(3)}(T_1) > 0$ if and only if $T_1 > \frac{1}{2}$. Consider first the two endpoints

$$h_1(T_1) = -\frac{1}{2}\log(1-T_1), \quad h_2(T_1) = \frac{1}{4}\frac{T_1}{1-T_1} \tag{6.129}$$

and observe that

$$h_1'(0) = \frac{1}{2} > h_2'(0) = \frac{1}{4}, \quad h_1'(1-\epsilon) = \frac{1}{2\epsilon} < h_2'(\epsilon) = \frac{1}{4\epsilon^2}. \tag{6.130}$$

Both $h_1(\cdot)$ and $h_2(\cdot)$ are increasing function on $[0,1]$. Hence there is a unique $\bar{T}_1$ such that $h_1(T_1) > h_2(T_1)$ for all $T_1 \in (0,\bar{T}_1)$ and $h_1(T_1) \le h_2(T_1)$ for all $T_1 \in [\bar{T}_1, 1)$. Numerically we find $\bar{T}_1 \approx 0.715$ (see also Figure 6.3). We distinguish three cases: $T_1 \in (0, \frac{1}{2}]$, $T_1 \in (\frac{1}{2}, \bar{T}_1]$ and $T_1 \in (\bar{T}_1, 1)$. The results that follow are not rigorously proven, but are derived by using numerical approximations.

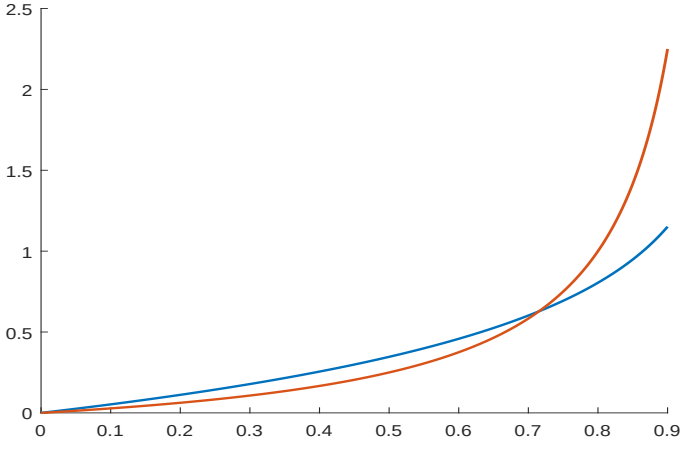



*Figure 6.3: Plot of the functions $h_1(\cdot)$ (blue line) and $h_2(\cdot)$ (red line).*

**Case $T_1 \in (0, \frac{1}{2}]$:** We have that $h_1(T_1) = f(T_1, -T_1) > h_2(T_1) = \lim_{x\uparrow 0} f(T_1, x)$. Moreover, $I^{(3)}(T_1) \le 0$, with equality at $T_1 = \frac{1}{2}$. Hence from (6.128) we have that $f(T_1,\cdot)$ decreases away from $f(T_1,-T_1)$ towards $\lim_{x\uparrow 0} f(T_1,x)$. We observe that it is also a decreasing function on $(-T_1, 0)$. Hence we have that

$$f(T_1,x) > \frac{1}{4}\frac{T_1}{1-T_1} \quad \forall x \in (-T_1, 0). \tag{6.131}$$

We illustrate this in Figure 6.4 and 6.5, where we plot $f(T_1,\cdot)$ for $T_1 = 0.1, 0.25, 0.4$ and $0.5$.

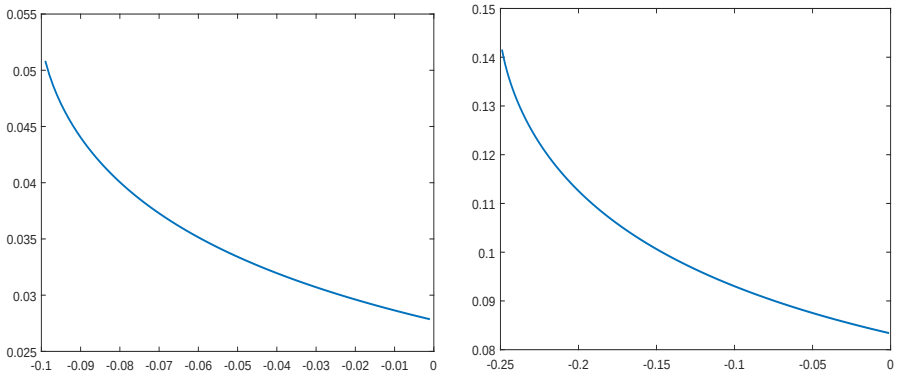



*Figure 6.4: Plot of the function $f(T_1, \cdot)$ for $T_1 = 0.1$ (left panel) and $T_1 = 0.25$ (right panel).*

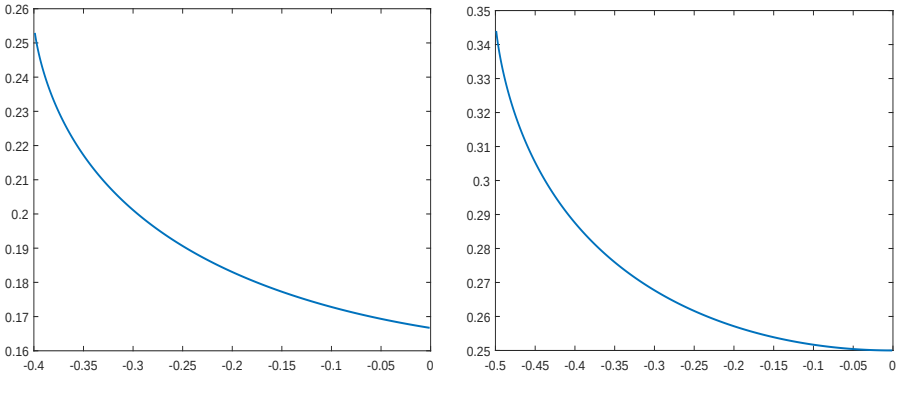



*Figure 6.5: Plot of the function $f(T_1, \cdot)$ for $T_1 = 0.4$ (left panel) and $T_1 = 0.5$ (right panel).*

**Case** $T_1 \in (\frac{1}{2}, \bar{T}_1]$**:** In this case we have that $I^{(3)}(T_1) > 0$, and so $f(T_1, x)$ increases into $\lim_{x\uparrow 0} f(T_1, x)$. A similar argument as above in Case 1 shows that there is at least one stationary point $\bar{T}_1^* \in (-T_1, 0)$, which is also a local minimum. Uniqueness of this local minimum is verified numerically, as depicted in Figure 6.6.

**Case** $T_1 \in (\bar{T}_1, 1)$**:** Using a similar reasoning as in Cases 1 and 2, we get that there is a stationary point $\bar{T}_1^* \in (-T_1, 0)$, which is a local minimum. Uniqueness of this minimum is verified numerically in in Figure 6.7.

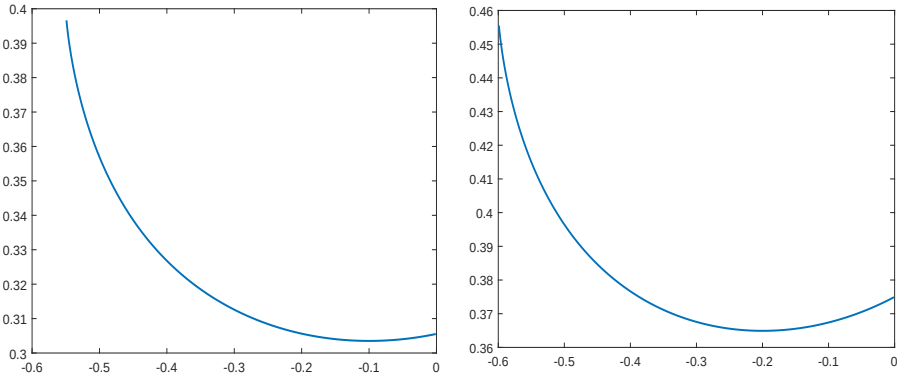



*Figure 6.6: Plot of the function* $f(T_1, \cdot)$ *for* $T_1 = 0.55$ *(left panel) and* $T_1 = 0.6$ *(right panel).*

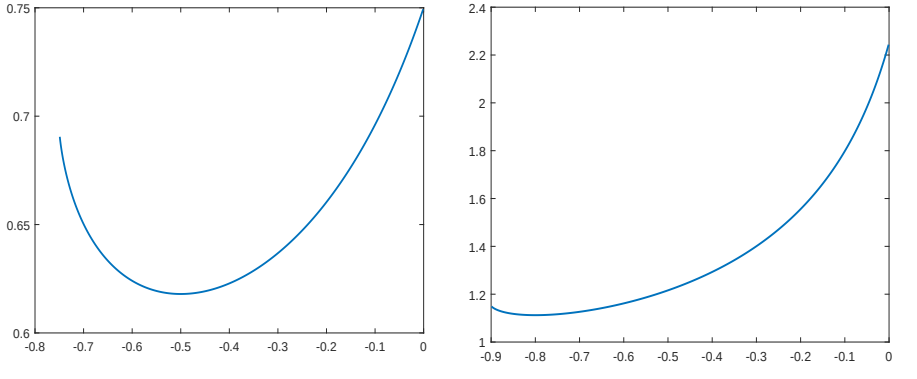



*Figure 6.7: Plot of the function* $f(T_1, \cdot)$ *for* $T_1 = 0.75$ *(left panel) and* $T_1 = 0.9$ *(right panel).*

From the three cases considered above we observe that if $T_1 \in (0, \frac{1}{2}]$, then

$$f(T_1, x) > \frac{1}{4}\frac{T_1}{1 - T_1}.$$

On the other hand, if $T_1 \in (\frac{1}{2}, 1)$, then the function $f(T_1, \cdot)$ attains a global minimum, denoted by $\bar{T}_1^* \in (-T_1, 0)$, and

$$f(T_1, \bar{T}_1^*) < \frac{1}{4}\frac{T_1}{1 - T_1}.$$

We finally return to (6.123). For $T_1 \in (0, \frac{1}{2}]$ the optimal graphon yields

$$I(T_1 + \Delta H_\epsilon) = I(T_1) + \frac{1}{4}\frac{T_1}{1-T_1}\epsilon^{2/3} + o(\epsilon^{2/3}). \tag{6.132}$$

For $T_1 \in (\frac{1}{2}, 1)$ the optimal graphon yields

$$I(T_1 + \Delta H_\epsilon) = I(T_1) + f(T_1, \bar{T}_1^*)\epsilon^{2/3} + o(\epsilon^{2/3}), \tag{6.133}$$

where $\bar{T}_1^* \in (-T_1, 0)$ is the unique minimizer of the function $f(T_1, \cdot)$ defined in (6.124).

# Bibliography

[1] K. Anand and G. Bianconi. Entropy measures for networks: Toward an information theory of complex topologies. *Phys. Rev. E*, 80:045102, Oct 2009.

[2] K. Anand and G. Bianconi. Gibbs entropy of network ensembles by cavity methods. *Phys. Rev. E*, 82:011116, Jul 2010.

[3] D. Aristoff and L. Zhu. Asymptotic structure and singularities in constrained directed graphs. *Stochastic Process. Appl.*, 125(11):4154–4177, 2015.

[4] D. Aristoff and L. Zhu. On the phase transition curve in a directed exponential random graph model. *Adv. in Appl. Probab.*, 50(1):272–301, 2018.

[5] R. Arratia and T. M. Liggett. How likely is an i.i.d. degree sequence to be graphical? *Ann. Appl. Probab.*, 15(1B):652–670, 2005.

[6] R. Balian. *From microphysics to macrophysics: methods and applications of statistical physics. Vol. II.* Theoretical and Mathematical Physics. Springer, Berlin, 2007.

[7] J. Barré and B. Gonçalves. Ensemble inequivalence in random graphs. *Phys. A*, 386(1):212–218, 2007.

[8] J. Barré, D. Mukamel, and S. Ruffo. Inequivalence of ensembles in a system with long-range interactions. *Phys. Rev. Lett.*, 87:030601, Jun 2001.

[9] A. Barvinok and J. A. Hartigan. The number of graphs and a random graph with a given degree sequence. *Random Struct. Algor.*, 42(3):301–348, 2013.

[10] E. A. Bender. The asymptotic number of non-negative integer matrices with given row and column sums. *Discrete Math.*, 10:217–223, 1974.

[11] E. A. Bender and E. R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Comb. Theory Ser. A*, 24(3):296–307, 1978.

[12] S. Bhamidi, G. Bresler, and A. Sly. Mixing time of exponential random graphs. *Ann. Appl. Probab.*, 21(6):2146–2170, 2011.

[13] G. Bianconi. Statistical mechanics of multiplex networks: Entropy and overlap. *Phys. Rev. E*, 87:062806, Jun 2013.

[14] G. Bianconi, A. C. C. Coolen, and C. J. Perez Vicente. Entropies of complex networks with hierarchically constrained topologies. *Phys. Rev. E (3)*, 78(1):016114, 11, 2008.

[15] M. Blume, V. J. Emery, and R. B. Griffiths. Ising model for the $\lambda$ transition and phase separation in $he^3$-$he^4$ mixtures. *Phys. Rev. A*, 4:1071–1077, Sep 1971.

[16] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Phys. Rep.*, 544(1):1–122, 2014.

[17] M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J. B*, 38(2):205–209, Mar 2004.

[18] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980.

[19] L. Boltzmann. Ueber die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik. *J. Reine Angew. Math.*, 100:201–212, 1887.

[20] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[21] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)*, 176(1):151–219, 2012.

[22] G. Caldarelli. *Scale-Free Networks: Complex Webs in Nature and Technology.* Oxford University Press, 2007.

[23] A. Campa, T. Dauxois, D. Fanelli, and S. Ruffo. *Physics of long-range interacting systems.* Oxford University Press, Oxford, 2014.

[24] A. Campa, T. Dauxois, and S. Ruffo. Statistical mechanics and dynamics of solvable models with long-range interactions. *Phys. Rep.*, 480(3-6):57–159, 2009.

[25] P. J. Carrington, J. Scott, and S. Wasserman. *Models and Methods in Social Network Analysis*, volume 28. Cambridge University Press, 2005.

[26] S. Chatterjee. An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc. (N.S.)*, 53(4):617–642, 2016.

[27] S. Chatterjee. *Large deviations for random graphs*, volume 2197 of *Lecture Notes in Mathematics.* Springer, Cham, 2017.

[28] S. Chatterjee and A. Dembo. Nonlinear large deviations. *Adv. Math.*, 299:396 – 450, 2016.

[29] S. Chatterjee and P. Diaconis. Estimating and understanding exponential random graph models. *Ann. Statist.*, 41(5):2428–2461, 2013.

[30] S. Chatterjee, P. Diaconis, and A. Sly. Random graphs with a given degree sequence. *Ann. Appl. Probab.*, 21(4):1400–1435, 2011.

[31] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the erdős-rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[32] Chavanis, P. H. Gravitational instability of isothermal and polytropic spheres. *A&A*, 401(1):15–42, 2003.

[33] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, 99(25):15879–15882, 2002.

[34] G. D'Agostino and A. Scala. *Networks of Networks: The Last Frontier of Complexity*, volume 340. Springer, 2014.

[35] M. D'Agostino, F. Gulminelli, P. Chomaz, M. Bruno, F. Cannata, R. Bougault, F. Gramegna, I. Iori, N. L. Neindre, G. Margagliotti, A. Moroni, and G. Vannini. Negative heat capacity in the critical region of nuclear fragmentation: an experimental evidence of the liquid-gas phase transition. *Phys. Lett. B*, 473(3):219 – 225, 2000.

[36] A. Dembo and E. Lubetzky. A large deviation principle for the erdős-rényi uniform random graph. *arXiv:1804.11327*.

[37] F. den Hollander, M. Mandjes, A. Roccaverde, and N. Starreveld. Breaking of ensemble equivalence for perturbed erdős-rényi random graphs. *arXiv:1807.07750*.

[38] F. den Hollander, M. Mandjes, A. Roccaverde, and N. J. Starreveld. Ensemble equivalence for dense graphs. *Electron. J. Probab.*, 23:Paper No. 12, 26, 2018.

[39] P. Diao, D. Guillot, A. Khare, and B. Rajaratnam. Differential calculus on graphon space. *J. Combin. Theory Ser. A*, 133:183–227, 2015.

[40] R. S. Ellis, K. Haven, and B. Turkington. Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles. *J. Statist. Phys.*, 101(5-6):999–1064, 2000.

[41] R. S. Ellis, K. Haven, and B. Turkington. Nonequivalent statistical equilibrium ensembles and refined stability theorems for most probable flows. *Nonlinearity*, 15(2):239–255, 2002.

[42] P. Erdős, D. J. Kleitman, and B. L. Rothschild. Asymptotic enumeration of $K_n$-free graphs. pages 19–27. Atti dei Convegni Lincei, No. 17, 1976.

[43] S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486(3-5):75–174, 2010.

[44] P. Fronczak, A. Fronczak, and M. Bujok. Exponential random graph models for networks with community structure. *Phys. Rev. E*, 88:032810, Sep 2013.

[45] D. Garlaschelli. The weighted random graph model. *New J. Phys.*, 11(7):073005, 2009.

[46] D. Garlaschelli, S. E. Ahnert, T. M. A. Fink, and G. Caldarelli. Low-temperature behaviour of social and economic networks. *Entropy*, 15(8):3148–3169, 2013.

[47] D. Garlaschelli, F. den Hollander, and A. Roccaverde. Complexe netwerken vanuit fysisch perspectief. *Nieuw Arch. Wisk.*, 5/16(3):207–209, 2015.

[48] D. Garlaschelli, F. den Hollander, and A. Roccaverde. Ensemble nonequivalence in random graphs with modular structure. *J. Phys. A*, 50(1):015001, 35, 2017.

[49] D. Garlaschelli, F. den Hollander, and A. Roccaverde. Statistical physics. `http://www.networkpages.nl/statistical-physics/`, 2017.

[50] D. Garlaschelli, F. den Hollander, and A. Roccaverde. Covariance structure behind breaking of ensemble equivalence in random graphs. *J. Stat. Phys.*, Jul 2018.

[51] D. Garlaschelli and M. I. Loffredo. Maximum likelihood: Extracting unbiased information from complex networks. *Phys. Rev. E*, 78:015101, Jul 2008.

[52] V. Gemmetto and D. Garlaschelli. Multiplexity versus correlation: the role of local constraints in real multiplexes. *Nature*, 5:9120, Mar 2015.

[53] J. W. Gibbs. *Elementary Principles of Statistical Mechanics*. Yale University Press, New Haven, Connecticut, 1902.

[54] J. W. Gibbs. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundation of thermodynamics*. Dover publications, Inc., New York, 1960.

[55] C. Greenhill, B. D. McKay, and X. Wang. Asymptotic enumeration of sparse 0-1 matrices with irregular row and column sums. *J. Combin. Theory Ser. A*, 113(2):291–324, 2006.

[56] P. Hertel and W. Thirring. A soluble model for a system with negative specific heat. *Ann. Phys.*, 63(2):520 – 533, 1971.

[57] P. W. Holland, K. B. Laskey, S. Leinhardt, and and. Stochastic blockmodels: first steps. *Soc. Networks*, 5(2):109–137, 1983.

[58] P. Holme and J. Saramäki. Temporal networks. *Phys. Rep.*, 519(3):97 – 125, 2012.

[59] K. Huang. *Statistical mechanics*. John Wiley & Sons, Inc., New York, second edition, 1987.

[60] I. Ipsen and D. Lee. Determinant approximations. *arXiv:1105.0437*.

[61] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev. (2)*, 106:620–630, 1957.

[62] B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E (3)*, 83(1):016107, 10, 2011.

[63] R. Kenyon, C. Radin, K. Ren, and L. Sadun. Multipodal structure and phase transitions in large constrained graphs. *J. Stat. Phys.*, 168(2):233–258, Jul 2017.

[64] R. Kenyon, C. Radin, K. Ren, and L. Sadun. The phases of large networks with edge and triangle constraints. *J. Phys. A*, 50(43):435001, 22, 2017.

[65] R. Kenyon, C. Radin, K. Ren, and L. Sadun. The phases of large networks with edge and triangle constraints. *J. Phys. A*, 50(43):435001, 2017.

[66] R. Kenyon, C. Radin, K. Ren, and L. Sadun. Bipodal structure in oversaturated random graphs. *Int. Math. Res. Notices*, 2018(4):1009–1044, 2018.

[67] R. Kenyon and M. Yin. On the asymptotics of constrained exponential random graphs. *J. Appl. Probab.*, 54(1):165–180, 2017.

[68] L. D. Landau and E. M. Lifshitz. *Course of theoretical physics. Vol. 5: Statistical physics.* Pergamon Press, Oxford-Edinburgh-New York, 1968.

[69] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[70] E. Lubetzky and Y. Zhao. On replica symmetry of large deviations in random graphs. *Rand. Struct. Algor.*, 47(1):109–146, 2015.

[71] E. Lubetzky and Y. Zhao. On the variational problem for upper tails in sparse random graphs. *Rand. Struct. Algor.*, 50(3):420–436, 2017.

[72] D. Lynden-Bell. Negative specific heat in astronomy, physics and chemistry. *Phys. A*, 263(1):293 – 304, 1999.

[73] D. Lynden-Bell, R. Wood, and A. Royal. The gravo-thermal catastrophe in isothermal spheres and the onset of red-giant structure for stellar systems. *Mon. Not. R. Astron. Soc.*, 138(4):495–525, 1968.

[74] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli. Enhanced reconstruction of weighted networks from strengths and degrees. *New J. Phys.*, 16(4):043022, 2014.

[75] R. Mavi and M. Yin. Ground states for exponential random graphs. *J. Math. Phys.*, 59(1):013303, 14, 2018.

[76] B. D. McKay and N. C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11(4):369–382, Dec 1991.

[77] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Struct. Algor.*, 6(2-3):161–179, 1995.

[78] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, Jul 2001.

[79] V. Nicosia, G. Bianconi, V. Latora, and M. Barthelemy. Growing multiplex networks. *Phys. Rev. Lett.*, 111:058701, Jul 2013.

[80] J. Park and M. E. J. Newman. Solution of the two-star model of a network. *Phys. Rev. E (3)*, 70(6):066146, 5, 2004.

[81] J. Park and M. E. J. Newman. Statistical mechanics of networks. *Phys. Rev. E*, 70:066117, Dec 2004.

[82] T. P. Peixoto. Entropy of stochastic blockmodel ensembles. *Phys. Rev. E*, 85:056122, May 2012.

[83] O. Pikhurko and A. Razborov. Asymptotic structure of graphs with the minimum number of triangles. *Combin. Probab. Comput.*, 26(1):138–160, 2017.

[84] M. A. Porter, J.-P. Onnela, and P. J. Mucha. Communities in networks. *Notices Amer. Math. Soc.*, 56(9):1082–1097, 2009.

[85] C. Radin and L. Sadun. Phase transitions in a complex network. *J. Phys. A*, 46(30):305002, 12, 2013.

[86] C. Radin and L. Sadun. Singularities in the entropy of asymptotically large simple graphs. *J. Stat. Phys.*, 158(4):853–865, 2015.

[87] C. Radin and M. Yin. Phase transitions in exponential random graphs. *Ann. Appl. Probab.*, 23(6):2458–2471, 2013.

[88] A. A. Razborov. On the minimal density of triangles in graphs. *Combin. Probab. Comput.*, 17(4):603–618, 2008.

[89] F. Reif. *Fundamentals of Statistical and Thermal Physics*. McGraw Hill, Tokyo, 1965.

[90] A. Roccaverde. Is breaking of ensemble equivalence monotone in the number of constraints? *Indagationes Mathematicae*, 2018.

[91] S. R. A. Salinas. *Introduction to statistical physics*. Graduate Texts in Contemporary Physics. Springer-Verlag, New York, 2001.

[92] T. Squartini, J. de Mol, F. den Hollander, and D. Garlaschelli. Breaking of ensemble equivalence in networks. *Phys. Rev. Lett.*, 115:268701, Dec 2015.

[93] T. Squartini and D. Garlaschelli. Reconnecting statistical physics and combinatorics beyond ensemble equivalence. *arXiv:1710.11422*.

[94] T. Squartini and D. Garlaschelli. Analytical maximum-likelihood method to detect patterns in real networks. *New Jour. Phys.*, 13(8):083001, 2011.

[95] T. Squartini, R. Mastrandrea, and D. Garlaschelli. Unbiased sampling of network ensembles. *New. J. Phys.*, 17(2):023052, 2015.

[96] W. Thirring. Systems with negative specific heat. *Z. Phys. A-Hadron. Nucl.*, 235(4):339–352, Aug 1970.

[97] H. Touchette. Equivalence and nonequivalence of ensembles: thermodynamic, macrostate, and measure levels. *J. Stat. Phys.*, 159(5):987–1016, 2015.

[98] H. Touchette, R. S. Ellis, and B. Turkington. An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles. *Phys. A*, 340(1-3):138–146, 2004. News and expectations in thermostatistics.

[99] R. van der Hofstad. *Random graphs and complex networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017.

[100] Y. H. Wang. On the number of successes in independent trials. *Statist. Sinica*, 3(2):295–312, 1993.

[101] M. Yin. Critical phenomena in exponential random graphs. *J. Stat. Phys.*, 153(6):1008–1021, 2013.

[102] M. Yin. Large deviations and exact asymptotics for constrained exponential random graphs. *Electron. Commun. Probab.*, 20:no. 56, 14, 2015.

[103] M. Yin and L. Zhu. Asymptotics for sparse exponential random graph models. *Braz. J. Probab. Stat.*, 31(2):394–412, 2017.

[104] X.-D. Zhang. The smallest eigenvalue for reversible Markov chains. *Linear Algebra Appl.*, 383:175–186, 2004.

[105] L. Zhu. Asymptotic structure of constrained exponential random graph models. *J. Stat. Phys.*, 166(6):1464–1482, 2017.

[106] R. M. Ziff, G. E. Uhlenbeck, and M. Kac. The ideal Bose-Einstein gas, revisited. *Phys. Rep.*, 32C(4):169–248, 1977.

# Samenvatting

Dit proefschrift presenteert nieuwe resultaten over het breken van ensemble equivalentie voor complexe netwerken met randcondities. We vergelijken het microkanonieke ensemble, waarbij de randcondities precies worden afgedwongen, met het kanonieke ensemble, waarbij de randcondities slechts gemiddeld worden afgedwongen. We geven voorbeelden van randcondities waarvoor beide ensembles niet equivalent zijn.

We bestuderen klassen van unipartite grafen in hoofdstuk 2 en laten zien dat het niet-equivalent zijn van de beide ensembles zich manifesteert via een nieuw mechanisme. Dit mechanisme is niet gerelateerd aan een faseovergang of niet-lineariteit, maar aan de aanwezigheid van een groot aantal locale topologische beperkingen.

Door een algemene klasse van random grafen met een variabel aantal randcondities te gebruiken, laten we zien dat een groot aantal randcondities de ensemble equivalentie breekt. We beginnen met de karakterisering van het breken van equivalentie in eenvoudige situaties met unipartite en bipartite grafen, zoals beschreven in [92]. Vervolgens beschouwen we een algemene klasse van grafen met een veellagenstructuur, waarbij de connectiviteit binnen de lagen en tussen de lagen kan worden aangepast.

De stellingen in hoofdstuk 2, die voornamelijk van toepassing zijn op het lagedichtheidsregime, karakteriseren niet alleen het breken equivalentie, maar geven ook een kwantitatieve formule voor de specifieke relatieve entropie. We onderzoeken belangrijke gevolgen van deze stellingen, zowel in de algemene context als in specifieke gevallen die van empirisch belang zijn. Bovendien geven we een interpretatie van de specifieke-relatieve-entropie formule in termen van Poissonisatie van de graden.

Het hogedichtheidsregime wordt bestudeerd in hoofdstuk 3, waar een formule voor de relatieve entropie gebaseerd op de covariantiestructuur van het kanonieke ensemble, zoals recent voorgesteld door Garlaschelli en Squartini [93], wordt bevestigd. We bekijken de situatie van een toevallige graaf met een gegeven gradenrij (een configuratie model) en laten zien dat de formule correct voorspelt dat de specifieke relatieve entropie wordt bepaald door de schaling van de determinant van de covariantie matrix van de randcondities in het zogenaamde $\delta$-tamme hogedichtheidregime, maar dat een extra correctieterm nodig is in het lagedichtheidregime en het ultrahogedichtheidsregime.

We laten zien dat het gedrag in de verschillende regimes correspondeert met het feit dat de graden asymptotisch Gaussisch zijn en asymptotisch Poisson in het lagedichtheidregime, en dat de duale graden asymptotisch Poisson zijn in het ultrahogedichtheidregime. We laten ook zien dat in het kanonieke ensemble de graden verdeeld zijn volgens een multivariate versie van de *Poisson-Binomiaal* verdeling [100], hetgeen een Gauss verdeling toelaat met de Poisson verdeling als de limiet in de corresponderende regimes.

We vervolgen onze studie van het configuratiemodel in hoofdstuk 4, waar een nieuwe vraag wordt beantwoord. Hoewel men kan laten zien dat extensiviteit van het aantal randcondities in het aantal knopen een cruciale rol speelt in het breken van ensemble equivalentie, is de vraag hoe het reduceren van het aantal beperkingen dit verschijnsel precies beïnvloedt nog open. In hoofdstuk 4 beantwoorden we deze vraag door het effect op de relatieve entropie van het reduceren van het aantal randcondities te bestuderen. We bekijken wat er gebeurt wanneer we slechts een deel van de knopen in hun graad beperken en de rest vrij laten.

Intuitief is de verwachting dat wanneer het aantal beperkingen afneemt de relatieve entropie ook afneemt. Dit is echter niet triviaal, omdat het aantal randcondities zowel het kanonieke als het microkanieke ensemble beïnvloedt. We bekijken toevallige grafen met een voorgeschreven partiële gradenrij (gereduceerde randconditie). Het breken van ensemble equivalentie wordt bestudeerd door te analyseren hoe de relatieve entropie verandert als een functie van het aantal randcondities. Meer precies laten we zien dat de relatieve entropie een monotone functie is van het aantal randcondities op *macroscopisch niveau*, d.w.z. wanneer een deel van de randcondities wordt weggelaten dat van de orde $n$ is, met $n$ het aantal knopen. Wanneer slecht $m$ knopen zijn beperkt en de resterende $n - m$ knopen niet, dan groeit de relatieve entropie als $m \log n$ voor $n \to \infty$.

Hoofdstukken 5 en 6 sluiten dit proefschrift af met een studie van dichte grafen met randcondities op deelgraafstructuren. In hoofdstuk 5 wordt het breken van ensemble equivalentie geanalyseerd door randcondities op te leggen op het aantal zijden en het aantal deelgrafen zoals hoeken, driehoeken, etc. Hier gebruiken we een grote afwijkingen principe voor graphons om te laten zien dat het breken van ensemble equivalentie ontstaat zodra de randcondities gefrustreerd zijn, d.w.z. met elkaar in conflict. Hoofdstuk 6 is een vervolg op hoofdstuk 5 voor het geval waar de randcondities worden opgelegd op het aantal zijden en driehoeken tegelijkertijd. In het bijzonder worden de randcondities zo gekozen dat ze de zogenaamde Erdős-Rényi-lijn benaderen, maar daar niet gelijk aan zijn. Het blijkt dat het gedrag van de relatieve entropie sterk verschilt tussen de situatie waarbij het totale aantal driehoeken groter is dan de derde macht van het totaal aantal zijden en de omgekeerde situatie.

# Acknowledgements

This thesis is the outcome of the research I did during the past four years. It would not exist without the help and support of many people that contributed to its realization.

I would like to express my sincere gratitude and thanks to my supervisors Frank den Hollander and Diego Garlaschelli, who I met on regular basis. They guided my journey by raising stimulating questions, by helping me to solve problems and by encouraging my research, allowing me to grow as a scientist. I really appreciated the scientific freedom and the precious advice they gave me both on doing research and on pursuing my career.

I am grateful to Michel Mandjes and Nicos Starreveld for the collaboration we had during the past two years. Together with them and Frank den Hollander, we had less frequent but long and intense discussions, which materialized into the two papers that led to Chapters 5 and 6.

I would like to thank the committee members for the time spent in reading the manuscript and providing feedback.

I want to thank the mathematics and physics departments at Leiden University, which funded my research through the Gravitation Program NETWORKS and a Leiden-Huygens fellowship, giving me the possibility to travel to many places in the Netherlands and abroad. I benefitted from the open-minded atmosphere and multicultural environment in Leiden and from the possibility to attend several summer schools and conferences. A sincere thanks goes to all the current and former members of the probability and dynamical systems group at the Mathematical Institute and the econophysics and network theory group at the Physics Institute for the precious and fruitful discussions we had in the past four years.

Besides science, the past four years consisted of many different distractions. I want to thank all those people I climbed, did sports, spent evenings and travelled with. Unfortunately, I cannot name them all, but I want to at least say a special thank to my two paranymphs, who were the first people I met in Leiden: Giulio and Marta.

Giulio, I remember the first day we met, when I had just arrived in Leiden and you immediately introduced me to all the people at the institute. Thanks for all the travels and unforgettable moments spent together.

Marta, you have been always present and of immense support. I want to thank you for all the wonderful travels, holidays, evenings, climbing sessions and for all the other beautiful moments shared together.

*Nonostante questi quattro anni di lontananza, mi siete sempre stati vicini. Grazie per gli innumerevoli e indimenticabili momenti passati assieme e per ricordarmi*

*sempre dove sono le mie radici: Eugenia, Francesco, Lorenzo, Maria Vittoria, Matteo, Nico, Silvia.*

Last but not least, I want to deeply thank my parents, who always offered me their never-ending support and a warm home.

# Curriculum Vitae

Andrea Roccaverde was born in Modena in 1990. He graduated in 2009 with a Scientific Degree from High School A.Tassoni in Modena. Afterwards he studied Mathematics at the University of Modena and Reggio Emilia. There he obtained a Bachelor of Science in Mathematics in 2012 and a Master of Science in Mathematics in 2014 (*cum laude*). He wrote his master thesis under the supervision of Dr. Emanuele Dolera, entitled *Probabilistic representations for solutions to various Boltzmann-like equations with a finite state space*. In December 2014 he moved to the University of Leiden to start his PhD research project under the supervision of Prof.dr. Frank den Hollander and Dr. Diego Garlaschelli. His research was funded by the mathematics and physics departments of the Science Faculty at Leiden University through a Leiden-Huygens fellowships, and was part of the Gravitation Program NETWORKS funded by the Dutch Ministry of Education, Culture and Science. He served as a teaching assistance in several courses. He presented his research at several scientific conferences in The Netherlands and India.