Ensemble nonequivalence in random graphs with modular structure

Diego Garlaschelli [1]
Frank den Hollander [2]
Andrea Roccaverde [12]

March 29, 2016

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 et al. (2015) it was shown that ensembles of random graphs are non-equivalent 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, block-model or community structure. We give a full classification of ensemble equivalence, proving that breakdown occurs if and only if 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 a formula for the specific relative entropy and provide an interpretation of this formula in terms of Poissonisation of the degrees.

MSC 2010. 60C05, 60K35, 82B20.

Key words and phrases. Random graph, community structure, multiplex network, multilayer network, stochastic blockmodel, constraints, microcanonical ensemble, canonical ensemble, relative entropy, equivalence vs. non-equivalence.

Acknowledgment. DG and AR are supported by EU-project 317532-MULTIPLEX. FdH and AR are supported by NWO Gravitation Grant 024.002.003–NETWORKS.

[1]Lorentz Institute for Theoretical Physics, Leiden University, P.O. Box 9504, 2300 RA Leiden, The Netherlands

[2]Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands

1 Introduction and main results

1.1 Background and outline

For systems with many interacting components a detailed microscopic description is unfeasible 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 vanishing of the fluctuations of the soft constraints, i.e., the soft constraints become asymptotically hard. This assumption of *ensemble equivalence*, which dates back to Gibbs [30], 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 [39], [52], [33], [38], [15], quantum phase separation [7], [3], [22], nuclear fragmentation [18], and fluid turbulence [20], [21]. 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 [22]; (ii) *macrostate equivalence* refers to the equivalence of the canonical and microcanonical sets of equilibrium values of macroscopic properties [54]; (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 [53]. The latter reference reviews the three definitions and shows that, under certain hypotheses, they are identical.

While there is consensus that ensemble 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 [12],

[13], [53]. 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 [2], 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 [47] and [48], 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 [51], [27] has shown that, surprisingly, ensemble nonequivalence can manifest itself via a different mechanism, totally unrelated to non-additivity or phase transitions. This previously unrecognised mechanism is the presence of an *extensive* number of local topological constraints, i.e., the degrees of all nodes. 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 [50]. 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 1.2 we give the definition of measure equivalence and show that it translates into a simple pointwise criterion for the large deviation properties of the microcanonical and canonical probabilities. In Section 1.3 we introduce our main theorems in pedagogical order, starting from the characterisation of nonequivalence in the simple cases of uni-partite and bi-partite graphs, and subsequently moving on to a very general class of graphs with arbitrary multilayer structure and tuneable intra-layer and inter-layer connectivity. Our main theorems not only characterise nonequivalence, they also provide a formula for the specific relative entropy. In Section 2 we discuss various important implications of our results, describing properties that are fully general bit also focussing on several special cases of empirical relevance. We confirm that, in all our models, ensemble nonequivalence arises from an extensive number of local topological constrains (in our setting, the degrees of a finite fraction of nodes) and is *not* related to the non-additivity of the system or to phase transitions. 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 a lot of interest in the literature, such as networks with a so-called multi-partite, multiplex [8], block-model [35], [37] or community structure [23], [46]. In Section 3, finally, we provide the proofs of our theorems.

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 subset of $\mathcal{G}_n$. 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 $n_1, \ldots, n_M$ 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, 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

$$g_{i,j}(\mathbf{G}) = \begin{cases} 1 & \text{if there is a link between node } i \text{ and node } j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

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

$$P_{\text{mic}}(\mathbf{G}) = \begin{cases} 1/\Omega_{\vec{C}^*}, & \text{if } \vec{C}(\mathbf{G}) = \vec{C}^*, \\ 0, & \text{else,} \end{cases} \tag{1.2}$$

where

$$\Omega_{\vec{C}^*} = |\{\mathbf{G} \in \mathcal{G}_n^\sharp \colon \vec{C}(\mathbf{G}) = \vec{C}^*\}| > 0 \tag{1.3}$$

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

$$S_n(P_{\text{can}}) = -\sum_{\mathbf{G} \in \mathcal{G}_n^\sharp} P_{\text{can}}(\mathbf{G}) \ln P_{\text{can}}(\mathbf{G}) \tag{1.4}$$

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

$$P_{\text{can}}(\mathbf{G}) = \frac{\exp[-H(\mathbf{G}, \vec{\theta}^*)]}{Z(\vec{\theta}^*)}, \tag{1.5}$$

where $H(\mathbf{G}, \vec{\theta}) = \vec{\theta} \cdot \vec{C}(\mathbf{G})$ is the *Hamiltonian* and

$$Z(\vec{\theta}) = \sum_{\mathbf{G} \in \mathcal{G}_n^\sharp} \exp[-H(\mathbf{G}, \vec{\theta})] \tag{1.6}$$

is the *partition function*. Note that in (1.5) 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 data [28].

It is worth mentioning that, in the social network analysis literature [14], 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 [44], [49], [50], [40], [25] ,[26], [37], [24], [45], [6], [1] 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. Apart for a few exceptions [1], [45], [51], 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

$$S_n(P_{\rm mic} \mid P_{\rm can}) = \sum_{\mathbf{G}\in\mathcal{G}_n^\sharp} P_{\rm mic}(\mathbf{G}) \ln \frac{P_{\rm mic}(\mathbf{G})}{P_{\rm can}(\mathbf{G})}, \tag{1.7}$$

Following [53], [51], we say that the two ensembles are measure equivalent if and only if the *specific relative entropy* is zero, i.e.,

$$s_\infty = \lim_{n\to\infty} n^{-1}\, S_n(P_{\rm mic} \mid P_{\rm can}) = 0. \tag{1.8}$$

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* $n \to \infty$, 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, (1.8) 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.

Before considering specific cases, we recall an important observation made in [51]. Noting from the form of $H(\mathbf{G}, \vec{\theta})$ that $P_{\rm can}(\mathbf{G}_1) = P_{\rm can}(\mathbf{G}_2)$ when $\vec{C}(\mathbf{G}_1) = \vec{C}(\mathbf{G}_2)$ (i.e., the canonical probability is the same for all graphs having the same value of the constraint), we may rewrite (1.7) as

$$S_n(P_{\rm mic} \mid P_{\rm can}) = \ln \frac{P_{\rm mic}(\mathbf{G}^*)}{P_{\rm can}(\mathbf{G}^*)}, \tag{1.9}$$

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 (1.8) then becomes

$$\lim_{n\to\infty} n^{-1} \left[\ln P_{\rm mic}(\mathbf{G}^*) - \ln P_{\rm can}(\mathbf{G}^*) \right] = 0, \tag{1.10}$$

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 (1.10), like (1.8), 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.

1.3 Main Theorems

1.3.1 Single layer: uni-partite graphs

The first class of (simple and undirected) random graphs we consider is specified by $M = 1$ and $\mathcal{G}_n^\sharp = \mathcal{G}_n$. This choice corresponds to the class of *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 (1.8) and (1.10) already well-defined.

Constraints on the degree sequence. An important constraint that we consider throughout the paper is the *degree sequence*. For a uni-partite graph $\mathbf{G} \in \mathcal{G}_n$, the degree sequence is defined as $\vec{k}(\mathbf{G}) = \big(k_i(\mathbf{G})\big)_{i=1}^n$ with $k_i(\mathbf{G}) = \sum_{j \neq i} g_{i,j}(\mathbf{G})$. We assume that the degree sequence is constrained to a specific (graphic) value $\vec{k}^*$. The constraints are therefore

$$\vec{C}^* = \vec{k}^* = (k_i^*)_{i=1}^n \in \mathbb{N}_0^n, \tag{1.11}$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ with $\mathbb{N} = \{1, 2, \ldots\}$. This class is also known as the *configuration model* ([5], [10], [42], [43], [16], [50]; see also [34, Chapter 7]). In [51] the breaking of ensemble equivalence was studied in the *sparse regime* defined by the condition

$$m^* = \max_{1 \leq i \leq n} k_i^* = o(\sqrt{n}). \tag{1.12}$$

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

$$f_n = n^{-1} \sum_{i=1}^n \delta_{k_i^*} \in \mathcal{P}(\mathbb{N}_0), \tag{1.13}$$

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

$$\lim_{n \to \infty} \|f_n - f\|_{\ell^1(g)} = 0, \tag{1.14}$$

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

$$g(k) = \log\left(\frac{k!}{k^k e^{-k}}\right), \qquad k \in \mathbb{N}_0, \tag{1.15}$$

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

Theorem 1.1. *Subject to* (1.11)–(1.12) *and* (1.14)*, the specific relative entropy equals*

$$s_\infty = \|f\|_{\ell^1(g)} > 0. \tag{1.16}$$

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

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

$$\vec{C}^* = L^*. \tag{1.17}$$

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 [51], in this case the usual result that the ensembles are asymptotically equivalent holds.

Theorem 1.2. *Subject to* (1.17)*, the specific relative entropy equals* $s_\infty = 0$.

1.3.2 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 Λ_1 and Λ_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 Λ_s to Λ_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 $L(\mathbf{G})$:

$$L(\mathbf{G}) = \sum_{i\in\Lambda_1} k_i(\mathbf{G}) = \sum_{j\in\Lambda_2} k'_j(\mathbf{G}). \tag{1.18}$$

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

$$\vec{C}^* = \{\vec{k}^*_{1\to2}, \vec{k}^*_{2\to1}\}. \tag{1.19}$$

We abbreviate

$$\begin{aligned} &m^* = \max_{i\in\Lambda_1} k^*_i, \quad m'^* = \max_{j\in\Lambda_2} k'^*_j, \\ &f^{(n_1)}_{1\to2} = {n_1}^{-1} \sum_{i\in\Lambda_1} \delta_{k^*_i}, \quad f^{(n_2)}_{2\to1} = {n_2}^{-1} \sum_{j\in\Lambda_2} \delta_{k'^*_j}, \end{aligned} \tag{1.20}$$

and assume the existence of

$$A_1 = \lim_{n_1,n_2\to\infty} \frac{n_1}{n_1+n_2}, \quad A_2 = \lim_{n_1,n_2\to\infty} \frac{n_2}{n_1+n_2}. \tag{1.21}$$

The *sparse regime* corresponds to

$$m^* m'^* = o(L^{*2/3}), \qquad n_1, n_2 \to \infty. \tag{1.22}$$

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

$$\lim_{n_1\to\infty} \|f^{(n_1)}_{1\to2} - f_{1\to2}\|_{\ell^1(g)} = 0, \quad \lim_{n_2\to\infty} \|f^{(n_2)}_{2\to1} - f_{2\to1}\|_{\ell^1(g)} = 0. \tag{1.23}$$

The relative entropy per node is

$$s_{n_1+n_2} = \frac{S_{n_1+n_2}(P_{\rm mic} \mid P_{\rm can})}{n_1+n_2}. \tag{1.24}$$

Theorem 1.3. *Subject to* (1.19) *and* (1.21)–(1.23),

$$s_\infty = \lim_{n_1,n_2\to\infty} \frac{S_{n_1+n_2}(P_{\rm mic} \mid P_{\rm can})}{n_1+n_2} = A_1 \, \|f_{1\to2}\|_{\ell^1(g)} + A_2 \, \|f_{2\to1}\|_{\ell^1(g)}. \tag{1.25}$$

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 $\vec{k}_{1\to 2}(\mathbf{G})$ to the value

$$\vec{C}^* = \vec{k}^*_{1\to 2} = \big(k_i^*\big)_{i\in\Lambda_1}, \tag{1.26}$$

while leaving $\vec{k}_{2\to 1}(\mathbf{G})$ unspecified (apart for the condition (1.18)). The microcanonical number of graphs satisfying the constraint is

$$\Omega_{\vec{k}^*_{1\to 2}} = \prod_{i\in\Lambda_1} \binom{n_2}{k_i^*}. \tag{1.27}$$

The canonical ensemble can be obtained from (1.5) by setting

$$H(\mathbf{G},\vec{\theta}) = \vec{\theta}\cdot\vec{k}_{1\to 2}(\mathbf{G}). \tag{1.28}$$

Setting $\vec{\theta} = \vec{\theta^*}$ in order that equation (1.5) is satisfied, we can write the canonical probability as

$$P_{\mathrm{can}}(\mathbf{G}) = \prod_{i\in\Lambda_1} (p_i^*)^{k_i(\mathbf{G})}(1-p_i^*)^{n_2-k_i(\mathbf{G})} \tag{1.29}$$

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

$$f_{n_1} = {n_2}^{-1}\sum_{i\in\Lambda_2}\delta_{k_i^*} \in \mathcal{P}(\mathbb{N}_0). \tag{1.30}$$

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

$$\lim_{n_1\to\infty}\|f_{n_1}-f\|_{\ell^1(g)} = 0. \tag{1.31}$$

The relative entropy per node can be written as

$$s_{n_1+n_2} = \frac{S_{n_1+n_2}(P_{\mathrm{mic}}\mid P_{\mathrm{can}})}{n_1+n_2} = \frac{n_1}{n_1+n_2}\|f_{n_1}\|_{\ell^1(g_{n_2})}, \tag{1.32}$$

with

$$g_{n_2}(k) = -\log\left[\mathrm{Bin}\left(n_2,\tfrac{k}{n_2}\right)(k)\right]\mathbb{I}_{0\le k\le n_2}, \qquad k\in\mathbb{N}_0, \tag{1.33}$$

and $\mathrm{Bin}(n_2,\frac{k}{n_2})(k) = \binom{n_2}{k}(\frac{k}{n_2})^k(\frac{n_2-k}{k})^{n_2-k}$ for $k=0,\dots,n_2$ 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 $c = \lim_{n_1,n_2\to\infty}\frac{n_2}{n_1}$ plays an important role.

Theorem 1.4. *Subject to* (1.26) *and* (1.31)*:*
(1) *If $n_2\to\infty$ with n_1 fixed ($c=\infty$), then $s_\infty = \lim_{n_2\to\infty} s_{n_1+n_2} = 0$.*
(2) *If $n_1, n_2\to\infty$ with $c=\infty$, then $s_\infty = \lim_{n_1,n_2\to\infty} s_{n_1+n_2} = 0$.*
(3) *If $n_1\to\infty$ with n_2 fixed ($c=0$), then*

$$s_\infty = \lim_{n_1\to\infty} s_{n_1+n_2} = \|f\|_{\ell^1(g_{n_2})}. \tag{1.34}$$

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

$$s_\infty = \frac{1}{1+c}\|f\|_{\ell^1(g)}. \tag{1.35}$$

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

$$\vec{C}^* = L^*. \tag{1.36}$$

In analogy with the corresponding result for the uni-partite case (Theorem 1.2), in this case ensemble equivalence is restored.

Theorem 1.5. *Subject to* (1.36)*, the specific relative entropy equals* $s_\infty = 0$*.*

1.3.3 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

$$\gamma_{s,t}(\mathbf{\Gamma}) = \begin{cases} 1 & \text{if a link between layers } s \text{ and } t \text{ exists} \\ 0 & \text{otherwise.} \end{cases} \tag{1.37}$$

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 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

$$\vec{C}^* = \Big\{\vec{k}^*_{s\to t} \colon\ s,t = 1,\dots,M \ \ \gamma_{s,t}(\mathbf{\Gamma}) = 1\Big\}. \tag{1.38}$$

We abbreviate

$$L^*_{s,t} = \sum_{i\in\Lambda_s} k_i^{*\,t} = \sum_{j\in\Lambda_t} k_j^{*\,s}, \quad m^*_{s\to t} = \max_{i\in\Lambda_s} k_i^{*\,t}, \quad f^{(n_s)}_{s\to t} = n_s^{-1} \sum_{i\in\Lambda_s} \delta_{k_i^{*\,t}}, \tag{1.39}$$

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

$$A_s = \lim_{n_1,\dots,n_M\to\infty} \frac{n_s}{n} \quad \forall\, s, \tag{1.40}$$

where $\sum_{s=1}^M A_s = 1$. The *sparse regime* corresponds to

$$\begin{aligned} &m^*_{s\to t} m^*_{t\to s} = o({L^*_{s,t}}^{2/3}), && n_s, n_t \to \infty \ \forall\, s \neq t, \\ &m^*_{s\to s} = o(n_s^{1/2}), && n_s \to \infty \ \forall\, s. \end{aligned} \tag{1.41}$$

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

$$\lim_{n_s\to\infty} \|f^{(n_s)}_{s\to t} - f_{s\to t}\|_{\ell^1(g)}, \qquad \lim_{n_s\to\infty} \|f^{(n_s)}_{s\to s} - f_{s\to s}\|_{\ell^1(g)} = 0. \tag{1.42}$$

Theorem 1.6. *Subject to* (1.38) *and* (1.40)–(1.42)*,*

$$s_\infty = \sum_{\substack{s,t=1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}}^{M} A_s \, \|f_{s\to t}\|_{\ell^1(g)}. \tag{1.43}$$

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.

1.3.4 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 Λ_s linking to Λ_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.

Theorem 1.7. *Subject to the above relaxation,*

$$s_\infty = \sum_{(s,t)\in\mathcal{D}} A_s \, \|f_{s\to t}\|_{\ell^1(g)}. \tag{1.44}$$

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 $\mathcal{D} = \emptyset$, 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 1.6 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 1.6 and the one in Theorem 1.4. For $s = 1, \dots, M$ we set $A_s = \lim_{n_1,n_2,\dots,n_M\to\infty} \frac{n_s}{n}$.

Theorem 1.8. *Subject to the above relaxation,*

$$s_\infty = \sum_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_1\cup\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} A_s \, \|f_{s\to t}\|_{\ell^1(g)}. \tag{1.45}$$

In particular,

$$s_\infty = 0 \quad \Longleftrightarrow \quad A_s = 0 \ \forall\, s \in \big\{u \in \mathcal{M}_1 \colon \exists\, t \in \mathcal{M}_1 \cup \mathcal{M}_2 \text{ with } \gamma_{u,t}(\mathbf{\Gamma}) = 1\big\}. \tag{1.46}$$

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 through which layer, possibly its own layer). In this case we lose the multi-layer structure: constraints are no longer involving pairs of layers. This is the same case described in the configuration model of Theorem 1.1. There are still an extensive number of local constraints, and the ensembles are nonequivalent.

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.

2.1 General considerations

Poissonisation. The function g in (1.15) has an interesting interpretation, namely,

$$g(k) = S\big(\,\delta[k] \mid \text{Poisson}[k]\,\big) \tag{2.1}$$

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 (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 (1.12), 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 1.3–1.4 and 1.6–1.8. The result in Theorem 1.4(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

$$s_n = \sum_{k=1}^n f(k) g_n(k), \qquad g_n(k) = S\big(\,\delta[k] \mid \text{Bin}(n, \tfrac{k}{n})\,\big). \tag{2.2}$$

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 of ensemble non-equivalence known so far in the literature, the generally accepted explanation for the breaking of equivalence is the presence of a non-additive energy, induced e.g. by long-range interactions [12], [13]. However, in all 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.

To that end, we introduce the notion of *additive Hamiltonian*. Define

$$H(\mathbf{G},\vec{\theta}) = \vec{\theta}\cdot\vec{C}(\mathbf{G}). \tag{2.3}$$

If m is the number of constraints, i.e., $\vec{C}$ is an m-valued vector, then we say that the Hamiltonian is additive if and only if (2.3) can be expressed as

$$H(\mathbf{G},\vec{\theta}) = \sum_{i=1}^{m} H(\mathbf{G}_i,\theta_i), \tag{2.4}$$

where $\mathbf{G}_1,\ldots,\mathbf{G}_m$ are subgraphs of $\mathbf{G}$ that have no links between them, and their union is $\mathbf{G}$.

The case of a bi-partite graph with a fixed degree sequence on the top layer is an example of an *additive* Hamiltonian with $m = n_1$. Indeed, from (1.28) we have

$$H(\mathbf{G},\vec{\theta}) = \vec{\theta}\cdot\vec{k}_{1\to 2}(\mathbf{G}) = \sum_{i=1}^{n_1}\theta_i k_{1\to 2}(\mathbf{G}_i), \tag{2.5}$$

where $\mathbf{G}_i$ is the bi-partite graph obtained from the i-th node of the top layer and all the nodes of the bottom layer, and $k_{1\to 2}(\mathbf{G}_i)$ is the degree sequence of the top layer of $\mathbf{G}_i$ (which in this case is just the degree of the single top node of $\mathbf{G}_i$ and also the total number of links in $\mathbf{G}_i$). As is clear from Theorem 1.4, there is equivalence or non-equivalence depending on the relative growth of n_1 and n_2, *independently* of the form of the Hamiltonian.

By contrast, the case of a uni-partite graph with a fixed degree sequence is an example of a *non-additive* Hamiltonian. In this case there always is non-equivalence. On the other hand, for a fixed total number of links there is equivalence, even though the Hamiltonian is non-additive.

These examples 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 [51], and it is confirmed in full generality for the multi-layer setting treated in the present paper. Indeed, we have shown that, whenever the number κ of constraints is *subextensive*, i.e., $\kappa = o(n)$ where n is the number of nodes, ensemble equivalence is restored.

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 the network with a suitable null model [49]. 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 [23], [46] (in Section 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 topologocal 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 [40]. 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 non-equivalence 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 non-equivalent 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 [50], and is here strengthened by our proof of nonequivalence.

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 $\mathcal{G}^\sharp_n$. 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 1.3.3 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 non-equivalent. 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 1.3.1. 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 [9]. Networks with a power-law degree distribution are said to be *scale-free* [11]. 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 (1.13)) of the form $f_n(k) = A_{\gamma,n} k^{-\gamma} \mathbb{I}_{1 \le k \le k_c(n)}$ 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 (1.12) [9]. Since $\lim_{n\to\infty} \|f_n - f\|_{\ell^1(g)} = 0$ with $f(k) = k^{-\gamma}/\zeta(\gamma)$ for $k \ge 1$ and 0 elsewhere, where ζ is the Riemann zeta-function, our result in (1.16) tells us that

$$s_\infty = \sum_{k\in\mathbb{N}} g(k)\, f(k) = \frac{1}{\zeta(\gamma)} \sum_{k\in\mathbb{N}} g(k)\, k^{-\gamma}. \tag{2.6}$$

Since $g(k) \sim \frac{1}{2}\log(2\pi k)$ as $k \to \infty$, we find that s_∞ 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 γ decreases, the degree distribution becomes broader and the degree of nonequivalence increases. A similar conclusion was drawn in [51].

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 1.3.2. 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 1.5) 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 1.6 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

$$s_\infty = \sum_{\substack{s,t=1 \\ s\neq t}}^{M} A_s \, \|f_{s\to t}\|_{\ell^1(g)} > 0, \tag{2.7}$$

which proves ensemble nonequivalence.

An empirical example where a multi-partite structure is relevant is the world-wide network of airports connected by direct international flights: airports in the same country (layer) are necessarily not mutually connected, while airports in different countries can be. However, since not all pairs of countries are connected by direct flights [32], this example suggests that, when useful, the all-to-all multi-partite structure can be conveniently constrained into an "incomplete" multi-partite structure by additionally forbidding inter-layer links between certain pairs of layers. This is easily done by setting the corresponding off-diagonal entries of $\mathbf{\Gamma}$ to zero, i.e., $\gamma_{s,t}(\mathbf{\Gamma}) = 0$ for some $s \neq t$. This restriction can induce structures that more closely resemble the properties of certain empirical networks. In particular, in the mentioned example, the master graph can be chosen to be the empirical network of countries (nodes of $\mathbf{\Gamma}$) connected by direct international flights (edges of $\mathbf{\Gamma}$). Note that international flights are bi-directional, i.e., they always exist in both ways between any two connected countries, consistently with our assumption of undirected graphs. The empirical properties of such a network of inter-country flights are well studied [32]. More importantly for our purposes here, the empirical properties of the "microscopic" network of all airports are also well studied [17], and they again include the presence of a broad empirical degree distribution, much broader than expected under the Erdős-Renyi random graph model. The nodes with largest degree in this network are international hubs, and so the main contribution to their degree comes from airports not in their own country. This means that a realistic model of the multi-partite international airport network should enforce the empirical degree sequence of each layer, and not only the number of links. The corresponding relative entropy is given by (2.7) minus the contribution of the missing edges of the master graph. Unless the master graph has only missing edges (a trivial choice that would induce only the empty graph in $\mathcal{G}_n$), the relative entropy is still strictly positive. Again, we conclude that the relevant ensemble for the mentioned example is nonequivalent.

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 [35], [37], [24]. The earliest and simplest stochastic block model [35] 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 1.7, with $\mathcal{D} = \emptyset$. It follows as a corollary that this model is ensemble equivalent:

$$s_\infty = 0. \tag{2.8}$$

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* [37], where an additional constraint is put on the degree of all nodes. An even more constrained variant of this model has been proposed in [24], 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 1.3.3, with the block structure given by the (complete) master graph. From Theorem 1.6 we calculate the relative entropy as

$$s_\infty = \sum_{s,t=1}^{M} A_s \, \|f_{s\to t}\|_{\ell^1(g)} > 0. \tag{2.9}$$

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 [24], the targeted degree-corrected block model is not just a reparametrisation 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 (2.9), further illustrating the difference between the two models. Still, we expect the relative entropy to be strictly positive, since there is still an extensive number of active constraints. This would contradict the result in [45], where it was claimed that the degree-corrected block models with soft and hard constraints are expected to become equivalent in the thermodynamic limit.

Networks with community structure. Another very important class of graphs that are studied intensively in the literature are graphs with community structure [23], [46]. 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* [23], [46]. 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

$$Q_{\mathbf{G}^*}(\vec{\sigma}) = K_{\mathbf{G}^*} \sum_{1\le i<j\le n} \left(g_{ij}(\mathbf{G}^*) - \langle g_{ij}\rangle\right) \delta_{\sigma_i,\sigma_j}, \tag{2.10}$$

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 (3.7) in the proof of Theorem 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 $i \in \Lambda_s$. Then, we require that the master graph $\mathbf{\Gamma}$ has all possible self-loops, plus a desired number of additional edges that need not be the maximum one (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., $\vec{\sigma}^\dagger = \mathrm{argmax}_{\vec{\sigma}}\, Q_{\mathbf{G}^*}(\vec{\sigma})$. To this end, we rewrite

$$\begin{aligned}
Q_{\mathbf{G}^*}(\vec{\sigma}^\dagger) &= K_{\mathbf{G}^*} \sum_{1\le i<j\le n} \left(g_{ij}(\mathbf{G}^*) - \langle g_{ij}\rangle\right) \delta_{\sigma_i^\dagger,\sigma_j^\dagger} \\
&= \frac{K_{\mathbf{G}^*}}{2} \sum_{1\le i,j\le n} \left(g_{ij}(\mathbf{G}^*) - \langle g_{ij}\rangle\right) \delta_{\sigma_i^\dagger,\sigma_j^\dagger} \\
&= \frac{K_{\mathbf{G}^*}}{2} \sum_{s=1}^{M} \sum_{i,j\in\Lambda_s} \left(g_{ij}(\mathbf{G}^*) - \frac{k_i^* k_j^*}{2L^*}\right) \\
&= \frac{K_{\mathbf{G}^*}}{2} \sum_{s=1}^{M} \left(L_{s,s}^* - \frac{1}{2L^*}\Big(\sum_{i\in\Lambda_s} k_i^*\Big)^2\right) \\
&= \frac{K_{\mathbf{G}^*}}{2} \sum_{s=1}^{M} \left(L_{s,s}^* - \frac{1}{\sum_{s,t=1}^{M} L_{s,t}^*}\Big(\sum_{t=1}^{M} L_{s,t}^*\Big)^2\right),
\end{aligned} \tag{2.11}$$

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 the $\{\vec{k}^*_{s\to t}\}$ satisfies the inequality

$$\sum_{s=1}^{M} L_{s,s}^* > \frac{\sum_{s=1}^{M}\Big(\sum_{t=1}^{M} L_{s,t}^*\Big)^2}{\sum_{s,t=1}^{M} L_{s,t}^*}, \tag{2.12}$$

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 $\vec{\sigma}^\dagger = \mathrm{argmax}_{\vec{\sigma}}\, Q_{\mathbf{G}^*}(\vec{\sigma})$ should instead be enforced by looking for the specific $\{\vec{k}^*_{s\to t}\}$ that maximises (2.11).

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 (1.43) where the degree distributions

$\{f_{s\to t}\}$ are induced by suitable t-targeted degree sequences that realise (2.12) and possibly also $\vec{\sigma}^\dagger = \mathrm{argmax}_{\vec{\sigma}}\, Q_{\mathbf{G}^*}(\vec{\sigma})$.

Multiplex networks and time-varying graphs. Two other important classes of graphs that have recently gained attention are those of *multiplex* networks [8] and *time-varying graphs* [36].

Multiplex networks are networks where the same set of nodes can be connected by M different types of links [8]. Two examples 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) [29]. 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 [29]. 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 [36]. 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. 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 Λ_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 [29]. Our results imply that the relevant ensembles are non-equivalent. In particular, as a corollary of Theorem 1.6 we have

$$s_\infty = \frac{1}{M}\sum_{s=1}^{M} \|f_{s\to s}\|_{\ell^1(g)}. \tag{2.13}$$

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 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

$$s_\infty = \|f\|_{\ell^1(g)}. \tag{2.14}$$

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 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 [8]. 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 (2.13). Indeed, Theorem 1.6 now leads to

$$s_\infty = \frac{1}{M} \sum_{\substack{s,t=1 \\ \gamma_{s,t}(\mathbf{\Gamma})=1}}^{M} \|f_{s\to t}\|_{\ell^1(g)}, \tag{2.15}$$

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 [19]. 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 (1.43) 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 1.7 with $\mathcal{D} = \{(s,s)\colon\ s = 1, M\}$. The relative entropy is therefore

$$s_\infty = \sum_{s=1}^{M} A_s \|f_{s\to s}\|_{\ell^1(g)} \tag{2.16}$$

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

3 Proof of the theorems

3.1 Proof of Theorem 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 (1.12). For this regime it was shown in [4], [41] that

$$\Omega_{\vec{k}^*} = \frac{\sqrt{2}\,(\frac{2L^*}{e})^{L^*}}{\prod_{i=1}^n k_i^*!}\, e^{-(\overline{k^{*2}}/2\overline{k^*})^2+\frac{1}{4}+o(n^{-1}\overline{k^*}^3)}, \tag{3.1}$$

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

$$\begin{aligned} Z(\theta) &= \sum_{\mathbf{G}\in\mathcal{G}_n} e^{-H(\mathbf{G},\vec{\theta})} = \sum_{\mathbf{G}\in\mathcal{G}_n}\prod_{1\leq i<j\leq n} e^{-\theta_i g_{i,j}(\mathbf{G})} \\ &= \sum_{\mathbf{G}\in\mathcal{G}_n}\prod_{1\leq i<j\leq n} e^{-(\theta_i+\theta_j) g_{i,j}(\mathbf{G})} = \prod_{1\leq i<j\leq n}(1+e^{-(\theta_i+\theta_j)}). \end{aligned} \tag{3.2}$$

The canonical probability equals

$$P_{\rm can}(\mathbf{G}\mid\vec{\theta}) = \frac{\prod_{1\leq i<j\leq n} e^{-(\theta_i+\theta_j)g_{i,j}(\mathbf{G})}}{Z(\vec{\theta})} = \prod_{1\leq i<j\leq n}\frac{e^{-(\theta_i+\theta_j)g_{i,j}(\mathbf{G})}}{1+e^{-(\theta_i+\theta_j)}}. \tag{3.3}$$

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

$$\sum_{j\neq i}\frac{e^{-\theta_i^*-\theta_j^*}}{1+e^{-\theta_i^*-\theta_j^*}} = k_i^* \quad \forall\, i \tag{3.4}$$

we have

$$P_{\rm can}(\mathbf{G}) = \prod_{1\leq i<j\leq n}(p_{ij}^*)^{g_{ij}}(1-p_{ij}^*)^{1-g_{ij}}. \tag{3.5}$$

It is ensured by (1.12) that $\lim_{n\to\infty}\frac{1}{n}\sum_{1\leq i<j\leq n}\widehat{p}_{ij}^{*\,2} = 0$, a condition under which we can show that (3.5) has the same asymptotic behaviour as

$$\widehat{P}_{\rm can}(\mathbf{G}) = \prod_{1\leq i<j\leq n}(\widehat{p}_{ij}^*)^{g_{ij}}(1-\widehat{p}_{ij}^*)^{1-g_{ij}}, \tag{3.6}$$

with

$$\widehat{p}_{ij}^* = e^{-\theta_i^*-\theta_j^*} = \frac{k_i^* k_j^*}{2L^*}. \tag{3.7}$$

Indeed,

$$\frac{1}{n}\log\left(\frac{\widehat{P}_{\rm can}(\mathbf{G})}{P_{\rm can}(\mathbf{G})}\right) = \frac{1}{n}\sum_{1\leq i<j\leq n} g_{i,j}\log(1-\widehat{p}_{ij}^*) - \frac{1}{n}\sum_{1\leq i<j\leq n}\log(1-\widehat{p}_{ij}^{*\,2}) \to 0, \qquad n\to\infty, \tag{3.8}$$

because

$$\sum_{1\le i<j\le n} g_{i,j}\log(1-\widehat{p}^*_{ij}) \le (m^*)^2 + O(\widehat{p}^{*\,2}_{ij}) \tag{3.9}$$

and

$$0 \le \frac{1}{n}\sum_{1\le i<j\le n}\widehat{p}^2_{ij} = \frac{1}{2}\left[\frac{\sum_{i=1}^n k_i^2}{\sqrt{n}\sum_{i=1}^n k_i}\right]^2 \le \frac{1}{2}\frac{(m^*)^2}{n} \to 0, \qquad n\to\infty. \tag{3.10}$$

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,

$$\ln P_{\rm can}(\mathbf{G}^*) \;\; = \;\; \sum_{i=1}^n k^*_i \ln k^*_i - L^*\ln(2L^*) - L^* + o(n). \tag{3.11}$$

Combining (3.1) and (3.11), we obtain

$$S_n(P_{\rm mic}\mid P_{\rm can}) = \sum_{i=1}^n g(k^*_i) + o(n), \qquad n\to\infty. \tag{3.12}$$

With the help of (1.13) this reads

$$n^{-1}\,S_n(P_{\rm mic}\mid P_{\rm can}) = \sum_{k\in\mathbb{N}_0} f_n(k)g(k) + o(1) = \|f_n\|_{\ell^1(g)} + o(1), \tag{3.13}$$

which together with (1.14) yields the claim. □

3.2 Proof of Theorem 1.2

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

$$\Omega_{L^*} = \binom{\binom{n}{2}}{L^*} = \binom{K}{\lambda K}, \qquad K = \binom{n}{2}. \tag{3.14}$$

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

$$Z(\theta) = \sum_{\mathbf{G}\in\mathcal{G}_n} e^{-H(\mathbf{G},\theta)} = \sum_{\mathbf{G}\in\mathcal{G}_n}\prod_{1\le i<j\le n} e^{-\theta g_{i,j}(\mathcal{G})} = \prod_{1\le i<j\le n}(1+e^{-\theta}). \tag{3.15}$$

The canonical probability equals

$$P_{\rm can}(\mathbf{G}\mid\theta) = \frac{e^{-\sum_{1\le i<j\le n}\theta g_{i,j}(\mathbf{G})}}{Z(\theta)} = \prod_{1\le i<j\le n}\frac{e^{-\theta g_{i,j}(\mathbf{G})}}{1+e^{-\theta}} = \prod_{1\le i<j\le n} p^{g_{i,j}(\mathbf{G})}(1-p)^{1-g_{i,j}(\mathbf{G})} \tag{3.16}$$

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

$$L^* = \sum_{1\le i<j\le n} p^*, \qquad p^* = \frac{e^{-\theta^*}}{1+e^{-\theta^*}}. \tag{3.17}$$

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

$$\begin{aligned}
\log P_{\mathrm{mic}}(\mathbf{G}^*) &= -\log(K)! + \log(\lambda K)! + \log((1-\lambda)K)! \\
&= -K[\log K - 1] + \lambda K[\log \lambda K - 1] + [(1-\lambda)K][\log((1-\lambda)K) - 1] + o(n) \\
&= K\log(1-\lambda) + \lambda K \log\left(\frac{\lambda}{1-\lambda}\right) + o(n), \\
\log P_{\mathrm{can}}(\mathbf{G}^*) &= \lambda K \log(\lambda) + (1-\lambda)K\log(1-\lambda).
\end{aligned} \tag{3.18}$$

This in turn implies that

$$\lim_{n\to\infty} \frac{S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}})}{n} = 0. \tag{3.19}$$

□

3.3 Proof of Theorem 1.3

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

$$\begin{aligned}
H(\mathbf{G}|\vec{\theta},\vec{\phi}) &= \sum_{i\in\Lambda_1} k_i(\mathbf{G})\theta_i + \sum_{j\in\Lambda_2} k'_j(\mathbf{G})\phi_j \\
&= \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2} \theta_i g_{i,j}(\mathbf{G}) + \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2} \phi_j g_{i,j}(\mathbf{G}) = \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2} (\theta_i+\phi_j) g_{i,j}(\mathbf{G}).
\end{aligned} \tag{3.20}$$

The partition function is

$$Z(\vec{\theta},\vec{\phi}) = \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}} e^{-\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}(\theta_i+\phi_j)g_{i,j}(\mathbf{G})} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2}\left(1+e^{-(\theta_i+\phi_j)}\right). \tag{3.21}$$

The canonical probability becomes

$$\begin{aligned}
P_{\mathrm{can}}(\mathbf{G} \mid \vec{\theta},\vec{\phi}) &= \frac{e^{-\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}(\theta_i+\phi_j)g_{i,j}(\mathbf{G})}}{Z(\vec{\theta},\vec{\phi})} \\
&= \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} \frac{e^{-(\theta_i+\phi_j)g_{i,j}(\mathbf{G})}}{1+e^{-(\theta_i+\phi_j)}} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} p_{i,j}^{g_{i,j}(\mathbf{G})}(1-p_{i,j})^{1-g_{i,j}(\mathbf{G})},
\end{aligned} \tag{3.22}$$

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

$$\begin{cases} \sum_{j\in\Lambda_2} p^*_{i,j} = k^*_i, \\ \sum_{i\in\Lambda_1} p^*_{i,j} = k'^*_j, \end{cases} \tag{3.23}$$

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

$$P_{\mathrm{can}}(\mathbf{G}) = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} p^{*\,g_{i,j}(\mathbf{G})}_{i,j}(1-p^*_{i,j})^{1-g_{i,j}(\mathbf{G})}. \tag{3.24}$$

Under the sparseness condition (1.22), 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

$$\begin{cases} \sum_{j\in\Lambda_2} \widehat{p}^*_{i,j} = k^*_i, \\ \sum_{i\in\Lambda_1} \widehat{p}^*_{i,j} = k'^*_j. \end{cases} \tag{3.25}$$

This has solution

$$\widehat{p}^*_{i,j} = \frac{k^*_i k'^*_j}{L^*}, \qquad L^* = \sum_{i\in\Lambda_1} k^*_i = \sum_{j\in\Lambda_2} k'^*_j. \tag{3.26}$$

We define

$$\widehat{P}_{\rm can}(\mathbf{G}) = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} (\widehat{p}^*_{i,j})^{g_{ij}(\mathbf{G})}(1-\widehat{p}^*_{i,j})^{1-g_{ij}(\mathbf{G})}, \tag{3.27}$$

and note that

$$\frac{1}{n_1+n_2}\log\left(\frac{\widehat{P}_{\rm can}(\mathbf{G})}{P_{\rm can}(\mathbf{G})}\right) \to 0, \qquad n_1, n_2 \to \infty. \tag{3.28}$$

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

$$\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\log(1-p^*_{i,j}) = -\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\frac{k^*_i k'^*_j}{L^*} + o(n_1+n_2), \qquad n_1, n_2 \to \infty. \tag{3.29}$$

Indeed,

$$0 \le \frac{1}{n_1+n_2}\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\widehat{p}^{*\,2}_{i,j} = \frac{1}{n_1+n_2}\frac{\sum_{i\in\Lambda_1}k^{*2}_i\sum_{j\in\Lambda_2}k'^{*2}_j}{\sum_{i\in\Lambda_1}k^*_i\sum_{j\in\Lambda_2}k'^*_j} \le \frac{m^*m'^*}{\sqrt{n_1n_2}}\frac{\sqrt{n_1n_2}}{n_1+n_2} \to 0, \tag{3.30}$$

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

Combining (3.24) and (3.29), we have

$$\begin{aligned}\log P_{\rm can}(\mathbf{G}^*) &= \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2} g_{i,j}(\mathbf{G}^*)\log\left(\frac{k^*_i k'^*_j}{L^*}\right) - \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\frac{k^*_i k'^*_j}{L^*} + o(n_1+n_2)\\ &= \sum_{i\in\Lambda_1} k^*_i\log\left(k^*_i\right) + \sum_{j\in\Lambda_2} k'^*_j\log\left(k'^*_j\right) - L^*\log L^* - L^* + o(n_1+n_2),\end{aligned} \tag{3.31}$$

which concludes our computation for the canonical ensemble.

Microcanonical probabilities come from the results in [31], where it is shown that, as $n_1, n_2 \to \infty$, the number of bi-partite graphs with degree sequences $\vec{k^*}, \vec{k'^*}$ on the two layers is given by

$$\Omega_{\vec{k^*},\vec{k'^*}} = \frac{L^*!}{\prod_{i\in\Lambda_1}k^*_i!\prod_{j\in\Lambda_2}k'^*_j!}\, e^{o(n_1+n_2)}. \tag{3.32}$$

Hence

$$\log P_{\rm mic}(\mathbf{G}^*) = -\log\Omega_{\vec{s^*},\vec{t^*}} = \sum_{i\in\Lambda_1}k^*_i! + \sum_{j\in\Lambda_2}k'^*_j! - \log(L^*!) + o(n_1+n_2). \tag{3.33}$$

From (3.31) and (3.33) we get

$$\begin{aligned}S_{n_1+n_2}(P_{\rm can}\mid P_{\rm mic}) &= \log P_{\rm mic}(\mathbf{G}^*) - \log P_{\rm can}(\mathbf{G}^*)\\ &= \sum_{i\in\Lambda_1}\log\left(\frac{k^*_i!}{k^{*k^*_i}_i}\right) + \sum_{j\in\Lambda_2}\log\left(\frac{k'^*_j!}{k'^{*k'^*_j}_j}\right)\\ &\quad + L^*\log L^* + L^* - \log(L^*!) + o(n_1+n_2)\\ &= \sum_{i\in\Lambda_1} g(k^*_i) + \sum_{j\in\Lambda_2} g(k'^*_j) + o(n_1+n_2),\end{aligned} \tag{3.34}$$

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

$$\begin{aligned} {n_1}^{-1} \sum_{i\in\Lambda_1} g(k_i^*) &= \sum_{k\in\mathbb{N}_0}^{n_2} f_{1\to 2}^{(n_1)}(k) g(k) = \|f_{1\to 2}^{(n_1)}\|_{\ell^1(g)}, \\ {n_2}^{-1} \sum_{j\in\Lambda_2} g(k_j'^*) &= \sum_{k\in\mathbb{N}_0}^{n_1} f_{2\to 1}^{(n_2)}(k) g(k) = \|f_{2\to 1}^{(n_2)}\|_{\ell^1(g)}, \end{aligned} \tag{3.35}$$

we get, with the help of (1.23),

$$\lim_{n_1,n_2\to\infty} \frac{S_{n_1+n_2}(P_{\text{can}} \mid P_{\text{mic}})}{n_1+n_2} = A_1 \|f_{1\to 2}\|_{\ell^1(g)} + A_2 \|f_{2\to 1}\|_{\ell^1(g)}, \tag{3.36}$$

which proves the claim. □

3.4 Proof of Theorem 1.4

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

$$\Omega_{\vec{k}^*} = \prod_{i\in\Lambda_1} \binom{n_2}{k_i^*}. \tag{3.37}$$

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

$$Z(\vec{\theta}) = \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}} e^{-\sum_{i\in\Lambda_1}\theta_i \sum_{j\in\Lambda_2} g_{i,j}(\mathbf{G})} = \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}} \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} e^{-\theta_i g_{i,j}(\mathbf{G})} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} [1+e^{-\theta_i}]. \tag{3.38}$$

The canonical probability becomes

$$P_{\text{can}}(\mathbf{G}|\vec{\theta}) = \frac{e^{-\sum_{i\in\Lambda_1}\theta_i\sum_{j\in\Lambda_2} g_{i,j}(\mathbf{G})}}{Z(\vec{\theta})} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} \frac{e^{-\theta_i g_{i,j}(\mathbf{G})}}{1+e^{-\theta_i}} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} p_i^{g_{i,j}(\mathbf{G})}(1-p_i)^{1-g_{i,j}(\mathbf{G})} \tag{3.39}$$

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

$$k_i^* = \sum_{j\in\Lambda_2} p_i^* = n_2 p_i^*, \qquad p_i^* = \frac{e^{-\theta_i^*}}{1+e^{-\theta_i^*}}. \tag{3.40}$$

It follows that $p_i = \frac{k_i^*}{n_2}$ (recall (1.29)). According to (1.9) we have

$$\begin{aligned} S_{n_1+n_2}(P_{\text{mic}} \mid P_{\text{can}}) &= \ln \frac{P_{\text{mic}}(\mathbf{G}^*)}{P_{\text{can}}(\mathbf{G}^*)} \\ &= -\sum_{i\in\Lambda_1} \log\binom{n_2}{k_i^*} - \sum_{i\in\Lambda_1} k_i^* \log\left(\frac{k_i^*}{n_2}\right) - \sum_{i\in\Lambda_1}(n_2-k_i^*)\log\left(1-\frac{k_i^*}{n_2}\right) \\ &= n_1 n_2 \log n_2 - \sum_{i\in\Lambda_1} \log\left[\binom{n_2}{k_i^*} {k_i^*}^{k_i^*} (n_2-k_i^*)^{(n_2-k_i^*)}\right]. \end{aligned} \tag{3.41}$$

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

$$S_{n_1+n_2}(P_{\text{mic}} \mid P_{\text{can}}) = n_1 n_2 \log n_2 - \sum_{i\in\Lambda_1} U_{n_2}(k_i^*) = n_1 n_2 \log n_2 - n_1 \sum_{k=0}^{n_2} f_{n_1}(k) U_{n_2}(k). \tag{3.42}$$

For the relative entropy per node this gives

$$\begin{aligned} s_{n_1+n_2} &= \frac{n_1}{n_1+n_2}\sum_{k=0}^{n_2} f_{n_1}(k)\, n_2 \log n_2 - \frac{n_1}{n_1+n_2}\sum_{k=0}^{n_2} f_{n_1}(k) U_{n_2}(k) \\ &= -\frac{n_1}{n_1+n_2}\sum_{k=0}^{n_2} f_{n_1}(k) \log \mathrm{Bin}\big(n_2, \tfrac{k}{n_2}\big)(k) = \frac{n_1}{n_1+n_2}\|f_{n_1}\|_{\ell^1(g_{n_2})}. \end{aligned} \tag{3.43}$$

Case (1). Recall (1.15). Note that $x \mapsto z(x) = e^{g(x)}$ is non-decreasing:

$$\frac{z(x-1)}{z(x)} = \left(\frac{x}{x-1}\right)^{x-1}\frac{1}{e} \le 1. \tag{3.44}$$

It therefore follows that

$$\begin{aligned} &\|f_{n_1}\|_{\ell^1(g_{n_2})} = -\sum_{k=0}^{n_2} f_{n_1}(k) \log \mathrm{Bin}\big(n_2, \tfrac{k}{n_2}\big)(k) = \sum_{k=0}^{n_2} f_{n_1}(k) \log\left(\frac{z(k)z(n_2-k)}{z(n_2)}\right) \\ &= \sum_{k\in\mathbb{N}_0} f_{n_1}(k) \log\left(\frac{z(k)z(n_2-k)}{z(n_2)}\right)\mathbb{I}_{k\le n_2} \le \sum_{k\in\mathbb{N}_0} \mathbb{I}_{0\le k\le n_2}\, f_{n_1}(k)\log z(k) \\ &\le \sum_{k\in\mathbb{N}_0} f_{n_1}(k)\log z(k) = \|f_{n_1}\|_{\ell^1(g)} < \infty. \end{aligned} \tag{3.45}$$

By (1.31) and dominated convergence, we may exchange limit and sum to obtain

$$\lim_{n_2\to\infty} s_{n_1,n_2} = \lim_{n_2\to\infty}\frac{n_1}{n_1+n_2}\sum_{k\in\mathbb{N}_0} f_{n_1}(k) \lim_{n_2\to\infty}\log\left(\frac{z(k)z(n_2-k)}{z(n_2)}\right)\mathbb{I}_{0\le k\le n_2} = 0, \tag{3.46}$$

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

Case (2). Using (3.45) and (1.31), we get

$$0 \le s_{n_1,n_2} = \frac{n_1}{n_1+n_2}\|f_{n_1}\|_{\ell^1(g)} \to^{n_1,n_2\to\infty} \frac{1}{1+c}\|f\|_{\ell^1(g)} = 0. \tag{3.47}$$

Case (3). Estimate

$$0 \le |\|f_{n_1}\|_{\ell^1(g_{n_2})} - \|f_{n_1}\|_{\ell^1(g_{n_2})}| \le \|f_{n_1} - f\|_{\ell^1(g_{n_2})} \le \|f_{n_1} - f\|_{\ell^1(g)} \to^{n_1\to\infty} 0. \tag{3.48}$$

Case (4).

$$0 \le |\|f_{n_1}\|_{\ell^1(g_{n_2})} - \|f\|_{\ell^1(g)}| \le \sum_{k\in\mathbb{N}_0} |f_{n_1}(k) - f(k)||g_{n_2}(k)\mathbb{I}_{k\le n_2} - g(k)| \le 2\|f_{n_1} - f\|_{\ell^1(g)}. \tag{3.49}$$

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

3.5 Proof of Theorem 1.5

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

$$\Omega_{L^*} = \binom{n_1 n_2}{L^*} = \binom{n_1 n_2}{\lambda n_1 n_2}. \tag{3.50}$$

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

$$Z(\theta) = \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}} e^{-H(\mathbf{G},\theta)} = \sum_{\mathbf{G}\in\mathcal{G}_{n_1,n_2}} \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} e^{-\theta g_{i,j}(\mathcal{G})} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2}(1+e^{-\theta}). \tag{3.51}$$

The canonical probability equals

$$P_{\mathrm{can}}(\mathbf{G}\mid\vec{\theta}) = \frac{e^{-\sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2}\theta g_{i,j}(\mathbf{G})}}{Z(\vec{\theta})} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2}\frac{e^{-\theta g_{i,j}(\mathbf{G})}}{1+e^{-\theta}} = \prod_{i\in\Lambda_1}\prod_{j\in\Lambda_2} p^{g_{i,j}(\mathbf{G})}(1-p)^{1-g_{i,j}(\mathbf{G})} \tag{3.52}$$

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

$$L^* = \sum_{i\in\Lambda_1}\sum_{j\in\Lambda_2} p^*, \qquad p^* = \frac{e^{-\theta^*}}{1+e^{-\theta^*}}. \tag{3.53}$$

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

$$\begin{aligned}
\log P_{\mathrm{mic}}(\mathbf{G}^*) &= -\log(n_1n_2)! + \log(\lambda n_1 n_2)! + \log((1-\lambda)n_1n_2)! \\
&= -n_1n_2[\log n_1n_2 - 1] + \lambda n_1n_2[\log \lambda n_1 n_2 - 1] \\
&\quad + [1-\lambda n_1 n_2][\log((1-\lambda)n_1n_2) - 1] + o(n_1+n_2) \\
&= n_1n_2\log(1-\lambda) + \lambda n_1 n_2 \log\left(\frac{\lambda}{1-\lambda}\right) + o(n_1+n_2), \\
\log P_{\mathrm{can}}(\mathbf{G}^*) &= n_1n_2\log(1-\lambda) + \lambda n_1 n_2 \log\left(\frac{\lambda}{1-\lambda}\right).
\end{aligned} \tag{3.54}$$

This in turn implies that

$$\lim_{n_1,n_2\to\infty}\frac{S_{n_1+n_2}(P_{\mathrm{mic}}\mid P_{\mathrm{can}})}{n_1+n_2} = 0. \tag{3.55}$$

□

3.6 Proof of Theorem 1.6

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

$\vec{\theta}_{s\to t} = (\theta_1^{(t)}, \ldots, \theta_{n_s}^{(t)})$. The Hamiltonian equals

$$\begin{aligned} & H\big(\mathbf{G} \mid \vec{\theta}_{s\to t};\ s,t=1,\ldots,M,\ \gamma_{s,t}(\mathbf{\Gamma})=1\big) \\ &= \sum_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \sum_{\substack{i\in\Lambda_s\\ j\in\Lambda_t}} (\theta_i^t+\theta_j^s) g_{i,j}(\mathbf{G}) + \sum_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \sum_{\substack{i,j\in\Lambda_s\\ i<j}} (\theta_i^s+\theta_j^s) g_{i,j}(\mathbf{G}) \\ &= \sum_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \sum_{\substack{i\in\Lambda_s\\ j\in\Lambda_t}} H_{s,t}(\mathbf{G}^{(st)} \mid \vec{\theta}_{s\to t}, \vec{\theta}_{t\to s}) + \sum_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \sum_{\substack{i,j\in\Lambda_s\\ i<j}} H_{s,s}(\mathbf{G}^{(ss)} \mid \vec{\theta}_{s\to s}), \end{aligned} \tag{3.56}$$

where

$$\begin{aligned} H_{s,t}(\mathbf{G}^{(st)} \mid \vec{\theta}_{s\to t}, \vec{\theta}_{t\to s}) &= \sum_{\substack{i\in\Lambda_s\\ j\in\Lambda_t}} (\theta_i^t+\theta_j^s) g_{i,j}(\mathbf{G}^{(st)}), \\ H_{s,s}(\mathbf{G}^{(ss)} \mid \vec{\theta}_{s\to s}) &= \sum_{\substack{i,j\in\Lambda_s\\ i<j}} (\theta_i^s+\theta_j^s) g_{i,j}(\mathbf{G}^{(ss)}), \end{aligned} \tag{3.57}$$

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

$$\begin{aligned} & Z\big(\vec{\theta}_{s\to t};\ s,t=1,\ldots,M,\ \gamma_{s,t}(\mathbf{\Gamma})=1\big) = \sum_{\mathbf{G}\in\mathcal{G}_{n_1,\ldots,n_M}(\mathbf{\Gamma})} e^{-H(\mathbf{G}\,|\,\vec{\theta}_{s\to t};\ s,t=1,2,\ldots,M:\ \gamma_{s,t}(\mathbf{\Gamma})=1)} \\ &= \prod_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \sum_{\mathbf{G}^{(st)}\in\mathcal{G}_{n_s,n_t}} e^{-H_{s,t}(\mathbf{G}^{(st)}\,|\,\vec{\theta}_{s\to t},\vec{\theta}_{t\to s})} \prod_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \sum_{\mathbf{G}^{(ss)}\in\mathcal{G}_{n_s,n_s}} e^{-H_{s,s}(\mathbf{G}^{(ss)}\,|\,\vec{\theta}_{s\to s})} \\ &= \prod_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} Z^{(st)}(\vec{\theta}_{s\to t}, \vec{\theta}_{t\to s}) \prod_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} Z^{(ss)}(\vec{\theta}_{s\to s}), \end{aligned} \tag{3.58}$$

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

$$P_{\rm can}(\mathbf{G}) = \prod_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} P^{(st)}_{\rm can}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} P^{(ss)}_{\rm can}(\mathbf{G}^{(ss)}), \tag{3.59}$$

where $P^{(st)}_{\rm can}(\mathbf{G}^{(st)})$ 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 $P^{(ss)}_{\rm can}(\mathbf{G}^{(ss)})$ 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 probabilities for simpler cases. The number of graphs with constraints $\vec{C}^*$ is

$$\begin{aligned}
\Omega_{\vec{k}^*_{s\to t};\, s,t\in\{1,\dots,M\},\ \gamma_{s,t}(\mathbf{\Gamma})=1} &= \left|\left\{ \mathbf{G}\in\mathcal{G}_{n_1,\dots,n_M}(\mathbf{\Gamma}) : \sum_{j\in\Lambda_t} g_{i,j}(\mathbf{G}) = k_i^{*\,t}\ \forall\, i\in\Lambda_s\, \forall\, \gamma_{s,t}=1 \right\}\right| \\
&= \prod_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \left|\left\{ \mathbf{G}^{(st)}\in\mathcal{G}_{n_s,n_k} : \sum_{j\in\Lambda_t} g_{i,j}(\mathbf{G}^{(st)}) = k_i^{*\,t}\ \forall\, i\in\Lambda_s, \sum_{i\in\Lambda_s} g_{i,j}(\mathbf{G}^{(st)}) = k_j^{*\,s}\ \forall\, j\in\Lambda_t \right\}\right| \\
&\quad \prod_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \left|\left\{ \mathbf{G}^{(ss)}\in\mathcal{G}_{n_s} : \sum_{j\in\Lambda_s} g_{i,j}(\mathbf{G}^{(ss)}) = s_i^{*\,h}\ \forall\, i\in\Lambda_s \right\}\right| \\
&= \prod_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \Omega_{\vec{k}^*_{s\to t},\vec{k}^*_{t\to s}} \prod_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \Omega_{\vec{k}^*_{s\to s}}.
\end{aligned} \tag{3.60}$$

This means the microcanonical probability can be factorised as

$$P_{\mathrm{mic}}(\mathbf{G}) = \prod_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} P^{(st)}_{\mathrm{mic}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} P^{(ss)}_{\mathrm{mic}}(\mathbf{G}^{(ss)}), \tag{3.61}$$

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 (3.59) and (3.61) imply that the relative entropy equals the sum

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = \sum_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} S_n(P^{(st)}_{\mathrm{mic}} \mid P^{(st)}_{\mathrm{can}}) + \sum_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} S_n(P^{(ss)}_{\mathrm{mic}} \mid P^{(ss)}_{\mathrm{can}}). \tag{3.62}$$

We can now apply Theorems 1.1 and 1.3 to get the asymptotic relative entropy per nodes as

$$\begin{aligned}
&\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}})}{n} \\
&= \sum_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{(st)}_{\mathrm{mic}} \mid P^{(st)}_{\mathrm{can}})}{n} + \sum_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{(ss)}_{\mathrm{mic}} \mid P^{(ss)}_{\mathrm{can}})}{n} \\
&= \sum_{\substack{1\le s<t\le M\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \left\{ A_s \left\| f_{s\to t}\right\|_{\ell^1(g)} + A_t \left\| f_{t\to s}\right\|_{\ell^1(g)} \right\} + \sum_{\substack{s=1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}}^{M} \left\{ A_s \left\| f_{s\to s}\right\|_{\ell^1(g)} \right\} \\
&= \sum_{\substack{s,t=1\\ \gamma_{s,t}(\mathbf{\Gamma})}}^{M} A_s \left\| f_{s\to t}\right\|_{\ell^1(g)}.
\end{aligned} \tag{3.63}$$

□

3.7 Proof of Theorem 1.7

Proof. We start by studying the Hamiltonian. For each pair (s,t) of layers in $\mathcal{D}$, we define Lagrange multipliers $\vec{\theta}_{s\to t} = (\theta_1^t, \dots, \theta_{n_s}^t)$. For each pair (s,t) of layers in $\mathcal{L}$, we define a Lagrange multiplier $\theta_{s,t}$. The Hamiltonian is

$$\begin{aligned} &H\big(\mathbf{G} \mid \vec{\theta}_{s\to t}, \theta_{l,m};\, (s,t)\in\mathcal{D}, (l,m)\in\mathcal{L}\big) \\ &= H_{\mathcal{D}}(\mathbf{G} \mid \vec{\theta}_{s\to t};\, (s,t)\in\mathcal{D}) + H_{\mathcal{L}}(\mathbf{G} \mid \theta_{l,m};\, (l,m)\in\mathcal{L}) \end{aligned} \tag{3.64}$$

with

$$\begin{aligned} H_{\mathcal{D}}(\mathbf{G} \mid \vec{\theta}_{s\to t};\, (s,t)\in\mathcal{D}) &= \sum_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{D}}} \sum_{\substack{i\in\Lambda_s\\ j\in\Lambda_t}} (\theta_i^t+\theta_j^s) g_{i,j}(\mathbf{G}) + \sum_{\substack{s=1\\(s,s)\in\mathcal{D}}}^{M} \sum_{\substack{i,j\in\Lambda_s\\ i<j}} (\theta_i^s+\theta_j^s) g_{i,j}(\mathbf{G}), \\ H_{\mathcal{L}}(\mathbf{G} \mid \theta_{s,t};\, (s,t)\in\mathcal{L}) &= \sum_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{L}}} \sum_{\substack{i\in\Lambda_s\\ j\in\Lambda_t}} (\theta_{s,t}) g_{i,j}(\mathbf{G}) + \sum_{\substack{s=1\\(s,s)\in\mathcal{L}}}^{M} \sum_{\substack{i,j\in\Lambda_s\\ i<j}} (\theta_{s,s}) g_{i,j}(\mathbf{G}). \end{aligned} \tag{3.65}$$

Consequently, the canonical ensemble is

$$P_{\text{can}}(\mathbf{G}) = P_{\text{can}}^{\mathcal{D}}(\mathbf{G}) P_{\text{can}}^{\mathcal{L}}(\mathbf{G}) \tag{3.66}$$

with

$$\begin{aligned} P_{\text{can}}^{\mathcal{D}}(\mathbf{G}) &= \prod_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{D}}} P_{\text{can}}^{(st)^{\mathcal{D}}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\(s,s)\in\mathcal{D}}}^{M} P_{\text{can}}^{(ss)^{\mathcal{D}}}(\mathbf{G}^{(ss)}), \\ P_{\text{can}}^{\mathcal{L}}(\mathbf{G}) &= \prod_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{L}}} P_{\text{can}}^{(st)^{\mathcal{L}}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\(s,s)\in\mathcal{L}}}^{M} P_{\text{can}}^{(ss)^{\mathcal{L}}}(\mathbf{G}^{(ss)}). \end{aligned} \tag{3.67}$$

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 $g_{i,j}(\mathbf{G}^{(st)}) = g_{i,j}(\mathbf{G})$ ($g_{i,j}(\mathbf{G}^{(ss)}) = g_{i,j}(\mathbf{G})$).

- $P_{\text{can}}^{(st)^{\mathcal{D}}}(\mathbf{G}^{(st)})$ ($P_{\text{can}}^{(ss)^{\mathcal{D}}}(\mathbf{G}^{(ss)})$) 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 $\vec{k}^*_{s\to s}$).

- $P_{\text{can}}^{(st)^{\mathcal{L}}}(\mathbf{G}^{(st)})$ ($P_{\text{can}}^{(ss)^{\mathcal{L}}}(\mathbf{G}^{(ss)})$) 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

$$\Omega_{\vec{k}^*_{s\to t}, L^*_{l,m};\,(s,t)\in\mathcal{D},(l,m)\in\mathcal{L}} = \Omega_{\vec{k}^*_{s\to t};(s,t)\in\mathcal{D}}\, \Omega_{L_{l,m};(l,m)\in\mathcal{L}}. \tag{3.68}$$

This means that the microcanonical probability can be factorised as

$$P_{\mathrm{mic}}(\mathbf{G}) = P^{\mathcal{D}}_{\mathrm{mic}}(\mathbf{G}) P^{\mathcal{L}}_{\mathrm{mic}}(\mathbf{G}) \tag{3.69}$$

with

$$\begin{aligned} P^{\mathcal{D}}_{\mathrm{mic}}(\mathbf{G}) &= \prod_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{D}}} {P^{(st)}_{\mathrm{mic}}}^{\mathcal{D}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\(s,s)\in\mathcal{D}}}^{M} {P^{(ss)}_{\mathrm{mic}}}^{\mathcal{D}}(\mathbf{G}^{(ss)}), \\ P^{\mathcal{L}}_{\mathrm{mic}}(\mathbf{G}) &= \prod_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{L}}} {P^{(st)}_{\mathrm{mic}}}^{\mathcal{L}}(\mathbf{G}^{(st)}) \prod_{\substack{s=1\\(s,s)\in\mathcal{L}}}^{M} {P^{(ss)}_{\mathrm{mic}}}^{\mathcal{L}}(\mathbf{G}^{(ss)}). \end{aligned} \tag{3.70}$$

Here,

- ${P^{(st)}_{\mathrm{mic}}}^{\mathcal{D}}(\mathbf{G}^{(st)})$ (${P^{(ss)}_{\mathrm{mic}}}^{\mathcal{D}}(\mathbf{G}^{(ss)})$) 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 $\vec{k}^*_{s\to s}$).
- ${P^{(st)}_{\mathrm{mic}}}^{\mathcal{L}}(\mathbf{G}^{(st)})$ (${P^{(ss)}_{\mathrm{mic}}}^{\mathcal{L}}(\mathbf{G}^{(ss)})$) 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

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = S_n(P^{\mathcal{D}}_{\mathrm{mic}} \mid P^{\mathcal{D}}_{\mathrm{can}}) + S_n(P^{\mathcal{L}}_{\mathrm{mic}} \mid P^{\mathcal{L}}_{\mathrm{can}}). \tag{3.71}$$

It follows that

$$\begin{aligned} &\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}})}{n} \\ &\quad= \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{\mathcal{D}}_{\mathrm{mic}} \mid P^{\mathcal{D}}_{\mathrm{can}})}{n} + \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{\mathcal{L}}_{\mathrm{mic}} \mid P^{\mathcal{L}}_{\mathrm{can}})}{n}. \end{aligned} \tag{3.72}$$

Using Theorem 1.6 we get

$$\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{\mathcal{D}}_{\mathrm{mic}} \mid P^{\mathcal{D}}_{\mathrm{can}})}{n} = \sum_{(s,t)\in\mathcal{D}} A_s \, \|f_{s\to t}\|_{\ell^1(g)}. \tag{3.73}$$

Moreover,

$$\begin{aligned} &\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P^{\mathcal{L}}_{\mathrm{mic}} \mid P^{\mathcal{L}}_{\mathrm{can}})}{n} \\ &= \lim_{n_1,\dots,n_M\to\infty} \sum_{\substack{1\le s<t\le M\\(s,t)\in\mathcal{L}}} \frac{S_n({P^{(st)}_{\mathrm{mic}}}^{\mathcal{L}} \mid {P^{(st)}_{\mathrm{can}}}^{\mathcal{L}})}{n} + \lim_{n_1,\dots,n_M\to\infty} \sum_{\substack{s=1\\(s,s)\in\mathcal{L}}}^{M} \frac{S_n({P^{(ss)}_{\mathrm{mic}}}^{\mathcal{L}} \mid {P^{(ss)}_{\mathrm{can}}}^{\mathcal{L}})}{n}. \end{aligned} \tag{3.74}$$

Using Theorems 1.2 and 1.5, we get

$$\lim_{n_1,\dots,n_M\to\infty} \frac{S_n({P^{(st)}_{\mathrm{mic}}}^{\mathcal{L}} \mid {P^{(st)}_{\mathrm{can}}}^{\mathcal{L}})}{n} = \lim_{n_1,\dots,n_M\to\infty} \frac{S_n({P^{(ss)}_{\mathrm{mic}}}^{\mathcal{L}} \mid {P^{(ss)}_{\mathrm{can}}}^{\mathcal{L}})}{n} = 0, \tag{3.75}$$

which proves the claim. □

3.8 Proof of Theorem 1.8

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

$$\begin{aligned} &H\big(\mathbf{G} \mid \vec{\theta}_{s\to t};\ s\in\mathcal{M}_1,\ t\in\mathcal{M}_1\cup\mathcal{M}_2,\ \gamma_{s,t}(\mathbf{\Gamma})=1\big) \\ &= \sum_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to t}\vec{s}_{s\to t}(\mathbf{G}) + \sum_{\substack{s\in\mathcal{M}_1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to s}\vec{s}_{s\to s}(\mathbf{G}) + \sum_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to t}\vec{s}_{s\to t}(\mathbf{G}) \\ &= H_{\mathcal{M}_1\to\mathcal{M}_1} + H_{\mathcal{M}_1\to\mathcal{M}_2}, \end{aligned} \tag{3.76}$$

with

$$\begin{aligned} H_{\mathcal{M}_1\to\mathcal{M}_1} &= \sum_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to t}\vec{s}_{s\to t}(\mathbf{G}) + \sum_{\substack{s\in\mathcal{M}_1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to s}\vec{s}_{s\to s}(\mathbf{G}), \\ H_{\mathcal{M}_1\to\mathcal{M}_2} &= \sum_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \vec{\theta}_{s\to t}\vec{s}_{s\to t}(\mathbf{G}). \end{aligned} \tag{3.77}$$

Consequently, the canonical ensemble is

$$P_{\rm can}(\mathbf{G}) = P_{\rm can}^{\mathcal{M}_1\to\mathcal{M}_1}(\mathbf{G}) P_{\rm can}^{\mathcal{M}_1\to\mathcal{M}_2}(\mathbf{G}) \tag{3.78}$$

with

$$\begin{aligned} P_{\rm can}^{\mathcal{M}_1\to\mathcal{M}_1}(\mathbf{G}) &= \prod_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} P_{\rm can}^{(st)^{top,bot}}(\mathbf{G}^{(st)}) \prod_{\substack{s\in\mathcal{M}_1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}} P_{\rm can}^{(ss)}(\mathbf{G}^{(ss)}), \\ P_{\rm can}^{\mathcal{M}_1\to\mathcal{M}_2}(\mathbf{G}) &= \prod_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} P_{\rm can}^{(st)^{top}}(\mathbf{G}^{(st)}). \end{aligned} \tag{3.79}$$

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

$$\Omega_{\vec{k}^*_{s\to t};\ s\in\mathcal{M}_1,\ t\in\mathcal{M}_1\cup\mathcal{M}_2,\ \gamma_{s,t}(\mathbf{\Gamma})=1} = \Omega_{\vec{k}^*_{s\to t};\, s,t\in\mathcal{M}_1,\ \gamma_{s,t}(\mathbf{\Gamma})=1} \Omega_{\vec{k}^*_{s\to t};\, s\in\mathcal{M}_1,\ t\in\mathcal{M}_2,\ \gamma_{s,t}(\mathbf{\Gamma})=1}. \tag{3.80}$$

This means that the microcanonical probability can be factorised as

$$P_{\mathrm{mic}}(\mathbf{G}) = P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_1}(\mathbf{G}) P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_2}(\mathbf{G}) \tag{3.81}$$

with

$$\begin{aligned} P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_1}(\mathbf{G}) &= \prod_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} {P_{\mathrm{mic}}^{(st)}}^{top,bot}(\mathbf{G}^{(st)}) \prod_{\substack{s\in\mathcal{M}_1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}} P_{\mathrm{mic}}^{(ss)}(\mathbf{G}^{(ss)}), \\ P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_2}(\mathbf{G}) &= \prod_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} {P_{\mathrm{mic}}^{(st)}}^{top}(\mathbf{G}^{(st)}). \end{aligned} \tag{3.82}$$

Here,

- ${P_{\mathrm{mic}}^{(st)}}^{top,bot}(\mathbf{G}^{(st)})$ 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.
- $P_{\mathrm{mic}}^{(ss)}(\mathbf{G}^{(ss)})$ is the microcanonical probability of the uni-partite graph $\mathbf{G}^{(ss)}$ with constraint $\vec{k}^*_{s\to s}$.
- ${P_{\mathrm{mic}}^{(st)}}^{top}(\mathbf{G}^{(st)})$ 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

$$S_n(P_{\mathrm{mic}} \mid P_{\mathrm{can}}) = S_n(P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_1} \mid P_{\mathrm{can}}^{\mathcal{M}_1\to\mathcal{M}_1}) + S_n(P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_2} \mid P_{\mathrm{can}}^{\mathcal{M}_1\to\mathcal{M}_2}). \tag{3.83}$$

It follows that

$$\begin{aligned} &\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\mathrm{can}} \mid P_{\mathrm{can}})}{n} \\ &= \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_1} \mid P_{\mathrm{can}}^{\mathcal{M}_1\to\mathcal{M}_1})}{n} + \lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_2} \mid P_{\mathrm{can}}^{\mathcal{M}_1\to\mathcal{M}_2})}{n}. \end{aligned} \tag{3.84}$$

Using again Theorem 1.6 we get

$$\begin{aligned} &\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_1} \mid P_{\mathrm{can}}^{\mathcal{M}_1\to\mathcal{M}_1})}{n} \\ &= \sum_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \left\{ A_s \left\| f_{s\to t} \right\|_{\ell^1(g)} + A_t \left\| f_{t\to s} \right\|_{\ell^1(g)} \right\} + \sum_{\substack{s\in\mathcal{M}_1\\ \gamma_{s,s}(\mathbf{\Gamma})=1}} A_s \left\| f_{s\to s} \right\|_{\ell^1(g)} \\ &= \sum_{\substack{s,t\in\mathcal{M}_1\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} A_s \left\| f_{s\to t} \right\|_{\ell^1(g)}. \end{aligned} \tag{3.85}$$

From Theorem 1.4 we get

$$\begin{aligned} &\lim_{n_1,\dots,n_M\to\infty} \frac{S_n(P_{\mathrm{mic}}^{\mathcal{M}_1\to\mathcal{M}_2} \mid P_{\mathrm{can}}^{\mathcal{M}_1\to\mathcal{M}_2})}{n} \\ &= \lim_{n_1,\dots,n_M\to\infty} \sum_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} \frac{S_n({P_{\mathrm{mic}}^{(st)}}^{\mathcal{M}_1\to\mathcal{M}_2} \mid {P_{\mathrm{can}}^{(st)}}^{\mathcal{M}_1\to\mathcal{M}_2})}{n} = \sum_{\substack{s\in\mathcal{M}_1,\ t\in\mathcal{M}_2\\ \gamma_{s,t}(\mathbf{\Gamma})=1}} A_s \left\| f_{s\to l} \right\|_{\ell^1(g)}, \end{aligned} \tag{3.86}$$

which concludes the proof. □

References

[1] K. Anand and G. Bianconi, *Phys. Rev. E* **80**, 045102 (2009).

[2] J. Barré and B. Goncalves, *Physica A* **386**, 212 (2007).

[3] J. Barré, D. Mukamel, and S. Ruffo, *Phys. Rev. Lett.* **87**, 030601 (2001).

[4] E.A. Bender, *Discrete Math.* **10**, 217 (1974).

[5] E.A. Bender and E.R. Canfield, *J. Comb. Theory (A)* **24**, 296 (1978).

[6] G. Bianconi, *Physical Review E* **87**, 062806 (2013).

[7] M. Blume, V.J. Emery, and R.B. Griffiths, *Phys. Rev. A* **4**, 1071 (1971).

[8] S. Boccaletti, G. Bianconi, R. Criado, C.I. Del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadalj, Z. Wang, and M. Zanin, *Physics Reports* **544**, 1 (2014).

[9] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, *The European Physical Journal B* **38**, 205-209 (2004).

[10] B. Bollobás, *European J. Combin.* **1**, 311 (1980).

[11] G. Caldarelli, *Scale-free networks: complex webs in nature and technology* (OUP, Oxford, 2007).

[12] A. Campa, T. Dauxois, D. Fanelli, and S. Ruffo, *Physics of long-range interacting systems* (OUP, Oxford, 2014).

[13] A. Campa, T. Dauxois, and S. Ruffo, *Physics Reports* **480**, 57 (2009).

[14] P.J. Carrington, J. Scott, and S. Wasserman, *Models and methods in social network analysis*, Vol. 28 (Cambridge university press, 2005).

[15] P.-H. Chavanis, *Astron. & Astrophys.* **401**, 15 (2003).

[16] F. Chung, and L. Lu, *Proceedings of the National Academy of Sciences* **99**, 15879 (2002).

[17] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2015 (2006).

[18] M. D'Agostino *et al.*, *Phys. Lett. B* **473**, 219 (2000).

[19] G. D'Agostino, and A. Scala, *Networks of networks: the last frontier of complexity*, Vol. 340 (Springer, Berlin, 2014).

[20] R.S. Ellis, K. Haven, and B. Turkington, *J. Stat. Phys.* **101**, 999 (2000).

[21] R.S. Ellis, K. Haven, and B. Turkington, *Nonlinearity* **15**, 239 (2002).

[22] R.S. Ellis, H. Touchette, and B. Turkington, *Physica A* **335**, 518 (2004).

[23] S. Fortunato, *Physics reports* **486**, 75 (2010).

[24] P. Fronczak, A. Fronczak, and M. Bujok, *Physical Review E* **88**, 032810 (2013).

[25] D. Garlaschelli, *New Journal of Physics* **11**, 073005 (2009).

[26] D. Garlaschelli, S.E. Ahnert, T. Fink, and G. Caldarelli, *Entropy* **15**, 3148 (2013).

[27] D. Garlaschelli, F. den Hollander, and A. Roccaverde, *Nieuw Archief voor Wiskunde* **5/16**, 207 (2015).

[28] D. Garlaschelli and M.I. Loffredo, *Physical Review E* **78**, 015101 (2008).

[29] V. Gemmetto, and D. Garlaschelli, *Scientific reports* **5**, 9120 (2015).

[30] J.W. Gibbs, *Elementary Principles of Statistical Mechanics*, Yale University Press, New Haven, Connecticut (1902).

[31] C. Greenhill, B.D. McKay, and X. Wang, *J. Comb. Theory, Series A* **113**, 291 (2006).

[32] R. Guimera, S. Mossa, A. Turtschi, and L.A. Amaral, *Proceedings of the National Academy of Sciences* **102**, 7794 (2005).

[33] P. Hertel and W. Thirring, *Annals of Physics* **63**, 520 (1971).

[34] R. van der Hofstad, *Random Graphs and Complex Networks, Volume I*, to appear with Cambridge University Press.

[35] P.W. Holland, K.B. Laskey, and S. Leinhardt, *Social networks* **5**, 109 (1983).

[36] P. Holme, and J. Saramäki, *Physics reports* **519**, 97 (2012).

[37] B. Karrer, and M.E.J. Newman, *Physical Review E* **83**, 016107 (2011).

[38] D. Lynden-Bell, *Physica A* **263**, 293 (1999).

[39] D. Lynden-Bell, R. Wood, *Monthly Notices of the Royal Astronomical Society* **138**, 495 (1968).

[40] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, *New Journal of Physics* **16**, 043022 (2014).

[41] B.D. McKay and N.C. Wormald, *Combinatorica* **11**, 369 (1991).

[42] M. Molloy, and B. Reed, *Random Structures & Algorithms* **6**, 161 (1995).

[43] M.E. Newman, S.H. Strogatz, and D.J. Watts, *Physical review E* **64**, 026118 (2001).

[44] J. Park, and M.E. Newman, *Physical Review E* **70**, 066117 (2004).

[45] T.P. Peixoto, *Physical Review E* **85**, 056122 (2012).

[46] M.A. Porter, J.P. Onnela, and P.J. Mucha, *Notices of the AMS* **56**, 1082-1097.

[47] C. Radin and L. Sadun, *J. Phys. A: Math. Theor.* **46**, 305002 (2013).

[48] C. Radin and M. Yi, *The Annals of Applied Probability* **23**, 2458-2471 (2013).

[49] T. Squartini, and D. Garlaschelli, *New Journal of Physics 13*, 083001 (2011).

[50] T. Squartini, R. Mastrandrea, and D. Garlaschelli, *New J. Phys.* **17**, 023052 (2015).

[51] T. Squartini, J. de Mol, F. den Hollander, and D. Garlaschelli, *Phys. Rev. Lett.* **115**, 268701 (2015).

[52] W. Thirring, *Zeitschrift für Physik* **235**, 339 (1970).

[53] H. Touchette, *J. Stat. Phys.* **159**, 987 (2015).

[54] H. Touchette, R.S. Ellis, and B. Turkington, *Physica A* **340**, 138 (2004).