Elect. Comm. in Probab. **14** (2009), 290–301

ELECTRONIC COMMUNICATIONS in PROBABILITY

LARGE DEVIATIONS IN RANDOMLY COLOURED RANDOM GRAPHS

J.D. BIGGINS
Dept of Probability and Statistics, The University of Sheffield, Sheffield, S3 7RH, U.K.
email: `j.biggins@sheffield.ac.uk`

D.B. PENMAN
Dept of Mathematics, University of Essex, Colchester C04 3SQ, U.K.
email: `dbpenman@essex.ac.uk`

Submitted December 12, 2008, *accepted in final form* March 30, 2009

AMS 2000 Subject classification: 05C80 60F10
Keywords: Large deviations, mixture, rate function, random graphs

Abstract

Models of random graphs are considered where the presence or absence of an edge depends on the random types (colours) of its vertices, so that whether or not edges are present can be dependent. The principal objective is to study large deviations in the number of edges. These graphs provide a natural example with two different non-degenerate large deviation regimes, one arising from large deviations in the colourings followed by typical edge placement and the other from large deviation in edge placement. A secondary objective is to illustrate the use of a general result on large deviations for mixtures.

1 Introduction

This paper considers, in the first instance, the following model for random graphs. Each vertex is independently assigned one of a finite number of colours, drawn from a set Σ, and then the presence of an edge is determined independently with a probability that depends on the colours of its two vertices. Thus there are two sources of variation, the generation of the colours and the subsequent generation of the edges. In such a model, the colouring induces some correlations in whether or not edges arise. The colouring itself is regarded as invisible: it is the characteristics of the resulting random graph which are the focus of attention.

These models with correlation structure were introduced in [10] and the motivation for the construction and areas of application for the model are discussed in [4]: however, familiarity with these two treatments is neither assumed nor needed here. This mechanism for producing correlated graphs has been considered also in [5] and [9], both of which refer to an earlier version of this paper. One of the referees reminded us that the model is closely related the model for inhomogeneous random graphs introduced, and examined closely, in [3] — it would fall within that framework if, when there were n vertices, n times the probability of an edge depended only on its vertices' colours.

Throughout, graphs will be finite and undirected, without loops or multiple edges. Given n, the number of vertices, and $k = |\Sigma|$, the finite number of colours, each vertex is coloured independently, receiving colour i with probability $s_i > 0$. The probability that an edge between a vertex of colour i and one of colour j arises is p_{ij}; thus $P = (p_{ij})$ is symmetric. We will consider the family of such models as n varies, with k, $\mathbf{s}^T = (s_1, s_2, \ldots, s_k)$ and P being fixed. Let α be the overall probability that an edge arises, so that $\alpha = \mathbf{s}^T P \mathbf{s}$. Note that if every entry in P is equal to α this framework will be equivalent to the classical model of random graphs $G(n,\alpha)$ (see [2]), once the colours are ignored. We use $\mathbb{P}$ and $\mathbb{E}$ for probability and expectation associated with these models. Let $\mathscr{E}$ be the number of edges present and let $\mathfrak{c}_n = n(n-1)/2$ be the number of potential edges. Then $\mathbb{E}\mathscr{E} = \mathfrak{c}_n\alpha$. Let $\mathbb{P}^n$ be the probability distribution of $\mathscr{E}/\mathfrak{c}_n$. The probability of a large deviation in the number of edges $\mathscr{E}$ is considered: $\mathbb{P}(\mathscr{E} > \mathfrak{c}_n\beta)$ $(= \mathbb{P}^n\{(\beta,\infty)\})$ for $\beta > \alpha$, for example. Large deviation theorems show that this probability is, very roughly, $e^{-I(\beta)a(n)}$ for suitable I and $a(n)$, and when such a result holds, ignoring some details of formulation supplied later, a large deviation principle (LDP) holds with constants $(a(n))$ and rate I. Here, qualitatively, two regimes may be anticipated. When some distribution of colours can produce an expected number of edges at least as large as $\mathfrak{c}_n\beta$ it will be enough to consider the probability of getting extreme colourings of this kind together with the typical number of edges for the colouring. Since this is an unusual outcome from the n independent random colourings it is plausible, based on standard large deviation theory, that $a(n) = n$ in this case. When β cannot be obtained in this way, it will be necessary for the number of edges (which arise from $\mathfrak{c}_n$ independent variables given the colouring) to be exceptional too, and then it is plausible that $a(n) = \mathfrak{c}_n$. The main result here confirms these speculations and specifies the function I in the two regimes.

Before the main theorem can be stated, notation to describe I is needed. The Fenchel dual of $\phi : \mathbb{R} \to \mathbb{R}$ is the convex function

$$\phi^*(y) = \sup_\theta(\theta y - \phi(\theta)).$$

Let $A(\theta)_{ij} = \log(1 - p_{ij} + p_{ij}e^\theta)$ and let Δ be the set of probability distributions on Σ, the set of colours, where elements of Δ are viewed as k-vectors with non-negative components adding to 1. Now let

$$\phi_{\mathbf{x}}(\theta) = \mathbf{x}^T A(\theta)\mathbf{x}, \quad \mathbf{x} \in \Delta \tag{1.1}$$

and

$$\psi(\mathbf{x}) = \begin{cases} \sum_i x_i \log(x_i/s_i) & \text{for } \mathbf{x} \in \Delta \\ \infty & \text{otherwise} \end{cases}. \tag{1.2}$$

Theorem 1.1. *Suppose vertices are coloured independently from k colours with probabilities $(s_1, s_2, \ldots, s_k)$.*

(i) $\mathbb{P}^n$ satisfies an LDP with constants (n) and rate $\Psi(y) = \inf\{\psi(\mathbf{x}) : \mathbf{x}^T P\mathbf{x} = y\}$.

(ii) $\mathbb{P}^n$ satisfies an LDP with constants $(\mathfrak{c}_n)$ and rate $\Phi(y) = \inf\{\phi^*_{\mathbf{x}}(y) : \mathbf{x} \in \Delta\}$.

Note that (i) and (ii) both hold for all y. This does not contradict the idea that there are two distinct regimes provided Φ is zero when Ψ is finite and non-zero. This is the relevance of the next result, which needs a little more notation. Define:

$$m = \inf\{\mathbf{x}^T P\mathbf{x} : \mathbf{x} \in \Delta\} \text{ and } M = \sup\{\mathbf{x}^T P\mathbf{x} : \mathbf{x} \in \Delta\};$$

$$l = \inf\{\mathbf{x}^T I(p_{ij} = 1)\mathbf{x} : \mathbf{x} \in \Delta\} \text{ and } L = \sup\{\mathbf{x}^T I(p_{ij} > 0)\mathbf{x} : \mathbf{x} \in \Delta\}.$$

These are attained because Δ is compact, and $l \le m \le M \le L$. Note that m and M are the extremes for the (normalised) expected number of edges as the proportions of colours vary.

Theorem 1.2.

(i) $\Psi(\alpha) = 0$.

(ii) Ψ *is finite and strictly monotone on* $[m, \alpha]$ *and on* $[\alpha, M]$.

(iii) Ψ *is infinite outside* $[m, M]$.

(iv) $\Phi(y) = 0$ *if and only if* $m \leq y \leq M$.

(v) $\Phi(y) < \infty$ *if and only if* $l \leq y \leq L$.

(vi) Φ *is strictly monotonic in* $[l, m]$ *and* $[M, L]$.

In $G(n, \alpha)$, $m = \alpha = M$, and so the 'inner' large deviation regime (with constants (n) and rate Ψ) is degenerate.

The approach produces, without much difficulty, results for some more general models. For example, it is not essential to colour each vertex independently, what matters is that the distributions of the numbers of colours obeys a suitable LDP (which it does in the independent case). This is formulated precisely in Theorem 5.1.

Section 2 illustrates how Theorem 1.1 applies in an example with two colours and a slight extension. Section 3 describes the general large deviation results used. Section 4 looks at the large deviation behaviour of the number of edges given that the colouring is according to some fixed proportions asymptotically. In Section 5, these results combine with suitable assumptions about the colouring to prove Theorem 1.1. Theorem 1.2 is proved in Section 6. Finally, in Section 7 the possibilities of drawing colours from a larger set and associating a random variable that takes values other than zero and one with every edge are discussed briefly.

2 An Example

To show the results above can be made more explicit in some cases, we let $|\Sigma| = 2$, and let p be the probability of an edge between vertices of the same colour and q the probability of an edge between opposite colours. Letting $\mathbf{x} = (x, 1-x)^T$, we have

$$\mathbf{x}^T P \mathbf{x} = p - 2x(1-x)(p-q)$$

and so

$$m = \min\left\{(p+q)/2, p\right\}, \quad M = \max\left\{(p+q)/2, p\right\},$$

$l = 0$ and $L = 1$. The regime described in part (i) of Theorem 1.1 is examined in the first lemma and that described in part (ii) in the second.

Lemma 2.1. *Suppose colours are picked independently using* $\mathbf{s} = (s, 1-s)^T$*, with* $s \leq 1/2$. *For* $y \in [m, M]$*, let*

$$x(y) = \frac{1}{2}\left(1 - \sqrt{1 - 2\frac{p-y}{p-q}}\right)$$

and let $\mathbf{x}(y) = (x(y), 1-x(y))^T$. *Then* $\Psi(y) = \psi(\mathbf{x}(y))$ *where* ψ *is given by (1.2).*

Proof. Let $\mathbf{x} = (x, 1-x)^T$. For $y \in [m, M]$ the condition that $\mathbf{x}^T P \mathbf{x} = y$, which is a quadratic in x, is easily solved to give the roots

$$\frac{1}{2}\left(1 \pm \sqrt{1 - 2\frac{p-y}{p-q}}\right).$$

Then $x(y)$ is the smaller of these two roots. There are just two candidate distributions to minimise over in the formula in Theorem 1.1(i): $(x(y), 1-x(y))^T$ and $(1-x(y), x(y))^T$. Using $s \le 1/2$ eliminates the second of these. $\square$

Lemma 2.2. *Assume $q < p$. The rate function $\Phi(y) = \inf\{\phi^*_{\mathbf{x}}(y) : \mathbf{x} \in \Delta\}$ in Theorem 1.1(ii) is given by*

$$\Phi(y) = \begin{cases} \phi^*_{(1/2,1/2)}(y) & \text{for } y \le m \\ 0 & \text{for } y \in [m, M] \\ \phi^*_{(1,0)}(y) & \text{for } y \ge M \end{cases} .$$

Hence the optimum choice for $\mathbf{x}$ in the definition is either 'all one colour' (when $y \ge M$) or 'equal proportions of the two colours' (when $y \le m$). Furthermore

$$\phi^*_{(1/2,1/2)}(y) = \sup_\theta \left(\theta y - \frac{1-p+\log(pe^\theta) + \log(1-q+qe^\theta)}{2} \right)$$

and

$$\phi^*_{(1,0)}(y) = \phi^*_{(0,1)}(y) = \sup_\theta \left(\theta y - \log\left(1-p+pe^\theta\right)\right).$$

If $q > p$ then $\phi^*_{(1/2,1/2)}$ and $\phi^*_{(1,0)}$ are interchanged in the formula above for Φ.

Proof. Let $\mathbf{x} = (x, 1-x)^T$. As we shall see in Lemma 4.5, $\phi_{\mathbf{x}}(\theta)$ is a differentiable convex function of θ. Since $\phi'_{\mathbf{x}}(0) = \mathbf{x}^T P \mathbf{x} \in [m, M]$, if $y \le m$

$$\phi^*_{\mathbf{x}}(y) = \sup_\theta(\theta y - \phi_{\mathbf{x}}(\theta)) = \sup_{\theta<0}(\theta y - \phi_{\mathbf{x}}(\theta)),$$

and similarly, when $y \ge M$, the supremum may be restricted to $\theta > 0$. By (1.1)

$$\begin{aligned} \phi_{\mathbf{x}}(\theta) &= (x^2 + (1-x)^2)\log(1-p+pe^\theta) + 2x(1-x)\log(1-q+qe^\theta) \\ &= \log(1-p+pe^\theta) + 2x(1-x)\log\left(\frac{1-q+qe^\theta}{1-p+pe^\theta}\right) \end{aligned}$$

Now note that $(1-q+qe^\theta)/(1-p+pe^\theta)$ is a monotonic function which is 1 for $\theta = 0$ and goes to $(1-q)/(1-p) > 1$ as $\theta \to -\infty$ (and to $q/p < 1$ as $\theta \to \infty$). Hence, for $0 \le x \le 1$

$$2x(1-x)\log\left(\frac{1-q+qe^\theta}{1-p+pe^\theta}\right) \tag{2.3}$$

is positive for $\theta < 0$ and negative for $\theta > 0$.

For $y < m$, a value of $2x(1-x)$ closer to its maximum will give a smaller value of $\phi^*_{\mathbf{x}}(y)$ for fixed y; and $2x(1-x)$ is maximised when $x = 1/2$. This gives the asserted form for $\Phi(y)$ in these cases. Similarly, when $y > M$, it is $\theta > 0$ that matters and now (2.3) has to minimised, which occurs when x is either 0 or 1.

When $q > p$, (2.3) is negative for $\theta < 0$ and positive for $\theta > 0$, interchanging the optimal mixtures. $\square$

In Lemma 2.2 the regime in Theorem 1.1(ii) arises by first picking the colouring to maximise (or minimise) the expected number of edges and then obtaining a large deviation in the number of edges from that colouring. However, the choice of the best colouring in Theorem 1.1(ii) need not arise in this way. To illustrate this, consider now three colours, with the matrix P given by

$$P = \begin{pmatrix} p & q & r \\ q & p & r \\ r & r & r \end{pmatrix}$$

with $(p+q)/2 < r < p$. Then it is easy to show that m and M are the same as if only the first two colours were considered. The minimum number of expected edges is obtained when the first two colours occur with probability one-half each.
Since $q < p$

$$\frac{p+q}{2} < 1 - \sqrt{(1-p)(1-q)} < p$$

and so in addition we can choose

$$r \in \left(\frac{p+q}{2}, 1 - \sqrt{(1-p)(1-q)} \right).$$

This ensures that $(1-r)^2 > (1-p)(1-q)$.
Suppose $y < m = (p+q)/2$. Theorem 1.1(ii) asserts that $\Phi(y)$ is obtained by minimising $\phi^*_{\mathbf{x}}(y)$ over the possible colourings. One possible $\mathbf{x}$ is $(1/2, 1/2, 0)^T$ and so Lemma 2.2 gives

$$\Phi(y) \le \sup_{\theta} \left(\theta y - \frac{\log(1-p+pe^{\theta}) + \log(1-q+qe^{\theta})}{2} \right).$$

Similarly, $(0,0,1)^T$ gives

$$\Phi(y) \le \sup_{\theta} \left(\theta y - \log(re^{\theta} + 1 - r) \right).$$

Considering $y = 0$,

$$\Phi(0) \le \min \left\{ -\frac{\log(1-p) + \log(1-q)}{2}, -\log(1-r) \right\} = -\log(1-r),$$

since, by arrangement, $(1-r)^2 > (1-p)(1-q)$. Thus, for $y \ge 0$, but sufficiently small, the colouring $(1/2, 1/2, 0)^T$, which gives the minimum expected number of edges, is certainly not the optimal one in Theorem 1.1(ii).

3 Large deviations: general results

The sequence of probability measures $(\mathbb{P}_n)$ on a (Polish) space Ω obeys a large deviation principle (LDP) with constants (a_n), tending to infinity, if there is a lower semi-continuous (lsc) non-negative function I on Ω (a rate function) such that for every open $G \subset \Omega$ and closed $F \subset \Omega$

$$-\inf_{y \in G} I(y) \le \liminf \frac{\log \mathbb{P}_n\{G\}}{a_n} \quad \text{and} \quad \limsup \frac{\log \mathbb{P}_n\{F\}}{a_n} \le -\inf_{y \in F} I(y). \tag{3.4}$$

The rate function I is called 'good' if for every finite β the set $\{x : I(x) \le \beta\}$ is compact.
The basic case of an LDP is when $\{X_i\}$ are independent and identically distributed and $\mathbb{P}_n$ is the distribution (on $\Omega = \mathbb{R}$) of $S_n/n = \sum_{i=1}^n X_i/n$, see [7, Theorem 2.2.3]. That result immediately yields the large deviation behaviour of $\mathscr{E}$ in $G(n, \alpha)$. The next result, which is a version of the Gärtner-Ellis Theorem, assumes much less about S_n.

Theorem 3.1. *Suppose (S_n) is a sequence of random variables, and (a_n) is a sequence of constants with $\lim_{n\to\infty} a_n = \infty$; let $\mathbb{P}_n$ be the distribution of S_n/a_n. Define $\phi_n(\theta) = a_n^{-1}\log(\mathbb{E}(e^{\theta S_n}))$. Assume that*

$$\lim_{n\to\infty} \phi_n(\theta) = \phi(\theta)$$

exists pointwise and is finite and differentiable for all θ. Then $(\mathbb{P}_n)$ satisfies the LDP with constants (a_n) and rate function ϕ^*.

Proof. See [7, Theorem 2.3.6 and Exercise 2.3.20]. □

Whenever $\mathbf{x}$ or $\mathbf{x}_n$ are used they are members of Δ, the probability distributions on the set of colours, even when this is not made explicit. Let $\mathbf{N}_n$ be the vector-valued random variable which gives the number of vertices of the various colours. Let $\Delta_n \subset \Delta$ be the set of possible values for $n^{-1}\mathbf{N}_n$. Let $\mathbb{Q}^n_{\mathbf{x}}$ be the conditional distribution of $\mathscr{E}/\mathfrak{c}_n$ when the number of vertices is n and the actual distribution of vertex colours is $\mathbf{x} \in \Delta_n$. Let μ^n be the distribution over Δ_n of $n^{-1}\mathbf{N}_n$ that arises from selecting colours independently according to $\mathbf{s}$. The marginal distribution of $\mathscr{E}/\mathfrak{c}_n$, $\mathbb{P}^n$, which is all that is seen when colours are invisible, is obtained by mixing over $\mathbf{x}$ with μ^n, and is given by

$$\mathbb{P}^n(A) = \int_{\Delta} \mathbb{Q}^n_{\mathbf{x}}(A)d\mu^n(\mathbf{x}) \ \left(= \int_{\Delta_n} \mathbb{Q}^n_{\mathbf{x}}(A)d\mu^n(\mathbf{x})\right). \tag{3.5}$$

This suggests that the behaviour of $\mathbb{P}^n$ should be understood through large deviation results for mixtures. That motivated a separate study reported in [1], extending results in [6] and [8]. The general result gives conditions for $(\mathbb{P}^n)$ given by (3.5) to obey an LDP when (μ^n) and $(\mathbb{Q}^n_{\mathbf{x}})$ satisfy suitable LDPs. Specialising the general result (specifically, Theorem 1 in [1] simplified using Lemma 3 and Proposition 4 there) to the framework here gives the following, which does not actually require that μ^n arises from independent colouring.

Theorem 3.2. *Suppose that*

- *(i) μ^n satisfies an LDP with constants (a_n) and good rate function ψ;*
- *(ii) whenever $\mathbf{x}_n \in \Delta_n$ and $\mathbf{x}_n \to \mathbf{x} \in \Delta$, $(\mathbb{Q}^n_{\mathbf{x}_n})$ satisfies an LDP with constants (a_n) and rate $\lambda_{\mathbf{x}}(y)$;*
- *(iii) $\lambda_{\mathbf{x}}(y)$ is lsc in $(y, \mathbf{x})$, jointly.*

Then $\mathbb{P}^n$ satisfies an LDP with constants (a_n) and rate

$$\lambda(y) = \inf\{\lambda_{\mathbf{x}}(y) + \psi(\mathbf{x}) : \mathbf{x} \in \Delta\}.$$

Furthermore, if, for each $\mathbf{x}$, $\lambda_{\mathbf{x}}(y)$ is a good rate function then λ is good.

Note that this theorem needs an LDP with the same constants for both the mixing distribution and the conditional distribution. In either case, these may not be the 'natural' constants for the LDP. To deal with this the next lemma gives a simple result on changing the constants in any LDP.

Lemma 3.3. *On $\mathscr{X}$, let $(\mathbb{P}_n)$ obey an LDP with constants (a_n) and rate I.*

- *(i) If $a_n/b_n \to 0$ and $\{x : I(x) < \infty\}$ is dense in $\mathscr{X}$ then $(\mathbb{P}_n)$ obeys an LDP with constants (b_n) and rate that is zero everywhere.*
- *(ii) If $b_n \to \infty$, $a_n/b_n \to \infty$ and I is zero at a single point z then $(\mathbb{P}_n)$ obeys an LDP with constants (b_n) and rate that is zero at z and infinity elsewhere.*

Proof. For the first part, since $\log \mathbb{P}_n\{F\} \leq 0$ the required bound on $\log \mathbb{P}_n\{F\}$ certainly holds. Now note that the bound on $\log \mathbb{P}_n\{G\}$ in (3.4) is finite, because every open set contains a member of any set that is dense in $\mathscr{X}$, and so converges to zero when multiplied by a_n/b_n.

For the second part, note first that the bound involving $\log \mathbb{P}_n\{G\}$ is automatic if $z \notin G$, as is the bound involving $\log \mathbb{P}_n\{F\}$ if $z \in F$. Now suppose $z \notin F$. Then, since I is lsc and z is its only zero, $\inf\{I(y) : y \in F\} > 0$ and so multiplying the bound in (3.4) by a_n/b_n gives the required infinite limit. The remaining case arises when $z \in G$. Then $z \notin G^c$, where G^c is the (closed) complement of G, and so $\inf\{I(y) : y \in G^c\} > 0$. Now the original LDP gives, rather comfortably, that $\mathbb{P}_n\{G^c\} \to 0$. This implies that $\log \mathbb{P}_n\{G\} \to 0$, which is more than enough to ensure that $b_n^{-1} \log \mathbb{P}_n\{G\} \to 0$. □

Clearly, information on the large deviation behaviour of $(\mathbb{Q}^n_{\mathbf{x}_n})$ is needed to apply Theorem 3.2. The independence of the edge placement given the colouring means this is fairly easy to deal with, using Theorem 3.1. That is the next topic.

4 LDPs conditional on the colouring

The main objective in this section is to prove the following theorem concerning the conditional distribution of the number of edges given the colouring.

Theorem 4.1. *When $\mathbf{x}_n \in \Delta_n$ and $\mathbf{x}_n \to \mathbf{x}$, $(\mathbb{Q}^n_{\mathbf{x}_n})$ obeys an LDP with constants $(\mathfrak{c}_n)$ and rate $\phi^*_{\mathbf{x}}$, and $\phi^*_{\mathbf{x}}(y)$ is lsc in $(y, \mathbf{x})$.*

The following lemma is a routine exercise using the independence of the edges given the colouring, that is given $\mathbf{N}_n$. The second and third are even easier.

Lemma 4.2. *Let $n^{-1}\mathbf{j} \in \Delta_n$. Then*

$$\mathbb{E}\left(e^{\theta \mathscr{E}(n)} \mid \mathbf{N}_n = \mathbf{j}\right) = \prod_{r<s} (p_{rs}e^\theta + 1 - p_{rs})^{j_r j_s} \prod_r (p_{rr}e^\theta + 1 - p_{rr})^{j_r(j_r-1)/2}.$$

Lemma 4.3. *If $\mathbf{x}_n \to \mathbf{x}$ then $\mathbf{x}_n^T A(\theta)\mathbf{x}_n \to \mathbf{x}^T A(\theta)\mathbf{x} = \phi_{\mathbf{x}}(\theta)$, where A and $\phi_{\mathbf{x}}$ are defined just before Theorem 1.1.*

Lemma 4.4. *The mean of the distribution $\mathbb{Q}^n_{\mathbf{x}}$ is $\mathbf{x}^T P\mathbf{x}$.*

Theorem 3.1 relies on properties of the (suitably scaled) logarithm of the moment generating function. With a view to applying that theorem, let

$$\phi_{n^{-1}\mathbf{j},n}(\theta) = \frac{\log\left(\mathbb{E}\left(e^{\theta \mathscr{E}(n)} \mid \mathbf{N}_n = \mathbf{j}\right)\right)}{\mathfrak{c}_n} = 2\frac{\log\left(\mathbb{E}\left(e^{\theta \mathscr{E}(n)} \mid \mathbf{N}_n = \mathbf{j}\right)\right)}{n(n-1)}.$$

Lemma 4.5. *If $n^{-1}\mathbf{j}_n \in \Delta_n$, so that $\mathbf{j}_n$ is a possible vector of numbers of vertices of each colour when there are n vertices, and $n^{-1}\mathbf{j}_n \to \mathbf{x}$, then*

$$\phi_{n^{-1}\mathbf{j}_n,n}(\theta) \longrightarrow \phi_{\mathbf{x}}(\theta),$$

and $\phi_{\mathbf{x}}(\theta)$ is an everywhere differentiable convex function of θ with $\phi_{\mathbf{x}}(0) = 0$. Also, $\phi^*_{\mathbf{x}}$ has a unique minimum of zero at $y = \mathbf{x}^T P\mathbf{x}$.

Proof. Let $\mathbf{d}(\theta)_i = \log(1 - p_{ii} + p_{ii}e^{\theta})$. Using Lemma 4.2,

$$\begin{aligned} \mathbb{E}\left(e^{\theta\mathscr{E}(n)} \mid \mathbf{N}_n = \mathbf{j}\right) &= \prod_{1\le s<t\le k}(1-p_{st}+p_{st}e^{\theta})^{j_s j_t}\prod_{1\le s\le k}(1-p_{ss}+p_{ss}e^{\theta})^{j_s(j_s-1)/2} \\ &= \exp(\mathbf{j}^T A(\theta)\mathbf{j}/2 - \mathbf{j}^T\mathbf{d}(\theta)/2). \end{aligned}$$

Hence, using Lemma 4.3,

$$\phi_{n^{-1}\mathbf{j}_n,n}(\theta) = \frac{\mathbf{j}_n^T A(\theta)\mathbf{j}_n - \mathbf{j}_n^T\mathbf{d}(\theta)}{n(n-1)} \longrightarrow \phi_{\mathbf{x}}(\theta).$$

The limit's differentiability is a consequence of $A'_{ij}(t)$ being bounded uniformly in (i,j) in any neighbourhood of θ and dominated convergence; its convexity follows from being the limit of convex functions; $A(0)=0$ implies that $\phi_{\mathbf{x}}(0)=0$ for every $\mathbf{x}$. For the last part, use [11, Theorem 12.2] to see that $\phi_{\mathbf{x}}^{**} = \phi_{\mathbf{x}}$ and then [11, Theorem 27.1(e)] to complete the proof. □

Lemma 4.6. *$\phi_{\mathbf{x}}^*(y)$ is a lsc function of the pair $(y,\mathbf{x})$ and infinite for every $y \notin [l,L]$.*

Proof. Suppose $\mathbf{x}_n \to \mathbf{x}$ and $y_n \to y$. Then, for any $\epsilon > 0$, there is a finite θ such that

$$\begin{aligned} \phi_{\mathbf{x}}^*(y) - \epsilon &\le \theta y - \phi_{\mathbf{x}}(\theta) \\ &= \theta y_n - \phi_{\mathbf{x}_n}(\theta) + \theta(y-y_n) - (\phi_{\mathbf{x}}(\theta) - \phi_{\mathbf{x}_n}(\theta)) \\ &\le \phi_{\mathbf{x}_n}^*(y_n) + \theta(y-y_n) - (\phi_{\mathbf{x}}(\theta) - \phi_{\mathbf{x}_n}(\theta)) \end{aligned}$$

and so, since $\phi_{\mathbf{x}_n}(\theta) \to \phi_{\mathbf{x}}(\theta)$ by Lemma 4.3,

$$\phi_{\mathbf{x}}^*(y) - \epsilon \le \liminf_n \phi_{\mathbf{x}_n}^*(y_n)$$

as required.
Using the explicit form of A

$$\lim_{\theta\to-\infty}\phi_{\mathbf{x}}'(\theta) = \lim_{\theta\to-\infty}\mathbf{x}^T A'(\theta)\mathbf{x} \ge l$$

and so $\phi_{\mathbf{x}}^*(y) = \infty$ when $y < l$. Similarly, letting $\theta \to \infty$, $\phi_{\mathbf{x}}^*(y) = \infty$ when $y > L$. □

Proof of Theorem 4.1. Apply Lemmas 4.5 and 4.6 and Theorem 3.1. □

Lemma 4.7. *Suppose that $b_n \to \infty$ and $b_n/\mathfrak{c}_n \to 0$. When $\mathbf{x}_n \in \Delta_n$ and $\mathbf{x}_n \to \mathbf{x}$, $(\mathbb{Q}_{\mathbf{x}_n}^n)$ obeys an LDP with constants (b_n) and rate $\lambda_{\mathbf{x}}$, where $\lambda_{\mathbf{x}}(\mathbf{x}^T P\mathbf{x}) = 0$ and $\lambda_{\mathbf{x}}(y) = \infty$ elsewhere. Furthermore $\lambda_{\mathbf{x}}(y)$ is lsc in $(y,\mathbf{x})$.*

Proof. The LDP and the form of $\lambda_{\mathbf{x}}$ follow directly from Lemma 3.3(ii) and the last part of Lemma 4.5. A routine verification shows $\lambda_{\mathbf{x}}(y)$ is lsc. □

5 LDPs for the number of edges

The next result does not presume that the distribution of colours on n vertices, μ^n arises from independent choices. Theorem 1.1 will be derived by specialisation from this one.

Theorem 5.1. *(i) Suppose the colouring (μ^n) obeys an LDP, with constants (b_n), and a good rate function, ψ. Suppose also that $(b_n/\mathfrak{c}_n) \to 0$. Then $(\mathbb{P}^n)$ satisfies an LDP with constants (b_n) and rate $\Psi(y) = \inf\{\psi(\mathbf{x}) : \mathbf{x}^T P\mathbf{x} = y\}$.*

(ii) Suppose the colouring (μ^n) obeys an LDP with constants $(\mathfrak{c}_n)$ and rate function that is zero throughout Δ. Then $(\mathbb{P}^n)$ satisfies an LDP with constants $(\mathfrak{c}_n)$ and rate $\Phi(y) = \inf\{\phi^*_{\mathbf{x}}(y) : \mathbf{x} \in \Delta\}$.

Proof. The first part is an application of Theorem 3.2 and Lemma 4.7. Note that, because Δ is compact, the rate function that is identically zero on Δ is good. Now the second part is an application of Theorems 3.2 and 4.1. □

The next result show that independent colourings produce (μ^n) with the right properties for both parts of Theorem 5.1.

Lemma 5.2. *For independent colouring with a finite number of possible colours, (μ^n) obeys an LDP with constants (n) and good rate function (on Δ). Furthermore, (μ^n) obeys an LDP with constants $(\mathfrak{c}_n)$ and rate function that is zero everywhere.*

Proof. When $\mathbf{N}_n$ has a multinomial distribution, the distributions of $\mathbf{n}_n = n^{-1}\mathbf{N}_n$ satisfy an LDP with constants (n) and rate given by (1.2) as a consequence of Sanov's Theorem, [7, Definition 2.1.5 et seq., and Theorem 2.1.10]. This rate function is continuous, convex and bounded on Δ. The final assertion is from Lemma 3.3(i). □

Proof of Theorem 1.1. This follows directly from Theorem 5.1 and Lemma 5.2. □

6 Properties of the rate functions

In this section Theorem 1.2 is proved after giving three preliminary Lemmas. The arguments rely heavily on the fact that Δ is compact.

Lemma 6.1. *Suppose the convex functions ϕ_n converge to ϕ, which is necessarily convex, as $n \to \infty$. Then*

$$\phi^*(y) \le \liminf \phi_n^*(y).$$

Proof. For any $\epsilon > 0$, there is a finite θ such that

$$\phi^*(y) - \epsilon \le \theta y - \phi(\theta) = \theta y - \phi_n(\theta) + (\phi_n(\theta) - \phi(\theta)) \le \phi_n^*(y) + (\phi_n(\theta) - \phi(\theta)).$$

Hence $\phi^*(y) - \epsilon \le \liminf_n \phi_n^*(y)$. □

Lemma 6.2. *The infimum in $\Phi(y) = \inf_{\mathbf{x}} \phi^*_{\mathbf{x}}(y)$ is attained.*

Proof. Let $\mathbf{x}_n$ be such that $\phi^*_{\mathbf{x}_n}(y) \to \Phi(y)$, with (using the compactness of Δ) $\mathbf{x}_n \to \mathbf{x}$. Then $\phi_{\mathbf{x}_n} \to \phi_{\mathbf{x}}$ and so, by Lemma 6.1

$$\Phi(y) = \liminf_n \phi^*_{\mathbf{x}_n}(y) \ge \phi^*_{\mathbf{x}}(y) \ge \Phi(y). \qquad \square$$

Lemma 6.3. *Suppose ψ is convex, finite on Δ and takes the value zero at a single $\mathbf{s}$. Then Ψ is strictly monotone on $[m, \alpha]$ and on $[\alpha, M]$. It is infinite outside $[m, M]$.*

Proof. Note first that $\Psi(\alpha) = \psi(\mathbf{s}) = \inf_{\mathbf{x}} \psi(\mathbf{x}) = 0$ and this infimum is attained only at $\mathbf{s}$. Take $y \in [m, \alpha)$ and $\tilde{\mathbf{x}}$ such that $\Psi(y) = \psi(\tilde{\mathbf{x}})$ and $\tilde{\mathbf{x}}^T P \tilde{\mathbf{x}} = y$. Since ψ is always finite, $\Psi(y) < \infty$. Take $z \in (y, \alpha)$. Then, for suitable $\delta > 0$, $(1-\delta)\tilde{\mathbf{x}} + \delta \mathbf{s} \in \{\mathbf{x} : \mathbf{x}^T P \mathbf{x} = z\}$ and, by convexity,

$$\psi((1-\delta)\tilde{\mathbf{x}} + \delta\mathbf{s}) \le (1-\delta)\psi(\tilde{\mathbf{x}}) + \delta\psi(\mathbf{s}) = (1-\delta)\psi(\tilde{\mathbf{x}}) < \psi(\tilde{\mathbf{x}})$$

Hence $\Psi(z) < \Psi(y)$ as required. The range from α to M is similar. □

Without the assumption that ψ is finite on Δ the proof of Lemma 6.3 still works to show that Ψ is monotone when finite either side of its minimum. However, it may be infinite for some values within $[m, M]$ which would mean that the two parts of Theorem 5.1 would leave a range of values (where Ψ is infinite and Φ is zero) where an LDP with constants intermediate between (b_n) and $(\mathfrak{c}_n)$ might be appropriate.

Proof of Theorem 1.2. For independent colouring, from (1.2), ψ is zero only at $\mathbf{s}$ and is finite throughout Δ. Then (i) holds because $\alpha = \mathbf{s}^T P \mathbf{s}$ and (ii) and (iii) hold by Lemma 6.3.
By Lemma 4.5, $\phi^*_{\mathbf{x}}(y) = 0$ exactly when $y = \mathbf{x}^T P \mathbf{x}$. Thus, since $\mathbf{x}^T P \mathbf{x}$ takes all values in $[m, M]$ as $\mathbf{x}$ varies, $\Phi(y) = 0$ for all $y \in [m, M]$. On the other hand, by Lemma 6.2, $\phi^*_{\mathbf{x}}(y) = 0$ for some $\mathbf{x}$ when $\Phi(y) = 0$, and then $y = \mathbf{x}^T P \mathbf{x} \in [m, M]$. This proves (iv).
Similarly, $\phi'_{\mathbf{x}}(\theta) \to \mathbf{x}^T I(P_{ij} = 1)\mathbf{x}$, as $\theta \to -\infty$ and so $\phi^*_{\mathbf{x}}(y) = \infty$ when $y < \mathbf{x}^T I(P_{ij} = 1)\mathbf{x}$. On the other hand, when $y = l$ direct calculation shows that $\phi^*_{\mathbf{x}}(l) < \infty$ for any $\mathbf{x}$ that provides the infimum in the definition of l. Considering $\theta \to \infty$ shows that $\Phi(y) = \infty$ for $y > L$ and $\Phi(L) < \infty$. This proves the 'only if' part of (v).
Take $y < m$. By Lemma 6.2 we know that for $\epsilon > 0$ there is a suitable $\mathbf{x}_*$ such that

$$\Phi(y - \epsilon) = \phi^*_{\mathbf{x}_*}(y - \epsilon) = \sup_\theta [\theta(y - \epsilon) - \mathbf{x}_*^T A(\theta)\mathbf{x}_*].$$

Since $y < m$, both $\phi^*_{\mathbf{x}_*}(y - \epsilon)$ and $\phi^*_{\mathbf{x}_*}(y)$ are strictly positive, by part (iv). Hence there must be some $\delta > 0$ such that

$$\begin{aligned}
\sup_\theta [\theta(y-\epsilon) - \mathbf{x}_*^T A(\theta)\mathbf{x}_*] &= \sup_{\theta \le -\delta} [\theta(y-\epsilon) - \mathbf{x}_*^T A(\theta)\mathbf{x}_*] \\
&\ge \sup_{\theta \le -\delta} [\theta y - \mathbf{x}_*^T A(\theta)\mathbf{x}_*] + \delta\epsilon \\
&= \phi^*_{\mathbf{x}_*}(y) + \delta\epsilon \\
&\ge \Phi(y) + \delta\epsilon.
\end{aligned}$$

Hence $\Phi(y - \epsilon) \ge \Phi(y) + \delta\epsilon$ which gives strict monotonicity when $\Phi(y)$ is finite. Furthermore taking $y - \epsilon = l$ and using $\Phi(l) < \infty$ shows that $\Phi(y)$ is finite for $l < y < m$. The range $M < y < L$ is handled similarly. This completes the proof of (v) and (vi). □

7 Extensions

Suppose that the set of available colours, Σ, is a Polish (i.e. complete, separable metric) space and P is continuous on $\Sigma \times \Sigma$. Let Δ be the set of probability distributions on Σ (equipped with the Lévy metric, which gives the topology of weak convergence of distributions and makes it a Polish space, see [7, Theorem D.8, p319]). Denote weak convergence of a sequence $(\mathbf{x}_n \in \Delta)$ of measures to another such $\mathbf{x}$ by $\mathbf{x}_n \xrightarrow{w} \mathbf{x}$. When there are only a finite number of possible colours $\mathbf{x}_n \xrightarrow{w} \mathbf{x}$ is identical to the convergence of the vectors in $\mathbb{R}^k$. The proofs in Section 4 are unchanged if $\mathbf{x}_n \to \mathbf{x}$ is replaced by $\mathbf{x}_n \xrightarrow{w} \mathbf{x}$. One should perhaps be made more explicit.

Proof of Lemma 4.3. For fixed $\theta, A(\theta)$ is a bounded continuous function on $\Sigma \times \Sigma$ and the product measure $\mathbf{x}_n \times \mathbf{x}_n$ on $\Sigma \times \Sigma$ converges weakly to $\mathbf{x} \times \mathbf{x}$. $\square$

Theorem 5.1 generalises in the obvious way: part (ii) needs the extra condition that Δ is compact (which is equivalent to Σ being compact [7, Theorem D.8.3]) to ensure the null rate function is good.

Theorem 7.1. *Suppose Σ is compact, colouring is independent according to* $\mathbf{s}$ *and* $\mathbf{s}$ *has support Σ. Then the conclusions of Theorem 1.1 hold.*

Proof. By Sanov's Theorem ([7, §6.2]) the first part of Lemma 5.2 holds, and the rate is finite exactly for those $\mathbf{x} \in \Delta$ that are absolutely continuous with respect to $\mathbf{s}$. These are dense in Δ. Thus, using Lemma 3.3(i), the second part of Lemma 5.2 holds also. Now apply the generalised version of Theorem 5.1. $\square$

The second direction for extension associates a random variable with each pair of vertices that is more general than the indicator variables for the presence of that edge. Let $M(\sigma_1, \sigma_2; \theta)$ be the moment generating function of a random variable associated with an edge with vertices of colours σ_1 and σ_2, with mean $m(\sigma_1, \sigma_2)$. Assume, for each θ, that $M(\sigma_1, \sigma_2; \theta)$ is a bounded continuous function on $\Sigma \times \Sigma$, with a derivative that is bounded uniformly in $\Sigma \times \Sigma$ on a neighbourhood of any θ. For $\mathbf{x} \in \Delta$ let s_1 and s_2 be independent colours selected using $\mathbf{x}$ and let $\phi_{\mathbf{x}}(\theta) = \log \mathbb{E} M(s_1, s_2; \theta)$ and $m_{\mathbf{x}} = \mathbb{E} m(s_1, s_2)$. Finally, let $\mathscr{E}$ be the sum of the variables over the edges and $\mathbb{P}^n$ be the distribution of $\mathscr{E}/\mathfrak{c}_n$. The next result is obtained by working through the details of the arguments leading to Theorem 5.1 and checking that nothing has changed.

Theorem 7.2. *(i) Suppose the colouring (μ^n) obeys an LDP, with constants (b_n), and good rate function, ψ, and $(b_n/\mathfrak{c}_n) \to 0$. Then $\mathbb{P}^n$ satisfies an LDP with constants (b_n) and rate $\Psi(y) = \inf\{\psi(\mathbf{x}) : m_{\mathbf{x}} = y\}$.*

(ii) Suppose Δ is compact and the colouring (μ^n) obeys an LDP with constants $(\mathfrak{c}_n)$ and rate function that is zero throughout Δ. Then $\mathbb{P}^n$ satisfies an LDP with constants $(\mathfrak{c}_n)$ and rate $\Phi(y) = \inf\{\phi^*_{\mathbf{x}}(y) : \mathbf{x} \in \Delta\}$.

8 Other models

In the spirit of [3], the probability that an edge between a vertex of colour i and one of colour j could be $p_{ij}(n)$ when there are n vertices, with $np_{ij}(n) \to c_{ij} < \infty$. Various large deviations for this model are obtained in [9]. Here we consider only the number of edges. Following the development in §4, large deviations arising from edge placement conditional on an asymptotic colouring will have constants (n), with a colouring of $\mathbf{x}$ giving

$$\phi^*_{\mathbf{x}}(y) = y(\log y - \log \mathbf{x}^t C \mathbf{x}) - y + \mathbf{x}^t C \mathbf{x},$$

which is, of course, the rate for a Poisson distribution with mean $\mathbf{x}^t C \mathbf{x}$. This is jointly lsc in $(y, \mathbf{x})$. Thus, when the colouring (μ^n) obeys an LDP with constants (n), and a good rate function, ψ, Theorem 3.2 give the overall large deviation rate for the number of edges with constants (n) to be

$$\lambda(y) = \inf\{\phi^*_{\mathbf{x}}(y) + \psi(\mathbf{x}) : \mathbf{x} \in \Delta\}.$$

References

[1] J. D. Biggins. Large deviations for mixtures. *Electron. Comm. Probab.*, 9:60–71 (electronic), 2004. MR2081460

[2] B. Bollobás. *Random graphs*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985. MR0809996

[3] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31(1):3–122, 2007. MR2337396

[4] C. Cannings and D. B. Penman. Models of random graphs and their applications. In *Stochastic processes: modelling and simulation*, volume 21 of *Handbook of Statist.*, pages 51–91. North-Holland, Amsterdam, 2003. MR1973541

[5] A. Cerquetti and S. Fortini. A Poisson approximation for coloured graphs under exchangeability. *Sankhyā*, 68(2):183–197, 2006. MR2303080

[6] N. R. Chaganty. Large deviations for joint distributions and statistical applications. *Sankhyā Ser. A*, 59(2):147–166, 1997. MR1665683

[7] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Jones and Bartlett Publishers, Boston, MA, 1993. MR1202429

[8] I. H. Dinwoodie and S. L. Zabell. Large deviations for exchangeable random vectors. *Ann. Probab.*, 20(3):1147–1166, 1992. MR1175254

[9] K. Doku-Amponsah and P. Mörters. Large deviation principles for empirical measures of coloured random graphs. `arXiv:math/0607545v1`, 2006.

[10] D. B. Penman. *Random graphs with correlation structure.* Ph.D thesis, University of Sheffield, 1998.

[11] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970. MR0274683