Large deviations of empirical neighborhood distribution in sparse random graphs

Charles Bordenave · Pietro Caputo

Abstract Consider the Erdős–Renyi random graph on n vertices where each edge is present independently with probability λ/n, with $\lambda > 0$ fixed. For large n, a typical random graph locally behaves like a Galton–Watson tree with Poisson offspring distribution with mean λ. Here, we study large deviations from this typical behavior within the framework of the local weak convergence of finite graph sequences. The associated rate function is expressed in terms of an entropy functional on unimodular measures and takes finite values only at measures supported on trees. We also establish large deviations for other commonly studied random graph ensembles such as the uniform random graph with given number of edges growing linearly with the number of vertices, or the uniform random graph with given degree sequence. To prove our results, we introduce a new configuration model which allows one to sample uniform random graphs with a given neighborhood distribution, provided the latter is supported on trees. We also introduce a new class of unimodular random trees, which generalizes the usual Galton Watson tree with given degree distribution to the case of neighborhoods of arbitrary finite depth. These generalized Galton Watson trees turn out to be useful in the analysis of unimodular random trees and may be considered to be of interest in their own right.

C. Bordenave (✉)
IMT UMR 5219 CNRS, Université Paul-Sabatier Toulouse III, Toulouse, France
e-mail: charles.bordenave@math.univ-toulouse.fr
URL: http://www.math.univ-toulouse.fr/~bordenave

P. Caputo
Dipartimento di Matematica, Università Roma Tre, Rome, Italy
e-mail: caputo@mat.uniroma3.it
URL: http://www.mat.uniroma3.it/users/caputo

🖉 Springer

1 Introduction and main results

Consider the Erdős–Renyi ensemble $\mathcal{G}(n, p)$, where a random graph is obtained from the vertex set $[n] = \{1, \ldots, n\}$ by adding each edge independently with probability p. In the sparse regime with $p = \lambda/n$, for a fixed $\lambda > 0$, it is well known that, for large n, a typical graph from $\mathcal{G}(n, p)$ locally looks like a Galton–Watson tree with Poisson offspring distribution with mean λ. In this work we study large deviations from this typical behavior. The problem is intimately related to the question: conditioned on having a certain neighborhood distribution, what does a typical element of $\mathcal{G}(n, p)$ locally look like? The same questions can be asked for other commonly studied random graph ensembles such as the uniform random graphs with fixed number of edges growing linearly with the number of vertices, or with given degree sequence. We formulate the problem within the theory of *local weak convergence* of graph sequences that was recently introduced by Benjamini and Schramm [4] and Aldous and Steele [2]. The associated local weak topology has now become a common tool for studying sparse graphs, see Aldous and Lyons [1] and Bollobàs and Riordan [8]. A surprising large variety of graph functionals are continuous for this topology. In Sect. 2 below, we will give more details on local weak convergence. In order to present our result, here we first introduce the main terminology.

1.1 Local weak convergence

A graph $G = (V, E)$, with V a countable set of vertices, is said to be locally finite if, for all $v \in V$, the degree of v in G is finite. A *rooted graph* (G, o) is a locally finite and connected graph $G = (V, E)$ with a distinguished vertex $o \in V$, called the root. For $t \geqslant 0$, we denote by $(G, o)_t$ the induced rooted graph with vertex set $\{u \in V : D(o, u) \leqslant t\}$, with $D(\cdot, \cdot)$ the natural graph distance. Two rooted graphs $(G_i, o_i) = (V_i, E_i, o_i)$, $i \in \{1, 2\}$, are *isomorphic* if there exists a bijection $\sigma : V_1 \to V_2$ such that $\sigma(o_1) = o_2$ and $\sigma(G_1) = G_2$, where σ acts on E_1 through $\sigma(\{u, v\}) = \{\sigma(u), \sigma(v)\}$. We will denote this equivalence relation by $(G_1, o_1) \simeq (G_2, o_2)$. An equivalence class of rooted graphs is often simply referred to as *unlabeled rooted graph*. We denote by $\mathcal{G}^*$ the set of (locally finite, connected) unlabeled rooted graphs. $\mathcal{T}^*$ will be the set of unlabeled rooted trees. To each unlabeled rooted graph $g \in \mathcal{G}^*$, we may associate a labeled rooted graph (G, o) with vertex set $V \subset \mathbb{Z}_+$, rooted at 0, in a canonical way; see e.g. [1]. For ease of notation, one sometimes identifies $g \in \mathcal{G}^*$ with its canonical rooted graph (G, o).

For $\gamma \in \mathcal{G}^*$ and $h \in \mathbb{N}$, we write γ_h for the truncation at h of the graph γ, namely the unlabeled rooted graph obtained by removing all vertices (together with the edges incident to them) that are at distance larger than h from the root. The *local topology* is the smallest topology such that for any $\gamma \in \mathcal{G}_*$ and $h \in \mathbb{N}$, the $\mathcal{G}^* \to \{0, 1\}$ function
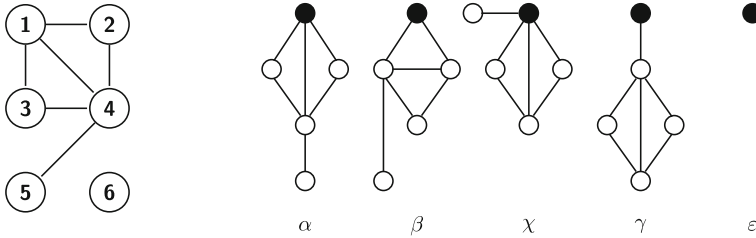
Fig. 1 Example of a graph G and its empirical neighborhood distribution. Here $U(G) = \frac{1}{6}(\delta_\alpha + 2\delta_\beta + \delta_\chi + \delta_\gamma + \delta_\varepsilon)$, where $\alpha, \beta, \chi, \gamma, \varepsilon \in \mathcal{G}^*$ are the unlabeled rooted graphs depicted above (the black vertex is the root), with $[G, 1] = \alpha$, $[G, 2] = [G, 3] = \beta$, $[G, 4] = \chi$, $[G, 5] = \gamma$, $[G, 6] = \varepsilon$

$f(g) = \mathbf{1}(g_h = \gamma_h)$ is continuous. Equivalently, a sequence $g_n \in \mathcal{G}^*$ converges locally to $g \in \mathcal{G}^*$ iff for all $h \in \mathbb{N}$ there exists $n_0(h)$ such that $(g_n)_h = g_h$ whenever $n \geqslant n_0(h)$. This topology is metrizable and the space $\mathcal{G}^*$ is separable and complete [1]. The space of probability measures on $\mathcal{G}^*$, denoted $\mathcal{P}(\mathcal{G}^*)$, is equipped with the topology of weak convergence. We often write $\rho_n \rightsquigarrow \rho$ to indicate that a sequence $\rho_n \in \mathcal{P}(\mathcal{G}^*)$ converges weakly to $\rho \in \mathcal{P}(\mathcal{G}^*)$.

For a finite graph $G = (V, E)$ and $v \in V$, one writes $G(v)$ for the connected component of G at v. The *empirical neighborhood distribution* $U(G)$ of G is the law of the equivalence class of the rooted graph $(G(o), o)$ where the root o is sampled uniformly at random from V, i.e. $U(G) \in \mathcal{P}(\mathcal{G}^*)$ is defined by

$$U(G) = \frac{1}{|V|} \sum_{v \in V} \delta_{[G,v]}, \tag{1}$$

where $[G, v] \in \mathcal{G}^*$ stands for the equivalence class of $(G(v), v)$ and δ_g is the Dirac mass at $g \in \mathcal{G}^*$; see Fig. 1 for an example. If $\{G_n\}$ is a sequence of finite graphs, we shall say that G_n has *local weak limit* $\rho \in \mathcal{P}(\mathcal{G}^*)$ if $U(G_n)$ converges to ρ in $\mathcal{P}(\mathcal{G}^*)$ as $n \to \infty$. A measure $\rho \in \mathcal{P}(\mathcal{G}^*)$ is called *sofic* if there exists a sequence of finite graphs $\{G_n\}$ whose local weak limit is ρ. In other words, the set of sofic measures is the closure of the set $\{U(G_n) : G_n \text{ finite graph}\}$. An example is the Dirac mass at the infinite regular tree with degree $d \in \mathbb{N}$, which is almost surely the local weak limit of a sequence of uniformly sampled random d-regular graphs on n vertices [31]. Another example is the law of the Galton–Watson tree with Poisson offspring distribution with mean $\lambda > 0$, which is almost surely the local weak limit of a sequence of random graphs sampled from $\mathcal{G}(n, p)$ when $p = \lambda/n$. Sofic measures form a closed subset of $\mathcal{P}(\mathcal{G}^*)$.

Sofic measures share a stationarity property called *unimodularity* [1]. To define the latter, consider the set $\mathcal{G}^{**}$ of unlabeled graphs with two distinguished roots, obtained as the set of equivalence classes of locally finite connected graphs with two distinguished vertices (G, u, v). The notion of local topology extends naturally to $\mathcal{G}^{**}$. A function f on $\mathcal{G}^{**}$ can be extended to a function on connected graphs with two distinguished roots (G, u, v) through the isomorphism classes. Then, a measure $\rho \in \mathcal{P}(\mathcal{G}^*)$ is called *unimodular* if for any Borel measurable function $f : \mathcal{G}^{**} \to \mathbb{R}_+$, we have

$$\mathbb{E}_\rho \sum_{v \in V} f(G, o, v) = \mathbb{E}_\rho \sum_{v \in V} f(G, v, o), \tag{2}$$

where (G, o) is the canonical rooted graph whose equivalence class $g \in \mathcal{G}^*$ has law ρ. It is not hard to check that if G is a finite graph then its neighborhood distribution $U(G)$ is unimodular. In particular, all sofic measures are unimodular. The converse is open; see [1]. We denote by $\mathcal{P}_u(\mathcal{G}^*)$ the set of unimodular probability measures. Similarly, we write $\mathcal{P}_u(\mathcal{T}^*)$ for unimodular probability measures supported by trees.

1.2 Unimodular Galton–Watson trees with given neighborhood

We now introduce a family of unimodular measures that will play a key role in what follows. As we will see, this is the natural generalization of the usual Galton–Watson trees with given degree distribution to the case of neighborhoods of arbitrary depth $h \in \mathbb{N}$. These measures will be shown to be sofic, and this fact can be used to give an alternative proof of the Bowen–Elek theorem [3,13,19] asserting that all $\rho \in \mathcal{P}_u(\mathcal{T}^*)$ are sofic, see Corollary 1.5 below.

Fix $h \in \mathbb{N}$, and recall that g_h denotes the truncation at depth h of $g \in \mathcal{G}^*$. Call $\mathcal{G}_h^*$ the set of unlabeled rooted graphs with depth h, i.e. the set of $g \in \mathcal{G}^*$ such that $g_h = g$. Similarly, call $\mathcal{T}_h^*$ the set of unlabeled rooted trees $t \in \mathcal{T}^*$ such that $t_h = t$. Given $\rho \in \mathcal{P}(\mathcal{G}^*)$, we write $\rho_h \in \mathcal{P}(\mathcal{G}_h^*)$ for the h-neighborhood marginal of ρ, i.e. the law of g_h when g has law ρ. Notice that if $t \in \mathcal{T}^*$ and $h = 1$, then t_h is simply the number of children of the root. In particular, $\mathcal{T}_1^*$ can be identified with $\mathbb{Z}_+$. When $h = 0$, it is understood that $\mathcal{G}_0^*$ contains only the trivial graph consisting of a single isolated vertex (the root), so that $|\mathcal{G}_0^*| = 1$.

If $G = (V, E)$ is a graph and $\{u, v\} \in E$ then define $G(u, v)$ as the rooted graph $(G'(v), v)$, where $G' = (V, E\backslash\{u, v\})$, i.e. $G(u, v)$ is the rooted graph obtained from G by removing the edge $\{u, v\}$ and taking the connected component at the root v. Next, given a rooted graph (G, o), and $g, g' \in \mathcal{G}_{h-1}^*$, define

$$E_h(g, g') = \left|\left\{ v \overset{G}{\sim} o : \ G(o, v)_{h-1} \simeq g, \ G(v, o)_{h-1} \simeq g' \right\}\right|. \qquad (3)$$

The notation $v \overset{G}{\sim} u$ indicates that the vertex v is a neighbor of u in G. Thus, $E_h(g, g')$ is the number of neighbors of the root in (G, o) which have the given patterns $G(o, v)_{h-1} \simeq g$ and $G(v, o)_{h-1} \simeq g'$. Notice that if $h = 1$, then necessarily $g, g' = o$ and $E_1(o, o) = \deg_G(o)$ is simply the degree of the root.

As an example, consider the the rooted graph α from Fig. 1. Fix $h = 2$, and call g_1, g_2 the elements of $\mathcal{G}_{h-1}^*$ consisting respectively of a rooted single edge and a rooted triangle. Then one has $E_h(g_1, g_2) = 2$ and $E_h(g_2, g_1) = 0$. Similarly, if the reference graph is β from Fig. 1, then $E_h(g_1, g_2) = 0$ while $E_h(g_2, g_1) = 1$.

We call a measure $P \in \mathcal{P}(\mathcal{G}_h^*)$ *admissible* if $\mathbb{E}_P \deg_G(o) < \infty$ and for all $g, g' \in \mathcal{G}_{h-1}^*$,

$$e_P(g, g') = e_P(g', g),$$

where

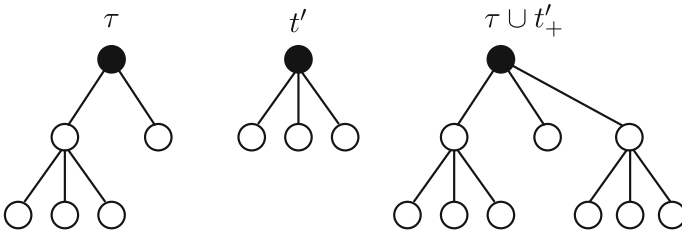$$e_P(g, g') := \mathbb{E}_P E_h(g, g'). \qquad (4)$$

Fig. 2 In this example one has $h = 2$ and $1 + |\{v \overset{\tau}{\sim} o : \tau(o, v) \simeq t'\}| = 2$. Notice that if $t = \tau_1$ denotes the rooted tree with two leaves, then in the rooted tree $\tau \cup t'_+$ one has $E_h(t', t) = 2$ according to the definition (3)

Here it is understood that (G, o) represents the canonical rooted graph whose equivalence class in $\mathcal{G}_h^*$ has law P. By applying the definition of unimodularity (2) to the function

$$f(G, u, v) = \mathbf{1}\big(v \overset{G}{\sim} u\big)\mathbf{1}\big(G(u, v)_{h-1} \simeq g; G(v, u)_{h-1} \simeq g'\big),$$

it is not hard to check that if ρ is unimodular and $\mathbb{E}_\rho \deg_G(o) < \infty$ then $\rho_h \in \mathcal{P}(\mathcal{G}_h^*)$ is admissible. In particular, for any finite graph G, the neighborhood distribution $U(G)_h$ truncated at depth h is admissible. Remark that, when $h = 1$, since $|\mathcal{G}_0^*| = 1$, all $P \in \mathcal{P}(\mathcal{T}_1^*) = \mathcal{P}(\mathbb{Z}_+)$ with finite mean are admissible.

We now define the measures $\mathrm{UGW}_h(P) \in \mathcal{P}(\mathcal{T}^*)$; see also Sect. 3 below for more details. Fix $P \in \mathcal{P}(\mathcal{T}_h^*)$ admissible. The probability $\mathrm{UGW}_h(P) \in \mathcal{P}(\mathcal{T}^*)$ is the law of the equivalence class of the random rooted tree (T, o) defined below.

For $t, t' \in \mathcal{T}_{h-1}^*$ such that $e_P(t, t') \neq 0$ define, for all $\tau \in \mathcal{T}_h^*$,

$$\widehat{P}_{t,t'}(\tau) = P(\tau \cup t'_+)\Big(1 + \big|\{v \overset{\tau}{\sim} o : \tau(o, v) \simeq t'\}\big|\Big)\frac{\mathbf{1}(\tau_{h-1} = t)}{e_P(t, t')}, \qquad (5)$$

where $\tau \cup t'_+$ denotes the tree obtained from τ by adding a new neighbor of the root whose rooted subtree is t'; see Fig. 2 for an example. The subtree $\tau(o, v)$ is defined before Eq. (3) with the graph G replaced by τ.

It can be checked that $\widehat{P}_{t,t'}$ is a probability, i.e. $\widehat{P}_{t,t'} \in \mathcal{P}(\mathcal{T}_h^*)$; see Sect. 3. We may now define the random rooted tree (T, o). First, $(T, o)_h$ is sampled according to P. Next, for each vertex v in the first generation of $(T, o)_h$, consider the subtree $t = T(o, v)_{h-1}$ with depth $h - 1$ rooted at v obtained by removing the edge $\{o, v\}$ and retaining the connected component up to distance $h - 1$ from v. We add a layer to t by replacing t with a new tree τ with depth h that coincides with t in the first $h - 1$ generations. The new tree τ is sampled according to $\widehat{P}_{t,t'}$ where t is as above while t' denotes the subtree $T(v, o)_{h-1}$ rooted at o obtained from $(T, o)_h$ by removing the edge $\{o, v\}$ and retaining the connected component up to distance $h - 1$ from o. This operation is repeated for each v in the first generation independently. After this step, we have overall added one layer to $(T, o)_h$, and thus we have sampled $(T, o)_{h+1}$.

We now proceed recursively, layer by layer, to obtain a sample of the full tree (T, o). Formally, this construction can be stated as follows. If u is the parent of v, we say

that v has *type* (t, t'), where $t, t' \in \mathcal{T}^*_{h-1}$, if $T(u, v)_{h-1} \simeq t$ and $T(v, u)_{h-1} \simeq t'$. The subtrees $T(u, v)$, and $T(v, u)$ are defined before Eq. (3) with G replaced by T. Denote by $1, \ldots, d$, with $d = \deg_T(o)$ the neighbors of the root in the canonical representation of the random variable with law P. Given $(T, o)_h$, the subtrees $T(o, v)$, $1 \leqslant v \leqslant d$, are independent random variables and, given that v has type (t, t'), then $T(o, v)_h$ has distribution $\widehat{P}_{t,t'}$. Once $T(o, v)_h$ is sampled, the type of a child v' of v is determined using only $T(o, v)_h$ and $T(v, o)_{h-2}$. For each child v' of v we sample the subtree $T(v, v')_h$ independently according to $\widehat{P}_{t,t'}$ where (t, t') is the type of v' and so on, recursively. This defines our random rooted tree (T, o).

If $h = 1$, then there is only one type possible and $\mathrm{UGW}_1(P)$ is the unimodular Galton–Watson tree with degree distribution $P \in \mathcal{P}(\mathbb{Z}_+)$, where the number d of children of the root is sampled according to P, and conditionally on d, the subtrees of the children of the root are independent Galton–Watson trees with offspring distribution given by the size-biased law $\widehat{P}$:

$$\widehat{P}(k) = \frac{(k+1)P(k+1)}{\sum_{\ell=1}^{\infty} \ell P(\ell)}. \tag{6}$$

If $P = \mathrm{Poi}(\lambda)$ is the Poisson distribution with mean λ, then $\widehat{P} = P$ and $\mathrm{UGW}_1(P)$ is the standard Galton–Watson tree with mean degree λ.

The following proposition summarizes the main properties of the measures $\mathrm{UGW}_h(P)$ for generic $h \in \mathbb{N}$ and $P \in \mathcal{P}(\mathcal{T}^*_h)$ admissible.

Proposition 1.1 *Fix $h \in \mathbb{N}$ and $P \in \mathcal{P}(\mathcal{T}^*_h)$ admissible. The measure $\mathrm{UGW}_h(P)$ is unimodular. Moreover, the following consistency relation is satisfied: for any $k \geqslant h$, $(\mathrm{UGW}_h(P))_k \in \mathcal{P}(\mathcal{T}^*_k)$ is admissible and*

$$\mathrm{UGW}_h(P) = \mathrm{UGW}_k((\mathrm{UGW}_h(P))_k).$$

1.3 Entropy of a measure $\rho \in \mathcal{P}(\mathcal{G}^*)$

It is convenient to work with uniformly distributed random graphs with a given number of edges. For any $n, m \in \mathbb{N}$, let $\mathcal{G}_{n,m}$ be the set of graphs on $V = [n]$ with $|E| = m$ edges. Fix $d > 0$, and a sequence $m = m(n)$ such that $m/n \to d/2$, as $n \to \infty$. Since

$$|\mathcal{G}_{n,m}| = \binom{n(n-1)/2}{m},$$

an application of Stirling's formula shows that

$$\log |\mathcal{G}_{n,m}| = m \log n + s(d)\, n + o(n), \qquad s(d) := \frac{d}{2} - \frac{d}{2} \log d. \tag{7}$$

If $\rho \in \mathcal{P}(\mathcal{G}^*)$, define

$$\mathcal{G}_{n,m}(\rho, \varepsilon) = \{ G \in \mathcal{G}_{n,m} : U(G) \in B(\rho, \varepsilon) \},$$

where $B(\rho, \varepsilon)$ denotes the open ball with radius ε around ρ with respect to the Lévy metric on $\mathcal{P}(\mathcal{G}^*)$. For $\varepsilon > 0$, define

$$\overline{\Sigma}(\rho, \varepsilon) = \limsup_{n \to \infty} \frac{\log |\mathcal{G}_{n,m}(\rho, \varepsilon)| - m \log n}{n}.$$

Since $\varepsilon \mapsto \overline{\Sigma}(\rho, \varepsilon)$ is non-decreasing, one defines

$$\overline{\Sigma}(\rho) = \lim_{\varepsilon \to 0} \downarrow \overline{\Sigma}(\rho, \varepsilon).$$

The extended real numbers $\underline{\Sigma}(\rho, \varepsilon)$ and $\underline{\Sigma}(\rho)$ are defined as above, with lim sup replaced by lim inf. If ρ is such that $\underline{\Sigma}(\rho) = \overline{\Sigma}(\rho)$, we set $\Sigma(\rho) := \overline{\Sigma}(\rho) = \underline{\Sigma}(\rho)$. The number $\Sigma(\rho)$ can be interpreted, up to an overall constant, as a microcanonical entropy associated to the state ρ. From (7), one has that $\Sigma(\rho) \in [-\infty, s(d)]$, whenever it is well defined.

Theorem 1.2 *Fix $d > 0$ and choose a sequence $m = m(n)$ such that $m/n \to d/2$. For any $\rho \in \mathcal{P}(\mathcal{G}^*)$, the entropy $\Sigma(\rho) \in [-\infty, s(d)]$ is well defined, it is upper semi-continuous, and it does not depend on the choice of the sequence $m(n)$. Moreover, $\Sigma(\rho) = -\infty$ if at least one of the following is satisfied:*

  (i) *ρ is not unimodular*
 (ii) *ρ is not supported on rooted trees.*
(iii) *$\mathbb{E}_\rho \deg_G(o) \neq d$.*

Notice that the definition of $\Sigma(\rho)$ depends on the parameter d. For simplicity, we do not write explicitly this dependence. In view of Theorem 1.2(iii), to avoid trivialities, unless otherwise stated, $\Sigma(\rho)$ will refer to the value at $d = \mathbb{E}_\rho \deg_G(o)$ (provided that the latter is finite). The next theorem computes the actual value of $\Sigma(\rho)$ for unimodular Galton–Watson trees and gives an expression for $\Sigma(\rho)$ for all $\rho \in \mathcal{P}_u(\mathcal{T}^*)$. Moreover, it shows that unimodular Galton–Watson trees maximize entropy under a h-neighborhood marginal constraint.

Let us introduce some additional notation. For any $P \in \mathcal{P}(\mathcal{T}_h^*)$, define the Shannon entropy

$$H(P) = -\sum_{t \in \mathcal{T}_h^*} P(t) \log P(t).$$

For $h \in \mathbb{N}$, call $\mathcal{P}_h$ the set of all $P \in \mathcal{P}(\mathcal{T}_h^*)$, with P admissible such that $H(P) < \infty$ and $\mathbb{E}_P[\deg_T(o) \log(\deg_T(o))] < \infty$.[1] For $P \in \mathcal{P}_h$, let π_P denote the probability on $\mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*$ defined by

$$\pi_P(s, s') = \frac{1}{d} e_P(s, s'), \qquad (s, s') \in \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*,$$

---

[1] We shall actually see with Lemma 5.10 below that $P \in \mathcal{P}_h$ is equivalent to P admissible and $\mathbb{E}_P[\deg_T(o) \log(\deg_T(o))] < \infty$.

where $d = \mathbb{E}_P \deg_G(o)$, $e_P(s, s') = \mathbb{E}_P E_h(s, s')$, and $E_h(s, s')$ is defined in (3). We write $H(\pi_P)$ for the Shannon entropy of π_P:

$$H(\pi_P) = - \sum_{(t,t') \in \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*} \pi_P(t, t') \log \pi_P(t, t').$$

Theorem 1.3 *Fix $h \in \mathbb{N}$. The expression*

$$J_h(P) = -s(d) + H(P) - \frac{d}{2} H(\pi_P) - \sum_{(s,s') \in \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*} \mathbb{E}_P \log(E_h(s, s')!), \quad (8)$$

defines a function $J_h : \mathcal{P}_h \mapsto [-\infty, s(d)]$, satisfying

$$\Sigma(\mathrm{UGW}_h(P)) = J_h(P),$$

for all $P \in \mathcal{P}_h$. Define $\overline{J}_h : \mathcal{P}(\mathcal{T}_h^*) \mapsto [-\infty, s(d)]$ by $\overline{J}_h(P) = J_h(P)$ if $P \in \mathcal{P}_h$, and $\overline{J}_h(P) = -\infty$ if $P \notin \mathcal{P}_h$. If $\rho \in \mathcal{P}_u(\mathcal{T}^*)$, then for all $h \in \mathbb{N}$,

$$\Sigma(\rho) \leqslant \overline{J}_h(\rho_h), \quad (9)$$

and, if ρ_1 has finite support, the inequality is strict unless $\rho = \mathrm{UGW}_h(\rho_h)$. Finally, for any $\rho \in \mathcal{P}_u(\mathcal{T}^*)$, $\overline{J}_h(\rho_h)$ is non-increasing in $h \in \mathbb{N}$, and

$$\Sigma(\rho) = \lim_{h \to \infty} \downarrow \overline{J}_h(\rho_h). \quad (10)$$

In Remark 5.13 below we provide an alternative expression for $J_h(P)$ in terms of relative entropies. Specializing to the case $h = 1$, we obtain the following corollary of Theorem 1.3.

Corollary 1.4 *If $P \in \mathcal{P}(\mathbb{Z}_+)$ has mean d, then*

$$\Sigma(\mathrm{UGW}_1(P)) = s(d) - H(P \,|\, \mathrm{Poi}(d)),$$

where $\mathrm{Poi}(d)$ stands for Poisson distribution with mean d, and $H(\cdot \,|\, \cdot)$ is the relative entropy.

In particular, the standard Galton–Watson tree $\rho = \mathrm{UGW}_1(\mathrm{Poi}(d))$ maximizes the entropy $\Sigma(\rho)$ among all measures ρ with mean degree d.

As a byproduct of our analysis, we will also obtain an alternative proof of the Bowen–Elek Theorem [3,13,19].

Corollary 1.5 *If $\rho \in \mathcal{P}_u(\mathcal{T}^*)$, then ρ is sofic.*

We observe finally that, from its definition, the map $\Sigma : \rho \mapsto \Sigma(\rho)$ is easily seen to be upper semi-continuous for the local weak topology (see Lemma 5.3). In Proposition 5.14 below, we will however prove that Σ fails to be continuous at any $\rho = \mathrm{UGW}_1(P)$ whenever $P \in \mathcal{P}(\mathbb{Z}_+)$ has finite support and satisfies $P(0) = P(1) = 0$, $P(2) < 1$.

1.4 Large deviations of uniform graphs with given degrees

Given a vector $\mathbf{d} \in \mathbb{Z}_+^n$, let $\mathcal{G}(\mathbf{d})$ denote the set of graphs $G = ([n], E)$ such that $\mathbf{d}$ is the degree sequence of G, i.e. if $\mathbf{d} = (d(1), \ldots, d(n))$, then for all $v \in [n]$, $\deg_G(v) = d(v)$. Consider a sequence $\mathbf{d}^{(n)}$, $n \in \mathbb{N}$, of degree vectors $(d^{(n)}(1), \ldots, d^{(n)}(n))$ such that, for some fixed $\theta \in \mathbb{N}$, and $P \in \mathcal{P}(\mathbb{Z}_+)$:

(C1) $\sum_{v=1}^n d^{(n)}(v)$ is even,
(C2) $\max_{1 \leqslant v \leqslant n} d^{(n)}(v) \leqslant \theta$,
(C3) $\frac{1}{n} \sum_{v \in [n]} \delta_{d^{(n)}(v)} \rightsquigarrow P$,

where $\rightsquigarrow$ denotes weak convergence in $\mathcal{P}(\mathbb{Z}_+)$. A consequence of Erdős and Gallai [21] is that if (C1)–(C3) above are satisfied, then $\mathcal{G}(\mathbf{d}^{(n)})$ is not empty for all n large enough. We shall consider a random graph G_n sampled uniformly from $\mathcal{G}(\mathbf{d}^{(n)})$. Models of this type are well known in the random graph literature; see e.g. Molloy and Reed [26]. In particular, it is a folklore fact that almost surely the neighborhood distribution $U(G_n)$ defined in (1) is weakly convergent to $\mathrm{UGW}_1(P)$; see also Theorem 4.8 below for a more general statement. One of our main results concerns the large deviations of $U(G_n)$. Here and below whenever we say that $U(G_n)$ satisfies the large deviation principle (LDP) in $\mathcal{P}(\mathcal{G}^*)$ with speed n and good rate function I, we mean that the function $I : \mathcal{P}(\mathcal{G}^*) \mapsto [0, \infty]$ is lower semi-continuous with compact level sets, and for every Borel set $B \subset \mathcal{P}(\mathcal{G}^*)$

$$-\inf_{\rho \in B^\circ} I(\rho) \leqslant \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(U(G_n) \in B\right) \leqslant \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(U(G_n) \in B\right)$$
$$\leqslant -\inf_{\rho \in \overline{B}} I(\rho), \tag{11}$$

where B° denotes the interior of B and $\overline{B}$ denotes the closure of B.

Theorem 1.6 *Let $\mathbf{d}^{(n)}$ be a sequence satisfying conditions (C1)–(C3) above. Let G_n be uniformly distributed on $\mathcal{G}(\mathbf{d}^{(n)})$. Then $U(G_n)$ satisfies the LDP in $\mathcal{P}(\mathcal{G}^*)$ with speed n and good rate function*

$$I(\rho) = \begin{cases} \Sigma(\mathrm{UGW}_1(P)) - \Sigma(\rho) & \text{if } \rho_1 = P, \\ \infty & \text{otherwise.} \end{cases}$$

It follows from Theorem 1.3 that for any integer $h \geqslant 1$, and $Q \in \mathcal{P}_h$ with $Q_1 = P$, then

$$\min\{I(\rho) : \rho_h = Q\} = J_1(P) - J_h(Q),$$

and the minimum is uniquely attained for $\rho = \mathrm{UGW}_h(Q)$. This allows one to compute large deviations of neighborhood measures $U(G_n)_h$ explicitly in terms of the function J_h.

On the other hand, consider the special case of d-regular graphs, where $\mathbf{d}^{(n)}$ is the constant vector $(d, \ldots, d)$, and $P = \delta_d$, for some fixed $d \in \mathbb{N}$. To have $\Sigma(\rho) > -\infty$,

ρ must be supported on trees by Theorem 1.2, and because of the constant degree constraint we find that the only $\rho \in \mathcal{P}(\mathcal{G}^*)$ such that $I(\rho) < \infty$ is the Dirac mass at the infinite rooted d-regular tree, which coincides with $\mathrm{UGW}_1(P)$, where the rate function is zero. Thus, for the d-regular random graph $I(\rho)$ is either zero or infinite, and one should look at faster speed than n here for non trivial large deviations.

We note finally that Theorem 1.6 establishes a large deviations principle with speed n. Other interesting large deviation events occur at higher speed. For example, for the proportion of vertices in a triangle in G_n, the speed would be $n \log n$.

1.5 Large deviations of Erdős–Rényi graphs

Next, we describe our main results for sparse Erdős–Rényi graphs such as the uniform random graph from $\mathcal{G}_{n,m}$, with $m \sim nd/2$ and the $\mathcal{G}(n, p)$ where each edge is independently present with probability $p = d/n$. It is well known that, in both cases, with probability one, $U(G_n)$ converges weakly to the standard Galton–Watson tree with mean degree d, i.e. $\rho = \mathrm{UGW}_1(\mathrm{Poi}(d))$, which by Corollary 1.4 satisfies $\Sigma(\rho) = s(d)$.

Theorem 1.7 *Fix $d > 0$ and a sequence $m = m(n)$ such that $m/n \to d/2$, as $n \to \infty$. Let G_n be uniformly distributed in $\mathcal{G}_{n,m}$. Then $U(G_n)$ satisfies the LDP in $\mathcal{P}(\mathcal{G}^*)$ with speed n and good rate function*

$$I(\rho) = \begin{cases} s(d) - \Sigma(\rho) & \text{if } \mathbb{E}_\rho \deg_G(o) = d, \\ \infty & \text{otherwise.} \end{cases} \tag{12}$$

Theorem 1.8 *Fix $\lambda > 0$ and take G_n with law $\mathcal{G}(n, \lambda/n)$. Then $U(G_n)$ satisfies the LDP in $\mathcal{P}(\mathcal{G}^*)$ with speed n and good rate function*

$$I(\rho) = \frac{\lambda}{2} - \frac{d}{2} \log \lambda - \Sigma(\rho), \tag{13}$$

where $d := \mathbb{E}_\rho \deg_G(o)$, with the convention that if $d = 0$ then $\Sigma(\rho) = s(0) = 0$.

In the special case of 1-neighborhoods, Theorem 1.7, Theorem 1.8 and Corollary 1.4 allow us to prove the following results. Let $u(G_n) \in \mathcal{P}(\mathbb{Z}_+)$ denote the empirical distribution of the degree: $u(G_n) = \frac{1}{n} \sum_{i=1}^n \delta_{\deg_{G_n}(i)}$.

Corollary 1.9 *Fix $d > 0$, a sequence $m = m(n)$ such that $m/n \to d/2$, and let G_n be uniformly distributed in $\mathcal{G}_{n,m}$. Then $u(G_n)$ satisfies the LDP in $\mathcal{P}(\mathbb{Z}_+)$ with good rate function*

$$K(P) = \begin{cases} H(P \mid \mathrm{Poi}(d)) & \text{if } \sum_k k P(k) = d, \\ \infty & \text{otherwise.} \end{cases}$$

Corollary 1.10 *Fix $\lambda > 0$ and take G_n with law $\mathcal{G}(n, \lambda/n)$. Then $u(G_n)$ satisfies the LDP in $\mathcal{P}(\mathbb{Z}_+)$ with speed n and good rate function*

$$
K(P) = \begin{cases} \frac{\lambda - d}{2} - \frac{d}{2} \log \frac{\lambda}{d} + H(P \mid \mathrm{Poi}(d)) & \text{if } d := \sum_k k P(k) < \infty \\ \infty & \text{otherwise.} \end{cases}
$$

1.6 Plan and methods

The proof of the main results discussed above is organized as follows. In Sect. 2 we review some basic facts about local weak convergence in the context of multi-graphs. We also establish a compactness criterion which parallels recent results of Benjamini, Lyons and Schramm [3]. In Sect. 3 we introduce the unimodular Galton Watson trees with given h-neighborhood distribution and prove the properties stated in Proposition 1.1. In Sect. 5 we prove our main results concerning the entropy $\Sigma(\rho)$, cf. Theorem 1.2 and Theorem 1.3. These are crucially based on the possibility of counting asymptotically the number of graphs in $\mathcal{G}_{n,m}$ which have a certain h-neighborhood distribution. To compute such things, we introduce what we call a *generalized configuration model*.

The standard configuration model, introduced in Bollobas [6], allows one to compute asymptotically the number of graphs with a given degree sequence. Since here we want to uncover the h-neighborhood of a vertex and not only its degree, we need to generalize the usual construction. To keep track of the h-neighborhood structure, we introduce directed multigraphs with colored edges and analyze the associated configuration model; see Sect. 4. This will allow us to sample a random graph with a given sequence of h-neighborhoods, as long as these neighborhoods are rooted trees. As an application, we prove Corollary 1.5 at the end of Sect. 4. It seems to us that this new configuration model may turn out to be a natural tool in other applications as well. Finally, Sect. 6 is devoted to the proof of large deviation principles in the classical random graphs ensembles. We stress that our methods allow in principle a much greater generality, since one could establish large deviation estimates for random graphs that are uniformly sampled from the class of all graphs with a given h-neighborhood distribution and not only with given degree sequences; see Remark 6.1.

1.7 Related work

Large deviations in random graphs is a rapidly growing topic. For dense graphs, e.g. $\mathcal{G}(n, p)$ with fixed $p \in (0, 1)$, a thorough treatment has been given recently by Chatterjee and Varadhan [14], in the framework of the cut topology introduced by Lovász and Szegedy [24], see also Borgs, Chayes, Lovász, Sós and Vesztergombi [10,11]. In the sparse regime, only a few partial results are known. O'Connell [27], Biskup et al. [5] and Puhalskii [28] have proven large deviation asymptotics for the connectivity and for the size of the connected components. Large deviations for degree sequences of Erdős–Rényi graphs has been studied in Doku-Amponsah and Mörters [17] and Boucheron et al. [12, Theorem 7.1]. Closer to our approach, large deviations in the local weak topology were obtained for critical multi-type Galton–Watson trees

by Dembo et al. [15]. Finally, large deviations for other models of statistical physics on Erdős–Rényi graphs have been considered in Rivoire [29] and Engel et al. [20].

As far as we know, this is the first time that large deviations of the neighborhood distribution are addressed in a systematic way. While our approach does not cover results on connectivity and the size of connected components such as [27], it does yield a simplification of some of the existing arguments concerning the large deviations for degree sequences. We point out that our Corollary 1.10 gives a corrected version of [17, Corollary 2.2]. Under a stronger sparsity assumption, large deviations of neighborhood distributions for random networks have been used in [9] to study the large deviations of the spectral measure of certain random matrices.

2 Local weak convergence

In this section, we first recall the basic notions of local weak convergence in the more general context of rooted multi-graphs; see [4], [2], and [1]. Then, we give a general tightness lemma.

2.1 Local convergence of rooted multi-graphs

Let V be a countable set, a *multi-graph* $G = (V, \omega)$ is a vertex set V together with a map ω from V^2 to $\mathbb{Z}_+$ such that for all $(u, v) \in V^2$, $\omega(u, u)$ is even and $\omega(u, v) = \omega(v, u)$. For ease of notation, we sometimes set $\omega(v) = \omega(v, v)$ for the weight of the loop at v. If $e = \{u, v\}$ is an unordered pair $(u \neq v)$, we may also write $\omega(e)$ in place of $\omega(u, v)$. The *edge set* E of G is the set of unordered pairs $e = \{u, v\}$ such that $\omega(e) \geqslant 1$, $\omega(e)$ being the *multiplicity* of the edge $e \in E$. Similarly, $\omega(v)/2$ is the number of *loops* attached to v. A multi-graph with no loop, and with no edge with multiplicity greater than 1 is a graph.

The *degree* of v in G is defined by

$$\deg(v) = \sum_{u \in V} \omega(v, u).$$

The multi-graph G is *locally finite* if for any vertex v, $\deg(v) < \infty$.

We denote by $\widehat{\mathcal{G}}$ the set of all locally finite multi-graphs. For a multi-graph $G \in \widehat{\mathcal{G}}$, to avoid possible confusion, we will often denote by $V_G, \omega_G, \deg_G$ the corresponding vertex set, weight and degree functions.

Recall that a path π from u to v of length k is a sequence $\pi = (u_0, \ldots, u_k)$ with $u_0 = u$, $u_k = v$ and, for $0 \leqslant i \leqslant k - 1$, $\{u_i, u_{i+1}\} \in E$. If such $\pi : u \to v$ exists, the distance $D(u, v)$ in G between u and v is defined as the minimal length of all paths from u to v. If there is no path $\pi : u \to v$, then the distance $D(u, v)$ is set to be infinite. A multi-graph is *connected* if $D(u, v) < \infty$ for any $u \neq v \in V$.

Below, a *rooted multi-graph* $(G, o) = (V, \omega, o)$ is a locally finite and connected multi-graph (V, ω) with a distinguished vertex $o \in V$, the root. For $t \geqslant 0$, we denote by $(G, o)_t$ the induced rooted multi-graph with vertex set $\{u \in V : D(o, u) \leqslant t\}$. Two rooted multi-graphs $(G_i, o_i) = (V_i, \omega_i, o_i)$, $i \in \{1, 2\}$, are *isomorphic* if there

exists a bijection $\sigma : V_1 \to V_2$ such that $\sigma(o_1) = o_2$ and $\sigma(G_1) = G_2$, where σ acts on G_1 through $\sigma(u, v) = (\sigma(u), \sigma(v))$ and $\sigma(\omega) = \omega \circ \sigma$. We will denote this equivalence relation by $(G_1, o_1) \simeq (G_2, o_2)$. The associated equivalence classes can be seen as unlabeled rooted multi-graphs. We call $\widehat{\mathcal{G}}^*$ the set of all such equivalence classes.

We define the semi-distance d between two rooted multi-graphs (G_1, o_1) and (G_2, o_2) as

$$d((G_1, o_1), (G_2, o_2)) = \frac{1}{1 + T},$$

where T is the supremum of those $t > 0$ such that $(G_1, o_1)_t$ and $(G_2, o_2)_t$ are isomorphic. On the space $\widehat{\mathcal{G}}^*$, d is a distance. The associated topology will be referred to as the *local topology*. The space $(\widehat{\mathcal{G}}^*, d)$ is Polish (i.e. separable and complete) [1].

Explicit compact subsets of $\widehat{\mathcal{G}}^*$ can be constructed as follows. If $g \in \widehat{\mathcal{G}}^*$, we define

$$|g| = \sum_{v \in V} \deg(v),$$

i.e. twice the total number of edges in g. For $g \in \widehat{\mathcal{G}}^*$, $t \in \mathbb{N}$, the truncation at distance t, g_t, is defined as the equivalence class of $(G, o)_t$ where the equivalence class of (G, o) is g.

Lemma 2.1 *Let $t_0 \geqslant 0$ and $\varphi : \mathbb{N} \to \mathbb{R}_+$ be a non-negative function. Then*

$$K = \{g \in \widehat{\mathcal{G}}^* : \forall t \geqslant t_0, \ |g_t| \leqslant \varphi(t)\},$$

is a compact subset of $\widehat{\mathcal{G}}^*$ for the local topology.

Proof For each $t \geqslant t_0$, there is a finite number of elements in $\widehat{\mathcal{G}}^*$, say $f_{t,1}, \ldots, f_{t,n_t}$, such that $|g| \leqslant \varphi(t)$ and for any vertex the distance to the root is at most t. Therefore, the collection $A_{t,1}, \ldots, A_{t,n_t}$ where $A_{t,k} = \{g \in \widehat{\mathcal{G}}^* : g_t = f_{t,k}\}$ is a finite covering of K of radius $1/(1 + t)$. $\square$

The notions of local weak convergence introduced in Sect. 1.1 are immediately extended to the present setting of multi-graphs. The definitions of $U(G)$ in (1) and unimodularity (2) easily carry over to $\mathcal{P}(\widehat{\mathcal{G}}^*)$. The next simple lemma is proved in [4].

Lemma 2.2 *The set $\mathcal{P}_u(\widehat{\mathcal{G}}^*)$ is closed in the local weak topology.*

2.2 Compactness lemma for the local weak topology

Let G_n be a sequence of finite multi-graphs. We now give a condition which guarantees that the sequence $U(G_n)$ is tight for the local weak topology. If $G = (V, \omega)$ is a multi-graph, we define the degree of a subset $S \subset V$ as

$$\deg_G(S) = \sum_{v \in S} \deg_G(v). \tag{14}$$

The next lemma is a sufficient condition for tightness in $\mathcal{P}_u(\widehat{\mathcal{G}}^*)$. A similar result appears in Benjamini et al. [3, Theorem 3.1]. We give an independent proof.

Lemma 2.3 *Let* $\delta : [0, 1] \to \mathbb{R}_+$ *be a continuous increasing function such that* $\delta(0) = 0$. *There exists a compact set* $\Pi = \Pi(\delta) \subset \mathcal{P}_u(\widehat{\mathcal{G}}^*)$ *such that if a finite multi-graph* $G = (V, \omega)$ *satisfies*

$$\deg_G(S) \leqslant |V| \delta \left(\frac{|S|}{|V|} \right) \tag{15}$$

for all $S \subset V$, *then* $U(G) \in \Pi$.

Considering a sequence $U(G_n)$, $n \geqslant 1$, condition (15) amounts to a uniform integrability of the degree sequences of the multi-graphs (G_n), $n \geqslant 1$. It may seem quite paradoxical that a sole condition on the degrees implies the tightness of the whole graph sequence. However, the unimodularity of $U(G)$ yields enough uniformity for this result to hold.

Proof of Lemma 2.3 Since $\widehat{\mathcal{G}}^*$ is a Polish space, from Prohorov's theorem, a set $\Pi \subset \mathcal{P}(\widehat{\mathcal{G}}^*)$ is relatively compact if and only if for any $\varepsilon > 0$, there exists a compact $K \subset \widehat{\mathcal{G}}^*$ such that for all $\mu \in \Pi$, $\mu(K^c) \leqslant \varepsilon$.

Set $c = \delta(1)$. Without loss of generality, we may assume $c > 1$. We consider the increasing function $[0, c] \mapsto [0, 1]$

$$f = \delta^{-1}.$$

Now, for each $\varepsilon > 0$, and integer $t \geqslant 1$, we set

$$h_\varepsilon(t) = (f \circ \cdots \circ f)(\varepsilon 2^{-t}) \quad \text{and} \quad \varphi_\varepsilon(t) = \frac{(c/h_\varepsilon(t))^t - 1}{1 - h_\varepsilon(t)/c},$$

where the composition holds t times. We now define Π as being the closure of the set of measures μ in $\mathcal{P}_u(\widehat{\mathcal{G}}^*)$ such that for any $\varepsilon > 0$, $\mu(K_\varepsilon^c) \leqslant \varepsilon$ where

$$K_\varepsilon = \left\{ g \in \widehat{\mathcal{G}}^* : \forall t \geqslant 1, \ |g_t| \leqslant \varphi_\varepsilon(t) \right\}.$$

By Lemma 2.1, K_ε is a compact set of $\widehat{\mathcal{G}}^*$. Hence, Prohorov's theorem asserts that Π is a compact set of $\mathcal{P}_u(\widehat{\mathcal{G}}^*)$.

We now check that $\rho = U(G) \in \Pi$. This will conclude the proof of our lemma. It is sufficient to prove that $\rho(K_\varepsilon) \geqslant 1 - \varepsilon$ for all $\varepsilon > 0$. Let $t \geqslant 0$ be an integer, for $S \subset V$, $B(S, t)$ denote the set of vertices at distance at most t from a vertex in S. In particular, if $v \in V$ and g is the equivalence class of $(G(v), v)$ we have

$$\deg(B(v, t)) = |g_t|.$$

Notice also that $|B(S, 1)| \leqslant \deg(S)$. Set $|V| = n$. By iteration on (15), it follows that if $S \subset V$ is such that $|S| \leqslant h_\varepsilon(t)n$ then

$$|B(S, t)| \leqslant \deg(B(S, t - 1)) \leqslant (f \circ \cdots \circ f)(|S|/n) \leqslant 2^{-t}\varepsilon n.$$

Moreover, from (15), we have

$$\deg(V) = \sum_{v \in V} \deg(v) \leqslant cn$$

Hence, using Markov inequality, we deduce that the set

$$S_t = \{v \in V : \deg(v) \geqslant c/h_\varepsilon(t)\}$$

has cardinality at most $h_\varepsilon(t)n$. From what precedes, the set

$$U_t = \{v \in V : \exists u \in B(v, t), \deg_{G_n}(u) \geqslant c/h_\varepsilon(t)\}$$

has cardinality at most $2^{-t}\varepsilon n$. Note that, if $v \notin U_t$, then $\deg(B(v, t))$ is bounded by

$$\frac{(c/h_\varepsilon(t))^t - 1}{c/h_\varepsilon(t) - 1} = \varphi_\varepsilon(t).$$

This implies that the set

$$V_t = \{v \in V : \deg(B(v, t)) \geqslant \varphi_\varepsilon(t)\}$$

has cardinality at most $2^{-t}\varepsilon n$. So finally, from the union bound, the set

$$W = \{v \in V : \forall t \geqslant 1, \deg(B(v, t)) \leqslant \varphi_\varepsilon(t)\}$$

has cardinality at least $(1 - \varepsilon)n$. We have thus checked that $\rho(K_\varepsilon) \geqslant 1 - \varepsilon$. $\quad\square$

3 Unimodular Galton–Watson trees with given neighborhood

The aim of this section is to prove Proposition 1.1. We thus fix $h \in \mathbb{N}$ and $P \in \mathcal{P}(\mathcal{T}_h^*)$ admissible. We start with some simple observations which ensure that $\mathrm{UGW}_h(P)$ is indeed well defined.

First observe that if $\tau \in \mathcal{T}_h^*$, $t' \in \mathcal{T}_{h-1}^*$ and $S = \tau \cup t'_+$, then (recall the definition of $\tau \cup t'_+$ and Fig. 2)

$$1 + \left|\{v \overset{\tau}{\sim} o : \tau(o, v) \simeq t'\}\right| = \left|\{v \overset{S}{\sim} o : S(v, o) \simeq \tau, S(o, v) \simeq t'\}\right|. \tag{16}$$

Therefore, for any $t, t' \in \mathcal{T}^*_{h-1}$,

$$\sum_{\tau \in \mathcal{T}^*_h} P(\tau \cup t'_+)\Big(1 + \big|\{v \overset{\tau}{\sim} o : \tau(o, v) \simeq t'\}\big|\Big)\mathbf{1}(\tau_{h-1} = t)$$

$$= \sum_{\tau \in \mathcal{T}^*_h} \sum_{S \in \mathcal{T}^*_h} P(S)\big|\{v \overset{S}{\sim} o : S(v, o) \simeq \tau, S(o, v) \simeq t'\}\big|\mathbf{1}(S = \tau \cup t'_+)\mathbf{1}(\tau_{h-1} = t)$$

$$= \sum_{S \in \mathcal{T}^*_h} P(S)\big|\{v \overset{S}{\sim} o : S(v, o)_{h-1} \simeq t, S(o, v) \simeq t'\}\big| = e_P(t, t'),$$

where e_P was defined by (4). We thus have checked that $\widehat{P}_{t,t'}$ defined by (5) is indeed a probability measure on $\mathcal{T}^*_h$. Consequently, the probability measure $\mathrm{UGW}_h(P)$ is well defined.

3.1 Unimodularity

The next lemma is a direct argument for the unimodularity of $\mathrm{UGW}_h(P)$, which establishes the first part of Proposition 1.1. We remark however that this fact could be derived indirectly from Theorem 4.8 and Lemma 4.9 below, which ensure in particular that $\mathrm{UGW}_h(P)$ is sofic (and hence unimodular).

Lemma 3.1 *Fix $h \in \mathbb{N}$ and $P \in \mathcal{P}(\mathcal{T}^*_h)$ admissible. The measure $\mathrm{UGW}_h(P)$ is unimodular.*

Proof It is sufficient to check the so-called involution invariance, i.e. that (2) holds with f restricted to functions $f : \mathcal{G}^{**} \to \mathbb{R}_+$ such that $f(G, u, v) = 0$ unless $\{u, v\} \in E_G$; see [1]. Recall that we may extend $f : \mathcal{G}^{**} \to \mathbb{R}_+$ to all connected graphs with two distinguished roots (G, u, v) through the isomorphism class.

Let (T, o) be the random rooted tree defined in the introduction whose equivalence class has law $\mathrm{UGW}_h(P)$. Recall that the neighbors of the root o are indexed by $1, \ldots, \deg_T(o)$ and that the vector of subtrees $(T(o, 1), \ldots, T(o, \deg_T(o)))$ is exchangeable. We write

$$\mathbb{E}\sum_{v \overset{T}{\sim} o} f(T, o, v) = \sum_{g \in \mathcal{T}^*_h} P(g)\sum_{v \overset{g}{\sim} o}\mathbb{E}[f(T, o, v) \mid (T, o)_h \simeq g]$$

$$= \sum_{\tau \in \mathcal{T}^*_h, \, t' \in \mathcal{T}^*_{h-1}} P(S)\big|\{v \overset{S}{\sim} o : S(o, v) \simeq t', S(v, o) \simeq \tau\}\big|$$

$$\times \mathbb{E}[f(T, o, 1) \mid T(o, 1)_{h-1} \simeq t', T(1, o)_h \simeq \tau],$$

where, in the summand, $S = \tau \cup t'_+$. Now, (5) and (16) imply

$$\mathbb{E}\sum_{v \overset{T}{\sim} o} f(T, o, v) = \sum_{t,t'} e_P(t, t')\sum_{\tau : \, \tau_{h-1}=t} \widehat{P}_{t,t'}(\tau)\mathbb{E}[f(T, o, 1) \mid T(o, 1)_{h-1}$$

$$\simeq t', T(1, o)_h \simeq \tau].$$

For $(t, t') \in \mathcal{T}^*_{h-1}$, we introduce a new random tree $H = H_{t,t'}$ defined as follows. Start with two vertices o and o' which are connected by an edge. Attach the tree t to o and the tree t' to o', so that the type of o is (t, t') and the type of o' is (t', t). Sample independently $H(o', o)_h$ according to $\widehat{P}_{t,t'}$ and $H(o, o')_h$ according to $\widehat{P}_{t',t}$. The subtrees $H(o', o)_h$ and $H(o, o')_h$ define the types of the children of o and o'. Next, sample independently their rooted subtrees, according to their types, i.e. $H(o, v)_h$ (resp. $H(o', v)_h$) is sampled according to $\widehat{P}_{a,b}$ if $v \sim o$ (resp. $v \sim o'$) has type (a, b). Repeating recursively for all children defines the random tree H. From the definition of $\mathrm{UGW}_h(P)$, one has

$$\sum_{\tau:\ \tau_{h-1}=t} \widehat{P}_{t,t'}(\tau) \mathbb{E}[f(T, o, 1)|T(o, 1)_{h-1} \simeq t', T(1, o)_h \simeq \tau] = \mathbb{E}_{t,t'}[f(H, o, o')],$$

where we use $\mathbb{E}_{t,t'}$ for expectation over the random $H = H_{t,t'}$ defined above. It follows that

$$\mathbb{E}\sum_{\substack{T \\ v\sim o}} f(T, o, v) = \sum_{t,t'} e_P(t, t') \mathbb{E}_{t,t'}[f(H, o, o')].$$

Similarly,

$$\mathbb{E}\sum_{\substack{T \\ v\sim o}} f(T, v, o) = \sum_{t,t'} e_P(t, t') \mathbb{E}_{t,t'}[f(H, o', o)] = \sum_{t,t'} e_P(t, t') \mathbb{E}_{t',t}[f(H, o, o')],$$

where the second identity follows from the symmetry in o, o' in the definition in H, which implies that $\mathbb{E}_{t,t'}[f(H, o', o)] = \mathbb{E}_{t',t}[f(H, o, o')]$.

Finally, the assumption $e_P(t, t') = e_P(t', t)$ yields

$$\mathbb{E}\sum_{\substack{T \\ v\sim o}} f(T, v, o) = \sum_{t,t'} e_P(t', t) \mathbb{E}_{t',t}[f(H, o, o')] = \mathbb{E}\sum_{\substack{T \\ v\sim o}} f(T, o, v).$$

$\square$

3.2 Consistency lemma

We turn to the second part of Proposition 1.1. The following lemma computes the law of the $(h + 1)$-neighborhood of a Galton–Watson tree with a given h-neighborhood.

Lemma 3.2 *Fix* $h \in \mathbb{N}$, $P \in \mathcal{P}(\mathcal{T}^*_h)$ *admissible and set* $\rho = \mathrm{UGW}_h(P)$. *For any* $\tau \in \mathcal{T}^*_{h+1}$ *with* $\deg_\tau(o) = d$, *we have*

$$\mathbb{P}_\rho((T, o)_{h+1} = \tau) = P(\tau_h) \prod_{a\in\mathcal{A}} \binom{n_a}{(k_{a,b})_{b\in\mathcal{B}_a}} \prod_{b\in\mathcal{B}_a} \widehat{P}_{s^a, s^{-a}}(t^{a,b})^{k_{a,b}},$$

where

- $(t^i \in \mathcal{T}_h^*, 1 \leqslant i \leqslant d)$ *are the subtrees of* τ *attached to the offspring of the root, and for* $1 \leqslant i \leqslant d$, $s^i = (t^i)_{h-1}$;
- $\{s^a\}_{a \in \mathcal{A}}$ *is set of distinct elements of* $(s^i, 1 \leqslant i \leqslant d)$, *and, for each* $a \in \mathcal{A}$, $\{t^{a,b}\}_{b \in \mathcal{B}_a}$ *is the set the distinct elements of* $(t^i, 1 \leqslant i \leqslant d)$, *such that* $(t^{a,b})_{h-1} = s^a$;
- n_a *is the cardinality of* s^i'*s equal to* s^a *and* $k_{a,b}$ *is the cardinality of* t^i'*s equal to* $t^{a,b}$;
- $s^{-a} = (t^{-a})_{h-1}$ *and* $t^{-a} \in \mathcal{T}_h^*$ *is the tree obtained from* τ_h *by removing one offsping with subtree equal to* s^a.

Proof Using $\rho_h = P$, for a fixed $\tau \in \mathcal{T}_{h+1}^*$, the above definitions allow us to write

$$\mathbb{P}_\rho\big((T,o)_{h+1} = \tau\big) = P(\tau_h)\mathbb{P}_\rho\big((T,o)_{h+1} = \tau|(T,o)_h = \tau_h\big)$$

$$= P(\tau_h)\mathbb{P}_\rho\Big(\forall a \in \mathcal{A}, b \in \mathcal{B}_a : \big|\{v \overset{T}{\sim} o : T(o,v)_h = t^{a,b}\}\big| = k_{a,b} \mid (T,o)_h = \tau_h\Big).$$

Observe that $T(o,v)_h = t^{a,b}$ implies that $T(o,v)_{h-1} = s^a$. Moreover, given $(T,o)_h = t$, $T(o,v)_{h-1} = s^a$ implies that $T(v,o)_{h-1} = s^{-a}$, i.e. the type of vertex v is (s^a, s^{-a}). The lemma is then a consequence of the conditional independence of the subtrees attached to the offspring of the root given $(T,o)_h$. □

Lemma 3.3 *Fix integers* $k > h \geqslant 1$, $P \in \mathcal{P}(\mathcal{T}_h^*)$ *admissible and set* $\rho = \mathrm{UGW}_h(P)$. *Then*

$$\rho = \mathrm{UGW}_k(\rho_k).$$

Proof By recursion, it suffices to prove the statement for $k = h + 1$. For $s \in \mathcal{T}_{h-1}^*$ such that $e_P(s, s') > 0$ for some $s' \in \mathcal{T}_{h-1}^*$, we may define the probability measure

$$\widehat{\mathbb{P}}_s(\cdot) = \frac{\mathbb{E}_\rho\big|\{v \overset{T}{\sim} o : T(o,v) \in \cdot, \ T(v,o)_{h-1} = s\}\big|}{\mathbb{E}_\rho\big|\{v \overset{T}{\sim} o : T(v,o)_{h-1} = s\}\big|}.$$

In words, $\widehat{\mathbb{P}}_s \in \mathcal{P}(\mathcal{T}^*)$ is the law of the whole subtree $T(o,v)$ of a neighbor v of the root given that $T(v,o)_{h-1} = s$, where (T,o) has law ρ. Next, we show that, for $s, s' \in \mathcal{T}_{h-1}^*$, $t \in \mathcal{T}_h^*$ such that $t_{h-1} = s$ and $e_P(s, s') > 0$, one has

$$\widehat{P}_{s,s'}(t) = \frac{\widehat{\mathbb{P}}_{s'}((T,o)_h = t)}{\widehat{\mathbb{P}}_{s'}((T,o)_{h-1} = s)}, \tag{17}$$

where (T,o) is now the random variable with law $\widehat{\mathbb{P}}_{s'}$. Since $P = \rho_h$ one has $e_P(s, s') = \mathbb{E}_\rho\big|\{v \overset{T}{\sim} o : T(o,v)_{h-1} = s, \ T(v,o)_{h-1} = s'\}\big|$, and therefore

$$\frac{\widehat{\mathbb{P}}_{s'}((T,o)_h = t)}{\widehat{\mathbb{P}}_{s'}((T,o)_{h-1} = s)} = \frac{\mathbb{E}_\rho\big|\{v \overset{T}{\sim} o : T(o,v)_h = t, \ T(v,o)_{h-1} = s'\}\big|}{e_P(s, s')}.$$

However, with $n = \left|\{v \overset{t}{\sim} o : t(o, v) = s'\}\right|$, we deduce from the unimodularity of ρ and (16) that

$$
\mathbb{E}_\rho \left|\{v \overset{T}{\sim} o : T(o, v)_h = t \,,\, T(v, o)_{h-1} = s'\}\right|
$$
$$
= \mathbb{E}_\rho \left|\{v \overset{T}{\sim} o : T(v, o)_h = t \,,\, T(o, v)_{h-1} = s'\}\right|
$$
$$
= (n + 1) P(t \cup s'_+). \tag{18}
$$

This proves (17).

Now we set $Q = \rho_{h+1}$ and $\rho' = \mathrm{UGW}_{h+1}(Q)$. Our aim is to prove that $\rho' = \rho$. It is sufficient to prove that for any $t, t' \in \mathcal{T}_h^*$ and $\tau \in \mathcal{T}_{h+1}^*$ such that $\tau_h = t$ and $e_Q(t, t') > 0$,

$$
\widehat{Q}_{t,t'}(\tau) = \frac{\widehat{\mathbb{P}}_{s'}((T, o)_{h+1} = \tau)}{\widehat{\mathbb{P}}_{s'}((T, o)_h = t)}, \tag{19}
$$

where $s' = t'_{h-1}$. Indeed, since ρ' and ρ have the same $h + 1$ neighborhood, this would prove that they have in fact the same $h + 2$ neighborhood and, by conditional independence, we would deduce that $\rho = \rho'$.

Let us prove (19). Set

$$
k = \left|\{v \overset{\tau}{\sim} o : \tau(o, v) = t'\}\right| \leqslant n = \left|\{v \overset{t}{\sim} o : t(o, v) = s'\}\right|,
$$

where, as above, $t = \tau_h$ and $s' = t'_{h-1}$. Since $(\tau \cup t'_+)_h = t \cup s'_+$, and $\rho'_{h+1} = \rho_{h+1}$, we have

$$
Q(\tau \cup t'_+) = P(t \cup s'_+) \mathbb{P}_{\rho'}((T, o)_{h+1} = \tau \cup t'_+ | (T, o)_h = t \cup s'_+)
$$
$$
= P(t \cup s'_+)) \mathbb{P}_\rho((T, o)_{h+1} = \tau \cup t'_+ | (T, o)_h = t \cup s'_+).
$$

As in Lemma 3.2, let $t^i \in \mathcal{T}_h^*$, $1 \leqslant i \leqslant d$ be the subtrees of τ attached to the offspring of the root and call s^i their restriction to $\mathcal{T}_{h-1}^*$. By construction, k elements of the t^i's are equal to t' and n elements of the s^i's are equal to s'. Let $(s^a)_a$ be the set of distinct elements of the set $\{s'\} \cup \{s^i, 1 \leqslant i \leqslant d\}$, and, for each a, let $(t^{a,b})_b$ denote the distinct elements of $\{t'\} \cup \{t^i, 1 \leqslant i \leqslant d\}$, such that $t^{a,b}$ restricted to $\mathcal{T}_{h-1}^*$ is s^a. We denote by n_a the cardinality of s^i's equal to s^a and $k_{a,b}$ the cardinality of t^i's equal to $t^{a,b}$. We set $n'_a = n_a + \mathbf{1}(s_a = s')$ and $k'_{a,b} = k_{a,b} + \mathbf{1}(t_{a,b} = t')$. Then, Lemma 3.2 yields

$$
\mathbb{P}_\rho((T, o)_{h+1} = \tau \cup t'_+ | (T, o)_h = t \cup s'_+) = \prod_a \binom{n'_a}{(k'_{a,b})_b} \prod_b \widehat{P}_{s^a, s^{-a}}(t^{a,b})^{k'_{a,b}}
$$
$$
= \frac{n+1}{k+1} \widehat{P}_{s', s}(t') \prod_a \binom{n_a}{(k_{a,b})_b} \prod_b \widehat{P}_{s^a, s^{-a}}(t^{a,b})^{k_{a,b}}, \tag{20}
$$

where, $s^{-a} = [(t \cup s'_+)^{-a}]_{h-1}$ and $(t \cup s'_+)^{-a} \in \mathcal{T}_h^*$ is the tree obtained from $t \cup s'_+$ by removing one of the offspring with subtree equal to s^a. Thus, we find

$$Q(\tau \cup t'_+) = P(t \cup s'_+) \frac{n+1}{k+1} \widehat{P}_{s',s}(t') \prod_a \binom{n_a}{(k_{a,b})_b} \prod_b \widehat{P}_{s^a,s^{-a}}(t^{a,b})^{k_{a,b}}. \quad (21)$$

Since $\rho_{h+1} = \rho'_{h+1}$, one has

$$e_Q(t,t') = e_Q(t',t) = \mathbb{E}_\rho \big| \{ v \overset{T}{\sim} o : T(o,v)_h = t' \,,\, T(v,o)_h = t \} \big|.$$

By sampling the h-neighborhood $(T,o)_h$ first, and using the number n as above, one has

$$e_Q(t,t') = (n+1) P(t \cup s'_+) \widehat{P}_{s',s}(t'). \quad (22)$$

From (21) and (22) we find

$$\widehat{Q}_{t,t'}(\tau) = \frac{(k+1) Q(\tau \cup t'_+)}{e_Q(t,t')} = \prod_a \binom{n_a}{(k_{a,b})_b} \prod_b \widehat{P}_{s^a,s^{-a}}(t^{a,b})^{k_{a,b}}. \quad (23)$$

Next, we show that the right hand side in (19) equals the above expression. We have

$$\frac{\widehat{\mathbb{P}}_{s'}((T,o)_{h+1} = \tau)}{\widehat{\mathbb{P}}_{s'}((T,o)_h = t)} = \frac{\mathbb{E}_\rho \big| \{ v \overset{T}{\sim} o : T(o,v)_{h+1} = \tau \,,\, T(v,o)_{h-1} = s' \} \big|}{\mathbb{E}_\rho \big| \{ v \overset{T}{\sim} o : T(o,v)_h = t \,,\, T(v,o)_{h-1} = s' \} \big|}$$

$$= \frac{\mathbb{E}_\rho \big| \{ v \overset{T}{\sim} o : T(v,o)_{h+1} = \tau \,,\, T(o,v)_{h-1} = s' \} \big|}{(n+1) P(t \cup s'_+)},$$

where we have used unimodularity and (18). Now, by sampling first the h-neighborhood $(T,o)_h$, one finds that

$$\mathbb{E}_\rho \big| \{ v \overset{T}{\sim} o : T(v,o)_{h+1} = \tau \,,\, T(o,v)_{h-1} = s' \} \big|$$

$$= P(t \cup s'_+) \sum_{t':t'_{h-1}=s'} \mathbb{E}_\rho \left[\sum_{v \overset{T}{\sim} o} \mathbf{1}(T(v,o)_{h+1} = \tau,\, T(o,v)_h = t') \,\big|\, (T,o)_h = t \cup s'_+ \right]$$

$$= P(t \cup s'_+) \sum_{t':t'_{h-1}=s'} (k+1) \mathbb{P}_\rho \big((T,o)_{h+1} = \tau \cup t'_+ \,|\, (T,o)_h = t \cup s'_+ \big),$$

where, as before, $k = k(t')$ stands for the number of $v \overset{\tau}{\sim} o$ such that $\tau(o, v) = t'$. Using Lemma 3.2 in the form (20), and the fact that $\sum_{t':t'_{h-1}=s'} \widehat{P}_{s',s}(t') = 1$, we find

$$
\mathbb{E}_\rho \big| \{ v \overset{T}{\sim} o : T(v, o)_{h+1} = \tau , T(o, v)_{h-1} = s' \} \big|
$$
$$
= (n + 1) P(t \cup s'_+) \prod_a \binom{n_a}{(k_{a,b})_b} \prod_b \widehat{P}_{s^a, s^{-a}}(t^{a,b})^{k_{a,b}}.
$$

Hence,

$$
\frac{\widehat{\mathbb{P}}_{s'}((T, o)_{h+1} = \tau)}{\widehat{\mathbb{P}}_{s'}((T, o)_h = t)} = \prod_a \binom{n_a}{(k_{a,b})_b} \prod_b \widehat{P}_{s^a, s^{-a}}(t^{a,b})^{k_{a,b}}. \tag{24}
$$

The identity (19) follows from (24) and (23). $\qquad\square$

Remark 3.4 From (22) one deduces the identity

$$
e_Q(t, t') = e_P(s, s') \widehat{P}_{s,s'}(t) \widehat{P}_{s',s}(t'),
$$

for any $t, t' \in \mathcal{T}_h^*$, with $s = t_{h-1}, s' = t'_{h-1}$, for any $P \in \mathcal{P}(\mathcal{T}_h^*)$ admissible, with $Q = [\mathrm{UGW}_h(P)]_{h+1}$.

4 Configuration model for directed graphs with colored edges

This section introduces a generalized configuration model, to be used later on to count the number of graphs with a given tree-like neighborhood distribution.

4.1 Directed multi-graphs with colors

We are now going to define a family of directed multi-graphs with colored edges. Let L be a fixed integer. Each pair (i, j) with $1 \leqslant i, j \leqslant L$ is interpreted as a color. Define the sets of colors

$$
\mathcal{C} = \{(i, j) : 1 \leqslant i, j \leqslant L\}
$$
$$
\mathcal{C}_< = \{(i, j) : 1 \leqslant i < j \leqslant L\}, \qquad \mathcal{C}_= = \{(i, i) : 1 \leqslant i \leqslant L\}.
$$

Also, define $\mathcal{C}_\leqslant = \mathcal{C}_< \cup \mathcal{C}_=, \mathcal{C}_> = \mathcal{C} \backslash \mathcal{C}_\leqslant$ and $\mathcal{C}_\neq = \mathcal{C} \backslash \mathcal{C}_=$. If $c = (i, j) \in \mathcal{C}$, then set $\bar{c} = (j, i)$ for the conjugate color.

We consider the class $\widehat{\mathcal{G}}(\mathcal{C})$ of directed multi-graphs with $\mathcal{C}$-colored edges defined as follows. We say that a directed multi-graph G is an element of $\widehat{\mathcal{G}}(\mathcal{C})$ if $G = (V, \omega)$ where $V = [n]$ for some $n \in \mathbb{N}$, $\omega = \{\omega_c\}_{c \in \mathcal{C}}$ and for each $c \in \mathcal{C}$, ω_c is a map $\omega_c : V^2 \to \mathbb{Z}_+$ with the following properties: if $c \in \mathcal{C}_=$, then $\omega_c(u, u)$ is even for all $u \in V$, and $\omega_c(u, v) = \omega_c(v, u)$ for all $u, v \in V$; if $c \in \mathcal{C}_\neq$, then $\omega_c(u, v) = \omega_{\bar{c}}(v, u)$ for all $u, v \in V$. The interpretation is that, for any $c \in \mathcal{C}$, if $u \neq v$ then $\omega_c(u, v)$ is the number of directed edges of color c from u to v; if $u = v$ and $c \in \mathcal{C}_=$, then $\frac{1}{2}\omega_c(u, u)$ is the number of loops of color c at u, while if $u = v$ and $c \in \mathcal{C}_<$ then
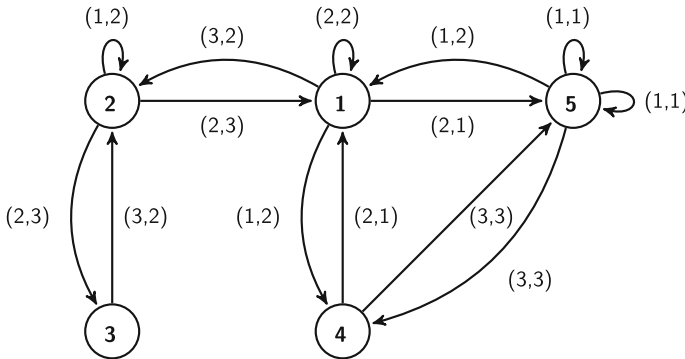
Fig. 3 An example of directed colored multi-graph in $\widehat{\mathcal{G}}(\mathcal{C})$. Here $n = 5$, $L = 3$, with: $\omega_{(2,1)}(4, 1) = \omega_{(1,2)}(1, 4) = \omega_{(2,1)}(1, 5) = \omega_{(1,2)}(5, 1) = 1$; $\omega_{(2,3)}(2, 3) = \omega_{(3,2)}(3, 2) = \omega_{(2,3)}(2, 1) = \omega_{(3,2)}(1, 2) = 1$; $\omega_{(3,3)}(4, 5) = \omega_{(3,3)}(5, 4) = 1$; $\omega_{(2,2)}(1, 1) = 2$; $\omega_{(1,1)}(5, 5) = 4$; $\omega_{(2,2)}(1, 1) = 2$; $\omega_{(1,2)}(2, 2) = \omega_{(2,1)}(2, 2) = 1$; all other entries of ω are zero

$\omega_c(u, u) = \omega_{\bar{c}}(u, u)$ is the number of loops of color c at u; we adopt the convention that there are no loops of color $c \in \mathcal{C}_>$ at any vertex. We call $\mathcal{G}(\mathcal{C})$ the subset of $\widehat{\mathcal{G}}(\mathcal{C})$ consisting of graphs, i.e. $G = (V, \omega)$ such that $\omega_c(u, v) \in \{0, 1\}$ for all $c \in \mathcal{C}$ and $u, v \in V$ (no multiple edges) and $\omega_c(u, u) = 0$ for all $c \in \mathcal{C}$ and $u \in V$ (no loop). See Fig. 3 for an example of an element of $\widehat{\mathcal{G}}(\mathcal{C})$.

If $G \in \widehat{\mathcal{G}}(\mathcal{C})$, one can define the colorblind multi-graph $\bar{G} = (V, \bar{\omega})$, by setting

$$\bar{\omega}(u, v) = \sum_{c \in \mathcal{C}} \omega_c(u, v). \tag{25}$$

The multi-graph $\bar{G} = (V, \bar{\omega})$ can be identified with an undirected multi-graph, in that by construction $\bar{\omega}(u, v) = \bar{\omega}(v, u)$ for all $u, v \in V$. We say that G is a simple graph if $\bar{G}$ has no loops and no multiple edges. Clearly, if $L = 1$ then there is only one color, so that any multi-graph $G \in \widehat{\mathcal{G}}(\mathcal{C})$ coincides with its own $\bar{G}$.

If $G \in \widehat{\mathcal{G}}(\mathcal{C})$, $c \in \mathcal{C}$ and $u \in V$, set

$$D_c(u) = \sum_v \omega_c(u, v), \tag{26}$$

and write $D(u) = \{D_c(u), c \in \mathcal{C}\}$. Note that $D(u)$ is an element of $\mathcal{M}_L$, defined as the set of $L \times L$ matrices with nonnegative integer valued entries. The vector $\mathbf{D} = \{D(u), u \in V\}$ of such matrices will be called the *degree sequence* of G.

4.2 Directed colored multi-graphs with given degree sequence

Fix $n \in \mathbb{N}$, and let $\mathcal{D}_n$ denote the set of all vectors $(D(1), \dots, D(n))$ such that $D(i) = \{D_c(i), c \in \mathcal{C}\} \in \mathcal{M}_L$ for all $i \in [n]$, and such that

$$S = \sum_{i=1}^n D(i) \tag{27}$$

is a symmetric matrix with even coefficients on the diagonal, i.e. $S = \{S_c, \; c \in \mathcal{C}\}$, $S_c = S_{\bar{c}}$ for all $c \in \mathcal{C}$, and $S_c \in 2\mathbb{Z}_+$ for all $c \in \mathcal{C}_=$. Clearly, if $G \in \widehat{\mathcal{G}}(\mathcal{C})$ then the vector $\mathbf{D}$ defined by (26) yields an element of $\mathcal{D}_n$ for some n. Next, for a given $\mathbf{D} \in \mathcal{D}_n$ we consider the set of all elements of $\widehat{\mathcal{G}}(\mathcal{C})$ which have $\mathbf{D}$ as their degree sequence.

Definition 4.1 Fix $n \in \mathbb{N}$ and $\mathbf{D} \in \mathcal{D}_n$.

- $\widehat{\mathcal{G}}(\mathbf{D})$ is the set of multi-graphs $G \in \widehat{\mathcal{G}}(\mathcal{C})$ with $V = [n]$ such that the degree sequence of G defined by (26) coincides with $\mathbf{D}$.
- $\mathcal{G}(\mathbf{D}, h)$ is the set of $G \in \widehat{\mathcal{G}}(\mathbf{D})$ such that the colorblind graph $\bar{G}$ defined in (25) contains no cycle of length $\ell \leqslant h$.

We also use the notation $\mathcal{G}(\mathbf{D})$ for the set of simple graphs in $\widehat{\mathcal{G}}(\mathbf{D})$. This set coincides with $\mathcal{G}(\mathbf{D}, 2)$, since loops are cycles with length 1 and multiple edges are cycles with length 2. The main goal of this section is to provide asymptotic formulas for the cardinality of $\mathcal{G}(\mathbf{D})$ and more generally of $\mathcal{G}(\mathbf{D}, h)$, for any $h \in \mathbb{N}$. To this end we introduce a natural extension of the usual configuration model from [7], see also [23].

Fix a multi-graph $G \in \widehat{\mathcal{G}}(\mathbf{D})$. For a fixed $c \in \mathcal{C}_=$, let G_c denote the subgraph of G obtained by removing all edges but the ones with color c. If $c \in \mathcal{C}_<$ instead, then define G_c as the subgraph of G obtained by removing all edges but the ones with color c or $\bar{c}$. Thus, every $G \in \widehat{\mathcal{G}}(\mathbf{D})$ is the result of the superposition of the multi-graphs G_c, $c \in \mathcal{C}_\leqslant$. We may then analyze each color separately.

4.2.1 Configuration model for $c \in \mathcal{C}_=$

When $c \in \mathcal{C}_=$, every pair u, v satisfies $\omega_c(u, v) = \omega_c(v, u)$, so G_c is actually a multi-graph with undirected edges, and we may use the usual construction [7, Section 2.4]. We provide the details for completeness. The degrees of G_c are fixed by the sequence $D_c(1), \ldots, D_c(n)$. Let $W_c = \cup_{i=1}^n W_c(i)$, be a fixed set of $S_c = \sum_{i=1}^n D_c(i)$ points, with the subsets $W_c(i)$ satisfying $|W_c(i)| = D_c(i)$. Recall that S_c is even by assumption. Let Σ_c be the set of all perfect matchings of the complete graph over the points of W_c, i.e. the set of all partitions of W_c into disjoint edges. Then,

$$|\Sigma_c| = (S_c - 1)!! = (S_c - 1)(S_c - 3) \cdots 1 = \frac{S_c!}{(S_c/2)! 2^{S_c/2}}.$$

Elements of Σ_c are called configurations. For any configuration $\sigma_c \in \Sigma_c$, call $\Gamma(\sigma_c)$ the multi-graph on $[n]$ with undirected edges obtained by including an edge $\{i, j\}$ iff σ_c has a pair with one element in $W_c(i)$ and the other in $W_c(j)$. Notice that $\Gamma(\sigma_c)$ has the same degree sequence $D_c(1), \ldots, D_c(n)$ of G_c. Moreover, any multi-graph with that degree sequence equals $\Gamma(\sigma_c)$ for some $\sigma_c \in \Sigma_c$.

Lemma 4.2 *Fix $c \in \mathcal{C}_=$. Let H be a multi-graph on $[n]$ with undirected edges and with degree sequence $D_c(1), \ldots, D_c(n)$. The number of $\sigma_c \in \Sigma_c$ such that $\Gamma(\sigma_c) = H$ is given by*

$$n_c(H) = \frac{\prod_{i=1}^n D_c(i)!}{\prod_{i=1}^n (\omega_c(i, i)/2)! 2^{(\omega_c(i,i)/2)} \prod_{i<j} \omega_c(i, j)!} \tag{28}$$

where $\omega_c(i, j)$ is the number of edges between nodes $\{i, j\}$ in H, while $\omega_c(i, i)/2$ is the number of loops at node i in H.

Proof We need to count the number of matchings $\sigma_c \in \Sigma_c$ such that for every $i < j$ one has $\omega_c(i, j)$ edges between $W_c(i)$ and $W_c(j)$, and such that for all i one has $\frac{1}{2}\omega_c(i, i)$ edges within $W_c(i)$. Fix $i < j$. Once we choose the $\omega_c(i, j)$ elements of $W_c(i)$ and the $\omega_c(i, j)$ elements of $W_c(j)$ to be matched together to produce the $\omega_c(i, j)$ edges, then there are $\omega_c(i, j)!$ distinct matchings that produce the same graph. Similarly, once we fix the $\omega_c(i, i)$ elements of $W_c(i)$ to be matched together to produce the $\frac{1}{2}\omega_c(i, i)$ loops at i, then there are $(\omega_c(i, i) - 1)!!$ distinct matchings that produce the same graph. On the other hand, for every node i there are

$$\binom{D_c(i)}{\omega_c(i, 1), \ldots, \omega_c(i, n)} = \frac{D_c(i)!}{\omega_c(i, 1)! \cdots \omega_c(i, n)!}$$

distinct ways of choosing the elements of $W_c(i)$ to be matched with $W_c(1), \ldots, W_c(n)$ respectively. Putting all together we arrive at the following expression for the total number of configurations producing the graph H:

$$\prod_{i=1}^n \binom{D_c(i)}{\omega_c(i, 1), \ldots, \omega_c(i, n)} \prod_{i<j} \omega_c(i, j)! \prod_{i=1}^n (\omega_c(i, i) - 1)!!,$$

which can be rewritten as (28).   □

4.2.2 Configuration model for $c \in \mathcal{C}_<$

When $c \in \mathcal{C}_<$, every pair u, v satisfies $\omega_c(u, v) = \omega_{\bar{c}}(v, u)$, so for the multi-graph G_c, $D_c(i)$ represents the number of outgoing edges at node i, which equals the number of incoming edges at that node. Here we use a bipartite version of the previous construction. Let $W_c = \cup_{i=1}^n W_c(i)$, be a fixed set of $S_c = \sum_{i=1}^n D_c(i)$ points, with the subsets $W_c(i)$ satisfying $|W_c(i)| = D_c(i)$. Similarly, set $\bar{W}_c = \cup_{i=1}^n \bar{W}_c(i)$, with $|\bar{W}_c(i)| = D_{\bar{c}}(i)$. Consider the set Σ_c of all perfect matchings of the complete bipartite graph over the sets $(W_c, \bar{W}_c)$, i.e. the set of perfect matchings containing only edges connecting an elements of W_c with an element of $\bar{W}_c$. Since $S_c = S_{\bar{c}}$, one has $|W_c| = |\bar{W}_c|$, and Σ_c can be identified with the set of permutations of S_c objects, or the set of bijective maps $W_c \mapsto \bar{W}_c$, and $|\Sigma_c| = S_c!$. A configuration is an element $\sigma_c \in \Sigma_c$. For any configuration σ_c, let $\Gamma(\sigma_c)$ denote the directed multi-graph on $[n]$ obtained by including the directed edge (i, j) with color c and the edge (j, i) with color $\bar{c}$ iff σ_c has a pair with one element in $W_c(i)$ and the other in $\bar{W}_c(j)$. Notice that $\Gamma(\sigma_c)$ has the same degree sequence $D_c(1), \ldots, D_c(n)$ of G_c, and any multi-graph with directed edges with colors with the same degree sequence equals $\Gamma(\sigma_c)$ for some $\sigma_c \in \Sigma_c$.

Lemma 4.3 *Fix $c \in \mathcal{C}_<$. Let H be a multi-graph on $[n]$ with directed edges with colors $(c, \bar{c})$ only and with degree sequence $D_c(1), \ldots, D_c(n)$. The number of $\sigma_c \in \Sigma_c$ such*

that $\Gamma(\sigma_c) = H$ is given by

$$n_c(H) = \frac{\prod_{i=1}^n D_c(i)! D_{\bar{c}}(i)!}{\prod_{i,j} \omega_c(i,j)!} \tag{29}$$

where $\omega_c(i,j) = \omega_{\bar{c}}(j,i)$ is the number of edges from i to j with color c in H.

Proof We have to count the number of bijective maps $W_c \mapsto \bar{W}_c$ such that for every $i, j \in [n]$ (including the case $i = j$), $\omega_c(i,j)$ elements of $W_c(i)$ are mapped to $\bar{W}_c(j)$. We begin by choosing, for every fixed node i, the subsets of $W_c(i)$ that are mapped into $\bar{W}_c(k)$, $k = 1, \ldots, n$, and the subsets of $\bar{W}_c(i)$ that are mapped into $W_c(k)$, $k = 1, \ldots, n$. This can be done in

$$\prod_{i=1}^n \binom{D_c(i)}{\omega_c(i,1), \ldots, \omega_c(i,n)} \binom{D_{\bar{c}}(i)}{\omega_{\bar{c}}(i,1), \ldots, \omega_{\bar{c}}(i,n)}$$

distinct ways. Once these subsets are chosen there remain, for every i, j, $\omega_c(i,j)!$ distinct bijections producing the same graph. Therefore, the total number of bijections from W_c to $\bar{W}_c$ which preserve the numbers $\omega_c(i,j) = \omega_{\bar{c}}(j,i)$ is given by

$$\prod_{i=1}^n \binom{D_c(i)}{\omega_c(i,1), \ldots, \omega_c(i,n)} \binom{D_{\bar{c}}(i)}{\omega_{\bar{c}}(i,1), \ldots, \omega_{\bar{c}}(i,n)} \prod_{i,j} \omega_c(i,j)!$$

The latter expression can be rewritten as (29). □

4.2.3 Generalized configuration model

We now define the configuration model for a generic degree sequence $\mathbf{D} \in \mathcal{D}_n$ by putting together the configuration models for all the colors. Let Σ denote the cartesian product of Σ_c, $c \in \mathcal{C}_\leqslant$, where, as defined above, Σ_c are the sets of configurations associated to the degree sequence $D_c(1), \ldots, D_c(n)$, that is Σ_c is the set of matchings of W_c if $c \in \mathcal{C}_=$ and Σ_c is the set of bijections $W_c \mapsto \bar{W}_c$ if $c \in \mathcal{C}_<$. A configuration is an element $\sigma = (\sigma_c)_{c \in \mathcal{C}_\leqslant}$ of Σ. The map $\Gamma(\cdot) : \Sigma \mapsto \widehat{\mathcal{G}}(\mathbf{D})$ is defined by calling $\Gamma(\sigma)$ the multi-graph obtained by superposition of the multi-graphs $\Gamma(\sigma_c)$ defined above. The *configuration model*, denoted CM($\mathbf{D}$), is the law of $\Gamma(\sigma)$ when $\sigma \in \Sigma$ is chosen uniformly at random.

Lemma 4.4 *Let $\mathbf{D} \in \mathcal{D}_n$, G with distribution CM($\mathbf{D}$) and $H \in \widehat{\mathcal{G}}(\mathbf{D})$. We have*

$$\mathbb{P}(G = H) = \frac{\prod_{c \in \mathcal{C}} \prod_{i=1}^n D_c(i)!}{b(H) \prod_{c \in \mathcal{C}_<} S_c! \prod_{c \in \mathcal{C}_=} (S_c - 1)!!}, \tag{30}$$

where $S_c = \sum_{i=1}^n D_c(i)$, and $b(H)$ is defined by

$$b(H) = \prod_{c \in \mathcal{C}_<} \prod_{i,j} \omega_c(i,j)! \prod_{c \in \mathcal{C}_=} \prod_{i=1}^n (\omega_c(i,i)/2)! 2^{(\omega_c(i,i)/2)} \prod_{i<j} \omega_c(i,j)! \tag{31}$$

In particular, for any $h \geqslant 2$, if $\mathcal{G}(\mathbf{D}, h)$ is not empty, the law of G conditioned on $\mathcal{G}(\mathbf{D}, h)$ is the uniform distribution on $\mathcal{G}(\mathbf{D}, h)$.

Proof The cardinality of Σ is given by $\prod_{c \in \mathcal{C}_<} S_c! \prod_{c \in \mathcal{C}_=} (S_c - 1)!!$. Thus, it suffices to check that $\Gamma^{-1}(H)$ has cardinality $b(H)^{-1} \prod_{c \in \mathcal{C}} \prod_i D_c(u)!$. This follows from Lemmas 4.2 and 4.3 by observing that $|\Gamma^{-1}(H)| = \prod_{c \in \mathcal{C}_\leqslant} n_c(H_c)$, where H_c denotes the multi-graph H after all edges with color $c' \notin \{c, \bar{c}\}$ are removed. This proves (30). If $H \in \mathcal{G}(\mathbf{D}, h)$, $h \geqslant 2$, then $\omega_c(i, i) = 0$ and $\omega_c(i, j) \in \{0, 1\}$ for all $i, j \in [n]$ and $c \in \mathcal{C}$, so that $b(H) = 1$. This proves the last assertion. $\square$

4.3 Probability of having no cycles of length $\ell \leqslant h$

Fix $\theta \in \mathbb{N}$ and call $\mathcal{M}_L^{(\theta)}$ the set of $L \times L$ matrices with nonnegative integer entries bounded by θ. Fix $P \in \mathcal{P}(\mathcal{M}_L^{(\theta)})$, a probability on $\mathcal{M}_L^{(\theta)}$. We consider a sequence $\mathbf{D}^{(n)} = (D^{(n)}(u))_{u \in [n]} \in \mathcal{D}_n, n \geqslant 1$ such that

(H1) for all $u \in [n]$, $D^{(n)}(u) \in \mathcal{M}_L^{(\theta)}$;
(H2) as $n \to \infty$, $\frac{1}{n} \sum_{u=1}^n \delta_{D^{(n)}(u)} \rightsquigarrow P$.

The main result of this section is the following

Theorem 4.5 *Fix $\theta \in \mathbb{N}$, $P \in \mathcal{P}(\mathcal{M}_L^{(\theta)})$, and a sequence $\mathbf{D}^{(n)}$ satisfying (H1)–(H2). Take $G_n \in \widehat{\mathcal{G}}(\mathbf{D}^{(n)})$ with distribution $\mathrm{CM}(\mathbf{D}^{(n)})$. For every $h \in \mathbb{N}$, there exists $\alpha_h > 0$ such that*

$$\lim_{n \to \infty} \mathbb{P}\left(G_n \in \mathcal{G}(\mathbf{D}^{(n)}, h)\right) = \alpha_h. \tag{32}$$

The actual value of α_h could be in principle computed in terms of P (see proof of Theorem 4.5). We will however not need that.

Corollary 4.6 *In the setting of Theorem 4.5, writing $S_c^{(n)} = \sum_{u \in [n]} D_c^{(n)}(u)$, for all $h \geqslant 2$:*

$$|\mathcal{G}(\mathbf{D}^{(n)}, h)| \sim \alpha_h \frac{\prod_{c \in \mathcal{C}_<} S_c^{(n)}! \prod_{c \in \mathcal{C}_=} (S_c^{(n)} - 1)!!}{\prod_{c \in \mathcal{C}} \prod_u D_c^{(n)}(u)!}.$$

Proof By definition of $\mathrm{CM}(\mathbf{D}^{(n)})$, one has

$$\mathbb{P}\left(G_n \in \mathcal{G}(\mathbf{D}^{(n)}, h)\right) = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} \mathbf{1}\left(\Gamma(\sigma) \in \mathcal{G}(\mathbf{D}^{(n)}, h)\right).$$

As in Lemma 4.4, for each $H \in \mathcal{G}(\mathbf{D}^{(n)}, h)$, $|\Gamma^{-1}(H)| = \prod_{c \in \mathcal{C}} \prod_u D_c^{(n)}(u)!$. Hence the sum in the right hand side above equals $|\mathcal{G}(\mathbf{D}^{(n)}, h)| \prod_{c \in \mathcal{C}} \prod_u D_c^{(n)}(u)!$. The conclusion follows from Theorem 4.5 and $|\Sigma| = \prod_{c \in \mathcal{C}_<} S_c^{(n)}! \prod_{c \in \mathcal{C}_=} (S_c^{(n)} - 1)!!$. $\square$

The proof of Theorem 4.5 will follow a well known strategy; see e.g. Bollobás [7, proof of Theorem 2.16] for a similar result. Our first lemma computes the number of copies of a subgraph in a graph sampled from $\mathrm{CM}(\mathbf{D}^{(n)})$. To formulate it, we need to introduce some more notation. Let $\widehat{\mathcal{G}}_n$ denote the set of $G \in \widehat{\mathcal{G}}(\mathcal{C})$ with vertex set $[n]$. If $G \in \widehat{\mathcal{G}}_n$, and $H \in \widehat{\mathcal{G}}(\mathcal{C})$ has vertex set $V \subset [n]$, we let $Y(H, G)$ be the number of times that $H \subset G$. When G is not a simple graph, then $Y(H, G)$ may be larger than 1. Indeed, one has

$$Y(H, G) = \mathbf{1}(H \subset G) \prod_{c \in \mathcal{C}_<} \prod_{u,v} B_c^{H,G}(u, v) \prod_{c \in \mathcal{C}_=} \prod_{u \leqslant v} B_c^{H,G}(u, v) \qquad (33)$$

where we use the notation $B_c^{H,G}(u, v)$ for the binomial coefficient $\binom{\omega_c^G(u,v)}{\omega_c^H(u,v)}$, with the convention that if $u = v$ and $c \in \mathcal{C}_=$, then $B_c^{H,G}(u, u)$ equals $\binom{(\omega_c^G(u,u)/2)}{(\omega_c^H(u,u)/2)}$.

Next, for $G \in \widehat{\mathcal{G}}_n$ and $H \in \widehat{\mathcal{G}}_k$, $1 \leqslant k \leqslant n$, define $X(H, G)$ as the number of distinct subgraphs of G that are isomorphic to H. If $a(H)$ denotes the cardinality of the automorphism group of H, i.e. the number of permutations of the vertex labels which leave H invariant, then

$$X(H, G) = \frac{1}{a(H)} \sum_\tau Y(\tau(H), G), \qquad (34)$$

where the sum is over all injective maps τ from $[k]$ to $[n]$, and $\tau(H)$ represents the multi-graph obtained by embedding H in $[n]$ through τ.

For $H \in \widehat{\mathcal{G}}_k$, the c-degree at vertex u is denoted

$$d_c^H(u) = \sum_v \omega_c^H(u, v).$$

The *excess* of H is defined by

$$\mathrm{exc}(H) = \left(\frac{1}{2} \sum_{c \in \mathcal{C}} \sum_{i=1}^k d_c^H(i) \right) - k,$$

Notice that $\mathrm{exc}(H) = |E(H)| - k$, where $E(H)$ is the total number of edges of $\bar{H}$ (counting 1 for each loop) where $\bar{H}$ is the colorblind undirected multi-graph obtained from H by (25). Notice that for H connected, then $\mathrm{exc}(H) \geqslant -1$, and $\mathrm{exc}(H) = -1$ iff $\bar{H}$ is a tree. If $n \geqslant k$ are positive integers, we use the notation $(n)_k = n!/(n-k)!$ for the number of injective maps $[k] \mapsto [n]$, with $(n)_0 = 1$.

Lemma 4.7 *Let $G_n \in \widehat{\mathcal{G}}(\mathbf{D}^{(n)})$ with distribution $\mathrm{CM}(\mathbf{D}^{(n)})$, where $\mathbf{D}^{(n)}$ satisfies assumptions (H1)–(H2). For any fixed $k \in \mathbb{N}$, $H \in \widehat{\mathcal{G}}_k$, as $n \to \infty$:*

$$\mathbb{E}X(H, G_n) \sim \frac{\prod_{i=1}^k \mathbb{E} \prod_{c \in \mathcal{C}} (D_c)_{d_c^H(i)}}{a(H)b(H) \prod_{c \in \mathcal{C}} (\mathbb{E}D_c)^{s_c^H/2}} n^{-\mathrm{exc}(H)},$$

where $D \in \mathcal{M}_L^{(\theta)}$ has distribution P and $s_c^H := \sum_{i=1}^k d_c^H(i)$.

Proof From (34), $\mathbb{E}X(H, G_n) = a(H)^{-1} \sum_\tau \mathbb{E}Y(\tau(H), G_n)$. Below, we fix a map τ and write H instead of $\tau(H)$ for simplicity. We start by showing that

$$\mathbb{E}Y(H, G_n) = \frac{\prod_{c \in \mathcal{C}} \prod_{i=1}^k (D_c^{(n)}(i))_{d_c^H(i)}}{b(H) \prod_{c \in \mathcal{C}_<} (S_c^{(n)})_{s_c^H} \prod_{c \in \mathcal{C}_=} ((S_c^{(n)}))_{s_c^H}}, \qquad (35)$$

where we use the notation $((n))_k = (n-1)!!/(n-k-1)!!$. Since $\mathrm{CM}(\mathbf{D}^{(n)})$ is a product measure over $c \in \mathcal{C}_\leqslant$, we may analyze one color at a time.

Consider first the case $c \in \mathcal{C}_<$. Set

$$Y_c(H, G) = \mathbf{1}(H_c \subset G) \prod_{u,v} \binom{\omega_c^G(u, v)}{\omega_c^H(u, v)},$$

where H_c is the graph H with all edges removed except for edges of color c or $\bar{c}$, and the condition $G \supset H_c$ indicates that $\omega_c^G(u, v) \geqslant \omega_c^H(u, v)$ for all $u, v \in [n]$. Then, as in Lemma 4.4

$$\mathbb{E}Y_c(H, G_n) = \sum_{G: G \supset H_c} \frac{\prod_{i=1}^n D_c(i)! D_{\bar{c}}(i)!}{S_c! \prod_{i,j} \omega_c^G(i, j)!} \prod_{u,v} \binom{\omega_c^G(u, v)}{\omega_c^H(u, v)}$$

where we drop the superscript (n) from $D_c(i)$ and S_c, and the sum runs over all $G \in \widehat{\mathcal{G}}_n$ with $(c, \bar{c})$ colors only, with degree sequence given by $(D_c(i), D_{\bar{c}}(i))_{i \in [n]}$. Therefore,

$$\mathbb{E}Y_c(H, G_n) = \frac{\prod_{i=1}^n D_c(i)! D_{\bar{c}}(i)!}{S_c! \prod_{i,j} \omega_c^H(u, v)!} \sum_{G: G \supset H_c} \prod_{u,v} \frac{1}{(\omega_c^G(u, v) - \omega_c^H(u, v))!}$$

On the other hand, applying (29) to the multi-graph $G \setminus H$ defined by $(\omega_c^G(u, v) - \omega_c^H(u, v))$, one has

$$\sum_{G: G \supset H_c} \prod_{u,v} \frac{1}{(\omega_c^G(u, v) - \omega_c^H(u, v))!} = \frac{(S_c - s_c^H)!}{\prod_i (D_c(i) - d_c^H(i))!(D_{\bar{c}}(i) - d_{\bar{c}}^H(i))!}$$

Thus, for $c \in \mathcal{C}_<$ one has

$$\mathbb{E}Y_c(H, G_n) = \frac{\prod_i (D_c^{(n)}(i))_{d_c^H(i)} (D_{\bar{c}}^{(n)}(i))_{d_{\bar{c}}^H(i)}}{(S_c)_{s_c^H} \prod_{i,j} \omega_c^H(u, v)!}. \qquad (36)$$

Next, consider the case $c \in \mathcal{C}_=$. Here

$$Y_c(H, G) = \mathbf{1}(H_c \subset G) \prod_{u<v} \binom{\omega_c^G(u, v)}{\omega_c^H(u, v)} \prod_u \binom{(\omega_c^G(u, u)/2)}{(\omega_c^H(u, u)/2)},$$

where H_c is the graph H with all edges removed except for edges of color c. Then,

$$\mathbb{E}Y_c(H, G_n) = \sum_{G:\, G \supset H_c} \frac{\prod_i D_c(i)! \prod_{u<v} \binom{\omega_c^G(u,v)}{\omega_c^H(u,v)} \prod_u \binom{\omega_c^G(u,u)/2}{\omega_c^H(u,u)/2}}{(S_c - 1)!! \prod_{i<j} \omega_c^G(i,j)! \prod_i \left(\omega_c^G(i,i)/2\right)! 2^{\left(\omega_c^G(i,i)/2\right)}}$$

Applying (28) to the multi-graph $G \smallsetminus H$ and simplifying, one arrives at

$$\mathbb{E}Y_c(H, G_n) = \frac{\prod_i (D_c^{(n)}(i))_{d_c^H(i)}}{((S_c))_{s_c^H} \prod_{i<j} \omega_c^H(i,j)! \prod_i \left(\omega_c^H(i,i)/2\right)! 2^{\left(\omega_c^H(i,i)/2\right)}}. \tag{37}$$

Finally, taking products over $c \in \mathcal{C}_<$ of (36) together with products over $c \in \mathcal{C}_=$ of (37), we arrive at (35).

Summing over the injective maps $\tau : [k] \mapsto [n]$, we deduce that

$$\mathbb{E}X(H, G_n) = \frac{(n)_k\, \mathbb{E} \prod_{c \in \mathcal{C}} \prod_{i=1}^k (M_c(i))_{d_c^H(i)}}{a(H)b(H) \prod_{c \in \mathcal{C}_<} (S_c^{(n)})_{s_c^H} \prod_{c \in \mathcal{C}_=} ((S_c^{(n)}))_{s_c^H}}, \tag{38}$$

where $(M(1), \ldots, M(k))$ is uniformly sampled without replacement on $(\mathbf{D}^{(n)}(1), \ldots, \mathbf{D}^{(n)}(n))$. From assumptions (H1)–(H2), for every fixed k and $H \in \widehat{\mathcal{G}}_k$, as $n \to \infty$:

$$\mathbb{E} \prod_{c \in \mathcal{C}} \prod_{i=1}^k (M_c(i))_{d_c^H(i)} \to \prod_{i=1}^k \mathbb{E} \prod_{c \in \mathcal{C}} (D_c)_{d_c^H(i)},$$

where $D \in \mathcal{M}_L^{(\theta)}$ has law P. Moreover, for $c \in \mathcal{C}_<$ and $c \in \mathcal{C}_=$ respectively,

$$(S_c^{(n)})_{s_c^H} \sim n^{s_c^H} (\mathbb{E}D_c)^{s_c^H} \quad \text{and} \quad ((S_c^{(n)}))_{s_c^H} \sim n^{s_c^H/2} (\mathbb{E}D_c)^{s_c^H/2}.$$

The desired conclusion now follows by using these asymptotics in (38) together with $(n)_k \sim n^k$ and

$$\sum_{c \in \mathcal{C}_<} s_c^H + \frac{1}{2} \sum_{c \in \mathcal{C}_=} s_c^H = \frac{1}{2} \sum_{c \in \mathcal{C}} s_c^H = \mathrm{exc}(H) + k.$$

$\square$

Proof of Theorem 4.5 For every $\ell \in \mathbb{N}$, call $\mathcal{L}_\ell$ the set of all $H \in \widehat{\mathcal{G}}_\ell$ such that the undirected graph $\bar{H}$ defined by (25) is a cycle of length ℓ. If $\ell = 1$, then $\mathcal{L}_\ell$ is the union over $c \in \mathcal{C}$ of the single loop graph at vertex $\{1\}$ with color c, if $\ell = 2$, then $\mathcal{L}_\ell$ is the union over $c, c' \in \mathcal{C}$ of the double edge graph at vertices $\{1, 2\}$ with $\omega_c(1, 2) = \omega_{\bar{c}}(2, 1) = 1$, $\omega_{c'}(1, 2) = \omega_{\bar{c}'}(2, 1) = 1$, and so on. Let $\mathcal{L}_{\leqslant h} = \cup_{\ell=1}^h \mathcal{L}_\ell$.

We define the random variable

$$Z = \sum_{H \in \mathcal{L}_{\leqslant h}} X(H, G_n), \tag{39}$$

where $G_n \in \widehat{\mathcal{G}}(\mathbf{D}^{(n)})$ has distribution $\mathrm{CM}(\mathbf{D}^{(n)})$. With this notation, we need to show that under the assumptions of the theorem one has

$$\lim_{n \to \infty} \mathbb{P}(Z = 0) = \alpha_h, \tag{40}$$

for some $\alpha_h > 0$.

If $H \in \mathcal{L}_{\leqslant h}$, then $\mathrm{exc}(H) = 0$. By Lemma 4.7, for some $\lambda_H \geqslant 0$, as $n \to \infty$, one has

$$\mathbb{E}X(H, G_n) \to \lambda_H, \tag{41}$$

and, setting $\lambda(h) = \sum_{H \in \mathcal{L}_{\leqslant h}} \lambda_H$, one finds

$$\lim_{n \to \infty} \mathbb{E}Z = \lambda(h). \tag{42}$$

We are going to prove that Z converges weakly to a Poisson random variable with mean $\lambda(h)$. This will prove (40) with $\alpha_h = e^{-\lambda(h)}$. To this end, by the well known moment method, it is sufficient to prove that for any integer $p \geqslant 1$:

$$\lim_{n \to \infty} \mathbb{E}\left[(Z)_p\right] = \lambda(h)^p, \tag{43}$$

where $(Z)_p = Z!/(Z - p)!$. The case $p = 1$ is (42). Below, we establish (43) for all $p \geqslant 2$.

For any $H \in \mathcal{L}_{\leqslant h}$, let $\mathcal{H}_H$ denote the set of multi-graphs $F \in \widehat{\mathcal{G}}(\mathcal{C})$ with vertex set $V_F \subset [n]$ which are isomorphic to H. If $\mathcal{H} = \cup_{H \in \mathcal{L}_{\leqslant h}} \mathcal{H}_H$, then one has

$$Z = \sum_{F \in \mathcal{H}} Y_F,$$

where $Y_F := Y(F, G_n)$ is defined by (33). The proof of (43) uses two elementary topological facts:

(i) if $F \neq F' \in \mathcal{H}$ and $F \cap F' \neq \emptyset$, i.e. $V_F \cap V_{F'} \neq \emptyset$, then $\mathrm{exc}(F \cup F') \geqslant 1$,
(ii) if $H \in \widehat{\mathcal{G}}_k$ and $H' \in \widehat{\mathcal{G}}_{k'}$, then $\mathrm{exc}(H \cup H') \geqslant \mathrm{exc}(H) + \mathrm{exc}(H')$ and $\mathrm{exc}(H \oplus H') = \mathrm{exc}(H) + \mathrm{exc}(H')$,

where $H \oplus H' \in \widehat{\mathcal{G}}_{k+k'}$ is the multigraph obtained from the disjoint union of H and an isomorphic copy of H' with vertex set $\{k + 1, \ldots, k + k'\}$. We also use two consequences of Lemma 4.7:

(iii) if $H \in \widehat{\mathcal{G}}_k$ and $\mathrm{exc}(H) \geqslant 1$ then $\mathbb{E}X(H, G_n) = o(1)$;
(iv) if $H \in \widehat{\mathcal{G}}_k$ and $H' \in \widehat{\mathcal{G}}_{k'}$, then $\mathbb{E}X(H \oplus H', G_n) \sim \mathbb{E}X(H, G_n)\,\mathbb{E}X(H', G_n)$.

We start by showing that for all $q \geqslant 1$, there exists $c = c(q) > 0$ such that

$$\mathbb{E}\left[Z^q\right] \leqslant c. \tag{44}$$

Write

$$Z^q = \sum_{(F_1,\ldots,F_q)\in\mathcal{H}^q} \prod_{i=1}^q Y_{F_i}.$$

By assumption (H1), $Y_F \leqslant c_0$ for some $c_0 = c_0(\theta, h)$, and hence, for some $c_1 = c_1(\theta, h, q)$, one has the crude bound

$$Z^q \leqslant c_1 \sum_{k=1}^q \sum_* \prod_{i=1}^k Y_{F_i},$$

where the sum $\sum_*$ is over all choices of pairwise distinct $F_1, \ldots, F_k$ in $\mathcal{H}$. We now decompose $\sum_*$ into the sum $\sum_{**}$ over all choices of k pairwise disjoint sets F_i in $\mathcal{H}$, and the sum $\sum_{***}$ over all choices of k pairwise distinct F_i in $\mathcal{H}$ such there exists $i \neq j$ with $F_i \cap F_j \neq \emptyset$. Notice that this last summation satisfies

$$\sum_{***} \prod_{i=1}^k Y_{F_i} \leqslant c_0 \sum_K X(K, G_n)$$

for some $c_0 = c_0(\theta, h)$, where K ranges over a finite collection (with cardinality independent of n) of multi-graphs which by facts (i–ii) satisfy $\mathrm{exc}(K) \geqslant 1$. In particular, fact (iii) implies that $\mathbb{E} \sum_{***} \prod_{i=1}^k Y_{F_i} = o(1)$ as $n \to \infty$. On the other hand,

$$\sum_{**} \prod_{i=1}^k Y_{F_i} = \sum_{(H_1,\ldots,H_k)\in(\mathcal{L}_{\leqslant h})^k} X(H_1 \oplus \cdots \oplus H_k, G_n).$$

Fact (iv) and (41) then imply that

$$\mathbb{E} \sum_* \prod_{i=1}^k Y_{F_i} = \sum_{(H_1,\ldots,H_k)\in(\mathcal{L}_{\leqslant h})^k} \prod_{i=1}^k \lambda_{H_i} + o(1) = \lambda(h)^k + o(1). \tag{45}$$

This ends the proof of (44).

Next, define $\tilde{Y}_F = \mathbf{1}(Y_F = 1)$ and $\tilde{Z} = \sum_{F\in\mathcal{H}} \tilde{Y}_F$. Let E be the event that for all $F \in \mathcal{H}$, $Y_F = \tilde{Y}_F$. Note that $\tilde{Z} = Z$ if E holds and $\mathbf{1}_{E^c} \leqslant \sum_K X(K, G_n)$, where K ranges over a finite collection of multi-graphs with $\mathrm{exc}(K) \geqslant 1$. From fact (iii), it follows that $\mathbb{P}(E^c) = o(1)$.

Clearly, $\tilde{Z} \leqslant Z$ and $(Z)_p \leqslant Z^p$. Cauchy-Schwarz' inequality yields

$$\left| \mathbb{E}(Z)_p - \mathbb{E}(\tilde{Z})_p \right| \leqslant \mathbb{E}[(Z)_p \mathbf{1}_{E^c}] \leqslant \sqrt{\mathbb{E}(Z^{2p})\mathbb{P}(E^c)}.$$

Therefore, using (44) and $\mathbb{P}(E^c) = o(1)$, we see that it suffices to prove that $\mathbb{E}(\tilde{Z})_p$ converges to $\lambda(h)^p$. Since $\tilde{Y}_F \in \{0, 1\}$, we write

$$(\tilde{Z})_p = \sum_* \prod_{i=1}^{p} \tilde{Y}_{F_i},$$

where the sum $\sum_*$ is over all choices of p pairwise distinct F_i in $\mathcal{H}$. By assumption (H1), Y_F is uniformly bounded, and therefore

$$\sum_* \left| \prod_{i=1}^{p} \tilde{Y}_{F_i} - \prod_{i=1}^{p} Y_{F_i} \right|$$

can be bounded by $c_1 \sum_K X(K, G)$ where K ranges over a finite collection of multi-graphs with $\mathrm{exc}(K) \geqslant 1$ and $c_1 = c_1(\theta, h, p)$. Therefore from fact (iii) we get

$$\mathbb{E}(\tilde{Z})_p = \mathbb{E} \sum_* \prod_{i=1}^{p} Y_{F_i} + o(1).$$

The conclusion $\mathbb{E}(\tilde{Z})_p \to \lambda(h)^p, n \to \infty$, then follows from (45).   □

4.4 Unimodular Galton–Watson trees with colors

Let $\widehat{\mathcal{G}}^*(\mathcal{C})$ denote the set of equivalence classes of rooted directed locally finite colored multi-graphs, i.e. the set of connected multi-graphs $G \in \widehat{\mathcal{G}}(\mathcal{C})$ with a distinguished vertex o (the root) where two rooted multi-graphs are identified if they only differ by a relabeling of the vertices. An element of $\widehat{\mathcal{G}}^*(\mathcal{C})$ is called a rooted directed colored tree if the corresponding colorblind multi-graph defined via (25) has no cycles. We now introduce a probability measure on $\widehat{\mathcal{G}}^*(\mathcal{C})$ supported on rooted colored directed trees. Let $P \in \mathcal{P}(\mathcal{M}_L)$ be a probability measure on $\mathcal{M}_L$, $|\mathcal{C}| = L^2$, such that for all $c \in \mathcal{C}$,

$$\mathbb{E}D_c = \mathbb{E}D_{\bar{c}}, \tag{46}$$

where $D \in \mathcal{M}_L$ has distribution P. For each $c \in \mathcal{C}$ such that $\mathbb{E}D_c > 0$, define the probability measure $\widehat{P}^c \in \mathcal{P}(\mathcal{M}_L)$ such that, for $M \in \mathcal{M}_L$,

$$\widehat{P}^c(M) = \frac{(M_{\bar{c}} + 1) P(M + E^{\bar{c}})}{\mathbb{E}D_c},$$

where D has distribution P, and for any $c \in \mathcal{C}$, E^c denotes the matrix with all entries equal to 0 except for the entry at c, which equals 1. Notice that $\widehat{P^c}$ is indeed a probability since

$$\sum_{M \in \mathcal{M}_L} (M_{\bar{c}} + 1) \, P(M + E^{\bar{c}}) = \sum_{M \in \mathcal{M}_L} M_{\bar{c}} \, P(M) = \mathbb{E}D_{\bar{c}} = \mathbb{E}D_c.$$

If $\mathbb{E}D_c = 0$ then we set $\widehat{P^c}(M) = \mathbf{1}(M = 0)$.

In a rooted directed colored tree (T, o), for all $v \neq o$, call $a(v)$ the parent of v in T. The *type* of a vertex $v \neq o$ in (T, o) is defined as the color of the edge $(a(v), v)$. The probability measure $\mathrm{UGW}(P) \in \mathcal{P}(\widehat{\mathcal{G}}^*(\mathcal{C}))$ is the law of the multi-type Galton–Watson tree defined as follows. The root o produces offspring according to the distribution P, i.e. the root has D_c children of type c, for all $c \in \mathcal{C}$, where $D \in \mathcal{M}_L$ has law P. Recursively, and independently, any $v \neq o$ of type c, produces offspring according to the distribution $\widehat{P^c}$, i.e. v has $D_{c'}$ children of type c', for all $c' \in \mathcal{C}$, where $D \in \mathcal{M}_L$ has law $\widehat{P^c}$. Notice that in the case of a single color ($L = 1$ and $\mathcal{C} = \{(1, 1)\}$), then P is a probability measure on $\mathbb{Z}_+$ and $\mathrm{UGW}(P)$ coincides with the Galton–Watson tree $\mathrm{UGW}_1(P)$ with degree distribution P, cf. (6).

Following the argument of Lemma 3.1, it could be proved that the measure $\mathrm{UGW}(P)$ is unimodular. However, in the next paragraph, Theorem 4.8 implies that $\mathrm{UGW}(P)$ is sofic (and hence unimodular).

4.5 Local weak convergence

It is straightforward to extend the local topology introduced in Sect. 2 to the case of rooted directed multi-graphs with colored edges $\widehat{\mathcal{G}}^*(\mathcal{C})$. The only difference is that the weight function ω is now matrix-valued.

Theorem 4.8 *If $G_n \in \widehat{\mathcal{G}}(\mathbf{D}^{(n)})$ has distribution $\mathrm{CM}(\mathbf{D}^{(n)})$, with $\mathbf{D}^{(n)}$ such that assumptions (H1)–(H2) hold, then with probability one $U(G_n) \rightsquigarrow \mathrm{UGW}(P)$. Moreover the same result holds if G_n is uniformly sampled on $\mathcal{G}(\mathbf{D}^{(n)}, h)$, for any fixed $h \geqslant 2$.*

In the case of a single color $L = 1$, Theorem 4.8 is folklore; see e.g. the monographs [18,23]. The proof of Theorem 4.8 in the general case is given in the appendix.

4.6 Graphs with given tree-like neighborhood

Here we show how the configuration model can be used to count the number of graphs with a given tree-like neighborhood structure.

Fix n and a graph $G = (V, E)$ with $V = [n]$. Call $\mathcal{G}_n$ the set of all such graphs. For $h \in \mathbb{N}$, define the h-neighborhood vector

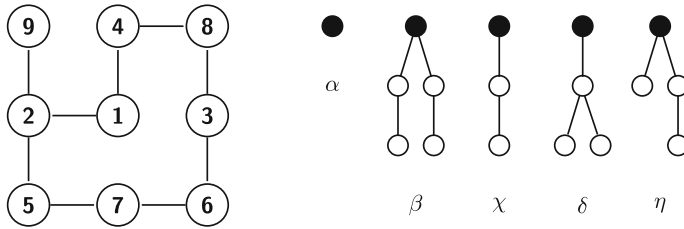$$\psi_h(G) = ([G, 1]_h, \ldots, [G, n]_h), \tag{47}$$

Fig. 4 A 3-tree-like graph $G \in \mathcal{G}_9$ (*left*). When $h = 3$, the associated set $\mathcal{F}$ of equivalence classes is given by the $L = 5$ rooted unlabeled graphs depicted on the *right* (the *black* vertex is the root). Here $G(2, 9)_2 = (\beta, \alpha)$, $G(9, 2)_2 = (\alpha, \beta)$; $G(1, 2)_2 = (\chi, \eta)$, $G(2, 1)_2 = (\eta, \chi)$; $G(5, 2)_2 = (\chi, \eta)$, $G(2, 5)_2 = (\eta, \chi)$; $G(7, 5)_2 = (\chi, \delta)$, $G(5, 7)_2 = (\delta, \chi)$; $G(6, 7)_2 = G(7, 6)_2 = G(3, 6)_2 = G(6, 3)_2 = G(3, 8)_2 = G(8, 3)_2 = G(4, 8)_2 = G(8, 4)_2 = (\chi, \chi)$; $G(4, 1)_2 = (\chi, \delta)$, $G(1, 4)_2 = (\delta, \chi)$

Fig. 5 The graph $\widetilde{G} \in \widehat{\mathcal{G}}(\mathcal{C})$ defined by (48) when G is the graph from Fig. 4. It is understood that if the directed edge (u, v) has color $(g, g') \in \mathcal{C}$, then the opposite edge (v, u) has color (g', g)
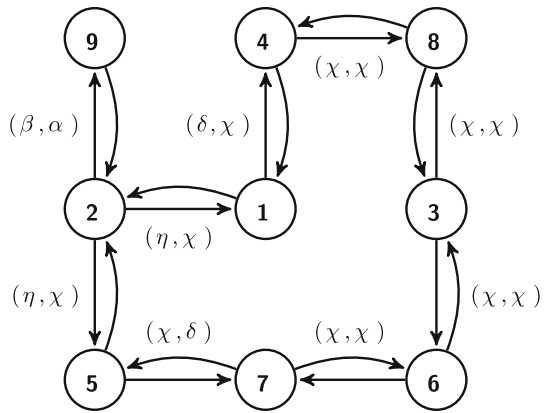
where $[G, u]_h$ stands for the equivalence class of the h-neighborhood of G at vertex u. We say that G is h-tree-like if $[G, u]_h$ is a tree for all $u \in [n]$.

We describe now a procedure which turns the given graph G into a directed colored graph $\widetilde{G}$ in $\mathcal{G}(\mathcal{C})$. The color set $\mathcal{C}$ is defined as follows. Let $\mathcal{F} \subset \mathcal{G}_{h-1}^*$ denote the collection of all equivalence classes of the subgraphs $G(u, v)_{h-1}$, where we recall that $G(u, v)$ is the rooted graph obtained from G by removing the edge $\{u, v\}$ and taking the root at v. For simplicity, below we will identify $G(u, v)_{h-1}$ with its equivalence class. If $L = |\mathcal{F}|$ denotes the cardinality of $\mathcal{F}$, we call $\mathcal{C}$ the set of L^2 pairs (g, g'), with $g, g' \in \mathcal{F}$; see Fig. 4 for an example. To construct the directed colored graph, for every pair u, v such that $\{u, v\}$ is an edge of G, we include a directed edge (u, v) with color

$$(g, g') = (G(u, v)_{h-1}, G(v, u)_{h-1}), \tag{48}$$

together with the directed edge (v, u) with color $(g', g) = (G(v, u)_{h-1}, G(u, v)_{h-1})$. This defines an element $\widetilde{G}$ of $\mathcal{G}(\mathcal{C})$; see Fig. 5. As such, we can define its degree sequence $\mathbf{D} = \mathbf{D}(\widetilde{G})$ as in (26) above. Notice that if G is h-tree-like, then the above construction yields an element of $\mathcal{G}(\mathbf{D}, 2h + 1)$ since being h-tree-like is equivalent to having no cycles with length $1 \leqslant \ell \leqslant 2h + 1$; see Fig. 6. A crucial property to be
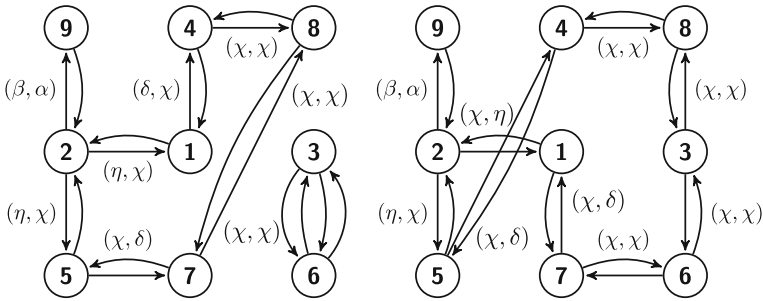
Fig. 6 Two examples of multigraphs $\Gamma_1, \Gamma_2 \in \widehat{\mathcal{G}}(\mathbf{D})$, where $\mathbf{D} = \mathbf{D}(\widetilde{G})$ is the degree sequence of $\widetilde{G}$ from Fig. 5. Notice that Γ_1 (left) yields a colorblind multigraph $\bar{\Gamma}_1$ with a double edge at $\{3, 6\}$, while Γ_2 (right) yields a 3-tree-like graph $\bar{\Gamma}_2$, i.e. $\Gamma_2 \in \mathcal{G}(\mathbf{D}, 7)$. In particular, as guaranteed by Lemma 4.9, $\bar{\Gamma}_2$ has the same 3-neighborhoods of G

used below is that, for this particular choice of $\mathbf{D}$, *all* elements of $\mathcal{G}(\mathbf{D}, 2h + 1)$ have the same h-neighborhoods.

Lemma 4.9 *Let $h \in \mathbb{N}$, let $G \in \mathcal{G}_n$ be a fixed h-tree-like graph and let $\mathbf{D} = \mathbf{D}(\widetilde{G})$ be the associated degree sequence as above. For any $\Gamma \in \mathcal{G}(\mathbf{D}, 2h + 1)$, the colorblind graph $\bar{\Gamma} \in \mathcal{G}_n$ defined via (25) satisfies $\psi_h(\bar{\Gamma}) = \psi_h(G)$.*

Proof Consider first the case $h = 1$. If $\Gamma \in \mathcal{G}(\mathbf{D}, 3)$, then for any node $i \in [n]$, the 1-neighborhood $(\bar{\Gamma}, i)_1$ at i is uniquely determined by the number of edges exiting node i. By (25), this number equals $\sum_{c \in \mathcal{C}} D_c(i)$, which is independent of Γ. Thus, all $\Gamma \in \mathcal{G}(\mathbf{D}, 3)$ satisfy necessarily $\psi_1(\bar{\Gamma}) = \psi_1(G)$.

Next, we assume that any $\Gamma \in \mathcal{G}(\mathbf{D}, 2h + 1)$ satisfies $\psi_{h-1}(\bar{\Gamma}) = \psi_{h-1}(G)$, and show that $\psi_h(\bar{\Gamma}) = \psi_h(G)$. Since $\mathcal{G}(\mathbf{D}, 2h + 1) \subset \mathcal{G}(\mathbf{D}, 2(h - 1) + 1)$, by induction over h this will prove the desired result.

Since there are no cycles of length $\ell \leqslant 2h + 1$ in G, $\mathcal{F}$ consists of unlabeled rooted trees of depth $h - 1$. For any $t \in \mathcal{F}$ we write t_k for the k-neighborhood of the root in t (truncation of t at depth k). Moreover, if t is a rooted tree, we write $t_{k,+}$ for the unlabeled rooted tree of depth $k + 1$ obtained from t_k by adding a new edge to the root and taking the other endpoint of that edge as the new root. If t, t' are finite rooted trees, we write $t \cup t'$ for the rooted tree obtained by joining t, t' at the common root. Since there are no cycles of length $\ell \leqslant 2h + 1$ in $\bar{\Gamma}$, to prove $\psi_h(\bar{\Gamma}) = \psi_h(G)$ it is sufficient to show that for any edge (u, v) with color (t, t') in Γ, with $t, t' \in \mathcal{F}$, one has $\bar{\Gamma}(u, v)_{h-1} = t'$ and $\bar{\Gamma}(v, u)_{h-1} = t$.

Let (u, v) be an edge in Γ with color (t, t'). Notice that in $\widetilde{G}$, u must have an edge $(u, \widetilde{v})$ with color (t, t') going out of u, and v must have an edge $(v, \widetilde{u})$ with color (t', t) going out of v. Therefore, $[G, u]_{h-1} = t \cup t'_{h-2,+}$ and $[G, v]_{h-1} = t' \cup t_{h-2,+}$. By assumption, $(\bar{\Gamma}, u)_{h-1} = [G, u]_{h-1}$ and $(\bar{\Gamma}, v)_{h-1} = [G, v]_{h-1}$. Therefore, the rooted trees $T := \bar{\Gamma}(v, u)_{h-1}$ and $T' := \bar{\Gamma}(u, v)_{h-1}$ must satisfy

$$T \cup T'_{h-2,+} = t \cup t'_{h-2,+}, \qquad T' \cup T_{h-2,+} = t' \cup t_{h-2,+}. \tag{49}$$

We need to show that $t = T$ and $t' = T'$. From (49), one has that it is sufficient to show that $T'_{h-2} = t'_{h-2}$ and $T_{h-2} = t_{h-2}$. Truncating (49) at depth $h - 2$ one has

$$T_{h-2} \cup T'_{h-3,+} = t_{h-2} \cup t'_{h-3,+}, \qquad T'_{h-2} \cup T_{h-3,+} = t'_{h-2} \cup t_{h-3,+}.$$

Thus, it is sufficient to show that $T'_{h-3} = t'_{h-3}$ and $T_{h-3} = t_{h-3}$. Iterating this reasoning, one finds that it suffices to show that $T'_1 = t'_1$ and $T_1 = t_1$. However, this is guaranteed by the fact that the degree of u in G and Γ is the same, for any $u \in [n]$. □

We turn to the problem of counting the number of graphs $G' \in \mathcal{G}_n$ whose h-neighborhood distribution coincides with that of a given h-tree-like graph G. The following is an important corollary of Lemma 4.9.

Corollary 4.10 *Fix an arbitrary h-tree-like graph $G \in \mathcal{G}_n$, and define*

$$N_h(G) = \big|\big\{G' \in \mathcal{G}_n : U(G')_h = U(G)_h\big\}\big|. \tag{50}$$

One has

$$N_h(G) = n(\mathbf{D})|\mathcal{G}(\mathbf{D}, 2h + 1)|, \tag{51}$$

where $\mathbf{D} = (D(1), \ldots, D(n))$ is the degree sequence associated to G via (48), and $n(\mathbf{D})$ denotes the number of distinct vectors $(D(\pi_1), \ldots, D(\pi_n)) \in \mathcal{D}_n$ as $\pi : [n] \mapsto [n]$ ranges over permutations of the labels.

Proof For a permutation $\pi : [n] \mapsto [n]$, let $\mathbf{D}^\pi = (D(\pi_1), \ldots, D(\pi_n))$. Since the cardinality of $\mathcal{G}(\mathbf{D}^\pi, 2h + 1)$ does not depend on π, $n(\mathbf{D})|\mathcal{G}(\mathbf{D}^\pi, 2h + 1)|$ coincides with the cardinality of $\cup_\pi \mathcal{G}(\mathbf{D}^\pi, 2h + 1)$. By Lemma 4.9, any two distinct elements $\Gamma_1, \Gamma_2 \in \cup_\pi \mathcal{G}(\mathbf{D}^\pi, 2h + 1)$ yield two distinct graphs $\bar{\Gamma}_1, \bar{\Gamma}_2$ such that $U(\bar{\Gamma}_i)_h = U(G)_h$, $i = 1, 2$. This proves that $N_h(G) \geqslant n(\mathbf{D})|\mathcal{G}(\mathbf{D}, 2h + 1)|$. On the other hand, any two distinct elements $G_1, G_2 \in \mathcal{G}_n$ with $U(G_i)_h = U(G)_h$, $i = 1, 2$, yield two distinct elements $\widetilde{G}_1, \widetilde{G}_2 \in \cup_\pi \mathcal{G}(\mathbf{D}^\pi, 2h + 1)$ with the map $G \mapsto \widetilde{G}$ defined by (48). This proves the other direction. □

Lemma 4.11 *Fix $h \in \mathbb{N}$, and $P \in \mathcal{P}(\mathcal{T}_h^*)$ admissible, with finite support and $\mathbb{E}_P \deg(o) = d$. Let $m = m(n)$ be a sequence such that $m/n \to d/2$ as $n \to \infty$. Then, there exist a finite set $\Delta \subset \mathcal{T}_h^*$ and a sequence of graphs $\Gamma_n \in \mathcal{G}_{n,m}$ such that the support of $U(\Gamma_n)_h$ is contained in Δ for all n and $U(\Gamma_n)_h \rightsquigarrow P$ as $n \to \infty$.*

Proof Let $S := \{t_1, \ldots, t_r\} \subset \mathcal{T}_h^*$ be the finite support of P. We define the vector $\mathbf{g}^{(n)} = (g^{(n)}(1), \ldots, g^{(n)}(n))$ with $g^{(n)}(i) \in S$ by setting $g^{(n)}(i) = t_k$ if $\sum_{\ell \leqslant k} P(t_\ell) > (i - 1)/n$ and $\sum_{\ell \leqslant k-1} P(t_\ell) \leqslant (i - 1)/n$ with the convention that the sum over an empty set is $-\infty$. The empirical measure of $\mathbf{g}^{(n)}$, say $P^{(n)}$, converges weakly to P.

Let $\mathcal{C}$ denote the set of all pairs $c = (t, t') \in \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*$ associated to any element $g \in S$ as in (48). In this manner, we associate to any $g^{(n)}(i)$ an integer valued matrix $D^{(n)}(i) \in \mathcal{M}_L$ where $|\mathcal{C}| = L^2$. We denote by $S_c^{(n)} = \sum_{i=1}^n D_c^{(n)}(i)$. We finally

set, for $c \in \mathcal{C}_=$, $\widetilde{S}_c^{(n)} = 2\lfloor S_c^{(n)}/2 \rfloor$ and, for $c \in \mathcal{C}_{\neq}$, $\widetilde{S}_c^{(n)} = S_c^{(n)} \wedge S_{\bar{c}}^{(n)}$. We may fix a sequence of integer-valued matrices $\widetilde{\mathbf{D}}^{(n)} = (\widetilde{D}^{(n)}(i))_{1 \leqslant i \leqslant n}$ such that component-wise $\widetilde{D}^{(n)}(i) \leqslant D^{(n)}(i)$ and, for all $c \in \mathcal{C}$, $\sum_{i=1}^{n} \widetilde{D}_c^{(n)}(i) = \widetilde{S}_c^{(n)}$. The properties $P^{(n)} \rightsquigarrow P$ and $\mathrm{supp}(P^{(n)}) \subset S$ imply that for all $c \in \mathcal{C}$, $\widetilde{S}_c^{(n)} - S_c^{(n)} = o(n)$ and for all but $o(n)$ vertices $\widetilde{D}^{(n)}(i) = D^{(n)}(i)$. Moreover,

$$\widetilde{m} = \frac{1}{2} \sum_{c \in \mathcal{C}} \widetilde{S}_c^{(n)} = m + o(n).$$

We consider the generalized configuration model on $\widetilde{\mathbf{D}}^{(n)}$. Corollary 4.6 implies the existence, for all n large enough, of an directed colored graph $\widetilde{\Gamma}_n$ with girth at least $2h+1$ and whose colored degree sequence is precisely given by $\widetilde{\mathbf{D}}^{(n)}$. Let $\bar{\Gamma}_n$ be the associated color-blind graph. The proof of Lemma 4.9 actually shows that if a vertex v of $\widetilde{\Gamma}_n$ is such that all vertices u in $(\widetilde{\Gamma}_n, v)_h$ satisfy $\widetilde{D}_c^{(n)}(u) = D_c^{(n)}(u)$ then the equivalence class of $(\bar{\Gamma}_n, v)_h$ is precisely $g^{(n)}(v)$. Now, let θ be the maximal degree of vertices in $t \in S$ and set $\kappa = \sum_{\ell=0}^{h} \theta^h$. Any vertex is in the h-neighborhood of at most κ vertices. Since for all but $o(n)$ vertices $\widetilde{D}^{(n)}(v) = D^{(n)}(v)$, we deduce that for all but $o(n)$ vertices, the equivalence class of $(\bar{\Gamma}_n, v)_h$ is $g^{(n)}(v)$. We thus have proved that $U(\bar{\Gamma}_n)_h \rightsquigarrow P$. Also, by construction, the support of $U(\bar{\Gamma}_n)_h$ is contained in the finite set $\Delta_{h,\theta}$ of unlabeled rooted trees $t \in \mathcal{T}_h^*$ such that all degrees of vertices in t are bounded by θ.

A last modification is needed: we have $\bar{\Gamma}_n \in \mathcal{G}_{n,\widetilde{m}}$ and we need a graph $\Gamma_n \in \mathcal{G}_{n,m}$. However, since the number of vertices in $(\bar{\Gamma}_n, v)_h$ is bounded by κ, adding or removing one edge in $\bar{\Gamma}_n$ will change the value of $(\bar{\Gamma}_n, v)_h$ for at most 2κ vertices. Let $\delta(n) = |\widetilde{m} - m| = o(n)$. Assume first that $\widetilde{m} < m$, then we need to add edges to $\bar{\Gamma}_n$. We may add $\delta(n)$ new edges to $\bar{\Gamma}_n$ such that any vertex has a most one new adjacent edge. From what precedes, we obtain a graph $\Gamma_n \in \mathcal{G}_{n,m}$ such that $U(\Gamma_n)_h \rightsquigarrow P$. Moreover the support $U(\Gamma_n)_h$ is contained in $\Delta_{h,\theta+1}$. If $\widetilde{m} > m$, we need to remove edges. We remove an arbitrary subset of them of cardinality $\delta(n)$. We get a graph $\Gamma_n \in \mathcal{G}_{n,m}$ such that $U(\Gamma_n)_h \rightsquigarrow P$ and the support of $U(\Gamma_n)_h$ is contained in $\Delta_{h,\theta}$. $\qquad \square$

4.7 Proof of Corollary 1.5

We note that the set, say $\mathcal{S}$, of sofic measures supported on trees is a closed subset of $\mathcal{P}_u(\mathcal{T}^*)$. Let B be the set of measures of the form $\rho = \mathrm{UGW}_h(P)$ with $P \in \mathcal{P}(\mathcal{T}_h^*)$ admissible with finite support and $h \in \mathbb{N}$. A consequence of Lemma 4.11 and Theorem 4.8 is that B is a subset of $\mathcal{S}$.

Let us first check that for any $h \in \mathbb{N}$ and $P \in \mathcal{P}(\mathcal{T}_h^*)$ admissible, $\rho = \mathrm{UGW}_h(P) \in \mathcal{S}$. For each $n \in \mathbb{N}$, consider the forest F_n obtained from (T, o), with law ρ, by removing all edges adjacent to a vertex with degree higher than n. We may define $\rho^{(n)}$ as the the law of $(F_n(o), o)$, the connected component of the root. It is easy to check that $\rho^{(n)}$ is a unimodular measure. We define $Q_n = \rho_h^{(n)}$, the law of its h-neighborhood. By construction, $\mathrm{UGW}_h(Q_n) \in B$ and Q_n converges weakly to P. We deduce that $\mathrm{UGW}_h(Q_n) \rightsquigarrow \mathrm{UGW}_h(P)$ and $\mathrm{UGW}_h(P) \in \mathcal{S}$.

Moreover, if $\rho \in \mathcal{P}_u(\mathcal{T}^*)$, then $\mathrm{UGW}_h(\rho_h) \rightsquigarrow \rho$, as $h \to \infty$. From what precedes $\mathrm{UGW}_h(\rho_h) \in \mathcal{S}$. Therefore, $\rho \in \mathcal{S}$ and $\mathcal{S} = \mathcal{P}_u(\mathcal{T}^*)$.

5 Graph counting and entropy

In this section we prove Theorem 1.2 and Theorem 1.3. The strategy will be as follows. We first establish the cases $\Sigma(\rho) = -\infty$ in Theorem 1.2. We then prove Theorem 1.3, and later complete the proof of Theorem 1.2. In what follows, we fix $d > 0$ and a sequence $m = m(n)$ such that $m/n \to d/2$ as $n \to \infty$.

5.1 Measures with $\Sigma(\rho) = -\infty$

Since unimodular measures form a closed subset of $\mathcal{P}(\mathcal{G}^*)$, if $\rho \notin \mathcal{P}_u(\mathcal{G}^*)$, then for some $\varepsilon > 0$ one has $B(\rho, \varepsilon) \subset \mathcal{P}(\mathcal{G}^*) \backslash \mathcal{P}_u(\mathcal{G}^*)$. Since $U(G_n) \in \mathcal{P}_u(\mathcal{G}^*)$, then $|\mathcal{G}_{n,m}(\rho, \varepsilon)| = 0$. Therefore $\overline{\Sigma}(\rho) = -\infty$ for all $\rho \notin \mathcal{P}_u(\mathcal{G}^*)$.

Next, we show that $\overline{\Sigma}(\rho) = -\infty$ whenever $\mathbb{E}_\rho \deg(o) \neq d$. We start with the case $\mathbb{E}_\rho \deg(o) > d$. Let $\rho \in \mathcal{P}_u(\mathcal{G}^*)$ and assume that $\overline{\Sigma}(\rho) > -\infty$. Then, by an extraction argument, there must exist a sequence of graphs $G_n \in \mathcal{G}_{n,m}$ such that $U(G_n) \rightsquigarrow \rho$. Weak convergence then implies that $\mathbb{E}_{U(G_n)}[\deg(o) \wedge t] \to \mathbb{E}_\rho[\deg(o) \wedge t]$ for any $t > 0$, and therefore, letting $n \to \infty$ and then $t \to \infty$:

$$\liminf_{n \to \infty} \mathbb{E}_{U(G_n)} \deg(o) \geqslant \mathbb{E}_\rho \deg(o).$$

On the other hand, by construction,

$$\mathbb{E}_{U(G_n)} \deg(o) = \frac{1}{n} \sum_{v=1}^{n} \deg_{G_n}(v) = \frac{2m}{n} = d + o(1). \tag{52}$$

We thus have checked that if $\mathbb{E}_\rho \deg(o) > d$, then $\overline{\Sigma}(\rho) = -\infty$.

The case $\mathbb{E}_\rho \deg(o, G) < d$ requires a little more care.

Lemma 5.1 *If* $\mathbb{E}_\rho \deg(o) < d$, *then* $\overline{\Sigma}(\rho) = -\infty$.

Proof From (7), it is sufficient to prove that, for any sequence $\varepsilon_n \to 0$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(U(G_n) \in B(\rho, \varepsilon_n)) = -\infty, \tag{53}$$

where G_n is a uniform random graph in $\mathcal{G}_{n,m}$. Define $d' = \mathbb{E}_\rho \deg(o)$ and $\delta = d - d' > 0$. If $U(G_n) \in B(\rho, \varepsilon_n)$ for all n, then for any $t > 0$:

$$\mathbb{E}_{U(G_n)}[\deg(o)\mathbf{1}(\deg(o) \leqslant t)] \to \mathbb{E}_\rho[\deg(o)\mathbf{1}(\deg(o) \leqslant t)].$$

Therefore, for some sequence $t_n \to \infty$, one has

$$\frac{1}{n} \sum_{v \in [n]} \deg_{G_n}(v)\mathbf{1}(\deg_{G_n}(v) \leqslant t_n) \to d'.$$

Define $A_n = \{i \in [n] : \deg_{G_n}(i) > t_n\}$. Using (52) one has

$$\frac{1}{n} \sum_{v \in A_n} \deg_{G_n}(v) \to \delta.$$

On the other hand, by Markov's inequality and (52), the cardinality of A_n satisfies $|A_n| \leqslant \alpha_n n$, where $\alpha_n = 2d/t_n$ for all n large enough. Thus $U(G_n) \in B(\rho, \varepsilon_n)$ implies that there exists $S \subset [n]$ with $|S| \leqslant \alpha_n n$ such that $\deg_{G_n}(S) := \sum_{v \in S} \deg_{G_n}(v)$ is larger than $\delta n/2$ for all n large enough. By the union bound one has

$$\mathbb{P}(U(G_n) \in B(\rho, \varepsilon_n)) \leqslant \binom{n}{\alpha_n n} \mathbb{P}\big(\deg_{G_n}([\alpha_n n]) \geqslant \delta n/2\big),$$

where $[\alpha_n n] = \{1, \ldots, \alpha_n n\}$. Next, we check that

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\big(\deg_{G_n}([\alpha_n n]) \geqslant \delta n/2\big) = -\infty. \tag{54}$$

To this end, observe that $\deg_{G_n}([\alpha_n n])$ is stochastically dominated by $2N$, where N denotes the binomial random variable $N = \mathrm{Bin}(\alpha_n n^2, 2d/n)$. Indeed, the number of potential edges incident to the set $[\alpha_n n]$ is trivially bounded by $\alpha_n n^2$ and each potential edge can be included in G_n recursively, where at each step the probability of inclusion is bounded above by

$$\frac{m}{\binom{n}{2} - \alpha_n n^2} \leqslant \frac{2d}{n},$$

if n is large enough, where we use $m/n \to d/2$ and $\alpha_n \to 0$. Therefore, from Chernov's bound, for any $x > 0$,

$$\mathbb{P}\big(\deg_{G_n}([\alpha_n n]) \geqslant \delta n/2\big) \leqslant \mathbb{P}(2N \geqslant \delta n/2) \leqslant e^{-\delta n x} \mathbb{E}[e^{4xN}]$$
$$= e^{-\delta n x}\Big(1 + (2d/n)(e^{4x} - 1)\Big)^{\alpha_n n^2} \leqslant e^{-\delta n x + 2d\alpha_n n e^{4x}}.$$

Taking e.g. $x = -\frac{1}{4} \log \alpha_n$, one obtains (54). Moreover, Stirling's formula implies

$$\frac{1}{n} \log \binom{n}{n\alpha_n} \sim -\alpha_n \log \alpha_n \to 0.$$

This implies (53). □

We turn to the claim that $\Sigma(\rho) = -\infty$ whenever ρ is not supported on trees.

Lemma 5.2 *Suppose $\rho \in \mathcal{P}_u(\mathcal{G}^*)$ is such that $\rho(\mathcal{T}^*) < 1$. Then there exists $\varepsilon_0 > 0$ such that if $0 < \varepsilon < \varepsilon_0$, then*

$$\limsup_{n \to \infty} \frac{\log |\mathcal{G}_{n,m}(\rho, \varepsilon)|}{n \log n} < \frac{d}{2}. \tag{55}$$

In particular, $\overline{\Sigma}(\rho, \varepsilon) = -\infty$, for any $0 < \varepsilon < \varepsilon_0$.

Proof Once (55) is established, the last assertion follows from (7) and $m/n \to d/2$. Let us prove (55). By assumption, there exist integers t and $\ell \geqslant 3$ such that

$$\mathbb{P}_\rho((G, o)_t \text{ contains a cycle of length } \ell) > 0.$$

For integer $k \geqslant 2$, let us say that a cycle is a (k, ℓ)-cycle if its length is ℓ and the degree of all vertices on the cycle is bounded by k. Since $(G, o)_t$ is ρ-a.s. locally finite, there exists an integer $k \geqslant 2$ such that

$$\mathbb{P}_\rho((G, o)_t \text{ contains a } (k, \ell)\text{-cycle}) > 0.$$

Consider the function $f(G, o, v) = \mathbf{1}(\text{dist}_G(o, v) \leqslant t ; v \text{ is in a } (k, \ell)\text{-cycle})$. From what precedes

$$\mathbb{E}_\rho \sum_{v \in V(G)} f(G, o, v) > 0.$$

Since ρ is unimodular, equation (2) applied to f implies that for some $\eta > 0$,

$$\mathbb{P}_\rho(o \text{ is in a } (k, \ell)\text{-cycle}) > 2\eta.$$

Thus, if $G \in \mathcal{G}_{n,m}(\rho, \varepsilon)$ and ε is small enough,

$$\mathbb{P}_{U(G)}(o \text{ is in a } (k, \ell)\text{-cycle}) > \eta.$$

By definition of $U(G)$, this implies that the number of vertices in a (k, ℓ)-cycle in G is at least ηn. Since degrees are bounded by k in a (k, ℓ)-cycle, we deduce that G contains at least δn mutually disjoint cycles of length ℓ, for some $\delta = \delta(\ell, k) > 0$. Therefore,

$$\left| \mathcal{G}_{n,m}(\rho, \varepsilon) \right| \leqslant C_{n,\ell} \left| \mathcal{G}_{n,m-\ell\lceil \delta n \rceil} \right|,$$

where $C_{n,\ell}$ is the number of ways to place $\lceil \delta n \rceil$ disjoint cycles of length ℓ on n vertices. One has

$$C_{n,\ell} \leqslant \frac{(n)_{\ell\lceil \delta n \rceil}}{\lceil \delta n \rceil!} \leqslant \frac{n^{\ell\lceil \delta n \rceil}}{\lceil \delta n \rceil!}.$$

Indeed $(n)_{\ell\lceil\delta n\rceil}$ counts the number of ordered choices of the ℓ vertices for each of $\lceil\delta n\rceil$ labeled cycles (the first ℓ vertices define the first cycle and so on), while division by $\lceil\delta n\rceil!$ is used to remove cycle labels. By Stirling's formula,

$$\log(C_{n,\ell}) \leqslant \ell\delta n \log n - \delta n \log n + o(n \log n).$$

On the other hand, from (7), we have

$$\log \left|\mathcal{G}_{n,m-\ell\lceil\delta n\rceil}\right| = \Big(\frac{d}{2} - \ell\delta\Big)n \log n + o(n \log n).$$

So finally,

$$\log \left|\mathcal{G}_{n,m}(\rho, \varepsilon)\right| \leqslant \frac{d}{2}n \log n - \delta n \log n + o(n \log n).$$

This proves (55).   □

5.2 Proof of Theorem 1.3 and Theorem 1.2

Notice that if $P \in \mathcal{P}_h$, then $J_h(P)$ is a well defined extended real number in $[-\infty, \infty)$. The fact that $J_h(P) \leqslant s(d)$ follows from Proposition 5.6 below and from the upper bound $\overline{\Sigma}(\rho) \leqslant s(d)$, cf. (7).

As before, we fix $d > 0$ and an integer sequence $m = m(n)$ such that $m/n \to d/2$ as $n \to \infty$. We start with three preliminary lemmas.

Lemma 5.3 *The function $\rho \mapsto \underline{\Sigma}(\rho)$ on $\mathcal{P}_u(\mathcal{G}_*)$ is upper semi-continuous.*

Proof Consider a sequence (ρ_k) converging to ρ. We should check that $\underline{\Sigma}(\rho) \geqslant \limsup \underline{\Sigma}(\rho_k)$. Observe that for any $\varepsilon > 0$, for all k large enough, $B(\rho, \varepsilon) \supset B(\rho_k, \varepsilon/2)$. We get for k large enough,

$$\underline{\Sigma}(\rho, \varepsilon) \geqslant \underline{\Sigma}(\rho_k, \varepsilon/2) \geqslant \underline{\Sigma}(\rho_k).$$

Letting k tend to infinity and then ε to 0, we obtain the claim.   □

We will also need two general lemmas.

Lemma 5.4 *Let $P = \{p_x, x \in \mathcal{X}\}$ be a probability measure on a discrete space $\mathcal{X}$ such that $H(P) < \infty$. Let $(\ell_x)_{x\in\mathcal{X}}$ be a sequence with $\ell_x \in \mathbb{Z}_+$, $x \in \mathcal{X}$, such that $\sum_x p_x\ell_x \log \ell_x < \infty$. Then $-\sum_x p_x\ell_x \log p_x < \infty$.*

Proof We can assume without loss of generality that $p_x \neq 0$ for all $x \in \mathcal{X}$. We look for the sequence (ℓ_x) which maximizes the linear function $-\sum_x p_x\ell_x \log p_x < \infty$ under the constraints $\ell_x \geqslant 1$ and $\sum_x p_x\ell_x \log \ell_x = c$. If the constraint $\ell_x \geqslant 1$ is not saturated, taking derivative, we find $0 = -p_x \log p_x - \lambda p_x - \lambda p_x \log \ell_x$ where

λ is the Lagrange mutliplier associated to the constraint $\sum_x p_x \ell_x \log \ell_x = c$. We get $\ell_x = e^{-1} p_x^{-1/\lambda}$. Let $\mathcal{X}_1$ be the set of x such that $\ell_x = 1$. We thus find

$$
\begin{aligned}
-\sum_x p_x \ell_x \log p_x &= -\sum_{x \in \mathcal{X}_1} p_x \log p_x - \sum_{x \notin \mathcal{X}_1} p_x \ell_x \log p_x \\
&\leqslant H(P) - \sum_{x \notin \mathcal{X}_1} p_x \ell_x \log e^{-\lambda} \ell_x^{-\lambda} \\
&\leqslant H(P) + \lambda \sum_x p_x \ell_x + \lambda \sum_x p_x \ell_x \log \ell_x.
\end{aligned}
$$

The conclusion follows.   □

Lemma 5.5 *Let p, κ be integers and $\mathcal{A}_\kappa \subset \mathcal{P}(\mathbb{Z}^p)$ the set of probability measures P on $\mathbb{Z}^p$ such that $\mathbb{E} \sum_{i=1}^p |X_i| \leqslant \kappa$ where $X = (X_1, \ldots, X_p)$ has law P. Then the map $P \mapsto H(P)$ is continuous on $\mathcal{A}_\kappa$ for the weak topology.*

Proof A simple truncation argument shows that $\mathcal{A}_\kappa$ is weakly closed. Let Q_n (resp. Q) be the law of $\|X\|_1 = \sum_{i=1}^p |X_i|$ where X has law P_n (resp. P). If P_n^k (resp. P^k) is the conditional law of P_n (resp. P) conditioned on $\|X\|_1 = k$, we have

$$
H(P_n) = \sum_{k \geqslant 0} Q_n(k) H(P_n^k) + H(Q_n),
$$

and similarly for P. Since P_n^k is a probability measure on a finite set of size $c_k \leqslant (2k+1)^p$, we have for any k, $Q_n(k) \to Q(k)$, $H(P_n^k) \to H(P^k)$ as $n \to \infty$. Also, $H(P_n^k) \leqslant \log(c_k) \leqslant p \log(2k+1)$. Since $\sum_k k Q_n(k) \leqslant \kappa$, using that $x/\log(2x+1)$ is increasing for $x \geqslant 1$, it follows that for $\theta \geqslant 1$,

$$
\sum_{k \geqslant \theta} Q_n(k) H(P_n^k) \leqslant \frac{p \log(2\theta + 1)}{\theta} \sum_{k \geqslant \theta} k Q_n(k) \leqslant \frac{p \kappa \log(2\theta + 1)}{\theta}.
$$

This proves the uniform integrability of $k \mapsto H(P_n^k)$ for the measures Q_n. Hence letting first n and then θ tend to infinity, we get

$$
\lim_{n \to \infty} \sum_{k \geqslant 0} Q_n(k) H(P_n^k) = \sum_{k \geqslant 0} Q(k) H(P^k).
$$

It thus remains to prove that $\lim_{n \to \infty} H(Q_n) = H(Q)$. The proof is similar. First, for any θ,

$$
\lim_{n \to \infty} -\sum_{k < \theta} Q_n(k) \log Q_n(k) = -\sum_{k < \theta} Q(k) \log Q(k).
$$

Then, we need to upper bound $-\sum_{k\geqslant\theta} Q_n(k)\log Q_n(k)$, uniformly in n. It can be done as follows. Observe that $\sum_{k\geqslant\theta}\sqrt{k}Q_n(k)\leqslant\kappa/\sqrt{\theta}$. We then compute

$$L(\delta) = \sup - \sum_{k\geqslant 0} x_k\log x_k,$$

under the linear constraints, $x_k\geqslant 0$, $\sum_k x_k\leqslant 1$ and $\sum_{k\geqslant 0}\sqrt{k}x_k = \delta$. Using Lagrange multipliers denoted by λ and μ, the solution of this convex optimization problem is of the form $x_k = e^{-\mu-\lambda\sqrt{k}}$ for $k\geqslant 0$ and $\sum_k x_k = 1$. It is then easy to check that as $\delta\to 0$, $\lambda\delta\to 0$ and $\mu\to 0$. It follows that $L(\delta) = \mu+\lambda\delta\to 0$. It implies that $-\sum_{k\geqslant\theta} Q_n(k)\log Q_n(k)\leqslant L(\kappa/\sqrt{\theta})$ goes to 0 as $\theta\to\infty$ uniformly in n. Letting n tend to infinity and then θ, it proves that $\lim_{n\to\infty} H(Q_n) = H(Q)$. This concludes the proof of Lemma 5.5. $\qquad\square$

We now compute the entropy of $\mathrm{UGW}_h(P)$.

Proposition 5.6 *For $h\in\mathbb{N}$, $P\in\mathcal{P}_h$ and $\mathbb{E}_P\deg(o) = d$:*

$$\Sigma(\mathrm{UGW}_h(P)) = J_h(P). \tag{56}$$

Proof Lower bound, finite support Consider first the lower bound $\underline{\Sigma}(\mathrm{UGW}_h(P))\geqslant J_h(P)$ when P has finite support. By Lemma 4.11, we may choose a sequence $\Gamma_n\in\mathcal{G}_{n,m}$ such that $U(\Gamma_n)_h\rightsquigarrow P$, $n\to\infty$ and $U(\Gamma_n)_h$ has support contained $\Delta :=\{t_1,\ldots,t_r\}\subset\mathcal{T}_h^*$ for all n. Let $N_h(\Gamma_n)$ denote the number of graphs $G\in\mathcal{G}_n$ such that $U(G)_h = U(\Gamma_n)_h$. Clearly, all such graphs have the same number m of edges. From Corollary 4.10, we know that $N_h(\Gamma_n) = n(\mathbf{D})|\mathcal{G}(\mathbf{D}, 2h+1)|$, where $\mathbf{D}$ is the neighborhood sequence associated to Γ_n, i.e. if $c = (t, t')\in\mathcal{T}_{h-1}^*\times\mathcal{T}_{h-1}^*$, then $D_c(i)$ is the number of $j\sim i$ in Γ_n such that $\Gamma_n(i, j)_{h-1} = t'$ and $\Gamma_n(j, i)_{h-1} = t$; see (48). Then,

$$n(\mathbf{D}) = \binom{n}{\alpha_1 n, \ldots, \alpha_r n},$$

where $\alpha_k = \alpha_k(n)$ stands for the probability of t_k under $U(\Gamma_n)_h$. Since $\alpha_k\to P(t_k)$ as $n\to\infty$, Stirling's formula yields

$$\lim_{n\to\infty}\frac{1}{n}\log n(\mathbf{D}) = -\sum_t P(t)\log P(t) = H(P) \tag{57}$$

On the other hand, from Corollary 4.6 we have

$$\log|\mathcal{G}(\mathbf{D}, 2h+1)| = \frac{1}{2}\sum_{c\in\mathcal{C}}(S_c\log S_c - S_c) - \sum_{u\in[n]}\sum_{c\in\mathcal{C}}\log D_c(u)! + o(n), \tag{58}$$

where $\mathcal{C}$ denotes the set of all pairs $c = (t, t')\in\mathcal{T}_{h-1}^*\times\mathcal{T}_{h-1}^*$ associated to Γ_n as in (48), $S_c = S_{\bar{c}} = \sum_{u\in[n]} D_c(u)$, $\bar{c} = (t', t)$ if $c = (t, t')$. Note that the size of $\mathcal{C}$

is finite and independent of n. For a given $c = (t, t')$, using the notation (3) one has $S_c/n \to e_P(t, t')$. Also, writing $2m = \sum_{c \in \mathcal{C}} S_c$, (58) can be rewritten as

$$m \log n - m + \frac{1}{2} n \sum_{(t,t')} e_P(t, t') \log e_P(t, t') - n \sum_{(t,t')} \mathbb{E}_P \log E_h(t, t')! + o(n). \qquad (59)$$

From (57) and (59),

$$\lim_{n \to \infty} \frac{1}{n} (\log N_h(\Gamma_n) - m \log n) = J_h(P). \qquad (60)$$

To prove the desired lower bound on $\underline{\Sigma}(\mathrm{UGW}_h(P))$, we may restrict to graphs $G \in \mathcal{G}_{n,m}$ with $U(G)_h = U(\Gamma_n)_h$ to obtain

$$|\mathcal{G}_{n,m}(\mathrm{UGW}_h(P), \varepsilon)| \geqslant N_h(\Gamma_n) \, \mathbb{P}(U(G_n) \in B(\mathrm{UGW}_h(P), \varepsilon)),$$

where G_n is uniformly distributed in $\mathcal{G}(\mathbf{D}, 2h + 1)$ with $\mathbf{D}$ as above. From Theorem 4.8, for all $\varepsilon > 0$ one has

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(U(G_n) \in B(\mathrm{UGW}_h(P), \varepsilon)\right) = 0.$$

Using (60), we have proved that for all $\varepsilon > 0$, $\underline{\Sigma}(\mathrm{UGW}_h(P), \varepsilon) \geqslant J_h(P)$. Therefore,

$$\underline{\Sigma}(\mathrm{UGW}_h(P)) \geqslant J_h(P). \qquad (61)$$

Lower bound, general case Set $\rho = \mathrm{UGW}_h(P)$. We can assume that $J_h(P) > -\infty$. For each $n \in \mathbb{N}$, consider the forest F_n obtained from (T, o) with law ρ by removing all edges adjacent to a vertex with degree higher than n. We may define $\rho^{(n)}$ as the the law of $(F_n(o), o)$, the connected component of the root. It is not hard to check that $\rho^{(n)}$ is a unimodular measure. We define $P_n = \rho_h^{(n)}$, the law of its h-neighborhood. By construction, P_n is finitely supported, admissible, P_n converges weakly to P and $d_n = \mathbb{E}_{Q_n} \deg_G(o) \leqslant d$ converges to d. We pick some fixed integer $D > d \vee 2$ and define $R = \delta_{t_\star}$ as the Dirac mass of the h-neighborhood of the D-regular tree. If n is large enough, there exists $p_n \to 1$ such that $Q_n = p_n P_n + (1 - p_n)R$ has mean root degree equal to d. Also, $Q_n \in \mathcal{P}(\mathcal{T}_h^*)$ is admissible (the set of admissible measures is convex) and has finite support. We apply Lemma 5.3 and the lower bound for finitely supported measures, to obtain

$$\underline{\Sigma}(\rho) \geqslant \limsup_{n \to \infty} \underline{\Sigma}(\mathrm{UGW}_h(Q_n)) \geqslant \limsup_{n \to \infty} J_h(Q_n).$$

By definition, $J_h(Q_n) = -s(d) + H(Q_n) - \frac{d}{2}H(\pi_{Q_n}) - \sum_{(s,s')} \mathbb{E}_{Q_n} \log E_h(s, s')!$. We need to prove that $\limsup J_h(Q_n) \geqslant J_h(P)$. It suffices to prove that

$$\liminf_{n \to \infty} J_h(P_n) \geqslant J_h(P). \qquad (62)$$

First, the lower semi-continuity of the entropy gives $\liminf_{n\to\infty} H(P_n) \geqslant H(P)$. We now check that

$$\lim_{n\to\infty} \sum_{t,t'} \mathbb{E}_{P_n} \log E_h(t, t')! = \sum_{t,t'} \mathbb{E}_P \log E_h(t, t')!. \tag{63}$$

For ease of notation, we write $\mathcal{C} = \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*$, $c = (t, t') \in \mathcal{C}$ and $E_h(c)(\tau)$ to make explicit the dependence in $\tau \in \mathcal{T}_h^*$. As above, F_n is the forest obtained from (T, o) with law ρ, so that

$$\sum_{t,t'} \mathbb{E}_{P_n} \log E_h(t, t')! = \mathbb{E}_P \varphi((F_n(o), o)),$$

where $\varphi(\tau) = \sum_c \log (E_h(c)(\tau)!)$ satisfies:

$$\varphi(\tau) \leqslant \sum_c E_h(c)(\tau) \log E_h(c)(\tau)$$

$$\leqslant \sum_c E_h(c)(\tau) \log \left(\sum_{c'} E_h(c')(\tau) \right) = \deg_\tau(o) \log \deg_\tau(o). \tag{64}$$

In particular, $\varphi((F_n(o), o)) \leqslant \bar{\varphi}(T, o) := \deg_T(o) \log \deg_T(o)$. The assumption $P \in \mathcal{P}_h$ implies that $\mathbb{E}_P \bar{\varphi}(T, o) < \infty$. Therefore (63) follows from the dominated convergence theorem.

To conclude the proof of (62), it remains to check that $\limsup H(\pi_{P_n}) \leqslant H(\pi_P)$, i.e.

$$\liminf_{n\to\infty} \sum_{c\in\mathcal{C}} e_{P_n}(c) \log e_{P_n}(c) \geqslant \sum_{c\in\mathcal{C}} e_P(c) \log e_P(c). \tag{65}$$

For $\theta \in \mathbb{N}$, we denote by $\mathcal{F}_\theta \subset \mathcal{T}_h^*$ the subset of trees whose root vertex has degree bounded by $\theta + 1$ and by $\mathcal{C}_\theta \subset \mathcal{C}$, the finite subset of pairs of trees with vertex degrees bounded by θ. The assumption $P \in \mathcal{P}_h$ and Lemma 5.4 imply that $-\sum_\tau \deg_\tau(o) P(\tau) \log P(\tau) < \infty$. Also, the assumption $J_h(P) > -\infty$ implies that $H(\pi_P) < \infty$ and $\sum_c |e_P(c) \log e_P(c)| < \infty$. It follows that for any $\varepsilon > 0$, there exists θ such that

$$\left| \sum_{c\notin\mathcal{C}_\theta} e_P(c) \log e_P(c) \right| \leqslant \varepsilon \quad \text{and} \quad -\sum_{\tau\notin\mathcal{F}_\theta} \deg_\tau(o) P(\tau) \log P(\tau) \leqslant \varepsilon.$$

By dominated convergence, for any $c \in \mathcal{C}$, $e_{P_n}(c) \to e_P(c)$. Since $\mathcal{C}_\theta$ is finite, we find

$$\limsup_{n\to\infty} \left| \sum_{c\in\mathcal{C}_\theta} e_{P_n}(c) \log e_{P_n}(c) - \sum_{c\in\mathcal{C}} e_P(c) \log e_P(c) \right| \leqslant \varepsilon.$$

Since $\varepsilon > 0$ is arbitrarily small, in order to complete (65), it suffices to prove that for any $n \in \mathbb{N}$,

$$\sum_{c \notin \mathcal{C}_\theta} e_{P_n}(c) \log e_{P_n}(c) \geqslant -\varepsilon.$$

We write

$$
\begin{aligned}
e_{P_n}(c) \log e_{P_n}(c) &= \sum_\tau P_n(\tau) E_h(c)(\tau) \log \left(\sum_{\tau'} P_n(\tau') E_h(c)(\tau') \right) \\
&\geqslant \sum_\tau P_n(\tau) E_h(c)(\tau) \log \left(P_n(\tau) E_h(c)(\tau) \right) \\
&\geqslant \sum_\tau P_n(\tau) E_h(c)(\tau) \log P_n(\tau).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\sum_{c \notin \mathcal{C}_\theta} e_{P_n}(c) \log e_{P_n}(c) &\geqslant \sum_\tau P_n(\tau) \log P_n(\tau) \sum_{c \notin \mathcal{C}_\theta} E_h(c)(\tau) \\
&\geqslant \sum_\tau \deg_\tau(o) \mathbf{1}(\tau \notin \mathcal{F}_\theta) P_n(\tau) \log P_n(\tau),
\end{aligned}
$$

where we use that $\sum_c E_h(c)(\tau) = \deg_\tau(o)$, and that if $\tau \notin \mathcal{F}_\theta$ and $c \in \mathcal{C}_\theta$ then $E_h(c)(\tau) = 0$. Now, by construction, there exists a partition $\cup_i \mathcal{X}_n^i$ of $\mathcal{T}_h^*$ and $\tau_n^i \in \mathcal{X}_n^i$ such that if $(T, o) \in \mathcal{X}_n^i$ then $(F_n(o), o) = \tau_n^i$. Also, $P_n(\tau_n^i) = P(\mathcal{X}_n^i) \geqslant P(\tau_n^i)$, and for all $\tau \in \mathcal{X}_n^i$, $\deg_\tau(o) \geqslant \deg_{\tau_n^i}(o)$, $\mathbf{1}(\tau \notin \mathcal{F}_\theta) \geqslant \mathbf{1}(\tau_n^i \notin \mathcal{F}_\theta)$. It follows that

$$
\begin{aligned}
\sum_{c \notin \mathcal{C}_\theta} e_{P_n}(c) \log e_{P_n}(c) &\geqslant \sum_i \sum_{\tau \in \mathcal{X}_n^i} \deg_\tau(o) \mathbf{1}(\tau \notin \mathcal{F}_\theta) P(\tau) \log P_n(\tau_n^i) \\
&\geqslant \sum_{\tau \notin \mathcal{F}_\theta} \deg_\tau(o) P(\tau) \log P(\tau) \geqslant -\varepsilon.
\end{aligned}
$$

This concludes the proof of (65).

Upper bound The upper bound $\overline{\Sigma}(\mathrm{UGW}_h(P)) \leqslant J_h(P)$ is a consequence of the general estimate of Lemma 5.7 below. $\qquad\square$

Lemma 5.7 *Fix $h \in \mathbb{N}$. If $\rho \in \mathcal{P}_u(\mathcal{T}^*)$ is such that $\rho_h \in \mathcal{P}_h$, then*

$$\overline{\Sigma}(\rho) \leqslant J_h(\rho_h). \tag{66}$$

Proof Finite support For clarity, we first assume that $P = \rho_h$ has finite support. The definition of local weak topology implies that for any $h \in \mathbb{N}$, any $\varepsilon > 0$, there exists

$\eta > 0$ such that $B(\rho, \eta) \subset \{\mu \in \mathcal{P}(\mathcal{G}^*) : d_{TV}(\mu_h, \rho_h) \leqslant \varepsilon\}$, where d_{TV} denotes the total variation distance. Define

$$A_{n,m}(P, \varepsilon) = \big\{G \in \mathcal{G}_{n,m} : d_{TV}(U(G)_h, P) \leqslant \varepsilon\big\}.$$

Therefore, (66) follows if we prove

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \big(\log |A_{n,m}(P, \varepsilon)| - m \log n\big) \leqslant J_h(P). \tag{67}$$

Let $\Delta \subset \mathcal{T}_h^*$ be the support of P. Define $\mathcal{F} \subset \mathcal{T}_{h-1}^*$ as the set of unlabeled rooted trees $t \in \mathcal{T}_{h-1}^*$ such that either $T(o, v)_{h-1} = t$ or $T(v, o)_{h-1} = t$ for some $T \in \Delta$. Set $L = |\mathcal{F}|$. Also, by adding a fictitious point $\star$ to $\mathcal{F}$, define $\bar{\mathcal{F}} = \mathcal{F} \cup \{\star\}$, and call $\bar{\mathcal{C}}$ the associated set of $(L + 1) \times (L + 1)$ colors $c = (t, t')$, $t, t' \in \bar{\mathcal{F}}$. To any graph $G \in \mathcal{G}_{n,m}$ we may associate a degree sequence $\bar{\mathbf{D}} = (\bar{D}(1), \dots, \bar{D}(n))$, where $\bar{D}(i)$ is a $(L + 1) \times (L + 1)$ matrix for each i, obtained as in (48) by identifying with $\star$ all neighborhoods that do not belong to $\mathcal{F}$. The precise construction is defined as follows. Fix an edge $\{u, v\}$ of G: if $G(u, v)_{h-1} = t'$ and $G(v, u)_{h-1} = t$, with $t, t' \in \mathcal{F}$, then we say that the oriented pair (u, v) has color $c = (t, t') \in \bar{\mathcal{C}}$; if either $G(u, v)_{h-1}$ or $G(v, u)_{h-1}$ are not in $\mathcal{F}$, then we say that the oriented pair (u, v) has color $(\star, \star) \in \bar{\mathcal{C}}$. This defines a directed colored graph $\widetilde{G}$ with colors from the set $\bar{\mathcal{C}}$. We call $\bar{\mathbf{D}}$ the corresponding degree sequence, i.e. $\bar{D}_c(i)$ is the number of directed edges with color c going out of vertex i. Note that by construction, if (u, v) has color c, then (v, u) has color $\bar{c}$, and that there is no edge with color $(t, \star)$ or $(\star, t)$ for any $t \in \mathcal{F}$.

In this way a graph $G \in \mathcal{G}_{n,m}$ yields an element $\widetilde{G}$ of $\widehat{\mathcal{G}}(\bar{\mathbf{D}})$. Let $\bar{Q}(G)$ denote the empirical degree law

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{\bar{D}(i)}. \tag{68}$$

Thus $\bar{Q}(G)$ is a probability measure on the set $\mathcal{M}_{L+1}$; see Eq. (26). Also, let $\bar{P}$ denote the probability measure on $\mathcal{M}_{L+1}$ induced by P. Namely, $\bar{P}$ is the law of the random matrix $\mathbf{D} \in \mathcal{M}_{L+1}$ defined as follows: for all $c = (t, \star)$, or $c = (\star, t)$ or $c = (\star, \star)$, set $D_c = 0$; and for $c = (t, t')$ with $t, t' \in \mathcal{F}$, set $D_c = E_h(t', t)$, where $E_h(t', t)$ is defined by (3) if the rooted graph (G, o) has law P. By contraction, one has $H(\bar{P}) \leqslant H(P)$ and

$$d_{TV}(\bar{Q}(G), \bar{P}) \leqslant d_{TV}(U(G)_h, P).$$

Let $\mathcal{P}_{n,m}(P, \varepsilon)$ denote the set of probability measures $Q \in \mathcal{P}(\mathcal{M}_{L+1})$ of the form (68), satisfying $\sum_{i \in [n]} \sum_{c \in \bar{\mathcal{C}}} \bar{D}_c(i) = 2m$, and such that $d_{TV}(Q, \bar{P}) \leqslant \varepsilon$. The above discussion shows that if $G \in A_{n,m}(P, \varepsilon)$, there must exist $Q \in \mathcal{P}_{n,m}(P, \varepsilon)$ such that $\bar{Q}(G) = Q$. Therefore, one obtains

$$|A_{n,m}(P, \varepsilon)| \leqslant |\mathcal{P}_{n,m}(P, \varepsilon)| \max_{Q \in \mathcal{P}_{n,m}(P, \varepsilon)} n(\bar{\mathbf{D}})|\widehat{\mathcal{G}}(\bar{\mathbf{D}})| \tag{69}$$

where $n(\bar{\mathbf{D}})$ is defined as in Corollary 4.10, and $\bar{\mathbf{D}}$ is the degree vector associated to Q as in (68).

Next, we claim that for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{P}_{n,m}(P, \varepsilon)| = 0. \tag{70}$$

Indeed, let $p = |\bar{\mathcal{C}}|$ and fix a vector $\ell \in \mathbb{Z}_+^p$. An integer partition of the vector ℓ is an unordered sequence $\{d(1), \ldots, d(k)\}$, with $d(i) \in \mathbb{Z}_+^p$ for all i, and such that $d(1) + \cdots + d(k) = \ell$ componentwise. By [17, Lemma 4.2], if $\sum_{i=1}^p \ell_i = 2m$ then the number of integer partitions of ℓ is $\exp(o(m))$. The number of vectors $\ell \in \mathbb{Z}_+^p$ such that $\sum_{i=1}^p \ell_i = m$ is bounded by $(m + 1)^p$. It follows that the number of unordered sequences $\{d(1), \ldots, d(n)\}$ in $\mathbb{Z}_+^p$ such that $\sum_{i=1}^n \sum_{c=1}^p d_c(i) = 2m$ is at most $\exp(o(n))$, for $m = O(n)$. Now, if Q is of the form (68) we may define $d_c(i) = \bar{D}_c(i)$, for every $c \in \bar{\mathcal{C}}$ and $i \in [n]$, which yields an injective map from $\mathcal{P}_{n,m}(P, \varepsilon)$ to the unordered sequences $\{d(1), \ldots, d(n)\}$, with $d(i) \in \mathbb{Z}_+^p$ such that $\sum_{i=1}^n \sum_{c \in \bar{\mathcal{C}}} d_c(i) = 2m$. This proves (70).

From (69) and (70), to prove (67), it remains to show that

$$\limsup_{n\to\infty} \max_{Q\in\mathcal{P}_{n,m}(P,\varepsilon)} \frac{1}{n} \left[\log \left(n(\bar{\mathbf{D}})|\widehat{\mathcal{G}}(\bar{\mathbf{D}})| \right) - m \log n \right] \leqslant J_h(P) + \eta(\varepsilon), \tag{71}$$

where we use the notation $\eta(\varepsilon)$ for an arbitrary function satisfying $\eta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Since $\bar{\mathcal{C}}$ is finite, reasoning as in (57) and using Lemma 5.5, it is easily seen that

$$\limsup_{n\to\infty} \max_{Q\in\mathcal{P}_{n,m}(P,\varepsilon)} \frac{1}{n} \log n(\bar{\mathbf{D}}) \leqslant H(\bar{P}) + \eta(\varepsilon) \leqslant H(P) + \eta(\varepsilon). \tag{72}$$

Moreover, as in (58) one has

$$\log \left| \widehat{\mathcal{G}}(\bar{\mathbf{D}}) \right| = \frac{1}{2} \sum_{c\in\bar{\mathcal{C}}} (\bar{S}_c \log \bar{S}_c - \bar{S}_c) - \sum_{u\in[n]} \sum_{c\in\bar{\mathcal{C}}} \log \bar{D}_c(u)! + o(n),$$

where $\bar{S}_c = \sum_{i\in[n]} \bar{D}_c(i)$. Observe that

$$\limsup_{n\to\infty} \max_{Q\in\mathcal{P}_{n,m}(P,\varepsilon)} |\bar{S}_c/n - e_P(c)| \leqslant \eta(\varepsilon). \tag{73}$$

Indeed, if Q_n is a sequence with $Q_n \in \mathcal{P}_{n,m}(P, \varepsilon)$, then $\bar{S}_c/n = \mathbb{E}_{Q_n} D_c$, where $D \in \mathcal{M}_{L+1}$ has law Q_n. Then, for any $k \in \mathbb{N}$, $\bar{S}_c/n \geqslant \mathbb{E}_{Q_n}[D_c \wedge k]$, and since $Q_n \in \mathcal{P}_{n,m}(P, \varepsilon)$ and $D_c \wedge k$ is a bounded function, taking first $\varepsilon \to 0$ and then $k \to \infty$, one has $\bar{S}_c/n \geqslant e_P(c) - \eta(\varepsilon)$ uniformly in n. Moreover, since $\sum_{c\in\bar{\mathcal{C}}} \bar{S}_c/n = 2m/n = d + o(1)$, one has $\bar{S}_c/n = d + o(1) - \sum_{c'\neq c} \bar{S}_{c'}/n$. Therefore, from the lower bound $\bar{S}_c/n \geqslant e_P(c) - \eta(\varepsilon)$ and the fact that $\sum_c e_P(c) = d$, one finds $\bar{S}_c/n \leqslant$

$e_P(c) + |\bar{C}|\eta(\varepsilon) + o(1)$. This ends the proof of (73). Moreover, with the same truncation argument as above one has that

$$\frac{1}{n} \sum_{u \in [n]} \log \bar{D}_c(u)! \geqslant \mathbb{E}_P[\log E_h(c)!] - \eta(\varepsilon),$$

for all $c \in \bar{C}$. This, together with (72)–(73) and the argument in (59) allows us to conclude the proof of (71). This ends the proof of (67).

General case We now come back to the case of arbitrary $P \in \mathcal{P}_h$. For any finite set $\Delta \subset \mathcal{T}_h^*$, we associate the sets $\mathcal{C} = \mathcal{C}(\Delta)$ and $\bar{C}$ as above. The above argument establishes that

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \left(\log \left| A_{n,m}(P, \varepsilon) \right| - m \log n \right) \leqslant J_h^\Delta(P), \qquad (74)$$

where $J_h^\Delta(P) := -s(d) + H(P) - \frac{d}{2} H(\pi_{\bar{P}}) - \sum_{(t,t') \in \mathcal{C}} \mathbb{E}_P \log E_h(t, t')! - \mathbb{E}_P \log E_h(\star, \star)!$ and $\pi_{\bar{P}} \in \mathcal{P}(\bar{C})$ is defined as follows: for $c = (t, t') \in \mathcal{C}$, $\pi_{\bar{P}}(t, t') = \pi_P(t, t')$ and for $c = (\star, \star)$,

$$\pi_{\bar{P}}(\star, \star) = 1 - \sum_{(t,t') \in \mathcal{C}} \pi_P(t, t') = \frac{1}{d} \mathbb{E}_P \left| \left\{ v \overset{T}{\sim} o : T(o, v) \text{ or } T(v, o)_{h-1} \text{ is not in } \mathcal{C} \right\} \right|.$$

Assume first that $J_h(P) > -\infty$. Using (64) at the second line, one has

$$J_h^\Delta(P) \leqslant J_h(P) - \frac{d}{2} \sum_{(t,t') \notin \mathcal{C}} \pi_P(t, t') \log \pi_P(t, t') + \sum_{(t,t') \notin \mathcal{C}} \mathbb{E}_P \log E_h(t, t')!$$

$$\leqslant J_h(P) - \frac{d}{2} \sum_{(t,t') \notin \mathcal{C}} \pi_P(t, t') \log \pi_P(t, t') + \mathbb{E}_P \left[\mathbf{1}(T \notin \Delta) \deg_T(o) \log \deg_T(o) \right].$$

We may then consider a sequence (Δ_k) of finite subsets in $\mathcal{T}_h^*$ such that $P(T \notin \Delta_k) \to 0$, and $\sum_{c \notin \mathcal{C}(\Delta_k)} \pi_P(c) \log \pi_P(c) \to 0$, as $k \to \infty$. Then as $k \to \infty$, the above expression converges to $J_h(P)$. This proves that (66) holds when $P \in \mathcal{P}_h$ and $J_h(P) > -\infty$.

If $P \in \mathcal{P}_h$ and $J_h(P) = -\infty$ then either $H(\pi_P) = \infty$ or $\sum_{(t,t')} \mathbb{E}_P \log E_h(t, t)! = \infty$. We use the upper bound

$$J_h^\Delta(P) \leqslant -s(d) + H(P) + \frac{d}{2} \sum_{(t,t') \in \mathcal{C}} \pi_P(t, t') \log \pi_P(t, t') - \sum_{(t,t') \in \mathcal{C}} \mathbb{E}_P \log E_h(t, t')!$$

We may consider a sequence (Δ_k) of finite subsets of $\mathcal{T}_h^*$ such that, as $k \to \infty$, one has $\sum_{c \in \mathcal{C}(\Delta_k)} \pi_P(c) \log \pi_P(c) \to -H(\pi_P)$ and $\sum_{c \in \mathcal{C}(\Delta_k)} \mathbb{E}_P \log E_h(c)! \to \sum_c \mathbb{E}_P \log E_h(c)!$, and therefore $J_h^{\Delta_k}(P) \to -\infty$, $k \to \infty$. This completes the proof of (66). $\qquad\square$

Next, we extend Lemma 5.7 to the case $\rho_h \notin \mathcal{P}_h$, i.e. $H(\rho_h) = \infty$ or $\mathbb{E}_\rho \deg_T(o) \log \deg_T(o) = \infty$. We start with the latter case.

Lemma 5.8 *If $\rho \in \mathcal{P}_u(\mathcal{T}^*)$ is such that $\mathbb{E}_\rho \deg_T(o) = d$ and $\mathbb{E}_\rho \deg_T(o) \log \deg_T(o) = \infty$ then*

$$\overline{\Sigma}(\rho) = -\infty.$$

Proof We set $P = \rho_1$ which can be identified with a probability measure on $\mathbb{Z}_+$. Since P has finite first moment, $H(P)$ is finite. The proof of Lemma 5.7 can be simplified for $h = 1$: since $\mathcal{T}^*_{h-1}$ has a unique element (the isolated root), one has $H(\pi_P) = 0$ and it is not necessary to consider the extra state $\star$. The bound (67) gives

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \left(\log \left| A_{n,m}(P, \varepsilon) \right| - m \log n \right) \leqslant -s(d) + H(P) - \mathbb{E}_P \log \deg_T(o)!$$

Now, from Stirling's approximation, for $n \geqslant 1$, $n! \geqslant c\sqrt{n}e^{-n}n^n$ for some constant $c > 0$. We deduce that $\log n! \geqslant c' - n + n \log n$ for some constant $c' > 0$. In particular, from $\mathbb{E}_P \deg_T(o) = d < \infty$ and $\mathbb{E}_\rho \deg_T(o) \log \deg_T(o) = \infty$, we get that $\mathbb{E}_P \log \deg_T(o)! = \infty$. □

The following statement is the extension of Lemma 5.7 to the case $\rho_h \notin \mathcal{P}_h$.

Proposition 5.9 *If $\rho \in \mathcal{P}_u(\mathcal{T}^*)$, then for any $h \in \mathbb{N}$,*

$$\overline{\Sigma}(\rho) \leqslant \overline{J}_h(\rho_h), \tag{75}$$

where $\overline{J}_h(\rho_h) = J_h(\rho_h)$ if $\rho_h \in \mathcal{P}_h$, and $\overline{J}_h(\rho_h) = -\infty$ otherwise.

In view of Lemma 5.7 and Lemma 5.8, Proposition 5.9 is a consequence of the following lemma.

Lemma 5.10 *Let $\rho \in \mathcal{P}_u(\mathcal{T}^*)$ be such that $\mathbb{E}_\rho \deg_T(o) \log \deg_T(o) < \infty$. Then for any $h \in \mathbb{N}$, $H(\rho_h) < \infty$. Consequently, for any $P \in \mathcal{P}(\mathcal{T}^*_h)$, $P \in \mathcal{P}_h$ is equivalent to P admissible and $\mathbb{E}_P \deg_T(o) \log \deg_T(o) < \infty$.*

Proof The second statement follows from the first applied to $\rho = \mathrm{UGW}_h(P)$ with P admissible. We now prove the first statement. Since $d := \mathbb{E}_\rho \deg(o)$ is finite, one has $H(\rho_1) < \infty$. To prove the lemma, we proceed by induction, and show that for any $h \in \mathbb{N}$, if $H(\rho_h) < \infty$ and $\mathbb{E}_\rho \deg_T(o) \log \deg_T(o) < \infty$ then $H(\rho_{h+1}) < \infty$. Set $P = \rho_h$, $Q = \rho_{h+1}$, and $Q^* = [\mathrm{UGW}_h(P)]_{h+1}$. Assume that $H(P) < \infty$. We are going to prove that $H(Q) < \infty$. Observe that

$$H(Q) = H(P) + \sum_{\gamma \in \mathcal{T}^*_h} P(\gamma) H(Q(\cdot|\gamma)),$$

where $Q(\cdot|\gamma)$ stands for the conditional distribution of the $(h+1)$-neighborhood given the h-neighborhood γ. Also,

$$\sum_{\gamma \in \mathcal{T}_h^*} P(\gamma) H(Q(\cdot|\gamma)) = - \sum_{\tau \in \mathcal{T}_{h+1}^*} Q(\tau) \log \frac{Q(\tau)}{P(\tau_h)}$$

$$= - \sum_{\tau \in \mathcal{T}_{h+1}^*} Q(\tau) \log \frac{Q^*(\tau)}{P(\tau_h)} - H(Q|Q^*)$$

$$\leqslant - \sum_{\tau \in \mathcal{T}_{h+1}^*} Q(\tau) \log \frac{Q^*(\tau)}{P(\tau_h)}.$$

Now recall that $\tau \in \mathcal{T}_{h+1}^*$ determines all the coefficients $E_{h+1}(t, t')$, $(t, t') \in \mathcal{T}_h^* \times \mathcal{T}_h^*$, and these can be partitioned according to the pairs $(s, s') \in \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*$ such that $t_{h-1} = s$, $t'_{h-1} = s'$. With this notation, by definition of Q^*, one has, for $\tau \in \mathcal{T}_{h+1}^*$ such that $\tau_h = \gamma$:

$$\frac{Q^*(\tau)}{P(\tau_h)} = Q^*(\tau|\gamma) = \prod_{(s,s') \in \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*} \binom{E_h(s, s')}{\{E_{h+1}(t, t')\}} \prod_{t \in \mathcal{T}_h^*} \widehat{P}_{s,s'}(t)^{k_{t,s'}(\tau)}, \quad (76)$$

where the terms $\{E_{h+1}(t, t')\}$ in the multinomial coefficient are all such that $t_{h-1} = s$, $t'_{h-1} = s'$, and we write $k_{t,s'}(\tau) := |\{v \overset{\tau}{\sim} o : \tau(o, v)_h = t, \tau(v, o)_{h-1} = s'\}|$, with $t_{h-1} = s$. Therefore,

$$- \sum_\tau Q(\tau) \log \frac{Q^*(\tau)}{P(\tau_h)} \leqslant - \sum_{s,s'} \sum_{t : t_{h-1}=s} \sum_\tau Q(\tau) k_{t,s'}(\tau) \log \widehat{P}_{s,s'}(t).$$

Moreover, unimodularity yields

$$\sum_\tau Q(\tau) k_{t,s'}(\tau) = \mathbb{E}_\rho |\{v \sim o : T(o, v)_{h-1} = s', T(v, o)_h = t\}| = \widehat{P}_{s,s'}(t) e_P(s, s')$$

$$(77)$$

Thus,

$$- \sum_\tau Q(\tau) \log \frac{Q^*(\tau)}{P(\tau_h)} \leqslant - \sum_{s,s'} \sum_{t : t_{h-1}=s} e_P(s, s') \widehat{P}_{s,s'}(t) \log \widehat{P}_{s,s'}(t)$$

$$= \sum_{s,s'} e_P(s, s') H(\widehat{P}_{s,s'}).$$

In conclusion, we have obtained that

$$H(Q) \leqslant H(P) + \sum_{s,s'} e_P(s, s') H(\widehat{P}_{s,s'}).$$

The proof will be complete once we show that $H(P) < \infty$ and $\mathbb{E}_\rho \deg_T(o) \log \deg_T(o)$ $< \infty$ imply that $\sum_{s,s'} e_P(s, s') H(\widehat{P}_{s,s'}) < \infty$.

Now, by definition, if $\gamma = t \cup s'_+$ and $n_{t,s'} = |\{v \overset{\gamma}{\sim} o : \gamma(v, o) = t, \gamma(o, v) = s'\}|$, we have

$$
\begin{aligned}
\sum_{s,s'} e_P(s, s') H(\widehat{P}_{s,s'}) &= \sum_{s,s'} \sum_{t:t_{h-1}=s} n_{t,s'} P(t \cup s'_+) \log \frac{e_P(s, s')}{n_{t,s'} P(t \cup s'_+)} \\
&\leqslant \sum_{s,s'} \sum_{t:t_{h-1}=s} n_{t,s'} P(t \cup s'_+) \log \frac{d}{P(t \cup s'_+)} \\
&= \sum_{\gamma \in \mathcal{T}_h^*} P(\gamma) \log \left(\frac{d}{P(\gamma)} \right) \sum_{s,s'} E_h(s', s)(\gamma),
\end{aligned}
$$

where we have used that $e_P(s, s') \leqslant d\, n_{t,s'}$, and that $1 \leqslant n_{t,s'} \leqslant E_h(s', s)(\gamma)$ where $\gamma = t \cup s'_+$ and $t_{h-1} = s$. Since $\deg_\gamma(o) = \sum_{s,s'} E_h(s, s')(\gamma)$, we find

$$
\sum_{s,s'} e_P(s, s') H(\widehat{P}_{s,s'}) \leqslant d \log d - \sum_{\gamma \in \mathcal{T}_h^*} \deg_\gamma(o) P(\gamma) \log P(\gamma).
$$

It remains to apply Lemma 5.4 with $\mathcal{X} = \mathcal{T}_h^*$ and $\ell_x = \deg_x(o)$ together with the assumption $\mathbb{E}_\rho \deg_T(o) \log \deg_T(o) < \infty$. □

Lemma 5.11 *Suppose $\rho \in \mathcal{P}_u(\mathcal{T}^*)$. Then $\overline{J}_k(\rho_k)$, $k \in \mathbb{N}$, is a non-increasing sequence. Assume moreover that ρ_1 has finite support. Then for fixed $k > h$, one has $\overline{J}_k(\rho_k) < \overline{J}_h(\rho_h)$ if and only if $\rho_k \neq [\mathrm{UGW}_h(\rho_h)]_k$. In particular, if ρ_1 has finite support, then for any $h \in \mathbb{N}$, one has $\overline{\Sigma}(\rho) < \overline{J}_h(\rho_h)$ if and only if $\rho \neq \mathrm{UGW}_h(\rho_h)$.*

Proof Fix $h \in \mathbb{N}$. To prove that $\overline{J}_{h+1}(\rho_{h+1}) \leqslant \overline{J}_h(\rho_h)$, we may assume that $\rho_{h+1} \in \mathcal{P}_{h+1}$. In this case one has also that $\rho_h \in \mathcal{P}_h$. From Proposition 5.6, we know that $\Sigma(\mathrm{UGW}_k(\rho_k)) = J_k(\rho_k)$ for both $k = h$ and $k = h + 1$. Therefore, using Lemma 5.7 one has

$$
J_{h+1}(\rho_{h+1}) = \Sigma(\mathrm{UGW}_{h+1}(\rho_{h+1})) \leqslant J_h([\mathrm{UGW}_{h+1}(\rho_{h+1})]_h) = J_h(\rho_h),
$$

where we use $[\mathrm{UGW}_k(\rho_k)]_h = \rho_h$, $k \geqslant h$. This proves that $\overline{J}_k(\rho_k)$ is non-increasing in k.

We now assume that ρ_1 has finite support. Then, by unimodularity it follows that ρ_h has finite support for all $h \in \mathbb{N}$. In particular, $\rho_h \in \mathcal{P}_h$ and $J_h(\rho_h) > -\infty$ for all $h \in \mathbb{N}$. Fix $k > h$. Suppose that $\overline{J}_k(\rho_k) < \overline{J}_h(\rho_h)$. One has $\Sigma(\mathrm{UGW}_k(\rho_k)) = J_k(\rho_k)$ by Proposition 5.6. From the consistency property of Lemma 3.3, one must then have $\rho_k \neq [\mathrm{UGW}_h(\rho_h)]_k$.

Next, suppose that ρ_1 has finite support and that $\rho_k \neq [\mathrm{UGW}_h(\rho_h)]_k$ and let us show that $J_k(\rho_k) < J_h(\rho_h)$. If $\Gamma_n \in \mathcal{G}_{n,m}$ is a sequence with $U(\Gamma_n)_k \rightsquigarrow \rho_k$, then also

$U(\Gamma_n)_h \rightsquigarrow \rho_h$ and by (60) one has

$$J_k(\rho_k) - J_h(\rho_h) = \lim_{n\to\infty} \frac{1}{n} \left(\log N_k(\Gamma_n) - \log N_h(\Gamma_n) \right). \tag{78}$$

Using Corollary 4.10, if $\widehat{G}_n$ denotes a random graph with uniform distribution in $\mathcal{G}(\mathbf{D}^{(n)}, 2h+1)$, $\mathbf{D}^{(n)}$ being the degree vector associated to the h-neighborhood of Γ_n, one also has

$$J_k(\rho_k) - J_h(\rho_h) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(U(\widehat{G}_n)_k = U(\Gamma_n)_k\big). \tag{79}$$

Since $\rho_k \neq \gamma_k := [\mathrm{UGW}_h(\rho_h)]_k$, there exist $\varepsilon > 0$ and an event A of the form $A = \{g \in \mathcal{G}^* : g_k = t\}$ for some $t \in \mathcal{T}_k^*$, such that $|\rho_k(A) - \gamma_k(A)| > \varepsilon$. Therefore, $U(\Gamma_n)_k \rightsquigarrow \rho_k$ implies that

$$J_k(\rho_k) - J_h(\rho_h) \leqslant \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(|U(\widehat{G}_n)_k(A) - \gamma_k(A)| > \varepsilon/2\big).$$

By Proposition 7.1, $\mathbb{E}U(\widehat{G}_n)(A)$ converges to $\gamma_k(A)$. It follows that

$$J_k(\rho_k) - J_h(\rho_h) \leqslant \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(|U(\widehat{G}_n)(A) - \mathbb{E}U(\widehat{G}_n)(A)| > \varepsilon/3\big).$$

The desired conclusion $J_k(\rho_k) - J_h(\rho_h) < 0$ now follows from (99) (in Appendix).

Finally, the assertion concerning $\overline{\Sigma}(\rho)$ follows easily from the results above. Indeed, from Proposition 5.6 we know that $\overline{\Sigma}(\rho) < J_h(\rho_h)$ implies that $\rho \neq \mathrm{UGW}_h(\rho_h)$. For the opposite direction, observe that if $\rho \neq \mathrm{UGW}_h(\rho_h)$, then $\rho_k \neq [\mathrm{UGW}_h(\rho_h)]_k$ for some $k > h$. From Lemma 5.7 one has $\overline{\Sigma}(\rho) \leqslant J_k(\rho_k)$, and the above implies $\overline{\Sigma}(\rho) < J_h(\rho_h)$. $\qquad\square$

Lemma 5.12 *Suppose* $\rho \in \mathcal{P}_u(\mathcal{T}^*)$. *Then*

$$\Sigma(\rho) = \overline{J}_\infty(\rho) := \lim_{k\to\infty} \overline{J}_k(\rho_k). \tag{80}$$

Proof The limit $\overline{J}_\infty(\rho)$ is well defined by the monotonicity in Lemma 5.11. The upper bound in Proposition 5.9 shows that $\overline{\Sigma}(\rho) \leqslant \overline{J}_\infty(\rho)$. Thus, all we have to prove is

$$\underline{\Sigma}(\rho) \geqslant \overline{J}_\infty(\rho). \tag{81}$$

We may assume that $\rho_k \in \mathcal{P}_k$ for all $k \in \mathbb{N}$. Fix $\eta > 0$ and set $\rho^h = \mathrm{UGW}_h(\rho_h)$. By the lower bound in Proposition 5.6, for any $h \in \mathbb{N}$, $\varepsilon > 0$ and $n \geqslant n_0(\varepsilon, h, \eta)$,

$$x(n, h, \varepsilon) := \frac{1}{n} \left(\log \left| \mathcal{G}_{n,m}(\rho^h, \varepsilon) \right| - m \log n \right) \geqslant \overline{J}_\infty(\rho) - \eta.$$

By diagonal extraction, there exist sequences $h_n \to \infty$ and $\varepsilon_n \to 0$ such that

$$\liminf_{n \to \infty} x(n, h_n, \varepsilon_n) \geqslant \overline{J}_\infty(\rho) - \eta.$$

Since $\rho^{h_n} \rightsquigarrow \rho$, for any fixed $\varepsilon > 0$ and all n large enough, $B(\rho^{h_n}, \varepsilon_n) \subset B(\rho, \varepsilon)$. In particular, $|\mathcal{G}_{n,m}(\rho, \varepsilon)| \geqslant |\mathcal{G}_{n,m}(\rho^{h_n}, \varepsilon_n)|$. It follows that $\underline{\Sigma}(\rho, \varepsilon) \geqslant \overline{J}_\infty(\rho) - \eta$. The latter holding for all $\varepsilon > 0$ and $\eta > 0$, we have checked that (81) holds.   □

All the statements in Theorem 1.3 are contained in Proposition 5.6, Proposition 5.9, Lemma 5.11 and Lemma 5.12. Moreover, Lemma 5.12 implies that $\Sigma(\rho)$ is well defined and equals $\overline{J}_\infty(\rho)$ for every $\rho \in \mathcal{P}_u(\mathcal{T}^*)$, independently of the choice of the sequence $m = m(n)$ with $m/n \to d/2$. This completes the proof of Theorem 1.2 and Theorem 1.3.

5.3 Proof of Corollary 1.4

In the special case $h = 1$, one has $P \in \mathcal{P}(\mathbb{Z}_+)$, and the condition $\sum_{n=0}^{\infty} n P(n) = d$ implies $H(P) < \infty$. By Proposition 5.6 one has $\Sigma(\mathrm{UGW}_1(P)) = J_1(P)$. Moreover, since $|\mathcal{T}_0^*| = 1$, there exists a unique type $(s, s') \in \mathcal{T}_0^* \times \mathcal{T}_0^*$ with $e_P(s, s') = d$ and therefore $H(\pi_P) = 0$, and

$$\sum_{(s,s') \in \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*} \mathbb{E}_P \log(E_h(s, s')!) = \sum_{n=0}^{\infty} P(n) \log(n!).$$

It follows that

$$J_1(P) = -s(d) - \sum_{n=0}^{\infty} P(n) \log P(n) - \sum_{n=0}^{\infty} P(n) \log(n!)$$

$$= -s(d) + d - d \log d - \sum_{n=0}^{\infty} P(n) \log \frac{P(n) n!}{d^n e^{-d}} = s(d) - H(P \mid \mathrm{Poi}(d)).$$

This ends the proof of Corollary 1.4.

Remark 5.13 Fix $h \in \mathbb{N}$, and suppose that $\rho \in \mathcal{P}_u(\mathcal{T}^*)$ is such that $\rho_h \in \mathcal{P}_h$. One can derive the following alternative expression for $J_h(\rho_h)$ in terms of relative entropies:

$$J_h(\rho) = s(d) - \sum_{k=1}^{h} \Delta_k(\rho), \tag{82}$$

where $\Delta_1(\rho) = H(\rho_1 \mid \mathrm{Poi}(d))$ and, for $k \geqslant 2$:

$$\Delta_k(\rho) = H(\rho_k \mid \rho_k^*) - \frac{d}{2} H(\pi_{\rho_k} \mid \pi_{\rho_k^*}) \geqslant 0, \tag{83}$$

where $\rho_k^* := [\mathrm{UGW}_{k-1}(\rho_{k-1})]_k$. To prove (82), thanks to Corollary 1.4, it suffices to prove that the increment $J_{k-1}(\rho_{k-1}) - J_k(\rho_k)$ equals (83) for $k \geqslant 2$. This in turn can be checked as follows.

Fix $h \in \mathbb{N}$, and write $Q = \rho_{h+1}$, $Q^* = \rho_{h+1}^*$, $P = \rho_h$. Simple manipulations show that

$$
J_h(P) - J_{h+1}(Q) = -\sum_t P(t) H(Q(\cdot|t)) + \frac{d}{2} \sum_{(s,s')} \pi_P(s, s') H(q(\cdot|s, s'))
$$
$$
- \sum_{(s,s')} \mathbb{E}_Q \left[\log \left(\frac{E_h(s, s')}{\{E_{h+1}(t, t')\}} \right) \right], \tag{84}
$$

where $t \in \mathcal{T}_h^*$ while $(s, s') \in \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*$, we use the multinomial coefficients introduced in (76), and we define the conditional probability $q(\cdot|s, s')$ on $\mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*$ by $\pi_Q(t, t') = \pi_P(s, s') q(t, t'|s, s')$. Using (76) and (77), one finds

$$
H(Q|Q^*) = -\sum_t P(t) H(Q(\cdot|t)) - \sum_{(s,s')} \mathbb{E}_Q \left[\log \left(\frac{E_h(s, s')}{\{E_{h+1}(t, t')\}} \right) \right]
$$
$$
+ d \sum_{(s,s')} \pi_P(s, s') H(\widehat{P}_{s,s'}).
$$

Therefore,

$$
J_h(P) - J_{h+1}(Q) = H(Q|Q^*) + \frac{d}{2} \sum_{(s,s')} \pi_P(s, s') [H(q(\cdot|s, s')) - 2H(\widehat{P}_{s,s'})]. \tag{85}
$$

Next observe that if $q^*(\cdot|s, s') := \widehat{P}_{s,s'}(t) \widehat{P}_{s',s}(t')$, then $\pi_{Q^*}(t, t') = \pi_P(s, s') q^*(\cdot|s, s')$, see Remark 3.4. Moreover, using

$$
\sum_{t' \in \mathcal{T}_h^*: t'_{h-1}=s'} q(t, t'|s, s') = \widehat{P}_{s,s'}(t) = \sum_{t' \in \mathcal{T}_h^*: t'_{h-1}=s'} q^*(t, t'|s, s'),
$$

one finds

$$
H(q(\cdot|s, s')) - 2H(\widehat{P}_{s,s'}) = -H(q(\cdot|s, s')|q^*(\cdot|s, s')).
$$

It follows that

$$
\frac{d}{2} \sum_{(s,s')} \pi_P(s, s') H(q(\cdot|s, s') \,|\, q^*(\cdot|s, s')) = \frac{1}{2} \sum_{(t,t')} \mathbb{E}_Q(E_{h+1}(t, t')) \log \frac{\mathbb{E}_Q(E_{h+1}(t, t'))}{\mathbb{E}_{\bar{\rho}}(E_{h+1}(t, t'))}
$$
$$
= \frac{d}{2} \sum_{(t,t')} \pi_Q(t, t') \log \frac{\pi_Q(t, t')}{\pi_{Q^*}(t, t')} = \frac{d}{2} H(\pi_Q \,|\, \pi_{Q^*}),
$$

where $(s, s') \in \mathcal{T}_{h-1}^* \times \mathcal{T}_{h-1}^*$, while $(t, t') \in \mathcal{T}_h^* \times \mathcal{T}_h^*$. From (85) we then obtain the desired conclusion $J_h(P) - J_{h+1}(Q) = \Delta_{h+1}(\rho)$. Clearly, the monotonicity in Lemma 5.11 implies that $\Delta_{h+1}(\rho) \geqslant 0$. This yields the seemingly nontrivial inequality $\frac{d}{2} H(\pi_Q \mid \pi_{Q^*}) \leqslant H(Q|Q^*)$.

5.4 Discontinuity of the entropy

The aim of this section is to prove that the $\mathcal{P}(\mathcal{G}^*) \to [-\infty, \infty)$ map $\rho \mapsto \Sigma(\rho)$ is discontinuous for the weak topology at $\rho = \mathrm{UGW}_1(P)$ for any finitely supported $P \in \mathcal{P}(\mathbb{Z}_+)$ with $P(0) = P(1) = 0$ and $P(2) < 1$.

Let P_1, P_2 be two probability measures on $\mathbb{Z}_+$ with finite positive means, say d_1 and d_2. For $i = 1, 2$, we set $p_i = d_{\bar{i}}/(d_1 + d_2)$, where $\bar{1} = 2$, $\bar{2} = 1$. We define $\mathrm{UGW}(P_1, P_2)$ as the law of the rooted tree (T, o) obtained as follows. We first build a rooted multi-type Galton–Watson tree $(\check{T}, o)$. The vertices can be of type 1 or of type 2. The root has type i with probability p_i. All offspring of a vertex of type i are of type $\bar{i}$. Conditioned on being of type i, the root has a number of offspring distributed according to P_i. Conditioned on being of type i, a vertex different from the root has a number of offspring distributed according to the size-biased law $\widehat{P_i}$ given by (6). The tree (T, o) is finally obtained from $(\check{T}, o)$ by removing the types.

The distribution of $(\check{T}, o)$ is unimodular. It implies that $\mathrm{UGW}(P_1, P_2)$ is also unimodular. It can be checked directly from the definition of unimodularity or by proving that it is the local weak limit of bipartite configuration models (they are especially of interest in coding theory, see e.g. Montanari and Mézard [25]).

Now, let $S \subset \mathbb{Z}_+$ be a finite set and P be a probability measure on S. Observe that $\mathrm{UGW}(P, P) = \mathrm{UGW}_1(P)$ and that if P_n is a sequence of probability measures on S such that $P_n \rightsquigarrow P$ then $\widehat{P_n} \to \widehat{P}$ and

$$\mathrm{UGW}(P, P_n) \rightsquigarrow \mathrm{UGW}_1(P).$$

However, we have the following discontinuity result:

Proposition 5.14 *Assume that $S \subset \mathbb{Z}_+\backslash\{0, 1\}$, $P, P_n \in \mathcal{P}(S)$, and $P_n \rightsquigarrow P$ as $n \to \infty$. Assume further that $P(2) < 1$ and that $P_n \neq P$ for all n large enough. Then,*

$$\limsup_{n \to \infty} \Sigma(\mathrm{UGW}(P, P_n)) < \Sigma(\mathrm{UGW}(P)).$$

The proposition is a consequence of the following upper bound on $\Sigma(\mathrm{UGW}(P, Q))$

Lemma 5.15 *Let $S \subset \mathbb{Z}_+\backslash\{0,1\}$ be a finite set and $P_1 \neq P_2$ be two probability measures on S. We have*

$$\Sigma(\text{UGW}(P_1, P_2)) \leqslant H((p_1, p_2)) + \sum_{i=1}^{2} p_i H(P_i) + \frac{d}{2} \log\left(\frac{d}{2}\right) - \frac{d}{2}$$

$$- \sum_{i=1}^{2} p_i \mathbb{E}_{P_i} \log(D!),$$

where D is the random variable with law P_1, P_2 respectively, and $H((p_1, p_2)) = -\sum_{i=1}^{2} p_i \log p_i$.

The idea will be to prove that if $U(G_n) \rightsquigarrow \text{UGW}(P_1, P_2)$, then G_n needs to be approximately bipartite. The constraint of being bipartite will be costly in terms of entropy.

Proof of Proposition 5.14 Using Lemma 5.15, with $P_1 = P_n$ and $P_2 = P$, we may upper bound $\Sigma(\text{UGW}(P, P_n))$ by

$$H((p_1(n), p_2(n))) + p_1(n)H(P_n) + p_2(n)H(P) + \frac{d}{2} \log\left(\frac{d}{2}\right) - \frac{d}{2}$$

$$- p_1(n)\mathbb{E}_{P_n} \log(D!) - p_2(n)\mathbb{E}_P \log(D!).$$

Since P_n and P have support in the finite set S, $p_1(n) \to 1/2$, $p_2(n) \to 1/2$, $H(P_n) \to H(P)$ and $\mathbb{E}_{P_n} \log(D!) \to \mathbb{E}_P \log(D!)$. So finally

$$\limsup_{n\to\infty} \Sigma(\text{UGW}(P, P_n)) \leqslant \log(2) + H(P) + \frac{d}{2} \log\left(\frac{d}{2}\right) - \frac{d}{2} - \mathbb{E}_P \log(D!)$$

$$= \Sigma(\text{UGW}_1(P)) - \left(\frac{d}{2} - 1\right) \log(2).$$

Since $P(0) = P(1) = 0$ and $P(2) < 1$, we have $d > 2$. $\qquad\qquad\square$

Proof of Lemma 5.15 Let us start by a remark. We denote by d_i and $\hat{d}_i$ the mean of P_i and $\widehat{P}_i$. Since $\{0,1\} \notin S$, the support of $\widehat{P}_i$ is included in $\{1, \ldots, \theta\}$ for some θ. It follows that $\hat{d}_i \geqslant 1$. Also, $\hat{d}_i = 1$ implies that $\widehat{P}_i = \delta_1$, hence $P_i = \delta_2$. Since $P_1 \neq P_2$, we have that either $\widehat{P}_1$ or $\widehat{P}_2$ is different from δ_1. In particular,

$$\alpha = \sqrt{\hat{d}_1 \hat{d}_2} > 1.$$

Let (T, o) be a rooted tree with distribution $\rho = \text{UGW}(P_1, P_2)$ obtained from a multi-type rooted tree $(\check{T}, o)$ as above whose law is denoted by $\check{\rho}$. We will assign to all vertices of T a type $\{a, b\}$: type a (resp. b) is supposed to be a good approximation for type 1 (resp. 2) in $\check{T}$.

Let $\mathcal{A}_1 \cup \mathcal{A}_2$ be a partition of $\mathcal{P}(\mathbb{Z}_+)$ such that $\widehat{P}_i$ is in the interior of $\mathcal{A}_i$ (it is possible since $\widehat{P}_1 \neq \widehat{P}_2$). Now, for $v \in V(T)$ and integer $h \geqslant 1$, $\partial B(v, h)$ is the set of

vertices at distance h from v in T. The assumption $\{0, 1\} \notin S$ implies that $\partial B(V, h)$ is not empty. Hence, we define

$$\mu_v^h = \frac{1}{|\partial B(v, 2h)|} \sum_{u \in \partial B(v, 2h)} \delta_{\deg_T(u)-1}.$$

Moreover $\alpha > 1$ implies that $\check{\rho}$-a.s.

$$\lim_{h \to \infty} \frac{1}{h} \log |\partial B(o, h)| = \log \alpha > 0. \tag{86}$$

Indeed, we consider a tree T' whose vertex set are the vertices at even distance (in T) from the root. T' is obtained by connecting vertices at distance $2h$ from the root to their grandchildren (the offspring of its own offspring), at distance $2(h + 1)$. Then, by construction, all vertices have the same type in T'. Moreover, conditioned on the root being of type i, T' is a Galton–Watson tree where the root has offspring distribution Q_i, the distribution of $\sum_{k=1}^{N} N_k$, where N has law P_i, independent of $(N_k)_k$ an i.i.d. sequence with law $\widehat{P}_2$ if $i = 1$ and $\widehat{P}_1$ if $i = 2$, and any other vertex in T' has offspring distribution Q_i', the distribution of $\sum_{k=1}^{\widehat{N}} N_k$, where $\widehat{N}$ has law $\widehat{P}_i$, independent of $(N_k)_k$ as above. By construction, Q_i' has mean $\alpha^2 = \hat{d}_1 \hat{d}_2$ and T' has extinction probability 0. Then (86) is a consequence of the Seneta–Heyde Theorem [22,30].

Also, conditioned on the root being of type i, all vertices $u \in B(o, 2h)$ are of type i. It follows that, conditioned on $|\partial B(o, 2h)|$ the vector $(\deg_T(u) - 1)_{u \in \partial B(o, 2h)}$ is i.i.d. with common law $\widehat{P}_i$. Hence, the strong law of large numbers implies that, $\check{\rho}$-a.s.

$$\mu_o^h \rightsquigarrow \widehat{P}_{c(o)}.$$

where $c(o)$ is the type of the root.

In the sequel, we fix $\delta > 0$ and take h large enough such that

$$\min_{i=1,2} \mathbb{P}_{\check{\rho}}(\mu_o^h \in \mathcal{A}_i \,|\, c(o) = i) \geq 1 - \delta.$$

Now, to a locally finite graph $G = (V, E)$, we attach to each vertex $v \in V$ the type $\omega(v) = a$ (resp. $\omega(v) = b$) if $\partial B(v, 2h)$ is not empty, $\deg(v) \in S$ and $\mu_v^h \in \mathcal{A}_1$ (resp. $\mu_v^h \in \mathcal{A}_2$). Otherwise, we set set $\omega(v) = \bullet$.

Let $\bar{a} = b$, $\bar{b} = a$, $\theta = \max(s \in S)$ and $\Theta = \{0, \ldots, \theta\}$. We also attach on the vertices of G a new type in the set $\mathcal{R} = \{\bullet, (a, k), (b, k) : k \in \Theta\}$ defined, for $c \in \{a, b\}$, by $\tau(u) = (c, k)$ if

(i) $\omega(u) = c$;
(ii) $\sum_{v \sim u}^{G} \mathbf{1}(\omega(v) = \bar{c}) = k$.

Otherwise, $\omega(u) = \bullet$ and we also set $\tau(u) = \bullet$. In words: a vertex has τ-type (c, k) if its ω-type is c and it has exactly k of its neighbors having ω-type $\bar{c}$. We may call this scalar k the ab-degree of the vertex.

By construction, $\mathbb{P}_{\check{\rho}}(\omega(o) = c(o)) \geqslant 1 - \delta$. Also, using the union bound and unimodularity,

$$\mathbb{P}_{\check{\rho}}(\exists\, v \overset{T}{\sim} o : \omega(v) \neq c(v)) \leqslant \mathbb{E}_{\check{\rho}} \sum_v \mathbf{1}(v \overset{T}{\sim} o)\mathbf{1}(\omega(v) \neq c(v))$$

$$= \mathbb{E}_{\check{\rho}} \sum_v \mathbf{1}(v \overset{T}{\sim} o)\mathbf{1}(\omega(o) \neq c(o))$$

$$\leqslant \theta\,\mathbb{P}_{\check{\rho}}(\omega(o) \neq c(o)) \leqslant \theta\,\delta.$$

We thus have proved that

$$\mathbb{P}_{\check{\rho}}(\tau(o) = (c(o), \deg(o))) \geqslant 1 - (\theta + 1)\delta.$$

It follows that, for any $k \in S, c \in \{a, b\}$,

$$\left|\mathbb{P}_\rho(\omega(o) = c) - p_i\right| \leqslant \delta \quad \text{and} \quad \left|\mathbb{P}_\rho(\tau(o) = (c, k)) - p_i P_i(k)\right| \leqslant (\theta + 1)\delta, \quad (87)$$

where $i = 1$ if $c = a$ and $i = 2$ if $c = b$. Equation (87) shows that we can nearly reconstruct the types and the bipartite structure from $2h$-neighborhoods.

Also, by construction, the maps $(G, o) \to \omega(o)$ and $(G, o) \to \tau(o)$ are continuous for the local topology. Hence, there exists $\eta(\varepsilon) > 0$ with $\eta(\varepsilon) \to 0$, $\varepsilon \to 0$, such that $\mu \in B(\rho, \varepsilon)$ implies that

$$\max_{c \in \{a, b, \bullet\}} \left|\mathbb{P}_\mu(\omega(o) = c) - \mathbb{P}_\rho(\omega(o) = c)\right| \leqslant \eta(\varepsilon), \quad \text{and}$$

$$\max_{r \in \mathcal{R}} \left|\mathbb{P}_\mu(\tau(o) = r) - \mathbb{P}_\rho(\tau(o) = r)\right| \leqslant \eta(\varepsilon).$$

For all $\varepsilon \leqslant \varepsilon(\delta)$ small enough, $\eta(\varepsilon) \leqslant \delta$.

All ingredients are now in order. Consider a sequence $m = m(n)$ such that $m(n)/n \to d/2$ where $d = 2p_1 d_1 = 2p_2 d_2 = 2d_1 d_2/(d_1 + d_2)$. Let $G_n \in \mathcal{G}_{n,m}(\rho, \varepsilon)$ with $\varepsilon \leqslant \varepsilon(\delta)$. For $c \in \{a, b, \bullet\}$ and $r \in \mathcal{R}$, we set

$$n_c = \sum_{v=1}^n \mathbf{1}(\omega(v) = c) \quad \text{and} \quad N_r = \sum_{v=1}^n \mathbf{1}(\tau(v) = r).$$

From what precedes and (87), for $c \in \{a, b\}$ and $k \in \Theta$,

$$|n_c - np_i| \leqslant 2\delta n \quad \text{and} \quad \left|N_{(c,k)} - np_i P_i(k)\right| \leqslant 2(\theta + 1)\delta n, \quad (88)$$

where $i = 1$ if $c = a$ and $i = 2$ if $c = b$. We notice also that $(n_a, n_b, n_\bullet)$ is an integer partition of n of length 3 and $(n_r)_{r \in \mathcal{R}}$ is an integer partition of length $|\mathcal{R}| = 2(\theta+1)+1$.

We now compute an upper bound for $|\mathcal{G}_{n,m}(\rho, \varepsilon)|$. Fix $\mathbf{n} = ((n_c)_{c \in \{a, b\}}, (N_r)_{r \in \mathcal{R}})$. We denote by $A(\mathbf{n})$ the set of vertex-labeled graphs $G = ([n], E, \omega', \tau')$ such that for any $c \in \{a, b\}$, $r \in \mathcal{R}$ and $v \in [n]$,

(i) $\omega'(v) \in \{a, b, \bullet\}$ and $\tau'(v) \in \mathcal{R}$;

(ii) $\tau'(v) = (c, k)$ iif $\omega'(v) = c$ and $\sum_{u \overset{G}{\sim} v} \mathbf{1}(\omega'(u) = \bar{c}) = k$;

(iii) $n_c = \sum_v \mathbf{1}(\omega'(v) = c)$ and $N_r = \sum_{v=1}^{n} \mathbf{1}(\tau'(v) = r)$.

From what precedes,

$$|\mathcal{G}_{n,m}(\rho, \varepsilon)| \leqslant n^2 n^{|\mathcal{R}|-1} \max_{\mathbf{n}} |A(\mathbf{n})|,$$

where the maximum is over all pairs of integer partitions $((n_c)_{c \in \{a,b,\bullet\}}, (N_r)_{r \in \mathcal{R}})$ satisfying (88).

We set

$$m_\circ = \sum_{k \in \Theta} k N_{(a,k)} = \sum_{k \in \Theta} k N_{(b,k)} \quad \text{and} \quad m_\bullet = m - m_\circ.$$

In words, $m_\circ$ is the number of ab-edges (i.e. adjacent to a vertex of ω'-type a and a vertex of ω'-type b), $m_\bullet$ counts all the other edges. Summing (88) over $c \in \{a, b\}$, $k \in \Theta$, yields

$$n_\bullet = n - n_a - n_b \leqslant 4n\delta,$$

and

$$\left| m_\circ - \frac{nd}{2} \right| = \left| \sum_{k \in \Theta} \left(k N_{(a,k)} - n p_1 k P_1(k) \right) \right| \leqslant 2\theta(\theta + 1)^2 \delta n = O(\delta n).$$

Since $m = m_\bullet + m_\circ = nd/2 + o(n)$. It follows that

$$m_\bullet = O(\delta n) + o(n),$$

where $O(\cdot)$ depends only on θ.

We find

$$|A(\mathbf{n})| \leqslant \binom{n}{(n_a, n_b, n_\bullet)} \binom{n_a}{(N_{(a,k)})_{k \in \Theta}} \binom{n_b}{(N_{(b,k)})_{k \in \Theta}}$$

$$\times \frac{m_\circ!}{\prod_{k \in \Theta}(k!)^{N_{(a,k)} + N_{(b,k)}}} \binom{\frac{n(n-1)}{2}}{m_\bullet},$$

where: the first term counts the number of ways to partition $[n]$ into three blocks of sizes n_a, n_b and $n_\bullet$; the second and third terms subdivide each of the blocks in terms of the ab-degrees of the vertices; the fourth term upper bounds the number of ways to realize the ab-degree sequence (reasoning as in Lemma 4.3); the last term bounds the number of ways to put the remaining $m_\bullet$ edges.

We set $p = (p_a, p_b, p_{\bullet})$ with $p_c = n_c/n$ and for $r = (c, k) \in \mathcal{R}$, $P_c(k) = N_{(c,k)}/n_c$. Using Stirling's approximation, we obtain

$$\log |A(\mathbf{n})| \leqslant nH(p) + np_a H(P_a) + np_b H(P_b) + m_{\circ} \log n + m_{\circ} \log \left(\frac{m_{\circ}}{n} \right) - m_{\circ}$$

$$- np_a \mathbb{E}_{P_a} \log(D!) - np_b \mathbb{E}_{P_b} \log(D!) + m_{\bullet} \log n - m_{\bullet} \log \left(\frac{2m_{\bullet}}{n(n-1)} \right)$$

$$+ m_{\bullet} + o(n),$$

where $o(\cdot)$ depends only on θ. Using our estimates in terms of δ, we get

$$\log |A(\mathbf{n})| - m \log n \leqslant nH((p_1, p_2)) + np_1 H(P_1) + np_2 H(P_2) + \frac{nd}{2} \log \left(\frac{d}{2} \right) - \frac{nd}{2}$$

$$- np_1 \mathbb{E}_{P_1} \log(D!) - np_2 \mathbb{E}_{P_2} \log(D!) + O(n\delta \log \delta^{-1}) + o(n).$$

Letting $n \to \infty$ and then $\delta \to 0$, the lemma follows. $\qquad\square$

6 Large deviation principles

6.1 Proof of Theorem 1.6

Fix a sequence $\mathbf{d} = \mathbf{d}^{(n)}$ as in (C1)–(C3), set $P_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{d(i)}$. The measure $P_n \in \mathcal{P}(\mathbb{Z}_+)$ may be viewed as a measure on rooted graphs with depth 1, i.e. $\mathcal{G}_1^*$, by assigning probability zero to any $g \in \mathcal{G}_1^* \backslash \mathcal{T}_1^*$, and by assigning the weight $P_n(k)$ to the unlabeled star with k neighbors (rooted at the center of the star). Define

$$m = \frac{1}{2} \sum_{i=1}^{n} d_i(n),$$

so that $m/n \to d/2$ as $n \to \infty$, and define the set

$$\mathcal{G}_{P_n} = \{G \in \mathcal{G}_{n,m} : U(G_n)_1 = P_n\}.$$

Each element of $\mathcal{G}(\mathbf{d}_n)$ is isomorphic to exactly $n(\mathbf{d})$ graphs in $\mathcal{G}_{P_n}$, i.e. $n(\mathbf{d})|\mathcal{G}(\mathbf{d})| = |\mathcal{G}_{P_n}|$, where $n(\mathbf{d})$ denotes the number of distinct vectors $(d(\pi_1), \ldots, d(\pi_n))$ as $\pi : [n] \mapsto [n]$ ranges over permutations of the vertex labels. Since $U(G)$ is invariant under isomorphisms, Theorem 1.6 is equivalent to the same statement where G_n is a random graph uniformly distributed in $\mathcal{G}_{P_n}$ rather than in $\mathcal{G}(\mathbf{d})$. Thus, for the rest of this proof G_n will denote a uniform graph in $\mathcal{G}_{P_n}$.

Since $U(G_n)$ is unimodular, we may restrict to the closed subspace $\mathcal{P}_u(\mathcal{G}^*)$. Let $\mathcal{K} \subset \mathcal{P}_u(\mathcal{G}^*)$ denote the compact set of unimodular probability measures supported by graphs with degree bounded by θ. Unimodularity implies that $\rho \in \mathcal{K}$ is equivalent to ρ being supported by graphs such that the degree at the root is bounded by θ. By construction, $U(G_n) \in \mathcal{K}$ and $P \in \mathcal{K}$. Therefore, if $\rho \in \mathcal{P}_u(\mathcal{G}^*)$ is such that $\rho_1 = P$,

then $\rho \in \mathcal{K}$. From general principles, see e.g. [16, Ch. 4], the theorem follows if we prove that: (i) for any $\rho \in \mathcal{K}$ with $\rho_1 = P$, $\delta > 0$,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(U(G_n) \in B(\rho, \delta)) \geqslant \Sigma(\rho) - \Sigma(\mathrm{UGW}_1(P)); \qquad (89)$$

and (ii) for any $\rho \in \mathcal{K}$

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(U(G_n) \in B(\rho, \varepsilon)) \leqslant \begin{cases} \Sigma(\rho) - \Sigma(\mathrm{UGW}_1(P)) & \text{if } \rho_1 = P, \\ -\infty & \text{otherwise.} \end{cases} \qquad (90)$$

To prove the lower bound (89), write

$$\mathbb{P}(U(G_n) \in B(\rho, \delta)) = \frac{|\{G \in \mathcal{G}_{n,m} : U(G_n)_1 = P_n, \ U(G_n) \in B(\rho, \delta)\}|}{|\{G \in \mathcal{G}_{n,m} : U(G_n)_1 = P_n\}|}. \qquad (91)$$

As a consequence of (67)

$$\frac{1}{n}\big(\log |\{G \in \mathcal{G}_{n,m} : U(G_n)_1 = P_n\}| - m \log n\big) \leqslant J_1(P) + o(1).$$

On the other hand, the lower bound in Proposition 5.6 proves that for fixed $\delta > 0$, one has

$$\frac{1}{n}\big(\log |\{G \in \mathcal{G}_{n,m} : U(G_n)_1 = P_n, \ U(G_n) \in B(\rho, \delta)\}| - m \log n\big) \geqslant J_h(\rho_h) + o(1)$$

for all h large enough. From Theorem 1.3 one has $\Sigma(\rho) = \lim_{h \to \infty} J_h(\rho_h)$, and $\Sigma(\mathrm{UGW}_1(P)) = J_1(P)$, and (89) follows.

We turn to the proof of the upper bound (90). We start with the case $\rho_1 \neq P$. For $\delta > 0$, consider the closure, say $F(\delta)$, of the probability measures $\rho \in \mathcal{K}$ such that $d_{TV}(\rho_1, P) \leqslant \delta$. For all n large enough, $U(G_n) \in F(\delta)$, since $U(G_n)_1 = P_n \rightsquigarrow P$. If $\rho_1 \neq P$, then $\rho \notin F(\delta)$ for some $\delta > 0$, and $\mathbb{P}(U(G_n) \in B(\rho, \varepsilon)) = 0$, for all ε small enough and n large enough. It follows that (90) is $-\infty$ in this case. Suppose now that $\rho_1 = P$. For the upper bound one may drop the constraint $U(G_n)_1 = P_n$ in the numerator of (91). Then, using the lower bound in Proposition 5.6 for the denominator and Theorem 1.3 for the numerator, one has the desired estimate.

Remark 6.1 The result of Theorem 1.6 can be extended with no difficulty to the case where G_n is uniformly distributed in the set of all graphs G with vertex set $[n]$ satisfying $U(G)_h = P_n$, where P_n is supported on some fixed set $\Delta = \{t_1, \dots, t_r\} \subset \mathcal{T}_h^*$ for all n, and such that $P_n \rightsquigarrow P$ for some admissible P. Theorem 1.6 is the special case $h = 1$. With the same proof, for any fixed $h \in \mathbb{N}$, one obtains that $U(G_n)$ satisfies the large deviation principle with speed n and good rate function $I(\rho) = J_h(P) - \Sigma(\rho)$ if $\rho_h = P$, and $I(\rho) = +\infty$ otherwise.

6.2 Proof of Theorem 1.7

We start with a proof of exponential tightness. Let $c \geqslant 1$ and let G_n be a random graph sampled uniformly on $\mathcal{G}_{n,m}$, where $m = m(n)$ is an arbitrary sequence satisfying

$$\frac{m(n)}{n} \leqslant \frac{c}{2}.$$

The random probability measure $\rho_n := U(G_n)$ is an element of $\mathcal{P}_u(\mathcal{G}^*)$.

Lemma 6.2 *The sequence of random variables ρ_n is exponentially tight in $\mathcal{P}_u(\mathcal{G}^*)$, i.e. for any $z \geqslant 1$, there exists a compact set $\Pi_z \subset \mathcal{P}_u(\mathcal{G}^*)$ such that*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(U(G_n) \notin \Pi_z) \leqslant -z.$$

Proof For $y \geqslant 1$ and $x \in (0, 1)$, we define

$$\delta_y(x) = -\frac{2y}{\log(cx)},$$

and consider the event,

$$\mathcal{E}_y(n) = \left\{ \forall S \subset [n] : |\deg_{G_n}(S)| \leqslant n \, \delta_y\left(\frac{|S|}{n}\right) \right\},$$

where $\deg_{G_n}(S)$ was defined in (14). We are going to prove that there exists a constant $L > 0$ such that for any real $y \geqslant 1$, for any integer $n \geqslant 1$,

$$\frac{1}{n} \log \mathbb{P}(\mathcal{E}_y(n)^c) \leqslant -y + L. \tag{92}$$

In view of Lemma 2.3, (92) implies the lemma.

To prove (92), we may restrict ourself to subsets $S \subset [n]$ of cardinality at most $|S| \leqslant n\varepsilon_0$, with $\varepsilon_0 = \delta_y^{-1}(1) = e^{-2y}/c \leqslant e^{-2y}$. From the union bound,

$$\mathbb{P}(\mathcal{E}_y(n)^c) \leqslant n \max_{0 < \varepsilon \leqslant \varepsilon_0} \mathbb{P}(\exists S \subset [n] : |S| = \varepsilon n, \ \deg_{G_n}(S) > n\delta_y(\varepsilon)).$$

By choosing y large enough we may assume that $\varepsilon_0 > 0$ is small enough. Choose $\varepsilon \in (0, \varepsilon_0]$ and $\delta := \delta_y(\varepsilon)$. We note that, as in the proof of (54), $\deg_{G_n}(S)$ is stochastically dominated by $2N$, where N has distribution $\text{Bin}(\varepsilon n^2, 2d/n)$. It follows that

$$\mathbb{P}(\mathcal{E}_y(n)^c) \leqslant n \max_{0 < \varepsilon \leqslant \varepsilon_0} \binom{n}{\varepsilon n} \mathbb{P}(N \geqslant \delta n/2).$$

For $x > 0$,

$$\mathbb{P}(N \geqslant \delta n/2) \leqslant e^{-\delta nx} \mathbb{E}[e^{2xN}] = e^{-\delta nx}\left(1 + (2d/n)(e^{2x} - 1)\right)^{\varepsilon n^2} \leqslant e^{-\delta nx + 2d\varepsilon n e^{2x}}.$$

Taking $x = -\frac{1}{2}\log(c\varepsilon)$ one finds

$$\frac{1}{n}\log \mathbb{P}(N \geqslant n\delta/2) \leqslant \frac{\delta}{2}\log(c\varepsilon) + \frac{2d}{c} = -y + \frac{2d}{c}.$$

On the other hand, from Stirling's formula, there exists a constant C such that

$$\binom{n}{n\varepsilon} \leqslant C\sqrt{n}e^{nH(\varepsilon)},$$

where $H(\varepsilon) = -\varepsilon \log \varepsilon - (1 - \varepsilon)\log(1 - \varepsilon)$. Since $\varepsilon \leqslant \varepsilon_0 = e^{-2y}$, these bounds imply the desired conclusion (92). $\qquad \square$

We turn to the proof of Theorem 1.7. Fix $d > 0$ and a sequence $m = m(n)$ such that $m/n \to d/2$, as $n \to \infty$. Thanks to Lemma 6.2, from general principles, see e.g. [16, Ch. 4], it is sufficient to establish: (i) for any $\rho \in \mathcal{P}_u(\mathcal{G}^*)$ and $\delta > 0$,

$$\liminf_{n \to \infty} \frac{1}{n}\log \mathbb{P}(U(G_n) \in B(\rho, \delta)) \geqslant \Sigma(\rho) - s(d); \qquad (93)$$

and (ii) for any $\rho \in \mathcal{P}_u(\mathcal{G}^*)$

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{1}{n}\log \mathbb{P}(U(G_n) \in B(\rho, \varepsilon)) \leqslant \Sigma(\rho) - s(d). \qquad (94)$$

However, both the lower bound (93) and the upper bound (94) follow immediately from the definition of $\Sigma(\rho)$, Theorem 1.2 and (7). This ends the proof.

6.3 Proof of Theorem 1.8

Theorem 1.8 is a simple consequence of Theorem 1.7. We argue as in [17]. Let G_n denote the random graph with distribution $\mathcal{G}(n, \lambda/n)$, and let $M(n)$ be the total number of edges in G_n. Then $M(n)$ is the binomial random variable $\text{Bin}(n(n-1)/2, \lambda/n)$. Conditioned on a given value $M(n) = m$, G_n has uniform distribution over $\mathcal{G}_{n,m}$. It follows that $\mathcal{G}(n, \lambda/n)$ is a mixture of the uniform distribution on $\mathcal{G}_{n,m}$, where m is sampled according to $\text{Bin}(n(n-1)/2, \lambda/n)$. We use the following simple lemma, whose proof is omitted.

Lemma 6.3 *The sequence $2M(n)/n$ satisfies the LDP in $[0, \infty)$ with speed n and good rate function*

$$j(x) = \frac{1}{2}(\lambda - x + x\log(x/\lambda)).$$

We need to prove that $\rho_n = U(G_n)$ satisfies a LDP on $\mathcal{P}_u(\mathcal{G}^*)$ with speed n and good rate function

$$I(\rho) = j(d) - \Sigma(\rho) + s(d) = \inf\{j(r) - \Sigma_r(\rho) + s(r) : r \geqslant 0\},$$

where $\rho \in \mathcal{P}_u(\mathcal{G}^*)$, $d = \mathbb{E}_\rho \deg_G(o)$ and $\Sigma_r(\rho)$ is the entropy of ρ associated to the mean degree r (which is equal to $-\infty$ if $r \neq d$ by Theorem 1.2). A simple adaptation of the proof of Lemma 6.2 shows that the random variable $\rho_n = U(G_n)$ is exponentially tight. The conclusion follows from a general result on large deviations for mixtures; see Biggins [**?**, Theorem 5(b)].

6.4 Proof of Corollary 1.9 and Corollary 1.10

The proof is an application of the contraction principle, cf. [16]. Concerning Corollary 1.9, by Theorem 1.7 one has that $u(G_n)$ satisfies the LDP in $\mathcal{P}(\mathbb{Z}_+)$ with speed n and good rate function

$$K(P) = \inf\{s(d) - \Sigma(\rho), \ \rho \in \mathcal{P}(\mathcal{G}^*) : \rho_1 = P\}.$$

From Theorem 1.3 and Corollary 1.4 this expression equals $s(d) - J_1(P) = H(P \mid \mathrm{Poi}(d))$.

As for Corollary 1.10, by Theorem 1.8 $u(G_n)$ satisfies the LDP in $\mathcal{P}(\mathbb{Z}_+)$ with speed n and good rate function

$$K(P) = \inf\{\phi(\lambda, d) - \Sigma(\rho), \ \rho \in \mathcal{P}(\mathcal{G}^*) : \rho_1 = P\},$$

where $\phi(\lambda, d) = \frac{\lambda}{2} - \frac{d}{2}\log\lambda$. Since all $\rho \in \mathcal{P}(\mathcal{G}^*)$ with $\rho_1 = P$ have the same expected degree at the root, this equals $\phi(\lambda, d) - J_1(P) = \phi(\lambda, d) - s(d) + H(P \mid \mathrm{Poi}(d))$.

7 Appendix : Local convergence for generalized configuration model

In this section we prove Theorem 4.8.

7.1 The exploration process

The first step is to prove convergence of the average measure $\mathbb{E}U(G_n)$, where G_n has distribution $\mathrm{CM}(\mathbf{D}^{(n)})$.

Proposition 7.1 *Let $G_n \in \widehat{\mathcal{G}}(\mathbf{D}^{(n)})$ with distribution $\mathrm{CM}(\mathbf{D}_n)$ such that assumptions (H1)–(H2) hold (see Sect. 4.3). Then, $\mathbb{E}U(G_n) \rightsquigarrow \mathrm{UGW}(P)$.*

The proof of Proposition 7.1 is based on an exploration process of the neighborhood of a vertex. We shall use the notation of Sect. 4. For ease of notation, we will often omit the dependence on n from our notation. Let $\mathbf{D} = (D(1), \ldots, D(n)) \in \mathcal{D}_n, \sigma \in \Sigma$ and $G = \Gamma(\sigma)$ the associated multigraph. To be precise, we specify the set $W = \cup_{c \in \mathcal{C}} W_c$ to be $W_c = \{(c, i, j) : i \in [n], 1 \leqslant j \leqslant D_c(i)\}$ and $W(i) = \{(c, i, j) : c \in \mathcal{C}, 1 \leqslant j \leqslant D_c(i)\}$ the set of half-edges of all colors starting from i. With a slight abuse of notation, we will sometimes write for $e = (c, i, j) \in W$, $\sigma(e)$ in place of $\sigma_c(i, j)$.

Let $\mathbb{N}^f = \cup_{k \geqslant 0} \mathbb{N}^k$ where $\mathbb{N}^0 = o$. We consider the total order on $\mathbb{N}^f$: $\mathbf{i} < \mathbf{j}$ with $\mathbf{i} = (i_1, \ldots, i_k), \mathbf{j} = (j_1, \ldots, j_\ell)$ if either $k < \ell$ or $k = \ell$ and $\mathbf{i} < \mathbf{j}$ for the lexicographical order. We will define a bijective map ϕ from a finite set $S \subset \mathbb{N}^f$ to the vertex set of $G(v)$. The value of ϕ is defined iteratively and if $\mathbf{i} < \mathbf{j}$ are in S then the value of $\phi(\mathbf{i})$ will be determined before the value of $\phi(\mathbf{j})$. Moreover $\phi(S \cap \mathbb{N}^k)$ will be the set of vertices at distance k from v.

The exploration is on the set of half-edges W and it is defined recursively. At integer step t, we partition W in 3 sets: an half-edge may belong to the *active* set $A(t)$, to the *unexplored* set $U(t)$ or to the *connected* set $C(t) = W \backslash (A(t) \cup U(t))$. At stage t, a vertex with an half-edge in $C(t) \cup A(t)$ will have a pre-image via ϕ in $\mathbb{N}^f$. We start with a given $v \in [n]$, and fix the initial conditions $A(0) = W(v), C(0) = \emptyset$, $U(0) = W \backslash W(v)$, and $\phi(o) = v$.

For integer $t \geqslant 0$, if $A(t) \neq \emptyset$, let $e_{t+1} = (c_t, \phi(\mathbf{i}_t), j_t)$ be an half-edge in $A(t)$ such that $\mathbf{i}_t$ is minimal for the total order on $\mathbb{N}^f$. Let $I(t+1) = (W(v_{t+1}) \backslash \{\sigma(e_{t+1})\}) \cap U(t)$ where v_{t+1} is the vertex such that $\sigma(e_{t+1}) \in W(v_{t+1})$. I_{t+1} is the set of new half-edges and our partition of W is updated as

$$
\begin{cases}
A(t+1) = A(t) \backslash \{e_{t+1}, \sigma(e_{t+1})\} \bigcup I_{t+1} \\
U(t+1) = U(t) \backslash (I(t+1) \cup \{\sigma(e_{t+1})\}) \\
C(t+1) = C(t) \cup \{e_{t+1}, \sigma(e_{t+1})\}.
\end{cases}
\tag{95}
$$

If $\sigma(e_{t+1}) \notin A(t)$, we also set $\phi((\mathbf{i}_t, j_t)) = v_{t+1}$. Finally, if $A(t) = \emptyset$, then the exploration process stops.

We notice that the elements in $C(t)$ are the half-edges for which we know by step t their matched half-edge. It implies that $\sigma(e_{t+1}) \in A(t) \cup U(t)$. Moreover, for any vertex u, we cannot have simultaneously $W(u) \cap U(t) \neq \emptyset$ and $W(u) \cap A(t) \neq \emptyset$. With a slight abuse, we may thus write $u \in U(t)$ or $u \in A(t)$ if, respectively, $W(u) \cap U(t) \neq \emptyset$ or $W(u) \cap A(t) \neq \emptyset$. Now, if $v_{t+1} \in U(t)$, then $I(t+1) = W(v_{t+1}) \backslash \{\sigma(e_{t+1})\}$, otherwise $v_{t+1} \in A(t)$ and $I(t+1) = \emptyset$. Note also that for integer k, the image by ϕ of the vertices of generation k in S, $\phi(S \cap \mathbb{N}^k)$, are the set of vertices in G at distance k from v (by recursion, this comes from the fact that $\mathbf{i}_t$ is minimal for the total order on $\mathbb{N}^f$).

We now define $X(0) = D(v)$ and for integer $t \geqslant 1$, $X_c(t+1) = |\{(i, j) : (c, i, j) \in I_{t+1}\}|$. Hence $X(t) \in \mathcal{M}_L$ gives the new colored half-edges attached to v_t. For ease of notation, we also set

$$
\varepsilon_c(t+1) = \mathbf{1}(v_{t+1} \in A(t), c_t = \bar{c}), \quad \delta_c(t+1) = \mathbf{1}(c_t = c),
$$

and

$$\tau = \inf\{t : A(t) = \emptyset\}.$$

Setting $A_c = A \cap W_c$, $U_c = U \cap W_c$ and $C_c = C \cap W_c$, we get

$$|A_c(t)| = D_c(v) + \sum_{k=1}^{t}(X_c(k) - \delta_c(k) - \varepsilon_c(k))$$

$$|U_c(t)| = |W_c| - D_c(v) - \sum_{k=1}^{t}(X_c(k) + \delta_{\bar{c}}(k) - \varepsilon_c(k))$$

$$|C_c(t)| = \sum_{k=1}^{t}(\delta_c(k) + \delta_{\bar{c}}(k)). \tag{96}$$

Note that $|C_c(t)| = |C_{\bar{c}}(t)|$ and, if $c \in \mathcal{C}_=$, $|C_c(t)|$ is even.

Now, as in the statement of Proposition 7.1, consider a random multi-graph G_n with distribution $\mathrm{CM}(\mathbf{D}^{(n)})$. For integer $t \geqslant 0$, we consider the filtration

$$\mathcal{F}_t = \sigma((A(0), U(0), C(0)), \dots, (A(t), U(t), C(t))).$$

The hitting time τ is a stopping time for this filtration. Also, given $\mathcal{F}_t$, if $\{t < \tau\}$ and $c_t = c \in \mathcal{C}_{\neq}$, then $\sigma(e_{t+1})$ is uniformly distributed on $U_{\bar{c}}(t) \cup A_{\bar{c}}(t)$. It follows that for $u \in [n]$,

$$\mathbb{P}(v_{t+1} = u | \mathcal{F}_t) = \frac{|W_{\bar{c}}(u) \cap (U_{\bar{c}}(t) \cup A_{\bar{c}}(t))|}{|U_{\bar{c}}(t)| + |A_{\bar{c}}(t)|}$$

$$= \frac{\mathbf{1}_{u \in U(t)} D_{\bar{c}}(u)}{|U_{\bar{c}}(t)| + |A_{\bar{c}}(t)|} + \frac{\mathbf{1}_{u \in A(t)} |W_{\bar{c}}(u) \cap A(t)|}{|U_{\bar{c}}(t)| + |A_{\bar{c}}(t)|}.$$

Similarly, given $\mathcal{F}_t$, if $\{t < \tau\}$ and $c_t = c \in \mathcal{C}_=$, $\sigma(e_{t+1})$ is uniformly distributed on $U_c(t) \cup A_c(t) \backslash \{e_{t+1}\}$. We find in this case,

$$\mathbb{P}(v_{t+1} = u | \mathcal{F}_t) = \frac{|W_c(u) \cap (U_c(t) \cup A_c(t) \backslash \{e_{t+1}\})|}{|U_c(t)| + |A_c(t)| - 1}$$

$$= \frac{\mathbf{1}_{u \in U(t)} D_c(u)}{|U_c(t)| + |A_c(t)| - 1} + \frac{\mathbf{1}_{u \in A(t)}(|W_c(u) \cap A_c(t)| - \mathbf{1}_{e_{t+1} \in W_c(u)})}{|U_c(t)| + |A_c(t)| - 1}.$$

In either case, for $c \in \mathcal{C}$, if $\sigma(e_{t+1}) \in U(t)$, then $X(t+1) = D(v_{t+1}) - E^{\bar{c}}$ otherwise, $\sigma(e_{t+1}) \in A(t)$ and $X(t+1) = 0$. We recall also that $|U_c(t)| + |A_c(t)| = |W_c| -$

$|C_c(t)| = |W_{\bar{c}}| - |C_{\bar{c}}(t)|$. We get, for $M \in \mathcal{M}_L$, if $c_t = c$ then

$$
\begin{aligned}
&\mathbb{P}(X(t+1) = M | \mathcal{F}_t) \\
&= \begin{cases}
\sum_{u \in U(t)} \frac{\mathbf{1}_{D(u)=M+E^{\bar{c}}}(M_{\bar{c}}+1)}{|W_c|-|C_c(t)|-\mathbf{1}(c \in \mathcal{C}_=)} & M \neq 0, \\
\sum_{u \in U(t)} \frac{\mathbf{1}_{D(u)=E^{\bar{c}}}}{|W_c|-|C_c(t)|-\mathbf{1}(c \in \mathcal{C}_=)} + \frac{|A_{\bar{c}}(t)|-\mathbf{1}(c \in \mathcal{C}_=)}{|W_c|-|C_c(t)|-\mathbf{1}(c \in \mathcal{C}_=)} & M = 0.
\end{cases}
\end{aligned}
\tag{97}
$$

Observe that, from (96) and assumption (H1), we find for any $c \in \mathcal{C}$,

$$
|A_c(t)| \leqslant \theta(t+1) \quad \text{and} \quad |C_c(t)| \leqslant 2t.
\tag{98}
$$

The next lemma computes the limiting marginals of the exploration process.

Lemma 7.2 *Under the assumption of Proposition* 7.1, *let o be uniformly distributed on* $[n]$, *independently of* G_n, *and consider the exploration process on the rooted graph* $(G_n(o), o)$. *For any integer* $t \geqslant 0$, *as* $n \to \infty$:

(i) $X(0)$ *converges weakly to* P.
(ii) *Let* $c \in \mathcal{C}$ *be such that* $\mathbb{E}D_c > 0$. *Given* $\mathcal{F}_t$, *if* $\{t < \tau\}$ *and* $c_t = c$, *then the conditional law of* $X(t+1)$ *given* $\mathcal{F}_t$ *converges weakly to* $\widehat{P}^c$.
(iii) *The probability that there exist* $c \in \mathcal{C}$ *and an integer* $1 \leqslant s \leqslant t \wedge \tau$ *such that* $\mathbb{E}D_c = 0$ *and* $c_s = c$ *goes to* 0.

Proof Since $X(0) = D(o)$, statement (i) is simply a restatement of the assumption (H2).

For statement (ii), we first note that the set $\{i \in [n] : i \notin U(t)\}$ has cardinality bounded by $1 + \theta L^2 t$. It follows by (97) that, if $\{t < \tau\}$ and $c_t = c$ hold, for any $M \in \mathcal{M}_L$,

$$\left| \mathbb{P}(X(t+1) = M | \mathcal{F}_t) - \frac{M_{\bar{c}} + 1}{|W_c| - |C_c(t)| - \mathbf{1}(c \in \mathcal{C}_=)} \sum_{i=1}^{n} \mathbf{1}_{D(i)=M+E^{\bar{c}}} \right|$$

$$\leqslant \frac{(M_{\bar{c}} + 1) \left(1 + \theta L^2 t + |A_{\bar{c}}(t)|\right)}{|W_c| - |C_c(t)| - \mathbf{1}(c \in \mathcal{C}_=)}.$$

Now, assumptions (H1)–(H2) imply that $|W_c|/n$ converges to $\mathbb{E}D_c$, where D has law P. Similarly, assumption (H2) implies that $\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{D(i)=M+E^{\bar{c}}}$ converges to $P(M+E^{\bar{c}})$. Hence, from (98), if $\mathbb{E}D_c > 0$, then $\mathbb{P}(X(t+1) = M | \mathcal{F}_t)$ converges to $\widehat{P}^c(M)$. This proves statement (ii).

We now turn to statement (iii). We set $\mathcal{C}_0 = \{c \in \mathcal{C} : D_c \equiv 0\}$ and $A_{\mathcal{C}_0}(t) = \cup_{c \in \mathcal{C}_0} A_c(t)$. We recall that $\mathbb{E}D_{\bar{c}} = \mathbb{E}D_c$, hence $c \in \mathcal{C}_0$ is equivalent to $\bar{c} \in \mathcal{C}_0$. We should prove that for any integer $t \geqslant 0$, $\mathbb{P}(|A_{\mathcal{C}_0}(t \wedge \tau)| \geqslant 1) \to 0$. First, statement (i) and the union bound implies that $\mathbb{P}(|A_{\mathcal{C}_0}(0)| \geqslant 1) \leqslant \sum_{c \in \mathcal{C}_0} \mathbb{P}(X_c(0) \geqslant 1) \to 0$. By recursion, it is thus sufficient to prove that for any integer $t \geqslant 0$, $c \in \mathcal{C}_0$, $c' \in \mathcal{C}\backslash\mathcal{C}_0$, if $\{t < \tau\}$ and $c_t = c'$ hold, then

$$\mathbb{P}(X_c(t+1) \geqslant 1 | \mathcal{F}_t) \to 0.$$

The latter follows from statement (ii) (recall that $\bar{c} \in \mathcal{C}_0$). $\qquad\square$

We introduce a variable that counts the number of times that two elements in the active sets are matched by step t:

$$E(t) = \sum_{k=1}^{t} \sum_{c \in \mathcal{C}} \varepsilon_c(k).$$

Lemma 7.3 *Under the assumption of Proposition 7.1, let o be uniformly distributed on $[n]$, independently of G_n, and consider the exploration process on the rooted graph $(G_n(o), o)$. For every integer $t \geqslant 0$, we have*

$$\lim_{n \to \infty} \mathbb{P}\left(E(t \wedge \tau) \neq 0\right) = 0.$$

If $t \leqslant \tau$ and $E(t) = 0$, the subgraph of G_n spanned by the vertices with all their half-edges in $C(t)$ is an directed colored tree.

Proof We start with the second statement. To every vertex u with an half-edge in $C(t) \cup A(t)$, there is an element $\mathbf{i}$ in $\mathbb{N}^f$ such that $\phi(\mathbf{i}) = u$. We may thus order these vertices by the order through ϕ^{-1} in $\mathbb{N}^f$. Every such vertex is adjacent to its parent. By construction if $E(t) = 0$ or equivalently if for all $1 \leqslant s \leqslant t$, all $c \in \mathcal{C}$, $\varepsilon_c(s) = 0$, then every vertex with an half-edge in $C(t) \cup A(t)$ has a unique adjacent vertex with a smaller index. It follows that there cannot be a cycle in the subgraph spanned by these vertices.

If $E(t \wedge \tau) \neq 0$, there exists an integer $1 \leqslant s \leqslant t \wedge \tau$ such that $\sigma(e_s) \in A(s-1)$. Using (97), it follows from the union bound and the fact that $\{s < \tau\} \in \mathcal{F}_s$,

$$\mathbb{P}(\exists 1 \leqslant s \leqslant t \wedge \tau : \sigma(e_s) \in A(s-1)) \leqslant \mathbb{E}\left[\sum_{s \geqslant 0} \mathbf{1}_{s < t \wedge \tau} \mathbb{P}(v_{s+1} \in A(s)|\mathcal{F}_s)\right]$$
$$\leqslant \mathbb{E} \sum_{s=0}^{t-1} \frac{|A_{\bar{c}_s}(s)|}{|W_{c_s}| - |C_{c_s}(s)| - \mathbf{1}(c_s \in \mathcal{C}_=)}.$$

From (98), for each $t \geqslant 0$, $|A_c(t)| \leqslant \theta(t+1)$. Also, by assumptions (H1)–(H2), $|W_c|/n$ converges to $\mathbb{E}D_c$, where D has law P. If $\mathbb{E}D_c = 0$, we may appeal to Lemma 7.2(iii). $\square$

All ingredients of the proof of Proposition 7.1 are now gathered.

Proof of Proposition 7.1 Let o be uniformly distributed on $[n]$, independently of G_n. We set $\overline{\rho}_n = \mathbb{E}U(G_n)$, and $\rho = \mathrm{UGW}(P)$. Define $B = \{g \in \widehat{\mathcal{G}}^* : g_t = \gamma\}$ where γ is the equivalence class of a finite rooted directed colored tree of depth at most t. It is sufficient to prove that for any integer $t \geqslant 1$ and any such γ, $\overline{\rho}_n(B)$ converges to $\rho(B)$.

For some $m = \sum_{k=0}^{t-1}(\theta L^2)^k$, $(G_n(o), o)_t$ has at most m vertices. However, by Lemma 7.3, with high probability, $E_{m \wedge \tau} = 0$ and $(G_n(o), o)_t$ is a rooted directed colored tree. Applying now Lemma 7.2, we deduce that

$$\lim_{n \to \infty} |\overline{\rho}_n(B) - \rho(B)| = \lim_{n \to \infty} |\mathbb{P}((G_n(o), o)_t \simeq \gamma) - \rho(B)| = 0.$$

The conclusion follows. $\square$

A.2 Concentration inequalities

We are going to state a concentration inequality for the configuration model. We use the notation of Sect. 4. We fix an integer $L \geqslant 1$ and consider a set of colors $\mathcal{C} = \{(i, j) : 1 \leqslant i, j \leqslant L\}$, $\mathbf{D} = (D(1), \dots, D(n)) \in \mathcal{D}_n$ and $\Sigma = \Sigma(\mathbf{D})$ be the set of configurations. We shall say that $m \in \Sigma$ and $m' \in \Sigma$ differ by at most one switch if there exists $c \in \mathcal{C}_\leqslant$ such that for all $c' \neq c$, $m_{c'} = m'_{c'}$ and a set $J \subset W_c$, with $|J| \leqslant 2$ if $c \in \mathcal{C}_\neq$ or $|J| \leqslant 4$ if $c \in \mathcal{C}_=$, and for all $x \in W_c \backslash J, m(x) = m'(x)$. In other words, if $c \in \mathcal{C}_\neq, m'_c \circ m_c^{-1}$ is either the identity ($|J| = 0$) or a transposition ($|J| = 2$). Similarly, for $c \in \mathcal{C}_=, m'_c \circ m_c^{-1}$ is either the identity ($|J| = 0$) or the composition of two disjoint transpositions ($|J| = 4$).

In the special case $L = 1$, the next proposition appears in Wormald [31, Theorem 2.19].

Proposition 7.4 *Let* $\mathbf{D} = (D(1), \dots, D(n)) \in \mathcal{D}_n$, $\Sigma = \Sigma(\mathbf{D})$ *be the set of configurations,* $S = \sum_{i=1}^n D(i)$ *and* $N = \sum_{c \in \mathcal{C}} S_c$. *Let* $F : \Sigma \to \mathbb{R}$ *be a function such that*

for some $\kappa > 0$ and any $m, m' \in \Sigma$ which differ by at most one switch, we have

$$|F(m) - F(m')| \leqslant \kappa.$$

Then, if σ is uniformly sampled from Σ, for any $t > 0$,

$$\mathbb{P}\left(|F(\sigma) - \mathbb{E}F(\sigma)| \geqslant t\right) \leqslant 2 \exp\left(\frac{-t^2}{\kappa^2 N}\right).$$

The proof will be given in Sect. 7.1.2 below.

Corollary 7.5 *Let $\mathbf{D} = (D(1), \dots, D(n)) \in \mathcal{D}_n$ such that (H1) holds (see Sect. 4.3). Let $k \geqslant 0$, γ be a rooted directed colored multigraph and $A = \{g : g_k = \gamma_k\}$. There exists a constant $\delta = \delta(\theta, k, L) > 0$, such that, if $G_n \overset{d}{\sim} CM(\mathbf{D})$, $\rho_n = U(G_n)$ and $t > 0$,*

$$\mathbb{P}\left(|\rho_n(A) - \mathbb{E}\rho_n(A)| \geqslant t\right) \leqslant 2 \exp\left(-\delta n t^2\right).$$

Proof By assumption we have for any $c \in \mathcal{C}, i \in [n]$, $D_c(i) \leqslant \theta$. We may thus assume without loss of generality that γ has degrees bounded by θ. We set

$$f(\sigma) := n\rho_n(A) = \sum_{i=1}^{n} \mathbf{1}((G_n(i), i)_k \simeq \gamma_k).$$

The number of vertices in G_n which are at distance at most k from both endpoints of any given edge is bounded by $\kappa = 2\sum_{s=0}^{k-1}(\theta L^2)^s$. If two configurations m, m' in Σ differ by at most one switch then $|f(m) - f(m')| \leqslant 4\kappa$. Indeed, a switch changes the status at most 4 edges and the addition or the removal of an edge can modify for at most κ vertices the value of $\mathbf{1}((G_n(i), i)_k \simeq \gamma_k)$. It remains to apply Proposition 7.4, with $F(\sigma) = f(\sigma)/n$ and $N = O(n)$. □

7.1.1 Proof of Theorem 4.8

Let us start with the case $G_n \overset{d}{\sim} CM(\mathbf{D}^{(n)})$. We set $\rho_n = U(G_n)$ and $\rho = UGW(P)$. Let $k \geqslant 0$, $\gamma \in \widehat{\mathcal{G}}_*$ and $A = \{g \in \widehat{\mathcal{G}}^* : g_k = \gamma_k\}$. Corollary 7.5, Proposition 7.1 and Borel–Cantelli's Lemma imply that with probability one, $\rho_n(A) \to \rho(A)$. The collection of sets $A = A(k, \gamma), k \geqslant 0, \gamma \in \widehat{\mathcal{G}}^*$, being a basis of the topology on $\widehat{\mathcal{G}}^*$, this proves the first statement of Theorem 4.8.

For the second statement, we notice that if G_n is uniformly distributed on $\mathcal{G}(\mathbf{D}^{(n)}, h)$ and $\widehat{G}_n \overset{d}{\sim} CM(\mathbf{D}^{(n)})$, then Lemma 4.4 implies for any subset $B \subset \mathcal{G}(\mathbf{D}^{(n)}, h)$ that

$$\mathbb{P}(G_n \in B) \leqslant \mathbb{P}(\widehat{G}_n \in B)/\mathbb{P}(\widehat{G}_n \in \mathcal{G}(\mathbf{D}^{(n)}, h)).$$

By Corollary 4.6, $\mathbb{P}(\widehat{G}_n \in \mathcal{G}(\mathbf{D}^{(n)}, h))$ is lower bounded by some $\alpha > 0$, uniformly in n (depending of the sequence $\mathbf{D}^{(n)}$ and h). Then, if $\rho_n = U(G_n)$ and $\widehat{\rho}_n = U(\widehat{G}_n)$,

from what precedes, for any $t > 0$, and A as above

$$\sum_n \mathbb{P}(|\rho_n(A) - \mathbb{E}\widehat{\rho}_n(A)| > t) \leqslant \alpha^{-1} \sum_n \mathbb{P}(|\widehat{\rho}_n(A) - \mathbb{E}\widehat{\rho}_n(A)| > t) < \infty.$$

It remains to apply again Borel-Cantelli's lemma and Proposition 7.1.   □

Remark 7.6 The proof of Theorem 4.8 actually shows that for a sequence $\mathbf{D}^{(n)}$ satisfying (H1)–(H2) the following holds. Let $k \geqslant 0$, γ fixed and $A = \{g : g_k = \gamma_k\}$. There exists a constant $\delta > 0$ (depending of the sequence $\mathbf{D}^{(n)}$, h and k), such that, if G_n is uniformly distributed on $\mathcal{G}(\mathbf{D}^{(n)}, h)$ and $\widehat{G}_n \overset{d}{\sim} \mathrm{CM}(\mathbf{D}^{(n)})$, we have for any $t > 0$

$$\mathbb{P}\left(|\rho_n(A) - \mathbb{E}\widehat{\rho}_n(A)| \geqslant t\right) \leqslant \delta^{-1} \exp\left(-\delta n t^2\right), \tag{99}$$

where $\rho_n = U(G_n)$ and $\widehat{\rho}_n = U(\widehat{G}_n)$.

7.1.2 Proof of Proposition 7.4

The proof is a consequence of Azuma–Hoeffding's inequality.

Case of random matchings   For clarity, we start with the case $L = 1$, i.e. $\mathcal{C} = \{(1, 1)\}$. Then $N = S_{(1,1)}$. We order the elements of $W = W_{(1,1)}$ in the lexicographic order. We may identify a matching of W as the set of $N/2$ matched pairs. We order these $N/2$ pairs by the index of their smallest element. We then define $\mathcal{F}_0$ as the trivial σ-algebra and for $1 \leqslant k \leqslant N/2$, we define $\mathcal{F}_k$ as the σ-algebra generated by the first k pairs of matched elements of σ. We set $Z_k = \mathbb{E}[F(\sigma)|\mathcal{F}_k]$, so that $Z_0 = \mathbb{E}F(\sigma)$, $Z_{N/2-1} = F(\sigma)$. By construction, Z_k is a Doob martingale.

If A is a finite set, we denote by $\mathbf{M}(A)$ the set of perfect matchings on A. With our previous notation $\Sigma = \mathbf{M}(W)$. For $1 \leqslant k \leqslant N/2$, an element σ of $\mathbf{M}(W)$ can be uniquely decomposed into $(\sigma_{k-1}^-, \sigma_k^+)$ where σ_{k-1}^- is the restriction of σ to the $k-1$ smallest pairs and σ_k^+ is the rest. Let W^{k-1} denote the subset of W such that σ_{k-1}^- is a perfect matching on W^{k-1}.

If v_k is the smallest element of $W \backslash W^{k-1}$, we set $w_k = \sigma(v_k) \in W \backslash W^{k-1}$, so that $W^k = W^{k-1} \cup \{v_k, w_k\}$. Now, for $w \in W \backslash (W^{k-1} \cup \{v_k\})$, let $\mathbf{M}_w$ denote the set of matchings of $W \backslash W^{k-1}$ such that $m(v_k) = w$. Then for any $w, w' \in W \backslash (W^{k-1} \cup \{v_k\})$, each $m \in \mathbf{M}_w$ corresponds to a unique $m' \in \mathbf{M}_{w'}$ through the switch $\{\{v_k, w\}, \{w', z\}\} \rightarrow \{\{v_k, w'\}, \{w, z\}\}$, where $m(w') = z$. This gives a bijection between $\mathbf{M}_w$ and $\mathbf{M}_{w'}$, and we set $N_k = |\mathbf{M}_w|$. By assumption, we deduce that for any w, w',

$$\left| \sum_{m \in \mathbf{M}_w} F(\sigma_k^-, m) - \sum_{m \in \mathbf{M}_{w'}} F(\sigma_k^-, m) \right| \leqslant \kappa.$$

Applying the above inequality to w_k, we deduce that

$$\left| \frac{1}{N_k} \sum_{m \in \mathbf{M}_{w_k}} F(\sigma_k^-, m) - \frac{1}{N - 2k + 1} \sum_{w \in W \setminus (W^{k-1} \cup \{v_k\})} \frac{1}{N_k} \sum_{m \in \mathbf{M}_w} F(\sigma_k^-, m) \right|$$

$$= |Z_k - Z_{k-1}| \leqslant \kappa.$$

We may then apply Azuma–Hoeffding's inequality to the martingale Z_k. We obtain that for any $t > 0$,

$$\mathbb{P}(Z_n - Z_0 \geqslant t) \leqslant \exp\left(\frac{-t^2}{2 \sum_{i=1}^{N/2-1} \kappa^2} \right) \leqslant \exp\left(\frac{-t^2}{N\kappa^2} \right).$$

This proves the proposition when $L = 1$.

Case of random bijections   Now, let σ be a uniformly drawn bijection from the set W to $\bar{W}$ of common cardinality N. We order the elements of the set W as $(x_1, \ldots, x_N)$. We introduce the filtration $\mathcal{F}_k$ generated by $\sigma(x_1), \ldots, \sigma(x_k)$. Now let F be a function on the set of bijections from W to $\bar{W}$ such that for any m, m' such that $m' \circ m^{-1}$ is a transposition, we have $|F(m) - F(m')| \leqslant \kappa$. We set $Z_k = \mathbb{E}[F(\sigma)|\mathcal{F}_k]$.

With minor modifications, the above argument shows that, for $1 \leqslant k \leqslant N$, $|Z_{k-1} - Z_k| \leqslant \kappa$. Then, by Azuma–Hoeffding's inequality, we find for $t \geqslant 0$,

$$\mathbb{P}(F(\sigma) - \mathbb{E}F(\sigma) \geqslant t) \leqslant \exp\left(\frac{-t^2}{2N\kappa^2} \right). \tag{100}$$

General case   It suffices to combine the two results. We first order the elements of $\mathcal{C}_\leqslant$ in an arbitrary way, say $c_1, \ldots, c_\ell$ with $\ell = L(L+1)/2$. Let $N_0 = 0$, for $1 \leqslant k \leqslant \ell$, $N_k = \sum_{i=1}^{k} S_{c_i}/(1 + \mathbf{1}_{c_i \in \mathcal{C}_=})$. We have $N_\ell = N/2$. We define the filtration, $(\mathcal{F}_t), 0 \leqslant t \leqslant N_\ell$, built as follows. $\mathcal{F}_0$ is the trivial σ-algebra, $\mathcal{F}_{N_k}$ is the filtration generated by the independent variables $(\sigma_{c_i}), 1 \leqslant i \leqslant k$. Finally for $1 \leqslant i < N_{k+1} - N_k$, $\mathcal{F}_{N_k+i}$ is the filtration generated by $\mathcal{F}_{N_k}$ and the first i matched pairs of $\sigma_{c_{k+1}}$.

As above, we set $Z_k = \mathbb{E}[F(\sigma)|\mathcal{F}_k]$, so that $Z_0 = \mathbb{E}F(\sigma)$, $Z_{N_\ell} = F(\sigma)$. By construction, using the independence of $(\sigma)_c, c \in \mathcal{C}$, we find, for $1 \leqslant k \leqslant N_\ell$, $|Z_{k-1} - Z_k| \leqslant \kappa$. Hence, Azuma–Hoeffding's inequality implies that (100) holds for σ uniform in Σ, with N replaced by $N_\ell = N/2$. This ends the proof of Proposition 7.4.

References

1. Aldous, D., Lyons, R.: Processes on unimodular random networks. Electron. J. Probab. **12**(54), 1454–1508 (2007)
2. Aldous, D., Steele, J.M.: The objective method: probabilistic combinatorial optimization and local weak convergence. In: Probability on discrete structures. Encyclopaedia of Mathematical Sciences, vol. 110, pp. 1–72. Springer, Berlin (2004)

3. Benjamini, I., Lyons, R., Schramm, O.: Unimodular random trees. arXiv:1207.1752 (2012)
4. Benjamini, I., Schramm, O.: Recurrence of distributional limits of finite planar graphs. Electron. J. Probab. **6**(23), 13 (2001)
5. Biskup, M., Chayes, L., Smith, S.A.: Large-deviations/thermodynamic approach to percolation on the complete graph. Random Struct. Algorithms **31**(3), 354–370 (2007)
6. Bollobás, B.: A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. Eur. J. Combin. **1**(4), 311–316 (1980)
7. Bollobás, B.: Random graphs. Cambridge Studies in Advanced Mathematics, vol. 73, 2nd edn. Cambridge University Press, Cambridge (2001)
8. Bollobás, B., Riordan, O.: Sparse graphs: metrics and random models. Random Struct. Algorithms **39**(1), 1–38 (2011)
9. Bordenave, C., Caputo, P.: A large deviation principle for Wigner matrices without Gaussian tails. Ann. Probab. **42**(6), 2454–2496 (2014)
10. Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T., Vesztergombi, K.: Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. Adv. Math. **219**(6), 1801–1851 (2008)
11. Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T., Vesztergombi, K.: Convergent sequences of dense graphs II. Multiway cuts and statistical physics. Ann. Math. (2), **176**(1), 151–219 (2012)
12. Boucheron, S., Gamboa, F., Léonard, C.: Bins and balls: large deviations of the empirical occupancy process. Ann. Appl. Probab. **12**(2), 607–636 (2002)
13. Bowen, L.: Periodicity and circle packings of the hyperbolic plane. Geom. Dedicata **102**, 213–236 (2003)
14. Chatterjee, S., Varadhan, S.R.S.: The large deviation principle for the Erdős–Rényi random graph. Eur. J. Combin. **32**(7), 1000–1017 (2011)
15. Dembo, A., Mörters, P., Sheffield, S.: Large deviations of Markov chains indexed by random trees. Ann. Inst. H. Poincaré Probab. Stat. **41**(6), 971–996 (2005)
16. Dembo, A., Zeitouni, O.: Large deviations techniques and applications. Springer, Berlin (2010). Reprint of second edition
17. Doku-Amponsah, K., Mörters, P.: Large deviation principles for empirical measures of colored random graphs. Ann. Appl. Probab. **20**(6), 1989–2021 (2010)
18. Durrett, R.: Random graph dynamics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (2010)
19. Elek, G.: On the limit of large girth graph sequences. Combinatorica **30**(5), 553–563 (2010)
20. Engel, A., Monasson, R., Hartmann, A.K.: On large deviation properties of Erdős–Rényi random graphs. J. Stat. Phys. **117**(3–4), 387–426 (2004)
21. Erdős, P., Gallai, T.: Graphs with prescribed degrees of vertices (hungarian). Mat. Lapok **11**, 264–274 (1960)
22. Heyde, C.: Extension of a result of Seneta for the super-critical Galton–Watson process. Ann. Math. Stat. **41**, 739–742 (1970)
23. Janson, S., Luczak, T., Rucinski, A.: Random graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, New York (2000)
24. Lovász, L., Szegedy, B.: Limits of dense graph sequences. J. Combin. Theory Ser. B **96**(6), 933–957 (2006)
25. Mézard, M., Montanari, A.: Information, physics, and computation. Oxford Graduate Texts. Oxford University Press, Oxford (2009)
26. Molloy, M., Reed, B.: A critical point for random graphs with a given degree sequence. In: Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993), vol. 6, pp. 161–179 (1995)
27. O'Connell, N.: Some large deviation results for sparse random graphs. Probab. Theory Related Fields **110**(3), 277–285 (1998)
28. Puhalskii, A.A.: Stochastic processes in random graphs. Ann. Probab. **33**(1), 337–412 (2005)
29. Rivoire, O.: Properties of atypical graphs from negative complexities. J. Stat. Phys. **117**(3–4), 453–476 (2004)
30. Seneta, E.: On recent theorems concerning the supercritical Galton–Watson process. Ann. Math. Stat. **39**, 2098–2102 (1968)
31. Wormald, N.C.: Models of random regular graphs. In: Surveys in Combinatorics, 1999 (Canterbury). London Mathematicala Society Lecture Note Series, vol. 267, pp. 239–298. Cambridge University Press, Cambridge (1999)