

Convergent sequences of sparse graphs: A large deviations approach

Christian Borgs*

Jennifer Chayes†

David Gamarnik‡

February 10, 2013

Abstract

In this paper we introduce a new notion of convergence of sparse graphs which we call Large Deviations or LD-convergence and which is based on the theory of large deviations. The notion is introduced by "decorating" the nodes of the graph with random uniform i.i.d. weights and constructing random measures on $[0, 1]$ and $[0, 1]^2$ based on the decoration of nodes and edges. A graph sequence is defined to be converging if the corresponding sequence of random measures satisfies the Large Deviations Principle with respect to the topology of weak convergence on bounded measures on $[0, 1]^d, d = 1, 2$. We then establish that LD-convergence implies several previous notions of convergence, namely so-called right-convergence, left-convergence, and partition-convergence. The corresponding large deviation rate function can be interpreted as the limit object of the sparse graph sequence. In particular, we can express the limiting free energies in terms of this limit object.

Finally, we establish several previously unknown relationships between the formerly defined notions of convergence. In particular, we show that partition-convergence does not imply left or right-convergence, and that right-convergence does not imply partition-convergence.

*Microsoft Research New England; e-mail: borgs@microsoft.com.

†Microsoft Research New England; e-mail: jchayes@microsoft.com.

‡MIT; e-mail: gamarnik@mit.edu. Research supported by the NSF grants CMMI-1031332.

1 Introduction

The theory of convergent graph sequences for dense graphs is by now a well-developed subject, with many *a priori* distinct notions of convergence proved to be equivalent. For sparse graph sequences (which in this paper we take to be sequences with bounded maximal degree), much less is known. While several of the notions defined in the context of dense graphs were generalized to the sparse setting, and some other interesting notions were introduced in the literature, the equivalence between the different notions was either unknown or known not to hold. In some cases, it was not even known whether one is stronger than the other or not. In this paper, we introduce a new natural notion of convergence for sparse graphs, which we call Large Deviations Convergence, and relate it to other notions of convergence. We hope that this new notion will also ultimately allow us to relate many of the previous notions to each other.

Let \mathbb{G}_n be a sequence graphs such that the number of vertices in \mathbb{G}_n goes to infinity. The question we address here is the notion of convergence and limit of such a sequence.

For dense graphs, i.e., graphs for which the average degree grows like the number of vertices, this question is by now well understood [BCL⁺06],[BCL⁺08],[BCL⁺12]. These works introduced and showed the equivalence of various notions of convergence: *left-convergence*, defined in terms of homomorphisms from a small graph \mathbb{F} into \mathbb{G}_n ; *right-convergence*, defined in terms of homomorphisms from \mathbb{G}_n into a weighted graph \mathbb{H} with strictly positive edge weights; convergence in terms of the so-called *cut metric*; and several other notions, including convergence of the set of *quotients* of the graphs \mathbb{G}_n , a notion which will play a role in this paper as well. The existence and properties of the limit object were established in [LS06], and its uniqueness was proved in [BCL10]. Some lovely follow up work on the dense case gave alternative proofs in terms of exchangeable random variables [DJ07], and provided applications using large deviations [CV11].

For sparse graph sequences much less is known. In [BCKL13], *left-convergence* was defined by requiring that the limit

$$t(\mathbb{F}) = \lim_{n \rightarrow \infty} \frac{1}{|V(\mathbb{G}_n)|} \text{hom}(\mathbb{F}, \mathbb{G}_n) \quad (1)$$

exists for every connected, simple graph \mathbb{F} , where $\text{hom}(\mathbb{F}, \mathbb{G})$ is used to denote the number of homomorphisms from \mathbb{F} into \mathbb{G} . Here and in the rest of this paper $V(\mathbb{G})$ denotes the vertex set of \mathbb{G} . Also $E(\mathbb{G})$ denotes the edge set of \mathbb{G} and for $u, v \in V(\mathbb{G})$ we write $u \sim v$ if $(u, v) \in E(\mathbb{G})$. It is easy to see that for sequences of graphs with bounded degrees, this notion is equivalent to an earlier notion of convergence, the notion of *local convergence* introduced by Benjamini and Schramm [BS01].

To define right-convergence, one again considers homomorphisms, but now from \mathbb{G}_n into a small “target” weighted graph \mathbb{H} . Each homomorphism is equipped with the weight induced by \mathbb{H} . Since the total weight of all homomorphisms denoted by $\text{hom}(\mathbb{G}_n, \mathbb{H})$ is at most exponential in $|V(\mathbb{G}_n)|$, it is natural to consider the sequence

$$f(\mathbb{G}_n, \mathbb{H}) = -\frac{1}{|V(\mathbb{G}_n)|} \log \text{hom}(\mathbb{G}_n, \mathbb{H}), \quad (2)$$

which in the terminology of statistical physics is the sequence of free energies. In [BCKL13] a sequence \mathbb{G}_n was defined to be right-convergent with respect to a given graph \mathbb{H} if the free energies $f(\mathbb{G}_n, \mathbb{H})$ converge as $n \rightarrow \infty$. It was shown that a sequence \mathbb{G}_n that is right-convergent on all simple graphs \mathbb{H} is also left-convergent. But the converse was shown not to be true. The

counterexample is simple: Let C_n be the cycle on n nodes, and let \mathbb{H} be a bipartite graph. Then $\text{hom}(C_n, \mathbb{H}) = 0$ for odd n , while $\text{hom}(C_n, \mathbb{H}) \geq 1$ for even n .

This example, however, relies crucially on the fact that \mathbb{H} encodes so-called *hard-core* constraints, meaning that some of the maps $\phi : V(\mathbb{G}_n) \rightarrow V(\mathbb{H})$ do not contribute to $\text{hom}(\mathbb{G}_n, \mathbb{H})$. As we discuss in Section 4, such hard-core constraints are not very natural in many respects, one of them being that the existence or value of the limit of the sequence (2) can change if we remove a sub-linear number of edges from \mathbb{G}_n (in the case of C_n , just removing a single edge makes the sequence convergent, as can be seen by a simple sub-additivity argument). Hard-core constraints are also not very natural from a point of view of applications in physics, where hard-core constraints usually represent an idealized limit like the zero-temperature limit, which is a mathematical idealization of something that in reality can never be realized.

We therefore propose to *define* right-convergence by restricting ourselves to so-called soft-core graphs, i.e., graphs \mathbb{H} with strictly positive edge weights. In other words, we define \mathbb{G}_n to be *right-convergent* iff the free energy defined in (2) has a limit for every soft-core graph \mathbb{H} . This raises the question of how to relate the existence of a limit for models with hard-core constraints to those on soft-core graphs, a question we will address in Section 4.2 below, but we do not make this part of the definition of right-convergence. Given our relaxed definition of the right-convergence, the question then remains whether left-convergence possibly implies right-convergence. The answer turns out to be still negative, and involves an example already considered in [BCKL13], see Section 5.1 below.

Next let us turn to the notion of partition-convergence of sparse graphs, a notion which was introduced by Bollobas and Riordan [BR11], motivated by a similar notion for dense graphs from [BCL⁺08], [BCL⁺12], where it was called convergence of quotients. We start by introducing quotients for sparse graphs. Fix a graph \mathbb{G} and let $\sigma = (V_1, \dots, V_k)$ be a partition of $V(\mathbb{G})$ into disjoint subsets (where some of the V_i 's may be empty). We think of this partition as a (non-proper) coloring $\sigma : V(\mathbb{G}) \rightarrow [k]$ with $V_i = \sigma^{-1}(i)$, $1 \leq i \leq k$. The partition σ then induces a weighted graph $\mathbb{F} = \mathbb{G}/\sigma$ on $[k] = \{1, \dots, k\}$ (called a k -quotient) by setting

$$x_i(\sigma) = \frac{|V_i|}{|V|} \quad \text{and} \quad X_{ij}(\sigma) = \frac{1}{|V|} |\{u \in V_i, v \in V_j : u \sim v\}|. \quad (3)$$

Here $x(\sigma) = (x_i(\sigma), 1 \leq i \leq k)$ and $X(\sigma) = (X_{i,j}(\sigma), 1 \leq i, j \leq k)$ are the set of node and edge weights of \mathbb{F} , respectively. Then for a given k , the set of all possible k -quotients $(x(\sigma), X(\sigma))$ is a discrete subset $\mathcal{S}_k(\mathbb{G}) \subset \mathbb{R}_+^{(k+1)k}$. Henceforth $\mathbb{R}(\mathbb{R}_+)$ denotes the set of all (non-negative) reals, and $\mathbb{Z}(\mathbb{Z}_+)$ denotes the set of all (non-negative) integers. Motivated by the corresponding notion from dense graphs, one might want to study whether the sequence of k -quotients $\mathcal{S}_k(\mathbb{G}_n)$ converges to a limiting set $\mathcal{S}_k \subset \mathbb{R}_+^{(k+1)k}$ with respect to some appropriate metric on the subsets of $\mathbb{R}_+^{(k+1)k}$. This leads to the notion of partition-convergence, a notion introduced in [BR11]. Since it is not immediate whether this notion of convergence implies, for example, left-convergence (in fact, one of the results in this paper is that it does not), Bollobas and Riordan [BR11] also introduced a stronger notion of colored-neighborhood-convergence. This notion was further studied by Hatami, Lovász and Szegedi [HLS12]. An implication of the results of these and earlier papers (details below) is that colored-neighborhood-convergence is strictly stronger than the notion of both partition- and left-convergence, but it is not known whether it implies right-convergence.

We now discuss the main contribution of this paper: the definition of convergent sparse graph sequences using the formalism of large deviations theory. We define a graph sequence to be Large Deviations (LD)-convergent if for every k , the weighted factor graphs $\mathbb{F} = G/\sigma$ defined above viewed as a vector $(x_i(\mathbb{F}), 1 \leq i \leq k)$ and matrix $(X_{i,j}(\mathbb{F}), 1 \leq i, j \leq k)$, satisfy the large deviation principle in \mathbb{R}^k and $\mathbb{R}^{k \times k}$, when the k -partition σ is chosen uniformly at random.¹ Intuitively, the large deviations rate associated with a given graph \mathbb{F} provides the limiting exponent for the number colorings σ such that the corresponding factor graphs \mathbb{G}/σ is approximately \mathbb{F} . It turns out that the large deviations rates provide enough information to "read off" the limiting partition sets \mathcal{S}_k . Those are obtained as partitions with finite large deviations rate. Similarly, one can "read off" the limits of free energies (2). As a consequence, the LD-convergence implies both the partition and right-convergence.

The deficiency of the definition above is that it requires that the large deviations principle holds for a infinite collection of probability measures associated with the choice of k . It turns out that there is a more elegant unifying way to introduce LD-convergence by defining just one large deviations rate, but on the space of random measures rather than the space of random weighted graphs \mathbb{G}_n/σ . This is done as follows. Given a sequence of sparse graphs \mathbb{G}_n , suppose the nodes of each graph are equipped with values $(\sigma(u), u \in V(\mathbb{G}))$ chosen independently uniformly at random from $[0, 1]$. We may think of these values as real valued "colors". From these values we construct a one-dimensional measure $\rho_n = \sum_{u \in V(\mathbb{G}_n)} |V(\mathbb{G}_n)|^{-1} \delta(\sigma(u))$ on $[0, 1]$ and a two-dimensional measure $\mu_n = \sum_{u,v: u \sim v} |V(\mathbb{G}_n)|^{-1} \delta(\sigma(u), \sigma(v))$ on $[0, 1]^2$. Here $\delta(x)$ is a measure with unit mass at x and zero elsewhere. As such we obtain a sequence of measures (ρ_n, μ_n) . We define the graph sequence to be LD-convergent if this sequence of random measures satisfy the Large Deviations Principle on the space of measures on $[0, 1]^d$, $d = 1, 2$ equipped with the weak convergence topology. One of our first result is establishing the equivalence of two definitions of the LD-convergence. To distinguish the two, the former mode of convergence is referred to as k -LD-convergence, and the latter simply as LD-convergence.

Our most important result concerning the new definition of graph convergence is that LD-convergence implies right (and therefore left) as well as partition-convergence. We conjecture that it also implies colored-neighborhood-convergence but we do not have a proof at this time. Finally, we conjecture that LD-convergence holds for a sequence of random D -regular graphs with high probability (w.h.p.), but at the present stage we are very far from proving this conjecture and we discuss important implications to the theory of spin glasses should this convergence be established.

Let us finally return to a question already alluded to when we defined right-convergence. From several points of view, it seems natural to consider two sequence \mathbb{G}_n and $\tilde{\mathbb{G}}_n$ on the same vertex set $V_n = V(\mathbb{G}_n) = V(\tilde{\mathbb{G}}_n)$ to be equivalent if the edge sets $E(\mathbb{G}_n)$ and $E(\tilde{\mathbb{G}}_n)$ differ on a set of $o(|V_n|)$ edges. One natural condition one might require of all definitions of convergence is the condition that convergence of \mathbb{G}_n implies that of $\tilde{\mathbb{G}}_n$ and vice versa, whenever \mathbb{G}_n and $\tilde{\mathbb{G}}_n$ are equivalent. We show that this holds for all notions of convergence considered in the paper, namely left-convergence, our modified version of right-convergence, partition-convergence, colored-neighborhood-convergence, and LD-convergence. We also show that for all right-convergent sequences \mathbb{G}_n , one can find an equivalent sequence $\tilde{\mathbb{G}}_n$ such that on $\tilde{\mathbb{G}}_n$, the limiting free energy $f(\mathbb{H}) = \lim_{n \rightarrow \infty} f(\tilde{\mathbb{G}}_n, \mathbb{H})$ exists even when \mathbb{H} contains hard-core constraints, see

¹Some necessary background on Large Deviations Principle and weak convergence of measures will be provided in Section 3.

Theorem 3 below. Our alternative definition of right-convergence can therefore be reexpressed by the condition that there exists a sequence $\tilde{\mathbb{G}}_n$ equivalent to \mathbb{G}_n such that the free energies $f(\tilde{\mathbb{G}}_n, \mathbb{H})$ converge for all weighted graphs \mathbb{H} .

The organization of this paper is as follows. In the next section we introduce notations and definitions and provide the necessary background on the large deviations theory. In Section 3 we introduce LD-convergence, establish some basic properties and consider two examples of LD-convergent graph sequences. Other notions of convergence are discussed in Section 4. In the same section we prove Theorem 3 relating free energy limits with respect to hard-core and soft models, discussed above. The relationship between different notions of convergence is considered in Section 5. The main result of this section is Theorem 4, which establishes relationships between different notions of convergence. Figures 3 and 4 are provided to illustrate the relations between the notions of convergence, both those known earlier and those established in this paper. In the last Section 6 we compare the expressions for the limiting free energy to those in the dense case (pointing out that the large-deviations rate function plays a role quite similar to the role the limiting graphon played in the dense case) and discuss some open questions.

2 Notations and definitions

For the convenience of the reader, we repeat some of the notations already used in the introduction. We consider in this paper finite simple undirected graphs \mathbb{G} with node and edge sets denoted by $V(\mathbb{G})$ and $E(\mathbb{G})$ respectively. We write $u \sim v$ for nodes u and v when $(u, v) \in E(\mathbb{G})$. For every node $u \in V(\mathbb{G})$, $\mathcal{N}(u)$ denotes the set of neighbors of u , namely all $v \in V(\mathbb{G})$ such that $u \sim v$. Sometimes we will write $\mathcal{N}_{\mathbb{G}}(u)$ in order to emphasize the underlying graph. The number $\Delta_{\mathbb{G}}(u)$ of vertices in $\mathcal{N}_{\mathbb{G}}(u)$ is called the degree of u in \mathbb{G} . We use $\Delta_{\mathbb{G}}$ to denote the maximum degree $\max_u \Delta_{\mathbb{G}}(u)$.

A path of length r is a sequence of nodes u_0, \dots, u_r such that $u_i \sim u_{i+1}$ for all $i = 0, 1, 2, \dots, r-1$. The distance between nodes u and v is the length of the shortest path u_0, \dots, u_r with $u_0 = u$ and $u_r = v$. The distance is assumed to be infinite if u and v belong to different connected components of \mathbb{G} . Given u and r , let $B(u, r) = B_{\mathbb{G}}(u, r)$ be the subgraph of \mathbb{G} induced by nodes with distance at most r from u . Given a set of nodes $A \subset V$ and $r \geq 1$, $B(A, r)$ is the graph induced by the set of nodes v with distance at most r from some node $u \in A$.

Given two graphs \mathbb{G}, \mathbb{H} a map $\sigma : V(\mathbb{G}) \rightarrow V(\mathbb{H})$ is a graph homomorphism if $(u, v) \in E(\mathbb{G})$ implies $(\sigma(u), \sigma(v)) \in E(\mathbb{H})$ for every $u, v \in V(\mathbb{G})$. Namely, σ maps edges into edges. $\sigma : \mathbb{G}_1 \rightarrow \mathbb{G}_2$ is called a graph isomorphism if it is a bijection and $(u, v) \in E(\mathbb{G}_1)$ if and only if $(\sigma(u), \sigma(v)) \in E(\mathbb{G}_2)$, i.e., σ is an isomorphism which maps edges into edges, and non-edges into non-edges. Two graphs are called isomorphic if there is a graph isomorphism mapping them into each other. Two rooted graphs (\mathbb{G}_1, x) and (\mathbb{G}_2, y) are called isomorphic if there is an isomorphism $\sigma : \mathbb{G}_1 \rightarrow \mathbb{G}_2$ such that $\sigma(x) = y$.

The edit distance between two graphs \mathbb{G} and $\tilde{\mathbb{G}}$ with the same number of vertices is defined as

$$\delta_{\text{edit}}(\mathbb{G}, \tilde{\mathbb{G}}) = \min_{\mathbb{G}'} \left(|E(\mathbb{G}') \setminus E(\tilde{\mathbb{G}})| + |E(\tilde{\mathbb{G}}) \setminus E(\mathbb{G}')| \right)$$

where the minimum goes over all graph \mathbb{G}' isomorphic to \mathbb{G} .

Given two graphs \mathbb{G}, \mathbb{H} , we use $\text{Hom}(\mathbb{G}, \mathbb{H})$ to denote the set of all homomorphisms from \mathbb{G} to \mathbb{H} , and $\text{hom}(\mathbb{G}, \mathbb{H}) = |\text{Hom}(\mathbb{G}, \mathbb{H})|$ to denote the number of homomorphisms from \mathbb{G} to \mathbb{H} .

If \mathbb{H} is weighted, with node weights given by a vector $(\alpha_i, i \in V(\mathbb{H}))$ and edge weights given by a symmetric matrix $(A_{ij}, (i, j) \in E(\mathbb{H}))$ (which we denote by $\alpha = \alpha(\mathbb{H})$ and $A = A(\mathbb{H})$ respectively), then $\text{hom}(\mathbb{G}, \mathbb{H})$ is defined by

$$\text{hom}(\mathbb{G}, \mathbb{H}) = \sum_{\sigma: V(\mathbb{G}) \rightarrow V(\mathbb{H})} \prod_{u \in V(\mathbb{G})} \alpha_{\sigma(u)} \prod_{(u, v) \in E(\mathbb{G})} A_{\sigma(u), \sigma(v)}. \quad (4)$$

Clearly, if α is a vector of ones and A is a matrix consisting of zeros and ones with zeros on the diagonal, then the definition reduces to the unweighted case where A denotes the adjacency matrix of a simple undirected graph \mathbb{H} .

A graph sequence $\mathbb{G}_n = (V(\mathbb{G}_n), E(\mathbb{G}_n)), n \geq 1$ is defined to be sparse if $\Delta_{\mathbb{G}_n} \leq D$ for some finite D for all n . When dealing with graph sequences we write V_n and E_n instead of $V(\mathbb{G}_n)$ and $E(\mathbb{G}_n)$, respectively. Two graph sequences \mathbb{G}_n and $\tilde{\mathbb{G}}_n$ are defined to be equivalent, in which case we write $\mathbb{G}_n \sim \tilde{\mathbb{G}}_n$, if $\tilde{\mathbb{G}}_n$ has the same number of nodes as \mathbb{G}_n and $\delta_{\text{edit}}(\mathbb{G}_n, \tilde{\mathbb{G}}_n) = o(|V(\mathbb{G}_n)|)$ as $n \rightarrow \infty$. Observe that \sim defines an equivalency relationship on the set of sequences of graphs.

Next we review some concepts from large deviations theory. Given a metric space S equipped with metric d , a sequence of probability measures \mathbb{P}_n on Borel sets of S is said to satisfy the Large Deviations Principle (LDP) with rate $\theta_n > 0, \lim_n \theta_n = \infty$, if there exists a lower semi-continuous function $I : S \rightarrow \mathbb{R}_+ \cup \{\infty\}$ such that for every (Borel) set $A \subset S$

$$-\inf_{x \in A^\circ} I(x) \leq \liminf_n \frac{\log \mathbb{P}_n(A^\circ)}{\theta_n} \leq \limsup_n \frac{\log \mathbb{P}_n(\bar{A})}{\theta_n} \leq -\inf_{x \in \bar{A}} I(x), \quad (5)$$

where A° and \bar{A} denote the interior and the closure of the set A , respectively. In this case, we say that (\mathbb{P}_n) obeys the LDP with rate function I . Typically $\theta_n = n$ is used in most cases, but in our case we will be considering the normalization $\theta_n = |V(\mathbb{G}_n)|$ corresponding to some sequence of graphs \mathbb{G}_n , and it is convenient not to assume that n is the number of nodes in \mathbb{G}_n .

The definition above immediately implies that the rate function I is uniquely recovered as

$$I(x) = -\lim_{\epsilon \rightarrow 0} \liminf_n \frac{\log \mathbb{P}_n(B(x, \epsilon))}{\theta_n} = -\lim_{\epsilon \rightarrow 0} \limsup_n \frac{\log \mathbb{P}_n(B(x, \epsilon))}{\theta_n}, \quad (6)$$

for every $x \in S$, where $B(x, \epsilon) = \{y \in S : d(x, y) \leq \epsilon\}$. Indeed, from lower semi-continuity we have

$$I(x) = \lim_{\epsilon \rightarrow 0} \inf_{y \in B^\circ(x, \epsilon)} I(y) = \lim_{\epsilon \rightarrow 0} \inf_{y \in \bar{B}(x, \epsilon)} I(y), \quad (7)$$

from which the claim follows. In fact the existence and the equality of double limits in (6) is also sufficient for the LDP to hold when S is compact. We provide a proof here for completeness.

Proposition 1. *Given a sequence of probability measures \mathbb{P}_n on a compact metric space S , and given a sequence $\theta_n > 0, \lim_n \theta_n = \infty$, suppose the limits in the second and third term in (6) exist and are equal for every x . Let $I(x)$ be defined as the negative of these terms for every $x \in S$. Then the LDP holds with rate function I and normalization θ_n .*

Proof. We first establish that I is lower semi-continuous. Fix $x, x_m \in S$ such that $x_m \rightarrow x$. For every ϵ , find m_0 such that $d(x, x_m) \leq \epsilon/2$ for $m \geq m_0$. We have for all $m \geq m_0$,

$$-I(x_m) \leq \limsup_n \frac{1}{\theta_n} \log \mathbb{P}_n(B(x_m, \epsilon/2)) \leq \limsup_n \frac{1}{\theta_n} \log \mathbb{P}_n(B(x, \epsilon)),$$

implying

$$\limsup_m (-I(x_m)) \leq \lim_{\epsilon \rightarrow 0} \limsup_n \frac{1}{\theta_n} \log \mathbb{P}_n(B(x, \epsilon)) = -I(x).$$

Therefore I is lower semi-continuous.

It remains to verify (5). Fix an arbitrary closed (and therefore compact) $A \subset S$, and let $\delta > 0$. Then

$$\limsup_n \frac{1}{\theta_n} \log \mathbb{P}_n(A) \leq \frac{1}{\theta_n} \log \mathbb{P}_n(A) + \delta$$

for an infinite number of values for n . Let N_0 be the set of n for which this holds. Further, given $m \in \mathbb{Z}_+$, we can find $x_1, \dots, x_{N(m)} \in A$ such that $A \subset \cup_{i \leq N(m)} B(x_i, 1/m)$. Let $i(m, n) \leq N(m)$ be the index corresponding to a largest value of $\mathbb{P}_n(B(x_i, 1/m))$. There exists an $i(m)$ such that $i(m, n) = i(m)$ for an infinite number of values $n \in N_0$. Denote the set of n for which this holds by N_m , and assume without loss of generality that $\theta_n^{-1} \log N(m) \leq \delta$ for all $n \in N_m$. For such n , we then have

$$\mathbb{P}_n(A) \leq \sum_{i \leq N(m)} \mathbb{P}_n(B(x_i, 1/m)) \leq N(m) \mathbb{P}_n(B(x_{i(m)}, 1/m))$$

and hence

$$\limsup_n \frac{1}{\theta_n} \log \mathbb{P}_n(A) \leq 2\delta + \frac{1}{\theta_n} \log \mathbb{P}_n(B(x_{i(m)}, 1/m)).$$

Find any limit point $x \in A$ of $x_{i(m)}$ as $m \rightarrow \infty$. Fix $\epsilon > 0$. Find m_1 such that $B(x_{i(m_1)}, 1/m_1) \subset B(x, \epsilon)$, and hence

$$\limsup_n \frac{1}{\theta_n} \log \mathbb{P}_n(A) \leq 2\delta + \frac{1}{\theta_n} \log \mathbb{P}_n(B(x, \epsilon))$$

whenever $n \in N_{m_1}$. Since N_{m_1} contains infinitely many $n \in \mathbb{Z}_+$, we conclude that

$$\limsup_n \frac{1}{\theta_n} \log \mathbb{P}_n(A) \leq 2\delta + \limsup_n \frac{1}{\theta_n} \log \mathbb{P}_n(B(x, \epsilon)).$$

Since this holds for every δ and ϵ , we may apply (6) to obtain

$$\limsup_n \frac{1}{\theta_n} \log \mathbb{P}_n(A) \leq -I(x) \leq -\inf_{y \in A} I(y),$$

thus verifying the upper bound in (5).

For the lower bound, suppose $A \subset S$ is open. Fix $\epsilon > 0$ and find $x \in A$ such that $I(x) \leq \inf_{y \in A} I(y) + \epsilon$. Find δ such that $B(x, \delta) \subset A$. We have

$$\begin{aligned} \liminf_n \frac{1}{\theta_n} \log \mathbb{P}_n(A) &\geq \liminf_n \frac{1}{\theta_n} \log \mathbb{P}_n(B(x, \delta)) \\ &\geq -I(x) \\ &\geq -\inf_{y \in A} I(y) - \epsilon. \end{aligned}$$

Since ϵ is arbitrary, this proves the lower bound. \square

We finish this section with some basic notions of weak convergence of measures. Let \mathcal{M}^d denote the space of Borel measures on $[0, 1]^d$ bounded by some constant $D > 0$. The space \mathcal{M}^d is equipped with the topology of weak convergence which can be realized using the so-called Prokhorov metric denoted by d defined as follows (see [Bil99] for details). For any two measures $\mu, \nu \in \mathcal{M}^d$, their distance $d(\mu, \nu)$ is the smallest τ such that for every measurable $A \subset [0, 1]^d$ we have $\mu(A) \leq \nu(A^\tau) + \tau$ and $\nu(A) \leq \mu(A^\tau) + \tau$. Here A^τ is the set of points with distance at most τ from A in the $\|\cdot\|_\infty$ norm on $[0, 1]^d$ (although the choice of norm on $[0, 1]^d$ is not relevant). It is known that the metric space \mathcal{M}^d is compact with respect to the topology of weak convergence. We will focus in this paper on the product $\mathcal{M}^1 \times \mathcal{M}^2$ which we equip with the following metric (which we also denote by d with some abuse of notation): for every two pairs of measures $(\rho_j, \mu_j) \in \mathcal{M}^1 \times \mathcal{M}^2$, $j = 1, 2$, the distance is $d((\rho_1, \mu_1), (\rho_2, \mu_2)) = \max(d(\rho_1, \rho_2), d(\mu_1, \mu_2))$, where d is the metric on \mathcal{M}^1 and \mathcal{M}^2 .

3 LD-convergence of sparse graphs

3.1 Definition and basic properties

Given a graph $\mathbb{G} = (V, E)$ and any mapping $\sigma : V \rightarrow [0, 1]$, construct $(\rho(\sigma), \mu(\sigma)) \in \mathcal{M}^1 \times \mathcal{M}^2$ as follows. For every $u \in V$ we put mass $1/|V|$ on $\sigma(u) \in [0, 1]$. This defines $|V|$ points on $[0, 1]$ with the total mass 1. The resulting measure is denoted by $\rho(\sigma) \in \mathcal{M}^1$. Similarly, we put mass $1/|V|$ on $(\sigma(u), \sigma(v)) \in [0, 1]^2$ and $(\sigma(v), \sigma(u)) \in [0, 1]^2$ for every edge $(u, v) \in E$. The resulting measure is denoted by $\mu(\sigma) \in \mathcal{M}^2$. Formally

$$\begin{aligned} \rho(\sigma) &= \sum_{u \in V} |V|^{-1} \delta_{\sigma(u)}, \\ \mu(\sigma) &= \sum_{(u,v) \in E} |V|^{-1} \delta_{(\sigma(u), \sigma(v))}, \end{aligned} \tag{8}$$

where the sum goes over oriented edges, i.e., ordered pairs (u, v) such that u is connected to v by an edge in E , and δ_x denotes a unit mass measure on $x \in \mathbb{R}^d$. For example, consider a graph with 4 nodes 1, 2, 3, 4 and edges (1, 2), (1, 3), (2, 3), (3, 4). Suppose a realization of σ is $\sigma(1) = .4, \sigma(2) = .8, \sigma(3) = .15, \sigma(4) = .5$, as shown on Figure 1. The corresponding $\rho(\sigma)$ and $\mu(\sigma)$ are illustrated on Figure 2.

Now given a sparse graph sequence $\mathbb{G}_n = (V_n, E_n)$ with a uniform degree bound $\max_n \Delta_{\mathbb{G}_n} \leq D$, let $\sigma_n : V_n \rightarrow [0, 1]$ be chosen independently uniformly at random with respect to the nodes

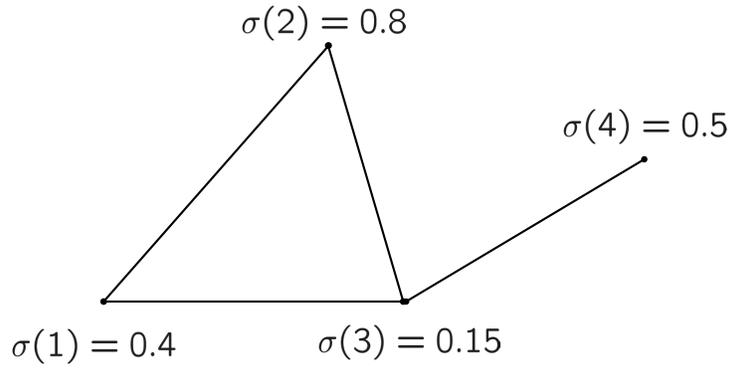


Figure 1: A four node graph decorated with real valued colors

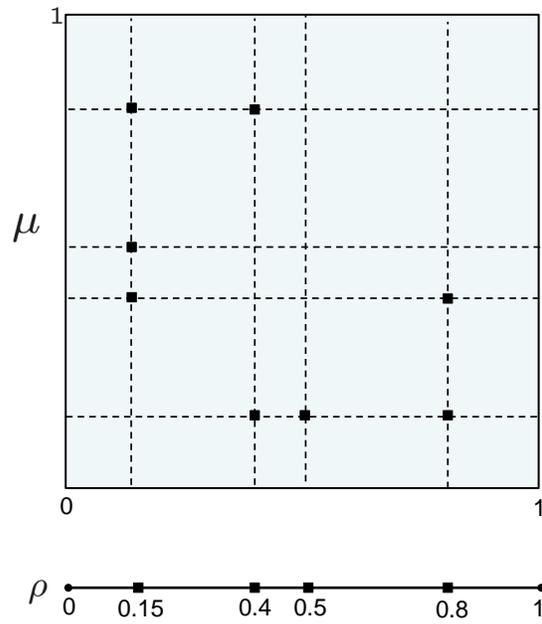


Figure 2: Measures ρ, μ associated with a four node graph with four edges

$u \in V_n$. Then we obtain a random element $(\rho(\sigma_n), \mu(\sigma_n)) \in \mathcal{M}^1 \times \mathcal{M}^2$. We will use notations ρ_n and μ_n instead of $\rho(\sigma_n)$ and $\mu(\sigma_n)$. Observe that $\rho_n([0, 1]) = 1$ and $0 \leq \mu_n([0, 1]^2) \leq D$. In particular, $\mu_n([0, 1]^2) = 0$ if and only if \mathbb{G}_n is an empty graph. We now introduce the notion of LD-convergence. In this definition and the Definition 2 below we assume $\theta_n = |V_n|$.

Definition 1. A graph sequence \mathbb{G}_n is defined to be LD-convergent if the sequence $(\rho(\sigma_n), \mu(\sigma_n))$ satisfies LDP in the metric space $\mathcal{M}^1 \times \mathcal{M}^2$.

Given a sparse graph sequence \mathbb{G}_n with degree bounded by D and a positive integer k , let $\sigma_n : V_n \rightarrow [k]$ be chosen uniformly at random. Consider the corresponding random weighted k -quotients $\mathbb{F}_n = \mathbb{G}_n/\sigma_n$ defined via (3) with random node weights $x(\sigma_n) = (x_i(\sigma_n), 1 \leq i \leq k)$ and random edge weights $X(\sigma_n) = (X_{ij}(\sigma_n), 1 \leq i, j \leq k)$. Note that $\sum_i x_i(\sigma_n) = 1$ and $\sum_{i,j} x_{ij}(\sigma_n) \leq D$. Assuming without loss of generality that $D \geq 1$, it will be convenient to view $(x(\sigma_n), X(\sigma_n))$ as a $(k+1) \times k$ matrix with elements bounded by D , namely an element in $[0, D]^{(k+1) \times k}$. We consider $[0, D]^{(k+1) \times k}$ equipped with the \mathbb{L}_∞ norm. That is the distance between (x, X) and (y, Y) is the maximum of $\max_{1 \leq i \leq k} |x_i - y_i|$ and $\max_{1 \leq i, j \leq k} |X_{i,j} - Y_{i,j}|$.

Definition 2. A graph sequence \mathbb{G}_n is defined to be k -LD-convergent if for every k the sequence $(x(\sigma_n), X(\sigma_n))$ satisfies the LDP in $[0, D]^{(k+1) \times k}$ with some rate function $I_k : [0, D]^{(k+1) \times k} \rightarrow \mathbb{R}_+ \cup \{\infty\}$.

The intuition behind this definition is as follows. Given (x, X) suppose there exists a k -coloring of a graph G_n such that the corresponding k -quotient weighted graph is "approximately" equal to (x, X) to within some additive error ϵ . Then in fact there are exponentially in $|V_n|$ many k -colorings of \mathbb{G}_n which achieve "nearly" the same quotient graph up to say an 2ϵ additive error (by arbitrarily recoloring $\epsilon|V_n|/D$ many nodes). In this case k -LD-convergence of the graph sequence \mathbb{G}_n means that the exponent of the number of colorings achieving the quotient (x, X) is well defined and is given by $\log k - I(x, X)$. In other words, the total number of k -coloring achieving the quotient (x, X) is approximately $\exp(|V_n|(\log k - I(x, X)))$. In case (x, X) is not "nearly" achievable by any coloring, we have $I(x, X) = \infty$, meaning that there are simply no coloring achieving the target quotient (x, X) .

While Definition 2 might be more intuitive than Definition 1, the latter is more powerful in the sense that it does not require a "for all k " condition, namely it is introduced through having the LDP on *one* as opposed to infinite (for every k) sequence of probability spaces. It is also more powerful, in the sense that the measure (ρ_n, σ_n) completely defines the underlying graph, whereas this is not the case for the k -quotients. Since the space $[0, D]^{(k+1) \times k}$ being a subset of a Euclidian space is a much "smaller" than the space of measures \mathcal{M}^d , it is not too surprising that the LD-convergence implies the k -LD-convergence. But it is perhaps surprising that the converse is true.

Theorem 1. A graph sequence \mathbb{G}_n is LD-convergent if and only if it is k -LD-convergent for every k .

Remark 1. Given the rate function I for LD-convergence, the rate function $I_k : [0, D]^{(k+1) \times k} \rightarrow \mathbb{R}_+ \cup \{\infty\}$ for k -LD-convergence is easy to calculate, and will be given by

$$I_k(x, X) = \inf_{(\rho, \mu)} I(\mu, \rho)$$

where the inf goes over all $(\rho, \mu) \in \mathcal{M}^1 \times \mathcal{M}^2$ such that

$$\begin{aligned} x_{i+1} &= \rho(\Delta_{i,k}), & 0 \leq i \leq k-1; \\ X_{i+1,j+1} &= \mu(\Delta_{i,k} \times \Delta_{j,k}), & 0 \leq i, j \leq k-1, \end{aligned} \tag{9}$$

with $\Delta_{i,k}$ denoting the open interval $\Delta_{i,k} = \left(\frac{i}{k}, \frac{(i+1)}{k}\right)$.

We will also give an explicit construction for the rate function I from the rate functions I_k , see Corollary 1 below.

Before proving the theorem, we reformulate the notion of k -LD-convergence in terms of random measures with piecewise constant densities. We need some notation. Given a positive integer k , let $O_k^d \subset [0, 1]^d$ be the open set of all points such that no coordinate is of the form $0, 1/k, 2/k, \dots, 1$. Let $\mathcal{M}_k^d \subset \mathcal{M}^d$ be the set of measures supported on O_k^d , and let $\mathcal{N}_k^d \subset \mathcal{M}_k^d$ be the set of measures in \mathcal{M}_k^d which have constant density on rectangles

$$\Delta_{i_1,k} \times \dots \times \Delta_{i_d,k}, \quad 0 \leq i_1, \dots, i_d \leq k-1.$$

We define a projection operator $T_k : \mathcal{M}^1 \times \mathcal{M}^2 \rightarrow \mathcal{N}_k^1 \times \mathcal{N}_k^2$ by mapping a pair of measures $(\rho, \mu) \in \mathcal{M}^1 \times \mathcal{M}^2$ into $(\rho_k, \mu_k) = T_k(\rho, \mu)$ defined as the pair of measures in $\mathcal{N}_k^1 \times \mathcal{N}_k^2$ for which

$$\begin{aligned} \rho_k(\Delta_{i,k}) &= \rho(\Delta_{i,k}) \quad \text{for } 0 \leq i \leq k-1 \\ \mu_k(\Delta_{i,k} \times \Delta_{j,k}) &= \mu(\Delta_{i,k} \times \Delta_{j,k}) \quad \text{for } 0 \leq i, j \leq k-1. \end{aligned} \tag{10}$$

Given a real valued coloring $\sigma : V_n \rightarrow [0, 1]$ let $\sigma_k = t_k(\sigma)$ be the coloring $\sigma_k : V_n \rightarrow [k]$ defined by $\sigma_k(u) = \lceil k\sigma(u) \rceil$ if $\sigma(u) > 0$ and $\sigma_k(u) = 1$ if $\sigma(u) = 0$. Note that $t_k(\sigma_n)$ is an i.i.d. uniformly random coloring with colors in $[k]$ if σ_n is chosen to be an i.i.d. uniformly random coloring with colors in $[0, 1]$.

Lemma 1. *Let $\mathbb{G}_n = (V_n, E_n)$, and let $\sigma_n : V_n \rightarrow [0, 1]$ be an i.i.d. uniformly random coloring of V_n . Then \mathbb{G}_n is k -LD-convergent if and only if $T_k(\rho(\sigma_n), \mu(\sigma_n))$ obeys a LDP on $\mathcal{N}_k^1 \times \mathcal{N}_k^2$.*

Proof. Let $\hat{\sigma}_n : V_n \rightarrow [k]$ be an i.i.d uniformly random coloring of V_n with colors in $[k]$, and define $F_k : [0, D]^{(k+1) \times k} \rightarrow \mathcal{N}_k^1 \times \mathcal{N}_k^2$ as follows: given $(x, X) \in [0, D]^{(k+1) \times k}$ with $x = (x_i, 1 \leq i \leq k)$ and $X = (X_{ij}, 1 \leq i, j \leq k)$, let $F_k((x, X)) := (\rho, \mu) \in \mathcal{N}_k^1 \times \mathcal{N}_k^2$ be the piece-wise constant measures with densities kx_i on $\Delta_{i-1,k}$ and $k^2 X_{i,j}$ on $\Delta_{i-1,k} \times \Delta_{j-1,k}$, respectively. Then F_k is an invertible continuous function with continuous inverse. Thus, applying the Contraction Principle, $(\rho_k(\hat{\sigma}_n), \mu_k(\hat{\sigma}_n)) = F_k((x(\hat{\sigma}_n), X(\hat{\sigma}_n)))$ satisfies the LDP on $\mathcal{N}_k^1 \times \mathcal{N}_k^2$ if and only if $(x(\hat{\sigma}_n), X(\hat{\sigma}_n))$ satisfies the LDP on $[0, D]^{(k+1) \times k}$.

To conclude the proof, we only need to show that $T_k(\rho(\sigma_n), \mu(\sigma_n))$ has the same distribution as $(\rho_k(\hat{\sigma}_n), \mu_k(\hat{\sigma}_n))$. But this follows immediately from the facts that $t_k(\sigma_n)$ has the same distribution as $\hat{\sigma}_n$ and that $T_k(\rho(\sigma_n), \mu(\sigma_n))$ depends only on the total mass in the intervals $\Delta_{k,i}$, so that $T_k(\rho(\sigma_n), \mu(\sigma_n)) = (\rho_k(t_k(\sigma_n)), \mu_k(t_k(\sigma_n)))$. \square

3.1.1 Proof of Theorem 1, Part 1: LD-convergence implies k -LD-convergence

In view of the last lemma, it seems quite intuitive that LD-convergence should imply k -LD-convergence, given that the set of measures $\mathcal{N}_k^1 \times \mathcal{N}_k^2$ is much smaller than the set of measures

$\mathcal{M}^1 \times \mathcal{M}^2$. More formally, one might hope to define a continuous map from $\mathcal{M}^1 \times \mathcal{M}^2$ to $\mathcal{N}_k^1 \times \mathcal{N}_k^2$ (or to $[0, D]^{(k+1) \times k}$) and use the Contraction Principle to prove the desired k -LD-convergence.

It turns out that we can't quite do that due to point masses on points of the form i/k , $1 \leq i \leq k$. To address this issue, we will first prove that we can restrict ourselves to $\mathcal{M}_k^1 \times \mathcal{M}_k^2$, and then define a suitable continuous function from $\mathcal{M}_k^1 \times \mathcal{M}_k^2$ to $[0, D]^{(k+1) \times k}$ to apply the Contraction Principle. We need the following lemma.

Lemma 2. *Suppose \mathbb{G}_n is LD-convergent with rate function I . Suppose ρ is such that $\rho(\{x\}) > 0$ for some $x \in [0, 1]$. Then for every $\mu \in \mathcal{M}^2$, $I(\rho, \mu) = \infty$. Similarly, suppose μ is such that $\mu(\{x\} \times [0, 1]) > 0$ or $\mu([0, 1] \times \{x\}) > 0$ for some $x \in [0, 1]$. Then for every $\rho \in \mathcal{M}^1$, $I(\rho, \mu) = \infty$.*

Proof. Suppose ρ is such that $\rho(\{x\}) = \alpha > 0$ for some $x \in [0, 1]$. Fix $\delta < \alpha/2$. Suppose ρ' is such that $d(\rho, \rho') \leq \delta$. Then, by definition of the Prokhorov metric,

$$\alpha = \rho(\{x\}) \leq \rho'(x - \delta, x + \delta) + \delta,$$

implying $\rho'(x - \delta, x + \delta) \geq \alpha/2$. Thus the event $(\rho(\sigma_n), \mu(\sigma_n)) \in B((\rho, \mu), \delta)$ implies the event $\rho(\sigma_n)(x - \delta, x + \delta) \geq \alpha/2$. We now estimate the probability of this event. For this event to occur we need to have at least $\alpha/2$ fraction of values $\sigma_n(u)$, $u \in V_n$ to fall into the interval $(x - \delta, x + \delta)$. This occurs with probability at most

$$\sum_{i \geq (\alpha/2)|V_n|} \binom{|V_n|}{i} (2\delta)^i \leq 2^{|V_n|} (2\delta)^{(\alpha/2)|V_n|}.$$

Therefore

$$\limsup_n |V_n|^{-1} \mathbb{P}(\rho(\sigma_n)(x - \delta, x + \delta) \geq \alpha/2) \leq \log 2 + (\alpha/2) \log(2\delta),$$

implying

$$\lim_{\delta \rightarrow 0} \limsup_n |V_n|^{-1} \mathbb{P}((\rho(\sigma_n), \mu(\sigma_n)) \in B((\rho, \mu), \delta)) = -\infty,$$

and thus $I(\rho, \mu) = \infty$ utilizing property (6). The proof for the case $\mu(\{x\} \times [0, 1]) > 0$ or $\mu([0, 1] \times \{x\}) > 0$ is similar. \square

We now establish that LD-convergence implies k -LD-convergence. To this end, we first observe that for an i.i.d. uniformly random coloring $\sigma_n : V_n \rightarrow [0, 1]$, we have that $\rho(\sigma_n) \in \mathcal{M}_k^1$ and $\mu(\sigma_n) \in \mathcal{M}_k^2$ almost surely (as the probability of hitting one of the points i/k is zero).

Next we claim that since $(\rho(\sigma_n), \mu(\sigma_n))$ satisfies the LDP in $\mathcal{M}^1 \times \mathcal{M}^2$, it also does so in $\mathcal{M}_k^1 \times \mathcal{M}_k^2$ with the same rate function. For this we will show that (5) holds when closures and interiors are taken with respect to the metric space $\mathcal{M}_k^1 \times \mathcal{M}_k^2$ as opposed to $\mathcal{M}^1 \times \mathcal{M}^2$. Indeed, fix any set $A \subset \mathcal{M}_k^1 \times \mathcal{M}_k^2$. Its closure in $\mathcal{M}_k^1 \times \mathcal{M}_k^2$ is a subset of its closure in $\mathcal{M}^1 \times \mathcal{M}^2$ and the set-theoretic difference between the two consists of measures (ρ, μ) such that ρ assigns a positive mass to some point i/k , or μ assigns a positive mass to some segment $\{i/k\} \times [0, 1]$ or segment $[0, 1] \times \{j/k\}$. By Lemma 2 the large deviations rate I at these points (ρ, μ) is infinite, so that the value $\inf I(\rho, \mu)$ over the closure of A in $\mathcal{M}_k^1 \times \mathcal{M}_k^2$ and $\mathcal{M}^1 \times \mathcal{M}^2$ is the same. Similarly, the interior of every set $A \subset \mathcal{M}_k^1 \times \mathcal{M}_k^2$ is a superset of its interior in $\mathcal{M}^1 \times \mathcal{M}^2$, and again the set

theoretic difference between the two interiors consists of points with infinite rate I . This proves the claim.

Consider $\phi_k : \mathcal{M}_k^1 \times \mathcal{M}_k^2 \rightarrow [0, D]^{(k+1) \times k}$, where ϕ_k maps every pair of measures ρ, μ into total measure assigned to interval $\Delta_{i,k}$ by ρ and assigned to rectangle $\Delta_{i,k} \times \Delta_{j,k}$ by μ , for $i, j = 0, 1, \dots, k-1$. Namely $\phi_k(\rho, \mu) = (x, X)$ where (x, X) is given by (9).

We claim that ϕ_k is continuous with respect to the respective metrics. Indeed, suppose $(\rho_n, \mu_n) \rightarrow (\rho, \mu)$ in $\mathcal{M}_k^1 \times \mathcal{M}_k^2$. Since \mathcal{M}^d is equipped with the weak topology, then for every closed (open) set $A \subset [0, 1]$, we have $\limsup_n \rho_n(A) \leq \rho(A)$ ($\liminf_n \rho_n(A) \geq \rho(A)$). The same applies to μ_n, μ . Setting A to be $[i/k, (i+1)/k]$ first and then $(i/k, (i+1)/k)$, and using the fact that ρ_n and ρ are supported on O_k^1 , namely $\rho_n(i/k) = \rho(i/k) = 0$ for all i , we obtain $\rho_n(i/k, (i+1)/k) \rightarrow \rho(i/k, (i+1)/k)$. A similar argument applies to μ_n, μ . This proves the claim.

Since ϕ_k is continuous, by the Contraction Principle [DZ98], the image of $(\rho(\sigma_n), \mu(\sigma_n))$ under ϕ_k satisfies the LDP as well. But this means precisely that we have k -LD-convergence with rate function I_k as defined in Remark 1 (strictly speaking, the Contraction Principle gives a rate function where the inf is taken over all $(\rho, \mu) \in \mathcal{M}_k^1 \times \mathcal{M}_k^2$, but by Lemma 2 we can extend the inf to that larger set $\mathcal{M}^1 \times \mathcal{M}^2$ without changing the value of I_k).

3.1.2 Proof of Theorem 1, Part 2: k -LD-convergence implies LD-convergence

In order to prove that k -LD-convergence for all k implies LD-convergence, we would like to use the rate functions $I_k : \mathcal{N}_k^1 \times \mathcal{N}_k^2 \rightarrow \mathbb{R}_+ \cup \{\infty\}$ associated with the LDP for $T_k(\rho(\sigma_n), \mu(\sigma_n))$ from Lemma 1 to construct a suitable rate function $I : \mathcal{M}^1 \times \mathcal{M}^2 \rightarrow \mathbb{R}_+ \cup \{\infty\}$ for the random variables $(\rho(\sigma_n), \mu(\sigma_n))$. The existence of such a rate function will follow from a suitable monotonicity property, see Lemma 4 and Corollary 1 below.

Before stating Lemma 4, we state and prove a more technical lemma which we will use at several places in this subsection. For two pairs of measures $(\rho, \mu), (\rho', \mu') \in \mathcal{N}_k^1 \times \mathcal{N}_k^2$ we define

$$d_{var}((\rho, \mu), (\rho', \mu')) = \max \left\{ \max_{A \subset [0,1]} |\rho(A) - \rho'(A)|, \max_{B \subset [0,1]^2} |\mu(B) - \mu'(B)| \right\}.$$

Lemma 3.

1) If $(\rho, \mu) \in \mathcal{M}_k^1 \times \mathcal{M}_k^2$, then

$$d((\rho, \mu), T_k(\rho, \mu)) \leq \frac{1}{k}. \quad (11)$$

2) If $(\rho, \mu), (\rho', \mu') \in \mathcal{N}_k^1 \times \mathcal{N}_k^2$, then

$$d((\rho, \mu), (\rho', \mu')) \leq d_{var}((\rho, \mu), (\rho', \mu')) \leq (4kD + 1)d((\rho, \mu), (\rho', \mu')). \quad (12)$$

3) If $(\rho, \mu), (\rho', \mu') \in \mathcal{N}_{2k}^1 \times \mathcal{N}_{2k}^2$, then

$$d_{var}(T_k(\rho, \mu), T_k(\rho', \mu')) \leq d_{var}((\rho, \mu), (\rho', \mu')). \quad (13)$$

Proof. 1) Set $(\rho_k, \mu_k) = T_k(\rho, \mu)$, and let $A \subset [0, 1]$. Define A_k to be the set that is obtained by replacing every non-empty set $A \cap \Delta_{i,k}$ with the entire interval $\Delta_{i,k}$. Then $A \subset A_k \subset A^{\frac{1}{k}}$. We have $\rho(A \cap \Delta_{i,k}) \leq \rho(\Delta_{i,k}) = \rho_k(\Delta_{i,k})$ by the definition of ρ_k . Let $\alpha = \{i : A \cap \Delta_{i,k} \neq \emptyset\}$. Then

$$\rho(A) = \sum_{i \in \alpha} \rho(A \cap \Delta_{i,k}) \leq \sum_{i \in \alpha} \rho_k(\Delta_{i,k}) = \rho_k(A_k) \leq \rho_k(A^{\frac{1}{k}}).$$

Conversely,

$$\rho_k(A) \leq \rho_k(A_k) = \rho(A_k) \leq \rho(A^{\frac{1}{k}}).$$

A similar argument is used for μ .

2) The lower bound follows immediately from the definitions of d and d_{var} . To prove the upper bound, we first show that

$$\max_{A \subset [0,1]} |\rho(A) - \rho'(A)| \leq (2kD + 1)d(\rho, \rho').$$

To this end, we note the maximum is obtained when A is of the form $A = \bigcup_{i \in \alpha} \Delta_{i,k}$ for some $\alpha \subset \{0, \dots, k-1\}$. On the other hand, for sets A of this form, we have that $\rho(A^\epsilon) \leq \rho(A) + 2k\epsilon\rho(A^c) \leq \rho(A) + 2k\epsilon D$, a bound which follows from the fact that $A^c = [0, 1] \setminus A$ is a union of intervals of length at least $1/k$. Choosing $\epsilon = d(\rho, \rho')$ we then have $\rho(A') \leq \rho(A^\epsilon) + \epsilon \leq \rho(A) + (2kD + 1)\epsilon$, which implies $\rho'(A) - \rho(A) \leq (2kD + 1)\epsilon = (2kD + 1)d(\rho, \rho')$. Exchanging the roles of ρ and ρ' gives $\rho(A) - \rho'(A) \leq (2kD + 1)d(\rho, \rho')$, which completes the claim. In a similar way, one proves that

$$\max_{B \subset [0,1]^2} |\mu(B) - \mu'(B)| \leq (4kD + 1)d(\mu, \mu').$$

The proof uses that for sets B which are unions of sets of the form $\Delta_{i,k} \times \Delta_{j,k}$ we have $\mu(B^\epsilon) \leq \mu(B) + 4k\epsilon D$.

3) Let $(\rho_k, \mu_k) = T_k(\rho, \mu)$ and $(\rho'_k, \mu'_k) = T_k(\rho', \mu')$. We need to show that $d_{var}(\rho_k, \rho'_k) \leq d_{var}(\rho, \rho')$ and $d_{var}(\mu_k, \mu'_k) \leq d_{var}(\mu, \mu')$. The first bound follows from the definition of T_k and the fact that the maximum in the definition $d_{var}(\rho_k, \rho'_k) = \max_{A \subset [0,1]} |\rho_k(A) - \rho'_k(A)|$ is obtained when A is of the form $A = \bigcup_{i \in \alpha} \Delta_{i,k}$ for some $\alpha \subset \{0, \dots, k-1\}$. Indeed, for such an A , one easily shows that $|\rho_k(A) - \rho'_k(A)| = |\rho(A) - \rho'(A)|$, which in turn implies that $d_{var}(\rho_k, \rho'_k) \leq d_{var}(\rho, \rho')$. The proof of the bound $d_{var}(\mu_k, \mu'_k) \leq d_{var}(\mu, \mu')$ is similar. \square

Lemma 4. *Suppose \mathbb{G}_n is k -LD-convergent, and let $I_k : \mathcal{N}_k^1 \times \mathcal{N}_k^2 \rightarrow \mathbb{R}_+ \cup \{\infty\}$ be the rate function associated with the LDP for $T_k(\rho(\sigma_n), \mu(\sigma_n))$. Then $I_k(T_k(\rho, \mu)) \leq I_{2k}(\rho, \mu)$ for every $(\rho, \mu) \in \mathcal{N}_{2k}^1 \times \mathcal{N}_{2k}^2$.*

Proof. Set $(\rho_k, \mu_k) = T_k(\rho, \mu)$, fix $\epsilon > 0$, and set $\epsilon' = (4kD + 1)\epsilon$. By the bounds (12) and (13), we have that

$$(\rho'_k, \mu'_k) = T_k(\rho', \mu') \in B(T_k(\rho, \mu), \epsilon') \tag{14}$$

whenever $(\rho', \mu') \in B((\rho, \mu), \epsilon)$. On the other hand, $T_k(\rho(\sigma_n), \mu(\sigma_n)) = T_k(T_{2k}(\rho(\sigma_n), \mu(\sigma_n)))$. By (14), this implies that

$$\mathbb{P}\left(T_k(\rho(\sigma_n), \mu(\sigma_n)) \in B\left(T_k(\rho, \mu), (4kD + 1)\epsilon\right)\right) \geq \mathbb{P}\left(T_{2k}(\rho(\sigma_n), \mu(\sigma_n)) \in B\left((\rho, \mu), \epsilon\right)\right).$$

Applying (6) we obtain the claim of the lemma. \square

Corollary 1. *Suppose \mathbb{G}_n is k -LD-convergent for all k , let $I_k : \mathcal{N}_k^1 \times \mathcal{N}_k^2 \rightarrow \mathbb{R}_+ \cup \{\infty\}$ be the rate function associated with the LDP for $T_k(\rho(\sigma_n), \mu(\sigma_n))$, let $(\rho, \mu) \in \mathcal{M}^1 \times \mathcal{M}^2$, and let B_k be the ball $B((\rho, \mu), 2k^{-1}) \cap \mathcal{N}_k^1 \times \mathcal{N}_k^2$. Then the limit*

$$I(\rho, \mu) = \lim_{\ell \rightarrow \infty} \inf_{(\rho', \mu') \in B_{2^\ell}} I_{2^\ell}(\rho', \mu')$$

exists.

Proof. Let $k \in \mathbb{Z}_+$. By Lemma 3, we have that $T_k(\rho', \mu') \in B_k$ whenever $(\rho', \mu') \in B_{2k}$. Combined with Lemma 4, this proves that

$$\inf_{(\rho', \mu') \in B_{2k}} I_{2k}(\rho', \mu') \geq \inf_{(\rho', \mu') \in B_{2k}} I_k(T_k(\rho', \mu')) \geq \inf_{(\rho'', \mu'') \in B_k} I_k(T_k(\rho'', \mu'')).$$

The claim now follows by monotonicity. \square

We are now ready to prove that k -LD-convergence for all k implies LD-convergence. Let $\sigma_n : V_n \rightarrow [0, 1]$ be an i.i.d. uniformly random coloring of V_n . Recalling Lemma 1, we will assume that $T_k(\rho(\sigma_n), \mu(\sigma_n))$ obeys a LDP with some rate function I_k . We will want to prove that $(\rho(\sigma_n), \mu(\sigma_n))$ obeys a LDP with the rate function I defined in Corollary 1.

Fix $(\rho, \mu) \in \mathcal{M}^1 \times \mathcal{M}^2$. By the Lemma 3, the fact that $(\rho(\sigma_n), \mu(\sigma_n)) \in \mathcal{M}_k^1 \times \mathcal{M}_k^2$ almost surely, and the triangle inequality, we have that

$$\begin{aligned} \liminf_n |V_n|^{-1} \log \mathbb{P}((\rho(\sigma_n), \mu(\sigma_n)) \in B((\rho, \mu), 3k^{-1})) \\ \geq \liminf_n |V_n|^{-1} \log \mathbb{P}(T_k(\rho(\sigma_n), \mu(\sigma_n)) \in B((\rho, \mu), 2k^{-1})) \\ \geq -\inf I_k(\rho', \mu'), \end{aligned}$$

where the infimum is taken over $(\rho', \mu') \in B((\rho, \mu), 2k^{-1}) \cap \mathcal{N}_k^1 \times \mathcal{N}_k^2$. Similarly,

$$\begin{aligned} \limsup_n |V_n|^{-1} \log \mathbb{P}((\rho(\sigma_n), \mu(\sigma_n)) \in B((\rho, \mu), k^{-1})) \\ \leq \limsup_n |V_n|^{-1} \log \mathbb{P}(T_k(\rho(\sigma_n), \mu(\sigma_n)) \in B((\rho, \mu), 2k^{-1})) \\ \leq -\inf I_k(\rho', \mu'), \end{aligned}$$

where the infimum is again taken over $(\rho', \mu') \in B((\rho, \mu), 2k^{-1}) \cap \mathcal{N}_k^1 \times \mathcal{N}_k^2$.

To complete the proof we apply Proposition 1 in conjunction with Corollary 1, yielding that $(\rho(\sigma_n), \mu(\sigma_n))$ obeys a LDP with rate function I .

3.2 Basic properties and examples

We begin by showing that the definition of LD-convergence is robust with respect to the equivalency relationship \sim on graphs.

Theorem 2. *If \mathbb{G}_n is LD-convergent and $\tilde{\mathbb{G}}_n \sim \mathbb{G}_n$, then $\tilde{\mathbb{G}}_n$ is also LD-convergent.*

Proof. We apply Theorem 1 and show that if \mathbb{G}_n is k -LD-convergent then $\tilde{\mathbb{G}}_n$ is also k -LD-convergent. Fix any k , any $(x, X) \in [0, D]^{(k+1) \times k}$ and $\epsilon > 0$. We denote by $\tilde{\mathbb{P}}$ the probability measure associated with $(\tilde{x}(\sigma_n), \tilde{X}(\sigma_n)) = \tilde{\mathbb{G}}_n / \sigma_n$ when $\sigma_n : V_n \rightarrow [k]$ is a random uniformly chosen map on V_n . The definition of \sim implies that for all sufficiently large n

$$\tilde{\mathbb{P}}\left((\tilde{x}(\sigma_n), \tilde{X}(\sigma_n)) \in B((x, X), \epsilon/2)\right) \leq \mathbb{P}\left((x(\sigma_n), X(\sigma_n)) \in B((x, X), \epsilon)\right),$$

implying

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \limsup_n |V_n|^{-1} \log \tilde{\mathbb{P}}\left((\tilde{x}(\sigma_n), \tilde{X}(\sigma_n)) \in B((x, X), \epsilon/2)\right) \\ \leq \lim_{\epsilon \rightarrow 0} \limsup_n |V_n|^{-1} \log \mathbb{P}\left((x(\sigma_n), X(\sigma_n)) \in B((x, X), \epsilon)\right). \end{aligned}$$

Similarly, we establish that

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \liminf_n |V_n|^{-1} \log \tilde{\mathbb{P}} \left((\tilde{x}(\sigma_n), \tilde{X}(\sigma_n)) \in B((x, X), \epsilon) \right) \\ \geq \lim_{\epsilon \rightarrow 0} \liminf_n |V_n|^{-1} \log \mathbb{P} \left((x(\sigma_n), X(\sigma_n)) \in B((x, X), \epsilon/2) \right). \end{aligned}$$

Since \mathbb{G}_n is k -LD-convergent, then from (6) we have

$$\lim_{\epsilon \rightarrow 0} \limsup_n |V_n|^{-1} \log \tilde{\mathbb{P}}(\cdot) = \lim_{\epsilon \rightarrow 0} \liminf_n |V_n|^{-1} \log \tilde{\mathbb{P}}(\cdot).$$

Thus the same identity applies to $\tilde{\mathbb{P}}$. Applying Proposition 1 we conclude that $\tilde{\mathbb{P}}$ is LD-convergent. \square

Let us give some examples of LD-convergent graph sequences. As our first example, consider a sequence of graphs which consists of a disjoint union of copies of a fixed graph.

Example 1. Let \mathbb{G}_0 be a fixed graph and let \mathbb{G}_n be a disjoint union of n copies of \mathbb{G}_0 . Then \mathbb{G}_n is LD-convergent.

Proof. We use Theorem 1 and prove that \mathbb{G}_n is k -LD-convergent for every k . Fix k , let Σ_k be the space of all possible k -colorings of \mathbb{G}_0 , and let $M = k^{|V(\mathbb{G}_0)|}$ denote the size of Σ_k . Every k -coloring $\sigma_n : V_n \rightarrow [k]$ of the nodes of \mathbb{G}_n can then be encoded as $\sigma_n = (\sigma_i^n)_{1 \leq i \leq n}$, where $\sigma_i^n \in \Sigma_k$ is a k -coloring of the i -th copy of \mathbb{G}_0 in \mathbb{G}_n . Consider the M -dimensional simplex S_M , i.e., the set of vectors (z_1, \dots, z_M) such that $\sum_{i \leq M} z_i = 1, z_i \geq 0$. For every $m = 1, \dots, M$ and $\sigma_n = (\sigma_i^n)_{1 \leq i \leq n}$, let $z_m(\sigma_n)$ be the number of σ_i^n -s which are equal to the m -element of Σ_k divided by n , and let $z(\sigma_n) = (z_m(\sigma_n))_{1 \leq m \leq M} \in S_M$. By Sanoff's Theorem, the sequence $z(\sigma_n), n \geq 1$ satisfies the LDP with respect to the metric space S_M . Now consider a natural mapping from S_M into $[0, D]^{(k+1) \times k}$, where each $z = (z_m)_{1 \leq m \leq M}$ is mapped into $(x, X) \in [0, D]^{(k+1) \times k}$ as follows. For each element of Σ_k encoded by some $m \leq M$ and each $i \leq k$, let $x(i, m)$ be the number of nodes of \mathbb{G}_0 colored i according to m , divided by $|V(\mathbb{G}_0)|$. In particular $\sum_i x(i, m) = 1$. Similarly, let $X(i, j, m)$ be 2 times the number of edges in \mathbb{G}_0 with end node colors i and j according to m , divided by $|V(\mathbb{G}_0)|$. In particular $\sum_{i,j} X(i, j, m) = 2|E(\mathbb{G}_0)|/|V(\mathbb{G}_0)|$. Setting $x_i = \sum_m x(i, m)z_m, X_{i,j} = \sum_m X(i, j, m)z_m$, this defines the mapping S_M into $[0, D]^{(k+1) \times k}$. This mapping is continuous. Observe that the composition $\sigma_n \rightarrow z(\sigma_n) \rightarrow (x, X)$ is precisely the construction of the factor graph $\mathbb{F} = \mathbb{G}_n/\sigma_n$ via (3). Applying the Contraction Principle, we obtain LD-convergence of the sequence \mathbb{G}_n . \square

Perhaps a more interesting example is the case of subgraphs of the d -dimensional lattice \mathbb{Z}^d , more precisely the case of graph sequences $L_{d,n}$ with vertex sets $V_{d,n} = \{-n, -n+1, \dots, n\}^d$. It is known that the free energy of statistical mechanics systems on these graphs has a limit [Sim93], [Geo88]. In our terminology this means that the sequence $(L_{d,n})$ is right-convergent (see Section 4). Here we show that this sequence is also LD-convergent.

Example 2. Let $d \geq 1$, let $V_{d,n} = \{-n, -n+1, \dots, n\}^d$, let $E_{n,d}$ be the set of pairs $\{x, y\} \subset V_{d,n}$ of ℓ_1 distance 1, and let $L_{d,n} = (V_{d,n}, E_{d,n})$. Then the sequence $(L_{d,n})$ is LD-convergent.

Proof. We again apply Theorem 1 and establish k -LD-convergence for every k . Fix $(x, X) \in [0, D]^{(k+1) \times k}$ and let

$$\bar{I}_k(x) = - \lim_{\delta \rightarrow 0} \limsup_n \frac{\log \mathbb{P} \left((x(\sigma_n), X(\sigma_n)) \in B((x, X), \delta) \right)}{|V_{d,n}|},$$

where $\sigma_n : V_{d,n} \rightarrow [k]$ is chosen uniformly at random. The limit $\lim_{\delta \rightarrow 0}$ exists by monotonicity. Fix an arbitrary $\delta > 0$. Then again by monotonicity

$$\limsup_n \frac{\log \mathbb{P} \left((x(\sigma_n), X(\sigma_n)) \in B((x, X), \delta/2) \right)}{|V_{d,n}|} \geq -\bar{I}_k(x, X),$$

and we can find n_0 so that

$$\frac{\log \mathbb{P} \left((x(\sigma_{n_0}), X(\sigma_{n_0})) \in B((x, X), \delta/2) \right)}{|V_{d,n_0}|} \geq -\bar{I}_k(x, X) - \delta. \quad (15)$$

Consider arbitrary $n \geq n_0$, and set $q = \lfloor \frac{2n+1}{2n_0+1} \rfloor$, $M = q^d$ and $m = (2n+1)^d - M(2n_0+1)^d$. The set of vertices $V_{d,n}$ can then be written as a disjoint union of M copies of V_{d,n_0} and m sets consisting of one node each. Let $\mathbb{H}_1, \dots, \mathbb{H}_M$ be the induced subgraphs on the copies of V_{d,n_0} , and let $\tilde{L}_{d,n} = (\tilde{V}_{d,n}, \tilde{E}_{d,n})$ be the union of $\mathbb{H}_1, \dots, \mathbb{H}_M$. Then

$$|\tilde{V}_{d,n}| = |V_{d,n}| - m = |V_{d,n}|(1 - O(n_0/n))$$

and

$$|\tilde{E}_{d,n}| = |E_{d,n}| - O(Mn_0^{d-1}) - O(m) = |E_{d,n}| - O(n^d n_0^{-1}) - O(n^{d-1} n_0).$$

A uniformly random coloring $\sigma_n : V_{d,n} \rightarrow [k]$ then induces a uniformly random coloring $\tilde{\sigma}_n : \tilde{V}_{d,n} \rightarrow [k]$. Due to the above bounds, the corresponding quotients $L_{d,n}/\sigma_n = (x(\sigma_n), X(\sigma_n))$ and $\tilde{L}_{d,n}/\tilde{\sigma}_n = (\tilde{x}(\tilde{\sigma}_n), \tilde{X}(\tilde{\sigma}_n))$ are close to each other. More precisely,

$$d \left((x(\sigma_n), X(\sigma_n)), (\tilde{x}(\tilde{\sigma}_n), \tilde{X}(\tilde{\sigma}_n)) \right) = O(n_0/n) + O(1/n_0).$$

Then we can find n_0 large enough so that (15) still holds, such that for all large enough n ,

$$d \left((x(\sigma_n), X(\sigma_n)), (\tilde{x}(\tilde{\sigma}_n), \tilde{X}(\tilde{\sigma}_n)) \right) \leq \delta/2.$$

Then

$$\mathbb{P} \left((x(\sigma_n), X(\sigma_n)) \in B((x, X), \delta) \right) \geq \mathbb{P} \left((\tilde{x}(\tilde{\sigma}_n), \tilde{X}(\tilde{\sigma}_n)) \in B((x, X), \delta/2) \right).$$

On the other hand, since $\tilde{X}_{i,j}(\tilde{\sigma}_n)$ are counted over a disjoint union of graphs identical to L_{d,n_0} , then if $\tilde{\sigma}_n$ is such that within each graph \mathbb{H}_r we have $d((x, X), (\tilde{x}(\tilde{\sigma}_n), \tilde{X}(\tilde{\sigma}_n))) \leq \delta/2$, then the same applies to the overall graph $\tilde{L}_{d,n}$. Namely,

$$\mathbb{P} \left((\tilde{x}(\tilde{\sigma}_n), \tilde{X}(\tilde{\sigma}_n)) \in B(x, \delta/2) \right) \geq \left(\mathbb{P} \left((x(\sigma_{n_0}), X(\sigma_{n_0})) \in B(x, X, \delta/2) \right) \right)^M,$$

where the second expectation is with respect to uniformly random coloring of L_{d,n_0} . Since $M \leq |V_{d,n}|/|V_{d,n_0}|$, we obtain

$$\begin{aligned} \frac{1}{|V_{d,n}|} \log \mathbb{P}\left(x(\sigma_n), X(\sigma_n) \in B(x, \delta)\right) &\geq \frac{1}{|V_{d,n_0}|} \log \mathbb{P}\left((x(\sigma_{n_0}), X(\sigma_{n_0})) \in B(x, X, \delta/2)\right) \\ &\geq -\bar{I}_k(x, X) - \delta, \end{aligned}$$

where the second inequality follows from (15). Since this holds for all large enough n , then from the bound above we obtain

$$\lim_{\delta \rightarrow 0} \liminf_n \frac{1}{|V_{d,n}|} \log \mathbb{P}\left((x(\sigma_n), X(\sigma_n)) \in B((x, X), \delta)\right) \geq -\bar{I}_k(x, X).$$

We conclude that the relation (6) holds. \square

4 Other notions of convergence and comparison with LD-convergence

In this section we consider other types of convergence discussed in the earlier literature and compare them with LD-convergence. Later we will show that every other mode of convergence, except for colored-neighborhood-convergence, is implied by LD-convergence. We begin with the notion of left-convergence.

4.1 Left-convergence

We start with the definition of left-convergence as introduced in [BCKL13]:

Definition 3. *A sequence of graphs \mathbb{G}_n with uniformly bounded degrees is left-convergent if the limit*

$$\lim_{n \rightarrow \infty} |V_n|^{-1} \text{hom}(\mathbb{H}, \mathbb{G}_n) \tag{16}$$

exists for every connected, finite graph \mathbb{H} .

As already noted in [BCKL13], this notion is equivalent to the notion of Benjamini-Schramm convergence of the sequence $\mathbb{G}_n = (V_n, E_n)$. To define the latter, consider a fixed positive integer r , and a rooted graph \mathbb{H} of diameter at most r . Define $p_n(\mathbb{H}, r)$ to be the probability that the r -ball around a vertex U chosen uniformly at random from V_n is isomorphic to \mathbb{H} . The sequence is called Benjamini-Schramm convergent if the limit $p(\mathbb{H}, r) = \lim_n p_n(\mathbb{H}, r)$ exists for all r and all rooted graphs \mathbb{H} of diameter r or less.

Recall our definition of relation \sim between graph sequences. It is immediate that if \mathbb{G}_n is left-convergent and $\tilde{\mathbb{G}}_n \sim \mathbb{G}_n$ then $\tilde{\mathbb{G}}_n$ is also left-convergent with the same set of limits $p(\mathbb{H}, r)$. Thus left-convergence can be defined on the equivalency classes of graph sequences.

4.2 Right-convergence

For sequences of bounded degree graphs, the notion of right-convergence for a fixed target graph \mathbb{H} was defined in [BCKL13]. In this paper, we define right-convergence without referring to a particular target graph, and require instead the existence of limits for all soft-core graphs \mathbb{H} . Recall that for a weighted graph \mathbb{H} with vertex and edge weights given by the vector $\alpha = (\alpha_1(\mathbb{H}), \dots, \alpha_k(\mathbb{H}))$ and the matrix $A = (A_{ij}(\mathbb{H}))_{1 \leq i, j \leq k}$, respectively, the homomorphism number $\text{hom}(\mathbb{G}, \mathbb{H})$ is given by (4). Without loss of generality we assume $\alpha_i(\mathbb{H}) > 0$ for all i . We say that \mathbb{H} is soft-core if also $A_{ij}(\mathbb{H}) > 0$ for all i, j .

To motivate our definition of right-convergence, let us note that for a soft-core graph \mathbb{H} on k nodes,

$$\alpha_{\min}^{(D+1)|V_n|} \leq \alpha_{\min}^{|V_n|+|E_n|} \leq k^{-|V_n|} \text{hom}(\mathbb{G}_n, \mathbb{H}) \leq \alpha_{\max}^{|V_n|+|E_n|} \leq \alpha_{\max}^{(D+1)|V_n|}$$

where

$$\alpha_{\max} = \max(1, \max \alpha_i, \max A_{ij}) \quad \text{and} \quad \alpha_{\min} = \min(1, \min \alpha_i, \min A_{ij}). \quad (17)$$

Thus $\log \text{hom}(\mathbb{G}_n, \mathbb{H})$ grows linearly with $|V_n|$. We will define a sequence to be right-convergent if the coefficient of proportionality converges for all soft-core graphs.

Definition 4. *A graph sequence \mathbb{G}_n is defined to be right-convergent if for every soft-core graph \mathbb{H} , the limit*

$$-f(\mathbb{H}) \triangleq \lim_{n \rightarrow \infty} \frac{\log \text{hom}(\mathbb{G}_n, \mathbb{H})}{|V_n|}, \quad (18)$$

exists.

In the language of statistical physics, the quantity $\text{hom}(\mathbb{G}, \mathbb{H})$ is usually called the partition function of the soft-core model with interaction \mathbb{H} on \mathbb{G} , and the quantity $f(\mathbb{H})$ is called its free energy. Sometimes we will write $f_{(\mathbb{G}_n)}(\mathbb{H})$ instead of $f(\mathbb{H})$ to emphasize the dependence on the underlying graph sequence \mathbb{G}_n .

Observe that our definition is robust with respect to the equivalency notion \sim on graph sequences. Namely if $\tilde{\mathbb{G}}_n \sim \mathbb{G}_n$, then the limits (18) are the same. Indeed, deleting or adding one edge in \mathbb{G}_n increases the partition function by a factor at most $\max A_{ij}$ and decreases it by a factor at most $\min A_{ij}$. Thus adding/deleting $o(|V_n|)$ edges changes the partition function of a soft-core model by a factor of $\exp(o(|V_n|))$.

By contrast, the partition function of a hard-core model can be quite sensitive to adding or deleting edges, as we already demonstrated in the introduction, where we discussed the case of cycles. The same observation can be extended to the case when \mathbb{G}_n is a d -dimensional cylinder or torus, again parity of n determining the two-colorability property, see [BCKL13] for details. An even more trivial example is the following: let \mathbb{G}_n be a graph on n isolated node (no edges) when n is even, and $n - 2$ isolated nodes plus one edge when n is odd. Then there exists 1-coloring of \mathbb{G}_n if and only if n is even, again leading to the conclusion that \mathbb{G}_n is not converging on all simple graphs \mathbb{H} .

We adapted Definition 4 to avoid these pathological examples. Nevertheless, we still would like to be able to define the limits (18) for hard-core graphs \mathbb{H} , and do it in a way which is insensitive to changing from \mathbb{G}_n to some equivalent sequence $\tilde{\mathbb{G}}_n \sim \mathbb{G}_n$. There are two ways to

achieve this which are equivalent, as we establish below. Given $\lambda > 0$ and a (not necessarily soft-core graph) \mathbb{H} with edge weights given by a matrix $A = (A_{ij}(\mathbb{H}))$, let A_λ be the matrix with positive entries

$$(A_\lambda)_{ij} = \max\{\lambda, A_{ij}(\mathbb{H})\}$$

and let \mathbb{H}_λ be the corresponding soft-core graph.

Theorem 3. *Given a right-convergent sequence \mathbb{G}_n and a weighted graph \mathbb{H} , the following limit exists*

$$f_{(\mathbb{G}_n)}(\mathbb{H}) \triangleq \lim_{\lambda \rightarrow 0} f_{(\mathbb{G}_n)}(\mathbb{H}_\lambda). \quad (19)$$

Furthermore, there exists a graph sequence $\tilde{\mathbb{G}}_n \sim \mathbb{G}_n$ such that for every graph \mathbb{H}

$$-f_{(\mathbb{G}_n)}(\mathbb{H}) = \lim_n \frac{\log \text{hom}(\tilde{\mathbb{G}}_n, \mathbb{H})}{|V_n|} \geq \sup_{\hat{\mathbb{G}}_n \sim \mathbb{G}_n} \limsup_n \frac{\log \text{hom}(\hat{\mathbb{G}}_n, \mathbb{H})}{|V_n|}, \quad (20)$$

where the supremum is over all graphs sequences $\hat{\mathbb{G}}_n$ which are equivalent to \mathbb{G}_n . In particular, the maximizing graph sequence $\hat{\mathbb{G}}_n$ exists and can be chosen in such a way that the lim sup is actually a limit.

The intuition of the definition above is that we define $f_{(\mathbb{G}_n)}(\mathbb{H})$ by slightly softening the "non-edge" requirement of zero elements of the weight matrix A . This turns out to be equivalent to the possibility of removing $o(|V_n|)$ edges in the underlying graph in trying to achieve the largest possible limit within the equivalency classes of graph sequences. One can further give a statistical physics interpretation of this definition as defining the free energy at zero temperature as a limit of free energies at positive temperatures.

Proof. The existence of the limit (19) follows immediately by monotonicity: note that $\text{hom}(\mathbb{G}, \mathbb{H})$ is monotonically non-decreasing in every element of the weight matrix A .

The proof of the second part is more involved, even though the intuition behind it is again simple: First, one easily shows that monotonicity in λ and the fact that the limit (18) is not changed when we change $o(|V_n|)$ edges implies that for every $\tilde{\mathbb{G}}_n \sim \mathbb{G}_n$

$$\limsup_n \frac{\log \text{hom}(\tilde{\mathbb{G}}_n, \mathbb{H})}{|V_n|} \leq -f_{(\mathbb{G}_n)}(\mathbb{H}). \quad (21)$$

To prove that $f_{(\mathbb{G}_n)}(\mathbb{H})$ is achieved asymptotically by some $\tilde{\mathbb{G}}_n \sim \mathbb{G}_n$ we may assume without loss of generality that $f_{(\mathbb{G}_n)}(\mathbb{H}) < \infty$, since otherwise the identity holds trivially. Consider a "typical configuration" $\sigma : V_n \rightarrow V(\mathbb{H}_\lambda)$ contributing to $\text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda)$, and let $E_0(\sigma)$ be the set of edges $\{u, v\} \in E_n$ such that $A_{\sigma(u), \sigma(v)}(\mathbb{H}) = 0$,

$$E_0(\sigma) = \{(u, v) \in E_n : A_{\sigma(u), \sigma(v)}(\mathbb{H}) = 0\}. \quad (22)$$

For small λ , we expect the size of this set to grow slowly with $|V_n|$, since otherwise $\text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda)$ would become too small to be consistent with $\lim_{\lambda \rightarrow 0} f_{(\mathbb{G}_n)}(\mathbb{H}_\lambda) < \infty$. We might therefore hope that we can find a subset $E_{0,n} \subset E_n$ such that (i) $|E_{0,n}| = o(|V_n|)$ and (ii) removing the set

of edges $E_{0,n}$ from E_n leads to a graph sequence \tilde{G}_n such that $|V_n|^{-1} \log \text{hom}(\tilde{G}_n, \mathbb{H})$ is close to $-f_{(\mathbb{G}_n)}(\mathbb{H})$.

It will require a little bit of work to make this rather vague argument precise. Before doing so, let us give the proof of (21), which is much simpler: Indeed, by monotonicity, for every $\lambda > 0$ we have

$$\begin{aligned} \limsup_n \frac{\log \text{hom}(\tilde{G}_n, \mathbb{H})}{|V_n|} &\leq \limsup_n \frac{\text{hom}(\tilde{G}_n, \mathbb{H}_\lambda)}{|V_n|} \\ &= \limsup_n \frac{\log \text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda)}{|V_n|} \\ &= \lim_n \frac{\log \text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda)}{|V_n|} \\ &= -f_{(\mathbb{G}_n)}(\mathbb{H}_\lambda), \end{aligned}$$

where the first equality follows since \mathbb{H}_λ is a soft-core graph. Passing to the limit $\lambda \rightarrow 0$, we obtain (21), which in turn implies

$$\sup_{\tilde{G}_n \sim \mathbb{G}_n} \limsup_n \frac{\log \text{hom}(\tilde{G}_n, \mathbb{H})}{|V_n|} \leq -f_{(\mathbb{G}_n)}(\mathbb{H}).$$

We now show that $f_{(\mathbb{G}_n)}(\mathbb{H})$ is achieved asymptotically by some $\tilde{G}_n \sim \mathbb{G}_n$. Our main technical result leading to the claim is as follows.

Lemma 5. *For every $\epsilon \in (0, 1)$, every graph \mathbb{H} , and for all large enough n , there exists a graph \tilde{G}_n which is obtained from \mathbb{G}_n by deleting at most $\epsilon|V_n|$ edges such that*

$$\text{hom}(\tilde{G}_n, \mathbb{H}) \geq \exp\left(-\left(f_{(\mathbb{G}_n)}(\mathbb{H}) + \epsilon\right)|V_n|\right). \quad (23)$$

We first show that this lemma implies the required claim. It relies on a standard diagonalization argument. Note that it suffices to establish the result for graphs \mathbb{H} with rational (possibly zero) weights. Let $\mathbb{H}_1, \mathbb{H}_2, \dots$ be an arbitrary enumeration of graphs with rational weights. For every $m = 1, 2, \dots$, we find large enough $n(m)$ such that for $n \geq n(m)$, the claim in lemma holds for $\epsilon = 1/m^2$ and for all $\mathbb{H} = \mathbb{H}_1, \dots, \mathbb{H}_m$. Namely, for every $n \geq n(m)$, graphs $\tilde{G}_{n,1}, \dots, \tilde{G}_{n,m}$ can be obtained from the graph \mathbb{G}_n each by deleting at most n/m^2 of edges in \mathbb{G}_n such that (23) holds for $\epsilon = 1/m^2$. Without loss of generality we may assume that $n(m)$ is strictly increasing in m . Let $\tilde{E}_{n,i}$ be the set of edges deleted from \mathbb{G}_n to obtain $\tilde{G}_{n,i}$ for $1 \leq i \leq m$ and \tilde{G}_n^m be the graph obtained from \mathbb{G}_n by deleting the edges in $\cup_{1 \leq i \leq m} \tilde{E}_{n,i}$. The number of deleted edges is at most $m(n/m^2) = n/m$. By monotonicity, (23) implies that

$$\text{hom}(\tilde{G}_n^m, \mathbb{H}_i) \geq \exp\left(-\left(f_{(\mathbb{G}_n)}(\mathbb{H}_i) + 1/m^2\right)|V_n|\right), \quad i = 1, 2, \dots, m.$$

We now construct the graph sequence \tilde{G}_n . For all $n < n(1)$ we simply set $\tilde{G}_n = \mathbb{G}_n$. For all other n , we find a unique m_n such that $n(m_n) \leq n < n(m_n + 1)$ and set $\tilde{G}_n = \tilde{G}_n^{m_n}$. Then $\tilde{G}_n \sim \mathbb{G}_n$ since the number of deleted edges is at most $|V_n|/m_n = o(|V_n|)$ as $n \rightarrow \infty$. For this sequence we have

$$|V_n|^{-1} \log \text{hom}(\tilde{G}_n, \mathbb{H}_i) \geq -f_{(\mathbb{G}_n)}(\mathbb{H}_i) - 1/m_n^2, \quad i = 1, 2, \dots, m_n.$$

Fixing i and taking \liminf_n of each side we obtain for each i

$$\liminf_n |V_n|^{-1} \log \text{hom}(\tilde{\mathbb{G}}_n, \mathbb{H}_i) \geq -f_{(\mathbb{G}_n)}(\mathbb{H}_i).$$

Combining with (21) we conclude that for each i

$$\lim_n |V_n|^{-1} \log \text{hom}(\tilde{\mathbb{G}}_n, \mathbb{H}_i) = -f_{(\mathbb{G}_n)}(\mathbb{H}_i).$$

This completes the proof of the theorem. \square

Proof of Lemma 5. For the ease of exposition we write $f(\mathbb{H})$ for $f_{(\mathbb{G}_n)}(\mathbb{H})$. If $f(\mathbb{H}) = \infty$, there is nothing to prove, so we assume $f(\mathbb{H}) < \infty$.

Let

$$\alpha_\lambda(\sigma) \triangleq \prod_{u \in V_n} \alpha_{\sigma(u)} \prod_{(u,v) \in E_n} \max(A_{\sigma(u), \sigma(v)}, \lambda)$$

and let $\Sigma(E)$ be the set of $\sigma : V_n \rightarrow [k]$ such that $E_0(\sigma) = E$. Then

$$\text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda) = \sum_{E \subset E_n} \sum_{\sigma \in \Sigma(E)} \alpha_\lambda(\sigma).$$

We will prove that for λ sufficiently small and n sufficiently large, we can find a set $E_0 \subset E_n$ such that $|E_0| \leq \epsilon |V_n|$ and

$$\sum_{\sigma \in \Sigma(E_0)} \alpha_\lambda(\sigma) \geq \exp(-(f(\mathbb{H}) + \epsilon)|V_n|). \quad (24)$$

Note that this bound immediately implies the claim of the lemma. Indeed, let $\tilde{\mathbb{G}}_n$ be the graph obtained from \mathbb{G}_n by deleting the edges in E_0 and assume with out loss of generality that λ is smaller than the smallest non-zero entry of A . Then every $\sigma \in \Sigma(E_0)$ satisfies $\alpha_\lambda(\sigma) = \lambda^{|E_0|} \tilde{\alpha}(\sigma)$, where $\tilde{\alpha}$ is the weight with respect to the graph $\tilde{\mathbb{G}}_n$, vector α and the matrix A . This implies

$$\begin{aligned} \text{hom}(\tilde{\mathbb{G}}_n, \mathbb{H}) &\geq \sum_{\sigma \in \Sigma(E_0)} \tilde{\alpha}(\sigma) = \lambda^{-|E_0|} \sum_{\sigma \in \Sigma(E_0)} \alpha_\lambda(\sigma) \\ &\geq \sum_{\sigma \in \Sigma(E_0)} \alpha_\lambda(\sigma) \geq \exp(-(f(\mathbb{H}) + \epsilon)|V_n|), \end{aligned}$$

as required. The proof of the lemma is therefore reduced to the proof of (24).

Given $0 < \epsilon < 1$ we chose $0 < \hat{\lambda} < \epsilon$ such that

$$3\hat{\lambda} + \hat{\lambda} \log D + \hat{\lambda} \log(1/\hat{\lambda})^{-1} < \epsilon, \quad (25)$$

and $0 < \lambda < \hat{\lambda}$ such that

$$\frac{f(\mathbb{H}) + 2 + (D + 1) \log \alpha_{\max} + \log k}{\log(\lambda^{-1})} < \hat{\lambda}. \quad (26)$$

Here we note that $-f(\mathbb{H}) \leq \log k + (D + 1) \log \alpha_{\max}$, and thus the numerator in the left-hand side is positive. Let $n_0 = n_0(\lambda)$ be large enough so that

$$\log(D|V_n|) \leq \lambda|V_n| \quad \text{and} \quad \text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda) \geq \exp((-f(\mathbb{H}_\lambda) - \lambda)|V_n|)$$

for all $n \geq n_0(\lambda)$. Note that by monotonicity in λ , the second bound implies that

$$\text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda) \geq \exp((-f(\mathbb{H}) - \lambda)|V_n|). \quad (27)$$

Fix an arbitrary such $n \geq n_0$, and let $\Sigma(r)$ be the set of mappings $\sigma : V_n \rightarrow V(\mathbb{H})$ such that precisely r edges of V_n are mapped into non-edges of \mathbb{H} ,

$$\Sigma(r) = \{\sigma : V_n \rightarrow V(\mathbb{H}) : |E_0(\sigma)| = r\}.$$

For all λ which are smaller than the smallest positive element of A we have

$$\text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda) = \sum_{0 \leq r \leq |E_n|} \sum_{\sigma \in \Sigma(r)} \alpha_\lambda(\sigma).$$

Let r_0 be such that

$$\sum_{\sigma \in \Sigma(r_0)} \alpha_\lambda(\sigma) \geq |E_n|^{-1} \text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda). \quad (28)$$

We claim that $r_0 \leq \hat{\lambda}|V_n|$. Indeed, since for every $\sigma \in \Sigma(r_0)$ we have $\alpha_\lambda(\sigma) \leq \lambda^{r_0} \alpha_{\max}^{|V_n|+|E_n|} \leq \lambda^{r_0} \alpha_{\max}^{(D+1)|V_n|}$, it follows from (28) that

$$k^{|V_n|} \lambda^{r_0} \alpha_{\max}^{(D+1)|V_n|} \geq |E_n|^{-1} \text{hom}(\mathbb{G}_n, \mathbb{H}_\lambda) \geq \exp((-f(\mathbb{H}) - 2\lambda)|V_n|),$$

where the second inequality follows from (27) and the fact that for $n \geq n_0$, we have $|E_n| \leq D|V_n| \leq e^{\lambda|V_n|}$. Rearranging, we get that

$$r_0 \leq (\log(1/\lambda))^{-1} \left(|V_n| (f(\mathbb{H}) + 2\lambda + (D + 1) \log \alpha_{\max} + \log k) \right).$$

Applying (26), this in turn implies that

$$\begin{aligned} r_0 &\leq (\log \lambda^{-1})^{-1} |V_n| (f(\mathbb{H}) + 2\lambda + (D + 1) \log \alpha_{\max} + \log k) \\ &\leq (\log \lambda^{-1})^{-1} |V_n| (f(\mathbb{H}) + 2 + (D + 1) \log \alpha_{\max} + \log k) \\ &\leq \hat{\lambda} |V_n|, \end{aligned} \quad (29)$$

as claimed.

Next we observe that

$$\sum_{\sigma \in \Sigma(r_0)} \alpha_\lambda(\sigma) = \sum_{\substack{E \subset E_n: \\ |E|=r_0}} \sum_{\sigma \in \Sigma(E)} \alpha_\lambda(\sigma).$$

Since the number of subsets of the edges of \mathbb{G}_n with cardinality r_0 is $\binom{|E_n|}{r_0}$, we can find a subset $E_0 \subset E_n$, $|E_0| = r_0$ such that

$$\begin{aligned} \sum_{\sigma \in \Sigma(E_0)} \alpha_\lambda(\sigma) &\geq \binom{|E_n|}{r_0}^{-1} \sum_{\sigma \in \Sigma(r_0)} \alpha_\lambda(\sigma) \\ &\geq \binom{|E_n|}{r_0}^{-1} \exp((-f(\mathbb{H}) - 2\lambda)|V_n|) \end{aligned} \quad (30)$$

where the second inequality follows from (28). Next, we note that $\binom{N}{k} \leq \left(\frac{Ne}{k}\right)^k$, a fact which can easily be proved by induction on k . Combining this bound with the fact that $|E_n| \leq D|V_n|$ and the bounds (29) and (25), we get

$$\binom{|E_n|}{r_0} \leq \binom{D|V_n|}{r_0} \leq \binom{D|V_n|}{\hat{\lambda}|V_n|} \leq \left(\frac{eD}{\hat{\lambda}}\right)^{\hat{\lambda}|V_n|} = e^{((1+\log D)\hat{\lambda} + \hat{\lambda} \log(\hat{\lambda}^{-1}))|V_n|} \leq e^{(\epsilon - 2\hat{\lambda})|V_n|}.$$

Combined with (30) this implies the bound (24). \square

4.3 Partition-convergence and colored-neighborhood-convergence

The notions of partition-convergence and colored-neighborhood-convergence were introduced in Bollobas and Riordan [BR11]. Recall the notion of a graph quotient $\mathbb{F} = \mathbb{G}/\sigma$ defined for every graph \mathbb{G} and mapping $\sigma : V(\mathbb{G}) \rightarrow [m]$. Then \mathbb{F} is a weighted graph on m nodes with node and edge weights $(x(\sigma), X(\sigma)) \in [0, D]^{(m+1) \times m}$ defined by (3). When \mathbb{G} is the n -th element of a graph sequence \mathbb{G}_n , we will use notations $x_n(\sigma)$ and $X_n(\sigma)$. Let $\Sigma_n(m)$ be the set of all pairs $(x_n(\sigma), X_n(\sigma))$, when we vary over all maps σ . As such $\Sigma_n(m) \in \mathcal{P}([0, D]^{(m+1) \times m})$, where $\mathcal{P}(A)$ is the set of closed subsets of A . For every set $A \subset [0, D]^{(m+1) \times m}$, let A^ϵ be the set of points in $[0, D]^{(m+1) \times m}$ with distance at most ϵ to A with respect to the \mathbb{L}_∞ norm on $[0, D]^{(m+1) \times m}$ (the actual choice of metric is not relevant). Define a metric ρ_m on closed sets in $\mathcal{P}([0, D]^{(m+1) \times m})$ via

$$\rho(A, B) = \inf_{\epsilon > 0} \{B \subset A^\epsilon, A \subset B^\epsilon\}. \quad (31)$$

Definition 5. *A graph sequence \mathbb{G}_n is partition-convergent if for every m , the sequence $\Sigma_n(m)$, $n \geq 1$ is a Cauchy sequence in the metric space $\mathcal{P}([0, D]^{(m+1) \times m})$.*

It is known that $\mathcal{P}([0, D]^{(m+1) \times m})$ is a compact (thus closed) metric space, which implies that $\Sigma_n(m)$ has a limit $\Sigma(m) \in \mathcal{P}([0, D]^{(m+1) \times m})$. Loosely speaking $\Sigma(m)$ describes the limiting set of graph quotients achievable through the coloring of nodes by m different colors. Thus, if for example some pair (x, X) belong to $\Sigma(m)$, this means that there is a sequence of partitions of \mathbb{G}_n into m colors such that the normalized number of nodes colored i converges to x_i for each i , and the normalized number of edges between colors i and j converges to $X_{i,j}$ for each i, j .

The definition above raises the question whether partition-convergence implies convergence of neighborhoods, namely left-convergence. We will show in the next section that this is not the case: partition-convergence does not imply left-convergence (whereas, as we mentioned

above, right-convergence does imply the left-convergence). In order to address this issue Bollobas and Riordan extended their definition of partition-convergence [BR11] to a richer notion called *colored-neighborhood-convergence*. This notion was further studied by Hatami, Lovász and Szegedi [HLS12] under the name *local-global* convergence, to stress the fact that this notion captures both global and local properties of the sequence \mathbb{G}_n . We prefer the original name, given that both right and LD-convergence capture local and global properties as well.

To define colored-neighborhood-convergence, we fix positive integers m and r . Let $\mathcal{H}_{m,r} \subset \mathcal{H}$ denote the finite set of all rooted colored graphs with radius at most r and degree at most Δ , colored using colors $1, \dots, m$. For every graph \mathbb{G} with degree $\leq \Delta$ and every node $u \in V(\mathbb{G})$, every coloring $\sigma : V(\mathbb{G}) \rightarrow [m]$ produces an element $\mathbb{H} \in \mathcal{H}_{m,r}$, which is the colored r -neighborhood $B_{\mathbb{G}}(u, r)$ of u in \mathbb{G} . For each $\mathbb{H} \in \mathcal{H}_{m,r}$, let $N(\mathbb{G}, m, r, \sigma, \mathbb{H})$ be the number of nodes u such that there exists a color matching isomorphism from $B_{\mathbb{G}}(u, r)$ to \mathbb{H} . Namely, it is the number of times that \mathbb{H} appears as a colored neighborhood $B_{\mathbb{G}}(u, r)$ when we vary u . Let $\pi(\mathbb{G}, m, r, \sigma)$ be the vector $(|V(\mathbb{G})|^{-1}N(\mathbb{G}, m, r, \sigma, \mathbb{H}), \mathbb{H} \in \mathcal{H}_{m,r})$. Namely, it is the vector of frequencies of observing graphs \mathbb{H} , as neighborhoods $B_{\mathbb{G}}(u, r)$ when we vary u . Then for every m , we obtain a finite and therefore closed set $\pi(\mathbb{G}, m, r) \subset [0, 1]^{|\mathcal{H}_{m,r}|}$ when we vary over all m -colorings σ in vectors $\pi(\mathbb{G}, m, r, \sigma)$. As such, $\pi(\mathbb{G}, m, r) \in \mathcal{P}([0, 1]^{|\mathcal{H}_{m,r}|})$. We consider $\mathcal{P}([0, 1]^{|\mathcal{H}_{m,r}|})$ as a metric space with metric given by (31).

Definition 6. *A graph sequence \mathbb{G}_n is defined to be colored-neighborhood-convergent if the sequence $\pi(\mathbb{G}_n, m, r)$ is convergent in $\mathcal{P}([0, 1]^{|\mathcal{H}_{m,r}|})$ for every m, r .*

It is not too hard to see that colored-neighborhood-convergence implies partition-convergence and left-convergence. As we show in the next section, the converse does not hold for either partition- or left-convergence.

5 Relationship between different notions of convergence

In this section we explore the relations between different notions of convergence. As we will see, most of the definitions are non-equivalent and many of them are not comparable in strength. Our findings are summarized in the following theorem.

Theorem 4. *The notions of left, right, partition, colored-neighborhood and LD-convergence satisfy the relations described on Figures 3 and 4, where the solid arrow between A and B means type A convergence implies type B convergence, and the striped arrow between A and B means type A convergence does not imply type B convergence.*

We have deliberately described the relations in two figures, since at this stage we do not know how LD and colored-neighborhood-convergence are related to each other. As a step towards this it would be interesting to see whether colored-neighborhood-convergence implies right-convergence. We will discuss this question along with some other open questions in Section 6.

Note that it was already proved in [BCKL13] that right-convergence implies left-convergence, and that it is obvious that colored-neighborhood-convergence implies partition and left-convergence. To prove the Theorem, it will therefore be enough to show that LD-convergence implies both right and partition-convergence, that left-convergence does not imply right-convergence, that partition-convergence does not imply left-convergence, and that right-convergence does not

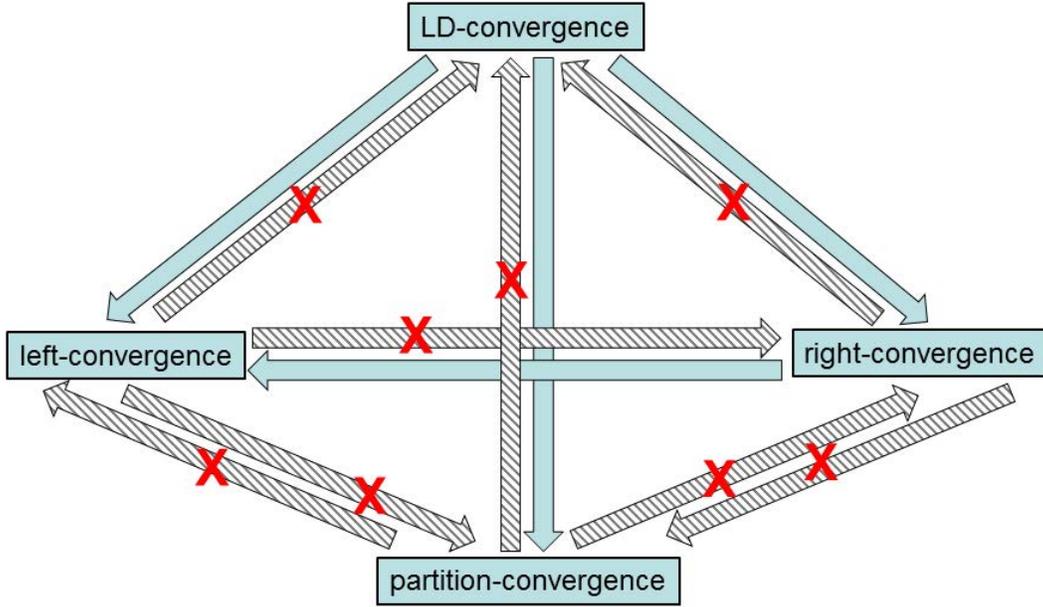


Figure 3: Implications between LD and other notions of convergence

imply partition-convergence (all other arrows are implied). We will start with the negative implications, which are all proved by giving counter-examples.

5.1 Left-convergence does not imply right-convergence

As discussed earlier left-convergence does not imply right-convergence on all hard-core graphs \mathbb{H} . But our modified definition of right-convergence involving only soft-core graphs raises the question whether with our new definition, left-convergence does imply right-convergence. It turns out that this is not the case since right-convergence implies convergence of certain global properties like the convergence of the maximal cut in a graph, a property which is not captured by the notion of left-convergence.

Indeed, let \mathbb{G}_n be right-convergent, let $\text{MaxCut}(\mathbb{G}_n)$ denotes the size of the maximal cut in \mathbb{G}_n , and let \mathbb{H} be a graph consisting of a single edge with weights $\alpha = (1, 1)$ and $A_{1,2} = A_{2,1} = e^\beta$, $A_{1,1} = A_{2,2} = 1$. Then

$$e^{\beta \text{MaxCut}(\mathbb{G}_n)} \leq \text{hom}(\mathbb{G}_n, \mathbb{H}) \leq 2^n e^{\beta \text{MaxCut}(\mathbb{G}_n)}.$$

Sending $\beta \rightarrow \infty$, we see that right-convergence implies convergence of $\frac{1}{|V(\mathbb{G}_n)|} \text{MaxCut}(\mathbb{G}_n)$.

As an immediate consequence, we obtain that left-convergence does not implies right-convergence.

Example 3 ([BCKL13]). *Let Δ be a positive integer, let $\mathbb{G}(n, \Delta)$ denote a random Δ -regular graph on n nodes, and let $B_{n, \Delta}$ denote a random Δ -regular bi-partite graph on n nodes. Let \mathbb{G}_n be equal to $\mathbb{G}(n, \Delta)$ when n is odd and equal to $B_{n, \Delta}$ when n is even. Then \mathbb{G}_n is left-convergent for all Δ , but for Δ sufficiently large, it is not right-convergent.*

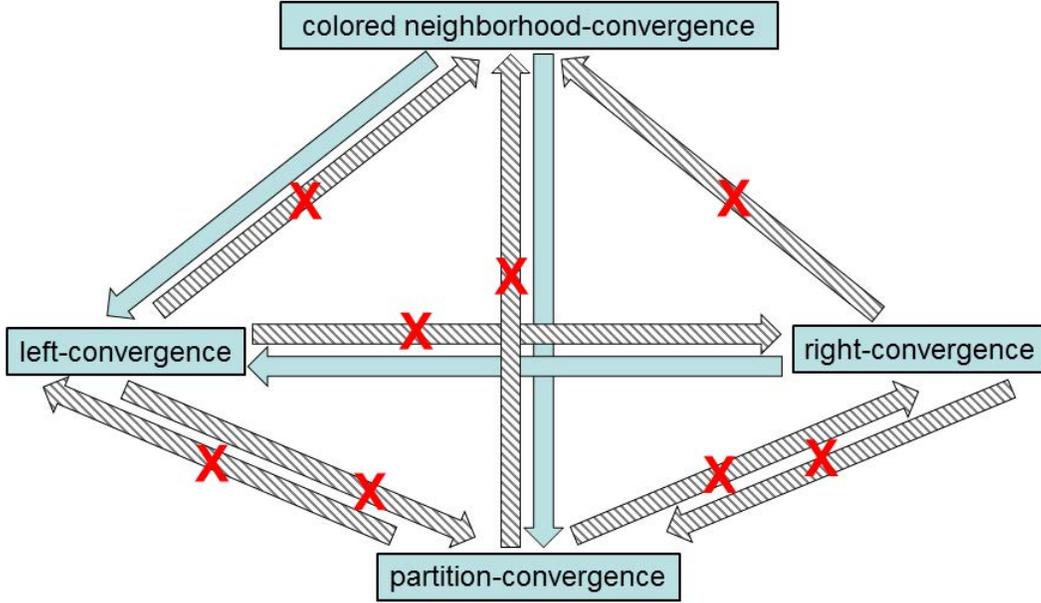


Figure 4: Implications between colored-neighborhood and other notions of convergence

Proof. \mathbb{G}_n is left-convergent since for every r , with high probability $B(U, r)$ is a Δ -Kelly (regular) tree truncated at depth r , where U is a random uniformly chosen node in \mathbb{G}_n . We now show that \mathbb{G}_n is not right-convergent for large Δ . Indeed, $\text{MaxCut}(\mathbb{G}_n) = \text{MaxCut}(B_{n,\Delta}) = \frac{1}{2}\Delta n$ for even n , while $\text{MaxCut}(\mathbb{G}_n) = \text{MaxCut}(\mathbb{G}(n, \Delta)) = \frac{1}{4}\Delta n + O(\sqrt{\Delta n})$ for odd n [BCP97]. This implies that for large Δ , the sequence \mathbb{G}_n is not right-convergent. \square

5.2 Right-convergence does not imply partition-convergence

In order to construct our counter example, we need the following lemma. We recall that the edge expansion constant of a sequence of graphs \mathbb{G}_n is the largest constant γ such that the number of edges between W and $V(\mathbb{G}_n) \setminus W$ is larger than $\gamma|W|$ whenever $|W| \leq |V(\mathbb{G}_n)|/2$.

Lemma 6. *Let $\Delta \geq 3$. Then there exists a $\gamma > 0$ and a right-convergent sequence of expanders \mathbb{G}_n with edge expansion constant γ and maximal degree bounded by Δ .*

Proof. Consider an arbitrary sequence $\mathbb{G}_n = (V_n, E_n)$ of expanders with edge expansion $\gamma > 0$ and maximal degree at most Δ , e.g., a sequence of 3-regular random graphs. Construct a right-convergent subsequence of this sequence by the following, standard diagonalization argument: enumerate the countable collection of weighted graphs with rational positive weights: $\mathbb{H}_1, \mathbb{H}_2, \dots$. For each $m = 1, 2, \dots$, construct a nested subsequence $n^{(m)} \subset n^{(m-1)}$, such that \mathbb{G}_n is right-convergent with respect to $\mathbb{H}_1, \dots, \mathbb{H}_m$ along the subsequence $n^{(m)}$, and then take the diagonal subsequence $n^{(m)}$ - the m -element of the m -th sequence. The constructed sequence of graphs is right-convergent with respect to every weighted graph \mathbb{H} with rational coefficients. A straightforward argument shows that then the sequence is right-convergent with respect to all graphs with positive real weights, namely it is right-convergent. \square

Example 4. Given a right-convergent sequence of expanders \mathbb{G}_n with bounded maximal degree and edge expansion $\gamma > 0$, let $\tilde{\mathbb{G}}_n = \mathbb{G}_m$ if $n = 2m$ is even, and let $\tilde{\mathbb{G}}_n$ be a disjoint union of two copies of \mathbb{G}_m if $n = 2m + 1$ is odd. Then $(\tilde{\mathbb{G}}_n)$ is right-convergent but not partition-convergent.

Proof. Let $G_n = (V_n, E_n)$. For every weighted graph \mathbb{H} ,

$$\begin{aligned} |\tilde{V}_{2n+1}|^{-1} \log Z(\tilde{\mathbb{G}}_{2n+1}, \mathbb{H}) &= (2|V_n|)^{-1} (2 \log Z(\mathbb{G}_n, \mathbb{H})) \\ &= |V_n|^{-1} \log Z(\mathbb{G}_n, \mathbb{H}) \\ &= |\tilde{V}_{2n}|^{-1} \log Z(\tilde{\mathbb{G}}_{2n}, \mathbb{H}). \end{aligned}$$

Since \mathbb{G}_n is right-convergent, $\tilde{\mathbb{G}}_n$ is right-convergent as well.

On the other hand, we claim that $\tilde{\mathbb{G}}_n$ is not partition-convergent, since every cut of $\tilde{\mathbb{G}}_n = \mathbb{G}_m$ into two equal parts has size at least $\gamma|V_n|/2$ if n is even, while $\tilde{\mathbb{G}}_n$ has a cut into equal parts with zero edges when n is odd. Indeed, denoting the node set of $\tilde{\mathbb{G}}_n$ by \tilde{V}_n , consider the set $\Sigma_n(2)$ of pairs $(x_n(\sigma), X_n(\sigma))$ of vectors $x_n(\sigma) = (x_1^n(\sigma), x_2^n(\sigma))$ and matrices $X_n(\sigma, 2) = (X_{i,j}^n(\sigma), 1 \leq i, j \leq 2)$ when we vary over $\sigma : \tilde{V}_n \rightarrow [2]$. Consider the projection of $\Sigma_n(2)$ onto $[0, D]^3$ corresponding to $(x_1^n(\sigma), x_2^n(\sigma), X_{1,2}^n(\sigma))$. Consider the point $a = (1/2, 1/2, 0) \in [0, D]^3$. Observe that for each n , there exists a σ such that $(x_1^{2n+1}(\sigma), x_2^{2n+1}(\sigma), X_{1,2}^{2n+1}(\sigma)) = a$, since we can split the graph $\tilde{\mathbb{G}}_{2n+1}$ into two parts with the same number of nodes in each part and no edges in between the parts. On the other hand for each graph $\tilde{\mathbb{G}}_{2n}$ and each σ , the distance between $(x_1^{2n}(\sigma), x_2^{2n}(\sigma), X_{1,2}^{2n}(\sigma))$ and a is bounded from below by $\min\{1/4, \gamma/4\}$ (indeed, if $|x_1^{2n}(\sigma) - 1/2| < 1/4$, then by the expansion property of G_n , $X_{1,2}^{2n}(\sigma) > \gamma/4$). This shows that the sets $\{(x_1^n(\sigma), x_2^n(\sigma), X_{1,2}^n(\sigma))\}$ obtained by varying σ are not converging as $n \rightarrow \infty$ and thus the sequence $\tilde{\mathbb{G}}_n$ is not partition-convergent. \square

5.3 Partition-convergence does not imply left-convergence

We now exhibit an example of a graph sequence which is partition-convergent but not left-convergent.

Example 5. Let \mathbb{G}_n be a disjoint union of n 4-cycles when n is even and a disjoint union of n 6-cycles when n is odd. Then \mathbb{G}_n is partition-convergent and not left-convergent.

Proof. \mathbb{G}_n is clearly not left-convergent. Let us show that it is partition-convergent. For this purpose set $D = 2$ and consider arbitrary k . Let $\Sigma(k)$ be the set of all $(x, X) \in [0, D]^{(k+1) \times k}$ such that $\sum_{1 \leq i \leq k} x_i = 1$, $X_{i,j} = X_{j,i}$ for all i, j , and

$$\sum_{j=1}^k X_{i,j} = 2x_i \quad \text{for all } i. \tag{32}$$

We will prove that $\Sigma_n(k) \rightarrow \Sigma(k)$ which shows that the sequence is partition-convergent with $\Sigma(k)$ being the limit.

To this end, we first show that $\Sigma_n(k) \subset \Sigma(k)$ for all n . Indeed, let $\sigma : V_n \rightarrow [k]$. By the definition (3), we have that

$$\sum_{j=1}^k X_{i,j}(\sigma) = \frac{1}{|V_n|} \sum_{u \in V_i} \sum_{\substack{v \in V_n: \\ (u,v) \in E_n}} 1 = \frac{1}{|V_n|} \sum_{u \in V_i} 2 = 2x_i(\sigma),$$

where we used that the degree of all vertices in \mathbb{G}_n is two. This proves that (32) holds for all $(x, X) \in \Sigma_n(k)$.

We now claim that for every element $(x, X) \in \Sigma(k)$ and every n there exists $\sigma_n : V_n \rightarrow [k]$ such that the \mathbb{L}_∞ distance between (x, X) and $(x(\sigma_n), X(\sigma_n))$ is at most k/n . This is clearly enough to complete the proof, since it implies that for every $\epsilon > 0$ and n large enough, we have $\Sigma(k) \subset (\Sigma_n(k))^\epsilon$.

To prove the claim, we first consider even n so that \mathbb{G}_n is a disjoint union of n 4-cycles. To this end, let $n_{i,j}$, $i, j \in [k]$, be such that

$$\left| n_{i,j} - \frac{n}{2} X_{i,j} \right| \leq 1 \quad \text{for all } i, j \in [k]$$

and

$$\sum_{i,j \in [k]} n_{i,j} = \sum_{i,j \in [k]} \frac{n}{2} X_{i,j} = n.$$

We consider the following coloring σ_n of \mathbb{G}_n : For each $i \neq j$ we take $n_{i,j}$ cycles and color their 4 vertices alternating between color i and j as we go around these cycles, while for $i = j$, we take $n_{i,i}$ cycles and color them with only one color, $i = j$. For this coloring, we have

$$X_{i,j}(\sigma) = X_{j,i}(\sigma) = 4 \frac{n_{i,j} + n_{j,i}}{4n},$$

so by our choice of $n_{i,j}$ we have that

$$\left| X_{i,j} - X_{i,j}(\sigma) \right| \leq \frac{2}{n}.$$

Since $\Sigma_n(k) \subset \Sigma(k)$, we have

$$|x_i - x_i(\sigma)| = \frac{1}{2} \left| \sum_{j=1}^k (X_{i,j} - X_{i,j}(\sigma)) \right| \leq \frac{k}{n},$$

completing the proof of the claim when n is even. A similar result is established for odd n , when \mathbb{G}_n is a disjoint union of n 6-cycles. \square

5.4 LD-convergence implies right-convergence

Let the sequence \mathbb{G}_n be LD-convergent. Applying Theorem 1 it is also k -LD-convergent with rate functions I_k , $k \geq 1$. Consider an arbitrary k -node weighted graph \mathbb{H} with strictly positive node weights $\alpha = (\alpha_i, 1 \leq i \leq k)$ and strictly positive edge weights $A = (A_{i,j}, 1 \leq i, j \leq k)$. We define an “energy functional” $\mathcal{E}_\mathbb{H}$ on $[0, D]^{(k+1) \times k}$ by

$$\mathcal{E}_\mathbb{H}(x, X) = - \sum_{1 \leq i \leq k} x_i \log \alpha_i - \frac{1}{2} \sum_{1 \leq i, j \leq k} X_{i,j} \log A_{i,j}. \quad (33)$$

We also introduce an “entropy functional” \mathcal{S}_k on $[0, D]^{(k+1) \times k}$ by setting

$$\mathcal{S}_k(x, X) = \log k - I_k(x, X). \quad (34)$$

Theorem 5. *Suppose the sequence of graphs \mathbb{G}_n is LD-convergent. Then it is also right-convergent and*

$$-\lim_{n \rightarrow \infty} \frac{1}{|V_n|} \log Z(\mathbb{G}_n, \mathbb{H}) = \inf_{(x, X) \in [0, D]^{(k+1) \times k}} \left(\mathcal{E}_{\mathbb{H}}(x, X) - \mathcal{S}_k(x, X) \right). \quad (35)$$

Remark 2. *By the lower semi-continuity of I_k , the infimum over $(x, X) \in [0, D]^{(k+1) \times k}$ is attained for some $(x^*, X^*) \in [0, D]^{(k+1) \times k}$. The right hand side can then be rewritten as $\mathcal{E}_{\mathbb{H}}(x^*, X^*) - \mathcal{S}_k(x^*, X^*)$. The theorem thus states that the limiting free energy exists, and is given as the energy minus entropy, an identity which is usually assumed axiomatically in traditional thermodynamics. Here it is a consequence of LD-convergence.*

Remark 3. *By Remark 1, we can express the limiting free energy directly in terms of the rate function I . Defining $\hat{\mathcal{E}}_{\mathbb{H}}(\rho, \mu) = \mathcal{E}_{\mathbb{H}}(x(\rho), X(\mu))$ where $x(\rho)$ and $X(\mu)$ are given by (9), this gives*

$$-\lim_{n \rightarrow \infty} \frac{1}{|V_n|} \log Z(\mathbb{G}_n, \mathbb{H}) = \inf_{(\rho, \mu) \in \mathcal{M}^1 \times \mathcal{M}^2} \left(\hat{\mathcal{E}}_{\mathbb{H}}(\rho, \mu) + I(\rho, \mu) - \log k \right).$$

Proof of Theorem 5. Fix $\delta > 0$. Recalling the definition (3), let $\Sigma(x, X)$ be the set of colorings $\sigma : V_n \rightarrow [k]$ such that

$$x_i - \delta \leq x_i(\sigma) \leq x_i + \delta.$$

for every $i = 1, \dots, k$ and

$$X_{i,j} - \delta \leq X_{i,j}(\sigma) \leq X_{i,j} + \delta.$$

for every $i, j = 1, 2, \dots, k$. Note that we have the identity

$$|\Sigma(x, X)| = k^{|V_n|} \mathbb{P}_{k,n} \left(B((x, X), \delta) \right),$$

where we recall that $B((x, X), \delta)$ denotes the closed δ -ball around (x, X) with respect to the \mathbb{L}_{∞} norm, see Subsection 3.1. Note also that

$$\left| \mathcal{E}_{\mathbb{H}}(x(\sigma), X(\sigma)) - \mathcal{E}_{\mathbb{H}}(x, X) \right| \leq K\delta, \quad K = \left(k + \frac{k^2}{2} \right) \max\{\log \alpha_i, \log A_{i,j}\},$$

whenever $\sigma \in \Sigma(x, X)$.

Define Γ_{δ} to be the set of pairs $(x, X) \in [0, D]^{(k+1) \times k}$ such that every x_i and every $X_{i,j}$ belongs to the set $\{0, \delta, 2\delta, \dots, \lceil D/\delta \rceil \delta\}$. Since $X_{i,j}(\sigma) \leq D$ for all $\sigma : V_n \rightarrow [k]$, every $\sigma : V_n \rightarrow [k]$ lies in $\cup_{(x, X) \in \Gamma_{\delta}} \Sigma(x, X)$. As a consequence,

$$\begin{aligned} Z(\mathbb{G}_n, \mathbb{H}) &\leq \sum_{(x, X) \in \Gamma_{\delta}} \sum_{\sigma \in \Sigma(x, X)} \prod_{u \in V_n} \alpha_{\sigma(u)} \prod_{(u, v) \in E_n} A_{\sigma(u), \sigma(v)} \\ &= \sum_{(x, X) \in \Gamma_{\delta}} \sum_{\sigma \in \Sigma(x, X)} e^{-\mathcal{E}_{\mathbb{H}}(x(\sigma), X(\sigma))|V_n|} \\ &\leq \sum_{(x, X) \in \Gamma_{\delta}} k^{|V_n|} \mathbb{P}_{k,n} \left(B((x, X), \delta) \right) e^{-\mathcal{E}_{\mathbb{H}}(x, X)|V_n| + K\delta|V_n|} \\ &\leq |\Gamma_{\delta}| \max_{(x, X) \in \Gamma_{\delta}} \left(k^{|V_n|} \mathbb{P}_{k,n} \left(B((x, X), \delta) \right) e^{-\mathcal{E}_{\mathbb{H}}(x, X)|V_n| + K\delta|V_n|} \right). \end{aligned}$$

We obtain

$$\begin{aligned} \frac{1}{|V_n|} \log Z(\mathbb{G}_n, \mathbb{H}) &\leq \frac{1}{|V_n|} \log |\Gamma_\delta| + \log k \\ &\quad + \max_{(x, X) \in \Gamma_\delta} \left(\frac{1}{|V_n|} \log \mathbb{P}_{k, n} \left(B((x, X), \delta) \right) - \mathcal{E}_{\mathbb{H}}(x, X) + K\delta \right), \end{aligned}$$

and hence

$$\begin{aligned} \limsup_n \frac{1}{|V_n|} \log Z(\mathbb{G}_n, \mathbb{H}) - \log k &\leq \limsup_n \max_{(x, X) \in \Gamma_\delta} \left(\frac{1}{|V_n|} \log \mathbb{P}_{k, n} \left(B((x, X), \delta) \right) - \mathcal{E}_{\mathbb{H}}(x, X) \right) + K\delta \\ &= \max_{(x, X) \in \Gamma_\delta} \left(\limsup_n \frac{1}{|V_n|} \log \mathbb{P}_{k, n} \left(B((x, X), \delta) \right) - \mathcal{E}_{\mathbb{H}}(x, X) \right) + K\delta \\ &\leq \max_{(x, X) \in \Gamma_\delta} \left(- \inf_{(y, Y) \in B((x, X), \delta)} I_k(y, Y) + \mathcal{E}_{\mathbb{H}}(x, X) \right) + K\delta \\ &\leq \max_{(x, X) \in \Gamma_\delta} \left(- \inf_{(y, Y) \in B((x, X), \delta)} \left(I_k(y, Y) + \mathcal{E}_{\mathbb{H}}(y, Y) \right) \right) + 2K\delta \\ &\leq \max_{(x, X) \in \Gamma_\delta} \left(- \inf_{(y, Y) \in [0, D]^{(k+1) \times k}} \left(I_k(y, Y) + \mathcal{E}_{\mathbb{H}}(y, Y) \right) \right) + 2K\delta \\ &= - \inf_{(y, Y) \in [0, D]^{(k+1) \times k}} \left(I_k(y, Y) + \mathcal{E}_{\mathbb{H}}(y, Y) \right) + 2K\delta. \end{aligned}$$

Since this inequality holds for arbitrary δ , we obtain

$$\limsup_n \frac{1}{|V_n|} \log Z(\mathbb{G}_n, \mathbb{H}) \leq - \inf_{(x, X) \in [0, D]^{(k+1) \times k}} \left(\mathcal{E}_{\mathbb{H}}(x, X) - \mathcal{S}_k(x, X) \right).$$

We now establish the matching lower bound. Fix $\delta > 0$ and let (x^*, X^*) be the minimizer of $\mathcal{E}_{\mathbb{H}} - \mathcal{S}_k$. We have

$$\begin{aligned} Z(\mathbb{G}_n, \mathbb{H}) &\geq \sum_{\sigma \in \Sigma(x^*, X^*)} e^{-\mathcal{E}_{\mathbb{H}}(x(\sigma), X(\sigma))|V_n|} \\ &\geq k^{|V_n|} \mathbb{P}_{k, n} \left(B((x^*, X^*), \delta) \right) e^{-\mathcal{E}_{\mathbb{H}}(x^*, X^*)|V_n| - K\delta|V_n|}, \end{aligned}$$

and thus

$$\begin{aligned} \liminf_n \frac{1}{|V_n|} \log Z(\mathbb{G}_n, \mathbb{H}) - \log k &\geq \liminf_n \frac{1}{|V_n|} \log \mathbb{P}_{k, n} \left(B((x^*, X^*), \delta) \right) - \mathcal{E}_{\mathbb{H}}(x^*, X^*) - K\delta \\ &\geq \liminf_n \frac{1}{|V_n|} \log \mathbb{P}_{k, n} \left(B^o((x^*, X^*), \delta) \right) - \mathcal{E}_{\mathbb{H}}(x^*, X^*) - K\delta \\ &\geq - \inf_{(y, Y) \in B^o((x^*, X^*), \delta)} I_k((y, Y)) - \mathcal{E}_{\mathbb{H}}(x^*, X^*) - K\delta \\ &\geq -I_k(x^*, X^*) - \mathcal{E}_{\mathbb{H}}(x^*, X^*) - K\delta. \end{aligned}$$

Letting $\delta \rightarrow 0$ we obtain the result. \square

5.5 LD-convergence implies partition-convergence

We prove this result by identifying the limiting sets $\Sigma(k)$, $k \geq 1$. In order to do this, we recall that by Theorem 1 every LD-convergent sequence is k -LD-convergent for all k . We denote the corresponding functions by I_k .

Theorem 6. *Let \mathbb{G}_n be a sequence of graphs which is LD-convergent. Then \mathbb{G}_n is partition-convergent. Moreover, for each k , the sets $\Sigma_n(k)$ converge to*

$$\Sigma(k) = \{(x, X) \in [0, D]^{(k+1) \times k} : I_k(x, X) < \infty\}. \quad (36)$$

The intuition behind this result is as follows. $\Sigma(k)$ is the set of "approximately achievable" partitions (x, X) . If some partition is approximately realizable as factor graph \mathbb{G}_n/σ for some $\sigma : V_n \rightarrow [k]$, then there are exponentially many such realizations, since changing a small linear in $|V_n|$ number of assignments of σ does not change the factor graph significantly. Therefore, a uniform random partition σ has at least an inverse exponential probability of creating a factor graph \mathbb{G}_n/σ near (x, X) and, as a result, the LD rate associated with (x, X) is positive.

On the other hand if (x, X) is ϵ -away from any factor graph \mathbb{G}_n/σ , then the likelihood that a uniform random partition σ results in a factor graph close to (x, X) is zero, implying that the LD rate associated with (x, X) is infinite. We now formalize this intuition.

Proof. We define $\Sigma(k)$ as in (36) and prove that it is the limit of $\Sigma_n(k)$ as $n \rightarrow \infty$ for each k . We assume the contrary. This means that either there exist k and $\epsilon_0 > 0$ such that for infinitely many n , $\Sigma_n(k)$ is not a subset of $(\Sigma(k))^{\epsilon_0}$, or there exist k and $\epsilon_0 > 0$ such that for infinitely many n , $\Sigma(k)$ is not a subset of $(\Sigma_n(k))^{\epsilon_0}$.

We begin with the first assumption. Then, for infinitely many n we can find $(x_n, X_n) \in \Sigma_n(k)$ such that (x_n, X_n) has distance at least ϵ_0 from $\Sigma(k)$. Denote this subsequence by (x_{n_r}, X_{n_r}) and find any limit point (\bar{x}, \bar{X}) of this sequence. For notational simplicity we assume that in fact (x_{n_r}, X_{n_r}) converges to (\bar{x}, \bar{X}) as $r \rightarrow \infty$ (as opposed to taking a further subsequence of n_r). Then the distance between (\bar{x}, \bar{X}) and $\Sigma(k)$ is at least ϵ_0 as well. We now show that in fact $I_k(\bar{x}, \bar{X}) < \infty$, implying that in fact (\bar{x}, \bar{X}) belongs to $\Sigma(k)$, which is a contradiction.

To prove the claim we fix $\delta > 0$. We have that \mathbb{L}_∞ distance between (x_{n_r}, X_{n_r}) and (\bar{x}, \bar{X}) is at most δ for all large enough r . Since $(x_{n_r}, X_{n_r}) \in \Sigma_{n_r}(k)$, then for all r there exists some $\hat{\sigma}_{n_r} : V_{n_r} \rightarrow [k]$ such that (x_{n_r}, X_{n_r}) arises from the partition $\hat{\sigma}_{n_r}$. Namely, $x_{n_r} = x(\mathbb{G}_{n_r}/\hat{\sigma}_{n_r})$ and $X_{n_r} = X(\mathbb{G}_{n_r}/\hat{\sigma}_{n_r})$. Since a randomly uniformly chosen map $\sigma_{n_r} : V_{n_r} \rightarrow [k]$ coincides with $\hat{\sigma}_{n_r}$ with probability exactly k^{-n_r} , we conclude that

$$\mathbb{P}((x(\sigma_{n_r}), X(\sigma_{n_r})) \in B((\bar{x}, \bar{X}), \delta)) \geq k^{-n_r},$$

for all sufficiently large r . This implies

$$\begin{aligned} \limsup_n \frac{1}{n} \log \mathbb{P}((x(\sigma_n), X(\sigma_n)) \in B((\bar{x}, \bar{X}), \delta)) \\ \geq \limsup_r \frac{1}{n_r} \log \mathbb{P}((x(\sigma_{n_r}), X(\sigma_{n_r})) \in B((\bar{x}, \bar{X}), \delta)) \\ \geq -\log k. \end{aligned}$$

Applying property (6) we conclude that $I_k(\bar{x}, \bar{X}) \leq \log k < \infty$, and the claim is established.

Now assume that k and $\epsilon_0 > 0$ are such that for infinitely many n , $\Sigma(k)$ is not a subset of $(\Sigma_n(k))^{\epsilon_0}$. Thus there is a subsequence of points $(x_{n_r}, X_{n_r}) \in \Sigma(k)$ which are at least ϵ_0 away from $\Sigma_{n_r}(k)$ for each r . Let $(\bar{x}, \bar{X}) \in [0, D]^{(k+1) \times k}$ be any limit point of $(x_{n_r}, X_{n_r}), r \geq 1$. By lower semi-continuity we have that the set $\Sigma(k)$ is closed, and thus compact. Therefore, we also have $(\bar{x}, \bar{X}) \in \Sigma(k)$. Again, without loss of generality, we assume that in fact (x_{n_r}, X_{n_r}) converges to (\bar{x}, \bar{X}) . Fix any $\delta < \epsilon_0/2$. We have $(x_{n_r}, X_{n_r}) \in B((\bar{x}, \bar{X}), \delta)$ for all sufficiently large r . Then for all sufficiently large r we have that the distance from (\bar{x}, \bar{X}) to $\Sigma_{n_r}(k)$ is strictly larger than δ (since otherwise the distance from (x_{n_r}, X_{n_r}) to this set is less than ϵ_0). This means that for all sufficiently large r ,

$$\mathbb{P}((x(\sigma_{n_r}), X(\sigma_{n_r})) \in B((\bar{x}, \bar{X}), \delta)) = 0,$$

implying

$$\begin{aligned} \liminf_n \frac{1}{n} \log \mathbb{P}((x(\sigma_n), X(\sigma_n)) \in B((\bar{x}, \bar{X}), \delta)) \\ \leq \liminf_r \frac{1}{n_r} \log \mathbb{P}((x(\sigma_{n_r}), X(\sigma_{n_r})) \in B((\bar{x}, \bar{X}), \delta)) \\ = -\infty. \end{aligned}$$

Applying property (6) we conclude that $I_k(\bar{x}, \bar{X}) = \infty$, contradicting the fact that $(\bar{x}, \bar{X}) \in \Sigma(k)$. This concludes the proof. \square

6 Discussion

6.1 Right-convergence for dense vs. sparse graphs

It is instructive to compare our results from Section 5.4, in particular the expression for the limiting free energy from Theorem 5, to the corresponding results from the theory of convergent sequences for dense graphs.

For sequences \mathbb{G}_n of dense graphs, i.e., graphs with average degree proportional to the number of vertices, the weighted homomorphism numbers $\text{hom}(\mathbb{G}_n, \mathbb{H})$ typically either grow or decay exponentially in $|V(\mathbb{G}_n)|^2$, implying that the limit of an expression of the form (2) would be either ∞ or $-\infty$. To avoid this problem, we adopt the usual definition from statistical physics, where the free energy is defined by

$$f(\mathbb{G}_n, \mathbb{H}) = -\frac{1}{|V(\mathbb{G}_n)|} \log Z(\mathbb{G}_n, \mathbb{H}), \quad (37)$$

where $Z(\mathbb{G}_n, \mathbb{H})$ is the *partition function*²

$$Z(\mathbb{G}_n, \mathbb{H}) = \sum_{\sigma: V(\mathbb{G}_n) \rightarrow V(\mathbb{H})} \prod_{u \in V(\mathbb{G}_n)} \alpha_{\sigma(u)} \prod_{(u,v) \in E(\mathbb{G}_n)} (A_{\sigma(u), \sigma(v)})^{1/|V(\mathbb{G}_n)|},$$

with α and A denoting the vector and matrix of the node and edges weights of \mathbb{H} , respectively.

²Note that the only difference with respect to (4) is the exponent $1/|V(\mathbb{G}_n)|$, which ensures that $\log Z(\mathbb{G}_n, \mathbb{H})$ is of order $O(|V_n|)$.

It was shown in [BCL⁺12] that for a left-convergent sequence of dense graphs, the limit of (37) exists, and can be expressed in terms of the limiting graphon. We refer the reader to the literature [LS06] for the definition of the limiting graphon; for us, it is only important that it is a measurable, symmetric function $W : [0, 1]^2 \rightarrow [0, 1]$ which can be thought of as the limit of a step function representing the adjacency matrices of \mathbb{G}_n (see Corollary 3.9 of [BCL⁺08] for a precise version of this statement).

To express the limit of the free energies (37) in terms of W , we need some notation. Assume $|V(\mathbb{H})| = k$, and let Ω_k be the set of measurable functions $\rho = (\rho_1, \dots, \rho_k) : [0, 1] \rightarrow [0, 1]^k$ such that ρ_i is non-negative for all $i \in [k]$ and $\sum_i \rho_i(x) = 1$ for all $x \in [0, 1]$. Setting $x_i(\rho) = \int_0^1 \rho_i(x) dx$ and $X_{i,j}(\rho) = \int_0^1 \int_0^1 \rho_i(x) \rho_j(y) W(x, y) dx dy$, we define the energy and entropy of a “configuration” $\rho \in \Omega_k$ as

$$\mathcal{E}_{\mathbb{H}}(\rho) = - \sum_{1 \leq i < j \leq k} x_i(\rho) \log \alpha_i - \frac{1}{2} \sum_{1 \leq i, j \leq k} X_{i,j}(\rho) \log A_{i,j}$$

and

$$\mathcal{S}_k(\rho) = - \int_0^1 dx \sum_{1 \leq i \leq k} \rho_i(x) \log \rho_i(x),$$

respectively. Then

$$\lim_{n \rightarrow \infty} f(\mathbb{G}_n, \mathbb{H}) = \inf_{\rho \in \Omega_k} \left(\mathcal{E}_{\mathbb{H}}(\rho) - \mathcal{S}_k(\rho) \right).$$

Note the striking similarity to the expressions (33), (34) and (35). There is, however, one important difference: in the dense case, the limit object W describes an effective edge-density and enters into the definition of the energy, while in the sparse case, the limit object is a large-deviation rate and enters into the entropy. By contrast, the entropy in the dense case is trivial, and does not depend on the sequence \mathbb{G}_n , similar to the energy in the sparse case, which does not depend on \mathbb{G}_n .

As we will see in the next subsection, this points to the fact that right-convergence is a much richer concept for sparse graphs than for dense graphs. Indeed, for many interesting sequences of dense graphs, one can explicitly construct the limiting graphon W , reducing free energy calculations to simple variational problems. By contrast, we do not even know how to calculate the rate functions for the simplest random sparse graph, the Erdős-Renyi random graph. We discuss this, and how it is related to spin glasses, in the next section.

6.2 Conjectures and open questions

While Theorem 4 establishes a complete picture regarding the relationships between LD, left, right and partition-convergence, several questions remain open regarding colored-neighborhood-convergence. In particular, not only do we not know whether LD implies colored-neighborhood-convergence or vice versa, but we not even know whether colored-neighborhood-convergence implies right-convergence.

Since colored-neighborhood-convergence does not appear to take into account the counting measures involved in the definition of right-convergence, we conjecture that in fact colored-neighborhood-convergence *does not* imply right-convergence. Should this be the case it would also imply that colored-neighborhood-convergence does not imply LD-convergence, via Theorem 4.

At the same time we conjecture that the LD-convergence implies the colored-neighborhood-convergence. It is an interesting open problem to resolve these conjectures.

Perhaps the most important outstanding conjecture regarding the theory of convergent sparse graph sequences concerns convergence of random sparse graphs. Consider two natural and much studied random graphs sequences, the sequence of Erdős-Renyi random graphs $\mathbb{G}(n, c/n)$, and the sequence of random regular graphs $\mathbb{G}(n, \Delta)$. As usual, $\mathbb{G}(n, c/n)$ is the random subgraph of the complete graph on n nodes where each edge is kept independently with probability c/n , where c is an n -independent constant, and $\mathbb{G}(n, \Delta)$ is a Δ -regular graph on n nodes, chosen uniformly at random from the family of all n -node Δ -regular graphs. We ask whether these graph sequences are convergent with respect to any of the definitions of graph convergence discussed above³.

It is well known and easy to show that these graphs sequences are left-convergent, since the r -depth neighborhoods of a random uniformly chosen point can be described in the limit as $n \rightarrow \infty$ as the first r generations of an appropriate branching process. The validity for other notions of convergence, however, presents a serious challenge as here we touch on the rich subject of mathematical aspects of the spin glass theory. The reason is that right-convergence of sparse graph sequences is nothing but convergence of normalized partition functions of the associated Gibbs measures. Such convergence means the existence of the associated thermodynamic limit and this existence question is one of the outstanding open areas in the theory of spin glasses.

Substantial progress has been achieved recently due to the powerful interpolation method introduced by Guerra and Toninelli [GT02] in the context of the Sherrington-Kirkpatrick model, and further by Franz and Leone [FL03], Panchenko and Talagrand [PT04], Bayati, Gamarnik and Tetali [BGT10], Abbe and Montanari [AM10] and Gamarnik [Gam]. In particular, a broad class of weighted graphs \mathbb{H} is identified, with respect to which the random graph sequence $\mathbb{G}(n, c/n)$ is right-convergent w.h.p. (in some appropriate sense). Some extensions of these results are also known for random regular graph sequences $\mathbb{G}(n, \Delta)$. A large general class of such weighted graphs \mathbb{H} is described in [Gam] in terms of suitable positive semi-definiteness properties of the graph \mathbb{H} . Another class of graphs \mathbb{H} with respect to which convergence is known and in fact the limit can be computed is when \mathbb{H} corresponds to the ferromagnetic Ising model and some extensions [DM10],[DMS11]. We note that for the latter models, it suffices that the graph sequence is left-convergent to some locally tree-like graphs satisfying some mild additional properties.

Still, the question whether the two most natural random graph sequences $\mathbb{G}(n, c/n)$ and $\mathbb{G}(n, \Delta)$ are right-convergent (with respect to all \mathbb{H}) is still open. The same applies to all other notions of convergence discussed in this paper (except for left-convergence). We conjecture that the answer is yes for all of them. Thanks to Theorem 4 this would follow from the following conjecture.

Conjecture 1. *The random graph sequences $\mathbb{G}(n, c/n)$ and $\mathbb{G}(n, \Delta)$ are almost surely LD-convergent.*

The quantifier "almost sure" needs some elaboration. Here we consider the product $\prod_n \mathcal{G}_n$, where \mathcal{G}_n is the set of all simple graphs on n nodes. We equip this product space with the natural product probability measure, arising from having each \mathcal{G}_n induced with random graph $\mathbb{G}(n, c/n)$

³In contrast to $\mathbb{G}(n, \Delta)$, the graphs $\mathbb{G}(n, c/n)$ do not have uniformly bounded degrees, which means that most of the theorems proved in this paper do not hold. Nevertheless, we can still pose the question of convergence of these graphs with respect to various notions.

measure. Our conjecture states that, almost surely with respect to this product probability space \mathcal{G} , the graph sequence realization is LD-convergent.

Acknowledgements

We gratefully acknowledge fruitful discussions with Laci Lovász and wish to thank him for providing the key ideas for the counterexample in Subsection 5.3. The third author also wishes to thank Microsoft Research New England for providing exciting and hospitable environment during his visit in 2012.

References

- [AM10] E. Abbe and A. Montanari, *On the concentration of the number of solutions of random satisfiability formulas*, <http://arxiv.org/abs/1006.3786> (2010).
- [BCKL13] C. Borgs, J.T. Chayes, J. Kahn, and L. Lovász, *Left and right convergence of graphs with bounded degree*, *Random Struct. Algorithms* **42** (2013), no. 1, 1–28.
- [BCL⁺06] C. Borgs, J. T. Chayes, L. Lovász, V. Sos, and K. Vesztergombi, *Counting graph homomorphisms*, *Topics in Discrete Mathematics* (eds. M. Klazar, J. Kratochvil, M. Loebel, J. Matousek, R. Thomas, P.Valtr), Springer, 2006, pp. 315–371.
- [BCL10] C. Borgs, J. Chayes, and L. Lovász, *Moments of two-variable functions and the uniqueness of graph limits*, *Geometric And Functional Analysis* **19** (2010), no. 6, 1597–1619.
- [BCL⁺12] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi, *Convergent graph sequences II: Multiway cuts and statistical physics*, *Ann. of Math.* **176** (2012), 151–219.
- [BCL⁺08] ———, *Convergent graph sequences I: Subgraph frequencies, metric properties, and testing*, *Advances in Math.* **219** (2008), 1801–1851.
- [BCP97] A. Bertoni, P. Campadelli, and R. Posenato, *An upper bound for the maximum cut mean value*, *Graph-theoretic concepts in computer science (Lecture Notes in Computer Science, Vol. 1335)*, Springer, Berlin, 1997, pp. 78–84.
- [BGT10] M. Bayati, D. Gamarnik, and P. Tetali, *Combinatorial approach to the interpolation method and scaling limits in sparse random graphs*, *Proc. 42nd Ann. Symposium on the Theory of Computing (STOC)*, <http://arxiv.org/abs/0912.2444>, 2010.
- [Bil99] P. Billingsley, *Convergence of probability measures*, Wiley-Interscience publication, 1999.
- [BR11] B. Bollobás and O. Riordan, *Sparse graphs: metrics and random models*, *Random Structures and Algorithms* **39** (2011), 1–38.
- [BS01] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, *Electronic Journal of Probability* **23** (2001), 1–13.

- [CV11] S. Chatterjee and S.R.S. Varadhan, *The large deviation principle for the Erdős-Rényi random graph*, European Journal of Combinatorics **32** (2011), 1000 – 1017.
- [DJ07] P. Diaconis and S. Janson, *Graph limits and exchangeable random graphs*, Arxiv preprint arXiv:0712.2749 (2007).
- [DM10] Amir Dembo and Andrea Montanari, *Ising models on locally tree-like graphs*, Annals of Applied Probability **20** (2010), 565–592.
- [DMS11] A. Dembo, A. Montanari, and N. Sun, *Factor models on locally tree-like graphs*, Arxiv preprint arXiv:1110.4821 (2011).
- [DZ98] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer, 1998.
- [FL03] S. Franz and M. Leone, *Replica bounds for optimization problems and diluted spin systems*, Journal of Statistical Physics **111** (2003), no. 3/4, 535–564.
- [Gam] D. Gamarnik, *Right-convergence of sparse random graphs*, Preprint at <http://arxiv.org/abs/1202.3123>.
- [Geo88] H. O. Georgii, *Gibbs measures and phase transitions*, de Gruyter Studies in Mathematics 9, Walter de Gruyter & Co., Berlin, 1988.
- [GT02] F. Guerra and F.L. Toninelli, *The thermodynamic limit in mean field spin glass models*, Commun. Math. Phys. **230** (2002), 71–79.
- [HLS12] H. Hatami, L. Lovász, and B. Szegedy, *Limits of local-global convergent graph sequences*, Preprint at <http://arxiv.org/abs/1205.4356> (2012).
- [LS06] L. Lovász and B. Szegedy, *Limits of dense graph sequences*, Journal of Combinatorial Theory, Series B **96** (2006), 933–957.
- [PT04] D. Panchenko and M. Talagrand, *Bounds for diluted mean-fields spin glass models*, Probability Theory and Related Fields **130** (2004), 312–336.
- [Sim93] B. Simon, *The statistical mechanics of lattice gases, Vol. I*, Princeton Series in Physics, Princeton University Press, Princeton, NJ, 1993.