Large deviation for uniform graphs with given degrees

Souvik Dhara†;‡, Subhabrata Sen†

† Department of Mathematics, Massachusetts Institute of Technology,
‡ Microsoft Research, New England

August 20, 2019

Abstract

Consider the random graph sampled uniformly from the set of all simple graphs with a given degree sequence. Under mild conditions on the degrees, we establish a Large Deviation Principle (LDP) for these random graphs, viewed as elements of the graphon space. As a corollary of our result, we obtain LDPs for functionals continuous with respect to the cut metric, and obtain an asymptotic enumeration formula for graphs with given degrees, subject to an additional constraint on the value of a continuous functional. Our assumptions on the degrees are identical to those of Chatterjee, Diaconis and Sly (2011), who derived the almost sure graphon limit for these random graphs.

1 Introduction

In a seminal paper, Chatterjee and Varadhan [12] initiated a study of large deviations for random graphs, and introduced a novel framework that synergizes the classical theory of Large Deviations with the theory of dense graph limits (Lovász [25]). They embedded Erdős-Rényi random graphs into the space of graphons, equipped with the cut-metric, and derived an LDP for the corresponding sequence of probability measures. As an important consequence, this yields LDPs for continuous functionals in the cut-metric topology, e.g. subgraph counts, largest eigenvalue, etc. Their result resolved a long-standing open question regarding large-deviations for sub-graph counts of dense Erdős-Rényi random graphs. This area has witnessed rapid developments subsequently — we refer the interested reader to Chatterjee’s Saint-Flour lecture notes [8] for a detailed history of these problems and an elaborate description of recent breakthroughs.

Numerous scientific applications naturally motivate the study of graphs with topological constraints, such as a fixed number of edges, triangles etc (see e.g. [14, 30, 38]). The desire to understand typical properties of constrained graphs motivates the study of random graphs, sampled uniformly, subject to these constraints. Natural examples include the Erdős-Rényi uniform random graph with a constrained number of edges, random regular
graphs \[19\], etc. In statistical physics parlance, these can be thought of as microcanonical ensembles, whereas unconstrained graphs, like Erdős-Rényi, correspond to canonical ensembles \[16, 34\]. A rigorous study of constrained graphs often turns out to be extremely challenging—in fact, even enumerating the total number of graphs, subject to combinatorial constraints, is exceedingly non-trivial, and has attracted significant attention recently in Probability, Combinatorics, and Statistical Physics (see e.g. \[2, 23, 31, 32, 33, 35, 37\]). The study of large deviations in this context is of natural interest—indeed, this has deep, natural connections to the problem of counting graphs with atypical properties, subject to the topological constraints. Recently, Dembo and Lubetzky \[15\] initiated a study of large deviations for constrained random graphs, and derived an LDP for dense Erdős-Rényi uniform random graphs, conditioned to have a fixed number of edges.

In this paper, we study the uniform random graph with a given degree sequence. The degrees are assumed to scale linearly in the number of vertices, so that we have a dense random graph. Such random graphs are used extensively in Physics \[34\] and Statistics \[4\], and have a rich history in Combinatorics \[2, 5, 36\]. In general, this model is intractable to theoretical analysis. In fact, characterizing the first order asymptotics of simple functionals like triangle counts is challenging in this case. In a breakthrough paper, Chatterjee, Diaconis and Sly \[11\] derived that, under fairly mild conditions (see Assumption \[1\]), these random graphs converge almost surely in the cut-metric, and identified the limit. Our main result, Theorem \[1.2\], establishes an LDP for uniform random graphs under identical conditions as \[11\]. This general theorem has two important corollaries. The first corollary (Corollary \[1.4\]) yields LDPs for continuous functionals such as subgraph counts. The second corollary (Corollary \[1.5\]) yields the convergence of the microcanonical partition function. Further, it provides the asymptotic count of graphs with given degrees, subject to an additional constraint on the value of a continuous functional, in terms of a variational formula.

Conceptually, the problem under consideration is significantly more challenging than the Erdős-Rényi case, due to the absence of edge-independence in these models. Further, in sharp contrast to the setting of Dembo and Lubetzky \[15\], the number of degree constraints grows linearly with the number of vertices in the graph. To overcome this issue, we crucially exploit a deep idea put forth in \[11\]— these random graphs may be sampled using appropriate inhomogeneous random graphs, conditioned to have the desired degrees (see Section \[4.1\]). Unfortunately, even with access to this ingredient, one still faces substantial technical obstacles due to the inhomogeneity of the unconstrained model. Our proofs require a very delicate understanding of the cut-topology, and deviate significantly from the established techniques for the dense Erdős-Rényi model. To the best of our knowledge, this is the first instance where an LDP has been derived in inhomogeneous settings. Finally, we remark that requisite analytic properties of the candidate rate function, such as lower-semicontinuity, are not obvious here, and require careful analysis.

The rest of the paper is organised as follows: In Section \[1.1\], we set up the framework necessary to state our main result. The statement of the main result and its corollaries is provided in Section \[1.2\]. In Section \[1.3\], we discuss the relevant literature and collect some open problems surfacing from our work. Section \[2\] derives important analytic properties of the rate function. In Section \[3\], we prove a large deviation upper bound for inhomogeneous random graphs. The proof of Theorem \[1.2\] is completed in Section \[4\]. Finally, we prove Corollaries \[1.4\] and \[1.5\] in Section \[5\].
1.1 Definitions and concepts

1.1.1 Graphons and the cut metric

A graphon is a measurable function \( W : [0,1]^2 \mapsto [0,1] \) that is symmetric, i.e., \( W(x,y) = W(y,x) \) for all \( x, y \in [0,1] \). To define the cut metric, let \( \mathcal{M} \) denote the set of all bijective, Lebesgue measure preserving maps \( \phi : [0,1] \mapsto [0,1] \). The cut distance between two graphons \( W_1 \) and \( W_2 \) is given by

\[
d_{\square}(W_1, W_2) = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} (W_1(x,y) - W_2(x,y)) \, dx \, dy \right|,
\]

and the cut metric is given by

\[
\delta_{\square}(W_1, W_2) = \inf_{\phi \in \mathcal{M}} d_{\square}(W_1, \phi^* W_2)
\]

where \( \phi^* W \) denotes the cut metric is given by \( \phi^* \). Setting \( \tilde{\mathcal{W}} \) to denote the space of all graphons, define the equivalence relation \( W_1 \sim W_2 \) if \( W_1 = \phi^* W_2 \) for some \( \phi \in \mathcal{M} \), and consider the quotient space \( \tilde{\mathcal{W}} = \mathcal{W} / \sim \). Note that \( (\tilde{\mathcal{W}}, \delta_{\square}) \) is a compact metric space [26, Theorem 5.1]. Henceforth, for any \( W \in \mathcal{W} \), we always write \( \tilde{W} \) to denote the equivalence class of \( W \) in \( \tilde{\mathcal{W}} \). Also, we simply write \( \delta_{\square}(W_1, W_2) \) for \( \delta_{\square}(\tilde{W}_1, \tilde{W}_2) \).

**Definition 1** (Empirical graphon). For a graph \( G_n \) with vertex set \([n] = \{1, \ldots, n\}\) and edge set \( E(G_n) \), the empirical graphon \( W^{G_n} \) is given by

\[
W^{G_n}(x,y) = \begin{cases} 
1 & \text{if } (i,j) \in E(G_n), \ (x,y) \in \left[ \frac{i-1}{n}, \frac{i}{n} \right) \times \left[ \frac{j-1}{n}, \frac{j}{n} \right), \\
0 & \text{otherwise}.
\end{cases}
\]

**Definition 2** (Graph Convergence). \( (G_n)_{n \geq 1} \) is said to converge in \( (\tilde{\mathcal{W}}, \delta_{\square}) \) if their empirical graphons converge.

**Definition 3** (Subgraph densities). For a finite simple graph \( H = (V(H), E(H)) \) with \( V(H) = [k] \), the subgraph density of \( H \) in \( W \) is defined as

\[
t(H,W) := \int_{[0,1]^k} \prod_{(i,j) \in E(H)} W(x_i,x_j) \prod_{i=1}^k dx_i.
\]

Note that \( t(H,W) = t(H,\phi^* W) \) for any \( \phi \in \mathcal{M} \), and thus \( t(H,\cdot) \) is well defined on \( \tilde{\mathcal{W}} \). Moreover, [7, Theorem 3.7] shows that \( t(H,\cdot) \) is Lipschitz continuous on \((\tilde{\mathcal{W}}, \delta_{\square})\) for any finite simple graph \( H \).

**Definition 4** (Degree distribution function). For any \( W \in \mathcal{W} \), the degree distribution function is defined by

\[
deg_{W}(\lambda) = \lambda \left\{ x : \int_0^1 W(x,y) \, dy \leq \lambda \right\},
\]

where \( \Lambda \) denotes the Lebesgue measure on \([0,1]\), and \( \lambda \in [0,1] \). Observe that \( \deg_{W} \) is well-defined on \( \tilde{\mathcal{W}} \). We write \( \deg_{\tilde{W}} \) to denote the degree distribution function of \( \tilde{W} \in \tilde{\mathcal{W}} \).

**Definition 5** (Graphons away from boundary). A graphon \( W \) is said to be \textit{away from boundary} if there exists an \( \eta > 0 \) such that \( \eta < W(x,y) < 1 - \eta \). A sequence \( (W_n)_{n \geq 1} \) is said to be away from boundary if for all \( n \geq 1 \), the above holds for some \( \eta > 0 \) (independent of \( n \)).
1.1.2 Uniform graphs with given degrees

Consider a sequence of degree sequences \((d^n)_{n \geq 1}, d^n = (d^n_i)_{i \in [n]}\). Without loss of generality, we will assume that the degree sequence is non-increasing, i.e., \(d^n_1 \geq \cdots \geq d^n_n\). For clarity of notation, we will simply write \(d^n_i = d_i^n\), and \(d^n = d^n\), and suppress the dependence of the degrees on \(n\). Let \(G_{n,d}\) denote the uniformly chosen random graph with degree sequence \(d^n\).

Of course, not all sequences \(d^n\) are valid degree sequences of simple graphs. Such sequences are called graphical, and they are characterized by the celebrated Erdős-Gallai theorem. This theorem establishes that \(d^n\) is graphical if and only if \(\sum_{i \in [n]} d_i^n\) is even and for all \(k \in [n]\)

\[
\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{i=k+1}^{n} \min\{d_i, k\}.
\] (1.6)

Thus \(G_{n,d}\) is defined whenever (1.6) holds. Chatterjee, Diaconis and Sly [11] obtained the graphon limit of \(G_{n,d}\) when the degrees converge, and the degree sequence lies in the interior of an asymptotic Erdős-Gallai boundary (1.6). We state below the precise assumptions from [11], which will also be the underlying assumption for our large deviation result:

**Assumption 1.** The degree sequence \(d^n\) satisfies the following:

1. There exists a non-decreasing function \(D : [0,1] \to [0,1]\) such that

\[
\lim_{n \to \infty} \left( \left| \frac{d_1^n}{n} - D(0) \right| + \left| \frac{d_2^n}{n} - D(1) \right| + \frac{1}{n} \sum_{i=1}^{n} \left| \frac{d_i^n}{n} - D\left(\frac{i}{n}\right) \right| \right) = 0.
\] (1.7)

2. There exists constants \(0 < c_1 < c_2 < 1\) such that, \(\forall x \in [0,1]\), \(c_1 \leq D(x) \leq c_2\), and

\[
\int_{x}^{1} \min\{D(y), x\}dy + x^2 - \int_{0}^{x} D(y)dy > 0.
\] (1.8)

We write \(\mathbb{P}_{n,d}\) to denote the probability measure on \(\mathcal{W}\) associated to the empirical graphon of \(G_{n,d}\), and write \(\hat{\mathbb{P}}_{n,d}\) to denote the corresponding push forward measure on \((\hat{\mathcal{W}}, \hat{\delta}_\infty)\). The following was proved in [11, Theorem 1.1]:

**Proposition 1.1 ([11, Theorem 1.1]).** Under Assumption 1, almost surely \(\bigotimes_{n \geq 1} \mathbb{P}_{n,d}\) \((G_{n,d})_{n \geq 1}\) converges to the graphon \(W_D\) in the cut-metric, as \(n \to \infty\), where \(W_D\) is given by

\[
W_D(x, y) := \frac{e^{\beta(x)+\beta(y)}}{1 + e^{\beta(x)+\beta(y)}},
\] (1.9)

where \(\beta : [0,1] \to \mathbb{R}\) is the unique function satisfying \(D(x) = \int_{0}^{1} \frac{e^{\beta(x)+\beta(y)}}{1 + e^{\beta(x)+\beta(y)}} dy\), for all \(x \in [0,1]\).

Note that \(W_D\) is away from the boundary for any degree function \(D\) satisfying Assumption 1 in the sense of Definition 5.
1.2 Main results

Our main result, Theorem 1.2, stated below, derives an LDP for the sequence of probability measures \( \mathbb{P}_{n,d} \). To this end, for the convenience of the reader, we start with recalling the formal notion of a large deviation principle (LDP). Let \( \mathcal{X} \) be a Polish space with Borel sigma-algebra \( \mathcal{B} \). Let \( I : \mathcal{X} \rightarrow [\rho, \infty) \) be a lower semi-continuous function. A sequence of probability measures \( (\mathbb{P}_n)_{n \geq 1} \) on \( (\mathcal{X}, \mathcal{B}) \) satisfies a large deviation principle (LDP) with speed \( s_n \searrow \infty \) and good rate function \( I \) if

(i) for all \( \alpha \geq 0 \), the level sets \( \{ x : I(x) \leq \alpha \} \) are compact,

(ii) for any closed set \( F \subset \mathcal{X} \) and open set \( U \subset \mathcal{X} \)

\[
\limsup_{n \to \infty} \frac{1}{s_n} \log \mathbb{P}_n(F) \leq -\inf_{x \in F} I(x) \quad \text{and} \quad \liminf_{n \to \infty} \frac{1}{s_n} \log \mathbb{P}_n(U) \geq -\inf_{x \in U} I(x). \quad (1.10)
\]

Next, we introduce the candidate rate function in our context. For \( W, W_0 \in \mathbb{W} \) with \( 0 < W_0 < 1 \) a.s., we define

\[
I_{W_0}(W) = \frac{1}{2} \int_{[0,1]^2} \left[ W(x,y) \log \left( \frac{W(x,y)}{W_0(x,y)} \right) + (1 - W(x,y)) \log \left( \frac{1 - W(x,y)}{1 - W_0(x,y)} \right) \right] dx dy
\]

\[
= \frac{1}{2} \sup_{a} \int_{[0,1]^2} \left[ a(x,y)W(x,y) - \log \left( W_0(x,y)e^{a(x,y)} + 1 - W_0(x,y) \right) \right] dx dy,
\]

where the supremum over \( a \) in the final term ranges over all functions in \( L^2([0,1]^2) \) satisfying \( a(x,y) = a(y,x) \) for all \( (x,y) \in [0,1]^2 \) [8, Lemma 5.2]. Unlike the rate function for the Erdős-Rényi random graph in [12, (7)], the function \( I_{W_0}(\cdot) \) is not well-defined on the quotient space \( \mathbb{W} \), i.e., \( I_{W_0}(W) \) is not necessarily equal to \( I_{W_0}(W^\phi) \), for \( \phi \in \mathbb{M} \). However, we can modify the function as follows. Let \( \mathbb{B}_\infty(\hat{W}, \delta) = \{ W' : \delta_{\mathbb{W}}(W,W') \leq \delta \} \), and define

\[
J_{W_0}(\hat{W}) = \sup_{\delta > 0} \inf_{W' \in \mathbb{B}_\infty(\hat{W}, \delta)} I_{W_0}(W') = \inf_{\phi \in \mathbb{M}} I_{W_0}(W^\phi) = \inf_{\phi \in \mathbb{M}} I_{W_0}(W^\phi). \quad (1.12)
\]

The second equality is not obvious, and is proved in Lemma 2.3 for graphons \( W_0 \) that are away from the boundary. The function \( J_{W_0} \) is indeed lower semi-continuous (see Lemma 2.1), i.e., the lower level sets \( \{ W : J_{W_0}(W) \leq \alpha \} \) are closed, and therefore compact due to the compactness of \( (\mathbb{W}, \delta_{\mathbb{W}}) \). Thus, \( J_{W_0} \) is a good rate function. Next recall the definition of \( W_D \) from (1.9). The degree distribution function of \( W_D \) is the inverse of \( D \), i.e.,

\[
\mu_D([0,\lambda]) = \deg_{W_D}(\lambda) = \Lambda \{ x : D(x) \leq \lambda \}. \quad (1.13)
\]

and define

\[
J_D(\hat{W}) = \begin{cases} J_{W_D}(\hat{W}) & \text{if } \deg_{\hat{W}}(\lambda) = \mu_D([0,\lambda]), \forall \lambda \in [0,1], \\ \infty & \text{otherwise.} \end{cases}
\]
Theorem 1.2. Under Assumption 1, the sequence of probability measures \((\mathbb{P}_{n,d})_{n \geq 1}\) on \((\mathcal{W}, \delta_\square)\) satisfies an LDP with speed \(n^2\) and good rate function \(J_D\) defined in (1.14).

For the particular case of a random \(d\)-regular graph, Assumption 1 holds when \(d = \lfloor np \rfloor\) for some \(p \in (0, 1)\) (see [11, Remark 3]), and thus Proposition 1.1 and Theorem 1.2 are applicable. In this case, \(W_D = p\), and the LDP rate function simplifies. This follows as \(I_p(W^\phi) = I_p(W)\) for any \(\phi \in \mathcal{M}\), and thus the rate function is obtained from the well-known rate function for the Erdős-Rényi random graph derived in [12], by constraining on the degrees to be constant. Define

\[
J_p(W) = \begin{cases} 
I_p(W) & \text{if } \deg_{\mathcal{W}}(\lambda) = \mathbb{1}\{p \leq \lambda\}, \forall \lambda \in [0, 1], \\
\infty & \text{otherwise}.
\end{cases} \tag{1.15}
\]

The following corollary states the corresponding LDP for the random regular graph. Let \(p \in (0, 1)\) and \(d = \lfloor np \rfloor\). Consider the degree sequence \(d = d_1\), and for this case simply denote the probability measure associated to the random regular graph by \(\mathbb{P}_{n,d}\).

Corollary 1.3. The sequence of probability measures \((\mathbb{P}_{n,d})_{n \geq 1}\) on \((\mathcal{W}, \delta_\square)\) satisfies an LDP with speed \(n^2\) and good rate function \(J_p\) defined in (1.15).

As the main application of their LDP, Chatterjee and Varadhan [12] derived the LDPs for subgraph counts of Erdős-Rényi random graphs. Under the constraint on the number of edges, Dembo and Lubetzky [15] also proved LDP results for subgraph counts. Below we state the corresponding results for \(G_{n,d}\).

Let \(\tau : \mathcal{W} \rightarrow \mathbb{R}_+\) be bounded and continuous in \((\mathcal{W}, \delta_\square)\). The LDP statement for \(\tau\) below will directly imply the LDP for subgraph counts of \(G_{n,d}\), using the continuity of subgraph counts. Define the rate functions

\[
\phi_\tau(D, r) = \inf \{ J_D(W) : \tau(W) \geq r \}, \tag{1.16}
\]

\[
\psi_\tau(D, r) = \inf \{ J_D(W) : \tau(W) \geq r \}. \tag{1.17}
\]

Also, let us denote \(\mathcal{W}_0 = \{ W \in \mathcal{W} : \deg_W \equiv \mu_D \}\) and

\[
l_\tau(D) = \tau(W_D), \quad r_\tau(D) = \sup \{ r : \{ W \in \mathcal{W}_0 : \tau(W) \geq r \} \neq \emptyset \}. \tag{1.18}
\]

Let \(\tau_{n,d}\) be the value of \(\tau\) computed on the empirical graphon of \(G_{n,d}\). Below we state the LDP result for \(\tau_{n,d}\):

Corollary 1.4. Let \(\tau\) be a bounded, continuous function on \((\mathcal{W}, \delta_\square)\). Then the following are true:

(1) The function \(\phi_\tau(D, \cdot)\) is zero on \([0, l_\tau(D)]\), and finite, strictly positive and strictly increasing on \((l_\tau(D), r_\tau(D))\). Further, \(\phi_\tau(D, \cdot)\) left-continuous.

(2) For \(r \in [0, l_\tau(D)]\), \(W_D\) is the unique minimizer in (1.16), and for \(r \in (l_\tau(D), r_\tau(D)]\), the set of minimizers in (1.16) and (1.17) coincide.

(3) Let \(r\) be any right continuity point of \(\phi_\tau(D, \cdot)\). Then, under Assumption 1,

\[
\lim_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}(\tau_{n,d} \geq r) = -\phi_\tau(D, r). \tag{1.19}
\]
Let $F_*$ be the set of minimizers in (1.16). Under Assumption 1, for all $\varepsilon > 0$, there exists a constant $C = C(\varepsilon, \tau, D, r) > 0$ such that
\[
\limsup_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}(\delta_\Box(W^{G_n,d}, F_*) \geq \varepsilon | \tau_n, d \geq r) \leq -C.
\]  

Chatterjee and Diaconis [10] used the LDP for Erdős-Rényi random graphs to evaluate the limit of the partition function associated with exponential random graphs [10, Theorem 3.1]. In a related direction, setting $G_{n,d}$ to be the set of all simple graphs on $n$ vertices with degree sequence $d$, we consider the probability measure on $G_{n,d}$ defined by
\[
P_{n,d,\tau}(G) = e^{n^2(\tau(\tilde{W}_G) - Z_{n,\tau})}, \quad \forall G \in G_{n,d},
\]  
where $\tau$ is a bounded continuous function on $(\tilde{W}, \delta_\Box)$, and $Z_{n,\tau} = \frac{1}{n^2} \log \sum_{G \in G_{n,d}} e^{n^2(\tilde{W}_G)}$.

We will refer to $Z_{n,\tau}$ as the microcanonical partition function. Its limiting value is naturally associated with the enumeration problem of graphs with given degrees and constrained sub-graph counts (see (1.24) below). Our next corollary derives the limit of the micro-canonical partition function. To this end, define the entropy function
\[
h_\varepsilon(W) = \frac{1}{2} \int_{[0,1]^2} W(x,y) \log(W(x,y)) + (1 - W(x,y)) \log(1 - W(x,y)) \, dx \, dy.
\]  

Finally, let $N_{n,\tau}(d, r)$ denote the number of graphs $G \in G_{n,d}$ with $\tau(\tilde{W}_G) \geq r$.

**Corollary 1.5.** Let $\tau$ be a bounded continuous function on $(\tilde{W}, \delta_\Box)$. Under Assumption 1,
\[
Z_{\tau} = \lim_{n \to \infty} Z_{n,\tau} = \sup_{\tilde{W} \in \mathcal{W}} (\tau(\tilde{W}) - J_D(\tilde{W})) + h_\varepsilon(W_D).
\]  

Moreover, for any continuity point $r$ of $\phi_\tau(D, \cdot)$,
\[
\lim_{n \to \infty} \frac{1}{n^2} \log N_{n,\tau}(d, r) = -\phi_\tau(D, r) + h_\varepsilon(W_D).
\]

**1.3 Discussion**

**The rate function.** In the theory of large deviations, the rate function corresponds to the minimum cost of changing the base measure to one where the rare event becomes typical. Unlike Erdős-Rényi random graphs, the typical graphon, as specified by Proposition 1.1, is not invariant under measure preserving transformations. This results in distinct costs of transportation, depending on the labeling of the vertices. This intuitively suggests the rate function $\inf_{\phi \in \mathcal{W}} I_{W_\phi}(\cdot)$. However, this is an infimum over an uncountable set, and thus lower continuity of the candidate rate function is not obvious. On the other hand, the lower semicontinuous envelope, given by $\sup_{\delta > 0} \inf_{W_\phi \in \mathcal{W} \cap [W_\delta, W]} I_{W_\phi}(W^n)$, provides an alternative natural candidate. Fortunately, one can show equalities in (1.12), and the intuition matches the analytical requirements of an LDP.
The variational problem. Corollary 1.4 characterizes the probability of a rare event in terms of a variational problem (1.16). From the perspective of large deviation theory, the natural follow up question concerns the structure of $G_{n,d_{r}}$ conditioned on the rare event. Using (1.20), this conditional structure corresponds to the minimizers of (1.16). The variational problem (1.16) has attracted significant attention in the Erdős-Rényi case. For instance, it is now understood that in the so-called replica symmetric regime, conditioned on the upper tail event for triangle counts, the graph is close to an Erdős-Rényi with a higher edge density [27]. Note that the replica symmetric regime is no longer tenable under exact constraints, such as a fixed number of edges, triangles, degrees, etc. In a set of related papers, [20, 21, 22, 32] study the structure of the minimizer under constraints on the edge, triangle or star counts, and discover intriguing characteristics of the minimizers. However, to the best of our knowledge, this problem has not been studied under degree constraints. We expect this case to be considerably more challenging than the prior settings.

A careful reader has noticed that Corollary 1.4 (3) holds when $r$ is a continuity point of $\phi_{\tau}(D, \cdot)$. For Erdős-Rényi random graphs, the continuity of this function has been established, when $\tau$ represents a subgraph density, the largest eigenvalue, etc. [8, 27]. Their proof is perturbative, and the idea does not generalize to the setting with given degrees. In fact, $\phi_{\tau}(D, \cdot)$ could be degenerate in constrained spaces. For example, the largest eigenvalue of random $d$-regular graphs equals $d$, and thus the rate function is degenerate. More generally, a deterministic function of the degrees, e.g. any $k$-star density, is constant in this case, and gives rise to degenerate rate functions.

Counting graphs with given degrees and subgraph densities. Counting graphs with given degrees has been studied extensively in Combinatorics [2, 24, 29, 35]. These counts are often described in terms of a variational problem such as (1.24). For example, [2, Theorem 1.4] evaluates the leading asymptotics of the number of graphs with given degrees, and expresses it in terms of an entropy maximization problem. Corollary 1.5 yields a formula for the asymptotic number of graphs with given degrees and a specified subgraph count. However, this description is completely implicit, and explicit solutions for general degree sequences could be significantly challenging.

The sparse regime. The breakthrough result of Chatterjee and Varadhan [12] completely resolved the question of large deviations for subgraph counts of dense Erdős-Rényi random graphs. The corresponding question for sparse Erdős-Rényi random graphs $G(n, p)$, $p \to 0$, is considerably more challenging, and has intrigued researchers in Probability and Combinatorics for a long time. For any fixed graph $H$ and $\delta > 0$, the infamous upper tail problem sought to understand the probability that the number of copies of $H$ in $G(n, p)$ exceeds $(1 + \delta)$ times its expectation. To address this challenging question, Chatterjee and Dembo [9] initiated the theory of non-linear large deviations. They establish that for any fixed subgraph $H$ and $\delta > 0$, the upper tail probability reduces to a variational problem on the space of weighted graphs whenever $p \to 0$, $p \geq n^{-\alpha_{H}}$. Remarkably, the variational problem was solved in the special case where $H$ is a clique by Lubetzky and Zhao [28] shortly thereafter. Subsequently, Bhattacharya et al. [3] resolved this question for all fixed subgraphs. Following the initial breakthrough of Chatterjee and Dembo [9], the exponent $\alpha_{H}$ was improved considerably by Eldan [17]. Recently, Cook and Dembo [13], Augeri [1], and Harel et al. [18] have further improved the bounds on $\alpha_{H}$, deriving the optimal exponent for certain specific subgraphs such as cycles, cliques, regular graphs etc.
These exciting recent developments have dramatically improved our understanding of the upper tail problem on sparse Erdős-Rényi random graphs. It would be fascinating to answer this question for sparse random graphs with a given degree sequence; unfortunately, this is wide open at the moment. In fact, the simpler question of enumeration of all graphs with a given degree sequence is not very well understood at present. We believe these questions furnish a fertile ground for future research.

2 Properties of the rate function

Recall the definition of \( J_{W_0} \) from (1.12). In this section, we will prove some elementary facts about \( J_{W_0} \) that will be crucial in our proofs. Throughout, we denote \( \overline{\mathcal{B}}(W, \eta) = \{ W' : \delta_\square(W, W') \leq \eta \} \) and \( \mathcal{B}(W, \eta) = \{ W' : \overline{W'} \in \overline{\mathcal{B}}(W, \eta) \} \).

**Lemma 2.1.** The function \( J_{W_0}(W) = \sup_{\delta > 0} \inf_{W' \in \overline{\mathcal{B}}(W, \delta)} I_{W_0}(W') \) is well-defined on the space \( \mathcal{W} \). Moreover, \( J_{W_0} \) is lower semi-continuous on \( (\mathcal{W}, \delta) \) for any \( W_0 \) that is away from boundary.

**Proof.** For any \( \phi \in \mathcal{M} \), it follows that \( \{ W' : \delta_\square(W, W') \leq \delta \} = \{ W' : \delta_\square(W^0, W') \leq \delta \} \), and therefore \( J_{W_0} \) is well-defined on \( \mathcal{W} \). Define the function \( H : \mathcal{W} \to \mathbb{R}_+ \) by \( H(W') = \inf_{\phi \in \mathcal{M}} I_{W_0}(W') \). Note that, since \( W_0 \) is away from boundary and \( |x \log x| \leq 1/e \) for all \( x \in [0,1] \), \( H \) is a bounded function. Now,

\[
J_{W_0}(\tilde{W}) = \sup_{\delta > 0} \inf_{W' \in \overline{\mathcal{B}}(W, \delta)} I_{W_0}(W') \geq \inf_{W' \in \mathcal{B}(W, \delta)} H(W') = \liminf_{W' \to \tilde{W}} H(W'),
\]

and it is a standard fact in analysis that the function obtained by taking pointwise \( \liminf \) of a bounded function must be lower semi-continuous. This completes the proof. \( \square \)

**Lemma 2.2.** \( J_{W_0}(W) = 0 \) if and only if \( \delta_\square(W, W_0) = 0 \).

**Proof.** The sufficiency part is obvious. To see the necessity, assume the contrapositive, i.e., there exists \( \tilde{W} \) such that \( \delta_\square(W, W_0) > 0 \) but \( J_{W_0}(\tilde{W}) = 0 \). Thus, there exists \( (\phi_n)_{n \geq 1} \subset \mathcal{M} \) such that \( I_{W_0}(W^\phi_n) \to 0 \). Now, observe that integrand in the first expression of (1.11) is the entropy of a Bernoulli \((W(x, y))\) with respect to Bernoulli \((W_0(x, y))\). By Pinsker’s inequality \( \|W^\phi_n - W_0\|_{\ell_2} \to 0 \). Now, using Cauchy-Schwarz inequality, \( \|W^\phi_n - W_0\|_{\ell_1} \to 0 \), and consequently \( \delta_\square(W^\phi_n, W_0) \to 0 \). But \( \delta_\square(W^\phi_n, W) = 0 \) for all \( n \). This yields a contradiction with \( \delta_\square(W, W_0) > 0 \). \( \square \)

**Lemma 2.3.** Suppose that \( W_0 \) is away from boundary. Then \( \sup_{\delta > 0} \inf_{W' \in \mathcal{B}(W, \delta)} I_{W_0}(W') = \inf_{\phi \in \mathcal{M}} I_{W_0}(W^\phi) \) for any \( W \in \mathcal{W} \).

**Proof.** First observe that, whenever \( d_\square(W_n, W) \to 0 \), we have

\[
\liminf_{n \to \infty} I_{W_0}(W_n) \geq I_{W_0}(W).
\]

This fact follows using the exact same arguments of [8, Corollary 5.1]. Now let us simply denote \( I(\delta) = \inf_{W' \in \mathcal{B}(\tilde{W}, \delta)} I_{W_0}(W') \). Thus we need to show that \( I(0) = \sup_{\delta > 0} I(\delta) \), i.e., for all \( \varepsilon > 0, \exists \delta(\varepsilon) > 0 \) such that \( I(\delta) > I(0) - \varepsilon \) for all \( \delta \in (0, \delta(\varepsilon)) \). Suppose that this does not hold. Thus, there exists \( \varepsilon > 0 \) and \( \delta_n \to 0 \) such that \( I(\delta_n) \leq I(0) - \varepsilon \) for all \( n \geq 1 \). Thus there exists \( (W_n)_{n \geq 1} \) such that \( d_\square(W_n, W) \to 0 \) but \( I_{W_0}(W_n) < I(0) - \varepsilon/2 \) for all \( n \geq 1 \). Now, using (2.2), we have that \( I_{W_0}(W) < I(0) - \varepsilon/2 \) which yields a contradiction because \( I(0) \leq I_{W_0}(W) \). \( \square \)
Next we will prove that if we have a sequence $W^n_0$ converging to $W_0$ in $L_1$, then the corresponding relative entropies converge as well. This fact is stated in Lemma 2.4 below. To this end, we first prove the following general assertion, which will be required to prove Lemma 2.4.

**Fact 1.** Let $X$, $Y$ be metric spaces with $f : X \times Y \to \mathbb{R}$, $f(x, \cdot)$ is lower semi-continuous for all $x \in X$. Suppose that there is an element $x \in X$ and a compact set $K \subset Y$ such that whenever $x_n \to x$, then $f(x_n, y) \to f(x, y)$ uniformly over $y \in K$. Then $\inf_{y \in K} f(x_n, y) \to \inf_{y \in K} f(x, y)$, as $n \to \infty$.

**Proof.** Let $y^*(x) := \arg\min_{y \in K} f(x, y)$. Such a $y^*(x)$ always exists as $K$ is compact and $f(x, \cdot)$ is lower semi-continuous. Thus we need to show that $f(x_n, y^*(x_n)) \to f(x, y^*(x))$. Fix $\varepsilon > 0$, and choose $n_0$ such that, for all $n \geq n_0$, $|f(x_n, y^*(x_n)) - f(x, y^*(x))| < \varepsilon$, $\forall y \in K$. Note that, for all $n \geq n_0$,

$$
\begin{align*}
&f(x_n, y^*(x_n)) - \varepsilon \geq f(x_n, y^*(x_n)) - \varepsilon \geq f(x_n, y^*(x_n)) - \varepsilon \\
&f(x_n, y^*(x_n)) - \varepsilon \geq f(x_n, y^*(x_n)) - \varepsilon.
\end{align*}
$$

(2.3)

(2.4)

The proof thus follows.

**Lemma 2.4.** Suppose that $\|W^n_0 - W_0\|_{L_1} \to 0$, and that $(W^n_0)_{n \geq 1}$ is away from the boundary. Then, for any $\delta \geq 0$, $\inf_{W' \in \mathcal{B}(W, \delta)} I_{W_0}(W') \to \inf_{W' \in \mathcal{B}(W, \delta)} I_{W_0}(W')$, as $n \to \infty$. Consequently, $\inf_{W' \in \mathcal{W}} I_{W_0}(W') \to \inf_{W' \in \mathcal{W}} I_{W_0}(W)$ for any $W' \in \mathcal{W}$.

**Proof.** Let $H_{W_0}(W) = \inf_{W' \in \mathcal{B}(W, \delta)} I_{W_0}(W')$, so that

$$
\inf_{W' \in \mathcal{B}(W, \delta)} I_{W_0}(W') = \inf_{W' \in \mathcal{B}(W, \delta)} H_{W_0}(W').
$$

(2.5)

Using the compactness of $\mathcal{B}(W, \delta)$ and Fact 1, the proof follows if one can show that

$$
H_{W_0}(W') \to H_{W_0}(\tilde{W}') \quad \text{uniformly over } \mathcal{B}(W, \delta).
$$

(2.6)

For some $\eta > 0$, let $\eta < W^n_0 < 1 - \eta$ for all $n \geq 1$. Thus, using the the Lipchitz continuity of the log function, it follows that for all $x, y$,

$$
\max \left\{ \log \left( \frac{W^n_0(x, y)}{W_0(x, y)} \right), \log \left( \frac{1 - W^n_0(x, y)}{1 - W_0(x, y)} \right) \right\} \leq c|W^n_0(x, y) - W_0(x, y)|,
$$

(2.7)

for some constant $c > 0$. Now,

$$
\begin{align*}
|I_{W_0}(W) - I_{W_0}(W')| &= \left| \int_{[0,1]^2} W(x, y) \log \left( \frac{W^n_0(x, y)}{W_0(x, y)} \right) + (1 - W(x, y)) \log \left( \frac{1 - W^n_0(x, y)}{1 - W_0(x, y)} \right) \, dx \, dy \right| \\
&\leq 2c \int_{[0,1]^2} W(x, y)|W^n_0(x, y) - W_0(x, y)| \, dx \, dy \\
&\leq 2c\|W\|_{\infty}\|W^n_0 - W_0\|_{L_1} \leq 2c\|W^n_0 - W_0\|_{L_1},
\end{align*}
$$

(2.8)

where the final step uses $\|W\|_{\infty} \leq 1$ for any $W \in \mathcal{W}$. The proof of (2.6) now follows by noting the the bound in the final term of (2.8) is uniform over $W$.
3 An upper bound for inhomogeneous random graphs

In this section, we obtain a large deviation upper bound for inhomogeneous random graphs. Define a piecewise constant graphon to be a graphon $g \in \mathcal{W}$ of the form

$$g(x, y) = \begin{cases} \frac{r_{ij}}{n}, & x, y \in \left[\frac{i-1}{n}, \frac{i}{n}\right) \times \left[\frac{j-1}{n}, \frac{j}{n}\right), \quad i \neq j \\ 0, & \text{otherwise.} \end{cases} \quad (3.1)$$

Let us denote the collection of graphons in (3.1) by $\mathcal{W}^n_{pc}$. Given any graphon $W^n_0 \in \mathcal{W}^n_{pc}$, consider the random graph $G_n = G_n(W^n_0)$ on vertex set $[n]$ obtained by keeping an edge between vertices $i$ and $j$ with probability $q^n_{ij} = W^n_0((i-1)/n,(j-1)/n)$. Let $\mathbb{P}_{n,W^n_0}$ denote the probability measure on $\mathcal{W}$ induced by $G_n,W^n_0$, and let $\mathcal{P}_{n,W^n_0}$ denote the corresponding measure on $\mathcal{W}$. The following proposition derives the LDP upper bound for $\mathbb{P}_{n,W^n_0}$. Recall that $\mathcal{B}(W,\eta) = \{W' : \delta_\mathcal{B}(W,W') \leq \eta\}$ and $\mathcal{B}(W,\eta) = \{W' : W' \in \mathcal{B}(W,\eta)\}$.

Proposition 3.1. Fix $\varepsilon > 0$. Let $W^n_0 \in \mathcal{W}^n_{pc}$ be such that $\|W^n_0 - W_0\|_{L_1} \to 0$, and further assume that $(W^n_0)_{n \geq 1}$ is away from boundary. Then, there exists a universal constant $c > 0$ such that for all $\eta \in (0,\varepsilon)$,

$$\limsup_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}_{n,W^n_0}(\mathcal{B}(W,\eta)) \leq - \inf_{f \in \mathcal{B}(W,\varepsilon)} I_{W^n_0}(f). \quad (3.2)$$

3.1 Upper bound in the weak topology

A graphon can always be viewed as an element of $L_2([0,1]^2)$. Denote $B_{L_2}(r) = \{f \in \mathcal{W} : \|f\|_{L_2} \leq r\}$. Next, consider the weak topology on $L_2([0,1]^2)$, i.e., $f_n \to f$ weakly if for every $h \in L_2([0,1]^2)$,

$$\int_{[0,1]^2} f_n(x,y)h(x,y)dxdy \to \int_{[0,1]^2} f(x,y)h(x,y)dxdy. \quad \text{(3.3)}$$

The Banach-Alaoglu Theorem implies that the set $\{f \in L_2([0,1]^2) : \|f\|_{L_2} \leq r\}$, equipped with the weak topology, is compact. Since $\mathcal{W}$ is closed in the weak topology, it also follows that $B_{L_2}(1)$ is compact. We now derive the following upper bound on the probabilities of weakly closed subsets of $B_{L_2}(1)$:

Proposition 3.2. Let $\|W^n_0 - W_0\|_{L_1} \to 0$ and further assume that $(W^n_0)_{n \geq 1}$ is away from boundary. Let $F \subseteq B_{L_2}(1)$ be relatively closed with respect to the weak topology on $B_{L_2}(1)$. For any $\varepsilon > 0$, setting $F^\varepsilon = \{h : \|h-g\|_{L_2} \leq \varepsilon, \text{ for some } g \in F\}$, the following holds:

$$\limsup_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}_{n,W^n_0} \left( \bigcup_{\sigma \in \mathcal{M}_n} \{W^{G^n_\sigma} \in F\} \right) \leq - \inf_{W \in F^\varepsilon} \inf_{\phi \in \mathcal{A}} I_{W^n_0}(W^\phi). \quad (3.4)$$

Proof. For each $k \geq 1$, define the dyadic squares $D_k = \left\{ \left[\frac{i-1}{2^k}, \frac{i}{2^k}\right) \times \left[\frac{j-1}{2^k}, \frac{j}{2^k}\right), 1 \leq i,j \leq 2^k \right\}$, and let $\mathcal{D} = \bigcup_{k \geq 1} D_k$. Let $D_1, D_2, \ldots$ be the natural lexicographic enumeration of the dyadic squares $\mathcal{D}$. For $f,g \in B_{L_2}(1)$, define

$$\delta(f,g) = \sum_{m=1}^{\infty} 2^{-m} \min \left\{ 1, \left| \int_{D_m} (f(x,y) - g(x,y))dxdy \right| \right\}. \quad (3.5)$$

11
Using [8, Proposition 2.7], it follows that $\delta$ metrizes the weak topology restricted to $B_{L_2}(1)$.

Note that $F$ is relatively closed with respect to the weak topology, and hence compact. Let $\eta > 0$, sufficiently small, to be specified later, and for each $g \in F$, consider $B_\delta(g, \eta) = \{ f \in B_{L_2}(1) : \delta(g, f) < \eta \}$. The collection $\{ B_\delta(g, \eta) : g \in F \}$ constructs an open cover of $F$, and the compactness of $F$ implies the existence of a finite sub-cover $\cup_{i=1}^{N(\eta)} B_\delta(g_i, \eta)$. Therefore,

$$\mathbb{P}_{n,W_0^\delta}\left( \bigcup_{\sigma \in \#_n} \{ W_{\sigma}^{G_\eta} \in F \} \right) \leq \sum_{i=1}^{N(\eta)} \sum_{\sigma \in \#_n} \mathbb{P}_{n,W_0^\delta}(W_{\sigma}^{G_\eta} \in B_\delta(g_i, \eta)).$$

(3.5)

Now, for any set $\mathcal{D}_k$, let $i(k) = \max\{ i : D_i \in \mathcal{D}_k \}$, and let us choose $k(\eta)$ such that $k(\eta) \to \infty$ and $2^{i(k(\eta))}4^{k(\eta)}\eta \to 0$ as $\eta \to 0$. Finally, for $S \subset [0,1]^2$ a Borel subset of $[0,1]^2$ and $f \in L^2([0,1]^2)$ we set

$$\text{Ave}_S(f) = \frac{1}{\text{vol}(S)} \int_S f(x)dx$$

(3.6)

to be the average value of $f$ on the set $S$. For $S = D \in \mathcal{D}_k$, $\text{vol}(D) = 4^{-k}$. Now, using the definition of the metric $\delta$ in (3.4), it follows that, for any $D_m \in \mathcal{D}_k(\eta)$ and $\delta(f, g) < \eta$,

$$|\text{Ave}_{D_m}(f) - \text{Ave}_{D_m}(g)| \leq 2^{i(k(\eta))}4^{k(\eta)}2^{-\eta l} \int_{D_m} (f - g) \leq 2^{i(k(\eta))}4^{k(\eta)}\eta = E(\eta),$$

(3.7)

where $\lim_{\eta \to 0} E(\eta) = 0$, by the choice of $k(\eta)$. Thus,

$$\mathbb{P}_{n,W_0^\delta}(W_{\sigma}^{G_\eta} \in B_\delta(g_i, \eta)) \leq \mathbb{P}_{n,W_0^\delta}(\forall D_m \in \mathcal{D}_k(\eta), |\text{Ave}_{D_m}(f) - \text{Ave}_{D_m}(g_i)| \leq E(\eta)).$$

(3.8)

Let $\mathcal{D}_{k(\eta)}$ denote the set of dyadic squares of $\mathcal{D}_k(\eta)$ above the diagonal. The independence of the edge occupancy random variables now implies

$$\mathbb{P}_{n,W_0^\delta}(W_{\sigma}^{G_\eta} \in B_\delta(g_i, \eta)) \leq \prod_{D_m \in \mathcal{D}_{k(\eta)}} \mathbb{P}_{n,W_0^\delta}(|\text{Ave}_{D_m}(f) - \text{Ave}_{D_m}(g_i)| \leq E(\eta)).$$

(3.9)

We now bound the RHS of (3.9) using Chernoff bound that states, for any independent sequence $(X_i)_{i \in [n]}$, $\mathbb{P}(\sum_{i \in [n]} X_i \geq \alpha) \leq e^{-\alpha^2} \prod_{i=1}^{\alpha} \mathbb{E}[e^{\lambda X_i}]$. Let $I_{ij}$ denote the indicator that there is an edge between $i$ and $j$. Note that

$$\int_{D_m} W_{\sigma}^{G_\eta} = \frac{1}{n^2} \sum_{(i,j) : \sigma(i/n), \sigma(j/n) \in D_m} I_{ij} + o(1),$$

(3.10)

where the $o(1)$ term accounts for the discrepancy in the area enclosed by $D_m$ and that covered by the $1/n^2$ sized boxes corresponding to $\{(i,j) : \sigma(i/n), \sigma(j/n) \in D_m\}$. Let $a$ be a piecewise constant graphon which takes constant values on the dyadic squares of $\mathcal{D}_k(\eta)$. Also, for any function $g$, define

$$\hat{g}^k(x,y) = 4^k \int_D g(x',y')dx'dy' \quad \forall (x,y) \in D, \quad \forall D \in \mathcal{D}_k.$$

(3.11)
Further, define \( \hat{g}^k = (\hat{g} \lor \eta) \land (1 - \eta) \). Thus, by Chernoff bound,

\[
\begin{align*}
\mathbb{P}_{n,W_0} \left( \left| \text{Ave}_{D_m}(f) - \text{Ave}_{D_m}(g_i) \right| \leq E(\eta) \right) \\
\leq \min \left\{ \mathbb{P}_{n,W_0} \left( \text{Ave}_{D_m}(f) \geq \text{Ave}_{D_m}(g_i) - E(\eta) \right), \\
\mathbb{P}_{n,W_0} \left( \text{Ave}_{D_m}(f) \leq \text{Ave}_{D_m}(g_i) + E(\eta) \right) \right\} \\
\leq \exp \left( -\frac{n^2}{2} a_m \int_{D_m} g_i + \frac{1}{n^2} \sum_{(i,j): (\sigma(i/n), \sigma(j/n)) \in D_m} \log (1 - W_0^n(i/n, j/n) + W_0^n(i/n, j/n) e^{a_m}) - E'(\eta) \right),
\end{align*}
\]

where also \( \lim_{\eta \to 0} |E'(\eta)| = 0 \) by the choice of \( \eta \). The final inequality follows by using Chernoff bound on the upper tail or the lower tail depending on \( \mathbb{P}_{n,W_0} \left[ \text{Ave}_{D_m}(f) \right] < \text{Ave}_{D_m}(g_i) \) or not. Let us now fix the choice of \( a_m \). Define,

\[
a(x, y) = \log \frac{\hat{g}_i^{k(\eta)}(x, y)}{1 - \hat{g}_i^{k(\eta)}(x, y)} - \log \frac{W_0^n,\sigma(x, y)}{1 - W_0^n,\sigma(x, y)}, \quad a_m = \text{Ave}_{D_m}(a).
\]

Note that \( a_m \int_{D_m} g_i = \int_{D_m} \hat{g}_i^{k(\eta)} + E_1(\eta) \), and

\[
\begin{align*}
\left| \frac{1}{n^2} \sum_{(i,j): (\sigma(i/n), \sigma(j/n)) \in D_m} \log (1 - W_0^n(i/n, j/n) + W_0^n(i/n, j/n) e^{a_m}) \\
- \int_{D_m} \log (1 - W_0^n,\sigma + W_0^n,\sigma e^\alpha) \right| = E_2(\eta),
\end{align*}
\]

where \( \lim_{\eta \to 0} E_1(\eta) = \lim_{\eta \to 0} E_2(\eta) = 0 \). (3.14) follows using the fact that \( (W_0^n)_{n \geq 1} \) is away from the boundary and \( \log(1 - p + pe^\alpha) \) is Lipschitz continuous in \( x \). Therefore, (3.9) reduces to

\[
\exp \left( -(1 + o(1)) \frac{n^2}{2} \left[ \int a\hat{g}_i^{k(\eta)} + \int \log (1 - W_0^n,\sigma + W_0^n,\sigma e^\alpha) - E''(\eta) \right] \right)
\]

\[
= \exp \left( -n^2(1 + o(1)) \left[ I_{W_0^n}(\hat{g}_i^{k(\eta)}) - E''(\eta) \right] \right)
\]

\[
\leq \exp \left( -n^2(1 + o(1)) \left[ \inf_{\phi \in \mathcal{A}} I_{W_0^n}((\hat{g}_i^{k(\eta)}),\phi) - E''(\eta) \right] \right),
\]

where \( \lim_{\eta \to 0} E''(\eta) = 0 \). Since \( \| W_0^n - W_0 \|_{L_1} \to 0 \), an application of Lemma 2.4 yields that \( \inf_{\phi \in \mathcal{A}} I_{W_0^n}(\hat{g}_i^{k(\eta)}),\phi) \to \inf_{\phi \in \mathcal{A}} I_{W_0}((\hat{g}_i^{k(\eta)}),\phi) \). Now, plugging the bound in (3.15) back into (3.5), it follows that

\[
\begin{align*}
\mathbb{P}_{n,W_0} \left( \bigcup_{\sigma \in \mathcal{A}_n} \{ W_{G^n} \in F \} \right) \\
\leq n! N(\eta) \exp \left( -n^2(1 + o(1)) \left[ \inf_{g \in F} \inf_{\phi \in \mathcal{A}} I_{W_0^n}(g,\phi) - E''(\eta) \right] \right).
\end{align*}
\]

Using the fact that \( n! \ll e^{n^2} \), this immediately implies that

\[
\limsup_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}_{n,W_0} \left( \bigcup_{\sigma \in \mathcal{A}_n} \{ W_{G^n} \in F \} \right) \leq - \inf_{g \in F} \inf_{\tau \in \mathcal{A}} I_{W_0^n}(g^{k(\eta)}) + E''(\eta).
\]
The proof is complete by taking $\eta \to 0$, and noting that $\hat{g}^{k(\eta)} \to g$ in $L_2$ as $\eta \to 0$ for all $g \in F$.

\[ \square \]

### 3.2 Proof of Proposition 3.1

Recall the setup of Proposition 3.1. First, note that

\[ \tilde{\mathbb{P}}_{n,W_0^n}(\tilde{\mathbb{B}}(\tilde{W}, \eta)) = \mathbb{P}_{n,W_0^n}(\mathbb{B}(W, \eta)). \]  

(3.18)

Next, we recall a version of Szemerédi’s regularity lemma from [8, Theorem 3.1] that will be crucial here. Fix $\varepsilon > 0$. Then there exists $C(\varepsilon) > 0$ and a set $\mathcal{W}(\varepsilon) \subset \mathcal{W}$ with $|\mathcal{W}(\varepsilon)| \leq C(\varepsilon)$ such that the following holds:

For any $f \in \mathcal{W}$, there exists $\phi \in \mathcal{M}$ and $h \in \mathcal{W}(\varepsilon)$ satisfying $d_{\mathcal{B}}(f^\phi, h) < \varepsilon$.

For empirical graphons corresponding to graphs, the above can be restated as below: Let $\mathcal{M}_n$ denote the set of all permutations of $[n]$, and $G_n^\sigma$ denote the graph by relabelling the vertex $i$ by $\sigma(i)$, for some $\sigma \in \mathcal{M}_n$. Also let us denote $B_\mathcal{B}(W, \eta) = \{W' : d_\mathcal{B}(W, W') \leq \eta\}$.

Then, for any graph $G_n$ on vertex set $[n]$, there exists $\sigma \in \mathcal{M}_n$ and $h \in \mathcal{W}(\varepsilon)$ such that

\[ W^{G_n} \in B_\mathcal{B}(h, \varepsilon). \]  

(3.19)

Let $G_n$ be the random graph sampled from the probability distribution $\mathbb{P}_{n,W_0^n}$. Since $\{W^{G_n} \in \mathbb{B}_\mathcal{B}(\tilde{W}, \eta)\} = \{W^{G_n} \in \mathbb{B}_\mathcal{B}(W, \eta)\}$ for any $\sigma \in \mathcal{M}_n$, the above version of regularity lemma implies that

\[ \{W^{G_n} \in \mathbb{B}_\mathcal{B}(\tilde{W}, \eta)\} \subseteq \{W^{G_n} \in \mathbb{B}_\mathcal{B}(W, \eta)\} \cap \left( \bigcup_{\sigma \in \mathcal{M}_n} \{W^{G_n} \in \mathbb{B}_\mathcal{B}(\mathcal{W}(\varepsilon), \varepsilon)\} \right) \]

\[ = \bigcup_{\sigma \in \mathcal{M}_n} \{W^{G_n} \in \mathbb{B}_\mathcal{B}(\tilde{W}, \eta) \cap B_\mathcal{B}(\mathcal{W}(\varepsilon), \varepsilon)\}. \]  

(3.20)

Now, $\mathcal{W}(\varepsilon)$ is a finite set and $n! = o(e^n)$. Therefore it is enough to show that

\[ \lim_{n \to \infty} \frac{1}{n^2} \log \mathbb{P} \left( \bigcup_{\sigma \in \mathcal{M}_n} \mathbb{B}_\mathcal{B}(\tilde{W}, \eta) \cap B_\mathcal{B}(g, \varepsilon) \right) \leq - \inf_{f \in B_\mathcal{B}(\tilde{W}, \varepsilon)} I_{\mathcal{W}_0}(f), \]  

(3.21)

where $g \in \mathcal{W}(\varepsilon)$ and $c > 0$ is a constant. Let $\eta < \varepsilon$, without loss of generality, let us assume that $\mathbb{B}_\mathcal{B}(\tilde{W}, \eta) \cap B_\mathcal{B}(g, \varepsilon) \neq \emptyset$. In this case, $\delta_\mathcal{B}(W, g) \leq 2\varepsilon$. Now, recall the statement of Proposition 3.2. Using [8, Lemma 5.4], it follows that $B_\mathcal{B}(g, \varepsilon)$ is relatively closed in $B_{l_2}(1)$ with respect to the weak topology, and thus Proposition 3.2 is applicable. It now follows that

\[ \lim_{n \to \infty} \frac{1}{n^2} \log \mathbb{P} \left( \bigcup_{\sigma \in \mathcal{M}_n} \{W^{G_n} \in B_\mathcal{B}(g, \varepsilon)\} \right) \leq - \inf_{f \in B_\mathcal{B}(g, 2\varepsilon)} \inf_{\phi \in \mathcal{M}} I_{\mathcal{W}_0}(f^\phi) \]

\[ \leq - \inf_{f \in B_\mathcal{B}(g, c\varepsilon)} I_{\mathcal{W}_0}(f) \leq - \inf_{f \in B_\mathcal{B}(g, (c+2)\varepsilon)} I_{\mathcal{W}_0}(f), \]  

(3.22)

where the second step follows using the fact that the $\varepsilon$ fattened set of $B_\mathcal{B}(g, \varepsilon)$ is contained in $B_\mathcal{B}(g, c\varepsilon)$, for some universal constant $c > 1$. The final step follows using the triangle inequality for the metric $\delta_\mathcal{B}$. This completes the proof.

\[ \square \]
4 Large deviation for uniform graphs with given degree

In this section, we complete the proof of Theorem 1.2. Using the fact that $(\tilde{\mathcal{W}}, \delta)$ is a compact metric space, it is sufficient (see [12, page 1004]) to show that for any $\tilde{W} \in \mathcal{W}$,

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{1}{n^2} \log \tilde{P}_n, d(\tilde{\mathcal{B}}_\mathcal{C}(\tilde{W}, \eta)) \leq -J_D(\tilde{W}),$$

(4.1)

and for any $\eta > 0$

$$\lim_{n \to \infty} \frac{1}{n^2} \log \tilde{P}_n, d(\tilde{\mathcal{B}}_\mathcal{C}(\tilde{W}, \eta)) \geq -J_D(\tilde{W}).$$

(4.2)

4.1 Key facts from Chatterjee, Diaconis, Sly [11]

Let us first recall a few key ingredients from [11], which were used to obtain the graphon limit of $G_{n, d}$. Let $\hat{\beta} = (\hat{\beta}_i)_{i \in [n]}$ be the solution to the system of equations

$$d_i = \sum_{j \neq i} e^{\hat{\beta}_i + \hat{\beta}_j} \frac{e^{\hat{\beta}_i + \hat{\beta}_j}}{1 + e^{\hat{\beta}_i + \hat{\beta}_j}}, \quad \forall i \in [n].$$

(4.3)

Due to [11, Lemma 4.1], $\hat{\beta}$ exists and $\|\hat{\beta}\|_\infty \leq C$ for some constant $C > 0$ for all sufficiently large $n$ under Assumption 1. It is not obvious that Assumption 1 yields the conditions in [11, Lemma 4.1], but that too was shown in the first part of the proof of [11, Theorem 1.1] in Section 6.2. Next, for any $i \neq j$, define

$$\hat{p}_{ij} = \frac{e^{\hat{\beta}_i + \hat{\beta}_j}}{1 + e^{\hat{\beta}_i + \hat{\beta}_j}},$$

(4.4)

and let $\hat{G}_n$ be the random graph on vertex set $[n]$ obtained by keeping an edge between vertices $i$ and $j$ with probability $\hat{p}_{ij}$, independently. Define

$$W_{n, d}(x, y) = \begin{cases} \hat{p}_{ij}, & x, y \in \left[\frac{i-1}{n}, \frac{i}{n}\right] \times \left[\frac{j-1}{n}, \frac{j}{n}\right], \quad i \neq j, \\ 0, & \text{otherwise.} \end{cases}$$

(4.5)

Since $\|\hat{\beta}\|_\infty \leq C$, it follows that $(W_{n, d})_{n \geq 1}$ is away from the boundary. Therefore, the results from Section 3 are applicable to $(W_{n, d})_{n \geq 1}$. Next, let $D_n : [0, 1] \to [0, 1]$ be the step function given by

$$D_n(x) = \frac{1}{n} \sum_{j \neq i} \hat{p}_{ij} = \frac{d_i}{n}, \quad \forall x \in \left[\frac{i-1}{n}, \frac{i}{n}\right], \quad \forall i \in [n],$$

(4.6)

and the degree distribution function is given by

$$\mu_{D_n}([0, \lambda]) = \Lambda\{x : D_n(x) \leq \lambda\}.$$

(4.7)

By Assumption 1, $\|D_n - D\|_{L_1} \to 0$, and thus

$$\mu_{D_n} \overset{w}{\to} \mu_D,$$

(4.8)
where $\mu_D$ is defined in (1.13), where $\Rightarrow$ denotes the weak-convergence of measures. Now, let $\beta_n(x) = \sum_{i=1}^n \beta_i \mathbb{1}\{x \in \left[\frac{i-1}{n}, \frac{i}{n}\right]\}$, and in [11, page 1430–1432], it was also shown that

$$\|\beta_n - \beta\|_{L_1} \to 0, \quad \|W_{n,d} - W_D\|_{L_1} \to 0,$$

(4.9)

where $\beta$ and $W_D$ are defined in Proposition 1.1.

Next, recall that $\mathcal{W}_0 = \{ W \in \mathcal{W} : \deg_{\tilde{W}} = \mu_D \}$ and define $\mathcal{W}_0'' = \{ W \in \mathcal{W} : \deg_{\tilde{W}} = \mu_{D_n} \}$. Recall the definition of the probability measure $P_{n,W_0'}$ from Section 3. Note that

$$P_{n,W_{n,d}}(\mathcal{W}_0'') = P_{n,d}(\cdot).$$

Next, we quote a key lemma from [11] which will be used in the proof: Let $(r_{ij})_{i \neq j}$ satisfy $r_{ij} = r_{ji}, r_{ii} = 0$ and $\sum_{i=1}^n r_{ij} = d_n$ and construct a random graph $G_n$ on vertex set $[n]$ by keeping an edge between $i$ and $j$ with probability $r_{ij}$.

**Lemma 4.1 ([11, Lemma 6.2]).** For all sufficiently large $n$, $G_n$ has degree sequence exactly $d$ with probability at least $e^{-n^{7/4}}$.

A direct corollary of Lemma 4.1 is the following:

$$P_{n,W_{n,d}}(\mathcal{W}_0'') \geq e^{-n^{7/4}},$$

(4.11)

for all sufficiently large $n$. We are now ready to prove our LDP result.

### 4.2 Proof of the upper bound (4.1)

Define the Lévy-Prokhorov distance between two distribution functions $F_1, F_2$ supported on $[0, 1]$ by

$$d_{LP}(F_1, F_2) = \inf \{ \varepsilon > 0 : F_2(\lambda - \varepsilon) - \varepsilon \leq F_1(\lambda) \leq F_2(\lambda + \varepsilon) + \varepsilon, \forall \lambda \in [0, 1] \}.$$

(4.12)

$d_{LP}$ metrizes weak convergence of probability measures. Using [6, Theorem 2.16], it follows that

$$d_{LP}(\deg_{\tilde{W}_1}, \deg_{\tilde{W}_2}) \leq (2d_{LP}(\tilde{W}_1, \tilde{W}_2))^{1/2}.$$

(4.13)

To prove (4.1), we will be assuming that $\tilde{W} \in \mathcal{W}_0$. If that is not the case, then the probability in (4.1) is $-\infty$ for all sufficiently large $n$ and small $\eta$. To see this, suppose $\tilde{W} \in \mathcal{W}$ is such that $d_{LP}(\deg_{\tilde{W}}, \mu_D) = c > 0$. Since $d_{LP}(\mu_{D_n}, \mu_D) \to 0$ by (4.8), it follows that, for all sufficiently large $n$, $d_{LP}(\deg_{\tilde{W}}, \mu_{D_n}) \geq c/2$. Take $\eta_0 = c^2/32$. Now, for any $\tilde{U} \in \overline{\mathcal{B}_{\mathcal{W}}}(\tilde{W}, \eta_0)$

$$\frac{c}{2} \leq d_{LP}(\deg_{\tilde{W}}, \mu_{D_n}) \leq d_{LP}(\deg_{\tilde{W}}, \deg_\mathcal{U}) + d_{LP}(\deg_\mathcal{U}, \mu_{D_n}) \leq \frac{c}{4} + d_{LP}(\deg_\mathcal{U}, \mu_{D_n}),$$

(4.14)

for all sufficiently large $n$, where the final step follows from (4.13). Thus $d_{LP}(\deg_\mathcal{U}, \mu_{D_n}) \geq c/4$ for all $\tilde{U} \in \overline{\mathcal{B}_{\mathcal{W}}}(\tilde{W}, \eta_0)$, and thus $P_{n,d}(\overline{\mathcal{B}_{\mathcal{W}}}(\tilde{W}, \eta)) = 0$ using (4.13).

Therefore, we will assume that $\deg_{\tilde{W}} = \mu_D$. Let $N(\mu_D, \varepsilon) = \{ W \in \mathcal{W} : d_{LP}(\deg_{\tilde{W}}, \mu_D) \leq \varepsilon \}$. By (4.8), $\mathcal{W}_0'' \subset N(\mu_D, \varepsilon)$ for all sufficiently large $n$. Now,

$$P_{n,d}(\overline{\mathcal{B}_{\mathcal{W}}}(\tilde{W}, \eta)) = P_{n,d}(\overline{\mathcal{B}_{\mathcal{W}}}(\tilde{W}, \eta)) = \frac{P_{n,W_{n,d}}(\overline{\mathcal{B}_{\mathcal{W}}}(\tilde{W}, \eta) \cap \mathcal{W}_0'')}{P_{n,W_{n,d}}(\mathcal{W}_0'')} \leq e^{n^{7/4}} P_{n,W_{n,d}}(\overline{\mathcal{B}_{\mathcal{W}}}(\tilde{W}, \eta) \cap N(\mu_D, \varepsilon)),$$

(4.15)
4.3 Proof of the lower bound (4.2)

Fix \( \hat{W} \in \mathcal{W} \) such that \( \text{deg}_{\hat{W}} = \mu_{D} \), otherwise the proof will be trivial using identical arguments as Section 4.2. Further, without loss of generality we can assume that \( W \) is away from boundary. This is because, for any \( W \in \mathcal{W}_{0} \), we can define \( W_{\eta} = \eta W_{D} + (1 - \eta)W \). Note that \( W_{\eta} \) is away from the boundary, and \( \| W_{\eta} - W \|_{1} < \eta \| \| W_{D} \|_{L_{1}} + \| W \|_{L_{1}} \) \), so that we can always take \( \eta' \) such that \( \mathbb{B}_{\square}(W_{\eta}, \eta') \subset \mathbb{B}_{\square}(W, \eta) \). In that case, it will be enough to lower bound \( \tilde{P}_{n,d}(\mathbb{B}_{\square}(W, \eta')) \).

Recall the definition of \( \hat{G}_{n} \) from Section 4.1, and define the event

\[
\mathcal{E}_{n} = \{ \exists g \in \mathcal{W} \text{ with } \| \delta_{\square}(g, W) \leq \eta \text{ such that } \delta_{\square}(W^{\hat{G}_{n}}, g) < 2\eta \}.
\]

Note that, if \( \mathcal{E}_{n} \) happens, then \( \delta_{\square}(W^{\hat{G}_{n}}, g) < 2\eta \), and therefore, by triangle inequality, \( \delta_{\square}(W^{\hat{G}_{n}}, \hat{W}) < 3\eta \). Next, note that for any collection of events \( (A_{\alpha})_{\alpha \in A} \), \( \mathbb{P}(\bigcup_{\alpha \in A} A_{\alpha}) \geq \max_{\alpha \in A} \mathbb{P}(A_{\alpha}) \). Thus, we have

\[
\tilde{P}_{n,d}(\mathbb{B}_{\square}(\hat{W}, 3\eta)) \geq \tilde{P}_{n,d}(\mathcal{E}_{n}) \geq \sup_{\tilde{g} \in \mathbb{B}_{\square}(W, \eta)} \tilde{P}_{n,d}(\mathbb{B}_{\square}(\tilde{g}, 2\eta)).
\]

Since we are considering lower bounds, we can restrict ourselves to \( g \)'s that are away from boundary. The following lemma states that the graphons with any fixed degree distribution function can be approximated by piecewise constant graphons with approximately same degree function. We first state this fact and complete the proof of the lower bound. The proof of the lemma is given at the end of this section.

**Lemma 4.2.** Let \( W \) be such that \( f(x) = \int_{0}^{1} g(x, y)dy \), \( g \) is away from boundary. Further, let \( f_{n} \) be a step function of the form \( f_{n}(x) = \sum_{i \in [n]} c_{i}^{n} \mathbb{1}(x \in \left[ \frac{i-1}{n}, \frac{i}{n} \right]) \) for some \( c_{i}^{n} \in \mathbb{R}_{+} \) such that \( \| f_{n} - f \|_{L_{1}} \to 0 \). Then there exists graphons \( (g_{n})_{n \geq 1} \) such that \( f_{n} = \int_{0}^{1} g_{n}(x, y)dy \), \( \| g_{n} - g \|_{L_{1}} \to 0 \), and \( g_{n} \) is of the form

\[
g_{n}(x, y) = \begin{cases} r_{ij}, & x, y \in \left[ \frac{i-1}{n}, \frac{i}{n} \right) \times \left[ \frac{j-1}{n}, \frac{j}{n} \right), \quad i \neq j \\ 0, & \text{otherwise}. \end{cases}
\]

Next, since \( \| D_{n} - D \|_{L_{1}} \to 0 \), using Lemma 4.2, we can construct a function \( g_{n} \) with \( \| g_{n} - g \|_{L_{1}} \to 0 \) such that (4.18) holds, and \( \sum_{j \in [n]\setminus\{i\}} r_{ij} = d_{i} \) for all \( i \in [n] \). Note that \( g_{n} \) is a function of \( g \) and \( \| g_{n} - g \|_{L_{1}} \to 0 \) so that \( \mathbb{B}_{\square}(g_{n}, \eta) \subset \mathbb{B}_{\square}(g, 2\eta) \) for all sufficiently large \( n \). Also let \( G_{n} \) denote the graph on vertex set \( [n] \), where an edge between vertices \( i \) and \( j \) are kept with probability \( r_{ij} \), independently, and let \( \tilde{P}_{n,g} \) denote the distribution of
\( W^{G_n} \). Thus
\[
\mathbb{P}_{n,d}(\mathbb{E}(W, 3\eta)) \geq \sup_{g \in \mathbb{E}(W, \eta)} \int_{\mathbb{E}(\tilde{g}, 2\eta) \cap \mathcal{W}_0^n} d\mathbb{P}_{n,W,d}
\]
\[
\geq \sup_{g \in \mathbb{E}(W, \eta)} \int_{\mathbb{E}(\tilde{g}, \eta) \cap \mathcal{W}_0^n} d\mathbb{P}_{n,W,d} = \sup_{g \in \mathbb{E}(W, \eta)} \int_{\mathbb{E}(\tilde{g}, \eta) \cap \mathcal{W}_0^n} e^{-\frac{\log d\mathbb{P}_{n,g}}{d\mathbb{P}_{n,W,d}}} d\mathbb{P}_{n,g,n}
\]
\[
= \sup_{g \in \mathbb{E}(W, \eta)} \mathbb{P}_{n,g,n}(\mathbb{E}(\tilde{g}, \eta) \cap \mathcal{W}_0^n) \frac{1}{\mathbb{P}_{n,g,n}(\mathbb{E}(\tilde{g}, \eta) \cap \mathcal{W}_0^n)} \int_{\mathbb{E}(\tilde{g}, \eta) \cap \mathcal{W}_0^n} e^{-\frac{\log d\mathbb{P}_{n,g,n}}{d\mathbb{P}_{n,W,d}}} d\mathbb{P}_{n,g,n}.
\]

(4.19)

Now, by Jensen’s inequality, the logarithm of the term inside the sup above is at least
\[
\log(\mathbb{P}_{n,g,n}(\mathbb{E}(\tilde{g}, \eta) \cap \mathcal{W}_0^n)) - \frac{1}{\mathbb{P}_{n,g,n}(\mathbb{E}(\tilde{g}, \eta) \cap \mathcal{W}_0^n)} \int_{\mathbb{E}(\tilde{g}, \eta) \cap \mathcal{W}_0^n} \frac{d\mathbb{P}_{n,g,n}}{d\mathbb{P}_{n,W,d}} \frac{d\mathbb{P}_{n,W,d}}{d\mathbb{P}_{n,g,n}}
\]

(4.20)

Denote the two terms above by \((I)\) and \((II)\) respectively. To deal with the term \((I)\), we need the following lemma:

**Lemma 4.3.** For any \( \eta > 0 \), as \( n \to \infty \),
\[
\mathbb{P}_{n,g,n}(\mathbb{E}(\tilde{g}, \eta) \mid \mathcal{W}_0^n) \to 1.
\]

(4.21)

**Proof.** In this proof, we denote by \( G_n \) and \( G_n' \) the random graphs sampled according to probability measures \( \mathbb{P}_{n,g,n}(\cdot) \) and \( \mathbb{P}_{n,g,n}(\cdot \mid \mathcal{W}_0) \) respectively. Since \( \|g_n - g\|_{L_1} \to 0 \), it is enough to show that, \( t(F, W^{G_n}) \to t(F, g) \) with high probability with respect to the measure \( (\mathbb{P}_{n,g,n}(\cdot \mid \mathcal{W}_0))_{n \geq 1} \) for any fixed finite simple graph \( F \). Firstly, it is obvious that \( \mathbb{E}[t(F, G_n)] = t(F, g_n) \to t(F, g) \), and a standard argument (see [11, Lemma 6.1]) using Azuma inequality yields for any \( \varepsilon > 0 \)
\[
\mathbb{P}_{n,g,n}(|t(F, G_n) - t(F, g)| > \varepsilon) \leq 2e^{-C\varepsilon^2 n^2},
\]

(4.22)

for some constant \( C > 0 \). Now, recall that \( \sum_{j \neq i} g_{ij} = d_i \) by construction. Thus, using Lemma 4.1, it follows that
\[
\mathbb{P}_{n,U}(|t(F, G_n) - t(F, g)| > \varepsilon \mid \mathcal{W}_0^n) = \frac{\mathbb{P}_{n,g,n}(|t(F, G_n) - t(F, g)| > \varepsilon \mid \mathcal{W}_0^n)}{\mathbb{P}_{n,g,n}(\mathcal{W}_0^n)} \leq \frac{\mathbb{P}_{n,U}(|t(F, G_n) - t(F, g)| > \varepsilon)}{\mathbb{P}_{n,g,n}(\mathcal{W}_0^n)} \leq e^{-Cn^2},
\]

(4.23)

for some constant \( C > 0 \). This completes the proof.

\( \square \)

**Completing the proof of the lower bound.** Note that, by Lemmas 4.1 and 4.3, the term \((I)\) in (4.20) simplifies to
\[
(I) \geq Cn^{7/4} = o(n^2),
\]

(4.24)

for some constant \( C > 0 \). To analyze term \((II)\), firstly note that
\[
\log \frac{d\mathbb{P}_{n,g,n}}{d\mathbb{P}_{n,W,d}} = \sum_{1 \leq i < j \leq n} \left( I_{ij} \log \left( \frac{g_{ij}}{\hat{p}_{ij}} \right) + (1 - I_{ij}) \log \left( \frac{1 - g_{ij}}{1 - \hat{p}_{ij}} \right) \right),
\]

(4.25)
where \( I_{ij} \sim \text{Ber}(g_{ij}) \) independently. By changing one \( I_{ij} \), this quantity can change by at most
\[
\max_{i,j} \left| \log \left( \frac{g_{ij}}{\hat{p}_{ij}} \right) + \log \left( \frac{1 - g_{ij}}{1 - \hat{p}_{ij}} \right) \right| \leq C, \tag{4.26}
\]
using the assumption in the beginning of the section that \( g \) (and therefore \((g_n)_{n\geq 1}\)) is away from the boundary, and by definition \( W_D \) (and therefore \((W_n,d)_{n\geq 1}\)) is also away from the boundary by definition. Therefore, an application of Azuma inequality yields that, for any \( \varepsilon > 0 \),
\[
\mathbb{P}_{n,g_n} \left( \log \frac{d\mathbb{P}_{n,g_n}}{d\mathbb{P}_{n,W,n,d}} - \mathbb{E}_{n,g_n} \left[ \log \frac{d\mathbb{P}_{n,g_n}}{d\mathbb{P}_{n,W,n,d}} \right] \right) > \varepsilon n^2 \right) \leq e^{-C\varepsilon^2 n^2}, \tag{4.27}
\]
for some constant \( C > 0 \) which depends on the constant in (4.26). Take \( \varepsilon_n = n^{-1/10} \). Also note that, by (4.26), the log derivative \( \log \frac{d\mathbb{P}_{n,g_n}}{d\mathbb{P}_{n,W,n,d}} \) is at most \( Cn^2 \). Therefore,
\[
(II) \leq \frac{1}{\mathbb{P}_{n,g_n}(E_0(\hat{g}_n, \eta) \cap W_0^n)} \int_{W_0^n} \log \frac{d\mathbb{P}_{n,g_n}}{d\mathbb{P}_{n,Y,n,d}} d\mathbb{P}_{n,g_n} \leq \frac{1}{\mathbb{P}_{n,g_n}(E_0(\hat{g}_n, \eta) \cap W_0^n)} \left( \mathbb{E}_{n,g_n} \left[ \log \frac{d\mathbb{P}_{n,g_n}}{d\mathbb{P}_{n,W,n,d}} \right] + n^{3/2} \right) \mathbb{P}_{n,g_n}(W_0^n) + Cn^2e^7/4 + Cn^9/5 \leq \frac{1}{\mathbb{P}_{n,g_n}(E_0(\hat{g}_n, \eta) \cap W_0^n)} \left( \mathbb{E}_{n,g_n} \left[ \log \frac{d\mathbb{P}_{n,g_n}}{d\mathbb{P}_{n,W,n,d}} \right] + o(n^2) \right) + o(n^2). \tag{4.28}
\]

Further, it follows directly from (4.25) that
\[
\lim_{n \to \infty} \frac{1}{n^2} \mathbb{E}_{n,g_n} \left[ \log \frac{d\mathbb{P}_{n,g_n}}{d\mathbb{P}_{n,W,n,d}} \right] = I_{W_D}(g). \tag{4.29}
\]
Thus, using Lemma 4.3, (4.28) and (4.29), (4.19) yields that
\[
\liminf_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}_{n,d}(B_{\varepsilon}(W, 3\eta)) \geq -\inf_{g \in B_{\varepsilon}(W, \eta)} I_{W_D}(g), \tag{4.30}
\]
this concludes the proof of the lower bound in (4.2).

**Proof of Lemma 4.2.** First, define
\[
g_{n_1}(x, y) = n^2 \int_{\left[ \frac{i-1}{n}, \frac{i}{n} \right] \times \left[ \frac{j-1}{n}, \frac{j}{n} \right]} g(x, y) \, dx \, dy, \quad \forall (x, y) \in \left[ \frac{i-1}{n}, \frac{i}{n} \right] \times \left[ \frac{j-1}{n}, \frac{j}{n} \right]. \tag{4.31}
\]

Note that \( \|g_{n_1} - g\|_{L_1} \to 0 \) and if \( f_{n_1}(x) = \int_0^1 g_{n_1}(x, y) \, dy \), then \( \|f_{n_1} - f\|_{L_1} \to 0 \). Next let \( \varepsilon_n(x) = f_{n_1}(x) - f_{n_1}(x) \) and define
\[
g_{n_2}(x, y) = g_{n_1}(x, y) + \frac{\varepsilon_n(x)\varepsilon_n(y)}{\int_0^1 \varepsilon_n(y) \, dy}. \tag{4.32}
\]

Clearly, \( \int_0^1 g_{n_2}(x, y) \, dy = f_{n_1}(x) \). Moreover, since \( \|\varepsilon_n\|_{L_1} \to 0 \), it follows that \( \|g_{n_2} - g\|_{L_1} \to 0 \). In order to complete the proof, we need to make sure that our function takes zero value on the diagonal blocks. For that, we need the following:
**Fact 2.** Given any sequence \( a = (a_i)_{i \in [n]} \), it is possible to find weights \( w = (w_{ij})_{i,j} \) with \( w_{ij} = w_{ji} \) for all \( i, j \), such that \( \sum_{j \in [n]\setminus \{i\}} w_{ij} = a_i \) for all \( i \in [n] \) and \( \|w\|_\infty \leq 2\|a\|_\infty / (n-2) \).

Let us first complete the proof of Lemma 4.2. Recall that \( f_n(x) = \sum_{i \in [n]} c_i^i \mathbb{1}\{x \in [i/n, 1/n]\} \), and let \( b_i^j \) be the values on the diagonal blocks of \( g_{n2} \). Define \( a_i = c_i - b_i \).

Now, we choose \( w \) according to Fact 2. Let \( w_{ij} = w_{ji} \), and define
\[
g_n(x, y) = \begin{cases} g_{n2}(x, y) + \frac{w_{ij} + w_{ji}}{2} & \forall (x, y) \in \left[\frac{i-1}{n}, \frac{i}{n}\right) \times \left[\frac{j}{n}, \frac{j+1}{n}\right), i \neq j, \\ 0 & \text{otherwise.} \end{cases}
\] (4.33)

By construction, \( g_n \) now satisfies all the conditions stated in Lemma 4.2 which completes the proof.

**Proof of Fact 2.** Let us view \( w \) as a vector with its elements indexed by \( (j, k), j < k \). We wish to find a solution of \( w \) in the equation \( Mw = a \), where \( M \) is an \( n \times \binom{n}{2} \) matrix with entries \( m_{i,j,k} = \mathbb{1}\{i \in \{j, k\}\} \). First let us find the inverse of \( MM^T \). Indeed,
\[
(MM^T)_{uv} = \sum_{j<k} \mathbb{1}\{u \in \{j, k\}\} \mathbb{1}\{v \in \{j, k\}\} = \begin{cases} 1 & \text{if } u \neq v, \\ n-1 & \text{if } u = v. \end{cases}
\] (4.34)

Thus \( MM^T = (n-2)I + 11^T \). An application of Sherman–Morrison formula yields that
\[
(MM^T)^{-1} = \frac{I}{n-2} - \frac{11^T}{2(n-1)(n-2)}.
\] (4.35)

Now, \( w = M^T(MM^T)^{-1}a \) is a solution to the equation \( Mw = a \). Also, the \( (j, k) \)-th column of \( M \) consists of 1 on the \( j \)-th and \( k \)-th entries and zero elsewhere. Thus, \( \|w\|_\infty \leq 2\|a\|_\infty / (n-2) \), and the proof follows.

## 5 Proofs of Corollaries 1.4 and 1.5

### 5.1 Large deviation for continuous functionals

In this section, we prove Corollary 1.4, leveraging the general techniques used in [12, Section 3] and [15, Section 3.2].

**Proof of Corollary 1.4 (1).** Let \( \Gamma_{\geq r} = \{ \bar{W} : \tau(\bar{W}) \geq r \} \). This is a closed set, since \( \tau \) is continuous. Also,
\[
\phi_\tau(D, r) = \inf_{\bar{W} \in \Gamma_{\geq r}} J_D(\bar{W}).
\] (5.1)

First, note that \( J_D(\bar{W}) = 0 \) if and only if \( \delta_{\square}(\bar{W}, W_D) = 0 \), which follows directly from Lemma 2.2. Thus, \( \phi_\tau(D, r) = 0 \) for \( r \in [0, \ell_\tau(D)] \). In this proof, let us henceforth assume \( r \in (\ell_\tau(D), r_\tau(D)) \). It follows that \( \Gamma_{\geq r} \cap W_0 \neq \emptyset \) and \( J_{W_D} \) is finite on \( \Gamma_{\geq r} \cap W_0 \). Consequently, \( \phi_\tau(D, r) < \infty \). For the strict positivity, since \( \Gamma_{\geq r} \cap W_0 \) is compact and \( J_D(\bar{W}) \) is lower semi-continuous, the infimum in (5.1) is attained at some point \( \bar{W}^* \). However, since \( \tau(\bar{W}^*) \geq r > \ell_\tau(D) \), it must be that \( \delta_{\square}(\bar{W}_D, \bar{W}^*) > 0 \) and thus \( J_D(\bar{W}^*) > 0 \). This shows \( \phi_\tau(D, r) \) is strictly positive.
To see that \( \phi_r(D, \cdot) \) is strictly increasing, let \( F_* \subset \Gamma_r \cap \mathcal{W}_0 \) be the set of minimizers of (5.1), which is shown to be non-empty above. Given \( W_r \in F_* \), fix \( W_r \) in equivalence class \( W_r \), and consider \( W_r(\lambda) \coloneqq \lambda W_r + (1 - \lambda) W_D \). Note that \( W_r \in \mathcal{W}_0 \). Also, \( \tau(W_r(\cdot)) \) is continuous in \([0, 1]\), using the continuity of \( \tau \) on \((\mathcal{W}, \delta_{\mathcal{W}})\). Thus, for any \( r' \in [l_r(D), \tau(W_r)) \), there exists \( \lambda' = \lambda'(r') \in [0, 1) \) such that \( \tau(W_r(\lambda')) = r' \). Using the convexity of \( J_r \),

\[
\phi_r(D, r') \leq J_D(W_r(\lambda')) \leq \lambda' J_D(W_r) < \phi_r(D, r)
\]

(5.2)

Since \( \tau(W_r) \geq r \), it follows that

\[
\phi_r(D, r') < \phi_r(D, r) \quad \forall r' \in [l_r(D), r),
\]

(5.3)

proving that \( \phi_r(D, \cdot) \) strictly increases. To prove left-continuity, let \( \alpha < \infty \) be such that \( \phi_r(D, r') \leq \alpha \) for all \( r' \leq r \). Note that \( J_D(W_r') \leq \alpha, \tau(W_r') \geq r' \), and further, \( \{W_r' : r' < r\} \) is precompact. Take a subsequence along which as \( r' \searrow r \), \( W_r' \to W_r \) in \((\mathcal{W}, \delta_{\mathcal{W}})\). Then, by the lower semi-continuity of \( J_D, J_D(W_r') \leq \alpha \), and by the continuity of \( \tau, \tau(W_r) \geq r \). Thus \( \phi_r(D, r) \leq \alpha \). This proves the left-continuity of \( \phi_r(D, \cdot) \).

**Proof of Corollary 1.4 (2).** For \( r \in [0, l_r(D)] \), since \( J_D \) is zero, the uniqueness of the solution in (1.16) follows from Lemma 2.2. Recall the notation \( W_r \) from the proof of Corollary 1.4 (1) and the fact from (5.2) that \( \phi_r(D, r') < \phi_r(D, r) \) for all \( r' \in [l_r(D), \tau(W_r)) \). Thus, if \( \tau(W_r) > r \), the \( \phi_r(D, r) < \phi_r(D, r) \), which is a contradiction. Thus \( \tau(W_r) = r \), proving that the solutions of (1.16) and (1.17) coincide.

**Proof of Corollary 1.4 (3).** Let \( \Gamma_{\geq r} \coloneqq \{ \tilde{W} : \tau(\tilde{W}) > r \} \). Then Theorem 1.2 yields,

\[
- \lim_{r' \searrow r} \phi_r(D, r) = - \inf_{W \in \Gamma_{\geq r}} J_D(W) \leq \lim_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}(\tau_{n,d} > r) \leq \lim_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}(\tau_{n,d} \geq r) \leq - \phi_r(D, r).
\]

(5.4)

Thus, if \( r \) is a right-continuity point of \( \phi_r(D, \cdot) \), then all the inequalities above hold with equality and the proof follows.

**Proof of Corollary 1.4 (4).** Let \( \alpha = \phi_r(D, r) \). Recall that \( \tilde{B}_\Box(W, \varepsilon) \) denotes the \( \varepsilon \) ball around \( W \) in \((\mathcal{W}, \delta_{\mathcal{W}})\). Define \( \Gamma_{r, \varepsilon} = \Gamma_{\geq r} \cap \{W \in F_* : \tilde{B}_\Box(W, \varepsilon) \} \). Define \( \{\delta_{\Box}(W_{G_{n,d}}^\varepsilon, d), F_*^\varepsilon \} \geq \varepsilon \text{ and } \tau_{n,d} \geq r \} = \{W_{G_{n,d}}^\varepsilon \in \Gamma_{r, \varepsilon} \} \).

(5.5)

It is enough to show that

\[
\limsup_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}(W_{G_{n,d}}^\varepsilon \in \Gamma_{r, \varepsilon}) < - \alpha.
\]

(5.6)

Since \( \Gamma_{r, \varepsilon} \) is a closed set, using Theorem 1.2, it is enough to show that \( \inf_{W \in \Gamma_{r, \varepsilon}} J_D(W) \leq \alpha \) yields a contradiction. Now, since \( \Gamma_{r, \varepsilon} \) is compact and \( J_D \) is lower semi-continuous, \( J_D(W_r) \leq \alpha \) for some \( W_r \in \Gamma_{r, \varepsilon} \). Further,

\[
F_* = \Gamma_{\geq r} \cap \{\tilde{W} : J_D(\tilde{W}) \leq \alpha \},
\]

(5.7)

so that \( \tilde{W}_r \in F_* \). Together with \( \tilde{W}_r \in \Gamma_{r, \varepsilon} \), this yields a contradiction.
5.2 Convergence of the microcanonical partition function

We now complete the proof of Corollary 1.5 in this section. We first need the following lemma:

**Lemma 5.1.** Recall that $G_{n,d}$ is the space of graphs with degree sequence $d$. Under Assumption 1, as $n \to \infty$,

$$\frac{1}{n^2} \log |G_{n,d}| \to h_e(W_D) = -\int_0^1 \beta(x) D(x) \, dx + \frac{1}{2} \int_{[0,1]^2} \log(1 + e^{\beta(x)+\beta(y)}) \, dx \, dy,$$

(5.8)

where $h_e$ is defined in (1.22), and $\beta$ is given by Proposition 1.1.

**Proof.** Recall the definitions of $\hat{\beta}, \hat{p}_{ij}, \hat{G}_n, W_{n,d}, D_n$ and $\beta_n$ from Section 4.1. Note that

$$\mathbb{P}(\hat{G}_n = G) = \frac{e^{\sum_{i \in [n]} \hat{\beta}_i d_i}}{\prod_{i < j} (1 + e^{\hat{\beta}_i + \hat{\beta}_j})}, \quad G \in G_{n,d}.$$

(5.9)

Thus, if $d(\hat{G}_n)$ denotes the degree sequence of $\hat{G}_n$, then

$$\mathbb{P}(d(\hat{G}_n) = d) = |G_{n,d}| \frac{e^{\sum_{i \in [n]} \hat{\beta}_i d_i}}{\prod_{i < j} (1 + e^{\hat{\beta}_i + \hat{\beta}_j})}.$$  

(5.10)

Now, using (4.9), $\beta_n \to \beta$ in $L_1$ and therefore

$$\frac{1}{n^2} \log \prod_{i < j} (1 + e^{\hat{\beta}_i + \hat{\beta}_j}) = \frac{1}{n^2} \sum_{i < j} \log(1 + e^{\hat{\beta}_i + \hat{\beta}_j})$$

$$= \frac{1}{2} \int_{[0,1]^2} \log(1 + e^{\beta(x)+\beta(y)}) \, dx \, dy - \frac{1}{n^2} \sum_{i \in [n]} \log(1 + e^{2\hat{\beta}_i})$$

$$\to \frac{1}{2} \int_{[0,1]^2} \log(1 + e^{\beta(x)+\beta(y)}) \, dx \, dy,$$

where the second term in the third equality goes to zero by dominated convergence theorem. Moreover, using the fact that $D_n \to D$ in $L_1$ from Assumption 1, and that $d_i < n$, $\|\beta\|_\infty \leq C$, it follows

$$\frac{1}{n^2} \sum_{i \in [n]} \hat{\beta}_i d_i = \int_0^1 \beta_n(x) D_n(x) \, dx \to \int_0^1 \beta(x) D(x) \, dx.$$

(5.12)

Now,

$$h_e(W_D)$$

$$= -\frac{1}{2} \int_{[0,1]^2} (W_D(x,y) \log(W_D(x,y)) + (1 - W_D(x,y)) \log(1 - W_D(x,y))) \, dx \, dy$$

(5.13)

$$= -\frac{1}{2} \int_0^1 \beta(x) D(x) \, dx + \frac{1}{2} \int_{[0,1]^2} \log(1 + e^{\beta(x)+\beta(y)}) \, dx \, dy.$$

Now, turning back to (5.10), let us recall from Lemma 4.1 that $\mathbb{P}(d(\hat{G}_n) = d)$ lies in $(e^{-n^{7/4}}, 1)$. Thus,

$$\frac{1}{n^2} \log |G_{n,d}| = -\frac{1}{n^2} \sum_{i \in [n]} \hat{\beta}_i d_i + \frac{1}{n^2} \log \prod_{i < j} (1 + e^{\hat{\beta}_i + \hat{\beta}_j}) + o(1) \to h_e(W_D),$$

(5.14)
where the last step follows from (5.13). The proof is now complete.

\[ \]  

**Proof of Corollary 1.5.** Let us view $\mathcal{G}_{n,d}$ as a subspace of $\mathcal{W}$ by identifying the graphs with the corresponding empirical graphons, and let $\tilde{\mathcal{G}}_{n,d}$ denote the corresponding subset of $\mathcal{W}$. For any $\tilde{A} \subseteq \mathcal{W}$, define $\tilde{A}_n = \tilde{A} \cap \tilde{\mathcal{G}}_{n,d}$, so that $|\tilde{A}_n| < \infty$ for all $n$. Observe that

$$\tilde{\mathbf{P}}_{n,d}(\tilde{A}) = \frac{|\tilde{A}_n|}{|\tilde{\mathcal{G}}_{n,d}|}. \quad (5.15)$$

Therefore, using Theorem 1.2 together with Lemma 5.1, for any closed set $\tilde{F} \subset \tilde{\mathcal{W}}$ and open set $\tilde{U} \subset \tilde{\mathcal{W}}$,

$$\limsup_{n \to \infty} \frac{1}{n^2} \log |\tilde{F}_n| \leq - \inf_{\tilde{W} \in \tilde{F}} J_D(\tilde{W}) + h_e(W_D), \quad (5.16)$$

$$\liminf_{n \to \infty} \frac{1}{n^2} \log |\tilde{U}_n| \geq - \inf_{\tilde{W} \in \tilde{U}} J_D(\tilde{W}) + h_e(W_D). \quad (5.17)$$

Fix $\varepsilon > 0$. Since $\tau$ is bounded, there exists $(a_i)_{i=1}^k$ such that the range of $\tau$ is a subset of $\cup_{i \in [k]} [a_i, a_i + \varepsilon]$. Now, let $\tilde{F}^{a_i} := \tau^{-1}([a_i, a_i + \varepsilon])$, which is closed due to the continuity of $\tau$. Thus,

$$e^{n^2Z_{n,\tau}} \leq \sum_{i \in [k]} e^{n^2(a_i + \varepsilon)} |\tilde{F}^{a_i}| \leq k \max_{i \in [k]} e^{n^2(a_i + \varepsilon)} |\tilde{F}^{a_i}|. \quad (5.18)$$

Thus, (5.16) implies that

$$\limsup_{n \to \infty} Z_{n,\tau} \leq \max_{i \in [k]} \left( a_i + \varepsilon - \inf_{\tilde{W} \in \tilde{F}^{a_i}} J_D(\tilde{W}) \right) + h_e(W_D)$$

$$\leq \varepsilon + \max_{i \in [k]} \sup_{\tilde{W} \in \tilde{F}^{a_i}} (\tau(\tilde{W}) - J_D(\tilde{W})) + h_e(W_D)$$

$$= \varepsilon + \sup_{\tilde{W} \in \tilde{\mathcal{W}}} (\tau(\tilde{W}) - J_D(\tilde{W})) + h_e(W_D), \quad (5.19)$$

where in the second step we have used the fact that $\tau(\tilde{W}) \geq a$ for all $\tilde{W} \in \tilde{F}^{a_i}$. For the lower bound, let $\tilde{U}^{b_i} = \tau^{-1}([b_i, b_i + \varepsilon])$ for $i \leq l$ be such that $\cup_{i \in [l]} (b_i, b_i + \varepsilon)$ covers the range of $\tau$. An identical computation to above yields that

$$\liminf_{n \to \infty} Z_{n,\tau} \geq - \varepsilon + \sup_{\tilde{W} \in \tilde{\mathcal{W}}} (\tau(\tilde{W}) - J_D(\tilde{W})) + h_e(W_D). \quad (5.20)$$

The proof of (1.23) now follows by taking $\varepsilon \to 0$. To see (1.24), the continuity of $\tau$, together with (5.16) implies that

$$\limsup_{n \to \infty} \frac{1}{n^2} \log N_{n,\tau}(d, r) \leq - \phi_r(D, r) + h_e(W_D). \quad (5.21)$$

Also, $|N_{n,\tau}(d, r)|$ is at least the number of graphs with degree sequence $d$ and $\tau(\tilde{W}) > r$. Thus, (5.17) implies that

$$\liminf_{n \to \infty} \frac{1}{n^2} \log N_{n,\tau}(d, r) \geq - \lim_{r' \uparrow r} \phi_r(D, r) + h_e(W_D). \quad (5.22)$$

The proof of (1.24) is now complete using the right continuity of $\phi_e(D, \cdot)$ at $r$. \qed
Acknowledgements. The authors sincerely thank Amir Dembo for stimulating discussions. This work was initiated during the BIRS workshop “Spin glasses and Related topics (2018)”. The authors thank the BIRS hosts, as well as the workshop organizers: Antonio Auffinger, Wei-Kuo Chen, Dmitry Panchenko, and Lenka Zdeborová. The authors also thank Christian Borgs and Jennifer Chayes for motivating this research direction, and Bhaswar Bhattacharya and Sumit Mukherjee for helpful comments on the first draft of this paper. The authors thank Julia Gaudio and Samantha Petti for pointing out minor errors in the proof of Proposition 3.2.

References

[1] Augeri, F. (2018). Nonlinear large deviation bounds with applications to traces of Wigner matrices and cycles counts in Erdős–Rényi graphs. arXiv:1810.01558.
[2] Barvinok, A. and Hartigan, J. A. (2013). The number of graphs and a random graph with a given degree sequence. Random Struct. Algor., 42(3):301–348.
[3] Bhattacharya, B. B., Ganguly, S., Lubetzky, E., and Zhao, Y. (2017). Upper tails and independence polynomials in random graphs. Adv. Math., 319:313–347.
[4] Blitzstein, J. and Diaconis, P. (2011). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. Internet Math., 6(4):489–522.
[5] Bollobás, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. European J. Combin., 1(4):311–316.
[6] Borgs, C., Chayes, J. T., Cohn, H., and Ganguly, S. (2015). Consistent nonparametric estimation for heavy-tailed sparse graphs. arXiv:1508.06675.
[7] Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., and Vesztergombi, K. (2008). Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. Adv. Math., 219(6):1801–1851.
[8] Chatterjee, S. (2017). Large Deviations for Random Graphs. Springer International Publishing, école d’été edition.
[9] Chatterjee, S. and Dembo, A. (2016). Nonlinear large deviations. Adv. Math., 299:396–450.
[10] Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. Ann. Statist., 41(5):2428–2461.
[11] Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. Ann. Appl. Probab., 21(4):1400–1435.
[12] Chatterjee, S. and Varadhan, S. (2011). The large deviation principle for the Erdős–Rényi random graph. European J. Combin., 32(7):1000–1017.
[13] Cook, N. A. and Dembo, A. (2018). Large deviations of subgraph counts for sparse Erdős–Rényi graphs. arXiv:1809.11148.
[14] Del Genio, C. I., Kim, H., Toroczkai, Z., and Bassler, K. E. (2010). Efficient and Exact Sampling of Simple Graphs with Given Arbitrary Degree Sequence. PLOS ONE, 5(4):1–7.
[15] Dembo, A. and Lubetzky, E. (2018). A large deviation principle for the Erdős–Rényi uniform random graph. Electron. Commun. Probab., 23.
[16] den Hollander, F., Mandjes, M., Roccaverde, A., and Starreveld, N. J. (2018). Ensemble equivalence for dense graphs. Electron. J. Probab., 23.
[17] Eldan, R. (2018). Gaussian-width gradient complexity, reverse log-
Sobolevinequalities and nonlinear large deviations. *Geom. Funct. Anal. (GAF A)*, 28(6):1548–1596.

[18] Harel, M., Mousset, F., and Samotij, W. (2019). Upper tails via high moments and entropic stability. *arXiv:1904.08212*.

[19] Janson, S., Łuczak, T., and Rucinski, A. (2000). Random Graphs. Wiley, New York.

[20] Kenyon, R., Radin, C., Ren, K., and Sadun, L. (2017a). Multipodal structure and phase transitions in large constrained graphs. *J. Stat. Phys.*, 168(2):233–258.

[21] Kenyon, R., Radin, C., Ren, K., and Sadun, L. (2017b). The phases of large networks with edge and triangle constraints. *J. Phys. A Math. Gen.*, 50(43):435001.

[22] Kenyon, R., Radin, C., Sadun, L., and Ren, K. (2016). Bipodal structure in oversaturated random graphs. *Int. Math. Res. Notices (IMRN)*, 2018(4):1009–1044.

[23] Kenyon, R. and Yin, M. (2017). On the asymptotics of constrained exponential random graphs. *J. Appl. Probab.*, 54(1):165–180.

[24] Liebenau, A. and Wormald, N. (2017). Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph. *arXiv:1702.08373*.

[25] Lovász, L. (2012). Large Networks and Graph Limits. American Mathematical Society.

[26] Lovász, L. and Szegedy, B. (2007). Szemerédi’s lemma for the analyst. *Geom. Funct. Anal. (GAF A)*, 17(1):252–270.

[27] Lubetzky, E. and Zhao, Y. (2015). On replica symmetry of large deviations in random graphs. *Random Struct. Algor.*, 47(1):109–146.

[28] Lubetzky, E. and Zhao, Y. (2017). On the variational problem for upper tails in sparse random graphs. *Random Struct. Algorithms*, 50(3):420–436.

[29] McKay, B. D. and Wormald, N. C. (1990). Asymptotic Enumeration by Degree Sequence of Graphs of High Degree. *European J. Combin.*, 11(6):565–580.

[30] Orsini, C., Dankulov, M. M., Colomer-de Simón, P., Jamakovic, A., Mahadevan, P., Vahdat, A., Bassler, K. E., Toroczkai, Z., Boguñá, M., Caldarelli, G., Fortunato, S., and Krioukov, D. (2015). Quantifying randomness in real networks. *Nat. Commun.*, 6:8627.

[31] Radin, C. (2018). Phases in large combinatorial systems. *Ann. Inst. H. Poincaré D*, 5(2):287–308.

[32] Radin, C. and Sadun, L. (2013). Phase transitions in a complex network. *J. Phys. A: Math. Theor.*, 46(30):305002.

[33] Radin, C. and Yin, M. (2013). Phase transitions in exponential random graphs. *Ann. Appl. Probab.*, 23(6):2458–2471.

[34] Squartini, T., de Mol, J., den Hollander, F., and Garlaschelli, D. (2015). Breaking of ensemble equivalence in networks. *Phys. Rev. Lett.*, 115(26):268701.

[35] van der Hoorn, P., Lippner, G., and Mossel, E. (2019). Regular graphs with linearly many triangles. *arXiv:1904.02212*.

[36] Wormald, N. C. (1999). Models of random regular graphs. In Lamb, J. D. and Preece, D. A., editors, Surveys in Combinatorics, 1999, pages 239–298. Cambridge University Press.

[37] Yin, M. (2015). Large deviations and exact asymptotics for constrained exponential random graphs. *Electron. Commun. Probab.*, 20.

[38] Ying, X. and Wu, X. (2009). Graph Generation with Prescribed Feature Constraints. In *Proc. SIAM Int Conf Data Min*, pages 966–977.