Statistical physics and information theory for systems with local constraints

Zhang, Q.

Chapter 4

New information-theoretic bounds for systems with local constraints

Abstract

The information-theoretic bounds are the limit of space to store the information generated by the information source and the limit of speed to reliable transmit information through a channel. In classical information theory, those bounds are determined by the Shannon entropy of the information sources. However, recent research shows that information sources in non-physical systems such as social networks or nervous systems are not a single variable with finite outcomes but a composition of numerous interacting units. Furthermore, these heterogeneous dependencies imply local constraints in those information sources. Thus, to find the new information-theoretical bounds of them, statistical ensembles with local constraints are used to describe those new information sources in this work. We find that under ensemble equivalence, information-theoretical bounds of information sources described by different statistical ensembles are equivalent. When heterogeneous dependencies implied local constraints break the ensemble equivalence, the information storage space of the information source described by the microcanonical ensemble with hard constraints is smaller than that of the canonical ensemble one with soft constraints. The extra sequences in the typical set of the canonical ensemble described sources have the same sum of *Hamiltonian* with the microcanonical ensemble one. But the constraints of each state in the sequence are not equal to hard constraints. Therefore, there is a tradeoff between the choosing of different ensembles to describe those information sources. Using the microcanonical one with hard constraints costs more calculation to obtain the probability distribution but requires less information storage space. Choosing the canonical one needs more space to store the information but requires less calculation to hold the 'soft' constraints[1].

[1]This chapter is based on the coming paper:
Qi Zhang, Diego Garlaschelli, "New information-theoretic bounds for systems with local constraints" (2021)

4.1 Introduction

The birth of modern information theory can be traced to 1948 when Shannon gave the first quantifiable definition of information in his fundamental paper [33]. It is believed to be promoted by the rapid progress of electronic communication systems in the first half of the 20th century. When all the engineers in the communication industry are desire to know what the smallest space needs to store the information that is generated by different information sources, and what is the maximum speed of reliable information transmission through a channel, i.e., the information-theoretical bounds of the communication systems [29].

To find the information-theoretical bounds of the communication systems, Shannon creatively divides the communication systems into three parts: the information source, the channel and the receiver. And all of them are described by the probability theory, e.g., information sources and the receivers are described by random variables; the channel used to transport the information is modelled by conditional probability. The smallest space needs to store the information generated by information sources is decided by the information entropy, which is the probabilistic uncertainty of the information source [33]. The maximum speed of reliable information transmission (channel capacity) is equal to the mutual information between the information source and receiver, which is determined by the conditional probability that is used to describe the channel [29].

Compared with 1948, the information needs to store and transmit in natural and artificial systems right now is much more complex, e.g., the activity of the neurons in the nervous system [9], the appearing of retweets and comments in a social network [36]. Information sources in these systems are not a single variable in the traditional information theory. Instead, they have numerous units, and almost all units are entangled with each other by different interactions. Thus, using the random variables with finite outcomes to model those new information sources is impossible, as the single variable can not describe the heterogeneous interactions among units.

Actually, signal generation by these information sources is not like the sampling of a random variable. It is closer to the change of the particles' status in the thermodynamic system under localized macroscopic properties [27]. Thus, the space to store the information that is generated by the billions of users in Twitter when there is break news in the real world, or the limit of information storage in the nervous system like the brain is equivalent to the quantify of the macroscopic property in thermodynamic systems with numerous particles. Fortunately, the signal generation at different times by those new information sources are still independent. Thus, we can not use the random variable with finite outcomes to describe those information sources. But we can use statistical ensembles from statistical physics to describe the status change of those heterogeneous interacted units and find the new information-theoretical bounds [1, 37, 27, 28].

In statistical physics, statistical ensembles are introduced by Gibbs to model the macroscopic properties of the numerous particles in the thermodynamic system from microscopic behaviour of them based on the probability theory [1]. In traditional statistical physics, the particles are identical, and this is why all the ensembles are under global constraints, such as the fixed total energy or temperature. However, in the new information sources, the heterogeneous interacted units are not identical. The heterogeneous interactions among all units will imply local constraints. Thus, to describe new information sources, we need the statistical ensembles with localized macroscopic constraints [27].

In ensemble theory, systems with different constraints will be described by different statistical ensembles [1]. The *microcanonical ensemble* is used to describe systems with fixed total energy E^*, and the *canonical ensemble* is used to describe the system with fixed temperature $\beta = 1/KT$, where K is the Boltzmann constant, T is the absolute temperature. Obviously, from the energy isolation, the restriction in the microcanonical ensemble is harder than that in the canonical ensemble [15]. Thus, when the local constraints implied by the heterogeneous interactions among the units in the information sources have a different macroscopic property (hard or soft), the information sources also need to be described by different statistical ensembles. When the local constraints are hard, those information sources need to be described by the microcanonical ensemble. When the local constraints are soft, those information sources need to be described by the canonical ensemble [26, 45]. The two ensembles will conjugate with each other by setting the parameter $\beta = \beta^*$ to make the average total energy in the canonical ensemble equal to the fixed total energy in the microcanonical ensemble,$\langle E \rangle = E^*$.

When the system has finite sizes, the two ensembles are certainly different. But in the thermodynamic limit (number of particles goes to infinite), the fluctuation of constraints in the canonical ensemble will vanish. The microcanonical ensemble can be replaced by the canonical ensemble, which is mathematically easy to calculate [8]. This phenomenon is called ensemble equivalence (EE). The existence of EE also shows that the information carried by different ensemble descriptions of the thermodynamic system is the same. However, recent research on networks and system with long-range interactions also show that in the boundary of phase transitions or when the system is under extensive local constraints, this ensemble equivalence will be broken [8, 5, 27]. This ensemble nonequivalence (EN) will directly influence the information-theoretical bounds of the new information sources with different statistical ensembles descriptions under local constraints. As detecting the limits needs the length of the sequences used to record the status changing in the information sources goes to infinite.

In Shannon's setting, the information generated by the random variable described information source is carried by the sequences use to record the status changing of the information source, and most of the information is carried by equiprobable sequences that belong to the typical set of it [29]. Thus, the smallest space of information storage is determined by the size of the typical set. And the influence of ensemble nonequivalence in information sources with heterogeneous interacted units will be manifested by the difference between the typical sets of different ensembles.

In this chapter, the statistical ensembles with local constraints are used to describe the information sources with heterogeneous interacting units to find the new information-theoretical bounds of them [27]. As the extensive local constraints in the new information source will lead the EN, we also need to check the influence of EN on the new information-theoretical bounds. We find that the information storage space of the microcanonical ensemble described information sources is smaller than the conjugate canonical ensemble one. As the typical set of the microcanonical ensemble is smaller than the conjugate canonical ensemble. The extra sequences in the typical set of the canonical ensemble have the same sum of *Hamilitonian* as sequences in the typical set of the microcanonical ensemble. But the local constraints of information sources are not the same. This result shows that using the microcanonical ensemble with hard constraints to describe information sources with heterogeneous interacting units needs less space to store the information generated by it than the conjugate canonical ensembles. But it needs cost more power on the calculations to obtain the probability distributions. Therefore, there is a tradeoff in the choosing of ensembles. Using the microcanonical ensemble needs less information storage space but cost more energy in the initial calculation. Using the canonical ensemble will reduce the energy cost in the initial calculation but requires more information storage space.

4.2 Ensemble described information sources

To describe those new information sources, we need a reasonable mathematical model to quantify the heterogeneous interactions. Actually, the quantification of interactions among different units is not a new problem in scientific research. Networks model as a specific case of the random matrix has already been widely used in different research fields to describe those interactions [12, 14]. Thus, in the following discussion, information sources with heterogeneous interacted units will be described by the random matrix with marginal sums as local constraints [27]. According to the traditional information theory, the limit of space to store the information generated by the random variable described information source is determined by the uncertainty of the information source. Therefore, to find new information-theoretical bounds, we also need to find the probability distribution of possible states of the random matrix with local constraints based on statistical ensembles.

According to statistical physics, the appearance of each state in the thermodynamic system with numerous particles is random. But the different macroscopic properties will determine the probability of each state's presence. For example, in the microcanonical ensemble, the total energy E^* of each state is equal to each other, so each state has the same probability. In the canonical ensemble, the total energy of each state is different, but the temperature β^* is fixed. Therefore, the probability of each state in the canonical ensemble is determined by β^* and the total energy of each state [1]. Thus, the different macroscopic properties of local constraints will affect the choice of ensembles to describe it.

Here, we use the $n \times m$ matrix $\mathbf{G}$ (m can equal to n) to represent the possible configurations of the interacted units in the information sources [27]. Each unit g_{ij} in the matrix $\mathbf{G}$ represents the degree of interaction in this system, and it will have different physical means when there is a different definition of i and j. The constraint of the matrix is $\vec{C}(\mathbf{G})$, and it is determined by the macroscopic property of the interactions among units in the information sources If the interactions are homogeneous, the constraints will be global, which is is the sum of all units in the matrix, $C(\mathbf{G}) = \sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}$. When interactions are heterogeneous, the constraints will be localized as the column and row sum of the matrix $\vec{C}(\mathbf{G}) = [\vec{c}, \vec{r}]$. The column local constraints is a vector with m units in it, $\vec{c} = [c_1, c_2, \cdots, c_j, \cdots, c_m]$, each unit is the sum of all the elements in column j of matrix $\mathbf{G}$ as $c_j = \sum_{i=1}^{n} g_{ij}$. This represents the sum of one kind of property of all the particles in the information source. The row constraints is $\vec{r} = [r_1, r_2, \cdots, r_i, \cdots, r_n]$, each unit is equal to $r_i = \sum_{j=1}^{m} g_{ij}$ [27]. It represents the total influence of the particle i in the information source. Two examples of information sources that need to be modelled by the matrix with local constraints are shown in FIG.4.1

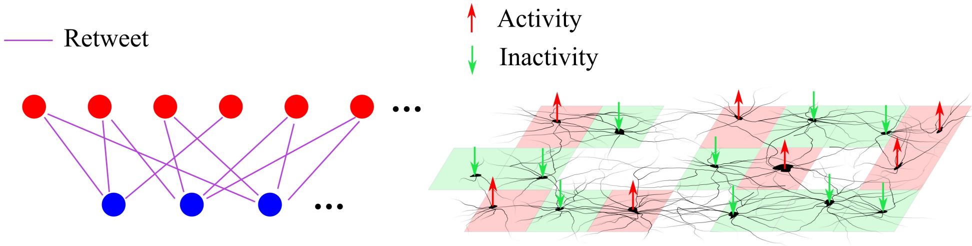

Figure 4.1. The left figure shows the retweets in social media. It can be treated as a bipartite network, where the m users retweet the n users' tweets, each g_{ji} here represents retweet from user j to user i. The right figure shows the activity of the neurons in the nervous system, i and j represents the spatial position of each neuron, g_{ji} here can be 0 or 1, to represents the activity of neurons in the position j, i. The local constraints in the two information sources can be the sum of all the activated neurons in the specific region in the nervous system or the fixed total number of retweets for each user in the social networks. The changing of all the units' state in the two systems is determined by the interactions with each other.

To analytically get the details of the interactions among the numerous units is difficult. Normally, we have the local constraints $\vec{C}(\mathbf{G})$ and the size of the information sources. Thus, to obtain the probability distribution of the states of the information source with the partial information, we need based on the maximum entropy principle introduced by Jaynes [15].

As the macroscopic property of the constraints will decide the ensemble used to describe it [7] so when the constraints are hard, each state of the information source have the same value of constraints as $\vec{C}^*$, the information source needs to be described by the microcanonical ensemble. The probability of each state in the microcanonical ensemble described information source $\mathcal{G}_{\text{mic}}$ is

$$P_{\text{mic}}(\mathbf{G}) = \Omega_{\vec{C}^*}^{-1}, \tag{4.1}$$

where $\Omega_{\vec{C}^*} = |\{\mathbf{G} \in \mathcal{G}_{\text{mic}} : \vec{C}(\mathbf{G}) = \vec{C}^*\}|$ is the number of states in microcanonical ensemble with hard constraints $\vec{C}^*$. The Shannon entropy of the microcanonical ensemble $S_{\text{mic}} = \ln \Omega_{\vec{C}^*}$ is also decided by the number of configurations in it.

If constraints are soft. Only the average value of the constraints in each state of the ensemble is required to equal to the hard constraints in the microcanonical ensemble as $\langle \vec{C}(\mathbf{G}) \rangle = \vec{C}^*$, then the information source needs to be described by the conjugate canonical ensemble. The probability of each state of the canonical ensemble described information source $\mathcal{G}_{\text{can}}$ is equal to

$$P_{\text{can}}(\mathbf{G}) = e^{-H(\mathbf{G},\vec{\beta}^*)}/Z(\vec{\theta}^*), \tag{4.2}$$

where $\vec{\beta}^*$ represnt parameters, which realize $\langle C(\mathbf{G}) \rangle_{\vec{\theta}}$ equal to C^* [15]. The partition function $Z(\vec{\beta}^*)$ is a normalization constant equal to $Z(\vec{\beta}^*) = \sum e^{-H(\mathbf{G},\vec{\beta}^*)}$, which is the sum of $e^{-H(\mathbf{G},\vec{\beta}^*)}$ of all the possible configuration of $\mathbf{G}$ in canonical ensemble. The *Hamiltonian* $H = \vec{C}(\mathbf{G}) \cdot \vec{\beta}^*$ is a liner combination of the constraints and parameter $\vec{\beta}^*$.

According to the definition of the partition function and the *Hamiltonian*, we can rewrite the probability $P_{\text{can}}(\mathbf{G})$ as a product of all the interactions' probability in the information source as

$$P_{\text{can}}(\mathbf{G}) = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{e^{-\beta_{ij}^* g_{ij}}}{\sum_{g_{ij} \in \mathbf{g}} e^{-\beta_{ij}^* g_{ij}}} \tag{4.3}$$

where $\mathbf{g}$ is the collection of all possible configuration of g_{ij} [27]. It means the localized interactions are independent. Thus, when the interaction is homogeneous, the parameter will be set equal to each other as $\beta_{ij}^* = \beta^*$. And then, the canonical ensemble is a $n \times m$ extension of the random variable. It means the classical description is a special case of the canonical ensemble descriptions.

In traditional information theory, the information generated by random variable described information sources are carried by sequences use to record the status changing of random variables [29]. In ensemble described information sources with local constraints, the information generated by it is carried by ensemble sequences. When using different ensembles to describe information sources, we will have different ensemble sequences. The structure of the microcanonical ensemble sequences $\mathbf{A}_{\text{mic}}^{(l)}$ and canonical ensemble sequence $\mathbf{A}_{\text{can}}^{(l)}$ with length l are shown in Fig.4.2.

The limit of information storage space is determined by the size of the corresponding typical set.

Let $\mathbf{G}^*$ denotes the state in the microcanonical ensemble with constraints $\vec{C}^*$, then probability of microcanonical ensemble sequence $\mathbf{A}_{\text{mic}}^{(l)}$ is equal to

$$P(\mathbf{A}_{\text{mic}}^{(l)}) = \prod_{k=1}^{l} P_{\text{mic}}(\mathbf{G}^*) = \Omega_{\vec{C}^*}^{-l} = e^{-lS_{\text{mic}}}. \tag{4.4}$$

As each state have the same probability, so all the microcanonical ensemble sequences have the same probability. According to the asymptotic equipartition property (AEP),

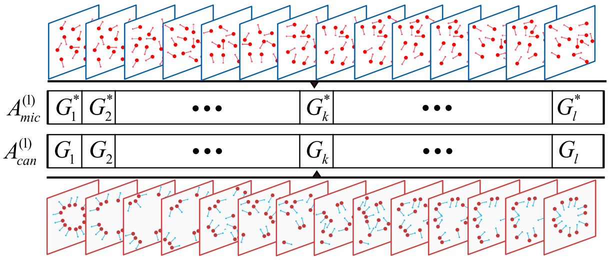

Figure 4.2. As we mentioned before, the states of the new information source are analogue to the recording of the behaviours of numerous particles. Thus, each ensemble sequence is the recording of the interacting units' behaviours in l times. In microcanonical ensemble sequences $\mathbf{A}_{\text{mic}}^{(l)}$, each state $\mathbf{G}_k^*$ have the same constraints $\vec{C}(\mathbf{G}_k^*) = \vec{C}^*$. But in canonical ensemble sequences $\mathbf{A}_{\text{can}}^{(l)}$, the total energy of each state is different. The average value of the constraints of all states should equal the hard constraints in the conjugate microcanonical ensemble $\langle \vec{C}(\mathbf{G}) \rangle = \vec{C}^*$.

microcanonical ensemble sequences have the same probability. Thus, all of them belong to the typical set of it [29].

The number of ensemble sequences in typical set of the microcanonical ensemble $|T_{\text{mic}}^{\epsilon}| = 1/P(\mathbf{A}_{\text{mic}}^{(l)})$ is equal to $e^{lS_{\text{mic}}}$. The smallest space need to store the information generated by the microcanonical ensemble described information source is equal to

$$\ln |T_{\text{mic}}^{\epsilon}| = l \times S_{\text{mic}}. \tag{4.5}$$

It is connected with the possible number of configurations in the microcanonical ensemble with hard constraints $\vec{C}^*$.

The canonical ensemble sequence $\mathbf{A}_{\text{can}}^{(l)}$ is generated by the canonical ensemble $\mathcal{G}_{\text{can}}$. According to Jaynes's work [15], the probability of each state in the canonical ensemble should realize the average value of constraints equal to hard constraints in the microcanonical ensemble as $\langle \vec{C}(\mathbf{G}) \rangle = \vec{C}^*$, and maximization the Shannon entropy S_{can} of it. Thus, probability of the canonical ensemble sequences $\mathbf{A}_{\text{can}}^{(l)}$ still equal to the production of all the l states as

$$P(\mathbf{A}_{\text{can}}^{(l)}) = \prod_{k=1}^{l} P_{\text{can}}(\mathbf{G}_k). \tag{4.6}$$

The smallest space to store the information carried by the canonical ensemble sequences $\mathbf{A}^{(l)}_{\text{can}}$ still can be estimated by the AEP as

$$\begin{aligned}\lim_{l\to\infty}\frac{1}{l}\ln P(\mathbf{A}^{(l)}_{\text{can}}) &= \lim_{l\to\infty}\frac{1}{l}\sum_{k=1}^{l}\ln P_{\text{can}}(\mathbf{G}_k)\\ &\to E[\ln P_{\text{can}}(\mathbf{G})]\\ &= S_{\text{can}}.\end{aligned} \tag{4.7}$$

The S_{can} is the Shannon entropy of the canonical ensemble. It is defined as $S_{\text{can}} = -\sum_{\mathbf{G}\in\mathcal{G}_{\text{can}}} P_{\text{can}}(\mathbf{G})\ln P_{\text{can}}(\mathbf{G})$. When symbol $\langle\cdot\rangle$ represents the average value, canonical entropy $S_{\text{can}} = \langle H + \ln Z(\vec{\theta}^*)\rangle$ will equal to $S_{\text{can}} = \langle\vec{C}\rangle\cdot\vec{\theta}^* + \ln Z(\vec{\theta}^*)$. In the setting of ensemble conjugation, the hard constraints equal to the average value of soft constraints as $\langle\vec{C}\rangle = \vec{C}^*$. It makes the value of S_{can} equals to $\vec{C}^*\cdot\vec{\theta}^* + \ln Z(\vec{\theta}^*)$, which is only based on the probability of state $\mathbf{G}^*$ and equal to logarithm of $P_{\text{can}}(\mathbf{G}^*)$.

As not all states in the canonical ensemble have the same probability, thus the probability of each canonical ensemble sequence may also be different. Therefore, to find the limit of information storage for the canonical ensemble described information source, we need to find the typical set of it. When use the ϵ to represent the bias between the canonical entropy function and the limit of the average value of $\ln P_{\text{can}}(\mathbf{G})$, probability of canonical ensemble sequence in typical set $\mathbf{A}^{(l)}_{T^{\epsilon}_{\text{can}}}$ have the property

$$e^{-l(S_{\text{can}}+\epsilon)} \le P(\mathbf{A}^{(l)}_{T^{\epsilon}_{\text{can}}}) \le e^{-l(S_{\text{can}}-\epsilon)}. \tag{4.8}$$

If the value of ϵ is equal to 0, then the ensemble sequences belong to the typical set of canonical ensembles can be identified by the sum of *Hamiltonian* in the ensemble sequences as

$$T^{\epsilon=0}_{\text{can}} = \{A^{(l)}_{\text{can}} | \sum_{k=1}^{l} H(\mathbf{G}_k, \vec{\theta}^*) = l\times H(\mathbf{G}^*, \vec{\theta}^*)\}. \tag{4.9}$$

This result shows that all sequences in the typical set of the conjugate microcanonical ensemble belong to the typical set of the canonical ensemble.

The number of ensemble sequences in the typical set is equal to $|T^{\epsilon=0}_{\text{can}}|$ and the smallest space needs to store the information is also equal to $\ln|T^{\epsilon=0}_{\text{can}}|$ as

$$\ln|T^{\epsilon=0}_{\text{can}}| = l\times S_{\text{can}} = l\times[\vec{C}^*\cdot\vec{\theta}^* + \ln Z(\vec{\theta}^*)]. \tag{4.10}$$

The result shows that states in the canonical ensemble with constraints $\vec{C}(\mathbf{G}) = \vec{C}^*$.determine the limit of information storage.

The generating of ensemble sequences by statistical ensembles with local constraints is independent. The space needs to store the information carried by the different kinds of ensemble sequences still decided by the Shannon entropy of the ensemble, which is used to describe the information source.

4.3 Information storage under ensemble nonequivalence

The measure-level ensemble equivalence is that the canonical probability distribution converges to the conjugate microcanonical one in the thermodynamic limit [8]. Under ensemble nonequivalence, there is always a difference between the two probability distributions, even in the thermodynamic limit. Specifically, probability of states with the constraints C^* in the canonical ensemble is smaller than that in the microcanonical ensemble $P_{\text{mic}}(\mathbf{G}^*) > P_{\text{can}}(\mathbf{G}^*)$. It is why the conjugate canonical ensemble always has a bigger Shannon entropy than the microcanonical ensemble [45].

The measure-level ensemble nonequivalence is easy to be shown in the probability distribution of the states as in FIG.4.3. And this difference can be quantified by the

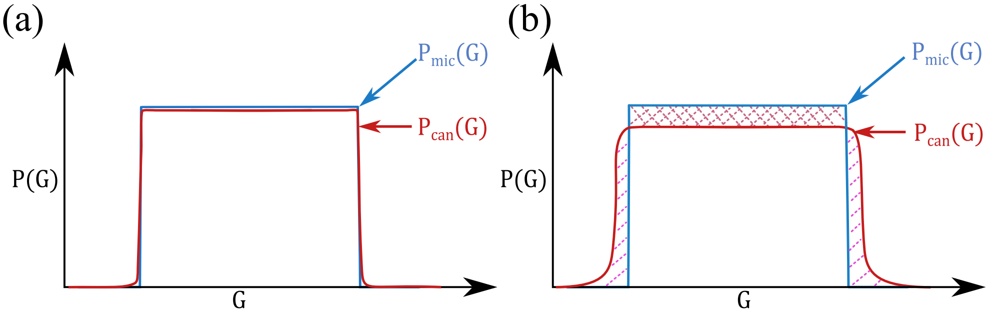

Figure 4.3. Probability distribution of states in the microcanonical and conjugate canonical ensemble under EE (a) and EN (b).

relative entropy between probability distributions of the microcanonical and canonical ensemble as

$$S(P_{\text{mic}}||P_{\text{can}}) = \sum_{\mathbf{G}\in\mathcal{G}_{\text{can}}} P_{\text{mic}}(\mathbf{G}) \ln \frac{P_{\text{mic}}(\mathbf{G})}{P_{\text{can}}(\mathbf{G})}. \tag{4.11}$$

The probability of states in microcanonical with constraints $\vec{C} \neq \vec{C}^*$ is equal to 0, so the relative entropy is decided by states in the two ensembles with constraints $\vec{C}^*$. And the value of it is equal to $S(P_{\text{mic}}||P_{\text{can}}) = \ln P_{\text{mic}}(\mathbf{G}^*) - \ln P_{\text{can}}(\mathbf{G}^*)$, which is the difference between the Shannon entropy of the two ensembles, $S(P_{\text{mic}}||P_{\text{can}}) = S_{\text{can}} - S_{\text{mic}}$. It directly connects with the difference of the typical set's size.

The relative entropy is difficult to obtain, as the value of $\Omega_{\vec{C}^*}$ is hard to calculate. However, according to the assumption that all the microscopic configurations in the microcanonical ensemble are the subset of the conjugate canonical ensemble, the number of configurations in the microcanonical ensemble can be estimated by the δ-function as $\Omega_{\vec{C}^*} = \sum_{\mathbf{G}\in\mathcal{G}} \int_{-\vec{\pi}}^{\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} e^{\mathrm{i}\vec{\psi}[\vec{C}^*-\vec{C}(\mathbf{G})]}$, which can be simplied as the function of the canonical probability $\Omega_{\vec{C}^*} = \int_{-\vec{\pi}}^{+\vec{\pi}} \frac{\mathrm{d}\vec{\psi}}{(2\pi)^K} P_{\text{can}}^{-1}(\mathbf{G}^*|\vec{\beta}^* + \mathrm{i}\vec{\psi})$ [37].

When the integration is hard to calculate we can still use the saddle-point technique to approach number of configurations in the microcanonical ensemble as

$$\Omega_{\vec{C}^*} = \frac{e^{S^*_{\text{can}}}}{\sqrt{\det(2\pi\boldsymbol{\Sigma}^*)}} \prod_{k=1}^{K} [1 + O(1/\lambda^*_k)], \tag{4.12}$$

which is based on the covariance matrix of constraints $\boldsymbol{\Sigma}^*$ [37]. Therefore, the relative entropy between the microcanonical and canonical ensemble is equal to

$$S(P_{\text{mic}}||P_{\text{can}}) = \frac{1}{2}\sum_{w=1}^{W} \ln \frac{2\pi\lambda^*_w}{[1 + O(1/\lambda^*_w)]^2}, \tag{4.13}$$

where λ^*_w is the wth no-zero eigenvalue of the covariance matrix of constraints in the canonical ensemble $\boldsymbol{\Sigma}^*$, W is the total number of constraints in the systems. When the matrix is under two-sided local constraints, the value of $W = n + m$ [27].

Each entry $\boldsymbol{\Sigma}^*_{kl}$ in the covariance matrix reprensents the covariance between local constraints C_k and C_l, $\boldsymbol{\Sigma}^*_{kl} = \text{Cov}[C_k, C_l]_{\vec{\beta}^*}$. The constraint C_k or C_l here can be the column local constraint r^*_i or c^*_j in the canonical ensemble. The value of $\boldsymbol{\Sigma}^*_{kl}$ equal to $\boldsymbol{\Sigma}^*_{kl} = \frac{\partial^2 \ln Z(\vec{\beta}^*)}{\partial\beta^*_k \partial\beta^*_l}$. It can be obtained from the partial differential of the logarithm of partition function of the canonical ensemble $\ln Z(\vec{\theta}^*)$. More details of the proofs can be found in [37].

According to Eq.(4.5) and Eq.(4.10), we can find the difference between the size of typical sets of different ensembles described information sources are connected with the relative entropy between the ensembles. This relative entropy is the indicator that is used to detect the measure-level ensemble nonequivalence. Thus, the EN in the information sources will directly affect the information-theoretical bounds of the new information sources.

As the Shannon entropy of the microcanonical ensemble is smaller than the conjugate canonical ensemble, the typical set of the microcanonical ensemble described information sources is also smaller than the canonical ensemble one. It means using the canonical ensemble to describe the information source with heterogeneous interacting units needs extra space to store the sequences that belong to the typical set of the canonical ensemble but not include in the typical set of the microcanonical ensemble. Moreover, this extra space is determined by the relative entropy between the two ensembles, or in other words, it will be affected by the degree of ensemble nonequivalence. Especially when each unit in the information source has a finite degree of freedom, there is a strong ensemble nonequivalence [27], the gap between the limit of space to store the information generated by the different ensemble described information source has the same order as the limit under the canonical ensemble descriptions.

The space requires to store the set of the extra sequences $T^{(l)}_{\tilde{\alpha}_n}$ in the typical set of the canonical ensemble but not belong to the typical set of the microcanonical ensemble is equal to

$$\ln |T^{(l)}_{\tilde{\alpha}_n}| = l \times \frac{1}{2} \sum_{w=1}^{W} \ln \frac{2\pi\lambda^*_w}{[1+O(1/\lambda^*_w)]^2}. \tag{4.14}$$

Since sequences belong to the typical set of the canonical ensemble should have the sum of each state's *Hamiltonian* as $l \times H(\mathbf{G}^*, \vec{\theta}^*)$ and ensemble sequences generate by the microcanonical ensemble are all belong to the typical set of the canonical ensemble. Thus the extra sequences $T^{\epsilon}_{\tilde{\alpha}_n}$ should satisfy the following condition:

$$T_{\tilde{\alpha}_n} = \{T^{(l)}_{\text{can}} | \sum_{k=1}^{l} \vec{C}(\mathbf{G}_k) = l \times \vec{C}^*, \vec{C}(\mathbf{G}_k) \neq \vec{C}^*\}. \tag{4.15}$$

If χ_{can} is a collection of all the ensemble sequences generated by the canonical ensemble described information source, and χ_{mic} represents all the ensemble sequences generated by the microcanonical ensemble, then the relationship between the typical set of them under ensemble equivalence or nonequivalence is shown as in FIG.4.4.

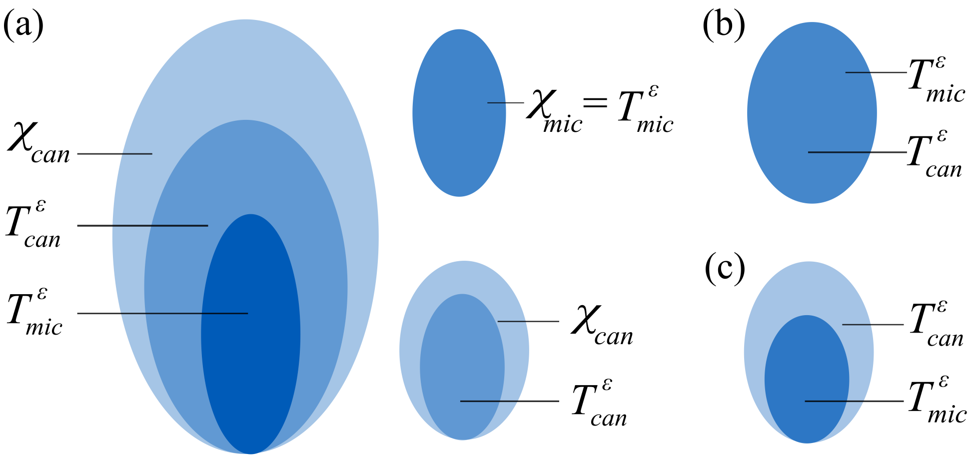

Figure 4.4. Every ensemble sequence generated by the microcanonical ensemble belongs to the typical set of it. But not all the ensemble sequences of the canonical ensemble are included in $T^{\epsilon}_{\text{can}}$. When the information source is under EE, the typical set of the microcanonical ensemble $T^{\epsilon}_{\text{mic}}$ is close to the typical set of the canonical ensemble as shown in (b). When the information sources are under EN, there is a non-vanished difference between the typical set of the canonical ensemble and the microcanonical ensemble like in (c).

Then we want to know if the difference between the information-theoretical bounds that is quantified by $\ln |T^{\epsilon}_{\tilde{\alpha}_n}|$ is bigger enough to affect the choice of the ensembles to describe information source? To solve this problem, we need to compare the limit of the extra space to store the set $T_{\tilde{\alpha}_n}$ with the space needs to store the information that is generated by the canonical ensemble as

$$\lim_{l\to\infty} r = \frac{\ln|T^{\epsilon}_{\tilde{\alpha}_n}|}{\ln|T^{\epsilon}_{\text{can}}|} = \frac{\frac{1}{2}\sum_{k=1}^{K} \frac{\ln(2\pi\lambda_k^*)}{[1+O(1/\lambda_w^*)]^2}}{-\vec{\theta}^* \cdot \vec{C}^* - \ln Z(\vec{\theta}^*)}. \tag{4.16}$$

If the limit value r is bigger than 0, then the space saved by choosing the microcanonical ensemble is bigger enough. This will happens when the information source with heterogeneous interacted units are under strong ensemble nonequivalence [27]. When the information source is under EE, the information-theoretical bounds of the different ensemble described information source is the same. Such as the information sources described by the ER model and the matrix with global constraints.

The proof details are in the appendix. When information sources need to be described by the matrix with local constraints, the ratio r is bigger than 0, and close to $r \sim \frac{\ln(2\pi m)}{m}$, where m is the degree of freedom for each unit in the information source [27]. Proof also can be found in the appendix.

4.4 Ensemble nonequivalence and channel capacity

As we already mentioned before, the information-theoretical bounds include the limit of information storage and the maximum speed of reliable information transmission through a channel. Here, to find the possible influence of EN on the channel capacity, we suppose there is a zero-less ensemble channel, which can transmit all the information generated by the ensemble described information sources correct. Then the information R should be transmitted through the channel is equal to

$$R = \begin{cases} l \times S_{\text{mic}}, & \text{microcanonical ensemble} \\ l \times S_{\text{can}}, & \text{canonical ensemble} \end{cases} \tag{4.17}$$

When the number of codes that can be used to carry the information is fixed, using the microcanonical ensemble to describe the information source will leave more redundancy for the information transmission. Furthermore, this redundancy can be quantified by the covariance of the matrix in the conjugate canonical ensemble.

4.5 Conclusions

In this chapter, we show that the information-theoretical bounds of the information sources with numerous heterogeneous interacting units are still decided by the uncertainty of the information sources, which are quantified by the Shannon entropy. As the random variable with finite outcomes is not enough to describe the new information sources, so we have introduced the statistical ensembles with local constraints to model those heterogeneous interactions. Thus, the new information-theoretical bounds are decided by the entropy of those statistical ensembles. As these new information sources have a huge number of units, the entropy of the statistical ensemble described information sources would be affected by the possible appearance of ensemble nonequivalence. Under this case, different ensemble descriptions have different Shannon entropy. Using the microcanonical ensemble to describe the information source with interacting units needs a smaller information storage space, but the probability distribution of the microcanonical ensemble is hard to calculate. Using the canonical ensemble to describe the new information sources is easy to get the probability distribution but requires more information storage space. The extra sequences that need to store in the canonical ensemble descriptions are those sequences with the same sum of *Hamiltonian* as the sequences in the typical set of the microcanonical ensemble, but the constraints of each state are not equal to the hard constraints of the microcanonical ensemble. The size of the extra typical sets can be approached by the covariance matrix of constraints in the canonical ensemble. It means the difference between the information-theoretical bounds of different ensemble descriptions of the ensemble nonequivalent information sources is affected by the fluctuation in the local constraints. It has reinforced the conclusions in the traditional information theory that the uncertainty of the information sources will affect the information-theoretical bounds.

Appendix 4.A ER model described information sources

ER model $G(n,p)$ represents the probability of the system with n units, and each pair of units have a probability of p to connect with each other. The constraint in the ER model is the total number of interactions among the units in the information source. It still can be described by the microcanonical and the canonical ensemble when constraints in it have different properties.

As the number of the possible links in the ER model is equal to $n(n-1)/2$, the hard constraint C^* in the microcanonical ensemble is equal to the expectation value of the total number of links as $p \times n(n-1)/2$. Thus, the probability of each state in the microcanonical ensemble with these hard constraints is

$$P_{\text{mic}}(\mathbf{G}^*) = 1/\binom{n(n-1)/2}{C^*}. \tag{4.18}$$

Each state has the same probability to appears in the process of signal generation. The value of the probability is decided by the total number of configurations with the same constraints.

When the ER model is under 'soft' constraints, each state of the information source does not need to have the same value of total interactions; only the average value of total interactions is equal to the C^*. Then the value of p is equal to $\frac{2C^*}{n(n-1)}$, probability of the state of the information source described by the canonical ensemble is

$$P_{\text{can}}(\mathbf{G}) = p^{C(\mathbf{G})}(1-p)^{\frac{1}{2}n(n-1)-C(\mathbf{G})}. \tag{4.19}$$

The canonical entropy is equal to $-\ln P_{\text{can}}(\mathbf{G}^*)$. When the value of soft constraints is equal to C^*, the canonical entropy is

$$\begin{aligned} S_{\text{can}} =& - C^* \ln 2C^* + \frac{n(n-1)}{2} \ln[n(n-1)] \\ & - (\frac{n(n-1)}{2} - C^*) \ln(n(n-1) - 2C^*). \end{aligned} \tag{4.20}$$

If the information source described by the microcanonical ER model ensembles is used to generate the information sequences with length l, the space to store the information that is generated by the microcanonical description is equal to

$$\ln |T_{\text{mic}}^{\epsilon=0}| = l \times \ln \binom{n(n-1)/2}{C^*}, \tag{4.21}$$

where $T_{\text{mic}}^{\epsilon=0}$ is the typical set of the microcanonical ensemble information sequence.

When information sources are described by the canonical ER model, the space to store the information that is generated by it is equal to

$$\ln |T_{\text{can}}^{\epsilon=0}| = l \times S_{\text{can}} = -l \times \ln P_{\text{can}}(\mathbf{G}^*). \tag{4.22}$$

It is decided by the canonical entropy of the ER model.

The difference between the limit of information storage of the information source described by different ensembles is related to the relative entropy between the microcanonical and canonical ER model, which is equal to $S_{\text{can}} - S_{\text{mic}}$ when the two ensembles are conjugate with each other. We can get the result of relative entropy based on Stirling's formula as

$$S(P_{\text{mic}}||P_{\text{can}}) \approx \ln \sqrt{2\pi C^*(1 - 2C^*/[n(n-1)])}. \tag{4.23}$$

In the thermodynamic limit, the two ER model ensembles are equivalent to each other, as the limit value of the relative entropy density is equal to 0

$$\lim_{n\to\infty} \frac{1}{n} S(P_{\text{mic}}||P_{\text{can}}) = 0. \tag{4.24}$$

The space can be saved from canonical ensemble description to microcanonical ensemble description is equal to $l \times S(P_{\text{mic}}||P_{\text{can}})$, but compared with the total space needs in the canonical ensemble description, it is not so important, as the space ratio r will equal to 0 in the thermodynamic limit

$$r = \frac{S(P_{\text{mic}}||P_{\text{can}})}{S_{\text{can}}} = 0. \tag{4.25}$$

Therefore, using the microcanonical ensemble only saves finite space of the information storage. The ensemble equivalence allows us to choose the canonical ensemble, which is mathematically easy to obtain.

Appendix 4.B Matrix described information sources

Matrix is a general model widely used to describe natural systems with heterogeneous interactions. [27]. The heterogeneous interactions among the units imply the local constraints in it. It means this information sources described by the matrix ensembles are under ensemble nonequivalence.

Thus, checking if the limit of information storage of the matrix ensemble described information source will be affected by the ensemble nonequivalence as we predicted is significant for our theory. We will start with matrice under global constraint, then extend the discussion to the local constrained one.

4.B.1 Matrix with global constraint

Global constraint is the sum of all the elements in each matrix is fixed as $C(\mathbf{G}) = \sum_{i=1,j=1}^{n,m} g_{ij}$. When the global constraint is hard, the constraint of each state in the matrix ensemble is equal to each other as C^*. The system can be described by the microcanonical ensemble. If the global constraint is soft, the average value of each matrix's constraints is equal to the hard constraints, $\langle C(\mathbf{G})\rangle = C^*$, the system can be described by the canonical ensemble. States both in the microcanonical and canonical ensemble with constraints equal to C^* is represented by $\mathbf{G}^*$.

The probability of each state in the matrix described by the microcanonical ensemble is equal to

$$P_{\text{mic}}(\mathbf{G}^*) = \frac{1}{\Omega_{C^*}}, \tag{4.26}$$

where Ω_{C^*} is the total number of states with global constraint equals to C^*.

In the matrix described by the canonical ensemble, the *Hamiltonian* of each matrix

$$H = \theta^* \cdot C(\mathbf{G}) \tag{4.27}$$

decides the probability of it. The θ^* is the maximum likelihood parameter realized $\langle C(\mathbf{G})\rangle = C^*$ and maximum the Shannon entropy. Thus, the probability of each state in the canonical ensemble is equal to

$$P_{\text{can}}(\mathbf{G}) = \frac{e^{-H(\mathbf{G},\theta^*)}}{Z(\theta^*)}, \tag{4.28}$$

$Z(\theta^*)$ is the partition function, and it is a normalization constant equal to $Z(\theta^*) = \sum_{\mathbf{G}\in\mathcal{G}} e^{-H}$.

If the matrix with global constraints is used as an information source, then the information generated from it is carried by a set of matrix ensemble sequences with length l.

When the global constrained matrix is described by the microcanonical ensemble, the space to store the information generated by it is equal to

$$\ln|T^{\epsilon}_{\text{mic}}| = l \times S_{\text{mic}} = l \times \ln \Omega_{C^*}. \tag{4.29}$$

For the canonical ensemble described matrix information sources, the space to store the information generated by it is equal to

$$\ln|T^{\epsilon}_{\mathrm{can}}| = l \times S_{\mathrm{can}} = -l \times \ln P_{\mathrm{can}}(\mathbf{G}^*). \tag{4.30}$$

Relative entropy between the two ensembles' probability distribution is equal to $S_{\mathrm{can}} - S_{\mathrm{mic}}$. It can be estimated by the determinant of the covariance matrix of constraints in the canonical ensemble as

$$S(P_{\mathrm{mic}}||P_{\mathrm{can}}) \approx \ln\sqrt{2\pi\mathbf{\Sigma}^*} = \frac{1}{2}\ln 2\pi\frac{\partial^2 \ln Z(\theta^*)}{\partial{\theta^*}^2}. \tag{4.31}$$

When the elements in the matrix is chosen from different set, partition function $Z(\theta^*)$ is also different. For example when the element in the matrix is equal to 1 or 0, this matrix is binary matrix, the partition function of it is equal to $Z(\theta^*) = (e^{-\theta^*}+1)^{mn}$. If the element in the matrix is chosen from the whole natural number set, the matrix is a weighted matrix, the partition function of it is equal to $Z(\theta^*) = (1-e^{-\theta^*})^{-mn}$. According to the relationship between θ^* and C^* in the two different matrices, we can find the value of relative entropy of binary matrix is equal to $S(P_{\mathrm{mic}}||P_{\mathrm{can}}) = \frac{1}{2}\ln[2\pi C^*(1-C^*/(mn))]$, and the relative entropy of weighted matrix is equal to $S(P_{\mathrm{mic}}||P_{\mathrm{can}}) = \frac{1}{2}\ln[2\pi C^*(1+C^*/(mn))]$.

Because the value of $S(P_{\mathrm{mic}}||P_{\mathrm{can}})$ for the two different kinds of matrices are both grows like $o(n)$, the two different ensemble descriptions are equivalent to each other in the thermodynamic limit.

The ratio of space that can be saved from canonical ensemble description to microcanonical ensemble description is

$$r = [\frac{1}{2}\ln 2\pi\frac{\partial^2 \ln Z(\theta^*)}{\partial{\theta^*}^2}]/[-\theta^*\cdot C^* - \ln Z(\theta^*)]. \tag{4.32}$$

The value r of the binary matrix under global constraint is equal to

$$r = \frac{1}{2}\frac{\ln[2\pi C^*(1-C^*/(mn))]}{mn\ln(mn) - C^*\ln C^* - (mn-C^*)\ln(mn-C^*)}. \tag{4.33}$$

The value of r for the weighted matrix ensemble under global constraints is equal to

$$r = \frac{1}{2}\frac{\ln[2\pi C^*(1+C^*/(mn))]}{(mn+C^*)\ln(mn+C^*) - mn\ln(mn) - C^*\ln C^*}, \tag{4.34}$$

When n goes to infinite, the r is equal to 0 both in the two matrices. Therefore, when the system with global constraint is used as the information source, it is under ensemble equivalence. The space saved from canonical to microcanonical ensemble description can be neglected.

4.B.2 Matrix with local constraints

The local constraints of the information sources are implied by the heterogeneous interactions among the units in it. The local constraint is the sum of all the elements in each row or column in the matrix ensemble. According to the research in [27], matrix ensemble with local constraints is under ensemble nonequivalence. When the value of rows in the matrix under local constraints is finite $m \ll \infty$, the ensemble nonequivalence is as strong as the one in the boundary of phase transition [8].

In this section, we will introduce how the heterogeneous interaction will affect the limit of information storage. As the coupled constraints only can be analytically solved in two particular cases, we will put all the calculations on the matrix with one-sided local constraints [27].

The $m \times n$ matrix ensemble under local column constraints $\vec{C}^* = [c_1^*, c_2^*, \cdots, c_i^*, \cdots, c_n^*]$ has n constraints in it. Each c_i^* is the sum of all the elements in the column i as $c_i^* = \sum_{j=1}^{m} g_{ij}$. The property of constraints decides which ensembles will be used to describe this local constrained matrix.

In the microcanonical ensemble description, each state still have the same value of constraints as $\vec{C}^*$, and the probability of it is equal to

$$P_{\text{mic}}(\mathbf{G}^*) = 1/\Omega_{\vec{C}^*}, \tag{4.35}$$

where $\Omega_{\vec{C}^*}$ is the number of states in the matrix described by the microcanonical ensemble. In binary matrix, $\Omega_{\vec{C}^*} = \prod_{i=1}^{n} \binom{m}{r_i^*}$. In weighted matrix, $\Omega_{\vec{C}^*} = \prod_{i=1}^{n} \binom{m+r_i^*-1}{r_i^*}$. The space to store the information generated by it is equal to

$$\ln |T_{\text{mic}}^{\epsilon=0}| = l \times S_{\text{mic}} = l \times \ln \Omega_{\vec{C}^*}, \tag{4.36}$$

both in the binary and weighted matrix.

When the local column constraints are soft, the matrix needs to be described by the canonical ensemble. The probability of states in the canonical ensemble is also based on the *Hamiltonian* of it, which is defined as $H = \sum_{i=1}^{n} \beta_i^* c_i^*$. Where β_i^* is the correspond parameter which maximum the Shannon entropy and realized the $\langle \vec{C}(\mathbf{G}) \rangle = \vec{C}^*$. Therefore, the probability of states in the canonical matrix ensemble is

$$P_{\text{can}}(\mathbf{G}) = \frac{e^{-H}}{Z(\vec{\beta}^*)}. \tag{4.37}$$

The information generated by it is also carried in the canonical matrix ensemble sequences. The space to store the information is equal to

$$\ln |T_{\text{can}}^{\epsilon=0}| = l \times S_{\text{can}} = -l \times \ln P_{\text{can}}(\mathbf{G}^*). \tag{4.38}$$

The space saved from the canonical description to the microcanonical ensemble description can be estimated by the function of the determinant of the covariance matrix of constraints in the canonical ensemble as

$$\begin{aligned} S(P_{\mathrm{mic}}||P_{\mathrm{can}}) &= \frac{1}{2}\sum_{k=1}^{n} \ln \frac{2\pi\lambda_k^*}{[1+O(1/\lambda_k^*)]^2} \\ &\approx \frac{1}{2}\sum_{k=1}^{n} \ln[2\pi \frac{\partial^2 \ln Z(\vec{\beta}^*)}{\partial \beta_k^{*2}}]. \end{aligned} \tag{4.39}$$

As different matrices have different difference of partition function, so the relative entropy is also different. Partition function of binary matrix under is equal to $Z(\vec{\beta}^*) = \prod_{i=1}^{n}(e^{-\beta_i^*}+1)^m$. In weighted matrix, partition function is equal to $Z(\vec{\beta}^*) = \prod_{i=1}^{n}(1-e^{-\beta^*})^{-m}$. Then the value of relative entropy for binary matrix is equal to $S(P_{\mathrm{mic}}||P_{\mathrm{can}}) = \frac{1}{2}\sum_{i=1}^{n}\ln[2\pi\frac{r_i^*(m-r_i^*)}{m}]$. In weighted matrix, the value of relative entropy is equal to $S(P_{\mathrm{mic}}||P_{\mathrm{can}}) = \frac{1}{2}\sum_{i=1}^{n}\ln[2\pi\frac{r_i^*(m+r_i^*)}{m}]$. Both of those two matrix are under ensemble nonequivalence. The space can be saved from the canonical ensemble description to the microcanonical ensemble description has the same order as the increase of the ensemble sequences' length. The ratio r is still defined as

$$r = [\frac{1}{2}\sum_{k=1}^{n}\ln[2\pi\frac{\partial^2 \ln Z(\vec{\beta}^*)}{\partial\beta_k^{*2}}]]/[-\vec{\beta}^*\cdot\vec{C}^* - \ln Z(\vec{\beta}^*)]. \tag{4.40}$$

In the binary matrix ensemble under local column constraints, the value of r is equal to

$$r = 1 - [\sum_{i=1}^{n}\ln\binom{m}{r_i^*}]/[\sum_{i=1}^{n}\ln[m^m/r_i^{*r_i^*}(m-r_i^*)^{m-r_i^*}]]. \tag{4.41}$$

In the weighted matrix ensemble, the value of r is equal to

$$r = 1 - [\sum_{i=1}^{n}\ln\binom{m+r_i^*-1}{r_i^*}]/[\sum_{i=1}^{n}\ln[\frac{(m-r_i^*)^{m-r_i^*}}{m^m r_i^{*r_i^*}}]. \tag{4.42}$$

When the matrix is in the thermodynamic limit, the limit value of r grows like $\frac{\ln(2\pi m)}{m}$. Thus, when the freedom of each element is finite $m \ll \infty$, the ratio is fixed. When the value of m is growing like $O(n)$, the ratio is close to 0. It means compare with canonical ensemble description, using the microcanonical ensemble will save $r\%$ of the space.

Under two-sided local constraints, it is impossible to calculate the number of states in the microcanonical ensemble. The increased number of constraints will decrease the possible configurations in the microcanonical ensemble, so the space to store the information generated by it is even smaller than the one with local column constraints [27]. The ensemble nonequivalence will affect the information-theoretical bounds.