LARGE DEVIATION FOR THE EMPIRICAL DEGREE DISTRIBUTION OF AN ERDÖS-RENYI GRAPH

SUMIT MUKHERJEE

ABSTRACT. With \((d_1, \ldots, d_n)\) denoting the labeled degrees of an Erdős Renyi graph with parameter \(\beta/n\), the large deviation principle for \(\frac{1}{n} \sum_{j=1}^{n} \delta_{d_j}\) (the empirical distribution of the degrees) is derived with a good rate function, with respect to a topology stronger than the weak topology.

As an application the degeneracy of some sparse ERGM models used in social networks is studied rigorously, showing in particular that using terms such as "gwd (geometrically weighted degree)" and alternating \(k\) stars does not cause degeneracy, whereas using a \(k\)-star term does.

1. INTRODUCTION

1.1. Brief outline of the paper. The large deviation result is contained in subsection 1.2 (Theorem 1.2), and is the main result of this paper. Subsection 1.3 contains the statements of Theorem 1.5 and Theorem 1.8, which are developed with application in mind. Theorem 1.5 demonstrates that using the number of \(k\)-stars as a sufficient statistic in an ERGM causes degeneracy (see remark 1.6). Theorem 1.8 characterizes a class of sufficient statistics which do not cause degeneracy. This class includes sufficient statistics like geometrically weighted degree and alternating \(k\) stars (see remarks 1.7 and 1.9) used in social sciences which are already known to not cause degeneracy at an empirical level (see [SPRH],[HIII], [MIII]). For an intuitive reasoning of why this class of statistics do not cause degeneracy see remark 1.4.

Section 2 explores some properties of the rate function associated with the large deviation. Section 3 is dedicated to proving Theorem 1.2 via a series of lemmas. Section 4 contains the proofs of Theorem 1.5 and Theorem 1.8.

Section 5 introduces an example of a particular ERGM, and analyzes the model in light of the above theory. This model, dubbed in this paper as the "sparse penalty model", is of interest in social science community, as conveyed to the author at an AIM conference on Exponential Random Graph models (ERGM).

1.2. Large Deviation Principle. Let \(G_n\) denote the space of all simple labelled undirected graphs on \(n\) vertices. For any \(G_n \in G_n\) let \(\mathbf{d} = \mathbf{d}(G_n) := (d_1, \ldots, d_n)\) denote the labeled degree sequence of \(G_n\). Also let \(E = E(G_n) := \frac{1}{2} \sum_{j=1}^{n} d_j\) denote the number of edges in \(G_n\). The empirical distribution of the degree sequence defined by \(\mu^{(n)} := \frac{1}{n} \sum_{j=1}^{n} \delta_{d_j}\), is a probability measure on \(\mathbb{N}_0 := \mathbb{N} \cup \{0\}\).
An equivalent definition of $\mu^{(n)}$ is the following:

Set $h_i := \# \{1 \leq j \leq n : d_j = i \}$, for $0 \leq i \leq n - 1$, i.e. $h_i$ is the number of vertices in $G_n$ of degree $i$. $\{h_i\}_{i=0}^{n-1}$ will henceforth be called the degree frequency vector. Note that $\mu^{(n)}$ is the probability measure which puts mass $h_i/n$ at $i$, for $0 \leq i \leq n - 1$. The degree frequency vector does not depend on the ordering of the vertices, and neither does $\mu^{(n)}$.

Set

$$S := \{\mu \in \mathbb{P}(\mathbb{N}_0) : \mu := \sum_{i=1}^{\infty} i \mu_i < \infty \}.$$ 

Thus $S$ is the set of all probability measures on $\mathbb{N}_0$ with finite mean. Equip $S$ with the following topology:

$$\nu_n, \nu \in S, \nu_n \rightarrow \nu \text{ if } \nu_n \overset{w}{\rightarrow} \nu, \text{ and } \nu_n \rightarrow \nu.$$ 

By Scheffe’s theorem, convergence in $S$ is also equivalent to convergence in the metric $d$ given by

$$d(\mu, \nu) = \sum_{i=1}^{\infty} i |\mu_i - \nu_i|.$$ 

Note in passing that $S$ is not compact with respect to weak convergence, and hence not compact with respect to $d(.,.)$.

Let $\mathbb{P}_n$ denote the Erdős Rényi distribution on $G_n$ with parameter $\beta/n$. It is easy to see that $d(\mu^{(n)},p_\beta) \overset{P}{\rightarrow} 0$ under $\mathbb{P}_n$, where $p_\beta \in S$ is the Poisson distribution with parameter $\beta$. The large deviation principle in the following theorem characterizes the probability that $\mu^{(n)}$ is away from $p_\beta$.

In a very recent paper in [BC], the authors derive a large deviation principle for Erdos Renyi graphs under the topology of local weak convergence. As a consequence, they deduce a large deviation principle for $\mu^{(n)}$ with a good rate function under the topology of weak convergence on $S$ (see [BC, Theorem 1.8]), which is given by the metric $\sum_{i=0}^{\infty} |\mu_i - \nu_i|$. See also [DM, Corollary 2.2] which gives the large deviation for $\mu^{(n)}$ with respect to weak topology, with a minor correction needed for the formula of the rate function in the case $\beta > \bar{\mu}$.

This paper proves a large deviation principle for $\mu^{(n)}$ with the same rate function as in [BC], under a stronger topology induced by the metric $d(.,.)$ above which also guarantees convergence of means. One motivation for the need for a stronger topology is to prove Theorem 1.8, which computes limiting log normalizing constants for a class of models of interest in social sciences, and also gives a natural sufficient condition to check the degeneracy of Exponential Random Graph Models whose sufficient statistics depend only on the degree distribution. This is described in more detail in section 1.3.

The following definition introduces the rate function for the large deviation principle, henceforth denoted by l.d.p. for convenience.
Definition 1.1. Define the function $I : \mathcal{S} \mapsto [-\infty, \infty]$ by

$$I(\mu) = \sum_{i=0}^{\infty} \mu_i \log(i!\mu_i) - \frac{\mu}{2} \log(\mu\beta) + \frac{\mu + \beta}{2} = D(\mu||p_\beta) + \frac{1}{2}(\mu - \beta) + \frac{\mu}{2} \log \beta - \frac{\mu}{2} \log \mu.$$  

It will be shown in section 2 that $I$ is non-negative, and a good rate function (i.e. its level sets are compact.) Thus $I(.)$ is a valid candidate for the rate function of a l.d.p.

The main result of the paper is now stated below. As a convention, the infimum over an empty set is taken to be $\infty$.

Theorem 1.2. For any $A \subset \mathcal{S}$,

$$\liminf_{n \to \infty} \frac{1}{n} \log P_n(\mu^{(n)} \in A) \geq - \inf_{\mu \in A^o} I(\mu).$$ (1.2)

In particular, $\mu^{(n)}$ satisfies a large deviation principle in $\mathcal{S}$ with the good rate function $I(.)$.

The l.d.p. follows from (1.1) and (1.2), which will be proved in section 3. Note that it is possible to assign probability to any $A \subset \mathcal{S}$, as the probability $P_n$ puts mass only on a finite set in $\mathcal{S}$. The l.d.p. bounds in Theorem 1.2 hold for all subsets of $\mathcal{S}$.

1.3. Applications. As an application of Theorem 1.2, section 4 explores some Exponential Random Graph models frequently used in social network studies. The following long definition introduces the class of probability distributions that can be handled using this approach.

Definition 1.3.  

- Let

$$F := \{f : \mathbb{N}_0 \mapsto \mathbb{R}, \ f(0) = 0, \ \limsup_{i \to \infty} \frac{f(i)}{i} < \infty\}.$$  

For any $f \in F$, $\mu \in \mathcal{S}$ define $\mu(f) := \sum_{i=0}^{\infty} \mu_i f(i)$. Note that $\mu(f) \in [-\infty, \infty]$ always exists, as the given condition implies there exists $C < \infty$ such that $f(i) \leq Ci$ for all $i \in \mathbb{N}_0$, and so $\mu(f) \leq C\mu < \infty$ for any $\mu \in \mathcal{S}$.

- For $f \in F$, $\theta \in [0, \infty)$, let

$$C(\theta, f) := \log \left( \sum_{i=0}^{\infty} \frac{1}{i!} \theta^i e^{f(i)} \right).$$

Using the bound $f(i) \leq Ci$ it follows that $C(\theta, f) < \infty$, and $\sigma = \sigma_{\theta, f}$ defined by

$$\sigma_i := \frac{1}{i!} e^{-C(\theta, f)} \theta^i e^{f(i)}$$

belongs to $\mathcal{S}$.

- Set $\Omega_f := \{\sigma_{\theta, f}, \theta \in [0, \infty)\} \subset \mathcal{S}$. 
• For \( f \in F \), define
\[
J(f) := \inf_{\theta \geq 0} \left\{ I(\sigma_{\theta,f}) - C(\theta, f) - \frac{m(\theta)}{2} \log(m(\theta)\beta) + \frac{m(\theta) + \beta}{2} \right\},
\]
where \( m(\theta) := \sigma_{\theta,f} \). That the last inequality holds follows by a direct computation using the formula for \( I(\cdot) \). Note that \( J(f) \) is defined in terms of a one dimensional optimization problem.

• Finally, denote by \( Q_{n,f}(\cdot) \) the following probability distribution on \( \mathbb{G}_n \):
\[
Q_{n,f}(G_n) := \frac{1}{Z_n(f)} \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{\binom{n}{2}} \left( \sum_{i=0}^{n-1} h_i f(i) \right) = \frac{1}{Z_n(f)} \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{\binom{n}{2}} e^{\gamma \mu(n)(f)},
\]
where \( Z_n(f) \) is the normalizing constant. As a comment, the assumption \( f(0) = 0 \) is not a restriction in any sense. This is because both the probability distributions \( \sigma_{\theta,f} \) and \( Q_{n,f} \) remain invariant if a constant is added to \( f \). Also note in passing that both \( Z_n(f) \) and \( Q_{n,f} \) also depend on \( \beta \), but this dependence is not made explicit as \( \beta \) will remain fixed throughout the paper, excluding section 5.

Remark 1.4. An intuitive explanation for the definition of \( F \) is that if either \( f \equiv 0 \) or \( f \) is linear, then the model \( Q_{n,f} \) is exactly Erdős Renyi, and so a choice of \( f \) growing at most linearly will not take it too far from an Erdős-Renyi, and hence should be reasonably well behaved.

As an example of a probability distribution of the form \( Q_{n,f} \), consider the \( k \) star model on \( \mathbb{G}_n \) (to be defined precisely in Theorem 1.5 below), where the sufficient statistic is the number of \( k \) stars \( T_k(G_n) \). Note that
\[
T_k = \sum_{j=1}^{n} \binom{d_j}{k} = \sum_{i=0}^{n-1} \binom{i}{k} h_i = n\mu(n)(g),
\]
where \( g : \mathbb{N}_0 \rightarrow \mathbb{N} \) is given by \( g(i) := \binom{i}{k} \). It is well known at an empirical level that models using sub-graph counts such as \( k \) stars usually leads to degeneracy. The next theorem gives a way of establishing this rigorously.

Theorem 1.5. Consider the \( k \) star probability distribution on \( \mathbb{G}_n \) given by \( Q_{n,\gamma g}(\cdot) \), where \( g(i) = \binom{i}{k} \), \( Q_{n,\cdot}(\cdot) \) is as in definition 1.3, and \( \gamma \in \mathbb{R} \) is a parameter. Mathematically the model can be written as
\[
Q_{n,\gamma g}(G_n) = \frac{1}{Z_n(\gamma g)} \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{\binom{n}{2}} e^{\gamma \sum_{i=0}^{n-1} h_i g(i)}.
\]

(a) If \( \gamma > 0 \) then
\[
\lim_{n \to \infty} \frac{1}{n} \log Z_n(\gamma g) = \infty.
\]

(b) If \( \gamma < 0 \) then
\[
\lim_{n \to \infty} \frac{1}{n} \log Z_n(\gamma g) = \sup_{\mu \in \mathcal{S}} \{ \gamma \mu(g) - I(\mu) \} = -J(\gamma g) < 0,
\]
(c) The supremum in part (b) is achieved on a finite set $A_{\gamma g} \subset \Omega_{\gamma g}$, and for any open $U$ containing $A_{\gamma g}$, 
\[ d(\mu^{(n)}(U^c)) \xrightarrow{P} 0 \]
under $Q_{n,\gamma g}$.

**Remark 1.6.** Theorem (1.5) says that the behavior of the model $Q_{n,\gamma g}$ changes drastically at $\gamma = 0$. For $\gamma < 0$ the degree distribution stabilizes, and any limit point of the degree distribution is of the form $\mu_i \propto e^{\gamma g(i)\theta^i/i!}$ for some $\theta \geq 0$. Also the normalizing constant goes to 0 exponentially fast.

On the other hand for $\gamma > 0$ the cascading effect takes over, and so the mass of the degree distribution escapes to $\infty$. Also the normalizing constant goes to $\infty$ super exponentially. Thus the limiting normalizing constant jumps from a negative number for $\gamma < 0$ to $\infty$ for $\gamma > 0$, bypassing all positive finite values, and there seems to be no uniform scaling in this model which prevents this behavior. Note that for $\gamma < 0$ the function $\gamma g$ is in $F$ as defined in 1.3, whereas for $\gamma > 0$ this is no longer true.

A possible alternative is to scale $\gamma$ differently for positive and negative values, but this will not be explored in this paper.

**Remark 1.7.** Thus to prevent degeneracy for both positive and negative values of the parameter one needs to impose a slightly stronger restriction than the class $F$. It is known in social network literature that using terms like geometrically weighted degree or alternating $k$ stars in an ERGM model does not cause degeneracy. In the class of probability distributions $Q_{n,f}$ the choice $f(i) = e^{-\lambda_1 i}$ gives the sufficient statistic 
\[ \sum_{i=0}^{n-1} e^{-\lambda_1 i} h_i, \]
which matches the form of geometrically weighted degree statistic given in [SPRH, (11)] with $\alpha = \lambda_1$. Similarly, the choice 
\[ f(i) = \sum_{k=2}^{i} (-1)^k \left( \begin{array}{c} i \\ k \end{array} \right) \lambda_2^{k-2} \]
gives the sufficient statistic 
\[ \sum_{i=0}^{n-1} h_i \sum_{k=2}^{i} (-1)^k \left( \begin{array}{c} i \\ k \end{array} \right) \lambda_2^{k-2} \]  
\[ = \sum_{k=2}^{n-1} (-1)^k \lambda_2^{k-2} \sum_{i=k}^{n-1} \left( \begin{array}{c} i \\ k \end{array} \right) h_i = \sum_{k=2}^{n-1} (-1)^k \lambda_2^{k-2} T_k, \]
which matches the form of alternating $k$ star statistic in [SPRH, (13)] with $\lambda = 1/\lambda_2$.

The next theorem establishes that if $|f|$ is assumed to grow at most linearly near $\infty$ the resulting degree distribution indeed stabilizes, and gives a description of the limit points of the degree distribution.

**Theorem 1.8.** Let $f : \mathbb{N}_0 \mapsto \mathbb{R}$ be such that 
\[ \limsup_{i \to \infty} \frac{|f(i)|}{i} < \infty, \]
and consider the probability distribution $Q_{n,\gamma f}$ on $\mathbb{G}_n$ on $\mathbb{G}_n$ for $\gamma \in \mathbb{R}$. Then
(a) \[
\frac{1}{n} \log Z_n(f) = \sup_{\mu \in S} \{ \gamma \mu(f) - I(\mu) \} = -J(f).
\]

(b) The supremum in part (a) is attained on a finite set \(A_{\gamma f} \subset \Omega_f\), and for any open set \(U\) containing \(A_{\gamma f}\),
\[
d(\mu^{(n)}, U^c) \xrightarrow{P} 0
\]
under \(Q_{n,\gamma f}\).

**Remark 1.9.** That theorem 1.8 covers both the geometrically weighted degree statistic with \(\lambda_1 > 0\) and the alternating k star statistic with \(\lambda_2 \in (0, 1)\) can be checked easily as follows:

For the gwd statistic \(f(i) = e^{-\lambda_1 i} \leq 1\) is bounded, and so the condition of Theorem 1.8 holds trivially.

For the alternating \(k\) star statistic, note that \(f(i) = \frac{1}{\lambda_2^2} [(1 - \lambda_2)^i - 1 + i\lambda_2]\)
satisfies \(\limsup_{i \to \infty} |f(i)|/i = 1/\lambda_2 < \infty\), and so again the condition of Theorem 1.8 holds.

In this way, Theorem 1.8 gives a large class of choices \(f\) for which the corresponding model \(Q_{n, f}\) is not degenerate.

By the theorem, any limit of the degree distribution is of the form \(\mu_i \propto e^{\gamma f(i)\theta^i/i!}\) for some \(\theta \geq 0\). Note that if \(f\) is not linear, then no distribution in \(A_{\gamma f}\) is a Poisson, which is the limiting degree distribution under an Erdős Renyi. Thus for non linear \(f\) the degree distribution of the graph does not look like that of an Erdős Renyi for large \(n\). If however \(f\) is linear, then the model \(Q_{n,\gamma f}\) is again an Erdős Renyi with a different parameter, and hence the degree distribution does converge to a Poisson with a different mean.

**Remark 1.10.** Even though Theorem 1.5 and Theorem 1.8 characterize the form of possible limits of the limiting degree distribution, they fall short of establishing weak convergence of the degree distribution, nor do they give a closed form expression for the minimizing value(s) of \(\theta\) in general. One way to establish weak convergence of the degree distribution under \(Q_{n, f}\) for a particular choice of \(f\) is to show that there is a unique \(\theta\) which is a minimizer in the definition of \(J_f\).

From a computational point of view, since \(J_f\) is defined via a minimum over a scaler parameter \(\theta\), it might be possible in a given problem to do a numerical optimization to solve for \(\theta\), and estimate the normalizing constant.

**1.4. Connections with dense graph l.d.p.** In [CV], the authors develop a large deviation principle for dense Erdős Renyi graphs (probability of an edge is \(p \in (0, 1)\) with respect to the cut metric. Using the techniques developed there, [CD] computed the limiting log partition function of a wide range of Exponential Random Graph models, where the sufficient statistic are sub graph counts (properly scaled). The limiting log partition function is expressed in terms of an infinite dimensional optimization problem, which can be reduced to one dimensional optimization problem for certain values of the parameters.
In comparison, this paper develops l.d.p. for (the degree distribution of) the sparse Erdős-Rényi graphs (probability of an edge is $\beta/n$), and uses the l.d.p. to compute limiting log partition functions of a class of ERGMs. Note that since this paper only develops an l.d.p. for the degree distribution (and not for the "entire graph" as done in [CV]), there is the added restriction that all the sufficient statistics must be functions of the degree distribution. For e.g. this approach can deal with $k$ stars (as demonstrated in Theorem 1.5), but not with triangles, as the number of triangles is not a function of the degree distribution.

2. Properties of the rate function

This section explores some properties of $I(\cdot)$. The first lemma shows that $I(\cdot)$ is a good rate function, and so a valid candidate for the rate function of a l.d.p.

Lemma 2.1. Let $I(\cdot)$ be as defined in (1.1). Then for any $\alpha \in \mathbb{R}$ the set $I_\alpha := \{\mu \in \mathcal{S} : I(\mu) \leq \alpha\}$ is compact. i.e. for any sequence $\nu^{(n)} \in I_\alpha$ there exists a further subsequence $\nu^{(n_k)}$ which converges in $I_\alpha$.

Proof. To begin, first note that $D(\cdot||\cdot)$ is lower semi continuous with respect to weak topology, and so lower semi continuous with respect to $d(\cdot,\cdot)$. Also by definition $\mu \mapsto \bar{\mu}$ is continuous with respect to $d(\cdot,\cdot)$, and so it follows trivially that $I$ is lower semi continuous, and so $I_\alpha$ is closed. Thus to show compactness, it suffices to show that there exists a subsequence $\nu^{(n_k)}$ of the original sequence $\nu^{(n)}$ which converges in $\mathcal{S}$.

Proceeding to show this, first note that

$$\log i! = \sum_{k=1}^{i} \log k \geq \int_{x=0}^{i} \log x dx = i \log i - i,$$

and so

$$\sum_{i=0}^{\infty} \mu_i \log i! \geq \sum_{i=0}^{\infty} i \log i \mu_i - \bar{\mu} \geq \bar{\mu} \log \bar{\mu} - \bar{\mu}, \tag{2.1}$$

where the last inequality follows by Jensen’s inequality on noting that $x \log x$ is convex.

Also, setting $\nu \in \mathcal{S}$ defined by $\nu_i := 2^{-(i+1)}$ it follows that

$$\sum_{i=0}^{\infty} \mu_i \log \mu_i = D(\mu||\nu) + \sum_{i=0}^{\infty} \mu_i \log \nu_i \geq -(\bar{\mu} + 1) \log 2. \tag{2.2}$$

Thus combining (2.1) and (2.2) gives

$$I(\mu) \geq \frac{1}{2} \bar{n} \log \bar{\mu} - \bar{\mu}\left( \log 2 + \frac{1 + \log \beta}{2} \right) + \frac{\beta}{2} - \log 2 = g_1(\bar{\mu}), \tag{2.3}$$

where $g_1(x) := \frac{1}{2} x \log x - x \left( \log 2 + \frac{1 + \log \beta}{2} \right) + \frac{\beta}{2} - \log 2$. Since $g_1(x)$ is continuous and diverges to $\infty$ as $x \to \infty$, it follows that $g_1(\bar{\mu}) \leq \alpha$ implies $\bar{\mu} \leq C_1$ for some $C_1 < \infty$. By a first moment argument using Chebyscheff’s inequality it follows that $I_\alpha$ is tight in weak topology, and so by Prohorov’s theorem there exists a subsequence $\nu^{(n_k)}$ which converges weakly in $\mathbb{P}(N_0)$ to $\nu$, say.
To complete the proof one has to show that $\nu < \infty$, and $\nu^{(n_k)} \rightarrow \nu$. For this it suffices to show that the laws $\nu^{(n_k)}$ are uniformly integrable. This will follow if it can be shown that $\sup_{\mu \in I_{\alpha}} \sum_{i=0}^{\infty} i \log \mu_i < \infty$. To check the last fact, note that by (2.1) and (2.2), for any $\mu \in I_{\alpha}$,

$$\sum_{i} i \log \mu_i \leq \mu + \sum_{i=0}^{\infty} \mu_i \log i! = I(\mu) + \frac{\mu - \beta}{2} + \frac{\mu}{2} \log(\mu \beta) - \sum_{i=0}^{\infty} \mu_i \log \mu_i$$

$$\leq I(\mu) + \frac{\mu - \beta}{2} + \frac{\mu}{2} \log(\mu \beta) + (\mu + 1) \log 2$$

$$\leq \alpha + \sup_{0 \leq x \leq C_1} g_2(x) =: C_2 < \infty,$$

where $g_2(x) := \frac{x - \beta}{2} + \frac{x}{2} \log(x \beta) + (x + 1) \log 2$.

\[ \square \]

The second lemma develops tools to be used in section 4 to derive theorem 1.5 and theorem 1.8.

**Lemma 2.2.** Let $f \in F$ be such that $\inf_{\mu \in \mathcal{S}} \{ I(\mu) - \mu(f) \} < \infty$.

(a) The function $\mu \mapsto \mu(f)$ is upper semi continuous.

(b) $\inf_{\mu \in \mathcal{S}} \{ I(\mu) - \mu(f) \} = J_f$, where $J_f$ is as defined in (1.3). The infimum in the definition of $J_f$ is attained over a finite set, and consequently the infimum of $\mu \mapsto I(\mu) - \mu(f)$ is attained on a non-empty finite set $A_f \subset \Omega_f$. Further any $\sigma_{\theta,f} \in A_f$ satisfies the relation $\theta = \sqrt{\beta \sigma_{\theta,f}}$.

(c) For any open $U$ containing $A_f$,

$$\inf_{\mu \in U^c} \{ I(\mu) - \mu(f) \} > \inf_{\mu \in \mathcal{S}} \{ I(\mu) - \mu(f) \}.$$

**Proof.** Since $f \in F$, there exists $C \in (0, \infty)$ such that $f(i) \leq Ci$ for all $i \in \mathbb{N}_0$.

(a) Let $\nu^{(k)} \rightarrow \nu$ in the metric $d(.,.)$. Then for any $N \geq 1$,

$$\sum_{i=0}^{\infty} \nu_i^{(k)} f(i) \leq \sum_{i=0}^{N} \nu_i^{(k)} f(i) + C \sum_{i=N+1}^{\infty} \nu_i^{(k)} f(i) + Cd(\nu^{(k)}, \nu) + C \sum_{i=N+1}^{\infty} i \nu_i$$

Taking limits as $k \rightarrow \infty$ gives

$$\limsup_{k \rightarrow \infty} \nu^{(k)}(f) \leq \sum_{i=0}^{N} \nu_i f(i) + C \sum_{i=N+1}^{\infty} i \nu_i.$$ 

Since $\nu < \infty$, letting $N \rightarrow \infty$ gives

$$\limsup_{n \rightarrow \infty} \nu^{(k)}(f) \leq \nu(f),$$

and so $\mu \mapsto \mu(f)$ is upper semi continuous.
(b) Since \( \alpha := \inf_{\mu \in S} \{ I(\mu) - \mu(f) \} \leq \infty \), it suffices to minimize \( I(\mu) - \mu(f) \) over \( \mu \) such that \( I(\mu) - \mu(f) \leq \alpha + 1 \). Also by (2.3), \( I(\mu) - \mu(f) \geq g_3(p) \), where \( g_3(x) := g_1(x) - Cx \) is continuous and diverges to \( \infty \) as \( x \to \infty \). Thus \( g_3(p) \leq \alpha + 1 \) implies \( p \leq C_3 \) for some \( C_3 < \infty \). But then the minimizing set is contained in

\[
\{ \mu : I(\mu) \leq \mu(f) + \alpha + 1 \} \subset \{ \mu : I(\mu) \leq CC_3 + \alpha + 1 \} = I_{CC_3+\alpha+1},
\]

which is compact by Lemma 2.1. Thus the infimum of the lower semi continuous function \( I(\mu) - \mu(f) \) is achieved on a non empty compact set \( A_f \).

Let \( \mu \in A_f \) be any point where the minimum is attained, and let \( \nu \in S \) be arbitrary. By convexity of \( S \), \( (1 - t)\mu + t\nu \in S \) for any \( t \in [0, 1] \), and so with \( \theta := \sqrt{p\beta} \),

\[
\frac{\partial}{\partial t} I((1-t)\mu + t\nu) - (1-t)\mu(f) - t\nu(f) \big|_{t=0} \geq 0
\]

\[
\Leftrightarrow \sum_{i=0}^{\infty} \left( 1 + \log \mu_i + \log i! - \frac{i}{2} \log p \right) - \frac{i}{2} \log \beta + \frac{i}{2} - f(i) (\nu_i - \mu_i) \geq 0
\]

\[
\Leftrightarrow \sum_{i=0}^{\infty} \left( \log \mu_i + \log i! - i \log \lambda - f(i) \right) (\nu_i - \mu_i) \geq 0
\]

\[
\Leftrightarrow D(\mu||\sigma_{\theta,f}) + D(\nu||\mu) \leq D(\nu||\sigma_{\theta,f}).
\]

where \( \sigma_{\theta,f} \) is as defined in definition (1.3). Since this holds for all \( \nu \in S \), setting \( \nu = \sigma_{\theta,f} \) gives \( D(\mu||\sigma_{\theta,f}) = 0 \), and so \( \mu = \sigma_{\theta,f} \). Thus \( A_f \subset \Omega_f \), and further any \( \sigma_{\theta,f} \in A_f \) satisfies \( \theta = \sqrt{3\beta^2} \).

Also compactness of \( A_f \) forces that the set of minimizers \( \theta \) in the definition of \( J_f \) is a compact subset of \([0, \infty)\). Finally since an analytic non constant function on a bounded domain cannot have infinitely many minimizers, the set of minimizers in \( \theta \) must be finite. This completes the proof of part (b).

(c) If \( \inf_{\mu \in U^c} \{ I(\mu) - \mu(f) \} = \infty \) then there is nothing to show. Otherwise by a similar argument as in part (b), to minimize \( I(\mu) - \mu(f) \) over \( U^c \) it is sufficient to minimize over \( I_{\alpha} \cap U^c \) for some \( \alpha < \infty \). Since \( U^c \) is closed, \( I_{\alpha} \cap U^c \) is compact and so the infimum over \( U^c \) is attained. But since none of these minimizers are in \( B_f \), it follows that \( \inf_{\mu \in U^c} \{ I(\mu) - \mu(f) \} > \inf_{\mu \in S} \{ I(\mu) - \mu(f) \} \).

As an immediate consequence of corollary of Lemma 2.2, the following corollary shows that the rate function \( I(.) \) is indeed non-negative, with a unique global minima at \( p_{\beta} \).

**Corollary 2.3.** The unique global minimizer of \( I(.) \) over \( S \) is at \( p_{\beta} \), with \( I(p_{\beta}) = 0 \).

**Proof.** Choosing \( f \) to be the identically 0 function, it follows by part (b) of Lemma 2.2 that the minimum of \( I(\mu) \) over \( S \) is attained over the class \( p_{\theta} \) for \( \theta \geq 0 \). Also

\[
I(p_{\theta}) = \frac{1}{2}(\beta - \theta + \theta \log \theta - \theta \log \beta) = \frac{1}{2}D(p_{\theta}||p_{\beta}),
\]

and so the unique global minimum occurs at \( p_{\beta} \), with \( I(p_{\beta}) = 0 \).

The last result of this section is the following proposition is exploratory and shows that \( I(.) \) is not continuous at any point in \( \{ \mu : I(\mu) < \infty \} \). This result will not be used in other sections.
Proposition 2.4. For any \( \nu \in S \) with \( I(\nu) < \infty \), there exists \( \nu^{(e)} \) such that \( d(\nu^{(e)}, \nu) \to 0 \) as \( \epsilon \to 0 \), and \( I(\nu^{(e)}) = \infty \).

Proof. Define \( \mu \in \mathbb{P}(\mathbb{N}_0) \) be defined by

\[
\mu_0 = \mu_1 = 0, \\
\mu_i = \frac{1}{Ci^2 \sqrt{\log i}}, \quad i \geq 2,
\]

where \( C := \sum_{i=2}^{\infty} \frac{i^{-2} \log^{-1/2}}{2} < \infty \), and let \( \nu^{(e)} := (1 - \epsilon) \nu + \epsilon \mu \). Then \( \mu \in S \), and so by convexity \( \nu^{(e)} \in S \). Also it is easy to check that \( \nu^{(e)} \xrightarrow{w} \nu \), and \( \nu^{(e)} \to \nu \), and so \( \nu^{(e)} \) converges to \( \nu \) in \( d(,..) \). However \( \nu^{(e)}(x \log x) = \mu(x \log x) = \infty \), and so \( I(\nu^{(e)}) = \infty \), completing the proof. \( \square \)

3. Proof of the main result

The proof strategy of Theorem 1.2 is carried out according to the following strategy:

Three auxiliary lemmas will be proved first, and then the theorem derived as a consequence of the lemmas.

The first of the three lemmas provides a crucial estimate, and helps in guessing the rate function.

Lemma 3.1. Let \( N(h) \) denote the number of simple graphs with given degree frequencies \( h = (h_0, \ldots, h_{n-1}) \). Then

(a) For any \( h \),

\[
N(h) \leq \frac{(2E)!}{E!2E^{n-1} \prod_{i=0}^{n-1} d_{th_i}} \times \frac{n!}{\prod_{i=0}^{n-1} h_i!}.
\]

(b) If \( h_i = 0 \) for all \( i > M \) with \( M < \infty \) then

\[
N(h) \geq C \frac{(2E)!}{E!2E^{n} \prod_{i=1}^{n} h_{i}^l} \times \frac{n!}{\prod_{i=0}^{n-1} h_i!}
\]

for some constant \( C := C(M) < \infty \).

Proof. Note that \( h \) determines the ordered degree sequence \( d_0 := (d_1(1) \leq d_2(2) \leq \cdots d_n(n)) \) uniquely. Thus if \( A(d_0) \) denote the number of graphs with degree sequence \( d_0 \), then the number of graphs with degree frequency \( h \) is \( A(d_0) \times \frac{n!}{\prod_{i=0}^{n-1} h_i!} \). It thus remains to estimate \( A(d_0) \).

(a) The upper bound follows trivially from the representation \( A(d_0) = P(d_0) \frac{(2E)!}{E!2E^{n} \prod_{j=1}^{n} d_{(j)}^l} \) with

\[
0 \leq P(d_0) \leq 1 \text{ (see [M])}, \quad \text{and noting that } \prod_{j=1}^{n} d_{(j)}^l = \prod_{i=0}^{n-1} d_{th_i},
\]

(b) Since the result holds when \( d_{(n)} = 0 \), w.l.o.g. assume \( E \geq d_{(n)} \geq 1 \). The assumption \( h_i = 0 \) for all \( i > M \) is equivalent to \( d_{(n)} \leq M \). Setting

\[
\lambda := \frac{1}{4E} \sum_{j=1}^{n} d_{(j)}(d_{(j)} - 1), \quad \hat{\Delta} := 2 + d_{(n)} + \frac{3}{2} d_{(n)}^2,
\]

3. Proof of the main result

The proof strategy of Theorem 1.2 is carried out according to the following strategy:

Three auxiliary lemmas will be proved first, and then the theorem derived as a consequence of the lemmas.

The first of the three lemmas provides a crucial estimate, and helps in guessing the rate function.

Lemma 3.1. Let \( N(h) \) denote the number of simple graphs with given degree frequencies \( h = (h_0, \ldots, h_{n-1}) \). Then

(a) For any \( h \),

\[
N(h) \leq \frac{(2E)!}{E!2E^{n-1} \prod_{i=0}^{n-1} d_{th_i}} \times \frac{n!}{\prod_{i=0}^{n-1} h_i!}.
\]

(b) If \( h_i = 0 \) for all \( i > M \) with \( M < \infty \) then

\[
N(h) \geq C \frac{(2E)!}{E!2E^{n} \prod_{i=1}^{n} h_{i}^l} \times \frac{n!}{\prod_{i=0}^{n-1} h_i!}
\]

for some constant \( C := C(M) < \infty \).

Proof. Note that \( h \) determines the ordered degree sequence \( d_0 := (d_1(1) \leq d_2(2) \leq \cdots d_n(n)) \) uniquely. Thus if \( A(d_0) \) denote the number of graphs with degree sequence \( d_0 \), then the number of graphs with degree frequency \( h \) is \( A(d_0) \times \frac{n!}{\prod_{i=0}^{n-1} h_i!} \). It thus remains to estimate \( A(d_0) \).

(a) The upper bound follows trivially from the representation \( A(d_0) = P(d_0) \frac{(2E)!}{E!2E^{n} \prod_{j=1}^{n} d_{(j)}^l} \) with

\[
0 \leq P(d_0) \leq 1 \text{ (see [M])}, \quad \text{and noting that } \prod_{j=1}^{n} d_{(j)}^l = \prod_{i=0}^{n-1} d_{th_i},
\]

(b) Since the result holds when \( d_{(n)} = 0 \), w.l.o.g. assume \( E \geq d_{(n)} \geq 1 \). The assumption \( h_i = 0 \) for all \( i > M \) is equivalent to \( d_{(n)} \leq M \). Setting

\[
\lambda := \frac{1}{4E} \sum_{j=1}^{n} d_{(j)}(d_{(j)} - 1), \quad \hat{\Delta} := 2 + d_{(n)} + \frac{3}{2} d_{(n)}^2,
\]
as in [M, Theorem 4.6] note that \( \hat{\Delta} \leq 2 + M + 2M^2 \), and

\[
\lambda \leq \frac{\sum_{j=1}^{n} d_j^2}{2 \sum_{j=1}^{n} d_j} \leq \frac{M}{2}.
\]

The conclusion then follows from [M, Theorem 4.6].

The second lemma shows that any \( \mu \in S \) is close to its truncation in the sense of the metric \( d(\ldots) \). This will be used during the proof of the lower bound in the third lemma. Part (b) of this Lemma uses the celebrated Erdős-Gallai criterion, which determines whether a given integer sequence is an ordered degree sequence corresponding to a simple graph. The following definition is required before stating the lemma.

**Definition 3.2.** For a degree frequency vector \( h \) let \( t_n(h) := (h/n, 0, 0, \ldots) \), i.e. \( t_n(h) \in S \) be an infinite tuple with \( h/n \) as its first \( n \) entries and appended by countably many 0’s. Also let \( H_n \subset S \) denote the set of all possible values of \( t_n(h) \) as \( G_n \) varies over the set of all simple graphs.

**Lemma 3.3.** (a) Given \( \nu \in S \) such that \( I(\nu) < \infty \), there exists a sequence \( \nu^{(k)} \in S \) satisfying the following conditions:

\[
I(\nu^{(k)}) \to I(\nu),
\]

\[
d(\nu^{(k)}, \nu) \to 0,
\]

\[
\nu_i^{(k)} = 0 \text{ for } i > k.
\]

(b) For any \( M \geq 2 \) and non negative vector \( (y_0, \ldots, y_M) \) such that \( y_0 > 0, \sum_{i=0}^{M} y_i = 1 \), there exists \( t_n(h) \in H_n \) such that

\[
|t_n(h)_i - y_i| \leq \frac{M}{n} \text{ for } 0 \leq i \leq M,
\]

\[
t_n(h) = 0 \text{ for } i > M.
\]

**Proof.** (a) Define \( \nu^{(k)} \in S \) by

\[
\nu_i^{(k)} := \frac{\nu_i}{\sum_{i=0}^{k} \nu_i}.
\]

Clearly \( \nu^{(k)} \) is well defined as soon as \( \sum_{i=0}^{k} \nu_i > 0 \), which is true for all large \( k \). Also \( \lim_{k \to \infty} \nu^{(k)}(i) = \nu_i \), and

\[
\nu^{(k)} = \frac{\sum_{i=1}^{k} \nu_i i}{\sum_{i=0}^{k} \nu_i} \to \frac{\sum_{i=1}^{\infty} \nu_i i}{\sum_{i=0}^{\infty} \nu_i} = \nu,
\]

\[
|t_n(h)_i - y_i| \leq \frac{M}{n} \text{ for } 0 \leq i \leq M,
\]

\[
t_n(h) = 0 \text{ for } i > M.
\]
and so $\nu^{(k)} \xrightarrow{d(k)} \nu$. Finally to check that $I(\nu^k) \to I(\nu)$ it suffices to check $D(\nu^{(k)}||p_\beta) \to D(\nu||p_\beta) < \infty$, which follows on noting that

$$D(\nu^{(k)}||p_\beta) = C_k \sum_{i=0}^{k} \nu_i \log \left( \frac{\nu_i}{p_{\beta_i}} \right) - C_k \log C_k$$

with $C_k := (\sum_{i=0}^k \nu_i)^{-1} \to 1$.

(b) If $y_0 = 1$ then set $h$ by $h_0 = n$ and $h_i = 0$ for $i > 0$. Thus w.l.o.g. assume $0 < y_0 < 1$. Define a candidate degree frequency $h$ as follows:

If $\sum_{i=1}^M \lfloor ny_i \rfloor$ is even, then set

$$h_i := \lfloor ny_i \rfloor \text{ for } 1 \leq i \leq M,$$
$$h_0 := n - \sum_{i=1}^M h_i,$$
$$h_i := 0 \text{ for } i > M.$$

If $\sum_{i=1}^M \lfloor ny_i \rfloor$ is odd, then set

$$h_i := \lfloor ny_i \rfloor \text{ for } 2 \leq i \leq M,$$
$$h_1 := \lfloor ny_1 \rfloor + 1,$$
$$h_0 := n - \sum_{i=1}^M h_i,$$
$$h_i := 0 \text{ for } i > M.$$

Note that

$$1 + \sum_{i=1}^M \lfloor ny_i \rfloor \leq 1 + n(1 - y_0) < n$$

for all large $n$, and so $h_0 > 0$. Also by construction $\sum_{i=1}^M ih_i$ is even, and

$$|t_n(h)_i - y_i| \leq \frac{1}{n} \text{ for } 1 \leq i \leq M, \quad |t_n(h)_0 - y_0| \leq \frac{M}{n}.$$

To complete the proof, it remains to check that the $h$ defined above is indeed is a valid degree frequency, i.e. the corresponding ordered degree sequence

$$\left( M, M, \cdots, M, M - 1, M - 1, \cdots, M - 1, \cdots, 0, 0, \cdots, 0 \right)$$
satisfies the Erdős-Gallai criterion, where \( i \) appears \( h_i \) times, for \( 0 \leq i \leq M \). With the ordered degree sequence \( (d(n) \geq d(n-1) \geq \cdots \geq d(1)) \), one needs to check that for all \( 1 \leq k \leq n \),
\[
\sum_{i=1}^{k} d(n+1-i) \leq k(k-1) + \sum_{i=k+1}^{n} \min(d(n+1-i), k).
\]
(3.1)

Note that the l.h.s. of (3.1) is bounded by \( kM \) and so (3.1) holds trivially for \( k > M \). Also for \( 1 \leq k \leq M \) the l.h.s. of (3.1) is bounded by \( M^2 \), whereas the r.h.s. is at least
\[
\sum_{i=1}^{M} [ny_i] - M,
\]
which goes to infinity as \( n \) grows, since \( y_i > 0 \) for some \( i \neq 0 \). Thus (3.1) holds for all \( k \leq M \) for all \( n \) large enough, and so the proof is complete.

\[\square\]

The third and final lemma uses Lemmas 3.1 and Lemma 3.3 to formalize the claim that it suffices to assume that maximum of the degrees are "not too large". This lemma will be used to derive the l.d.p.

Lemma 3.4. For any set \( A \subseteq S \),
\[
\liminf_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \max_{h : t_n(h) \in A \setminus \{h_j = 0, j > n\delta\}} \mathcal{N}(h) \leq - \inf_{\mu \in A} I(\mu),
\]
(3.2)

\[
\limsup_{M \to \infty} \liminf_{n \to \infty} \frac{1}{n} \log \max_{h : t_n(h) \in A \setminus \{h_j = 0, j > M\}} \mathcal{N}(h) \geq - \inf_{\mu \in A} I(\mu),
\]
(3.3)

where
\[
\mathcal{N}(h) := \frac{(2E)!}{E!2^E} \prod_{i=0}^{n-1} h_i! \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{\left( \frac{n}{2} \right)^2 - E}.
\]

Proof. For proving (3.2), first note that Stirling’s approximation gives
\[
| \log n! - n \log n + n | = 0 \text{ if } n = 0,
\]
\[
= 1 \text{ if } n = 1
\]
\[
\leq C_1 \log n \text{ for } n \geq 2,
\]
for some \( C_1 < \infty \). Using this along with the assumption \( h_j = 0 \) for \( j > n\delta \) gives
\[
| \log \mathcal{N}(h) + I(t_n(h)) | \leq C_2 \left( \frac{1}{n} \log n + \delta + \frac{1}{n} \max_{h : h_j = 0, j < n\delta} \sum_{i=1}^{n} \log(h_j \vee 1) \right)
\]
for some \( C_2 := C_2(\beta) < \infty \). By A.M-G.M. inequality the maximum for the last term on the r.h.s. occurs when all the non zero \( h_j \)'s are equal, giving the bound \( C_2(\log n/n + \delta + \delta \log(1/\delta)) \), and so
\[
\liminf_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathcal{N}(h) \leq - \inf_{\mu \in A} I(\mu),
\]
where the last step uses the fact that the infimum is taken over a larger collection of \( \mu \)'s.
Turning to prove (3.3), fix $\epsilon > 0$ arbitrary. By a similar argument as above the assumption $h_j = 0$ for $j > M$ gives

$$|\log \overline{N}(h) + I(t_n(h))| \leq \frac{1}{n}(C_2 + M) \log n,$$

and so

$$\limsup_{M \to \infty} \liminf_{n \to \infty} \frac{1}{n} \log \max_{h: t_n(h) \in A} \overline{N}(h) \geq -\liminf_{M \to \infty} \limsup_{n \to \infty} \min_{\mu \in H_n \cap A: \{p_j = 0, j > M\}} I(\mu),$$

and so it suffices to prove that given $\epsilon > 0, M_0 < \infty$ arbitrary, there exists $M > M_0$ such that

$$\limsup_{n \to \infty} \min_{\mu \in H_n \cap A: \{p_j = 0, j > M\}} I(\mu) \leq \inf_{\mu \in A^0} I(\mu) + \epsilon \quad (3.4)$$

If $\inf_{\mu \in A^0} I(\mu) = \infty$ then there is nothing to prove. So w.l.o.g. fix $\mu \in A^0$ such that $I(\mu) < \infty$. Since $A^0$ is open, there exists $\nu \in A^0$ such that $I(\nu) \leq I(\mu) + 2\epsilon$, and $\nu_0 > 0$. For this $\nu$, let $\nu^{(k)}$ be the corresponding sequence as constructed in part (a) of Lemma 3.3. Thus by Lemma 3.3 and openness of $A^0$ there exists $M > M_0$ such that $I(\nu^{(M)}) \leq I(\nu) + \epsilon$, and $\nu^{(M)} \in A^0$. Since $\nu^{(M)}_0 > 0$, by part (b) of Lemma 3.3 there exists $\sigma^{(n)} \in S \cap \mathbb{H}_n$ such that $\sigma^{(n)}_i \to \nu^{(M)}_i$ for $0 \leq i \leq M$, and $\sigma^{(n)}_i = 0$ for $i > M$. But this readily gives

$$d(\sigma^{(n)}, \nu^{(M)}) \to 0, \quad I(\sigma^{(n)}) \to I(\nu^{(M)}).$$

Finally since $\sigma^{(n)} \to \nu^{(M)} \in A^0$ open, for all $n$ large enough $\sigma^{(n)} \in A^0 \cap \mathbb{H}_n$, and so

$$\limsup_{n \to \infty} \min_{\mu \in H_n \cap A: \{p_j = 0, j > M\}} I(\mu) \leq \limsup_{n \to \infty} I(\sigma^{(n)}) = I(\nu^{(M)}) \leq I(\nu) + \epsilon \leq I(\mu) + 2\epsilon,$$

from which (3.4) follows on taking infimum over $\mu \in A^0$.

\[\Box\]

**Proof of Theorem 1.2.** Fix $\delta \in (0, \log 2)$ and note that $d_i$ has a Binomial distribution with parameters $n-1$ and $\beta/n$, and so

$$\mathbb{P}_n(\max_{1 \leq i \leq n} d_i > n\delta) \leq n\mathbb{P}_n(d_1 > n\delta) \leq n \sum_{r > n\delta} \binom{n-1}{r} \left(\frac{\beta}{n-1}\right)^r \leq n(2\beta^\delta)^n n^{-n\delta}$$

for all large $n$. Also since $E$ has a Binomial distribution with parameters $\binom{n}{2}$ and $\beta/n$, and so by Hoeffding’s inequality,

$$\mathbb{P}_n(E > nK) \leq e^{-2n/\beta^2},$$

where $K = K_\delta := (\beta/2) + (1/\delta)$. Since the ordered degrees $d_0 := (d_{(1)} \leq d_{(2)} \leq \cdots d_{(n)})$ is a partition of $2E$, and the number of partitions of an integer $n$ is bounded by $e^{3\sqrt{n}}$ for all large $n$ (for a proof of this classical result see [HR] or [E]), the number of possible ordered degree sequences satisfying $E \leq Kn$ is bounded by $Kn e^{3\sqrt{2Kn}}$ for all large $n$. Noting that the ordered degree sequences $d_0$ are in 1-1 correspondence with the degree frequency vector $h$, with $N(h)$ denoting
the number of graphs in \( \mathbb{G}_n \) with degree frequency vector \( \mathbf{h} \), for any set \( A \subseteq \mathcal{S} \),

\[
\mathbb{P}_n(\mu_n \in A) \\
\leq \mathbb{P}_n(\max_{1 \leq i \leq n} d_i > n\delta) + \mathbb{P}_n(E > Kn) + \mathbb{P}_n(t_n(\mathbf{h}) \in A, E \leq Kn, h_j = 0, j > n\delta)
\]

\[
\leq n(2\beta^n) n^{-\delta} + e^{-2n/\beta^2} + \sum_{t_n(\mathbf{h}) \in A, E \leq Kn, h_j = 0, j > n\delta} N(\mathbf{h}) \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{(\frac{n}{2})-E}
\]

\[
\leq n(2\beta^n) n^{-\delta} + e^{-2n/\beta^2} + \sum_{t_n(\mathbf{h}) \in A, E \leq Kn, h_j = 0, j > n\delta} \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{(\frac{n}{2})-E} \left( \frac{2E}{E!2^E} \prod_{i=0}^{n-1} \frac{i!h_i}{i!h_i!} \right)
\]

\[
\leq n(2\beta^n) n^{-\delta} + e^{-2n/\beta^2} + Kne^{3\sqrt{2Kn}} \max_{t_n(\mathbf{h}) \in A, h_j = 0, j > n\delta} \overline{N}(\mathbf{h})
\]

where the last but one step requires part (a) of Lemma 3.1, and \( \overline{N}(\mathbf{h}) \) is as defined in Lemma 3.4. Taking log, dividing by \( n \) and taking \( n \to \infty \) gives

\[
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n(\mu_n \in A) \leq \limsup_{n \to \infty} \frac{1}{n} \log \max_{h: t_n(\mathbf{h}) \in A, \{h_j = 0, j > n\delta\}} \overline{N}(\mathbf{h}).
\]

Letting \( \delta \to 0 \) and using (3.2), (1.1) follows.

Proceeding to prove the lower bound fix \( M < \infty \) and note that for any \( A \),

\[
\mathbb{P}_n(\mu^{(n)} \in A) \geq \mathbb{P}_n(\mu_n \in A, h_i = 0, i > M)
\]

\[
= \sum_{t_n(\mathbf{h}) \in A, t_i = 0, i > M} N(\mathbf{h}) \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{(\frac{n}{2})-E}
\]

\[
\geq C \max_{t_n(\mathbf{h}) \in A, t_i = 0, i > M} \overline{N}(\mathbf{h}),
\]

with \( C = C(M) < \infty \), where the last step requires part (b) of Lemma 3.1. As before taking log, dividing by \( n \) and taking limits gives

\[
\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n(\mu_n \in A) \geq \liminf_{n \to \infty} \frac{1}{n} \log \max_{h: t_n(\mathbf{h}) \in A, \{h_j = 0, j > M\}} \overline{N}(\mathbf{h}).
\]

On letting \( M \to \infty \) and using (3.3), (1.2) follows. \( \square \)

**Corollary 3.5.** If \( A \) is open,

\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n(\mu^{(n)} \in A) = - \inf_{\mu \in A} I(\mu).
\]

**Proof.** Follows trivially from (1.1) and (1.2) of Theorem 1.2. \( \square \)

4. **Proofs of Theorem 1.5 and Theorem 1.8**

The first lemma of this section gives an upper bound which will be used in proving both theorems.

**Lemma 4.1.** Let \( f \in F \), and \( U \subseteq \mathcal{S} \) be open. Then

\[
\limsup_{n \to \infty} \mathbb{E}_{\mathbb{P}_n} e^{n\mu^{(n)}} 1_{\mu^{(n)} \in U^c} \leq \sup_{\mu \in U^c} \{\mu(f) - I(\mu)\}.
\]
Proof. Since \( f \in F \), there exists \( C < \infty \) such that \( f(i) \leq Ci \) for all \( i \in \mathbb{N}_0 \). To begin, first note that
\[
E_P e^{n\mu(n)} 1_{U^c} = E_P e^{nT(\mu(n))},
\]
where \( T : S \mapsto [-\infty, \infty) \) defined by
\[
T(\mu) := \mu(f) \text{ if } \mu \in U^c,
\]
\[= \infty \text{ otherwise}
\]
is upper semi continuous, as \( \mu \mapsto \mu(f) \) is upper semi continuous by part (a) of Lemma 2.2. Thus an application of [DZ, Lemma 4.3.6] along with Theorem 1.2 gives
\[
\limsup_{n \to \infty} E_P e^{nT(\mu(n))} \leq \sup_{\mu \in S} \{ T(\mu) - I(\mu) \} = \sup_{\mu \in U^c} \{ \mu(f) - I(\mu) \},
\]
which is the desired conclusion. Condition (4.3.3) of [DZ, Page 137] is verified below with \( \gamma = 2 \) to check that [DZ, Lemma 4.3.6] is indeed applicable:

\[
\limsup_{n \to \infty} \frac{1}{n} \log E_P e^{2 \sum_{i=0}^{n-1} f(i) h_i} \leq \limsup_{n \to \infty} \frac{1}{n} \log E_P e^{2C \sum_{i=0}^{n-1} i h_i} = \limsup_{n \to \infty} \frac{1}{n} \log E_P e^{4CE} = \frac{\beta}{2} (e^{4C} - 1) < \infty.
\]

\[\square\]

Remark 4.2. Note that even though [DZ, Lemma 4.3.6] requires \( T \) to be finitely defined everywhere on \( S \), the proof goes through as long as \( T \in [-\infty, \infty) \), which holds here as \( \mu(f) \leq C' \mu < \infty \).

Proof of Theorem 1.5. Note that
\[
Z_n(\gamma) = \sum_{G_n \in G_n} \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{\binom{n}{2} - E} e^{n\gamma \mu(n)}(g) = E_P e^{n\gamma \mu(n)}(g),
\]
where \( g(i) = \binom{i}{k} \) for \( i \in \mathbb{N}_0 \).

(a) For part (a), setting \( r_n := \lfloor n^{1/k} M \rfloor \) for any \( M > 0 \), note that \( r_n \leq n - 1 \) for all large \( n \). Also recall that
\[
n\mu(n)(g) = \sum_{i=0}^{n-1} h_i g(i) = \sum_{j=1}^{n} \binom{d_j}{k},
\]
and so
\[
E_P e^{n\gamma \mu(n)}(g) \geq e^{\gamma \binom{r_n}{k}} P_n(d_1 = r_n) = e^{\gamma \binom{r_n}{k}} \left( \frac{n - 1}{r_n} \right) \left( \frac{\beta}{n} \right)^{r_n} \left( 1 - \frac{\beta}{n} \right)^{\binom{n}{2} - r_n}
\]
(since \( d_1 \), which is the degree of vertex 1, has a Binomial distribution with parameters \( (n - 1, \beta/n) \)) and so
\[
\frac{1}{n} \log \liminf_{n \to \infty} E_P e^{n\gamma \mu(n)}(g) \geq \gamma \frac{M^k}{k!} - \frac{\beta}{2}.
\]
Since this holds for all \( M > 0 \), the conclusion follows on letting \( M \to \infty \).

(b) For part (b), since \( \gamma < 0 \), the function \( \gamma g \in F \) (since \( \gamma g \leq 0 \)), and so by Lemma 4.1 with \( U = \phi \),
\[
\limsup_{n \to \infty} \frac{1}{n} \log E_P e^{n\mu(n)} \leq \sup_{\mu \in S} \{ \gamma \mu(g) - I(\mu) \}.
\]
For the lower bound, first note that for any \( M \in \mathbb{N} \) the function \( G_n \mapsto \sum_{i=0}^{M} h_i g(i) \) is non decreasing in the sense of graphs, and so

\[
\mathbb{E}_n e^{n\gamma \mu^{(n)}(g)} e^{n\gamma \mu^{(n)}(g)} 1_{\max_1 \leq i \leq d_i \leq M} = \mathbb{E}_n e^{n\gamma \sum_{i=0}^{M} \mu_i^{(n)} g(i)} 1_{\max_1 \leq i \leq d_i \leq M} \geq \mathbb{E}_n e^{n\gamma \sum_{i=0}^{M} \mu_i^{(n)} g(i)} \mathbb{P}(d_1 \leq M)^n,
\]

where the last step follows by the FKG inequality, on noting that \( \gamma < 0 \), and the function \( G_n \mapsto \sum_{i=0}^{M} \mu_i^{(n)} g(i) \) is non decreasing on the space \( \mathcal{G}_n \).

Since \( \mu \mapsto \gamma \sum_{i=0}^{M} \mu_i g(i) \) is bounded and continuous with respect to \( d(.,.) \), it follows by Theorem 1.2 and an application of Varadhan’s lemma that

\[
\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E}_n e^{n\gamma \mu^{(n)}(g)} \geq \sup_{\mu \in \mathcal{S}} \left\{ \sum_{i=0}^{M} \mu_i g(i) - I(\mu) \right\} + \log p_\beta[0, M],
\]

where \( p_\beta[0, M] \) is the probability that a Poisson random variable with parameter \( \beta \) is at most \( M \). Finally note that \( \gamma \mu(g) \leq \sum_{i=0}^{M} \mu_i g(i) \), and so the r.h.s. of the above inequality is bounded below by \( \sup_{\mu \in \mathcal{S}} \{ \mu(g) - I(\mu) \} + \log p_\beta[0, M] \). The lower bound follows on letting \( M \to \infty \) and noting that \( p_\beta[0, M] \to 1 \).

(c) By part (b) of Lemma 2.2, the supremum in part (b) equals \(-J(\gamma g)\), and is attained on a finite set \( A_{\gamma g} \subset \Omega_{\gamma g} \). Finally by Lemma 4.1 for any open \( U \),

\[
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{Q}_{n, \gamma g}(U^c) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}_n e^{n\gamma \mu^{(n)}(g)} 1_{\mu^{(n)} \in U^c} - \liminf_{n \to \infty} \frac{1}{n} \log Z_n(\gamma g) \leq \sup_{\mu \in U^c} \{ \gamma \mu(g) - I(\mu) \} - \sup_{\mu \in \mathcal{S}} \{ \gamma \mu(g) - I(\mu) \}.
\]

The last quantity above is negative by part (c) of Lemma 2.2, and so the conclusion follows.

\(\square\)

**Proof of Theorem 1.8.** (a) Since \( \limsup_{i \to \infty} \frac{|f(i)|}{i} < \infty \), there exists \( C < \infty \) such that \( |f(i)| \leq Ci \) for all \( i \in \mathbb{N}_0 \). This readily gives that the function \( \mu \mapsto \mu(f) \) is continuous with respect to \( d(.,.) \). To see this, note that if \( \nu^{(k)} \) converges to \( \nu \) in \( d(.,.) \), then

\[
\left| \sum_{i=1}^{\infty} f(i) \nu_i^{(k)} - \sum_{i=1}^{\infty} f(i) \nu_i^{(k)} \right| \leq C \sum_{i=1}^{\infty} |\nu_i^{(k)} - \nu_i^{(k)}| = C d(\nu^{(k)}, \nu)
\]

which goes to 0 as \( k \) goes to \( \infty \).
Since
\[ Z_n(f, \beta) = \mathbb{E}_{P_n} e^{\gamma \sum_{i=0}^{n-1} f(i) h_i} = \mathbb{E}_{P_n} e^{n \gamma \psi(n)}(f), \]
an application of Varadhan’s Lemma [DZ, Theorem 4.3.1] along with Theorem 1.2 proves part (a). Note that since \( f \in F \), condition (4.3.3.) of [DZ, Page 137] holds, as verified during the proof of Lemma 4.1, and so Varadhan’s lemma is applicable.

(b) The proof of part (b) follows exactly the same lines as the proof of part (c) of Theorem 1.5, and is not repeated here.

□

5. A PARTICULAR EXAMPLE: THE SPARSE PENALTY MODEL

This section uses the theory developed in the previous sections to analyze a particular ERGM on sparse graphs. The sufficient statistic for this model is the number of edges, and the number of isolated vertices in the graph. The isolated vertices term can also be viewed as a penalty term which prefers or dislikes isolated vertices, depending on the sign of the associated parameter \( \gamma \). This model is probably the most simple model that can be handled by the theory above (see Theorem 1.8), and is of some interest in the social science community.

The edge parameter in this model has been scaled with \( n \) to force this model to put most of its mass on sparse graphs. Denoting \( h_0 \) to be the number of empty vertices as before, the probability mass function is proportional to
\[
\left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{\binom{n}{2}} - E e^{-\gamma h_0}.
\]

This model has two parameters, the edge parameter \( \beta > 0 \) and the sparse penalty parameter \( \gamma \in \mathbb{R} \). The negative sign in front of \( \gamma \) is taken for convenience in applying the theoretical results, and does not lose generality in anyway.

To connect with the theory, note that this model is same as \( Q_{n, \gamma \psi}(\cdot) \) (see Definition 1.3) with \( \psi(i) := 1 - 1_{i=0} \) for \( i \in \mathbb{N}_0 = \mathbb{N} \cup \{0\} \). Using the notation of Definition 1.3, the model can be written as
\[
Q_{n, \gamma \psi}(G_n) = \frac{1}{Z_{n, \gamma \psi}} \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{\binom{n}{2}} - E \sum_{i=1}^{n-1} \psi(i) h_i e^{\gamma \psi(i)} = \frac{e^{n \gamma}}{Z_{n, \psi}} \left( \frac{\beta}{n} \right)^E \left( 1 - \frac{\beta}{n} \right)^{\binom{n}{2}} - E e^{-\gamma h_0}.
\]

Thus to estimate the normalizing constant, it suffices to estimate \( Z_{n, \psi} \).

Analogous to this model, for every \( \theta > 0 \) consider the following probability distribution on non negative integers as follows:
\[
\sigma_{\theta, \gamma \psi}(i) = \frac{1}{i!} e^{-C(\theta, \gamma \psi)} \theta^i e^{\gamma \psi(i)},
\]
where \( e^{C(\theta, \gamma \psi)} \) is the appropriate normalizing constant. Also denote by \( m(\theta) \) the mean of \( \sigma_{\theta, \gamma \psi} \). For this particular choice of \( \psi \), a direct calculation reveals that
\[
e^{C(\theta, \gamma \psi)} = 1 + e^{\gamma (e^\theta - 1)}, \quad m(\theta) = \frac{\theta e^{\gamma + \theta}}{1 + e^{\gamma (e^\theta - 1)}}.
\]
Since \(\psi(i)/i \to 0\), part (a) of Theorem 1.8 gives an asymptotic estimate of \(Z_{n,\psi}\) as follows:

\[
\lim_{n \to \infty} \frac{1}{n} \log Z_{n,\gamma\psi} = -\inf_{\theta \geq 0} H_{\gamma,\beta}(\theta),
\]

where

\[
H_{\gamma,\beta}(\theta) := \left\{ m(\theta) \log \theta - C(\theta, \gamma\psi) - \frac{m(\theta)}{2} \log(m(\theta)\beta) + \frac{m(\theta) + \beta}{2} \right\}.
\]

Thus the problem of estimating the normalizing constant reduces to a one dimensional optimization problem, and can be done on a computer by a one dimensional grid search.

To understand how the degree distribution looks like under this model, one needs to understand the set of minimizers of \(H_{\gamma,\beta}(\theta)\). Differentiating with respect to \(\theta\) and equating it to 0 gives

\[
m'(\theta) \log \left( \frac{\theta}{\sqrt{m(\theta)\beta}} \right) = 0,
\]

which gives that either \(m'(\theta) = 0\) or \(\theta = \sqrt{m(\theta)\beta}\). Also it is easy to check that \(m'(\theta) = \theta C(\theta, \gamma\psi) > 0\), and so \(\theta = \sqrt{m(\theta)\beta}\).

(On a more rigorous note, \(m'(\theta) = 0\) is possible only if \(\theta = 0\), as \(C(\theta, f) > 0\) for \(\theta > 0\). But \(m(\theta) \approx \theta e^\gamma\) for \(\theta \approx 0\), and so the function \(\theta \mapsto \log \left( \frac{\theta}{\sqrt{m(\theta)\beta}} \right)\) diverges to \(-\infty\) as \(\theta\) approaches 0. Since \(m'(\theta) \geq 0\), it follows that \(H'_{\gamma,\beta}(\theta) < 0\) for \(\theta \approx 0\), and so 0 cannot be a local minima of \(H_{\gamma,\beta}(\theta)\).

Substituting the expression of \(m(\theta)\) in the relation \(\theta = \sqrt{m(\theta)\beta}\) gives

\[
\theta = \frac{\beta e^{\gamma + \theta}}{1 + e^{\gamma}(e^{\theta} - 1)} = h_{a,b}(\theta)
\]

where \(a = \beta, b = e^\gamma\), and

\[
h_{a,b}(x) := \frac{abe^x}{1 + (e^x - 1)^2}.
\]

The following simple Lemma 5.1 analyzes the roots of the function \(x = h_{a,b}(x)\) for \(a > 0, b > 0\).

**Lemma 5.1.** For \(a, b > 0\) and \(b \neq 1\) consider the function \(h_{a,b}(x) = \frac{abe^x}{1 + (e^x - 1)}\) for \(x > 0\).

(a) The equation \(h_{a,b}(x) = x\) has either one or three roots (counting multiplicity).

(b) If either \(b > 1\) or \(a < 4\) then the equation \(h_{a,b}(x) = x\) has exactly one root.

**Proof.** (a) Since \(h_{a,b}(0) = ab > 0\) and \(\lim_{x \to \infty} h_{a,b}(x) = a < \infty\), the given equation has at least one root, and the number of roots are odd. Differentiation gives

\[
h'_{a,b}(x) - 1 = \frac{ab(1 - b)e^x}{[1 + (e^x - 1)]^2} - 1
\]

Thus \(h'_{a,b}(x) = 1\) is a quadratic equation in \(e^x\), and so can have at most two real roots and so by Rolle’s theorem the equation \(h_{a,b}(x) = x\) has at most three real roots, thus concluding the proof of part (a).

(b) Since \(h'_{a,b}(x) - 1 < 0\) if \(b > 1\), \(h_{a,b}(x) - x\) is monotone decreasing and so has exactly one root. Thus w.l.o.g. assume \(b < 1\) and \(a < 4\), and note that

\[
h'_{a,b}(x) = a \frac{be^x}{1 - b + be^x} \frac{1 - b}{1 - b + be^x} \leq \frac{a}{4} < 1,
\]
and so $h_{a,b}(x) - x$ is again monotone decreasing, thus concluding the proof.

The case $b = 1$ is not considered, as that corresponds to the Erdős Renyi model with parameter $\beta/n$ and is well understood. From Lemma 5.1, the equation $\theta = h_{a,b}(\theta)$ has either one or three roots. Thus there are two sub cases:
(i) If \( h_{a,b}(\theta) = \theta \) has exactly one root \( \theta_0 \), then \( \theta_0 \) is the unique local and global minima of \( H_{\gamma,\beta}(\cdot) \). By part (b) of Theorem 1.8, the degree distribution converges to \( \sigma_{\theta_0,\gamma\psi} \). By part (b) of Lemma 5.1 this happens if either \( \gamma > 0 \) or \( \beta < 4 \).

An example of this case is Figure 1, where the plot of \( H_{\beta,\gamma}(\cdot) \) is given for \( \beta = 1.2 \) and \( e^\gamma = 0.5 \).

\[\text{Figure 1. Unique local minima}\]

(ii) If \( h_{a,b}(\theta) = \theta \) has exactly three roots \( \theta_1 < \theta_2 < \theta_3 \), then \( \theta_1 \) and \( \theta_3 \) are local minima of the \( H_{\gamma,\beta}(\cdot) \), and \( \theta_2 \) is a local maxima. This gives two further sub cases.
(a) If out of $(\theta_1, \theta_3)$ only one is a local minima and the other is a global minima, then denoting the global minimizer by $\theta_0$, the degree distribution converges to $\sigma_{\theta_0, \gamma\psi}$.

An example of this is Figure 2, where the plot of $H_{\beta, \gamma}(.)$ is given for $\beta = 6.5$ and $e^{\gamma} = 0.04$.

---

**Figure 2.** Non unique local minima, Unique global minima

(b) If both $(\theta_1, \theta_3)$ are global minimizers of $H_{a,b}(\theta)$, then the degree distribution converges to a mixture of $\sigma_{\theta_1, \gamma\psi}$ and $\sigma_{\theta_3, \gamma\psi}$.

An approximate example of this is Figure 3, where the plot of $H_{\beta, \gamma}(.)$ is given for $\beta = 5.89$ and $e^{\gamma} = 0.05$. 
Remark 5.2. Thus even in this simple model there is a "phase transition", namely for some values of the parameters \((\beta, \gamma)\) the degree distribution has a unique limit, whereas for other parameter values the limiting degree distribution is a mixture distribution. Also, even though phase transition is established, the exact phase transition boundary for this problem (i.e. the parameter values for which the limit is a mixture distribution) has not been characterized in this paper, and is a scope of possible future research.

The exact same analysis works for any model of the form \(Q_{n, \gamma \psi}\) for any function \(\psi\) satisfying the conditions of Theorem 1.8. One possible difference is that for a general \(\psi\) the functions \(C(\theta, \gamma \psi)\) and \(m(\theta)\) might not be computable in closed form, and need to be estimated numerically as well.

6. ACKNOWLEDGEMENT

I would like to thank my advisor Prof. Persi Diaconis for his inspirational guidance throughout my Ph.D, and Prof. Amir Dembo for helpful discussions on this paper. This paper also benefitted from some helpful discussions during an AIM conference on ERGM held in June 2013, for which I would like to thank all participants of the conference.
References

[BC] Bordenave, C., and Caputo, P., Large deviations of empirical neighborhood distribution in sparse random graphs. arXiv:1308.5725

[Cs] Csiszar, I., I-Divergence Geometry of Probability Distributions and Minimization Problems. The Annals of Probability, 3 (1): 146-158. (1975)

[CD] Chatterjee, S. and Diaconis, P., Estimating and understanding exponential random graph models. To appear: The Annals Statistics.

[CV] Chatterjee, S. and Varadhan, S. R. S., The large deviation principle for the Erdős-Renyi random graph. European Journal of Combinatorics (special issue on Homomorphisms and Limits), 32, 7, 1000-1017. (2011)

[DM] Doku-Amponsah, K., and Mörters, P., Large deviation principles for empirical measures of colored random graphs. Annals of Applied Probability, 20(6):1989-2021. (2010)

[DZ] Dembo, A. and Zeitouni, O., Large deviations techniques and applications. second edition. Springer, Application of Mathematics, vol. 38. (1998)

[HH] Hunter, D.R. and Handcock, M.S., Inference in Curved Exponential Family Models for Networks. Journal of Computational and Graphical Statistics;15(3):565583. (2006)

[E] Erdős, P., On an elementary proof of some asymptotic formulas in the theory of partitions. The Annals of Mathematics. 43 (3): 437-450. (1942)

[HR] Hardy, G. H. and Ramanujan, S., Asymptotic Formulae in Combinatory Analysis. Proceedings of the London Mathematical Society. s2-17 (1): 75-115. (1918)

[M] Mckay, B. D., Asymptotics for symmetric 0 − 1 matrices with prescribed row sums. Ars Combinatoria, 19A 15-25. (1985)

[MHH] Morris, M., Hunter, D., and Handcock, M., Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. Journal of Statistical Software 42(i04). (2008)

[SPRH] Snijders, T.A.B., Pattison, P, Robins, G.L., and Handcock, M.S., New Specifications for Exponential Random Graph Models. Sociological Methodology;36:99153 (2006)

Department of Statistics, Stanford University
Sequoia Hall, 390 Serra Mall, Stanford, California 94305