Strong ensemble nonequivalence in systems with local constraints

Zhang, Q.; Garlaschelli, D.

PAPER • OPEN ACCESS

Strong ensemble nonequivalence in systems with local constraints

View the article online for updates and enhancements.

New Journal of Physics
The open access journal at the forefront of physics

PAPER

Strong ensemble nonequivalence in systems with local constraints

Qi Zhang[1,2,*] ⏺ and Diego Garlaschelli[2,3,*] ⏺

[1] School of Science, Jiangsu University of Science and Technology, Zhenjiang 212100, People's Republic of China
[2] Lorentz Institute for Theoretical Physics, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
[3] IMT School for Advanced Studies, Piazza San Francesco 19, 55100 Lucca, Italy
[*] Authors to whom any correspondence should be addressed.

E-mail: Qi.Zhang@just.edu.cn and garlaschelli@lorentz.leidenuniv.nl

Abstract

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

1. Introduction

In statistical physics, systems with different constraints can be described by different ensembles. For example, systems with fixed energy can be described by the *microcanonical ensemble*, where all microscopic configurations have precisely the same value of the energy and are equiprobable, thereby modelling large isolated systems. In this case, the energy is treated as a 'hard' constraint enforced separately on each configuration. By contrast, systems with fixed temperature (which is the 'dual' thermodynamic quantity conjugated with the energy) can be described by the *canonical ensemble*, where individual microscopic configurations can have different values of the energy and are assigned different probabilities, but in such a way that the average value of the energy coincides with the one defining the corresponding microcanonical ensemble [1]. This ensemble represents systems that can exchange energy with their environment, and the energy is in fact treated as a 'soft' constraint which is enforced only as an ensemble average.

When the size of the system is finite, the two ensembles are necessarily different. However, in the simplest and most traditional situation, the microcanonical description as a function of the energy becomes equivalent with the canonical description as a function of the temperature in the *thermodynamic limit* (i.e., when the number of particles in the system tends to infinity). This phenomenon is called *ensemble*

equivalence (EE) and is a basic concept in statistical mechanics as already established by Gibbs [2]. The property of EE justifies the replacement of the (typically unfeasible) asymptotic calculations in the microcanonical ensemble with the corresponding (much easier) calculations in the canonical ensemble, i.e. to choose the ensemble based on mathematical convenience.

However, over the past decades, the breakdown of EE has been observed in various physical systems, including models of gravitation, fluid turbulence, quantum phase separation, and networks [3–5]. When the system is under *ensemble nonequivalence* (EN), the microcanonical description can no longer be replaced by the canonical description in the thermodynamic limit. In this situation, many assumptions and calculations that are based on EE in statistical mechanics do not hold anymore. Thus checking for the breaking of EE is important for both practical applications and theoretical research. Quantitatively, EN can be defined as a nonvanishing relative entropy density between the microcanonical and canonical probability distributions of microscopic configurations [5, 6]. This is equivalent to a nonvanishing difference between the canonical and microcanonical entropy densities [7]. Technically, this is the so-called *measure-level* EN, which (under mild assumptions) has been shown to coincide with other (*macrostate-level* and *thermodynamic-level*) definitions as well [6]. Importantly, the traditional criterion for EE based on the vanishing of the relative canonical fluctuations of the constraints has been recently found to break down when the latter are local in nature and extensive in number [8].

Indeed, for the most studied systems in statistical physics, the number of constraints defining the ensembles of interest is finite. Traditional physical examples are *global* constraints such as the total energy and the total number of particles. Non-physical examples of systems under global constraints have also been considered, e.g. networks with given total numbers of edges and triangles and/or wedges [9]. In order to observe the breakdown of EE in these systems, one typically needs to introduce long-range interactions implying the non-additivity of the energy and possibly associated with the onset of phase transitions [10] (in the example of networks, the underlying mechanism is a sort of 'frustration' in the simultaneous realizability of the desired numbers of edges and triangles [9]). In this form of EN, the relative entropy between canonical and microcanonical ensembles is of the same order as the canonical entropy itself [6, 9]. This is what we will refer to as a 'strong' form of EN. At the same time, this form of EN is also 'restricted', because it is confined to a selected (e.g. critical) region of the space of parameters. Outside this region, EE is restored.

Recently, a new manifestation of EN has been observed in a different class of network ensembles, where a constraint is enforced on the *degree* (number of links) and/or the *strength* (total weight of all incident links) of each node [5, 8, 11, 12]: unlike systems with global constraints, in these models the number of constraints grows linearly in the number of nodes. This crucial difference implies that, at variance with the more 'traditional' situation described above, the onset of EN in this class of models is completely unrelated to phase transitions and is instead the result of the presence of an *extensive* number of *local* constraints [7, 13, 14]. This situation is by far less studied, because systems with local constraints are not the traditional focus of statistical physics and have attracted attention only recently as models of complex systems with built-in spatial heterogeneity [15] and/or temporal non-stationarity [16]. In this different form of EN, the relative entropy between microcanonical and canonical ensembles is, at least for the cases studied so far, of lower order (i.e. subleading) with respect to the canonical entropy. For this reason, we may refer to this situation as a 'weak' (i.e. weaker than the one found in the presence of phase transitions) form of EN. However, this form of EN is 'unrestricted', precisely because it is not confined to specific values of the control parameters and holds in the entire parameter space. Rather than a property of a phase (or a phase boundary), in this case EN appears to be an intrinsic property of the system itself. In these models, no parameter value can restore EE.

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

In this paper, we will address these problems by exploring the effects of the presence of an extensive number of local constraints on more general ensembles than the ones that have been considered so far to model random networks with given node degrees [5, 11, 13, 14]. In particular, we consider the general setting where each of the n units of the system has a number m of 'state variables' (or 'degrees of freedom'),

and where constraints are defined as sums over these state variables. Surprisingly, besides confirming the onset of an unrestricted form of EN in the thermodynamic limit where n diverges, we also find its simultaneous manifestation in strong form. This happens when each element of the system retains only a finite number m of degrees of freedom in the thermodynamic limit. For brevity, we will denote this situation as the 'strong and unrestricted' form of EN. To the best of our knowledge, this finding provides the first evidence that EN, *even in its strong form*, does not need phase transitions and can appear in the entire parameter space as an intrinsic property of the system, if the latter is subject to an extensive number of local constraints. This simultaneously 'strong and unrestricted' form of EN is the most robust among the ones studied so far. Spatial heterogeneity and temporal non-stationarity are simple candidate mechanisms that can lead to this phenomenon.

To emphasize the general and important consequences of this form of EN for a diverse range of practical applications, we consider generic ensembles of random matrices with fixed margins. These ensembles, which include matrices with 0/1 (or equivalently ±1) and non-negative integer entries subject to global or local constraints, arise for instance in studies of multi-cell gene expression profiles [17], multiplex (online) social activity [18], multi-channel communication systems [19], complex networks [20], and multivariate time series in finance [16], neuroscience [21] or other disciplines. Our results imply that, in many practical situations, the assumption of EE is incorrect and leads to mathematically wrong conclusions. For the benefit of the aforementioned applications, we compensate for the 'disconnection' between the two ensembles by calculating explicitly the correct canonical and microcanonical quantities of interest via a generalized relationship that is either analytically computable or asymptotically determined by the covariance matrix of the constraints in the canonical ensemble. These calculations represent a practical tool for properly dealing with the consequences of EN in all real-world situations.
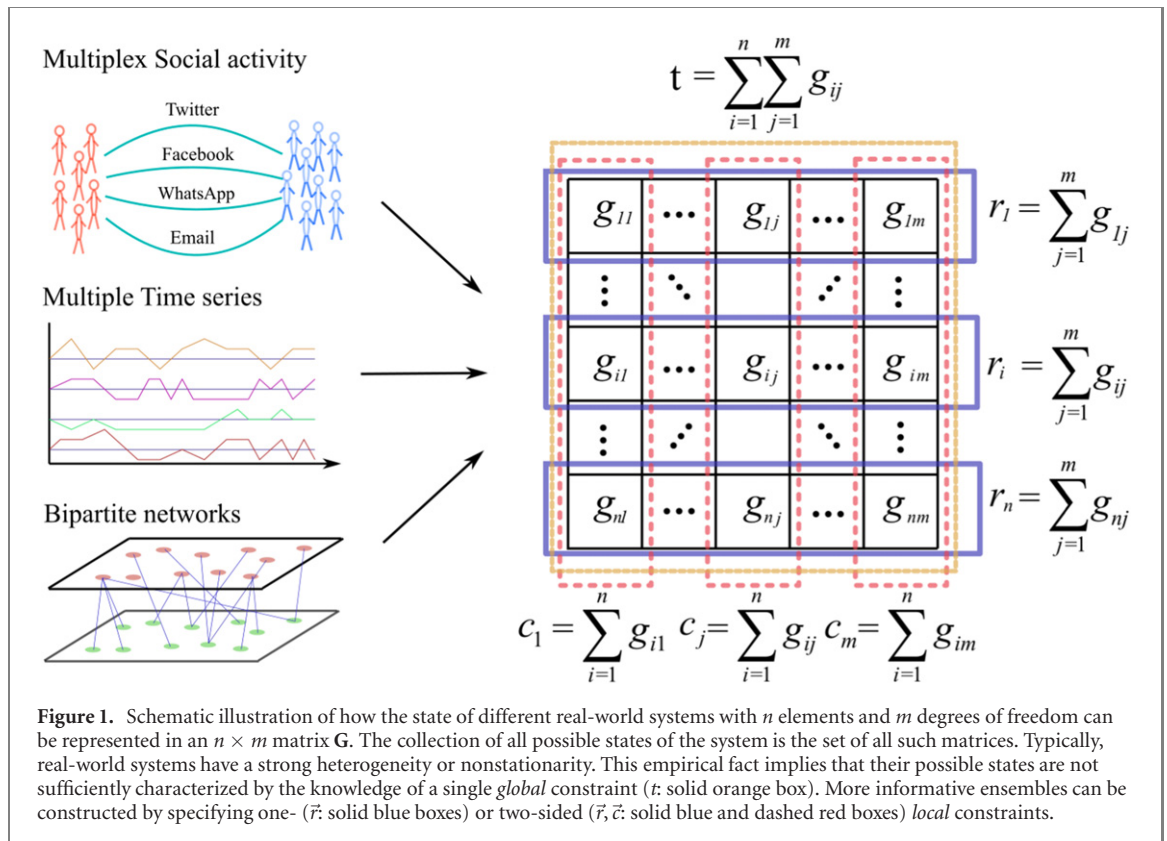
2. General formalism

2.1. Matrix ensembles
A discrete $n \times m$ matrix ensemble is a set $\mathcal{G}$ of available configurations for an $n \times m$ integer-valued matrix $\mathbf{G}$, endowed with a suitably chosen probability distribution $P(\mathbf{G})$ over such configurations, such that $\sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) = 1$. An entry of the matrix is denoted by g_{ij} (with $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$). We distinguish two main cases, the *binary* case where g_{ij} takes one of the two values $\{0, 1\}$ and the *weighted* case where g_{ij} takes a value in the set $\{0, 1, 2, \ldots\}$ of non-negative integer values. The number n of rows in each matrix represents the number of elements (i.e. the *size*) of the system being modelled. The number m of columns represents instead the number of state variables, or degrees of freedom, for each element.

In general, each matrix $\mathbf{G}$ can represent one of the possible states of a (large) real-world system. For instance, $\mathbf{G}$ may represent the realization of a multivariate time series, where m is the number of time series (e.g. brain regions, financial stocks, etc) and n is the number of time steps during which the signals are recorded. $\mathbf{G}$ may also represent a multi-cell array of gene expression profiles, where m is the number of cells and n the number of genes for which expression levels are being measured. Similarly, $\mathbf{G}$ may represent the state of a multi-channel communication systems, where n is the length of the sequences being transmitted from sender to receiver (in information theory, such length defines the 'size' of the communication process) and m is the number of channels. Finally, $\mathbf{G}$ may represent the adjacency matrix of a *bipartite graph*, where n is the number of nodes in the layer of interest (e.g. people in a co-affiliation network), while m is the number of possible dimensions where nodes can co-occur (e.g. work, family, sport, friendship, etc). In the special case $m = n$, the network can also be interpreted as a (binary or weighted) *directed unipartite graph*, i.e. one where there is a single set of n nodes that can be linked to each other via directed edges (note that, by contrast, undirected unipartite graphs are associated with a symmetric adjacency matrix, a property that we do not enforce in this paper; the nonequivalence of ensembles of binary or weighted undirected graphs with given constraints has been studied previously in [5, 8, 13, 14]).

2.2. Global and local constraints
In each of the examples mentioned above, the 'true' microscopic configuration (or *microstate*) of the system can be uniquely represented by a specific 'empirical' matrix $\mathbf{G}^*$ in the set $\mathcal{G}$ of all possible states. A schematic illustration is shown in figure 1. As ordinary in statistical mechanics, when the size of the system is large one no longer focuses on the specific microstate $\mathbf{G}^*$ (which becomes not empirically accessible), but rather on the *macrostate* defined by a collection of microscopic configurations compatible with the empirical value $\vec{C}^* \equiv \vec{C}(\mathbf{G}^*)$ of a certain observable quantify $\vec{C}(\mathbf{G})$ playing the role of a constraint. The choice of $\vec{C}(\mathbf{G})$ determines the probability distribution $P(\mathbf{G})$ over $\mathcal{G}$ conditional on our knowledge of $\vec{C}^*$. In other words, it determines how our estimate of the microstate of the system concentrates around the

Figure 1. Schematic illustration of how the state of different real-world systems with n elements and m degrees of freedom can be represented in an $n \times m$ matrix **G**. The collection of all possible states of the system is the set of all such matrices. Typically, real-world systems have a strong heterogeneity or nonstationarity. This empirical fact implies that their possible states are not sufficiently characterized by the knowledge of a single *global* constraint (t: solid orange box). More informative ensembles can be constructed by specifying one- ($\vec{r}$: solid blue boxes) or two-sided ($\vec{r}, \vec{c}$: solid blue and dashed red boxes) *local* constraints.

compatible configurations once we observe $\vec{C}^*$. Intuitively, before anything is observed, $P(\mathbf{G})$ is uniform on $\mathcal{G}$.

In ordinary statistical physics, the quantity $\vec{C}(\mathbf{G})$ is typically scalar (e.g. the total energy) or at most low-dimensional (e.g. a vector containing the total energy and the total number of particles), reflecting a few global conservation laws applying to a large homogeneous system at thermodynamic equilibrium. However, in models of complex systems $\vec{C}(\mathbf{G})$ can be high-dimensional, as it may encode a large number of local constraints reflecting separate conservation laws imposed by spatial heterogeneity and/or temporal non-stationarity. For instance, if $\mathbf{G}^*$ is the observed configuration of a complex network with n nodes (i.e. the empirical $n \times n$ adjacency matrix), it is well known that the knowledge of purely global properties such as the overall number of links is insufficient in order to produce a statistical ensemble of networks with properties similar to those found in $\mathbf{G}^*$. Indeed, enforcing only the total number of links produces the popular Erdős–Rényi random graph model, whose topological properties are way too homogenous as compared with those of real-world networks. By contrast, if the number of links of each node is enforced separately (as in the so-called *configuration model*), the resulting ensemble of graphs is found to successfully replicate many higher-order empirical topological properties [20]. As another example, if $\mathbf{G}^*$ represents a set of synchronous time series or sequences produced by m sources of a non-stationary system observed over n time steps, then the statistical properties of these time series will change over time. As a result, overall time-independent constraints will not be enough in order to produce ensembles of multivariate time series with properties close to those of $\mathbf{G}^*$, and time-dependent (i.e. local in time) constraints will in general be needed [16]. In this case, n is the number of local (in time) constraints to consider.

In our setting, we consider the general case where the distribution $P(\mathbf{G})$ defining the (binary or weighted) matrix ensemble is induced by a K-dimensional vector $\vec{C}(\mathbf{G})$ of *constraints* imposed on the matrices. We will assume that the K constraints are all non-redundant, e.g. they are not trivial copies or linear combinations of each other [7]. We will consider both *global* and *local* constraints. As global constraint we will consider the scalar quantity $t(\mathbf{G})$ defined as the total sum of all the entries of the matrix $\mathbf{G}$, i.e. $t(\mathbf{G}) = \sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij}$. The number of constraints in this case is $K = 1$ and the 'empirical' value of t will be denoted as $t^* \equiv t(\mathbf{G}^*)$. As local constraints, we will consider two possibilities: *one-sided* local constraints and *two-sided* local constraints. A one-sided local constraint is the n-dimensional vector $\vec{r}(\mathbf{G})$ where the entry $r_i(\mathbf{G}) = \sum_{j=1}^{m} g_{ij}$ ($i = 1, n$) represents the sum of the entries of the matrix $\mathbf{G}$ along its ith row. The number of constraints is in this case $K = n$ and the empirical value of $\vec{r}$ will be denoted as $\vec{r}^* \equiv \vec{r}(\mathbf{G}^*)$. A two-sided local constraint is a pair of vectors $(\vec{r}(\mathbf{G}), \vec{c}(\mathbf{G}))$, where $\vec{r}(\mathbf{G})$ is still the n-dimensional vector representing the n row sums of $\mathbf{G}$, while $\vec{c}(\mathbf{G})$ is the m-dimensional vector

representing the m column sums of $\mathbf{G}$, i.e. where each entry $c_j(\mathbf{G}) = \sum_{i=1}^{n} g_{ij}$ $(j = 1, m)$ is the sum of the entries of $\mathbf{G}$ along its jth column. The number of constraints is therefore $K = n + m$ and the empirical value of the pair $(\vec{r}, \vec{c})$ will be denoted as $(\vec{r}^*, \vec{c}^*) \equiv (\vec{r}(\mathbf{G}^*), \vec{c}(\mathbf{G}^*))$. A visual illustration of these constraints for possible data structures of practical interest is shown in figure 1.

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

2.3. Soft constraints: canonical ensemble

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

$$\langle \vec{C} \rangle \equiv \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \vec{C}(\mathbf{G}), \tag{1}$$

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

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

$$S_n[P] \equiv -\sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \ln P(\mathbf{G}), \tag{2}$$

(where the subscript n indicates that the entropy is calculated for given n), subject to the condition $\langle \vec{C} \rangle = \vec{C}^*$. The result [22] of this constrained maximization problem is the parametric solution

$$P_{\text{can}}(\mathbf{G}|\vec{\theta}) = \frac{e^{-H(\mathbf{G}, \vec{\theta})}}{Z(\vec{\theta})}, \tag{3}$$

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

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

$$P_{\text{can}}^*(\mathbf{G}) \equiv P_{\text{can}}(\mathbf{G}|\vec{\theta}^*), \tag{4}$$

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

$$\langle \vec{C} \rangle_{\vec{\theta}^*} = \vec{C}^*, \tag{5}$$

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

$$\langle \vec{C} \rangle_{\vec{\theta}} = \sum_{\mathbf{G} \in \mathcal{G}} P_{\text{can}}(\mathbf{G}|\vec{\theta}) \vec{C}(\mathbf{G}). \tag{6}$$

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

$$\mathcal{L}^*(\vec{\theta}) \equiv \ln P_{\text{can}}(\mathbf{G}^*|\vec{\theta}), \tag{7}$$

where $\mathbf{G}^*$ is the empirical configuration, or equivalently any configuration that realizes the empirical constraint exactly, i.e. such that $\vec{C}(\mathbf{G}^*) = \vec{C}^*$ [22]. The uniqueness of $\vec{\theta}^*$ follows whenever $\mathcal{L}^*(\vec{\theta})$ can be differentiated at least twice [7], as we confirm below for all the models considered in this paper.

Inserting equation (4) into equation (2), we obtain the value of the canonical entropy

$$S_{\text{can}}^* \equiv S_n[P_{\text{can}}^*] = -\mathcal{L}^*(\vec{\theta}^*) = -\ln P_{\text{can}}^*(\mathbf{G}^*), \tag{8}$$

where we have omitted the dependence on n to simplify the notation. The last equality is very useful, as it show that S_{can}^* can be calculated by simply evaluating $P_{\text{can}}^*(\mathbf{G})$ on the single configuration $\mathbf{G}^*$ [7].

2.4. Hard constraints: microcanonical ensemble

In the case of hard constraints, one requires that each individual configuration realizes the value $\vec{C}^*$. This means that the 'soft' constraint in equation (5) is replaced by the much stricter constraint

$$\vec{C}(\mathbf{G}) = \vec{C}^*, \tag{9}$$

for each allowed configuration $\mathbf{G}$. The microcanonical probability P_{mic} is found by enforcing this stronger requirement, while still maximizing the entropy $S_n[P]$ defined in equation (2). The result is the uniform distribution

$$P_{\text{mic}}^*(\mathbf{G}) = \begin{cases} \Omega_{\vec{C}^*}^{-1} & \vec{C}(\mathbf{G}) = \vec{C}^* \\ 0 & \vec{C}(\mathbf{G}) \neq \vec{C}^*, \end{cases} \tag{10}$$

where $\Omega_{\vec{C}^*}$ is the number of configurations in $\mathcal{G}$ realizing the 'hard' constraint in equation (9). The corresponding microcanonical entropy is obtained by inserting equation (10) into equation (2):

$$S_{\text{mic}}^* \equiv S_n[P_{\text{mic}}^*] = \ln \Omega_{\vec{C}^*}, \tag{11}$$

which is also known as Boltzmann entropy.

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

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

2.5. The thermodynamic limit

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

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

- m remains finite as $n \to \infty$: in this case, we have $m = O(1)$ where $O(x)$ indicates a quantity that has a finite limit if divided by x as $n \to +\infty$, i.e. $O(x)$ is asymptotically of the same leading order[4] as x;

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

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

---

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

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

- We define *sparse matrices* those for which each of the m column sums (irrespective of whether such sums are chosen as constraints) is *finite* in the thermodynamic limit, i.e. $c_j^* = O(1)$ $(j = 1, \ldots, m)$. This implies that, in the canonical ensemble, the expected value of any entry g_{ij} of the matrix $\mathbf{G}$ is on average $O(1/n)$; correspondingly, in the microcanonical ensemble the allowed matrices are dominated by zeroes (whence the name 'sparse matrices'). Note that for the row sums one has $r_i^* = O(m/n)$ for all i. If m grew slower than n, these row sums would vanish as $n \to +\infty$, which would imply that asymptotically no microcanonical configuration would realize the local constraints. Since we require $\Omega_{\vec{C}^*} > 0$ (see above), this means that in the sparse case we necessarily need $m = O(n)$. Consequently, $r_i^* = O(1)$, $r_i^*/m = O(1/n)$ $(i = 1, \ldots, n)$, $t^* = O(n)$, and $t^*/mn = O(1/n)$.
- By contrast, we define *dense matrices* those for which each of the m column sums (again, irrespective of whether they are chosen as constraints) *diverges* proportionally to n in the thermodynamic limit, i.e. $c_j^* = O(n)$ $(j = 1, \ldots, m)$. In the canonical ensemble, the expected value of g_{ij} is therefore $O(1)$, which makes the allowed matrices in the corresponding microcanonical ensemble 'dense'. The row sums are now $r_i^* = O(m)$ and we have $r_i^*/m = O(1)$ (for all i) and $t^*/mn = O(1)$. In this case, we consider m as either remaining finite, in which case we have $m = O(1)$, $r_i^* = O(1)$ $(i = 1, \ldots, n)$ and $t^* = O(n)$, or diverging proportionally to n (see above), in which case we have $m = O(n)$, $r_i^* = O(n)$ $(i = 1, \ldots, n)$, and $t^* = O(n^2)$.
- Note that, in principle, in the weighted case we may even consider a sort of *superdense* regime where some of the individual entries of the matrix diverge in the thermodynamic limit. This possibility is related to a Bose–Einstein condensation concentrating a finite fraction of the total weight t^* of the matrix in a finite number of entries [8]. However, we will not consider this extreme case here for simplicity, as it would not arise in most real-world applications.

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

2.6. Ensemble (non)equivalence

There are various ways to mathematically define the property of ensemble (non)equivalence. These include the notions of EE in the *thermodynamic*, *macrostate* and *measure* sense which, under mild assumptions, can be proven to be equivalent [6]. We will adopt the definition in the measure sense, which states that the ensembles are equivalent if the relative entropy

$$S_n[P_{\text{mic}}^* \| P_{\text{can}}^*] \equiv \sum_{\mathbf{G} \in \mathcal{G}} P_{\text{mic}}^*(\mathbf{G}) \ln \frac{P_{\text{mic}}^*(\mathbf{G})}{P_{\text{can}}^*(\mathbf{G})}, \tag{12}$$

(which is the Kullback–Leibler divergence for given n between the microcanonical and canonical entropies and is guaranteed to be non-negative [23]), when rescaled by n, vanishes in the thermodynamic limit [6], i.e. if the *specific relative entropy* vanishes:

$$s[P_{\text{mic}}^* \| P_{\text{can}}^*] \equiv \lim_{n \to +\infty} \frac{S_n[P_{\text{mic}}^* \| P_{\text{can}}^*]}{n} = 0, \tag{13}$$

or equivalently

$$S_n[P_{\text{mic}}^* \| P_{\text{can}}^*] = o(n), \tag{14}$$

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

Importantly, it can be shown [5, 7] that

$$S_n[P_{\text{mic}}^* \| P_{\text{can}}^*] = \ln \frac{P_{\text{mic}}^*(\mathbf{G}^*)}{P_{\text{can}}^*(\mathbf{G}^*)} = S_{\text{can}}^* - S_{\text{mic}}^*. \tag{15}$$

The inequality $S_n[P_{\text{mic}}^* \| P_{\text{can}}^*] \geqslant 0$, which is a general property of the relative entropy, implies therefore $S_{\text{can}}^* \geqslant S_{\text{mic}}^*$ and indicates the presence of an 'extra entropy' in the canonical ensemble. This extra entropy is due to the fact that, while in the microcanonical ensemble the constraint $\vec{C}$ is a deterministic quantity fixed to the value $\vec{C}^*$ through the hard constraint introduced in equation (9), in the canonical ensemble it is a random variable fluctuating around the expected value $\vec{C}^*$ as dictated by the soft constraint defined in

equation (5). With respect to the canonical ensemble, the hardness of the constraint in the microcanonical ensemble implies additional dependencies (i.e. smaller entropy) among the entries of **G**.

The definition of EE in equation (13) states that, if the extra entropy $S_n[P^*_{\text{mic}}\|P^*_{\text{can}}]$, once divided by n, vanishes in the thermodynamic limit, then the ensembles are equivalent. We can reformulate this definition of EE in terms of the large deviations of the canonical ensemble as follows. We may wonder what is the probability $Q^*_{\text{can}}(\vec{C}^*)$ to observe, in the canonical ensemble with soft constraints, the 'microcanonical macrostate' characterized by the hard value $\vec{C}^*$ of the constraint. This macrostate is the set of all microscopic configurations **G** with given value $\vec{C}(\mathbf{G}) = \vec{C}^*$, i.e. all the $\Omega_{\vec{C}^*}$ allowed microstates in the microcanonical ensemble. Since each such microstate occurs with the same probability $P^*_{\text{can}}(\mathbf{G}^*)$ in the canonical ensemble, the canonical probability of the microcanonical macrostate is

$$Q^*_{\text{can}}(\vec{C}^*) = \Omega_{\vec{C}^*} P^*_{\text{can}}(\mathbf{G}^*) = \frac{P^*_{\text{can}}(\mathbf{G}^*)}{P^*_{\text{mic}}(\mathbf{G}^*)}, \tag{16}$$

where we have used $P^*_{\text{mic}}(\mathbf{G}^*) = \Omega_{\vec{C}^*}^{-1}$ from equation (10). Taking logarithms and using equation (15), we obtain

$$\ln Q^*_{\text{can}}(\vec{C}^*) = -\ln \frac{P^*_{\text{mic}}(\mathbf{G}^*)}{P^*_{\text{can}}(\mathbf{G}^*)} = -S_n[P^*_{\text{mic}}\|P^*_{\text{can}}], \tag{17}$$

which generalizes a similar result for the case of binary graphs with given degree sequence [13]. Dividing the above expression by n and taking the thermodynamic limit, we get

$$\lim_{n\to\infty} \frac{\ln Q^*_{\text{can}}(\vec{C}^*)}{n} = -s_n[P^*_{\text{mic}}\|P^*_{\text{can}}]. \tag{18}$$

In the language of large deviations theory, the above result states that the specific relative entropy $s_n[P^*_{\text{mic}}\|P^*_{\text{can}}]$ between the microcanonical and canonical ensembles coincides with the *rate function* for the probability of observing a certain macrostate $\vec{C}$ in the canonical ensemble, evaluated at the specific value $\vec{C} = \vec{C}^*$ corresponding to the microcanonical ensemble. Clearly, the criterion for EE in equation (13) means that, when EE holds, the rate function vanishes when evaluated at $\vec{C}^*$, i.e. $\ln Q^*_{\text{can}}(\vec{C}^*) = o(n)$. This means that the canonical ensemble concentrates on the microcanonically allowed microstates, as the probability of the microcanonical macrostate approaches one and its logarithm approaches zero. When EE does not hold, the rate function for the microcanonical macrostate does not vanish and the above concentration of the canonical probability does not occur, which means that there are necessarily other macrostates over which the canonical probability concentrates.

From equation (15) it is clear that equation (13) is equivalent to the condition

$$\lim_{n\to+\infty} \frac{S_n[P^*_{\text{can}}] - S_n[P^*_{\text{mic}}]}{n} = 0 \tag{19}$$

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

$$S^*_{\text{mic}} = S^*_{\text{can}} - o(n). \tag{20}$$

This implies

$$\Omega_{\vec{C}^*} = e^{S^*_{\text{can}} - o(n)}, \tag{21}$$

i.e. $\Omega_{\vec{C}^*}$ is approximated by $e^{S^*_{\text{can}}}$ up to a subexponential (in n) correction factor. The above asymptotics is used in statistical mechanics textbooks whenever the property of EE is believed to hold, i.e. in absence of phase transitions or long-range interactions. When EE does not hold, equation (21) breaks down. In this case, the extra entropy in the canonical ensemble grows at least as fast as n. Recent research has shown that this breakdown can happen even in complete absence of phase transitions, hence also in situations where EE was typically believed to hold. Here we are going to show that, additionally, the breakdown can occur with previously undocumented strength, i.e. the extra entropy can grow as fast as the entropy itself.

Combining equations (11) and (15), one obtains the following exact generalization of equation (21), valid irrespective of whether EE holds:

$$\Omega_{\vec{C}^*} = e^{S^*_{\text{can}} - S_n[P^*_{\text{mic}}\|P^*_{\text{can}}]}. \tag{22}$$

Clearly, the above expression reduces to equation (21) in case of EE, i.e. when equation (14) holds. Although exact, equation (22) is not very useful unless one can calculate $S_n[P^*_{\text{mic}}\|P^*_{\text{can}}]$ explicitly. An

equivalent exact expression, which only requires the knowledge of P_{can}^* and is again valid even when EE does not hold, has been derived [7]:

$$\Omega_{\vec{C}^*} = \sum_{\mathbf{G} \in \mathcal{G}} \int_{-\vec{\pi}}^{\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} e^{i\vec{\psi}[\vec{C}^* - \vec{C}(\mathbf{G})]}$$

$$= \int_{-\vec{\pi}}^{+\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} P_{can}^{-1}(\mathbf{G}^* | \vec{\theta}^* + i\vec{\psi}), \tag{23}$$

(where $\int_{-\vec{\pi}}^{+\vec{\pi}} \mathrm{d}\vec{\psi} \equiv \prod_{k=1}^K \int_{-\pi}^{+\pi} \mathrm{d}\psi_k$). We will confirm that the above expression provides the exact result in cases where the complex integral can be calculated explicitly and $\Omega_{\vec{C}^*}$ can be evaluated independently via combinatorial enumeration. Indeed, equation (23) highlights a beautiful connection between canonical and microcanonical probabilities through an extension to complex numbers.

When the integral in equation (23) cannot be calculated directly, it is still possible to use a saddle-point technique leading to [7]

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

$$= e^{S_{can}^*} \prod_{k=1}^K \frac{1 + O(1/\lambda_k^*)}{\sqrt{2\pi \lambda_k^*}}, \tag{24}$$

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

$$\Sigma_{ij}^* = \Sigma_{ij}\big|_{\vec{\theta} = \vec{\theta}^*}, \tag{25}$$

with

$$\Sigma_{ij} \equiv -\frac{\partial^2 \mathcal{L}^*(\vec{\theta})}{\partial \theta_i \partial \theta_j}$$

$$= \frac{\partial^2 \ln Z(\vec{\theta})}{\partial \theta_i \partial \theta_j}$$

$$= \langle C_i C_j \rangle_{\vec{\theta}} - \langle C_i \rangle_{\vec{\theta}} \langle C_j \rangle_{\vec{\theta}}$$

$$= \mathrm{Cov}_{\vec{\theta}}[C_i, C_j], \tag{26}$$

and $\{\lambda_k^*\}_{k=1}^K$ are the eigenvalues of $\boldsymbol{\Sigma}^*$. We recall that covariance matrices are positive-semidefinite, so all their eigenvalues are non-negative. If λ_k^* is finite, then the quantity $O(1/\lambda_k^*)$ in equation (24) cannot in general be calculated explicitly, although it generates a correction that does not change the leading order of $\Omega_{\vec{C}^*}$ and S_{mic}^*. If λ_k^* is infinite (i.e., if it diverges in the thermodynamic limit), then $O(1/\lambda_k^*)$ will vanish asymptotically and we have $1 + O(1/\lambda_k^*) = 1 + o(1)$. This implies that, if all the eigenvalues of $\boldsymbol{\Sigma}^*$ diverge, then equation (24), when inserted into certain expressions, can lead to an exact result. This includes the case of local constraints, for which K diverges in the thermodynamic limit. We will therefore discuss the asymptotic behaviour of the eigenvalues of $\boldsymbol{\Sigma}^*$ in each of the examples considered later.

Equation (24) generalizes equation (21) to the case where EE does not necessarily hold. Note that our initial assumption that the K constraints are non-redundant implies that $\lambda_k^* > 0$ for all k, i.e. $\boldsymbol{\Sigma}^*$ is positive-definite [7]. Keeping this assumption also in the thermodynamic limit (as ensured by our choice of both global and local constraints defined above), we note two consequences. First, since equation (26) shows that $\boldsymbol{\Sigma}^*$ is the Hessian matrix of second derivatives of $-\mathcal{L}^*(\vec{\theta})$, the fact that $\boldsymbol{\Sigma}^*$ is positive-definite implies that $\vec{\theta}^*$ is a unique global maximum for $\mathcal{L}^*(\vec{\theta})$ [7], confirming what we had anticipated previously. Second, the product in equation (24) is at most of the same order as the denominator. Therefore, in full generality, we can exploit equation (24) to rewrite equation (22) as

$$\Omega_{\vec{C}^*} = e^{S_{can}^* - O(\alpha_n)}, \tag{27}$$

where we have defined [7]

$$\alpha_n \equiv \ln \sqrt{\det(2\pi \boldsymbol{\Sigma}^*)} = \frac{1}{2} \sum_{k=1}^K \ln(2\pi \lambda_k^*). \tag{28}$$

We can now make three important considerations. First, equation (27) means that

$$S_n[P_{mic}^* \| P_{can}^*] = O(\alpha_n), \tag{29}$$

showing that the speed of growth of $S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}]$ with n can be calculated explicitly through equation (28) using the knowledge of $\boldsymbol{\Sigma}^*$, which requires only the canonical ensemble. This is useful when microcanonical calculations are unfeasible. Second, if K is finite, or if K diverges but all (except possibly a finite number of) the eigenvalues of $\boldsymbol{\Sigma}^*$ diverge, then the product inside equation (24) gives a subleading contribution to $S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}]$, which therefore has the same asymptotic behaviour as α_n:

$$S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}] = \alpha_n[1 + o(1)]. \tag{30}$$

This result, which is stronger than equation (29), means that in such a case one can obtain exact estimates of quantities that depend on $S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}]$, using only the knowledge of α_n. Third, equation (29) shows that the definition of EE given by equation (14) coincides with

$$\alpha_n = o(n), \tag{31}$$

which, again, can be ascertained by evaluating only $\boldsymbol{\Sigma}^*$ and avoiding any microcanonical calculation. Indeed, equation (31) can be formulated as an equivalent definition of EE in the measure sense [7]. If α_n grows faster than $o(n)$, then the system is under EN.

3. Weak and strong EN

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

3.1. Relative entropy ratio

Equation (24) reveals that the asymptotic behaviour of $\Omega_{\vec{C}^*}$ depends on that of K and of the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}^*$. We can indeed convince ourselves of this fact by looking at results of previous studies from a novel perspective.

Specifically, if $K = o(n)$ and if we exclude phase transitions, then equation (24) leads to equation (21), i.e. the ensembles are equivalent. This includes the traditional situation where one has a finite number of constraints, as well as more complicated cases where the number of constraints is subextensive (e.g. random graphs with constraints on a subextensive subset of node degrees [14]). In order to break EE in this case, one needs phase transitions corresponding to singularities of the partition function [6]. For instance, in the case of graphs with fixed numbers of edges and triangles (or wedges) [9], there is a region in parameter space where one gets $S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}] = O(n^2)$ and therefore $\Omega_{\vec{C}^*} = e^{S^*_{\mathrm{can}} - O(n^2)}$. Since also S^*_{can} and S^*_{mic} are $O(n^2)$ in this case, it follows that

$$S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}] = O(S^*_{\mathrm{can}}), \tag{32}$$

(note that in general $S^*_{\mathrm{can}} \geqslant S^*_{\mathrm{mic}}$ due to the non-negativity of the Kullback–Leibler divergence and to equation (15), therefore $O(S^*_{\mathrm{can}})$ is necessarily the leading order). This is what we have previously referred to as a form of EN that is 'restricted' (i.e. valid only in a certain region in parameter space arising from a phase transition and outside which EE is restored) but 'strong' (i.e. where the relative entropy is of the same order as the entropy itself).

If $K = O(n)$, then equation (21) is in general no longer valid. For instance, in the case of *sparse* graphs with fixed degrees ($K = n$), all the eigenvalues of $\boldsymbol{\Sigma}^*$ are finite in the thermodynamic limit [5, 13]; one indeed obtains $S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}] = O(n)$ [5] and hence $\Omega_{\vec{C}^*} = e^{S^*_{\mathrm{can}} - O(n)}$. Note that in this case the product in equation (24) (which in general cannot be calculated exactly) is of the same order as the denominator and should be taken into account. In the case of *dense* graphs with fixed degrees (again $K = n$), all the eigenvalues of $\boldsymbol{\Sigma}^*$ are instead $O(n)$ [13]; one indeed obtains $S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}] = O(n \ln n)$ [13] and hence $\Omega_{\vec{C}^*} = e^{S^*_{\mathrm{can}} - O(n \ln n)}$. The product in equation (24) is in this case negligible with respect to the denominator, which can be calculated exactly. In any case, since S^*_{can} and S^*_{mic} are still $O(n^2)$ for both sparse and dense networks with fixed degrees, these situations correspond to

$$S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}] = o(S^*_{\mathrm{can}}), \tag{33}$$

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

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

$$R_n \equiv \frac{S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}]}{S_n[P^*_{\mathrm{can}}]} = 1 - \frac{S_n[P^*_{\mathrm{mic}}]}{S_n[P^*_{\mathrm{can}}]}, \tag{34}$$

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

$$\begin{aligned} R_\infty &\equiv \lim_{n\to\infty} R_n \\ &= \lim_{n\to\infty} \frac{S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}]}{S_n[P^*_{\mathrm{can}}]} \\ &= 1 - \lim_{n\to\infty} \frac{S_n[P^*_{\mathrm{mic}}]}{S_n[P^*_{\mathrm{can}}]}. \end{aligned} \tag{35}$$

For brevity, we will call R_n the *relative entropy ratio* and R_∞ the *limiting relative entropy ratio*. Note that the inequality $S^*_{\mathrm{can}} \geqslant S^*_{\mathrm{mic}} \geqslant 0$ implies $0 \leqslant R_n \leqslant 1$ for all $n > 0$. The condition characterizing our notion of strong EN in equation (32) coincides with R_∞ being strictly positive. The value of R_∞ in that case quantifies exactly the asymptotic proportionality between $S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}]$ and $S_n[P^*_{\mathrm{can}}]$, which is otherwise left unquantified by equation (32) alone. We will therefore adopt the strict inequality

$$R_\infty > 0, \tag{36}$$

(which in turns implies the breakdown of equation (31), the converse being in general not true) as our definition of *strong EN*. By contrast, the condition characterizing our notion of weak EN in equation (33) can be rephrased as the equality $R_\infty = 0$. Note that one may have $R_\infty = 0$ also in cases where the ensembles are equivalent. We will therefore adopt the condition $R_\infty = 0$, *in conjunction with the violation of* equation (31), as our definition of *weak EN*. Note that our discussion following equation (24) implies that, if all but at most a finite number of the eigenvalues of $\boldsymbol{\Sigma}^*$ diverge, then the exact value of R_∞ can be retrieved by replacing $S_n[P^*_{\mathrm{mic}}\|P^*_{\mathrm{can}}]$ with α_n given by equation (28), i.e. using only the canonical covariances between the constraints, without microcanonical calculations.

Note that equation (22) implies

$$\Omega_{\vec{C}^*} = e^{S^*_{\mathrm{can}}(1-R_n)} = O\left(\left(e^{S^*_{\mathrm{can}}}\right)^{1-R_\infty}\right). \tag{37}$$

So, in presence of strong nonequivalence ($R_\infty > 0$), $\Omega_{\vec{C}^*}$ is of strictly smaller order compared with the ordinary estimate in equation (21). This is actually due to the canonical ensemble having much bigger entropy than the microcanonical one: indeed, equation (15) implies

$$S^*_{\mathrm{mic}} = S^*_{\mathrm{can}}(1 - R_n), \tag{38}$$

and, inverting,

$$S^*_{\mathrm{can}} = \frac{1}{1-R_n} S^*_{\mathrm{mic}}. \tag{39}$$

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

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

3.2. Global constraints

As already discussed, ensembles of (binary or weighted) $n \times m$ matrices with a global constraint are defined by requiring that the single quantity $t(\mathbf{G}) = \sum_{i=1}^n \sum_{j=1}^m g_{ij}$ takes, either 'hardly' or 'softly', a specific value $t^* \equiv t(\mathbf{G}^*)$. For this simple choice of the constraint, both S^*_{can} and S^*_{mic} can be calculated exactly. This allows us to check that the complex integral in equation (23) indeed provides the exact value of Ω_{t^*}. Moreover, we can confirm the correctness of the asymptotic formula in equation (24). All these approaches show that for both binary and weighted matrices with a global constraint the canonical and microcanonical ensembles are equivalent.

3.2.1. Binary matrices under a global constraint

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

$$P_{\text{can}}(\mathbf{G}^*|\theta) = \frac{e^{-\theta\, t^*}}{(1 + e^{-\theta})^{mn}}, \tag{40}$$

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

$$\theta^* = \ln \frac{mn - t^*}{t^*}. \tag{41}$$

Using equation (8), we can then easily evaluate S_{can}^* from equations (40) and (41) as

$$S_{\text{can}}^* = -\ln P_{\text{can}}(\mathbf{G}^*|\theta^*) = \ln \frac{(mn)^{mn}}{(t^*)^{t^*}(mn - t^*)^{mn-t^*}}. \tag{42}$$

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

$$S_{\text{mic}}^* = \ln \Omega_{t^*} = \ln \binom{mn}{t^*}. \tag{43}$$

Importantly, it is possible to confirm that, upon extending the argument of the likelihood to the complex domain and calculating $P_{\text{can}}(\mathbf{G}^*|\theta^* + i\psi)$, the integral formula in equation (23) returns a value of Ω_{t^*} that produces the exact value of the microcanonical entropy S_{mic}^* given in equation (43):

$$\begin{aligned}
S_{\text{mic}}^* &= \ln \int_{-\pi}^{+\pi} \frac{d\psi}{2\pi} P_{\text{can}}^{-1}(\mathbf{G}^*|\theta^* + i\psi) \\
&= \ln \int_{-\pi}^{+\pi} \frac{d\psi}{2\pi} \frac{(1 + e^{-\theta^* - i\psi})^{mn}}{e^{-(\theta^* + i\psi)\, t^*}} \\
&= \ln \binom{mn}{t^*},
\end{aligned} \tag{44}$$

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

Combining the expressions for S_{mic}^* and S_{can}^* into equation (15), we obtain the relative entropy between the two ensembles:

$$S_n[P_{\text{mic}}^* \| P_{\text{can}}^*] = \ln \frac{(mn)^{mn}}{\binom{mn}{t^*}(t^*)^{t^*}(mn - t^*)^{mn-t^*}}. \tag{45}$$

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

$$p(g_{ij}|\theta^*) = \frac{e^{-\theta^* g_{ij}}}{1 + e^{-\theta^*}}, \quad g_{ij} \in \{0, 1\}, \tag{46}$$

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

$$\Sigma^* = \text{Var}_{\theta^*}[t] = nm \frac{e^{-\theta^*}}{(1 + e^{-\theta^*})^2} = t^*\left(1 - \frac{t^*}{mn}\right), \tag{47}$$

(see appendix), where Σ^* is the only (recall that here $K = 1$) of the covariance matrix $\boldsymbol{\Sigma}^*$ introduced in equation (26).

As discussed in subsection 2.6, $S_n[P_{\text{mic}}^* \| P_{\text{can}}^*]$ and Σ^* are asymptotically related through equation (29), and the (non)equivalence of canonical and microcanonical ensembles is decided by the asymptotic behaviour of these two quantities. We will confirm both results in the particular case under consideration here. However, for compactness, we do this in conjunction with the weighted case, after introducing the latter below.

3.2.2. Weighted matrices under a global constraint

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

$$P_{\text{can}}(\mathbf{G}^*|\theta) = e^{-\theta\, t^*}(1 - e^{-\theta})^{mn}, \tag{48}$$

and is maximised by the parameter value

$$\theta^* = \ln \frac{mn + t^*}{t^*},$$ (49)

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

$$S_{\mathrm{can}}^* = -\ln P_{\mathrm{can}}(\mathbf{G}^* | \theta^*) = \ln \frac{(mn + t^*)^{mn + t^*}}{(t^*)^{t^*}(mn)^{mn}}.$$ (50)

In the microcanonical ensemble, the number Ω_{t^*} of configurations realizing the hard constraint $t(\mathbf{G}) = t^*$ coincides with the number of so-called *weak compositions* of the positive integer t^* into exactly mn parts, i.e. the number of ways of writing the positive integer t^* as the sum of an ordered sequence of mn non-negative integers (note that two sequences that differ in the order of their terms represent different configurations). This number is given by the *negative binomial* coefficient $\Omega_{t^*} = \binom{mn + t^* - 1}{t^*}$ [24], whence

$$S_{\mathrm{mic}}^* = \ln \Omega_{t^*} = \ln \binom{mn + t^* - 1}{t^*}.$$ (51)

In this case as well, one can confirm that the integration of the complex quantity $P_{\mathrm{can}}^{-1}(\mathbf{G}^* | \theta^* + i\psi)$ as specified in equation (23) produces precisely the same value of Ω_{t^*} used in equation (51) (see appendix), thus retrieving the exact entropy

$$\begin{aligned} S_{\mathrm{mic}}^* &= \ln \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi}{2\pi} P_{\mathrm{can}}^{-1}(\mathbf{G}^* | \theta^* + i\psi) \\ &= \ln \int_{-\pi}^{\pi} \frac{\mathrm{d}\psi}{2\pi} \frac{(1 - e^{-\theta^* - i\psi})^{-mn}}{e^{-(\theta^* + i\psi)t^*}} \\ &= \ln \binom{mn + t^* - 1}{t^*}. \end{aligned}$$ (52)

The relative entropy $S_n[P_{\mathrm{mic}}^* \| P_{\mathrm{can}}^*]$, calculated using equation (15), equals

$$S_n[P_{\mathrm{mic}}^* \| P_{\mathrm{can}}^*] = \ln \frac{(mn + t^*)^{mn + t^*}}{\binom{mn + t^* - 1}{t^*}(t^*)^{t^*}(mn)^{mn}}.$$ (53)

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

$$p(g_{ij} | \theta^*) = e^{-\theta^* g_{ij}}(1 - e^{-\theta^*}), \quad g_{ij} \in \{0, 1, 2, \dots\}.$$ (54)

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

$$\Sigma^* = \mathrm{Var}_{\theta^*}[t] = nm \frac{e^{-\theta^*}}{(1 - e^{-\theta^*})^2} = t^* \left(1 + \frac{t^*}{mn} \right),$$ (55)

(see appendix).

3.2.3. EE for matrices under a global constraint

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

$$\binom{k + l - 1}{l} = (-1)^l \binom{-k}{l}.$$ (56)

The above relation allows us to conveniently rewrite the relative entropy for the weighted case appearing in equation (53) as

$$S_n[P_{\mathrm{mic}}^* \| P_{\mathrm{can}}^*] = \ln \frac{(-mn)^{-mn}}{\binom{-mn}{t^*}(t^*)^{t^*}(-mn - t^*)^{-mn - t^*}}.$$ (57)

Upon comparison with the corresponding equation (45) valid in the binary case, we can express the relative entropy in general as

$$S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*] = \ln \frac{(\pm mn)^{\pm mn}}{\binom{\pm mn}{t^*}(t^*)^{t^*}(\pm mn - t^*)^{\pm mn - t^*}}, \tag{58}$$

where the superscript '+' applies to binary matrices (note that $t^* \leqslant mn$ in this case) and the superscript '−' applies to weighted matrices. Note that the expression for the weighted case can be formally retrieved by changing the sign of m in the expression valid for the binary case.

As we discussed in subsection 2.6, checking for (non)equivalence requires studying the asymptotic behaviour of the relative entropy. In this case, we can calculate the asymptotic behaviour of $S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*]$ explicitly from the exact expression given by equation (58). Note that, in both the sparse and dense case (see subsection 2.5), t^* and mn diverge in the thermodynamic limit. We can therefore apply Stirling's formula

$$k! = \sqrt{2\pi k}\left(\frac{k}{e}\right)^k [1 + o(1)], \tag{59}$$

to equation (58), which yields

$$S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*] = \frac{1}{2}\ln\left[2\pi t^*\left(1 - \frac{t^*}{\pm mn}\right)\right][1 + o(1)]. \tag{60}$$

For purely pedagogical reasons, we check that the asymptotic behaviour found above is consistent with the one we would retrieve by using the expansion in equation (24), which leads to equation (29) and reduces the problem of the calculation of $S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*]$ to that of its leading order α_n. To this end, we note that in this case the matrix $\boldsymbol{\Sigma}^*$, being a 1×1 matrix, coincides with its only eigenvalue

$$(\lambda^*)^{\pm} = t^*\left(1 - \frac{t^*}{\pm mn}\right), \tag{61}$$

where we have used equation (47) for binary (+) and equation (55) for weighted (−) matrices. Therefore

$$\begin{aligned}\alpha_n^{\pm} &= \ln\sqrt{2\pi(\lambda^*)^{\pm}} \\ &= \frac{1}{2}\ln\left[2\pi t^*\left(1 - \frac{t^*}{\pm mn}\right)\right],\end{aligned} \tag{62}$$

which has indeed the same leading order as equation (60), thereby confirming the correctness of the saddle-point calculation. As an even stronger result, we are under the conditions for which equation (30) holds, a relationship that can be confirmed by comparing equations (60) and (62). It should also be noted that, since t^* diverges in the thermodynamic limit, so does $(\lambda^*)^{\pm}$ and equation (24) leads to

$$\Omega_{t^*}^{\pm} = \frac{e^{S_n^{\pm}[P_{\text{can}}^*]}}{\sqrt{2\pi t^*\left(1 - \frac{t^*}{\pm mn}\right)}}[1 + o(1)], \tag{63}$$

which is precisely what we get by applying equation (59) to the binomial and negative binomial coefficients appearing in the exact expression for $\Omega_{t^*}^{\pm}$ in the binary and weighted case respectively.

As stated in equation (31), checking whether the ensembles are equivalent boils down to checking whether $\alpha_n = o(n)$. Note that the only effect of the asymptotic scaling of t^* is that the quantity $1 - t^*/(\pm mn)$ in equation (62) converges to 1 in the sparse case $t^*/mn = O(1/n)$ and to a different, but still finite and positive constant in the dense case $t^*/mn = O(1)$ (see subsection 2.5). Therefore in both cases we have $\alpha_n = O(\ln t^*) = O(\ln mn)$. This implies $\alpha_n = o(n)$ independently of the asymptotic behaviour of m. This result shows that, *in presence of a global constraint, both binary and weighted matrices are under EE, irrespective of the scaling of t^* and m.* This finding confirms, in a generalized setting, the result obtained for networks with a given total number of links [5]. Since EE is preserved, we avoid the calculation of the limiting relative entropy ratio R_{∞} defined in equation (35) in this case.

3.3. One-sided local constraints

We now consider ensembles of binary and weighted $n \times m$ matrices with one-sided local constraints, i.e. under the requirement that the n-dimensional vector $\vec{r}(\mathbf{G})$ with entries $r_i(\mathbf{G}) = \sum_{j=1}^m g_{ij}$ ($i = 1, n$) takes a specific value $\vec{r}^* \equiv \vec{r}(\mathbf{G}^*)$. Note that, unlike the case of global constraints, here the number of constraints is extensive. Nonetheless, it turns out that both S_{can}^* and S_{mic}^* can still be calculated exactly. Therefore we can again confirm the correctness of both the exact integral formula in equation (23) and the asymptotic expansion in equation (24). Despite these extensions are mathematically straightforward, we find a deep

physical difference with respect to the case with global constraints: the presence of an extensive number of local constraints implies the breaking of the equivalence of canonical and microcanonical ensembles for both binary and weighted matrices. The calculation of the limiting relative entropy ratio R_∞ allows us to quantify the strength of nonequivalence and also to identify the conditions leading to the 'strong and unrestricted' form.

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

$$P_{\mathrm{can}}(\mathbf{G}^*|\vec{\theta}) = \frac{e^{-\vec{\theta}\cdot\vec{r}^*}}{\prod_{i=1}^n (1 + e^{-\theta_i})^m}, \tag{64}$$

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

$$\theta_i^* = \ln \frac{m - r_i^*}{r_i^*} \quad i = 1, n, \tag{65}$$

corresponding to the soft constraint $\langle \vec{r} \rangle_{\vec{\theta}^*} = \vec{r}^*$. Substituting equation (65) into equation (64), we obtain the canonical entropy as

$$\begin{aligned} S_{\mathrm{can}}^* &= -\ln P_{\mathrm{can}}(\mathbf{G}^*|\vec{\theta}^*) \\ &= \sum_{i=1}^n \ln \frac{m^m}{(r_i^*)^{r_i^*}(m - r_i^*)^{m - r_i^*}}. \end{aligned} \tag{66}$$

Let us now turn to the microcanonical ensemble. Since the constraints are only one-sided, it is immediate to realize that the number $\Omega_{\vec{r}^*}$ of configurations realizing the hard constraint $\vec{r}(\mathbf{G}) = \vec{r}^*$ is a product of row-specific binomial coefficients, so that the microcanonical entropy S_{mic}^* can still be calculated exactly as the following simple generalization of equation (43):

$$S_{\mathrm{mic}}^* = \ln \Omega_{\vec{r}^*} = \ln \prod_{i=1}^n \binom{m}{r_i^*} = \sum_{i=1}^n \ln \binom{m}{r_i^*}. \tag{67}$$

For the same reason, S_{mic}^* can also be exactly retrieved by explicitly integrating the complex-valued quantity $P_{\mathrm{can}}(\mathbf{G}^*|\vec{\theta}^* + i\vec{\psi})$ as prescribed by equation (23) (see appendix):

$$\begin{aligned} S_{\mathrm{mic}}^* &= \ln \int_{-\vec{\pi}}^{+\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^n} P_{\mathrm{can}}^{-1}(\mathbf{G}^*|\vec{\theta}^* + i\vec{\psi}) \\ &= \ln \prod_{i=1}^n \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi_i}{2\pi} \frac{(1 + e^{-\theta_i^* - i\psi_i})^m}{e^{-(\theta_i^* + i\psi_i)r_i^*}} \\ &= \sum_{i=1}^n \ln \binom{m}{r_i^*}. \end{aligned} \tag{68}$$

Combining the above results, we can calculate the relative entropy from equation (15) as

$$S_n[P_{\mathrm{mic}}^* \| P_{\mathrm{can}}^*] = \sum_{i=1}^n \ln \frac{m^m}{\binom{m}{r_i^*}(r_i^*)^{r_i^*}(m - r_i^*)^{m - r_i^*}}. \tag{69}$$

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

$$p(g_{ij}|\vec{\theta}^*) = \frac{e^{-\theta_i^* g_{ij}}}{1 + e^{-\theta_i^*}}, \quad g_{ij} \in \{0, 1\}, \tag{70}$$

as we show in the appendix. Correspondingly, in the microcanonical ensemble $\vec{r}$ is a deterministic vector fixed to the value $\vec{r}^*$, while in the canonical ensemble it is a random vector with expected value $\vec{r}^*$. The

covariance matrix Σ^* between the entries of $\vec{r}$ (i.e. between the n constraints) in the canonical ensemble is a diagonal matrix with entries

$$\Sigma^*_{ij} = \delta_{ij} r^*_i \left(1 - \frac{r^*_i}{m} \right), \tag{71}$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$ (see appendix). This implies that the eigenvalues $\{\lambda^*_i\}^n_{i=1}$ of Σ^* are

$$\lambda^*_i = r^*_i \left(1 - \frac{r^*_i}{m} \right) \quad i = 1, n. \tag{72}$$

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

3.3.2. Weighted matrices under one-sided local constraints

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

$$P_{\text{can}}(\mathbf{G}^* | \vec{\theta}) = \frac{e^{-\vec{\theta} \cdot \vec{r}^*}}{\prod^n_{i=1}(1 - e^{-\theta_i})^{-m}}, \tag{73}$$

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

$$\theta^*_i = \ln \frac{m + r^*_i}{r^*_i} \quad i = 1, n, \tag{74}$$

realizing the soft constraint $\langle \vec{r} \rangle_{\vec{\theta}^*} = \vec{r}^*$ (see appendix). If we insert equation (74) into equation (73), we get

$$
\begin{aligned}
S^*_{\text{can}} &= -\ln P_{\text{can}}(\mathbf{G}^* | \vec{\theta}^*) \\
&= \sum^n_{i=1} \ln \frac{(m + r^*_i)^{m + r^*_i}}{(r^*_i)^{r^*_i} m^m}.
\end{aligned} \tag{75}
$$

The microcanonical entropy S^*_{mic} is instead given by the following generalization of equation (51):

$$S^*_{\text{mic}} = \ln \Omega_{\vec{r}^*} = \sum^n_{i=1} \ln \binom{m + r^*_i - 1}{r^*_i}, \tag{76}$$

where we have expressed the number $\Omega_{\vec{r}^*}$ of configurations realizing the hard constraint $\vec{r}(\mathbf{G}) = \vec{r}^*$ as a product of row-specific negative binomial coefficients. Again, the microcanonical entropy can be obtained equivalently from equation (23) as follows (see appendix):

$$
\begin{aligned}
S^*_{\text{mic}} &= \ln \int^{+\vec{\pi}}_{-\vec{\pi}} \frac{d\vec{\psi}}{(2\pi)^n} P^{-1}_{\text{can}}(\mathbf{G}^* | \vec{\theta}^* + i\vec{\psi}) \\
&= \ln \prod^n_{i=1} \int^{+\pi}_{-\pi} \frac{d\psi_i}{2\pi} \frac{(1 - e^{-\theta^*_i - i\psi_i})^{-m}}{e^{-(\theta^*_i + i\psi_i)r^*_i}} \\
&= \sum^n_{i=1} \ln \binom{m + r^*_i - 1}{r^*_i}.
\end{aligned} \tag{77}
$$

The relative entropy, which can be obtained from equation (15) as usual, equals

$$S_n[P^*_{\text{mic}} \| P^*_{\text{can}}] = \sum^n_{i=1} \ln \frac{(m + r^*_i)^{m + r^*_i}}{\binom{m + r^*_i - 1}{r^*_i} (r^*_i)^{r^*_i} m^m}, \tag{78}$$

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

$$p(g_{ij} | \vec{\theta}^*) = e^{-\theta^*_i g_{ij}} (1 - e^{-\theta^*_i}), \quad g_{ij} \in \{0, 1, 2, \dots\}. \tag{79}$$

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

$$\Sigma_{ij}^* = \delta_{ij} r_i^* \left(1 + \frac{r_i^*}{m} \right), \tag{80}$$

(see appendix) and eigenvalues

$$\lambda_i^* = r_i^* \left(1 + \frac{r_i^*}{m} \right) \quad i = 1, n. \tag{81}$$

3.3.3. EN for matrices under one-sided local constraints

We can now compactly discuss the (non)equivalence of canonical and microcanonical ensembles of both binary and weighted matrices under one-sided local constraints. As for the case of global constraints discussed in subsection 3.2.3, we still have an exact knowledge of the canonical entropy, the microcanonical entropy, and the relative entropy. Moreover, these quantities can all be written, using equation (56), in compact expressions formally valid for both binary ($+$) and weighted ($-$) matrices. Indeed, the canonical entropy can be expressed by combining the expressions for $S_n[P_{\text{can}}^*] = S_{\text{can}}^*$ in equations (66) and (75) into the unified formula

$$S_n^{\pm}[P_{\text{can}}^*] = \sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}}, \tag{82}$$

and, similarly, the microcanonical entropy can be obtained by formally combining equations (67) and (76) into

$$S_n^{\pm}[P_{\text{mic}}^*] = \sum_{i=1}^{n} \ln \left[(\pm 1)^{r_i^*} \binom{\pm m}{r_i^*} \right]. \tag{83}$$

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

$$S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*] = \sum_{i=1}^{n} \ln \frac{(\pm m)^{\pm m}}{\binom{\pm m}{r_i^*} (r_i^*)^{r_i^*} (\pm m - r_i^*)^{\pm m - r_i^*}}, \tag{84}$$

which indeed combines the expressions given in equation (69) for binary ($+$) matrices and equation (78) for weighted ($-$) matrices. Equation (84) extends equation (58) to the case of one-sided local constraints. We now consider different regimes.

- In the sparse case where $r_i^* = O(1)$ (for all i) and $m = O(n)$ (see subsection 2.5), we can use Stirling's formula, given by equation (59), to expand $m!$ (but not $r_i^*!$) appearing in the (negative) binomial coefficient to get

$$\binom{\pm m}{r_i^*} \approx \frac{(\pm m)^{r_i^*}}{r_i^*!}, \tag{85}$$

and consequently

$$S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*] \approx \sum_{i=1}^{n} \ln \frac{e^{r_i^*} r_i^*!}{(r_i^*)^{r_i^*}} = O(n). \tag{86}$$

- In the dense case where both r_i^* (for all i) and m are $O(n)$, as discussed in subsection 2.5 (so that r_i^*/m converges to a finite constant), we can use Stirling's formula to expand both $m!$ and $r_i^*!$ into equation (84) to obtain

$$S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*] = \frac{1}{2} \sum_{i=1}^{n} \ln \left[2\pi r_i^* \left(1 - \frac{r_i^*}{\pm m} \right) \right] [1 + o(1)]$$

$$= O(n \ln n), \tag{87}$$

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

- In the dense case where both m and r_i^* are finite, there is no asymptotic expansion that allows to simplify equation (84) in general, so $S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*]$ has to be evaluated explicitly (simple examples are provided below). The important general consideration is that, irrespective of the specific values of m and r_i^*,

$$S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*] = O(n). \tag{88}$$

Again, we can confirm that the above asymptotic expressions are consistent with what we would obtain from equation (29), which follows from the saddle-point approximation given by equation (24). To see this,

noting that here $K = n$ and that equations (71) and (80) indicate that $(\mathbf{\Sigma}^*)^\pm$ is a diagonal matrix with entries

$$(\Sigma_{ij}^*)^\pm = \delta_{ij} r_i^* \left(1 - \frac{r_i^*}{\pm m} \right), \tag{89}$$

for binary $(+)$ and weighted $(-)$ one-sided constraints respectively, we can compactly express equation (28) as

$$
\begin{aligned}
\alpha_n^\pm &= \ln \sqrt{\det \left[2\pi (\mathbf{\Sigma}^*)^\pm \right]} \\
&= \frac{1}{2} \sum_{i=1}^{n} \ln \left[2\pi (\Sigma_{ii}^*)^\pm \right] \\
&= \frac{1}{2} \sum_{i=1}^{n} \ln \left[2\pi r_i^* \left(1 - \frac{r_i^*}{\pm m} \right) \right].
\end{aligned} \tag{90}
$$

In the dense regime with diverging m, the conditions guaranteeing the strong result in equation (30) (all but a finite number of diverging eigenvalues of $\mathbf{\Sigma}^*$) hold, as can be confirmed by comparing equations (87) and (90). Moreover, noticing from equation (59) that $e^k k!/k^k = \sqrt{2\pi k}[1 + o(1)]$, we see that equation (90) confirms the asymptotic behaviour of the relative entropy obtained also in equations (86) and (88) for the other two regimes, under the respective assumptions on the scaling of m and k. Coincidentally, we also see that the stronger result in equation (30) turns out to be a very good approximation for the relative entropy even in these two regimes where, technically, the required conditions are not met. This means that, for the one-sided dense case with finite m, we can rewrite equation (29) asymptotically (i.e. for large n) as

$$
\begin{aligned}
S_n^\pm [P_{\text{mic}}^* \| P_{\text{can}}^*] &= C_1(m) \alpha_n^\pm \\
&= C_1(m) \ln \sqrt{\det[2\pi(\mathbf{\Sigma}^*)^\pm]},
\end{aligned} \tag{91}
$$

where $C_1(m)$ is a finite and positive constant. Moreover, from the known inequality $e^k k!/k^k \geqslant \sqrt{2\pi k}$ for the factorial, we see that $C_1(m) \geqslant 1$ as implied by comparing equations (84) and (90). Finally, we also know from Stirling's approximation that $C_1(m)$ is not much bigger than 1, i.e. $C_1(m) \gtrsim 1$, and that it rapidly approaches 1: indeed when m diverges equation (30) holds exactly, which implies

$$\lim_{m \to \infty} C_1(m) = 1. \tag{92}$$

The fact that, in all regimes, $S_n^\pm [P_{\text{mic}}^* \| P_{\text{can}}^*]$ (or equivalently α_n) is at least of order $O(n)$ shows that equation (31) is violated and that *EE breaks down for both binary and weighted matrices under one-sided local constraints*, irrespective of the density and of the behaviour of m. This important finding generalizes the result, documented so far only for ensembles of binary graphs with given degree sequence [5, 13, 14] (and possibly modular structure [11]) and weighted graphs with given strength sequence [8], that EE breaks down in the presence of an extensive (i.e. growing like n) number of local constraints. Here, this result is extended to more general ensembles of matrices, i.e. asymmetric, rectangular matrices describing e.g. bipartite graphs, multivariate time series, multiplex social activity, multi-cast communication systems and multi-cell gene expression profiles with variable m. More importantly, this generalized setting allows for a qualitatively new phenomenon to emerge, namely the onset of 'strong' EN, as we now show.

Indeed, we can investigate the 'strength' of nonequivalence by comparing the asymptotic behaviour of the relative entropy with that of the canonical entropy given by equation (82). This expression can be evaluated in the usual three regimes as follows.

In the sparse case with $r_i^* = O(1)$ and $m = O(n)$, noticing that asymptotically (for large n) we have $(m \mp r_i^*)^{\pm m - r_i^*} \approx m^{\pm m - r_i^*} e^{\mp r_i^*}$, equation (82) reduces to

$$S_n^\pm [P_{\text{can}}^*] \approx \sum_{i=1}^{n} r_i^* \ln \frac{e^{\pm 1} m}{r_i^*} = O(n \ln n), \tag{93}$$

which dominates over the order $O(n)$ of the corresponding relative entropy $S_n^\pm [P_{\text{mic}}^* \| P_{\text{can}}^*]$ calculated previously in equation (86) for the sparse case. This implies that the limiting relative entropy ratio defined in equation (35) is $R_\infty^\pm = 0$ for both binary $(+)$ and weighted $(-)$ constraints, meaning that in this case the breaking of EE is still 'weak' as in the case of graphs with local constraints.

In the dense case with $r_i^* = O(n)$ and $m = O(n)$, equation (82) can be evaluated as

$$S_n^{\pm}[P_{\text{can}}^*] = \sum_{i=1}^{n} \left[\pm m \ln \frac{m}{m \mp r_i^*} + r_i^* \ln \frac{m \mp r_i^*}{r_i^*} \right]$$

$$= O(n^2), \tag{94}$$

which, again, dominates over the order $O(n \ln n)$ of the corresponding relative entropy calculated in equation (87). Therefore we still have $R_\infty^{\pm} = 0$ (weak nonequivalence).

Finally, the dense case where both m and r_i^* remain finite as $n \to \infty$ is the subject of the rest of this section. Equation (82) implies that

$$S_n^{\pm}[P_{\text{can}}^*] = O(n), \tag{95}$$

which, upon comparison with equation (88), shows that now the relative entropy grows as fast as the canonical entropy, signalling the 'strong' form of EN. Using the combined expressions given in equations (82) and (83), we can explicitly calculate the relative entropy ratio introduced in equation (34) as follows:

$$R_n^{\pm} = 1 - \frac{\sum_{i=1}^{n} \ln \left[(\pm 1)^{r_i^*} \binom{\pm m}{r_i^*} \right]}{\sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}}} > 0. \tag{96}$$

Using equations (90) and (91), we obtain the alternative asymptotic (for large n) expression

$$R_n^{\pm} = \frac{S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*]}{S_n^{\pm}[P_{\text{can}}^*]}$$

$$= C_1(m) \frac{\alpha_n^{\pm}}{\sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}}}$$

$$= \frac{C_1(m)}{2} \frac{\sum_{i=1}^{n} \ln \left[2\pi r_i^* \left(1 - \frac{r_i^*}{\pm m} \right) \right]}{\sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}}}. \tag{97}$$

Comparing equations (96) and (97) confirms that, as noticed above, $C_1(m) \approx 1$ also for finite m.

In general, taking the thermodynamic limit $n \to \infty$ in equations (96) or (97) requires the specification of the value of r_i^* for all i. For the sake of illustration, we can consider the simplest case where each constraint has the same value $r_i^* = r^*$ ($i = 1, n$). Note that the resulting canonical entropy of matrices with constant one-sided constraint r^*, given by equation (82), coincides with the canonical entropy of matrices with the implied global constraint $t^* = nr^*$, given by equations (42) and (50) in the binary and weighted case respectively. However, the microcanonical entropy in the one-sided case, given by equation (83), is strictly smaller than the corresponding one for matrices with the implied global constraint $t^* = nr^*$, given by equations (43) and (51) in the binary and weighted case respectively. From equation (96) we immediately find

$$R_\infty^{\pm} = \lim_{n \to \infty} R_n^{\pm} = 1 - \frac{\ln \left[(\pm 1)^{r^*} \binom{\pm m}{r^*} \right]}{\ln \frac{m^{\pm m}}{(r^*)^{r^*} (m \mp r^*)^{\pm m - r^*}}} > 0, \tag{98}$$

confirming strong nonequivalence as defined in equation (36). To gain numerical and visual insight about the behaviour of $R_\infty^{\pm}$ in equation (98), let us consider the binary and weighted cases separately.
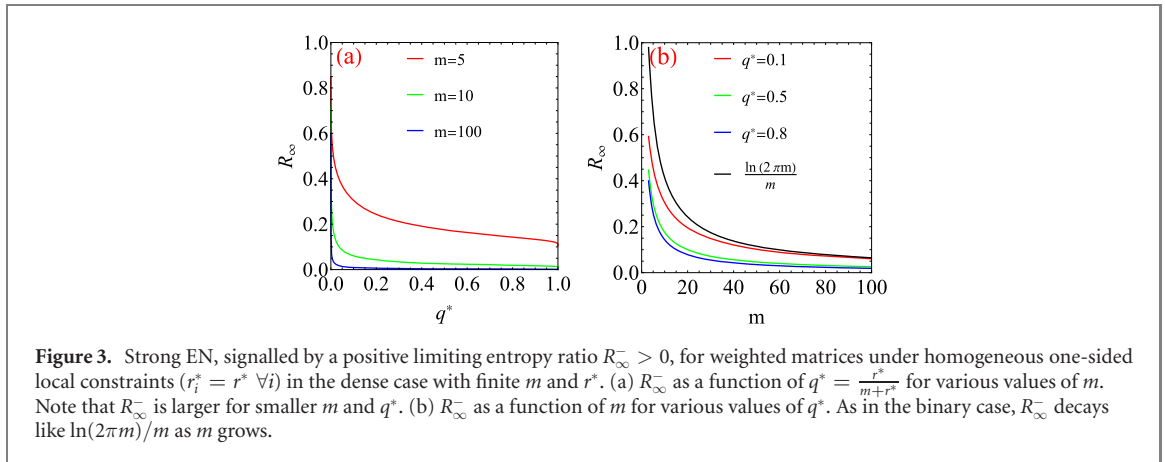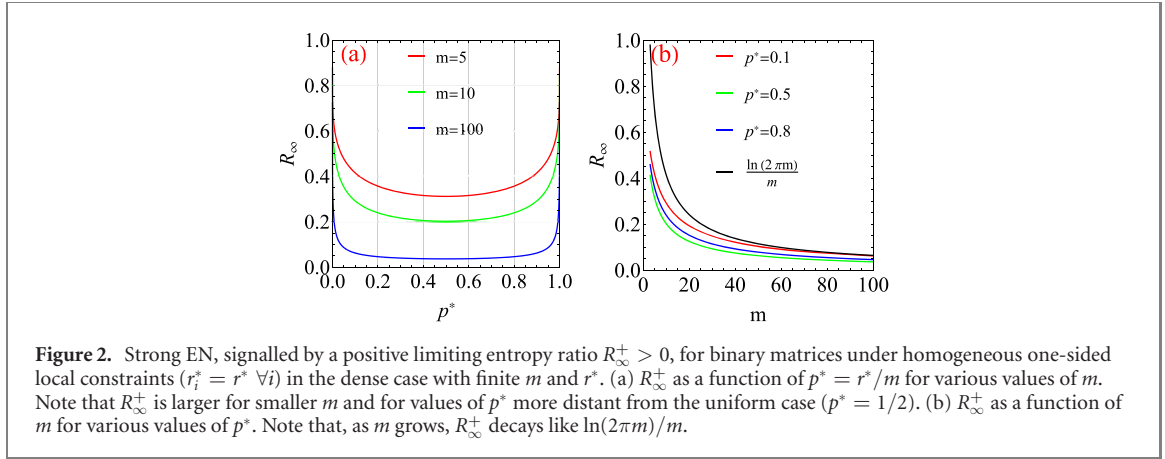
In the binary case, equation (65) implies that, if $r_i^* = r^*$ for all i, then the Lagrange multipliers θ_i^* are all equal to

$$\theta_+^* \equiv \ln \frac{m - r^*}{r^*}. \tag{99}$$

Then, writing $p^* \equiv r^*/m = e^{-\theta_+^*}/(1 + e^{-\theta_+^*}) \in (0, 1)$, from equation (98) we obtain

$$R_\infty^+ = 1 - \frac{\ln \binom{m}{r^*}}{\ln \frac{m^m}{(r^*)^{r^*} (m - r^*)^{m - r^*}}}$$

$$= 1 - \frac{\ln \binom{m}{p^* m}}{\ln \frac{m^m}{(p^* m)^{p^* m} (m - p^* m)^{m - p^* m}}} > 0. \tag{100}$$

Using the above expression, in figure 2 we plot R_∞^+ as a function of either p^* (for fixed m) or m (for fixed p^*). We see that, for a wide range of values of p^*, R_∞^+ remains appreciably large for values of m up to one

Figure 2. Strong EN, signalled by a positive limiting entropy ratio $R_\infty^+ > 0$, for binary matrices under homogeneous one-sided local constraints ($r_i^* = r^* \; \forall i$) in the dense case with finite m and r^*. (a) R_∞^+ as a function of $p^* = r^*/m$ for various values of m. Note that R_∞^+ is larger for smaller m and for values of p^* more distant from the uniform case ($p^* = 1/2$). (b) R_∞^+ as a function of m for various values of p^*. Note that, as m grows, R_∞^+ decays like $\ln(2\pi m)/m$.

Figure 3. Strong EN, signalled by a positive limiting entropy ratio $R_\infty^- > 0$, for weighted matrices under homogeneous one-sided local constraints ($r_i^* = r^* \; \forall i$) in the dense case with finite m and r^*. (a) R_∞^- as a function of $q^* = \frac{r^*}{m+r^*}$ for various values of m. Note that R_∞^- is larger for smaller m and q^*. (b) R_∞^- as a function of m for various values of q^*. As in the binary case, R_∞^- decays like $\ln(2\pi m)/m$ as m grows.

hundred. Moreover, values of p^* closer to 0 or 1 than to 1/2 make R_∞^+ larger. So, for empirical applications where the level of 'multiplexity' is moderate (i.e. small m), and especially away from the uniform case ($p^* = 1/2$), there is a significant entropy reduction from the canonical to the microcanonical ensemble. By contrast, as m increases while p^* remains fixed, R_∞^+ decreases like $\frac{\ln 2\pi m}{m}$, as can be easily realized by applying Stirling's formula to equation (100). This coincides with the system progressively moving to the different regime where both m and r_i^* grow as n grows, which results in weak EN and $R_\infty^+ = 0$ as previously noticed. Similarly, if r_i^* remains finite while m grows, we enter the sparse regime for which $R_\infty^+ = 0$ as previously noticed.

In the weighted case, equation (74) implies that if $r_i^* = r^*$ for all i, then $\theta_i^* = \theta_-^*$ for all i with

$$\theta_-^* \equiv \ln \frac{m + r^*}{r^*}. \tag{101}$$

Then, writing $q^* \equiv e^{-\theta_-^*} = \frac{r^*}{m+r^*} \in (0,1)$, from equation (98) we obtain

$$
\begin{aligned}
R_\infty^- &= 1 - \frac{\ln \binom{m+r^*-1}{r^*}}{\ln \frac{(m+r^*)^{m+r^*}}{m^m (r^*)^{r^*}}} \\
&= 1 - \frac{\ln \binom{(m-1+q^*)/(1-q^*)}{mq^*/(1-q^*)}}{\frac{m}{1-q^*} \ln \frac{m}{1-q^*} - m \ln m - \frac{mq^*}{1-q^*} \ln \frac{mq^*}{1-q^*}} > 0.
\end{aligned}
\tag{102}
$$

Figure 3 shows the behaviour of R_∞^- either as a function of q^* (with m fixed) or as a function of m (with q^* fixed). Considerations similar to the binary case apply. The main difference is that, while R_∞^+ has a symmetric behaviour around the value $p^* = 1/2$ (arising from the fundamental symmetry of exchanging p^* with $1 - p^*$ and $g_{ij} = 1$ with $g_{ij} = 0$ in the binary case), R_∞^- decreases monotonically as a function of q^* (due to the lack of any symmetry of that sort in the weighted case). So now R_∞^- is larger for smaller m and q^*.

The above results illustrate what we had anticipated previously, i.e. that if m is finite (and the matrices are necessarily dense) then the ensembles feature a strong form of EN. Here, this form of EN is also 'unrestricted', as it holds irrespective of the value of $\vec{C}^*$ or $\vec{\theta}^*$, i.e. throughout the parameter space. To the best of our knowledge, this is the first evidence of a situation for which EN occurs in a simultaneously 'strong and unrestricted' form, i.e. the most robust manifestation of the breaking of EE documented so far. Its ultimate origin is the presence of an extensive number of local constraints, and not of phase transitions.

Note that the case with one-sided constraints corresponds, purely formally, to a collection of n independent subsystems, each with m degrees of freedom. In this situation, as clear from equations (82) and (83), both the canonical and microcanonical entropies are obtained as sums over the entropies of the n independent systems. Therefore the overall entropies are on average n times bigger than the average entropies of each independent subsystem. The same applies to the relative entropy, as clear from equation (84). When dividing the overall relative entropy by n, we then obtain the average relative entropy for a single subsystem. This relative entropy does not vanish, unless it is further divided by m and, additionally, m is taken to diverge in the thermodynamic limit (weak EN). But this further division is legitimate only when the fundamental physical units of the system are identified with the individual entries of the matrix $\mathbf{G}$. By contrast, in our situation the m entries of each of the n rows of the matrix represent the possible states of each of the n units (as anticipated, these units are defined either in space or in time). So the physical size of the system is always taken to be n in our setting. Also note that, in the regime when m remains finite as n diverges (strong EN), the relative entropy would not vanish even after a further division by m.

3.4. Two-sided local constraints

We now discuss binary and weighted matrices under two-sided local constraints $(\vec{r}(\mathbf{G}), \vec{c}(\mathbf{G}))$, where $\vec{r}(\mathbf{G})$ is still the n-dimensional vector of row sums while $\vec{c}(\mathbf{G})$ is the m-dimensional vector of column sums, with entries $c_j(\mathbf{G}) = \sum_{i=1}^{n} g_{ij}$ $(j = 1, \ldots, m)$. We constrain both vectors to a given value $(\vec{r}^*, \vec{c}^*) \equiv (\vec{r}(\mathbf{G}^*), \vec{c}(\mathbf{G}^*))$. Note that the number of constraints is still extensive. Unlike the case of one-sided constraints, for two-sided constraints it is not possible to calculate the exact number $\Omega_{\vec{r}^*, \vec{c}^*}$ of configurations in the microcanonical ensemble. By contrast, all canonical calculations can still be carried out analytically (although the value of the Lagrange multipliers can be determined only via implicit expressions). Therefore, as we show below, the matrix $\mathbf{\Sigma}^*$ of canonical covariances between the constraints becomes a crucial tool to calculate the asymptotic behaviour of the relevant microcanonical quantities.

3.4.1. Binary matrices under two-sided local constraints

As usual, we start from the binary case. As shown in the appendix, in the canonical ensemble the likelihood is

$$P_{\text{can}}(\mathbf{G}^* | \vec{\alpha}, \vec{\beta}) = \frac{e^{-\vec{\alpha} \cdot \vec{r}^* - \vec{\beta} \cdot \vec{c}^*}}{\prod_{i=1}^{n} \prod_{j=1}^{m} [1 + e^{-(\alpha_i + \beta_j)}]}, \tag{103}$$

and is maximized by the parameter values $(\vec{\alpha}^*, \vec{\beta}^*)$ defined implicitly by the following set of $n + m$ coupled nonlinear equations:

$$r_i^* = \sum_{j=1}^{m} \frac{e^{-(\alpha_i^* + \beta_j^*)}}{1 + e^{-(\alpha_i^* + \beta_j^*)}}, \quad i = 1, n, \tag{104}$$

$$c_j^* = \sum_{i=1}^{n} \frac{e^{-(\alpha_i^* + \beta_j^*)}}{1 + e^{-(\alpha_i^* + \beta_j^*)}}, \quad j = 1, m. \tag{105}$$

Unfortunately, in general these equations cannot be solved analytically to express $(\vec{\alpha}^*, \vec{\beta}^*)$ as an explicit function of $(\vec{r}^*, \vec{c}^*)$. This is due to the fact that the presence of both row and column constraints couples all parameters. However, the equations can be solved numerically and the unique solution $(\vec{\alpha}^*, \vec{\beta}^*)$ can then be inserted into $P_{\text{can}}(\mathbf{G} | \vec{\alpha}, \vec{\beta})$. This gives complete analytical control over the canonical ensemble. In particular, the canonical entropy is

$$S_{\text{can}}^* = - \ln P_{\text{can}}(\mathbf{G}^* | \vec{\alpha}^*, \vec{\beta}^*)$$

$$= \vec{\alpha}^* \cdot \vec{r}^* + \vec{\beta}^* \cdot \vec{c}^* + \sum_{i=1}^{n} \sum_{j=1}^{m} \ln[1 + e^{-(\alpha_i^* + \beta_j^*)}]. \tag{106}$$

Note that, since this model has additional constraints with respect to the one-sided case with the same row sums $\vec{r}^*$, the canonical entropy above cannot be larger than the corresponding one-sided canonical entropy given by equation (66), i.e. we have the following upper bound:

$$S_{\text{can}}^* \leqslant \sum_{i=1}^{n} \ln \frac{m^m}{(r_i^*)^{r_i^*}(m - r_i^*)^{m - r_i^*}}. \tag{107}$$

On the other hand, the microcanonical entropy S_{mic}^* cannot be computed analytically, although the asymptotic formulas based on equations (27) and (28) can be used to estimate it from the canonical covariance matrix Σ^*. As we show later, this leads to an asymptotic estimate of the relative entropy based on equation (29). As for the canonical entropy, the microcanonical one cannot be larger than the corresponding one given by equation (67) in the one-sided case with the same row sums, with the only difference that now the resulting upper bound is tight:

$$S_{\text{mic}}^* < \sum_{i=1}^{n} \ln \binom{m}{r_i^*}. \tag{108}$$

Indeed, the configurations matching both the row and the column constraints in the two-sided case form a proper subset of the configurations matching only the row constraints in the one-sided case.

It should be noted that in the microcanonical ensemble both $\vec{r}$ and $\vec{c}$ are deterministic vectors fixed to the values $\vec{r}^*$ and $\vec{c}^*$ respectively, while in the canonical ensemble they are random vectors with expected values $\vec{r}^*$ and $\vec{c}^*$. In the microcanonical ensemble, the hard constraints $(\vec{r}(\mathbf{G}), \vec{c}(\mathbf{G})) = (\vec{r}^*, \vec{c}^*)$ create mutual dependencies among all the entries of $\mathbf{G}$. In the canonical ensemble, the soft constraint $(\langle \vec{r} \rangle_{\vec{\alpha}^*}, \langle \vec{c} \rangle_{\vec{\beta}^*}) = (\vec{r}^*, \vec{c}^*)$ leaves all entries of $\mathbf{G}$ independent. As in all other canonical binary ensembles considered above, each entry g_{ij} is Bernoulli-distributed, but now with its specific parameters:

$$p(g_{ij}|\vec{\alpha}^*, \vec{\beta}^*) = \frac{e^{-(\alpha_i^* + \beta_i^*)g_{ij}}}{1 + e^{-(\alpha_i^*, \beta_i^*)}}, \quad g_{ij} \in \{0, 1\}, \tag{109}$$

as shown in the appendix.

For illustration, we consider the special case where the column sums are all equal to each other, i.e. $c_j^* = c^*$ for all j. In this case, since the corresponding Lagrange multipliers must also be all equal to each other ($\beta_j^* = \beta^*$ for all j), it is indeed possible to solve for the parameters explicitly. Indeed, equations (104) and (105) reduce to the $n + 1$ independent equations

$$r_i^* = m \frac{e^{-(\alpha_i^* + \beta^*)}}{1 + e^{-(\alpha_i^* + \beta^*)}}, \quad i = 1, n, \tag{110}$$

$$c^* = \sum_{i=1}^{n} \frac{e^{-(\alpha_i^* + \beta^*)}}{1 + e^{-(\alpha_i^* + \beta^*)}}, \tag{111}$$

where the second equation is simply the consistency condition $c^* = \sum_{i=1}^{n} r_i^*/m$ implied by the first one. This means that the parameter β^* is actually redundant, as it could in principle be reabsorbed into a shift of all the α_i^*'s. In any case, the combination $\alpha_i^* + \beta^*$ is found explicitly by inverting equation (110):

$$\alpha_i^* + \beta^* = \ln \frac{m - r_i^*}{r_i^*}. \tag{112}$$

Note that, inserting this value into the expression for S_{can}^* in equation (106), we obtain exactly the canonical entropy found previously in equation (66) for the binary ensemble with one-sided (row) constraints specified by the same vector $\vec{r}^*$, i.e. the bound in equation (107) is fully saturated:

$$S_{\text{can}}^* = \vec{\alpha}^* \cdot \vec{r}^* + m\beta^* c^* + m \sum_{i=1}^{n} \ln[1 + e^{-(\alpha_i^* + \beta^*)}]$$

$$= \sum_{i=1}^{n} \ln \frac{m^m}{(r_i^*)^{r_i^*}(m - r_i^*)^{m - r_i^*}}. \tag{113}$$

Indeed, the two canonical ensembles are indistinguishable and all their properties are the same. However, the corresponding microcanonical ensembles remain very different, because in the two-sided case each of the m column sums has to match the exact value c^* separately, while in the one-sided case only the total sum mc^* of all the m column sums (which is necessarily implied by the row constraints) has to be matched exactly. Indeed equation (108) is a tight bound that cannot be saturated. Similarly, the covariance matrix

Σ^* is now a $(n + m) \times (n + m)$ matrix (calculated later) and its determinant is different from the one obtained in the one-sided case, where the matrix is $n \times n$.

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

3.4.2. Weighted matrices under two-sided local constraints

We now discuss EN in weighted matrices with two-sided local constraints. The likelihood (see appendix) is now

$$P_{\text{can}}(\mathbf{G}^*|\vec{\alpha}, \vec{\beta}) = \frac{e^{-\vec{\alpha}\cdot\vec{r}^* - \vec{\beta}\cdot\vec{c}^*}}{\prod_{i=1}^{n}\prod_{j=1}^{m}[1 - e^{-(\alpha_i + \beta_j)}]^{-1}}, \tag{114}$$

and is maximized by the unique parameter values $(\vec{\alpha}^*, \vec{\beta}^*)$ defined implicitly through the $n + m$ coupled nonlinear equations

$$r_i^* = \sum_{j=1}^{m} \frac{e^{-(\alpha_i^* + \beta_j^*)}}{1 - e^{-(\alpha_i^* + \beta_j^*)}}, \quad i = 1, n, \tag{115}$$

$$c_j^* = \sum_{i=1}^{n} \frac{e^{-(\alpha_i^* + \beta_j^*)}}{1 - e^{-(\alpha_i^* + \beta_j^*)}}, \quad j = 1, m, \tag{116}$$

that can be solved numerically. The solution $(\vec{\alpha}^*, \vec{\beta}^*)$, when inserted into $P_{\text{can}}(\mathbf{G}|\vec{\alpha}, \vec{\beta})$, completely characterizes the canonical ensemble. The resulting canonical entropy is

$$S_{\text{can}}^* = -\ln P_{\text{can}}(\mathbf{G}^*|\vec{\alpha}^*, \vec{\beta}^*)$$

$$= \vec{\alpha}^* \cdot \vec{r}^* + \vec{\beta}^* \cdot \vec{c}^* - \sum_{i=1}^{n}\sum_{j=1}^{m} \ln[1 - e^{-(\alpha_i^* + \beta_j^*)}], \tag{117}$$

and an upper bound is provided by the canonical entropy given in equation (75) for the one-sided case with the same row constraints $\vec{r}^*$:

$$S_{\text{can}}^* \leqslant \sum_{i=1}^{n} \ln \frac{(m + r_i^*)^{m + r_i^*}}{(r_i^*)^{r_i^*} m^m}. \tag{118}$$

As in the binary two-sided case, the microcanonical entropy S_{mic}^* cannot be computed explicitly, but it can still be evaluated asymptotically from the determinant of the canonical covariance matrix Σ^* using equations (27) and (28). Correspondingly, the relative entropy can be computed using equation (29). The microcanonical entropy given by equation (76) for the corresponding one-sided case is still a strict upper bound for the two-sided entropy:

$$S_{\text{mic}}^* < \sum_{i=1}^{n} \ln \binom{m + r_i^* - 1}{r_i^*}. \tag{119}$$

As in the corresponding binary case, in the microcanonical ensemble both $\vec{r}$ and $\vec{c}$ are deterministic and fixed to the values $\vec{r}^*$ and $\vec{c}^*$, while in the canonical ensemble they are random with expected values $\vec{r}^*$ and $\vec{c}^*$. The coupled hard constraints $(\vec{r}(\mathbf{G}), \vec{c}(\mathbf{G})) = (\vec{r}^*, \vec{c}^*)$ create mutual dependencies among all the entries of $\mathbf{G}$ in the microcanonical ensemble. By contrast, the soft constraint $(\langle\vec{r}\rangle_{\vec{\alpha}^*}, \langle\vec{c}\rangle_{\vec{\beta}^*}) = (\vec{r}^*, \vec{c}^*)$ leaves all entries of $\mathbf{G}$ independent in the canonical ensemble. In the latter, as for all weighted matrices discussed so far, each entry g_{ij} is geometrically distributed, but now with its specific parameters:

$$p(g_{ij}|\vec{\alpha}^*, \vec{\beta}^*) = e^{-(\alpha_i^* + \beta_i^*)g_{ij}} \left[1 - e^{-(\alpha_i^* + \beta_i^*)}\right], \tag{120}$$

for $g_{ij} \in \{0, 1, 2, \ldots\}$, as we show in the appendix.

Here as well, the special case where the column sums are all equal to each other ($c_j^* = c^*$ for all j) provides a nice example. The corresponding Lagrange multipliers are in this case all equal to each other ($\beta_j^* = \beta^*$ for all j) and this allows us to solve for all parameters explicitly. In particular, equations (115) and (116) reduce to the $n + 1$ independent equations

$$r_i^* = m \frac{e^{-(\alpha_i^* + \beta^*)}}{1 - e^{-(\alpha_i^* + \beta^*)}}, \quad i = 1, n, \tag{121}$$

$$c^* = \sum_{i=1}^{n} \frac{e^{-(\alpha_i^* + \beta^*)}}{1 - e^{-(\alpha_i^* + \beta^*)}}, \tag{122}$$

where, again, the second equation is equivalent to the consistency condition $c^* = \sum_{i=1}^{n} r_i^* / m$. Inverting equation (121), we obtain explicitly

$$\alpha_i^* + \beta^* = \ln \frac{m + r_i^*}{r_i^*},$$ (123)

which, if inserted into the expression for S_{can}^* in equation (117), produces exactly the canonical entropy found previously in equation (75) for the weighted ensemble with one-sided (row) constraints specified by the same vector $\vec{r}^*$:

$$S_{\text{can}}^* = \vec{\alpha}^* \cdot \vec{r}^* + m\beta^* c^* - m \sum_{i=1}^{n} \ln[1 - e^{-(\alpha_i^* + \beta^*)}]$$

$$= \sum_{i=1}^{n} \ln \frac{(m + r_i^*)^{m + r_i^*}}{(r_i^*)^{r_i^*} m^m}.$$ (124)

The upper bound in equation (118) is therefore fully saturated. Again, while the canonical ensembles are identical for the two cases, the microcanonical ensembles remain very different and the microcanonical entropy under two-sided constraints is strictly smaller than the one under one-sided constraints: the upper bound in equation (119) cannot be saturated. Similarly, the determinant of the covariance matrix $\mathbf{\Sigma}^*$, which here is a $(n + m) \times (n + m)$ matrix (that we calculate later on), is different from the one obtained in the one-sided case.

The (non)equivalence of the two ensembles is discussed below, in conjunction with the case of two-sided binary matrices.

3.4.3. EN for matrices under two-sided local constraints

To investigate EN in the two-sided case, it is convenient to preliminary combine the results obtained so far in the binary $(+)$ and weighted $(-)$ cases as follows.

The canonical entropy $S_n^{\pm}[P_{\text{can}}^*]$ can be evaluated by combining equations (106) and (117), as well as the corresponding upper bounds given by equations (107) and (118), into

$$S_n^{\pm}[P_{\text{can}}^*] = \vec{\alpha}^* \cdot \vec{r}^* + \vec{\beta}^* \cdot \vec{c}^* \pm \sum_{i=1}^{n} \sum_{j=1}^{m} \ln[1 \pm e^{-(\alpha_i^* + \beta_j^*)}]$$

$$\leqslant \sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m \mp r_i^*}},$$ (125)

(see equation (82) for a comparison). It is easy to check that, in all the three regimes considered (sparse, dense with diverging m, dense with finite m), the above canonical entropy has the same qualitative behaviour as the corresponding quantity obtained previously in equation (82) for the one-sided case.

Unlike the one-sided case, the microcanonical entropy cannot be evaluated exactly, neither through a direct combinatorial formula nor via the complex integral approach, and we only have strict upper bounds given by equations (108) and (119) in the binary and weighted case respectively, which we can combine as follows:

$$S_n^{\pm}[P_{\text{mic}}^*] < \sum_{i=1}^{n} \ln \left[(\pm 1)^{r_i^*} \binom{\pm m}{r_i^*} \right],$$ (126)

(see equation (83) for a comparison).

We can now discuss EN in a combined fashion for binary and weighted matrices. While we cannot calculate the relative entropy exactly, we can correctly evaluate its asymptotic scaling via equation (29), because the canonical covariance matrix $(\mathbf{\Sigma}^*)^{\pm}$ between the constraints can still be calculated analytically as a function of the parameters $(\vec{\alpha}^*, \vec{\beta}^*)$, in both the binary and weighted cases. In particular, it is easy to see that the entries $(\Sigma_{ij}^*)^{\pm}$ are arranged into a block structure, with a square $n \times n$ diagonal block $(i, j \in [1, n])$ representing the covariance matrix between pairs of row sums, a square $m \times m$ diagonal block $(i, j \in [n + 1, n + m])$ representing the covariance matrix between pairs of column sums, and two rectangular $(n \times m$ and $m \times n)$ off-diagonal blocks representing the covariances between row and column

sums ($i \in [1, n], j \in [n + 1, n + m]$ and $i \in [n + 1, n + m], j \in [1, n]$). As we show in the appendix, these entries are

$$
(\Sigma_{ij}^*)^{\pm} = \begin{cases}
\delta_{ij} \sum_{k=1}^{m} \dfrac{e^{-(\alpha_i^* + \beta_k^*)}}{\left[1 \pm e^{-(\alpha_i^* + \beta_k^*)}\right]^2} & i, j \in [1, n], \\[3ex]
\dfrac{e^{-(\alpha_i^* + \beta_{j-n}^*)}}{\left[1 \pm e^{-(\alpha_i^* + \beta_{j-n}^*)}\right]^2} & i \in [1, n], j \in [n + 1, n + m] \\[3ex]
\dfrac{e^{-(\alpha_j^* + \beta_{i-n}^*)}}{\left[1 \pm e^{-(\alpha_j^* + \beta_{i-n}^*)}\right]^2} & i \in [n + 1, n + m], j \in [1, n] \\[3ex]
\delta_{ij} \sum_{k=1}^{n} \dfrac{e^{-(\alpha_k^* + \beta_{j-n}^*)}}{\left[1 \pm e^{-(\alpha_k^* + \beta_{j-n}^*)}\right]^2} & i, j \in [n + 1, n + m].
\end{cases}
\tag{127}
$$

The above expression is the generalization of equation (89) to the case of two-sided constraints. Once the values of $\vec{r}^*$ and $\vec{c}^*$ are specified, one can calculate the determinant of the above matrix and, through equation (29), the leading order of the relative entropy $S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*]$. As we show in the appendix, the order of $\alpha_n^{\pm}$ confirms the same scalings for the relative entropy found previously in equations (86), (87) and (88) for the one-sided case: namely, $\alpha_n^{\pm} = O(n)$ in the sparse regime, $\alpha_n^{\pm} = O(n \ln n)$ in the dense regime with $m = O(n)$, and $\alpha_n^{\pm} = O(n)$ in the dense regime with finite m.

In practice, unlike the one-sided case, calculating the values of $S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*]$ and $R_{\infty}^{\pm}$ (or bounds for them) as explicit functions of the constraints is not easy in general. It is however possible, and instructive, to consider a special case where $S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*]$ and $R_{\infty}^{\pm}$ in this two-sided case (given the vectors $\vec{r}^*$ and $\vec{c}^*$) can be related to the corresponding values obtained in the one-sided case with the same vector $\vec{r}^*$ (but without a constraint on $\vec{c}^*$). Indeed, if we consider again the special case with constant column constraints ($c_j^* = c^*$, $j = 1, m$) then from our previous results in equations (113) and (124) we recall that, for any given value of n, the two-sided canonical entropy $S_n^{\pm}[P_{\text{can}}^*]$ is exactly equal to the one-sided canonical entropy given in equation (82) corresponding to the same vector $\vec{r}^*$, while of course the two-sided microcanonical entropy $S_n^{\pm}[P_{\text{mic}}^*]$ is strictly smaller than the one-sided one given in equation (83). This automatically implies that $S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*]$ in the two-sided case is strictly larger than the corresponding one-sided relative entropy given in equation (84). This proves that the scaling of the relative entropy is always at least $O(n)$, irrespective of the density and of the value of m: in all regimes, EE breaks down for binary and weighted matrices under two-sided local constraints, as found in the one-sided case. The presence of the extra column constraints is not changing the qualitative behaviour of the relative entropy, but only its numerical value. Since the assumption of constant column sums only changes the values, but not the order, of the relative entropy, we expect that the scalings remain unchanged in the general case as well.

Moreover, EN has again the strong form ($R_{\infty}^{\pm} > 0$) in the sparse regime with finite m, because the value of $R_n^{\pm} = 1 - S_n^{\pm}[P_{\text{mic}}^*]/S_n^{\pm}[P_{\text{can}}^*]$ in the two-sided case is strictly larger than the corresponding one calculated previously for the one-sided case. In particular, we can use equation (96) to establish the following lower bound in the two-sided case with constant column constraints and finite m:

$$
R_n^{\pm} > 1 - \frac{\sum_{i=1}^{n} \ln \left[(\pm 1)^{r_i^*} \binom{\pm m}{r_i^*} \right]}{\sum_{i=1}^{n} \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}}} > 0.
\tag{128}
$$

The above inequality proves strong EN in this case as well. Again, we expect that relaxing the assumption of constant column sums will change only the value of $R_n^{\pm}$, but not its strict positivity.

Finally, we can also establish an upper bound for $R_n^{\pm}$ by rewriting equation (29) asymptotically for large n, in analogy with equation (91), as

$$
S_n^{\pm}[P_{\text{mic}}^* \| P_{\text{can}}^*] = C_2(m) \alpha_n^{\pm},
\tag{129}
$$

where $C_2(m)$ is a finite positive constant and noticing that, since the covariance matrix $(\mathbf{\Sigma}^*)^{\pm}$ is positive-definite, we can use Hadamard's inequality stating that the determinant of a positive-definite matrix is less than or equal to the product of the diagonal entries of the matrix. This means

$$
\alpha_n^{\pm} = \ln \sqrt{\det \left[2\pi (\mathbf{\Sigma}^*)^{\pm} \right]} \leqslant \tilde{\alpha}_n^{\pm},
\tag{130}
$$

where, using equation (127), we have introduced

$$
\tilde{\alpha}_n^{\pm} = \frac{1}{2} \sum_{i=1}^{n} \ln \left[2\pi (\Sigma_{ii}^*)^{\pm} \right] + \frac{1}{2} \sum_{j=1}^{m} \ln \left[2\pi (\Sigma_{jj}^*)^{\pm} \right].
\tag{131}
$$

Now, using equations (112) and (123) in the binary and weighted case respectively, it is easy to show that, in the two-sided case with constant column sums, the first n diagonal entries of $(\Sigma^*)^\pm$ are identical to the n diagonal entries of the covariance matrix in the corresponding one-sided case given by equation (89), i.e.

$$(\Sigma_{ii}^*)^\pm = r_i^* \left(1 - \frac{r_i^*}{\pm m} \right), \quad i = 1, n. \tag{132}$$

Inserting the above expression into equation (131), and noticing that the last sum in the latter is strictly positive, we can write

$$\tilde{\alpha}_n^\pm < \frac{1}{2} \sum_{i=1}^n \ln \left[2\pi r_i^* \left(1 - \frac{r_i^*}{\pm m} \right) \right]. \tag{133}$$

Combining equations (129), (130) and (133) we obtain the upper bound (for large n)

$$
\begin{aligned}
R_n^\pm &= C_2(m) \frac{\alpha_n^\pm}{S_n^\pm[P_{\text{can}}^*]} \\
&\leqslant C_2(m) \frac{\tilde{\alpha}_n^\pm}{S_n^\pm[P_{\text{can}}^*]} \\
&< \frac{C_2(m)}{2} \frac{\sum_{i=1}^n \ln \left[2\pi r_i^* \left(1 - \frac{r_i^*}{\pm m} \right) \right]}{\sum_{i=1}^n \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}}} \\
&= \frac{C_2(m)}{C_1(m)} \left[1 - \frac{\sum_{i=1}^n \ln \left[(\pm 1)^{r_i^*} \binom{\pm m}{r_i^*} \right]}{\sum_{i=1}^n \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}}} \right],
\end{aligned}
\tag{134}
$$

where we have used equations (96) and (97) established in the one-sided case. Comparing equation (134) with equation (128) we see that we must have $C_2(m)/C_1(m) > 1$. We conjecture that, in analogy with $C_1(m)$ in the one-sided case, $C_2(m) \gtrsim 1$. Moreover, here as well we know that $\lim_{m \to \infty} C_2(m) = 1$ as in equation (92). This means that we expect that, for n large, $C_2(m)/C_1(m) \approx 1$ so that the upper bound in equation (134) approaches the lower bound in equation (128), which is therefore a very good estimate of the actual value of $R_n^\pm$ in the two-sided case with constant column constraints:

$$R_n^\pm \approx 1 - \frac{\sum_{i=1}^n \ln \left[(\pm 1)^{r_i^*} \binom{\pm m}{r_i^*} \right]}{\sum_{i=1}^n \ln \frac{m^{\pm m}}{(r_i^*)^{r_i^*} (m \mp r_i^*)^{\pm m - r_i^*}}}. \tag{135}$$

Upon comparison with equation (96), we see that $R_n^\pm$ remains practically unchanged with respect to the one-sided case with the same value of $\vec{r}^*$.

Note that the above result means that the decrease Δ_{mic} in microcanonical entropy introduced by the extra column constraints is subleading with respect to the canonical entropy. Indeed, denoting with $\{\cdot\}_h$ a quantity evaluated in the h-sided case (where $h = 1, 2$), and exploiting again the identity of the canonical entropies $\{S_n^\pm[P_{\text{can}}^*]\}_1 = \{S_n^\pm[P_{\text{can}}^*]\}_2$ and our conjecture $C_2(m)/C_1(m) \approx 1$, we can use equation (15) and the results obtained so far to express the decrease in microcanonical entropy as

$$
\begin{aligned}
\Delta_{\text{mic}} &= \{S_n^\pm[P_{\text{mic}}^*]\}_1 - \{S_n^\pm[P_{\text{mic}}^*]\}_2 \\
&= \{S_n^\pm[P_{\text{mic}}^* \| P_{\text{can}}^*]\}_2 - \{S_n^\pm[P_{\text{mic}}^* \| P_{\text{can}}^*]\}_1 \\
&= C_2(m) \{\alpha_n^\pm\}_2 - C_1(m) \{\alpha_n^\pm\}_1 \\
&< C_2(m) \{\tilde{\alpha}_n^\pm\}_2 - C_1(m) \{\alpha_n^\pm\}_1 \\
&\approx \frac{C_1(m)}{2} \sum_{j=1}^m \ln \left[2\pi (\Sigma_{jj}^*)^\pm \right] \\
&= C_1(m) \frac{m}{2} \ln \sum_{k=1}^n \frac{2\pi e^{-(\alpha_k^* + \beta^*)}}{\left[1 \pm e^{-(\alpha_k^* + \beta^*)} \right]^2} \\
&= C_1(m) \frac{m}{2} \ln \sum_{k=1}^n \left[2\pi \frac{r_i^*}{m} \left(1 - \frac{r_i^*}{\pm m} \right) \right],
\end{aligned}
\tag{136}
$$

which is of order $O(\ln n)$, while the canonical entropy is of order $O(n)$ in the dense case with finite m considered here.

Clearly, if we additionally consider constant row constraints, i.e. $r_i^* = r^*$ for $i = 1, n$ (where necessarily $r^* = c^* m/n$), then in analogy with equation (98) we can establish the following explicit lower bound for the value of $R_\infty^\pm$ in the two-sided case with constant row and column constraints:

$$R_\infty^\pm > 1 - \frac{\ln\left[(\pm 1)^{r^*} \binom{\pm m}{r^*}\right]}{\ln \frac{m^{\pm m}}{(r^*)^{r^*}(m \mp r^*)^{\pm m - r^*}}} > 0. \tag{137}$$

Our expectation in equation (135) suggests that the above lower bound is a very good approximation for the actual value of $R_\infty^\pm$:

$$R_\infty^\pm \approx 1 - \frac{\ln\left[(\pm 1)^{r^*} \binom{\pm m}{r^*}\right]}{\ln \frac{m^{\pm m}}{(r^*)^{r^*}(m \mp r^*)^{\pm m - r^*}}}, \tag{138}$$

leading to the same result as in equation (98) for the one-sided case.

The above results generalize the finding of strong EN to the two-sided case, again in the dense regime with finite m. The results do not change qualitatively, and apparently only slightly quantitatively, with respect to the one-sided case. This result points again at the fact that it is the extensivity of the constraints that plays the key role for EN: adding a finite number m of (column) constraints does not relevantly change the picture already obtained in the one-sided case.

4. Discussion and conclusions

We have studied the problem of EN in the general context of $n \times m$ matrices with given constraints. Such matrices can represent high-dimensional data such as multivariate time series, expression profiles, multiplex social activity, and other relational or structured data encountered in many settings. Their entries can either be binary (Boolean) or weighted (non-negative integers). The constraints imposed on these matrices represent sums over either all the entries of the matrix (single global constraint) or over individual rows (local one-sided constraints) and possibly also columns (local two-sided constraints). These constraints take the form of linear terms into the Hamiltonian at the exponent of the maximum-entropy probability distribution characterizing the matrix ensemble.

Global constraints do not account for the heterogeneity (either spatial or temporal, i.e. nonstationarity) in the physical data-generating process, as they lead to probability distributions with identical parameters for all the entries of the matrix. By contrast, local constraints produce probability distributions with different local (row- and possibly column-specific) parameters. Most modern data structures are heterogeneous and/or nonstationary, and are therefore characterized by (at least) the type of local constraints considered here. Indeed, maximum-entropy ensembles with local constraints are being increasingly used, either as null models for pattern detection or as imputation methods whenever there is only partial, local information available about the system [15, 20].

We have shown that local constraints break the asymptotic (i.e. for large n) equivalence of canonical and microcanonical ensembles, where the constraints are enforced in a soft and hard manner respectively. By contrast, global constraints preserve EE. Mathematically, EE is encountered when the relative entropy between the canonical and microcanonical probability distributions is $o(n)$. Importantly, the breakdown of EE observed here under local constraints occurs without phase transitions, which would require nonlinear constraints in the Hamiltonian and are therefore deliberately excluded from the cases we considered. The form of EN we observe under local constraints is also 'unrestricted', i.e. it holds for any value of the model parameters (here, for any graphical value of the constraints), while the mechanism for EN based on phase transitions requires specific parameters or phases. Our results hold in all regimes of density and for all values of m, and therefore generalize a recently discovered, alternative mechanism for the breakdown of EE observed so far in ensembles of binary graphs with given degree sequence [5, 11, 13, 14] and weighted graphs with given strength sequence [8].

At the same time, our results highlight a qualitatively new finding. While the systems with local constraints studied in the past exhibited a 'weak' degree of EN (where the relative entropy is of smaller order compared with the canonical entropy, while still growing at least linearly in n), here we identified a regime for which EN is as 'strong' as in presence of phase transitions (i.e. with the relative entropy being of the same order as the canonical entropy). This regime is obtained when both m and the expected value of each entry of the matrix are finite, i.e. $O(1)$. In practice, this means that the data structure is one where n grows as the size of the system grows, while m remains finite. This circumstance is naturally encountered e.g. when n represents a large number of timesteps during which a small number m of synchronous time

series are observed (e.g. for EEG signals), or when n represents a large number of genes for which expression levels are observed in a small number m of cells at the same time, or when n represents a large number of users whose activities or preferences are recorded for a small number m of platforms, items, or other dimensions. The simultaneously 'strong' and 'unrestricted' form of EN discussed here has never been documented so far, to the best of our knowledge. Indeed, in all the settings that had been studied previously, m was necessarily equal to n since the matrices represented the special case of square adjacency matrices of graphs, therefore the regime leading to strong EN could not be observed.

EN has important practical consequences, for both physics and data science. A traditional expectation in statistical physics is that, in absence of phase transitions or long-range interactions, ensembles are equivalent and it is therefore legitimate to freely choose the ensemble to work with, e.g. based purely on mathematical or computational convenience. So, even if in principle one knows that for an isolated system the appropriate ensemble is the microcanonical one with a hard constraint on the total energy, while for a system that exchanges heat with the environment the appropriate ensemble is the canonical one with a soft constraint on the total energy, for practical purposes the choice of the ensemble is irrelevant, if the ensembles are equivalent. Similarly in data science, where statistical ensembles are often used as null models for real systems, EE would guarantee an analogous freedom of choice: for instance, one may either want to randomize the data numerically by keeping certain quantities fixed (in which case the microcanonical ensemble is generally the computationally cheapest choice) or prefer an exact mathematical characterization of the probability of each configuration in the ensemble (in which case the canonical ensemble is typically the easiest to work with). However, the discovery of EN under local constraints challenges this view and imposes a principled choice of the ensemble, that cannot be based on practical convenience.

Therefore, if one has reasons to believe that the hypothesis underlying the null model, or the partial information available about the system, should be translated into hard local constraints, then one should necessarily choose the microcanonical ensemble. By contrast, if one believes the local constraints should be treated as soft (for instance to account for possible measurement errors leading to noisy values of the constraints in the data), then one should take the canonical route. In analogy with statistical physics, the expectation that the measured constraints are error-free can be regarded as the assumption that each constraint does not interact with any external source of noise that can produce either spurious or missing entries (isolated system), while the expectation of noisy constraints corresponds to the assumption of the existence of a 'noise bath' that, in the case of local constraints, produces fluctuations around the expected value of each constraint separately, along with the corresponding covariances between constraints (system at thermal equilibrium). In presence of EN, a principled choice of the ensemble realising the null model becomes crucial.

Indeed, recent studies have illustrated different (sometimes even opposite) statistical conclusions that can derive from the use of microcanonical and canonical ensembles. For instance, analyses of bipartite networks in ecology [25, 26] have calculated z-scores for various empirical structural properties, according to both the canonical and microcanonical versions of the null model with local constraints. The findings show that certain structural properties (such as the so-called nestedness and niche overlap) are statistically significant with respect to the canonical ensemble but not with respect to the microcanonical one, while for other properties (such as the so-called functional complementarity and certain small motifs) the opposite is true [25, 26]. In certain cases, the nestedness even has z-scores of opposite signs in the two null models and alternative definitions of nestedness turn out to be negatively correlated in the microcanonical one and positively correlated in the canonical one [25]. These results provide an extreme example of different statistical conclusions that can derive from the choice of the null model: with respect to the ensemble with hard constraints, a structural property is interpreted as underrepresented in the data, while with respect to the ensemble with soft constraints, the opposite is true (or vice versa). In a different context, EN has been shown to determine a different expected value for the leading eigenvalue of the adjacency matrix of a graph in the canonical and microcanonical ensembles [27]. Since eigenvalues of adjacency matrices are used for many purposes, from the detection of communities to the study of dynamical processes taking place on a network, the documented difference in the expected eigenvalue can imply different conclusions for many network properties. In general, EN in the *measure* sense (which is the notion we adopt here) is one-to-one connected to EN in the *macrostate* sense, which states that there must be certain macroscopic quantities that have different expectation values in the two ensembles [6]. The leading eigenvalue, the nestedness and all the other quantities mentioned above turn out to be examples of such properties. So, loosely speaking, the existence of different statistical conclusions implied by the use of canonical and microcanonical null models with local constraints is virtually guaranteed mathematically.

The finding of strong EN shows that the quantitative differences between the canonical and microcanonical descriptions of the same system under local constraints are even bigger than previously encountered in the case of weak EN. These differences are exemplified by equations (37), (38) and (39). A

'wrong' choice of the ensemble can therefore lead to substantial errors in the estimation of the probability distribution characterizing the ensemble, of the resulting entropy, of the expected values of higher-order properties that are nonlinear functions of the constraints, etc. Conclusions of statistical analyses can be highly biased if not formulated through the appropriate null model. At the same time, besides providing such a clear warning, the findings presented here are also intended to offer a constructive solution. The fact that it is possible to rigorously quantify the differences between the two ensembles via the explicit calculation of the relative entropy ratio R_n and its limiting value R_∞ implies that one can still make a convenient choice of the ensemble, while at the same time being able to retrieve the desired results for the other ensemble via the calculated value of R_∞. In other words, besides being a warning signal for strong EN, R_∞ is also a concrete tool allowing researchers to switch more easily between alternative descriptions of the same system, by compensating for their irreducible differences. The calculations carried out here can hopefully serve as useful references for future quantitative research in a variety of domains.

Acknowledgments

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Appendix A. Global constraints

Here we derive the main mathematical expressions for the case of global constraints on our ensembles of binary and weighted matrices. This corresponds to the case where the constraint $\vec{C}(\mathbf{G})$ is a simple scalar quantity $C(\mathbf{G})$ defined as the total value $C(\mathbf{G}) \equiv t(\mathbf{G}) \equiv \sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}$. There is only one scalar Lagrange multiplier θ entering the definition of the Hamiltonian

$$H(\mathbf{G}, \theta) = \theta \, t(\mathbf{G}) = \theta \sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij}. \tag{A1}$$

The above Hamiltonian is the same for both binary and weighted matrices under a global constraint. However, the calculation of the partition function (hence of all the other properties) is different in the two cases.

A.1. Binary matrices under a global constraint
Let us consider binary matrices first ($g_{ij} = 0, 1$). The partition function can be calculated as follows:

$$\begin{aligned}
Z(\theta) &= \sum_{\mathbf{G} \in \mathcal{G}} e^{-H(\mathbf{G}, \theta)} \\
&= \sum_{\mathbf{G} \in \mathcal{G}} e^{-\theta \sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij}} \\
&= \sum_{\mathbf{G} \in \mathcal{G}} \prod_{i=1}^{n} \prod_{j=1}^{m} e^{-\theta g_{ij}} \\
&= \prod_{i=1}^{n} \prod_{j=1}^{m} \sum_{g_{ij}=0,1} e^{-\theta g_{ij}} \\
&= \prod_{i=1}^{n} \prod_{j=1}^{m} (1 + e^{-\theta}) \\
&= (1 + e^{-\theta})^{mn}.
\end{aligned} \tag{A2}$$

This leads to

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

$$= \frac{e^{-\theta\, t(\mathbf{G})}}{(1 + e^{-\theta})^{mn}}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{e^{-\theta g_{ij}}}{1 + e^{-\theta}}, \tag{A3}$$

and to equation (40) in the main text. Notice that equation (A3) reveals that each entry g_{ij} of the matrix $\mathbf{G}$ is a Bernoulli-distributed random variable taking the value $g_{ij} = 1$ with probability $p(1|\theta) = e^{-\theta}/(1 + e^{-\theta})$ and the value $g_{ij} = 0$ with probability $p(0|\theta) = 1/(1 + e^{-\theta})$, i.e.

$$p(g_{ij}|\theta) = \frac{e^{-\theta g_{ij}}}{1 + e^{-\theta}}. \tag{A4}$$

(the entries of $\mathbf{G}$ are therefore *i.i.d.*). The expected value of g_{ij} is

$$\langle g_{ij}\rangle_\theta \equiv \sum_{g_{ij}=0,1} g_{ij} p(g_{ij}|\theta) = \frac{e^{-\theta}}{1 + e^{-\theta}}, \tag{A5}$$

while its variance is

$$\text{Var}_\theta[g_{ij}] \equiv \langle g_{ij}^2\rangle_\theta - \langle g_{ij}\rangle_\theta^2 = \frac{e^{-\theta}}{(1 + e^{-\theta})^2}. \tag{A6}$$

The resulting expected value and variance of the constraint $t(\mathbf{G})$ are

$$\langle t\rangle_\theta = \sum_{i=1}^{n} \sum_{j=1}^{m} \langle g_{ij}\rangle_\theta = nm\frac{e^{-\theta}}{1 + e^{-\theta}}, \tag{A7}$$

$$\text{Var}_\theta[t] = \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Var}_\theta[g_{ij}] = nm\frac{e^{-\theta}}{(1 + e^{-\theta})^2}, \tag{A8}$$

the latter identity following from the fact that, since all the entries of $\mathbf{G}$ are mutually independent, the variance of the constraint $t(\mathbf{G})$ is the sum of all variances.

Now, we have to find the parameter value θ^* that solves equation (5) or equivalently maximizes the log-likelihood $\ln P_{\text{can}}(\mathbf{G}^*|\theta)$. This can be done by setting the expected value $\langle t\rangle_{\theta^*}$ equal to the desired value t^*, which leads to

$$e^{-\theta^*} = \frac{t^*}{mn - t^*}, \tag{A9}$$

and to equation (41) in the main text. Inserting equation (A9) into the expressions for $P_{\text{can}}(\mathbf{G}^*|\theta)$ and $\text{Var}_\theta[t]$ leads to the values of S_{can}^* and $\text{Var}_{\theta^*}[t]$ given in equations (42) and (47) in the main text. One can easily confirm that $\text{Var}_{\theta^*}[t]$ coincides with the only ($K = 1$) entry of the 1×1 covariance matrix $\mathbf{\Sigma}^*$ obtained through equation (26), i.e.

$$\mathbf{\Sigma}^* = \left.\frac{\partial^2 \ln Z(\theta)}{\partial \theta^2}\right|_{\theta=\theta^*}$$

$$= t^*\left(1 - \frac{t^*}{mn}\right)$$

$$= \text{Var}_{\theta^*}[t], \tag{A10}$$

and, trivially,

$$\det(\mathbf{\Sigma}^*) = t^*\left(1 - \frac{t^*}{mn}\right). \tag{A11}$$

The calculation of the microcanonical entropy S_{mic}^* is in this case trivial and given by equation (43) in the main text. Given the simplicity of this example, it is instructive to show explicitly that the integral formula

in equation (23), which can be calculated exactly in this case, gives the correct value of Ω_{t^*}. To do this, we use equation (A3) to obtain the complex-valued quantity

$$P_{\text{can}}(\mathbf{G}^*|\theta^* + i\psi) = \frac{e^{-(\theta^*+i\psi)t^*}}{[1 + e^{-(\theta^*+i\psi)}]^{mn}}, \tag{A12}$$

and use equation (23) to calculate Ω_{t^*} as

$$\Omega_{t^*} = \frac{1}{2\pi}\int_{-\pi}^{\pi}[1 + e^{-(\theta^*+i\psi)}]^{mn}e^{(\theta^*+i\psi)t^*}\,\mathrm{d}\psi. \tag{A13}$$

To calculate the above integral, we change variable from ψ to $z \equiv e^{-(\theta^*+i\psi)}$, so that $\mathrm{d}z = \mathrm{d}e^{-(\theta^*+i\psi)} = -iz\mathrm{d}\psi$ and $\mathrm{d}\psi = i\mathrm{d}z/z$. Then the integral becomes

$$\Omega_{t^*} = \frac{i}{2\pi}\int_{e^{-(\theta^*-i\pi)}}^{e^{-(\theta^*+i\pi)}}(1 + z)^{mn}z^{-(t^*+1)}\mathrm{d}z, \tag{A14}$$

and, using the binomial formula

$$(1 + x)^l = \sum_{k=0}^{l}\binom{l}{k}x^{l-k}, \tag{A15}$$

we obtain

$$\Omega_{t^*} = \frac{i}{2\pi}\sum_{k=1}^{mn}\binom{mn}{k}\int_{e^{-(\theta^*-i\pi)}}^{e^{-(\theta^*+i\pi)}}z^{mn-k-t^*-1}\mathrm{d}z. \tag{A16}$$

Now, each integral in the above sum can be calculated using Cauchy's residue theorem, from which we know that the integral is non-zero only when the exponent of z is -1, in which case it equals $-2\pi i$. This selects the only value $k = mn - t^*$ in the sum, so that

$$\Omega_{t^*} = \frac{i}{2\pi}\binom{mn}{mn-t^*}(-2\pi i) = \binom{mn}{t^*}, \tag{A17}$$

which coincides with the binomial coefficient used in equation (43).

A.2. Weighted matrices under a global constraint

We now consider the case of weighted matrices ($g_{ij} = 0, 1, 2, \ldots, +\infty$) with a global constraint t^*. The Hamiltonian is still given by equation (A1), while the partition function is now calculated differently as follows:

$$\begin{aligned}
Z(\theta) &= \sum_{\mathbf{G}\in\mathcal{G}}e^{-H(\mathbf{G},\theta)} \\
&= \sum_{\mathbf{G}\in\mathcal{G}}e^{-\theta\sum_{i=1}^{n}\sum_{j=1}^{m}g_{ij}} \\
&= \sum_{\mathbf{G}\in\mathcal{G}}\prod_{i=1}^{n}\prod_{j=1}^{m}e^{-\theta g_{ij}} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{g_{ij}=0}^{+\infty}e^{-\theta g_{ij}} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m}\frac{1}{1 - e^{-\theta}} \\
&= \frac{1}{(1 - e^{-\theta})^{mn}}.
\end{aligned} \tag{A18}$$

The canonical probability is therefore

$$
\begin{aligned}
P_{\mathrm{can}}(\mathbf{G}|\theta) &= \frac{e^{-H(\mathbf{G},\theta)}}{Z(\theta)} \\
&= e^{-\theta\, t(\mathbf{G})}(1 - e^{-\theta})^{mn} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m} e^{-\theta g_{ij}}(1 - e^{-\theta}),
\end{aligned}
\tag{A19}
$$

which leads to equation (48) in the main text. Equation (A19) shows that all the entries of $\mathbf{G}$ are *i.i.d.* random variables, in this case distributed according to a geometric distribution with success probability $e^{-\theta}$:

$$
p(g_{ij}|\theta) = e^{-\theta g_{ij}}(1 - e^{-\theta}).
\tag{A20}
$$

The expected value of g_{ij} is now

$$
\langle g_{ij}\rangle_\theta = \sum_{g_{ij}=0}^{+\infty} g_{ij}\, p(g_{ij}|\theta) = \frac{e^{-\theta}}{1 - e^{-\theta}},
\tag{A21}
$$

and its variance is

$$
\mathrm{Var}_\theta[g_{ij}] \equiv \langle g_{ij}^2\rangle_\theta - \langle g_{ij}\rangle_\theta^2 = \frac{e^{-\theta}}{(1 - e^{-\theta})^2},
\tag{A22}
$$

(note the change of sign at the denominator with respect to equation (A6)), from which we calculate the expected value and variance of the constraint $t(\mathbf{G})$ as

$$
\langle t\rangle_\theta = \sum_{i=1}^{n}\sum_{j=1}^{m}\langle g_{ij}\rangle_\theta = nm\frac{e^{-\theta}}{1 - e^{-\theta}},
\tag{A23}
$$

$$
\mathrm{Var}_\theta[t] = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{Var}_\theta[g_{ij}] = nm\frac{e^{-\theta}}{(1 - e^{-\theta})^2}.
\tag{A24}
$$

The maximum-likelihood parameter value θ^* is found by setting the expected value $\langle t\rangle_{\theta^*}$ equal to t^*, resulting in

$$
e^{-\theta^*} = \frac{t^*}{mn + t^*},
\tag{A25}
$$

(notice again the change of sign with respect to the binary case) and to equation (49) in the main text. Substituting equation (A25) into equations (A19) and (A24) produces the expressions for S_{can}^* and $\mathrm{Var}_{\theta^*}[t]$ shown in equations (50) and (55) in the main text. As for the binary case, one can easily confirm that $\mathrm{Var}_{\theta^*}[t]$ coincides with

$$
\begin{aligned}
\Sigma^* &= \left.\frac{\partial^2\ln Z(\theta)}{\partial\theta^2}\right|_{\theta=\theta^*} \\
&= t^*\left(1 + \frac{t^*}{mn}\right) \\
&= \mathrm{Var}_{\theta^*}[t],
\end{aligned}
\tag{A26}
$$

so that

$$
\det(\Sigma^*) = t^*\left(1 + \frac{t^*}{mn}\right).
\tag{A27}
$$

Again, it is instructive to show that the complex integral in equation (23) gives the exact result corresponding to the microcanonical entropy reported in equation (51). Calculating the quantity

$$
P_{\mathrm{can}}(\mathbf{G}^*|\theta^* + \mathrm{i}\psi) = \frac{e^{-(\theta^*+\mathrm{i}\psi)t^*}}{[1 - e^{-(\theta^*+\mathrm{i}\psi)}]^{-mn}},
\tag{A28}
$$

and inserting it into equation (23) yields

$$
\Omega_{t^*} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{(\theta^*+\mathrm{i}\psi)t^*}}{[1 - e^{-(\theta^*+\mathrm{i}\psi)}]^{mn}}\,\mathrm{d}\psi.
\tag{A29}
$$

We first perform the change of variable $y \equiv e^{-(\theta^* + i\psi)}$, $d\psi = i \, dy/y$ and rearrange the integral as

$$
\begin{aligned}
\Omega_{t^*} &= \frac{i}{2\pi} \int_{e^{-(\theta^* - i\pi)}}^{e^{-(\theta^* + i\pi)}} y^{-(t^* + 1)} \left(\frac{1}{1-y} \right)^{mn} dy \\
&= \frac{i}{2\pi} \int_{e^{-(\theta^* - i\pi)}}^{e^{-(\theta^* + i\pi)}} y^{-(t^* + 1)} \left(1 + \frac{y}{1-y} \right)^{mn} dy.
\end{aligned}
\tag{A30}
$$

Then we perform a second change of variable $z \equiv y/(1-y)$, $dy = dz/(z+1)^2$ and apply the binomial formula in equation (A15) twice to obtain

$$
\begin{aligned}
\Omega_{t^*} &= \frac{i}{2\pi} \sum_{k=0}^{mn} \binom{mn}{k} \int_{z_-}^{z_+} \left(\frac{z}{z+1} \right)^{-(t^*+1)} \frac{z^k}{(z+1)^2} \, dz \\
&= \frac{i}{2\pi} \sum_{k=0}^{mn} \binom{mn}{k} \int_{z_-}^{z_+} z^{k-t^*-1} (z+1)^{t^*-1} \, dz \\
&= \frac{i}{2\pi} \sum_{k=0}^{mn} \binom{mn}{k} \sum_{h=0}^{t^*-1} \binom{t^*-1}{h} \int_{z_-}^{z_+} z^{k+h-t^*-1} \, dz,
\end{aligned}
$$

where we have defined

$$
z_\pm \equiv \frac{e^{-(\theta^* \pm i\pi)}}{1 - e^{-(\theta^* \pm i\pi)}}.
\tag{A31}
$$

Using again the residue theorem, the only non-zero integral is obtained for $h = t^* - k$, which selects the value

$$
\begin{aligned}
\Omega_{t^*} &= \frac{1}{2\pi i} \sum_{k=0}^{mn} \binom{mn}{k} \binom{t^*-1}{k-1} (-2\pi i) \\
&= \sum_{k=0}^{mn} \binom{mn}{k} \binom{t^*-1}{k-1} \\
&= \sum_{k=0}^{mn} \binom{mn}{k} \binom{t^*-1}{t^*-k} \\
&= \binom{t^*+mn-1}{t^*},
\end{aligned}
\tag{A32}
$$

(where we have used the generalized Vandermonde's identity). The above calculation retrieves exactly the negative binomial coefficient used in equation (51).

Appendix B. One-sided local constraints

We now consider the case of one-sided local constraints on ensembles of binary and weighted $n \times m$ matrices. The constraint $\vec{C}(\mathbf{G})$ is now an n-dimensional ($K = n$) vector $\vec{r}(\mathbf{G})$ where the entry $r_i(\mathbf{G}) = \sum_{j=1}^{m} g_{ij}$ is the ith row sum of the matrix $\mathbf{G}$. Correspondingly, there is an n-dimensional vector $\vec{\theta}$ of Lagrange multipliers and the Hamiltonian is

$$
H(\mathbf{G}, \vec{\theta}) = \vec{\theta} \cdot \vec{r}(\mathbf{G}) = \sum_{i=1}^{n} \theta_i r_i(\mathbf{G}) = \sum_{i=1}^{n} \theta_i \sum_{j=1}^{m} g_{ij},
\tag{B1}
$$

for both binary and weighted matrices. The calculation of the resulting properties of binary and weighted ensembles is discussed separately below.

B.1. Binary matrices under one-sided local constraints
In the binary case, the partition function $Z(\vec{\theta})$ can be calculated from equation (B1) according to the following generalization of equation (A2):

$$
\begin{aligned}
Z(\vec{\theta}) &= \sum_{\mathbf{G} \in \mathcal{G}} e^{-H(\mathbf{G}, \vec{\theta})} \\
&= \sum_{\mathbf{G} \in \mathcal{G}} e^{-\sum_{i=1}^{n} \theta_i \sum_{j=1}^{m} g_{ij}} \\
&= \sum_{\mathbf{G} \in \mathcal{G}} \prod_{i=1}^{n} \prod_{j=1}^{m} e^{-\theta_i g_{ij}} \\
&= \prod_{i=1}^{n} \prod_{j=1}^{m} \sum_{g_{ij}=0,1} e^{-\theta_i g_{ij}} \\
&= \prod_{i=1}^{n} \prod_{j=1}^{m} (1 + e^{-\theta_i}) \\
&= \prod_{i=1}^{n} (1 + e^{-\theta_i})^m.
\end{aligned}
\tag{B2}
$$

The resulting canonical probability is

$$
\begin{aligned}
P_{\text{can}}(\mathbf{G}|\vec{\theta}) &= \frac{e^{-H(\mathbf{G}, \vec{\theta})}}{Z(\vec{\theta})} \\
&= \frac{e^{-\vec{\theta} \cdot \vec{r}(\mathbf{G})}}{\prod_{i=1}^{n} (1 + e^{-\theta_i})^m} \\
&= \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{e^{-\theta_i g_{ij}}}{1 + e^{-\theta_i}},
\end{aligned}
\tag{B3}
$$

which leads to equation (64) in the main text. As in the case of binary matrices under a global constraint, each entry g_{ij} of the matrix $\mathbf{G}$ is a Bernoulli-distributed random variable. However, while all these entries are still independent, the parameter of the distribution depends on the row being considered:

$$
p(g_{ij}|\vec{\theta}) = \frac{e^{-\theta_i g_{ij}}}{1 + e^{-\theta_i}}.
\tag{B4}
$$

Consequently, equations (A5) and (A6) generalize to

$$
\langle g_{ij} \rangle_{\vec{\theta}} \equiv \sum_{g_{ij}=0,1} g_{ij} p(g_{ij}|\vec{\theta}) = \frac{e^{-\theta_i}}{1 + e^{-\theta_i}},
\tag{B5}
$$

$$
\text{Var}_{\vec{\theta}}[g_{ij}] \equiv \langle g_{ij}^2 \rangle_{\vec{\theta}} - \langle g_{ij} \rangle_{\vec{\theta}}^2 = \frac{e^{-\theta_i}}{(1 + e^{-\theta_i})^2}.
\tag{B6}
$$

We can therefore calculate the expected value of each constraint $r_i(\mathbf{G})$ as

$$
\langle r_i \rangle_{\vec{\theta}} = \sum_{j=1}^{m} \langle g_{ij} \rangle_{\vec{\theta}} = m \frac{e^{-\theta_i}}{1 + e^{-\theta_i}}, \quad i = 1, n.
\tag{B7}
$$

Similarly, the variance of r_i is

$$
\text{Var}_{\vec{\theta}}[r_i] = \sum_{j=1}^{m} \text{Var}_{\vec{\theta}}[g_{ij}] = m \frac{e^{-\theta_i}}{(1 + e^{-\theta_i})^2}, \quad i = 1, n,
\tag{B8}
$$

while all covariances between different constraints are zero, because of the independence of distinct entries of $\mathbf{G}$:

$$
\text{Cov}_{\vec{\theta}}[r_i, r_j] = \sum_{k=1}^{m} \sum_{l=1}^{m} \text{Cov}_{\vec{\theta}}[g_{ik}, g_{jl}] = 0, \quad i \neq j.
\tag{B9}
$$

We can combine equations (B8) and (B9) as follows:

$$
\text{Cov}_{\vec{\theta}}[r_i, r_j] = \delta_{ij} m \frac{e^{-\theta_i}}{(1 + e^{-\theta_i})^2},
\tag{B10}
$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

Now, the parameter value $\vec{\theta}^*$ that maximizes the log-likelihood is found by equating the expected value $\langle \vec{r} \rangle_{\vec{\theta}^*}$ with the desired value $\vec{r}^*$. Inverting equation (B7), this leads to

$$e^{-\theta_i^*} = \frac{r_i^*}{m - r_i^*} \quad i = 1, n,$$

(B11)

or equivalently to equation (65) in the main text. The expression for the canonical entropy S_{can}^* given in equation (66) in the main text follows from substituting equation (B11) into equation (B3). Similarly, the expression for the entries of the $n \times n$ covariance matrix $\mathbf{\Sigma}^*$ given in equation (71) in the main text follows from combining equations (B10) and (B11). Note that equation (71) can also be obtained by differentiating the logarithm of equation (B2) as prescribed by equation (26):

$$
\begin{aligned}
\Sigma_{ij}^* &= \left. \frac{\partial^2 \ln Z(\vec{\theta})}{\partial \theta_i \partial \theta_j} \right|_{\vec{\theta} = \vec{\theta}^*} \\
&= \delta_{ij} r_i^* \left(1 - \frac{r_i^*}{m} \right) \\
&= \mathrm{Cov}_{\vec{\theta}^*}[r_i, r_j].
\end{aligned}
$$

(B12)

These results imply

$$\det(\mathbf{\Sigma}^*) = \prod_{i=1}^n r_i^* \left(1 - \frac{r_i^*}{m} \right).$$

(B13)

The microcanonical entropy S_{mic}^* can be directly calculated as equation (67) in the main text. We can still confirm that its value is exactly retrieved by using the integral formula in equation (23). From equation (B3) we obtain

$$P_{\mathrm{can}}(\mathbf{G}^* | \vec{\theta}^* + i\vec{\psi}) = \prod_{i=1}^n \frac{e^{-(\theta_i^* + i\psi_i) r_i^*}}{[1 + e^{-(\theta_i^* + i\psi_i)}]^m}.$$

(B14)

Using equation (23), we can calculate $\Omega_{\vec{r}^*}$ by exploiting again the binomial theorem, a change of variables $(z_i \equiv e^{-(\theta_i^* + i\psi_i)}, \mathrm{d}z_i = -iz_i \mathrm{d}\psi_i)$ and the residue theorem as in equations (A14), (A16) and (A17):

$$
\begin{aligned}
\Omega_{\vec{r}^*} &= \int_{-\vec{\pi}}^{+\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^n} \prod_{i=1}^n \frac{[1 + e^{-(\theta_i^* + i\psi_i)}]^m}{e^{-(\theta_i^* + i\psi_i) r_i^*}} \\
&= \prod_{i=1}^n \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi_i}{2\pi} \frac{[1 + e^{-(\theta_i^* + i\psi_i)}]^m}{e^{-(\theta_i^* + i\psi_i) r_i^*}} \\
&= \prod_{i=1}^n \int_{-\pi}^{+\pi} \frac{\mathrm{d}\psi_i}{2\pi} \sum_{k=0}^m \binom{m}{k} e^{-(\theta_i^* + i\psi_i)(k - r_i^*)} \\
&= \prod_{i=1}^n \int_{\theta_i^* - i\pi}^{\theta_i^* + i\pi} \frac{\mathrm{d}z_i}{(-2\pi i)} \sum_{k=0}^m \binom{m}{k} z_i^{k - r_i^* - 1} \\
&= \prod_{i=1}^n \frac{(-2\pi i)}{(-2\pi i)} \binom{m}{r_i^*} \\
&= \prod_{i=1}^n \binom{m}{r_i^*},
\end{aligned}
$$

(B15)

which coincides with equation (67).

B.2. Weighted matrices under one-sided local constraints

In the weighted case, the partition function is given by the following generalization of equation (A18):

$$Z(\vec{\theta}) = \sum_{\mathbf{G} \in \mathcal{G}} e^{-H(\mathbf{G}, \vec{\theta})}$$

$$= \sum_{\mathbf{G} \in \mathcal{G}} e^{-\sum_{i=1}^{n} \theta_i \sum_{j=1}^{m} g_{ij}}$$

$$= \sum_{\mathbf{G} \in \mathcal{G}} \prod_{i=1}^{n} \prod_{j=1}^{m} e^{-\theta_i g_{ij}}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \sum_{g_{ij}=0}^{+\infty} e^{-\theta_i g_{ij}}$$

$$= \prod_{i=1}^{n} \frac{1}{(1 - e^{-\theta_i})^m}. \tag{B16}$$

The resulting canonical probability is

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

$$= \frac{e^{-\vec{\theta} \cdot \vec{r}(\mathbf{G})}}{\prod_{i=1}^{n}(1 - e^{-\theta_i})^{-m}}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{e^{-\theta_i g_{ij}}}{(1 - e^{-\theta_i})^{-1}}, \tag{B17}$$

leading to equation (73) in the main text. As in the case of weighted matrices under a global constraint, each entry g_{ij} of the matrix $\mathbf{G}$ is an independent and geometrically distributed random variable. On the other hand, as in the case of binary matrices under local constraints, the parameter of the distribution depends on the row being considered:

$$p(g_{ij}|\vec{\theta}) = e^{-\theta_i g_{ij}}(1 - e^{-\theta_i}). \tag{B18}$$

The resulting expected value and variance of g_{ij} are given by the following generalizations of equations (A21) and (A22):

$$\langle g_{ij} \rangle_{\vec{\theta}} \equiv \sum_{g_{ij}=0,1} g_{ij} p(g_{ij}|\vec{\theta}) = \frac{e^{-\theta_i}}{1 - e^{-\theta_i}}, \tag{B19}$$

$$\text{Var}_{\vec{\theta}}[g_{ij}] \equiv \langle g_{ij}^2 \rangle_{\vec{\theta}} - \langle g_{ij} \rangle_{\vec{\theta}}^2 = \frac{e^{-\theta_i}}{(1 - e^{-\theta_i})^2}. \tag{B20}$$

The expected value of each constraint $r_i(\mathbf{G})$ is therefore

$$\langle r_i \rangle_{\vec{\theta}} = \sum_{j=1}^{m} \langle g_{ij} \rangle_{\vec{\theta}} = m \frac{e^{-\theta_i}}{1 - e^{-\theta_i}}, \quad i = 1, n, \tag{B21}$$

while the covariances between different constraints are

$$\text{Cov}_{\vec{\theta}}[r_i, r_j] = \delta_{ij} m \frac{e^{-\theta_i}}{(1 - e^{-\theta_i})^2}. \tag{B22}$$

As usual, note the change of sign at the denominator of equations (B21) and (B22) with respect to the corresponding equations (B7) and (B10) valid in the binary case.

Using equation (B21), we set $\langle \vec{r} \rangle_{\vec{\theta}^*} = \vec{r}^*$ and solve for $\vec{\theta}^*$, finding

$$e^{-\theta_i^*} = \frac{r_i^*}{m + r_i^*} \quad i = 1, n, \tag{B23}$$

as the parameter value that maximizes the log-likelihood. From the above expression, we get equation (74) and, using equation (B17), equation (75) in the main text.

Similarly, the expression for the entries of the $n \times n$ covariance matrix $\mathbf{\Sigma}^*$ given in equation (71) in the main text follows from combining equations (B8), (B9) and (B11). Note that equation (71) can also be obtained by differentiating the logarithm of equation (B2) as prescribed by equation (26):

$$
\begin{aligned}
\Sigma_{ij}^* &= \left.\frac{\partial^2 \ln Z(\vec{\theta})}{\partial \theta_i \partial \theta_j}\right|_{\vec{\theta}=\vec{\theta}^*} \\
&= \delta_{ij} r_i^* \left(1 + \frac{r_i^*}{m}\right) \\
&= \mathrm{Cov}_{\vec{\theta}^*}[r_i, r_j].
\end{aligned}
\tag{B24}
$$

So, in analogy with the binary case,

$$
\det(\mathbf{\Sigma}^*) = \prod_{i=1}^{n} r_i^* \left(1 + \frac{r_i^*}{m}\right).
\tag{B25}
$$

The microcanonical entropy S_{mic}^* can be directly calculated as equation (76) in the main text. We can still confirm that its value is correctly retrieved by using the integral formula in equation (23). From equation (B17) we obtain

$$
P_{\mathrm{can}}(\mathbf{G}^* | \vec{\theta}^* + \mathrm{i}\vec{\psi}) = \prod_{i=1}^{n} \frac{e^{-(\theta_i^* + \mathrm{i}\psi_i) r_i^*}}{[1 - e^{-(\theta_i^* + \mathrm{i}\psi_i)}]^{-m}}.
\tag{B26}
$$

Using equation (23), we can calculate $\Omega_{\vec{r}^*}$ by exploiting again the binomial theorem as

$$
\Omega_{\vec{r}^*} = \int_{-\vec{\pi}}^{\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^n} \prod_{i=1}^{n} \frac{e^{(\beta_i^* + \mathrm{i}\psi_i)(r_i^*)}}{[1 - e^{-(\beta_i^* + \mathrm{i}\psi_i)}]^m}.
\tag{B27}
$$

We can use the change of variables $y_i \equiv e^{-(\beta_i^* + \mathrm{i}\psi_i)}$, $\mathrm{d}\psi_i = -\mathrm{i}\mathrm{d}y_i/y_i$ and the relation $(1 - y_i)^{-m} = (1 + \frac{y_i}{1-y_i})^m$ to calculate $\Omega_{\vec{r}^*}$ as

$$
\begin{aligned}
\Omega_{\vec{r}^*} &= \prod_{i=1}^{n} \int_{\beta_i^* - \mathrm{i}\pi}^{\beta_i^* + \mathrm{i}\pi} \frac{\mathrm{d}y_i}{2\pi\mathrm{i}} \left(1 + \frac{y_i}{1 - y_i}\right)^m y_i^{-r_i^* - 1} \\
&= \prod_{i=1}^{n} \int_{\beta_i^* - \mathrm{i}\pi}^{\beta_i^* + \mathrm{i}\pi} \frac{\mathrm{d}y_i}{2\pi\mathrm{i}} \sum_{k=0}^{m} \binom{m}{k} y_i^{-r_i^* - 1} \left(\frac{y_i}{1 - y_i}\right)^k.
\end{aligned}
\tag{B28}
$$

Using another change of variables $z_i = y_i/(1 - y_i)$, $y_i = z_i/(z_i + 1)$, $\mathrm{d}y_i = \mathrm{d}z_i/(z_i + 1)^2$, and denoting $u_i^* = \frac{e^{-(\beta_i^* + \mathrm{i}\pi)}}{1 - e^{-(\beta_i^* + \mathrm{i}\pi)}}$, we find

$$
\begin{aligned}
\Omega_{\vec{r}^*} &= \prod_{i=1}^{n} \int_{-u_i^*}^{u_i^*} \frac{\mathrm{d}z_i}{2\pi\mathrm{i}} \sum_{k=0}^{m} \binom{m}{k} \left(\frac{z_i}{z_i + 1}\right)^{-r_i^* - 1} \frac{(z_i)^k}{(z_i + 1)^2} \\
&= \prod_{i=1}^{n} \int_{-u_i^*}^{u_i^*} \frac{\mathrm{d}z_i}{2\pi\mathrm{i}} \sum_{k=0}^{m} \binom{m}{k} (z_i + 1)^{r_i^* - 1} (z_i)^{k - r_i^* - 1} \\
&= \prod_{i=1}^{n} \int_{-u_i^*}^{u_i^*} \frac{\mathrm{d}z_i}{2\pi\mathrm{i}} \sum_{k=0}^{m} \binom{m}{k} \sum_{l=0}^{r_i^* - 1} \binom{r_i^* - 1}{l} z_i^{l + k - r_i^* - 1}.
\end{aligned}
$$

Now, according to Cauchy's residue theorem, only when $l + k = r_i^*$ we get a non-zero value. This allows us to further write

$$
\Omega_{\vec{r}^*} = \prod_{i=1}^{n} \sum_{k=1}^{m} \binom{m}{k} \binom{r_i^* - 1}{k - 1} = \prod_{i=1}^{n} \binom{m + r_i^* - 1}{r_i^*},
$$

which coincides with equation (77) in the main text.

Appendix C. Two-sided local constraints

We now discuss ensembles of binary and weighted $n \times m$ matrices with two-sided local constraints. In this case $\vec{C}(\mathbf{G})$ is $(n + m)$-dimensional ($K = n + m$) and specified by the two vectors $(\vec{r}(\mathbf{G}), \vec{c}(\mathbf{G}))$, where $\vec{r}(\mathbf{G})$ is

still the n-dimensional vector of row sums of the matrix $\mathbf{G}$ (as in the one-sided case) and, additionally, $\vec{c}(\mathbf{G})$ is the m-dimensional vector of column sums of $\mathbf{G}$, with entries $c_j(\mathbf{G}) = \sum_{i=1}^{n} g_{ij}$ $(j = 1, m)$. The corresponding Lagrange multipliers take the form $(\vec{\alpha}, \vec{\beta})$ where $\vec{\alpha}$ is n-dimensional and coupled to $\vec{r}(\mathbf{G})$, while $\vec{\beta}$ is m-dimensional and coupled to $\vec{c}(\mathbf{G})$. The corresponding Hamiltonian is

$$H(\mathbf{G}, \vec{\alpha}, \vec{\beta}) = \sum_{i=1}^{n} \alpha_i r_i(\mathbf{G}) + \sum_{j=1}^{m} \beta_j c_j(\mathbf{G})$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) g_{ij}. \tag{C1}$$

As usual, the binary and weighted cases are discussed separately below.

C.1. Binary matrices under two-sided local constraints

Starting from the Hamiltonian in equation (C1), the partition function of the canonical binary matrix ensemble can be still calculated exactly as a simple generalization of equation (B2):

$$Z(\vec{\alpha}, \vec{\beta}) = \sum_{\mathbf{G} \in \mathcal{G}} e^{-H(\mathbf{G}, \vec{\alpha}, \vec{\beta})}$$

$$= \sum_{\mathbf{G} \in \mathcal{G}} e^{-\sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) g_{ij}}$$

$$= \sum_{\mathbf{G} \in \mathcal{G}} \prod_{i=1}^{n} \prod_{j=1}^{m} e^{-(\alpha_i + \beta_j) g_{ij}}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \sum_{g_{ij}=0,1} e^{-(\alpha_i + \beta_j) g_{ij}}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} [1 + e^{-(\alpha_i + \beta_j)}]. \tag{C2}$$

The resulting probability is

$$P_{\text{can}}(\mathbf{G}|\vec{\alpha}, \vec{\beta}) = \frac{e^{-H(\mathbf{G}, \vec{\alpha}, \vec{\beta})}}{Z(\vec{\alpha}, \vec{\beta})}$$

$$= \frac{e^{-\vec{\alpha} \cdot \vec{r}(\mathbf{G}) - \vec{\beta} \cdot \vec{c}(\mathbf{G})}}{\prod_{i=1}^{n} \prod_{j=1}^{m} [1 + e^{-(\alpha_i + \beta_j)}]}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{e^{-(\alpha_i + \beta_j) g_{ij}}}{1 + e^{-(\alpha_i + \beta_j)}}. \tag{C3}$$

As in the case of binary matrices under global and one-sided local constraints, each entry g_{ij} of the matrix $\mathbf{G}$ is still an independent and Bernoulli-distributed random variable, now controlled by the two parameters α_i and β_j. We can write the probability of g_{ij} as

$$p(g_{ij}|\vec{\alpha}, \vec{\beta}) = \frac{e^{-(\alpha_i + \beta_j) g_{ij}}}{1 + e^{-(\alpha_i + \beta_j)}}. \tag{C4}$$

The expected value of g_{ij} is now

$$\langle g_{ij} \rangle_{\vec{\alpha}, \vec{\beta}} \equiv \sum_{g_{ij}=0,1} g_{ij} \, p(g_{ij}|\vec{\alpha}, \vec{\beta}) = \frac{e^{-(\alpha_i + \beta_j)}}{1 + e^{-(\alpha_i + \beta_j)}}, \tag{C5}$$

and the variance is

$$\text{Var}_{\vec{\alpha}, \vec{\beta}}[g_{ij}] \equiv \langle g_{ij}^2 \rangle_{\vec{\alpha}, \vec{\beta}} - \langle g_{ij} \rangle_{\vec{\alpha}, \vec{\beta}}^2$$

$$= \frac{e^{-(\alpha_i + \beta_j)}}{[1 + e^{-(\alpha_i + \beta_j)}]^2}. \tag{C6}$$

The resulting expected values of the constraints are

$$\langle r_i \rangle_{\vec{\alpha},\vec{\beta}} = \sum_{j=1}^{m} \frac{e^{-(\alpha_i+\beta_j)}}{1+e^{-(\alpha_i+\beta_j)}}, \quad i=1,n, \tag{C7}$$

$$\langle c_j \rangle_{\vec{\alpha},\vec{\beta}} = \sum_{i=1}^{n} \frac{e^{-(\alpha_i+\beta_j)}}{1+e^{-(\alpha_i+\beta_j)}}, \quad j=1,m. \tag{C8}$$

The unique parameter values $(\vec{\alpha}^*, \vec{\beta}^*)$ that maximize the likelihood are found as usual by imposing that the expected values $(\langle \vec{r} \rangle_{\vec{\alpha}^*,\vec{\beta}^*}, \langle \vec{c} \rangle_{\vec{\alpha}^*,\vec{\beta}^*})$ match the desired values $(\vec{r}^*, \vec{c}^*)$. Unfortunately, in this case the values $(\vec{\alpha}^*, \vec{\beta}^*)$ cannot be determined analytically as a function of $(\vec{r}^*, \vec{c}^*)$, but they are defined implicitly by imposing

$$(\vec{r}^*, \vec{c}^*) = (\langle \vec{r} \rangle_{\vec{\alpha}^*,\vec{\beta}^*}, \langle \vec{c} \rangle_{\vec{\alpha}^*,\vec{\beta}^*}), \tag{C9}$$

which leads to equations (104) and (105) in the main text.

C.2. Weighted matrices under two-sided local constraints

In the canonical ensemble of weighted matrices under two-sided local constraints, the partition function is the following generalization of equation (B16):

$$\begin{aligned}
Z(\vec{\alpha},\vec{\beta}) &= \sum_{\mathbf{G}\in\mathcal{G}} e^{-H(\mathbf{G},\vec{\alpha},\vec{\beta})} \\
&= \sum_{\mathbf{G}\in\mathcal{G}} e^{-\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i+\beta_j)g_{ij}} \\
&= \sum_{\mathbf{G}\in\mathcal{G}} \prod_{i=1}^{n}\prod_{j=1}^{m} e^{-\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i+\beta_j)g_{ij}} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m} \sum_{g_{ij}=0}^{+\infty} e^{-(\alpha_i+\beta_j)g_{ij}} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m} \frac{1}{1-e^{-(\alpha_i+\beta_j)}}.
\end{aligned} \tag{C10}$$

The resulting canonical probability is

$$\begin{aligned}
P_{\text{can}}(\mathbf{G}|\vec{\alpha},\vec{\beta}) &= \frac{e^{-H(\mathbf{G},\vec{\alpha},\vec{\beta})}}{Z(\vec{\alpha},\vec{\beta})} \\
&= \frac{e^{-\vec{\alpha}\cdot\vec{r}(\mathbf{G})-\vec{\beta}\cdot\vec{c}(\mathbf{G})}}{\prod_{i=1}^{n}\prod_{j=1}^{m}[1-e^{-(\alpha_i+\beta_j)}]^{-1}} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m} \frac{e^{-(\alpha_i+\beta_j)g_{ij}}}{[1-e^{-(\alpha_i+\beta_j)}]^{-1}}.
\end{aligned} \tag{C11}$$

As in the case of weighted matrices under global and one-sided local constraints, each entry g_{ij} of the matrix $\mathbf{G}$ is an independent geometrically distributed random variable defined by the probability

$$p(g_{ij}|\vec{\alpha},\vec{\beta}) = e^{-(\alpha_i+\beta_j)g_{ij}}[1-e^{-(\alpha_i+\beta_j)}], \tag{C12}$$

which is now controlled by the entry-specific pair of parameters α_i, β_j. The expected value of g_{ij} is

$$\langle g_{ij} \rangle_{\vec{\alpha},\vec{\beta}} \equiv \sum_{g_{ij}=0}^{+\infty} g_{ij} p(g_{ij}|\vec{\alpha},\vec{\beta}) = \frac{e^{-(\alpha_i+\beta_j)}}{1-e^{-(\alpha_i+\beta_j)}}, \tag{C13}$$

and the variance is

$$\begin{aligned}
\text{Var}_{\vec{\alpha},\vec{\beta}}[g_{ij}] &\equiv \langle g_{ij}^2 \rangle_{\vec{\alpha},\vec{\beta}} - \langle g_{ij} \rangle_{\vec{\alpha},\vec{\beta}}^2 \\
&= \frac{e^{-(\alpha_i+\beta_j)}}{[1-e^{-(\alpha_i+\beta_j)}]^2}.
\end{aligned} \tag{C14}$$

The expected values of the constraints are

$$\langle r_i \rangle_{\vec{\alpha}, \vec{\beta}} = \sum_{j=1}^{m} \frac{e^{-(\alpha_i + \beta_j)}}{1 - e^{-(\alpha_i + \beta_j)}}, \quad i = 1, n, \tag{C15}$$

$$\langle c_j \rangle_{\vec{\alpha}, \vec{\beta}} = \sum_{i=1}^{n} \frac{e^{-(\alpha_i + \beta_j)}}{1 - e^{-(\alpha_i + \beta_j)}}, \quad j = 1, m. \tag{C16}$$

As in the two-sided binary case, the values $(\vec{\alpha}^*, \vec{\beta}^*)$ maximizing the likelihood cannot be determined analytically as a function of the empirical values $(\vec{r}^*, \vec{c}^*)$, but they are defined implicitly by imposing the equality

$$(\vec{r}^*, \vec{c}^*) = (\langle \vec{r} \rangle_{\vec{\alpha}^*, \vec{\beta}^*}, \langle \vec{c} \rangle_{\vec{\alpha}^*, \vec{\beta}^*}), \tag{C17}$$

between the empirical and the expected values of the constraints. This equality leads to equations (115) and (116) in the main text.

C.3. Determinant of the covariance matrix for two-sided local constraints

The covariance matrix $(\Sigma^*)^{\pm}$ in binary $(+)$ and weighted $(-)$ ensembles of matrices under two-sided local constraints is an $(n + m) \times (n + m)$ matrix. It contains all covariances among the n row sums, all covariances among the m column sums, and all the covariances between row and column sums. If we order the constraints by considering first the n row sums $\vec{r}^*$ and then the m column sums $\vec{c}^*$ into the $(n + m)$-dimensional vector $\vec{C}^* = (\vec{r}^*, \vec{c}^*)$, and combine equations (C2) and (C10) into the general partition function

$$Z^{\pm}(\vec{\alpha}, \vec{\beta}) = \prod_{i=1}^{n} \prod_{j=1}^{m} [1 \pm e^{-(\alpha_i + \beta_j)}]^{\pm 1}, \tag{C18}$$

valid for binary $(+)$ and weighted $(-)$ matrices, we can determine the entries of $\Sigma^{\pm}$ by applying the definition in equation (26). This yields

$$\Sigma_{ij}^{\pm} = \begin{cases} \dfrac{\partial^2 \ln Z^{\pm}(\vec{\alpha}, \vec{\beta})}{\partial \alpha_i \partial \alpha_j} & i, j \in [1, n], \\[2ex] \dfrac{\partial^2 \ln Z^{\pm}(\vec{\alpha}, \vec{\beta})}{\partial \alpha_i \partial \beta_{j-n}} & i \in [1, n], j \in [n+1, n+m] \\[2ex] \dfrac{\partial^2 \ln Z^{\pm}(\vec{\alpha}, \vec{\beta})}{\partial \alpha_{i-n} \partial \beta_j} & i \in [n+1, n+m], j \in [1, n] \\[2ex] \dfrac{\partial^2 \ln Z^{\pm}(\vec{\alpha}, \vec{\beta})}{\partial \beta_{i-n} \partial \beta_{j-n}} & i, j \in [n+1, n+m] \end{cases}$$

$$= \begin{cases} \text{Cov}_{\vec{\alpha}, \vec{\beta}}^{\pm}[r_i, r_j] & i, j \in [1, n], \\[1.5ex] \text{Cov}_{\vec{\alpha}, \vec{\beta}}^{\pm}[r_i, c_{j-n}] & i \in [1, n], j \in [n+1, n+m] \\[1.5ex] \text{Cov}_{\vec{\alpha}, \vec{\beta}}^{\pm}[c_{i-n}, r_j] & i \in [n+1, n+m], j \in [1, n] \\[1.5ex] \text{Cov}_{\vec{\alpha}, \vec{\beta}}^{\pm}[c_{i-n}, c_{j-n}] & i, j \in [n+1, n+m], \end{cases}$$

and, following equation (25),

$$(\Sigma_{ij}^*)^{\pm} = (\Sigma_{ij})^{\pm} \big|_{(\vec{\alpha}, \vec{\beta}) = (\vec{\alpha}^*, \vec{\beta}^*)}. \tag{C19}$$

It is easy to see that $(\Sigma^*)^{\pm}$ is a combination of four blocks

$$(\Sigma^*)^{\pm} = \begin{bmatrix} (\mathbf{A}^*)^{\pm} & (\mathbf{B}^*)^{\pm} \\ (\mathbf{C}^*)^{\pm} & (\mathbf{D}^*)^{\pm} \end{bmatrix}, \tag{C20}$$

where each block has entries as described below. What determines these entries is the elements g_{ij} of the (binary or weighted) adjacency matrix $\mathbf{G}$ that different constraints have in common. The covariance between constraints that have no g_{ij} in common is zero (as such constraints are independent), while the covariance between constraints that share a term g_{ij} receives from that term a contribution equal to

$$\text{Var}_{\vec{\alpha}^*, \vec{\beta}^*}^{\pm}[g_{ij}] = \frac{e^{-(\alpha_i^* + \beta_j^*)}}{[1 \pm e^{-(\alpha_i^* + \beta_j^*)}]^2}, \tag{C21}$$

obtained combining equations (C6) and (C14).

- Block $(\mathbf{A}^*)^{\pm}$ is the $n \times n$ covariance matrix between the row sums, with entries

$$
\begin{aligned}
(A_{ij}^*)^{\pm} &= \mathrm{Cov}_{\vec{\alpha}^*, \vec{\beta}^*}^{\pm}[r_i, r_j] \\
&= \left. \frac{\partial^2 \, \ln Z^{\pm}(\vec{\alpha}, \vec{\beta})}{\partial \alpha_i \partial \alpha_j} \right|_{(\vec{\alpha}, \vec{\beta}) = (\vec{\alpha}^*, \vec{\beta}^*)} \\
&= \delta_{ij} \sum_{k=1}^{m} \frac{e^{-(\alpha_i^* + \beta_k^*)}}{\left[1 \pm e^{-(\alpha_i^* + \beta_k^*)} \right]^2}.
\end{aligned}
\tag{C22}
$$

Note that $(\mathbf{A}^*)^{\pm}$ is a diagonal matrix, since different row sums are all independent.

- Block $(\mathbf{B}^*)^{\pm}$ is the $n \times m$ matrix of covariances between row sums and column sums, with entries

$$
\begin{aligned}
(B_{ij}^*)^{\pm} &= \mathrm{Cov}_{\vec{\alpha}^*, \vec{\beta}^*}^{\pm}[r_i, c_j] \\
&= \left. \frac{\partial^2 \, \ln Z^{\pm}(\vec{\alpha}, \vec{\beta})}{\partial \alpha_i \partial \beta_j} \right|_{(\vec{\alpha}, \vec{\beta}) = (\vec{\alpha}^*, \vec{\beta}^*)} \\
&= \frac{e^{-(\alpha_i^* + \beta_j^*)}}{\left[1 \pm e^{-(\alpha_i^* + \beta_j^*)} \right]^2},
\end{aligned}
\tag{C23}
$$

where we now see that the matrix is not diagonal, as reach row sum r_i shares the entry g_{ij} with the column sum c_j.

- Similarly, block $(\mathbf{C}^*)^{\pm}$ is the $m \times n$ matrix of covariances between column sums and row sums, and is therefore the transpose of $(\mathbf{B}^*)^{\pm}$, as follows also from the fact that $(\mathbf{\Sigma}^*)^{\pm}$ must be symmetric. Indeed its entries are

$$
\begin{aligned}
(C_{ij}^*)^{\pm} &= \mathrm{Cov}_{\vec{\alpha}^*, \vec{\beta}^*}^{\pm}[c_i, r_j] \\
&= \left. \frac{\partial^2 \, \ln Z^{\pm}(\vec{\alpha}, \vec{\beta})}{\partial \alpha_j \partial \beta_i} \right|_{(\vec{\alpha}, \vec{\beta}) = (\vec{\alpha}^*, \vec{\beta}^*)} \\
&= \frac{e^{-(\alpha_j^* + \beta_i^*)}}{\left[1 \pm e^{-(\alpha_j^* + \beta_i^*)} \right]^2}.
\end{aligned}
\tag{C24}
$$

- Finally, block $(\mathbf{D}^*)^{\pm}$ is the $m \times m$ matrix of covariances among the column sums, with entries

$$
\begin{aligned}
(D_{ij}^*)^{\pm} &= \mathrm{Cov}_{\vec{\alpha}^*, \vec{\beta}^*}^{\pm}[c_i, c_j] \\
&= \left. \frac{\partial^2 \, \ln Z^{\pm}(\vec{\alpha}, \vec{\beta})}{\partial \beta_i \partial \beta_j} \right|_{(\vec{\alpha}, \vec{\beta}) = (\vec{\alpha}^*, \vec{\beta}^*)} \\
&= \delta_{ij} \sum_{k=1}^{n} \frac{e^{-(\alpha_k^* + \beta_j^*)}}{\left[1 \pm e^{-(\alpha_k^* + \beta_j^*)} \right]^2}.
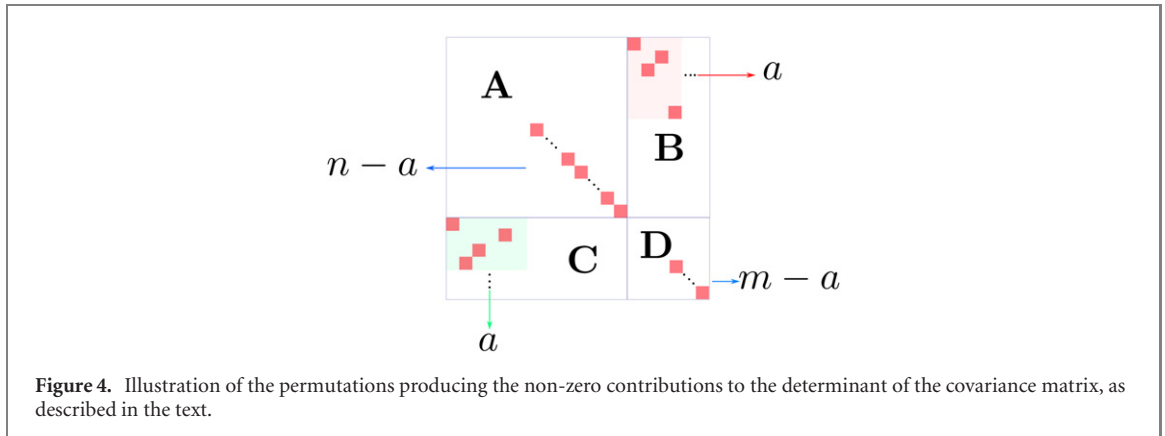\end{aligned}
\tag{C25}
$$

Like $(\mathbf{A}^*)^{\pm}$, $(\mathbf{D}^*)^{\pm}$ is a diagonal matrix, since different column sums are all independent.

Combining equations (C20), (C22), (C23), (C24) and (C25) proves equation (127) in the main text.

Now, in order to calculate the scaling of the determinant of $(\mathbf{\Sigma}^*)^{\pm}$, we follow the definition by Leibniz as

$$
\det[(\mathbf{\Sigma}^*)^{\pm}] = \sum_{\sigma \in \mathbf{Z}_{n+m}} \mathrm{sgn}(\sigma) \prod_{l=1}^{n+m} (\Sigma_{l,\sigma_l}^*)^{\pm},
\tag{C26}
$$

where σ is a permutation of the first $n + m$ integers that exchanges (without replacement) each of these integers i with another such integer $j = \sigma_i$, $\mathbf{Z}_{n+m}$ is the set of all such $(n + m)!$ permutations, and the symbol $\mathrm{sgn}(\sigma)$ represents the *parity* of σ: $\mathrm{sgn}(\sigma) = +1$ when σ is an even permutation (i.e. obtained by combining an even number of pairwise exchanges of the type $j = \sigma_i$ and $i = \sigma_j$) and $\mathrm{sgn}(\sigma) = -1$ when σ is an odd permutation (i.e. obtained by combining an odd number of pairwise exchanges). Let us call σ^0 the identity permutation, i.e. the one such that $\sigma_i^0 = i$ for all i, and $\mathbf{Z}_{n+m}^0 \equiv \mathbf{Z}_{n+m} \backslash \sigma^0$ the set of all other

Figure 4. Illustration of the permutations producing the non-zero contributions to the determinant of the covariance matrix, as described in the text.

permutations. Clearly, $\text{sgn}(\sigma^0) = +1$ because σ^0 involves an even number (zero) of exchanges. We can therefore rewrite equation (C26) as

$$\det[(\mathbf{\Sigma}^*)^\pm] = \Delta^0 + \Delta', \tag{C27}$$

where

$$\Delta^0 = \prod_{l=1}^{n+m} (\Sigma^*_{l,\sigma^0_l})^\pm = \prod_{i=1}^{n} (A^*_{ii})^\pm \prod_{j=1}^{m} (D^*_{jj})^\pm, \tag{C28}$$

is the product of the diagonal entries of $(\mathbf{\Sigma}^*)^\pm$ and

$$\Delta' = \sum_{\sigma \in \mathbf{Z}^0_{n+m}} \text{sgn}(\sigma) \prod_{l=1}^{n+m} (\Sigma^*_{l,\sigma_l})^\pm. \tag{C29}$$

We are going to show that Δ' is at most of the same order of Δ^0. Setting $c_n = 1/n$ in the sparse regime (for which we recall that $m = O(n)$ necessarily) and $c_n = 1$ in the dense regime (for which m can be either finite or $O(n)$), we note that each entry of the blocks $(\mathbf{B}^*)^\pm$ and $(\mathbf{C}^*)^\pm$ is of order $O(c_n)$, while each of the diagonal entries of block $(\mathbf{A}^*)^\pm$ is of order $O(c_n m)$ and each of the diagonal entries of block $(\mathbf{D}^*)^\pm$ is of order $O(c_n n)$. In general, the order of Δ^0 is therefore

$$\Delta^0 = O\left((c_n n)^m (c_n m)^n\right), \tag{C30}$$

as clear from equation (C28).

To this end, we note that each permutation σ appearing in equation (C29) can be expressed as a combination of a certain number (say $a > 0$) of exchanges of pairs of the first $n + m$ integers. It is easy to see that all the $\binom{n}{2}$ exchanges of pairs of the first n integers give a zero contribution to Δ', because they lead to terms of the type $(\Sigma^*_{i,j})^\pm = 0$ where $i, j \in [0, n]$ with $i \neq j$ (combination of exchanges that lead again to $i = j$ are such that $j = \sigma_i = i$ and therefore do not lead to new permutations: they are already accounted for in permutations with lower a). Similarly, all the $\binom{m}{2}$ exchanges of pairs of the next m integers give a zero contribution to Δ', because they lead to terms of the type $(\Sigma^*_{i,j})^\pm = 0$ where $i, j \in [n + 1, n + m]$ with $i \neq j$. Therefore the only exchanges leading to nonzero contributions to Δ' are the nm exchanges across the first n integers and the next m integers, i.e. those that lead to terms $(\Sigma^*_{i,j})^\pm > 0$ where $i \in [0, n]$ and $j \in [n + 1, n + m]$ or $j \in [0, n]$ and $i \in [n + 1, n + m]$ in equation (C29). Each of these nontrivial contributing permutations involves a (unrepeated) exchanges of integers, where $a \in [1, nm]$. Compared with the identity σ_0, each of these permutations replaces a of the first n diagonal entries and a of the next m diagonal entries of $(\mathbf{\Sigma}^*)^\pm$ appearing in equation (C28) with a number $2a$ of non-zero off-diagonal entries in blocks $(\mathbf{B}^*)^\pm$ and $(\mathbf{C}^*)^\pm$ (see figure 4).

Each such permutation therefore gives a contribution of order $(c_n n)^{m-a}(c_n m)^{n-a}$ to the summation in equation (C29). Individually, each such contribution is subleading with respect to the term Δ^0. However, collectively all the contributions involving the same number a of exchanges contribute a term of order $E_a(c_n n)^{m-a}(c_n m)^{n-a}$ where E_a is the number of unrepeated exchanges of a pairs. The order of E_a is given by the number of distinct choices of a exchanges out of the nm possible ones, which is $\binom{nm}{a}$. This estimate does not control for the fact that, for each a, some of the exchanges reduce to simpler permutations already accounted for by smaller values of a, however the leading order is correct. Since nm is large, we can apply

Stirling's approximation to $(nm)!$ and estimate the order of E_a as

$$E_a = O\left(\binom{nm}{a}\right) = O\left(\frac{(nm)^a}{a!}\right), \tag{C31}$$

as in equation (85). Therefore all the permutations realized by a exchanges collectively give a contribution of order

$$E_a(c_n n)^{m-a}(c_n m)^{n-a} = O\left(\frac{(c_n n)^m (c_n m)^n}{a!}\right), \tag{C32}$$

and sign $(-1)^a$ to the sum in equation (C29), so Δ' can be rewritten as a sum over a (with $a = 1, nm$) of terms of alternating sign. Qualitatively, and with an abuse of notation, the order of the entire sum defining $\det[(\boldsymbol{\Sigma}^*)^{\pm}]$ in equation (C26) is (except for accidental cancellations due to particular combinations of values of the entries of $(\boldsymbol{\Sigma}^*)^{\pm}$)

$$O(\det[(\boldsymbol{\Sigma}^*)^{\pm}]) = \sum_{a=1}^{nm} O\left(\frac{(-1)^a (c_n n)^m (c_n m)^n}{a!}\right)$$
$$= O\left((c_n n)^m (c_n m)^n e^{-1}\right)$$
$$= O\left((c_n n)^m (c_n m)^n\right). \tag{C33}$$

We therefore see that the order of Δ' does not exceed that of Δ^0, so the leading order of $\det[(\boldsymbol{\Sigma}^*)^{\pm}]$ is

$$\det[(\boldsymbol{\Sigma}^*)^{\pm}] = O(\Delta^0) = O\left((c_n n)^m (c_n m)^n\right). \tag{C34}$$

In other words, the off-diagonal terms of $(\boldsymbol{\Sigma}^*)^{\pm}$ do not alter the order obtained by multiplying the diagonal terms. For finite m, we therefore have

$$\alpha_n^{\pm} = \ln\sqrt{\det\left[2\pi(\boldsymbol{\Sigma}^*)^{\pm}\right]}$$
$$= O\left(m \ln(c_n n) + n \ln(c_n m)\right). \tag{C35}$$

In the sparse case where $c_n = 1/n$ and $m = O(n)$, we have

$$\alpha_n^{\pm} = O(n), \tag{C36}$$

while in the dense case with $c_n = 1$ and $m = O(n)$ we have

$$\alpha_n^{\pm} = O(n \ln n), \tag{C37}$$

and finally in the dense case with $c_n = 1$ and finite m we have

$$\alpha_n^{\pm} = O(n), \tag{C38}$$

confirming the same scalings for the relative entropy obtained in equations (86), (87) and (88) for the one-sided case.

ORCID iDs

Qi Zhang ⓘ https://orcid.org/0000-0002-6692-0737
Diego Garlaschelli ⓘ https://orcid.org/0000-0001-6035-1783

References

[1] Boltzmann L 2012 *Lectures on Gas Theory* (Massachusetts: Courier Corporation)
[2] Gibbs J 1902 *The Collected Works of JW Gibbs* vol 2 (New Haven, CT: Yale University)
[3] Ellis R S, Haven K and Turkington B 2002 *Nonlinearity* **15** 239
[4] Kastner M 2010 *Phys. Rev. Lett.* **104** 240403
[5] Squartini T, de Mol J, den Hollander F and Garlaschelli D 2015 *Phys. Rev. Lett.* **115** 268701
[6] Touchette H 2015 *J. Stat. Phys.* **159** 987
[7] Squartini T and Garlaschelli D 2017 arXiv:1710.11422
[8] Zhang Q and Garlaschelli D 2020 arXiv:2012.09998
[9] den Hollander F, Mandjes M, Roccaverde A and Starreveld N 2018 *Electron. J. Probab.* **23** 1
[10] Campa A, Dauxois T and Ruffo S 2009 *Phys. Rep.* **480** 57
[11] Garlaschelli D, Den Hollander F and Roccaverde A 2016 *J. Phys. A: Math. Theor.* **50** 015001
[12] Squartini T, Mastrandrea R and Garlaschelli D 2015 *New J. Phys.* **17** 023052

[13] Garlaschelli D, den Hollander F and Roccaverde A 2018 *J. Stat. Phys.* **173** 644

[14] Roccaverde A 2019 *Indag. Math.* **30** 7

[15] Cimini G, Squartini T, Saracco F, Garlaschelli D, Gabrielli A and Caldarelli G 2019 *Nat. Rev. Phys.* **1** 58

[16] Almog A and Garlaschelli D 2014 *New J. Phys.* **16** 093015

[17] Mircea M, Hochane M, Fan X, Chuva de Sousa Lopes S M, Garlaschelli D and Semrau S 2022 *Genome Biol.* **23** 18

[18] Wang P, He W and Zhao J 2013 *IEEE Internet Comput.* **18** 10

[19] Cover T M and Thomas J A 2012 *Elements of Information Theory* (New York: Wiley)

[20] Squartini T and Garlaschelli D 2017 *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics* (Berlin: Springer)

[21] Almog A, Buijink M R, Roethler O, Michel S, Meijer J H, Rohling J H T and Garlaschelli D 2019 *PLoS Comput. Biol.* **15** e1006934

[22] Jaynes E T 1957 *Phys. Rev.* **106** 620

[23] Kullback S and Leibler R A 1951 *Ann. Math. Stat.* **22** 79

[24] Heubach S and Mansour T 2009 *Combinatorics of Compositions and Words* (Boca Raton, FL: CRC Press)

[25] Bruno M, Saracco F, Garlaschelli D, Tessone C J and Caldarelli G 2020 *Sci. Rep.* **10** 19903

[26] Caruso T, Clemente G V, Rillig M C and Garlaschelli D 2021 arXiv:2106.13565

[27] Dionigi P, Garlaschelli D, den Hollander F and Mandjes M 2021 *Electron. Commun. Probab.* **26** 1