

# Large Deviation Analysis for Layered Percolation Problems on the Complete Graph\*

Lincoln Chayes, S. Alex Smith

Department of Mathematics, University of California at Los Angeles, Los Angeles, California; e-mail: lchayes@math.ucla.edu

Received 3 June 2008; accepted 16 September 2010; received in final form 24 June 2011  
Published online 23 November 2011 in Wiley Online Library (wileyonlinelibrary.com).  
DOI 10.1002/rsa.20387

**ABSTRACT:** We analyze the large deviation properties for the (multitype) version of percolation on the complete graph – the simplest substitutive generalization of the Erdős-Rényi random graph that was treated in article by Bollobás et al. (Random Structures Algorithms 31 (2007), 3–122). Here the vertices of the graph are divided into a fixed finite number of sets (called *layers*) the probability of  $\{u, v\}$  being in our edge set depends on the respective layers of  $u$  and  $v$ . We determine the exponential rate function for the probability that a giant component occupies a fixed fraction of the graph, while all other components are small. We also determine the exponential rate function for the probability that a particular exploration process on the random graph will discover a certain fraction of vertices in each layer, without encountering a giant component. © 2011 Wiley Periodicals, Inc. Random Struct. Alg., 40, 460–492, 2012

*Keywords:* Erdős-Rényi random graph; percolation; exponential rate function

## 1. INTRODUCTION

The principle object of study in this paper is a generalized version of the Erdős-Rényi random graph. Here, instead of a percolation process occurring homogeneously throughout the graph, the vertex set is first divided into a finite number of differing types. Then edges are placed randomly and independently with a probability that depends on the types of vertices it might connect (and an overall scaling by the inverse of the total number of vertices). Models of this form were treated in [6] where some preliminary results were derived. Moreover, this model is a special case of the general systems defined in [3] – specifically Example 4.3 – where more detailed properties were obtained.

---

Correspondence to: L. Chayes

\*Supported by NSF (DMS-0306167, DMS-0805486)

© 2011 Wiley Periodicals, Inc.

From our perspective (which is slightly different than that of the above mentioned) the differing types correspond to *layers* which, by analogy with spin–systems, are physically displaced copies of some regular infinite lattice. The scaling of the interaction/edge probabilities is the *mean–field* approximation and, in the context of this approximation, for various reasons the interaction might be non–homogeneous relative to the layers of the constituents.

The subject of layered mean–field spin–systems is rather ancient (see, e.g., [5]). While it is unlikely that a complete mathematical treatment of these models has ever been enacted, it is still regarded as a reliable guide for the behavior of systems in thin geometries (see, e.g., [7]). For the layered problems, as in the homogenous cases, the mean–field analysis is greatly facilitated by the construction of a free energy function: In general one may write down, explicitly or implicitly what amounts to the large deviation rate for the probability of observing a particular magnetization fraction (or other general order–parameter) in each layer. This function is then minimized whereupon quantities of interest e.g., the actual limiting magnetization profile, are obtained. Such an approach was taken in [4] for the standard Erdős–Rényi random graph and was, to some extent, a facet in the study of the random cluster model on the complete graph [2].

While the strategic philosophy that we follow in this work has some of its roots in [4], the multiple layers represent additional dimensions to the problem. Thus, to find the paramount large deviation rates, we must optimize over *trajectories* in a density–parameter space which, at the discrete level are generated by the systematic removal of clusters. This leads to a constrained variational problem the (semi-explicit) solution of which provides the desired rates.

We close this section with the basic notations we will be using and, in the next section, we will state our main results. Proofs will emerge in Sections 3–5.

**1.1. Notation**

For the purposes of this paper, we define a *layered* set to be a set in which each element  $v$  has an associated integer, which we designate by  $\text{layer}(v)$ . We similarly define a layered graph to be a graph whose vertex set is a layered set. Since we think of  $\text{layer}(v)$  as having something to do with the position of  $v$ , we will freely use language such as “ $v$  is in the  $i^{\text{th}}$  layer” to mean that  $\text{layer}(v) = i$ . If  $S$  is a layered set, we let  $S_\ell = \{v \in S : \text{layer}(v) = \ell\}$  be the set of elements of  $S$  in layer  $\ell$ . If  $G$  is a layered graph with vertex set  $V$ , we let  $G_\ell$  be the subgraph of  $G$  restricted to  $V_\ell$ .

We will typically use  $L$  to indicate the number of layers in a graph and we will use the hat symbol ( $\hat{\cdot}$ ) to indicate vectors with  $L$  components. If  $\hat{\eta}$  is such a vector, we will let  $|\hat{\eta}|$  be the  $\mathcal{L}^1$  norm on  $\hat{\eta}$ . If  $S$  is a set or a graph, we let  $|S|$  be the number of elements or vertices in  $S$ , respectively. We will let  $\langle S \rangle$  be the vector of dimension  $L$  whose  $\ell^{\text{th}}$  component is  $|S_\ell|$ . We will also use the componentwise partial ordering on vectors, so (for example)  $\hat{\eta} > 0$  will mean that  $\hat{\eta} \in (0, \infty)^L$ . If  $v$  is a vertex in a graph  $G$ , we will let  $C(v) = C_G(v)$  be the component of  $G$  containing  $v$ .

We now introduce the main model of the paper: Let  $L \in \mathbb{N}$  and  $\hat{n} \in \mathbb{N}^L$  be given, and consider the layered vertex set

$$\mathcal{V} = \{(\ell, k) : 1 \leq \ell \leq L, 1 \leq k \leq \hat{n}_\ell\}, \tag{1.1}$$

where  $\text{layer}((\ell, k)) = \ell$ . Then given probabilities  $(p_{ij})_{i,j \in \{1, \dots, L\}}$  with  $p_{ij} = p_{ji}$  for all  $i$  and  $j$ , let  $\mathcal{E}$  be the (random) edge set so that each edge  $\{u, v\}$  appears (or not) in  $\mathcal{E}$  independently,

and  $P(\{u, v\} \in \mathcal{E}) = p_{\text{layer}(u)\text{layer}(v)}$ , so that the probability that an edge exists between given vertices in layers  $i$  and  $j$  is  $p_{ij}$ . In this paper we will restrict ourselves to the case in which the number of vertices in each layer scales proportionally to the number of vertices in every other layer; so let us for the remainder of the paper take  $\hat{n} = (\lfloor \hat{\rho}_1 n \rfloor, \dots, \lfloor \hat{\rho}_L n \rfloor)$  for some  $\hat{\rho} \in (0, \infty)^L$  and  $n > 0$ . Additionally, let us take  $A = (\alpha_{ij})$  to be a symmetric, non-negative, irreducible  $L \times L$  matrix, and let  $p_{ij} = \frac{\alpha_{ij}}{n}$ . We will generally consider  $L$  and  $A$  to be fixed throughout the paper, and therefore we designate the resulting random graph by  $\mathcal{G}(n, \hat{\rho})$  and the corresponding probability measure by  $P_{n, \hat{\rho}}$  – although we will allow  $n$  and  $\hat{\rho}$  to be implicit when it is clear from context. We let  $E_{n, \hat{\rho}}$  indicate expectation with respect to  $P_{n, \hat{\rho}}$ .

Throughout this note we shall, emphatically, *not* adhere to the summation convention (concerning repeated indices) as this would be a cause for much confusion.

## 2. MAIN RESULTS

### 2.1. Background

Here we shall summarize various properties of the model which will be needed later. The vast majority of these have been proved in [3] and [6], often by comparison to branching processes. The seminal result, stated below, can now be derived by the alternative method of minimizing the appropriate free energy/rate function.

**Theorem 2.1.** *Let  $\hat{\theta}(r, n, \hat{\rho})$  be the (random) portion of sites in each layer of  $\mathcal{G}(n, \hat{\rho})$  which are in components of size greater than  $r$  (that is to say  $\hat{\theta}_k = \hat{\rho}_k \times$  the fraction of sites satisfying the above clause).*

*Then there exists a  $\hat{\theta}^*(\hat{\rho})$  so that*

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \hat{\theta}(\epsilon n, n, \hat{\rho}) = \hat{\theta}^*(\hat{\rho}) \tag{2.1}$$

*in the sense of convergence in distribution. Furthermore,  $\hat{\theta}^*(\hat{\rho})$  is the maximum solution to the system of equations given by*

$$\hat{\theta}_k = \hat{\rho}_k \left[ 1 - \exp \left( - \sum_{i=1}^L \alpha_{ik} \hat{\theta}_i \right) \right] \tag{2.2}$$

**Remark .** This is a special case of Theorem 3.1 in [3] where, as explained therein, “maximal” actually turns out to be a well defined concept as will be described in the next lemma. The above was also obtained as Eq. (7) – and Theorem 1 – of [6] for the particular model at hand.

**Lemma 2.2.** *The system of equations given by (2.2) has a maximal solution in  $\mathbb{R}^L$  for all  $\hat{\rho} > 0$ , which is either zero or strictly positive. This maximal solution is an increasing function of  $\hat{\rho}$ . Moreover, if a strictly positive solution to (2.2) exists, it is unique.*

The proof of Lemma 2.2 can be found in the proofs Theorems 6.1 and 6.2 of [3] and is adequately discussed in [6]. It also is discussed in [1] (and references therein) in the context of Theorem 8.

In the first two citations, the critical condition for the model on  $\mathcal{G}(n, \hat{\rho})$  is also spelled out. This, in our language, is described as follows:

We will say that  $\hat{\rho}$  is a *supercritical* density if  $\hat{\theta}^*(\hat{\rho})$  is nonzero, a *subcritical* density if there is a neighborhood of  $\hat{\rho}$  in  $(0, \infty)^L$  on which  $\hat{\theta}^*$  is zero, and a *critical* density otherwise. In order to characterize these three regions of density further, we introduce an attachment-susceptibility matrix, whose  $ij^{\text{th}}$  entry is the expected number of edges from a vertex in layer  $i$  to vertices in layer  $j$  (up to  $O(1/n)$  corrections):

$$B_{\hat{\rho}} = \begin{bmatrix} \alpha_{11}\hat{\rho}_1 & \alpha_{12}\hat{\rho}_2 & \cdots & \alpha_{1L}\hat{\rho}_L \\ \alpha_{21}\hat{\rho}_1 & \alpha_{22}\hat{\rho}_2 & \cdots & \alpha_{2L}\hat{\rho}_L \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{L1}\hat{\rho}_1 & \alpha_{L2}\hat{\rho}_2 & \cdots & \alpha_{LL}\hat{\rho}_L \end{bmatrix}. \tag{2.3}$$

We then have:

**Lemma 2.3.** *Let  $\kappa_0(\hat{\rho})$  be the maximal eigenvalue of  $B_{\hat{\rho}}$ . Then  $\hat{\rho}$  is subcritical if  $\kappa_0(\hat{\rho}) < 1$ , critical if  $\kappa_0(\hat{\rho}) = 1$ , and supercritical if  $\kappa_0(\hat{\rho}) > 1$ .*

We omit a formal proof since results of this form have appeared previously (c.f. the Remark below).

The behavior above, below and at criticality is described by the following:

**Proposition 2.4.** *Let  $L, A$ , and  $\hat{\rho} > 0$  be given. Suppose that for each  $n$ , a vertex is chosen uniformly – so that  $P_{n,\hat{\rho}}(\text{layer}(v) = i) = |\hat{\rho}|^{-1}\hat{\rho}_i$ . Then if  $\hat{\rho}$  is subcritical,*

$$\lim_{n \rightarrow \infty} E_{n,\hat{\rho}}[|C(v)|] = (\mathbf{I} - B_{\hat{\rho}}^T)^{-1} \frac{\hat{\rho}}{|\hat{\rho}|}. \tag{2.4}$$

*If  $\hat{\rho}$  is critical then*

$$\lim_{n \rightarrow \infty} E_{n,\hat{\rho}}[|C(v)|] = \infty, \tag{2.5}$$

*but*

$$\lim_{n \rightarrow \infty} P_{n,\hat{\rho}}(|C(v)| > \epsilon n) = 0 \tag{2.6}$$

*for all  $\epsilon > 0$ . If  $\hat{\rho}$  is supercritical then there is some  $\epsilon > 0$  such that*

$$\liminf_{n \rightarrow \infty} P_{n,\hat{\rho}}(|C(v)| > \epsilon n) > 0. \tag{2.7}$$

**Remark.** The critical and supercritical results here and in Lemma 2.3 are part of the main theorem (Theorem 3.1) in [3] and in various places in [6]. The subcritical result is not readily found in these references. The proof is not particularly difficult: As in [3] Section 7 (see also [6]) we may compare to a relevant branching process as described in their Sections 5 and 6 and in [1]. In sub-critical instance the desired result is readily obtained [1] since the process dies out at an exponential rate.

It may be of slight concern (or interest) to see the transpose of  $B_{\hat{\rho}}$  playing the seminal rôle in the subcritical – as well as the critical – cases. We demonstrate by direct appeal to the above mentioned branching process; note that this is peculiar to the uniform selection

of the vertex. Let  $X_{ij}$  denote the *matrix* element which represents the average number of vertices of type  $j$  attached to the root when it is a single vertex of type  $i$ . Then, as one would expect,

$$X_{ij} = \delta_{ij} + \sum_{k=1}^L B_{ik} X_{kj}. \quad (2.8)$$

But, defining

$$V_{ij} := \frac{\hat{\rho}_i}{|\hat{\rho}|} X_{ij}$$

we have

$$V_{ij} = \frac{\hat{\rho}_i}{|\hat{\rho}|} \delta_{ij} + \sum_{k=1}^L \frac{\hat{\rho}_i}{|\hat{\rho}|} B_{ik} X_{kj}. \quad (2.9)$$

By the special attribute of uniform selection – as opposed to some other definition of  $V_{ij}$  – we see

$$\frac{\hat{\rho}_i}{|\hat{\rho}|} B_{ik} X_{kj} = \frac{\hat{\rho}_i}{|\hat{\rho}|} A_{ik} \hat{\rho}_k X_{kj} = B_{ik}^T V_{kj}.$$

Thus, multiplying (2.9) on the right by a vector with unity in each component we obtain

$$\sum_{j=1}^L V_{ij} =: \hat{v}_i = \frac{\hat{\rho}_i}{|\hat{\rho}|} + \sum_{k=1}^L B_{ik}^T \hat{v}_k \quad (2.10)$$

which is equivalent to (2.4). It is noted that if for fixed  $\hat{\rho} \neq 0$  a solution to (2.4) has been acquired with  $\hat{v}_i \geq 0$  then it can be concluded that the model defined by  $B_{\hat{\rho}}$  is subcritical. Indeed, since  $B_{\hat{\rho}}$  is irreducible, by the Perron–Frobenius theorem, the maximum eigenvalue corresponds to a one–dimensional space and therefore cannot be orthogonal to  $\hat{\rho}$ . It follows e.g., from the version of the above exhibited in (2.4) that the maximum eigenvalue of  $B_{\hat{\rho}}$  (or  $B_{\hat{\rho}}^T$ ) is strictly less than 1.

## 2.2. Main Theorems

To state the first theorem, we let  $S$  be the entropy function given by

$$S(\hat{\eta}, \hat{\rho}) = \sum_{k=1}^L (\hat{\rho}_k \log \hat{\rho}_k - \hat{\eta}_k \log \hat{\eta}_k - (\hat{\rho}_k - \hat{\eta}_k) \log(\hat{\rho}_k - \hat{\eta}_k)), \quad (2.11)$$

and let

$$\Xi(\hat{\eta}) = \sum_{k=1}^L \hat{\eta}_k \log \left( 1 - e^{-\sum_{i=1}^L \alpha_{ik} \hat{\eta}_i} \right). \quad (2.12)$$

With these, we can calculate the large deviation rate of large components existing in  $\mathcal{G}(n, \hat{\rho})$ :

**Theorem 2.5.** *Let  $L, A,$  and  $\hat{\rho}$  be given, and for each  $n$  let  $v$  be a random vertex of  $\mathcal{G}(n, \hat{\rho})$  chosen by any distribution desired which is independent from  $\mathcal{E}$ . Also let  $\mathcal{S} \subset \prod_i [0, \hat{\rho}_i]$  be a set of densities which is open in  $\prod_i [0, \hat{\rho}_i]$ . Then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{n, \hat{\rho}} \left( \frac{1}{n} \langle C(v) \rangle \in \mathcal{S} \right) = \sup_{\hat{\eta} \in \mathcal{S}} [S(\hat{\eta}, \hat{\rho}) + \Xi(\hat{\eta}) - \hat{\eta}^T A(\hat{\rho} - \hat{\eta})]. \tag{2.13}$$

We are also interested in the rate of  $\mathcal{G}(n, \hat{\rho})$  not having large components:

**Theorem 2.6.** *Let  $L, A,$  and  $\hat{\rho} > 0$  be given. We define the continuous function*

$$\Psi(\hat{\rho}) = \begin{cases} \sum_{i=1}^L \left[ \hat{\rho}_i \log \frac{\hat{\rho}_i}{\hat{\rho}_i^*} - \frac{1}{2} (\hat{\rho}_i - \hat{\rho}_i^*) \left( 1 + \sum_{j=1}^L \alpha_{ij} \hat{\rho}_j \right) \right] & \text{if } \hat{\rho} \text{ is supercritical} \\ 0 & \text{otherwise} \end{cases}, \tag{2.14}$$

where for  $\hat{\rho}$  supercritical,  $\hat{\rho}^*$  is the unique critical density such that  $(I - B_{\hat{\rho}^*}^T)(\hat{\rho} - \hat{\rho}^*) = 0$ . Then if  $\mathbb{S}_r$  is the event that all components in  $\mathcal{G}(\hat{\rho}, n)$  are of size smaller than  $r$ , we have

$$\begin{aligned} \lim_{r \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \log P_{n, \hat{\rho}}(\mathbb{S}_r) &= \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log P_{n, \hat{\rho}}(\mathbb{S}_{\epsilon n}) \\ &= \Psi(\hat{\rho}). \end{aligned} \tag{2.15}$$

Moreover, the convergence is uniform for  $\hat{\rho}$  bounded above.

As a kind of combination of Theorem 2.5 and Theorem 2.6, we find the exponential rate of  $\hat{\theta}$  taking a specified value:

**Theorem 2.7.** *Let  $L, A,$  and  $\hat{\rho} > 0$  be given, and let  $\mathcal{S} \subset \prod_i [0, \hat{\rho}_i]$  be a set of densities open in  $\prod_i [0, \hat{\rho}_i]$ . Then*

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log P(\hat{\theta}(\epsilon n, n, \hat{\rho}) \in \mathcal{S}) = \sup_{\hat{\eta} \in \mathcal{S}} [S(\hat{\eta}, \hat{\rho}) + \Xi(\hat{\eta}) + \Psi(\hat{\rho} - \hat{\eta}) - \hat{\eta}^T A(\hat{\rho} - \hat{\eta})]. \tag{2.16}$$

Finally, we supplement the results in [3] (Theorems 3.1 and 3.6) concerning the uniqueness of the giant component:

**Theorem 2.8.** *Let  $L, A,$   $\hat{\rho} > 0,$  and  $\delta > 0$  be given. Let  $Q_\epsilon$  be the event that  $\mathcal{G}(n, \hat{\rho})$  has two or more components of size at least  $\epsilon n$ . Then*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_{n, \hat{\rho}}(Q_\epsilon \mid \hat{\theta}(\epsilon n, n, \hat{\rho}) \in [\delta, 1]^L) < 0. \tag{2.17}$$

Thus the giant component is unique with exponential probability.

### 3. ALL SITES CONNECTED

The goal of this section is to prove Theorem 2.5 and the following:

**Theorem 3.1.** *Let  $L$  and  $A$  be given, and let  $K_m$  be the event that  $\mathcal{G}(n, \hat{\rho})$  has  $m$  or fewer components. Then for each  $m \geq 1$  and  $\hat{\rho} > 0$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{n, \hat{\rho}}(K_m) = \Xi(\hat{\rho}), \tag{3.1}$$

and the convergence is uniform for  $\hat{\rho}$  bounded above and bounded away from zero.

Notice that if  $m$  is chosen to be 1, Theorem 3.1 gives the exponential rate of  $\mathcal{G}(n, \hat{\rho})$  being connected.

In order to prove Theorem 3.1, we will convert the problem into an equivalent problem for directed graphs. To this end, for all vertex sets  $\mathcal{U}$  and  $\mathcal{W}$ , and all directed edge sets  $\vec{\mathcal{E}}$ , let  $F(\mathcal{U}, \mathcal{W}, \vec{\mathcal{E}})$  be the event that for every vertex in  $\mathcal{U}$  there is a path of (strictly) positive length in  $\vec{\mathcal{E}}$  to a vertex in  $\mathcal{W}$ .

**Lemma 3.2.** *Let vertex sets  $\mathcal{U}$  and  $\mathcal{W}$ , and  $(p_v \in [0, 1])_{v \in \mathcal{U} \cup \mathcal{W}}$  be given with  $\sum_{v \in \mathcal{U} \cup \mathcal{W}} p_v = 1$ . Let  $(\omega_v)_{v \in \mathcal{U}}$  be i.i.d. random elements of  $\mathcal{U} \cup \mathcal{W}$  with*

$$P(\omega_v = u) = p_u. \tag{3.2}$$

Then let  $\vec{\mathcal{E}}$  be the random edge set  $\vec{\mathcal{E}} = \{(v, \omega_v) : v \in \mathcal{U}\}$ . Then

$$P[F(\mathcal{U}, \mathcal{W}, \vec{\mathcal{E}})] = \sum_{v \in \mathcal{W}} p_v. \tag{3.3}$$

*Proof.* Given our construction of  $\vec{\mathcal{E}}$  (in which each vertex has an outgoing edge), we have that

$$F(\mathcal{U}, \mathcal{W}, \vec{\mathcal{E}}) = F(\mathcal{U} \setminus \mathcal{W}, \mathcal{W}, \vec{\mathcal{E}}). \tag{3.4}$$

Indeed, in the left hand event each vertex in  $\mathcal{W}$  is either directly connected to  $\mathcal{W}$ , or is indirectly connected to  $\mathcal{W}$  through  $\mathcal{U} \setminus \mathcal{W}$ . Thus we may assume without loss of generality that  $\mathcal{U}$  and  $\mathcal{W}$  are disjoint. We will proceed by induction on  $|\mathcal{U}|$ ; the size of  $\mathcal{W}$  is unimportant. (Indeed the system is manifestly equivalent to the one where  $\mathcal{W} = \{w\}$  with  $p_w$  given by the right hand side of Eq. (3.3)). The case with  $|\mathcal{U}| = 1$  is trivial, so we assume that the lemma holds for  $|\mathcal{U}| \leq N$ , and take  $|\mathcal{U}| = N + 1$ . Let  $u \in \mathcal{U}$ , and note that if  $\omega_u = u$  then  $F(\mathcal{U}, \mathcal{W}, \vec{\mathcal{E}})$  does not occur. On the other hand – that is, if  $\omega_u \neq u$  – we let

$$\omega'_v = \begin{cases} \omega_u & \text{if } \omega_v = u \\ \omega_v & \text{otherwise} \end{cases}, \tag{3.5}$$

for each  $v \in \mathcal{U} \setminus \{u\}$ , and let  $\vec{\mathcal{E}}' = \{(v, \omega'_v) : v \in \mathcal{U} \setminus \{u\}\}$ . Then every path in  $\vec{\mathcal{E}}$  which does not start or end with  $u$  has a naturally corresponding path in  $\vec{\mathcal{E}}'$ : simply remove all instances of  $u$  from the path. By this correspondence, we see that  $F(\mathcal{U} \setminus \{u\}, \mathcal{W}, \vec{\mathcal{E}}')$  occurs if and only if  $F(\mathcal{U}, \mathcal{W}, \vec{\mathcal{E}})$  occurs. We note further that in the case  $\omega_u \in \mathcal{W}$  we have  $P(\omega'_v \in \mathcal{W}) = p_u + \sum_{y \in \mathcal{W}} p_y$  for  $v \in \mathcal{U} \setminus \{u\}$ .

Thus we have

$$\begin{aligned}
 P[F(\mathcal{U}, \mathcal{W}, \vec{\mathcal{E}})] &= \sum_{y \in \mathcal{W}} p_y P[F(\mathcal{U} \setminus \{u\}, \mathcal{W}, \vec{\mathcal{E}}') \mid \omega_u = y] \\
 &+ \sum_{y \in \mathcal{U} \setminus \{u\}} p_y P[F(\mathcal{U} \setminus \{u\}, \mathcal{W}, \vec{\mathcal{E}}') \mid \omega_u = y].
 \end{aligned}
 \tag{3.6}$$

Since  $(\omega'_v)_{v \in \mathcal{U} \setminus \{u\}}$  are i.i.d. once  $\omega_u$  is conditioned upon, our induction assumption gives that

$$P[F(\mathcal{U} \setminus \{u\}, \mathcal{W}, \vec{\mathcal{E}}') \mid \omega_u = y] = \begin{cases} p_u + \sum_{v \in \mathcal{W}} p_v & \text{if } y \in \mathcal{W} \\ \sum_{v \in \mathcal{W}} p_v & \text{if } y \in \mathcal{U} \setminus \{u\} \end{cases}.
 \tag{3.7}$$

Substituting these into (3.6) gives the result for  $|\mathcal{U}| = N + 1$ , and completes the inductive step of the proof. ■

**Lemma 3.3.** *Let  $L, A$ , and  $\hat{\rho} > 0$  be given, and let  $\vec{\mathcal{E}}$  be the random directed edge set on vertex set  $\mathcal{V}$  so that each directed edge  $(u, v)$  appears (or not) in  $\vec{\mathcal{E}}$  independently with  $P((u, v) \in \vec{\mathcal{E}}) = p_{\text{layer}(u)\text{layer}(v)}$ . Then, letting  $\mathbb{P}_{n, \hat{\rho}}$  be the probability measure associated with  $\vec{\mathcal{E}}$ , we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_{n, \hat{\rho}}[F(\mathcal{V}, \{v\}, \vec{\mathcal{E}})] = \Xi(\hat{\rho}) = \lim_{n \rightarrow \infty} \mathbb{P}_{n, \hat{\rho}}[F(\mathcal{V}, \mathcal{V}, \vec{\mathcal{E}})]^{1/n}
 \tag{3.8}$$

for all  $v \in \mathcal{V}$ . Moreover, if  $\hat{\rho}$  is bounded above and bounded away from zero, the convergence is uniform.

*Proof.* We will prove the lemma by showing

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n, \hat{\rho}}[F(\mathcal{V}, \{v\}, \vec{\mathcal{E}})]^{1/n} = \lim_{n \rightarrow \infty} \mathbb{P}_{n, \hat{\rho}}[F(\mathcal{V}, \mathcal{V}, \vec{\mathcal{E}})]^{1/n}.
 \tag{3.9}$$

Indeed, once this is established, the existence and value of the mutual limit is clear. In particular, the event on the right hand side is the event that each vertex in  $\mathcal{V}$  has an outgoing edge in  $\vec{\mathcal{E}}$ ; for finite  $n$  this may be calculated explicitly:

$$\mathbb{P}_{n, \hat{\rho}}[F(\mathcal{V}, \mathcal{V}, \vec{\mathcal{E}})] = \prod_{k=1}^L \left( 1 - \prod_{i=1}^L \left( 1 - \frac{\alpha_{ik}}{n} \right)^{\lfloor \hat{\rho}_i n \rfloor} \right)^{\lfloor \hat{\rho}_k n \rfloor}.
 \tag{3.10}$$

and the value of the limit follows.

We turn to the substantive task which is to establish Eq. (3.9). First, note that since  $F(\mathcal{V}, \mathcal{W}, \vec{\mathcal{E}})$  is increasing in  $\mathcal{W}$ , in the sense that  $F(\mathcal{V}, \mathcal{W}, \vec{\mathcal{E}}) \subset F(\mathcal{V}, \mathcal{U}, \vec{\mathcal{E}})$  whenever  $\mathcal{W} \subset \mathcal{U}$ . In particular, we have

$$\mathbb{P}_{n, \hat{\rho}}[F(\mathcal{V}, \{v\}, \vec{\mathcal{E}})] \leq \mathbb{P}_{n, \hat{\rho}}[F(\mathcal{V}, \mathcal{V}, \vec{\mathcal{E}})].
 \tag{3.11}$$

To get a lower bound, we let  $\vec{\mathcal{F}}$  be a subset of  $\vec{\mathcal{E}}$  gotten by discarding all but one outgoing edge (randomly selected) from each vertex. Then let  $\tilde{\mathbb{P}}(\cdot) = \mathbb{P}_{n, \hat{\rho}}[\cdot \mid F(\mathcal{V}, \mathcal{V}, \vec{\mathcal{E}})]$ , and note that since  $F(\mathcal{V}, \{v\}, \vec{\mathcal{E}})$  is an increasing event with respect to  $\vec{\mathcal{E}}$ , we have

$$\tilde{\mathbb{P}}[F(\mathcal{V}, \{v\}, \vec{\mathcal{E}})] \geq \tilde{\mathbb{P}}[F(\mathcal{V}, \{v\}, \vec{\mathcal{F}})].
 \tag{3.12}$$



We also note that under the measure  $\tilde{\mathbb{P}}$ , the set  $\vec{\mathcal{F}}$  contains exactly one outgoing edge for each vertex. We let

$$b_{ij} = \lim_{n \rightarrow \infty} \tilde{\mathbb{P}}[(x, y) \in \vec{\mathcal{F}} \text{ for some } y \in \mathcal{V}_j] \tag{3.13}$$

where  $x \in \mathcal{V}_i$ . Then, as is readily estimated using the limiting Poisson statistics,

$$b_{ij} \geq \frac{[1 - e^{-\alpha_{ij}\hat{\rho}_j}]e^{-\sum_{k \neq j} \alpha_{ik}\hat{\rho}_k}}{1 - e^{-\sum_k \alpha_{ik}\hat{\rho}_k}}. \tag{3.14}$$

Specifically, if the probability of  $x$  having an edge in  $\vec{\mathcal{E}}$  to a given layer is uniformly bounded away from zero, the probability of  $x$  having such an edge in  $\vec{\mathcal{F}}$  is also uniformly bounded away from zero.

We will now use induction on  $L$  to show that

$$\tilde{\mathbb{P}}[F(\mathcal{V}, \{v\}, \vec{\mathcal{F}})] = c/n + o(1/n), \tag{3.15}$$

for some constant  $c > 0$  depending on  $L, A$ , and  $\hat{\rho}$ . If  $L = 1$ , this is a straightforward application of Lemma 3.2. We then assume that (3.15) holds for  $L \leq M$  and take  $L = M + 1$ . Assume without loss of generality that  $v$  is not in layer  $M + 1$  and apply Lemma 3.2 to conclude

$$\tilde{\mathbb{P}} \left[ F \left( \mathcal{V}_{M+1}, \bigcup_{i=1}^M \mathcal{V}_i, \vec{\mathcal{F}} \right) \right] = 1 - b_{(M+1)(M+1)} + o(1). \tag{3.16}$$

Here the fact that  $A$  is irreducible (and  $\hat{\rho} > 0$ ) implies that  $1 - b_{(M+1)(M+1)} > 0$ . Next, if  $F(\mathcal{V}_{M+1}, \bigcup_{i=1}^M \mathcal{V}_i, \vec{\mathcal{F}})$  occurs, we let  $\vec{\mathcal{F}}'$  be a copy of  $\vec{\mathcal{F}}$  in which the edges terminating in  $\mathcal{V}_{M+1}$  are remapped to edges terminating in  $\bigcup_{i=1}^M \mathcal{V}_i$  à la the proof of Lemma 3.2. The rates of connections in  $\vec{\mathcal{F}}'$  are given by  $(b'_{ij})$ , where  $b'_{ij} = b_{ij} + b_{i(M+1)}b_{(M+1)j}$  for  $i, j \leq M$ . Once again  $A$  being irreducible implies  $(b'_{ij})$  must be as well. Thus we may use our inductive assumption to prove that (3.15) holds for all  $L$ .

Now using (3.12) and taking limits we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n, \hat{\rho}}[F(\mathcal{V}, \{v\}, \vec{\mathcal{E}})]^{1/n} \geq \lim_{n \rightarrow \infty} \mathbb{P}_{n, \hat{\rho}}[F(\mathcal{V}, \mathcal{V}, \vec{\mathcal{E}})]^{1/n}, \tag{3.17}$$

which proves (3.9); and since nonzero  $b_{ij}$  are uniformly bounded away from zero for  $\hat{\rho}$  bounded above and bounded away from zero, we have uniform convergence. ■

We will also use Lemma 3.2 from [4], which we reproduce here (with slightly modified notation):

**Lemma 3.4.** *For a collection of vertices  $\mathcal{W} = \{1, \dots, n\}$  with an associated set of edge probabilities  $(p_{kl})_{1 \leq k < l \leq n}$ , let  $\mathcal{G}$  be the inhomogeneous undirected random graph over  $\mathcal{W}$ . Similarly, let  $\vec{\mathcal{G}}$  denote the inhomogeneous directed graph with the restriction that the two possible (directed) edges between  $k$  and  $l$  occur independently, each with probability  $p_{kl}$ . Letting  $\mathcal{E}_{\vec{\mathcal{G}}}$  be the edge set of  $\vec{\mathcal{G}}$ , we have  $P(\mathcal{G} \text{ is connected}) = P(F(\mathcal{W} \setminus \{1\}, \{1\}, \mathcal{E}_{\vec{\mathcal{G}}}))$ .*

*Proof of Theorem 3.1.* For a vertex set  $\mathcal{W}$ , let  $\mathcal{E}^+(\mathcal{W})$  be the complete set of edges on  $\mathcal{W}$ . We then note for  $n \geq m$

$$K_m = \bigcup_{\substack{\mathcal{W} \subset \mathcal{V} \\ |\mathcal{W}|=m}} \{\mathcal{G}(n, \hat{\rho}) \cup \mathcal{E}^+(\mathcal{W}) \text{ is connected}\}, \tag{3.18}$$

and thus

$$P_{n,\hat{\rho}}(K_m) = n^{O(m)} \sup_{\substack{\mathcal{W} \subset \mathcal{V} \\ |\mathcal{W}|=m}} P_{n,\hat{\rho}}(\mathcal{G}(n, \hat{\rho}) \cup \mathcal{E}^+(\mathcal{W}) \text{ is connected}). \tag{3.19}$$

Moreover, since the vertices are a priori interchangeable within a given layer, we need only consider a finite number of  $\mathcal{W}$  in the supremum. Thus we will have proved the theorem once we have shown that  $\frac{1}{n} \log P_{n,\hat{\rho}}(\mathcal{G}(n, \hat{\rho}) \cup \mathcal{E}^+(\mathcal{W}) \text{ is connected})$  converges uniformly to  $\Xi(\hat{\rho})$  for  $\hat{\rho}$  bounded above and bounded away from zero. Thus let us fix  $\mathcal{S} \subset \mathcal{V}$  with  $|\mathcal{S}| = m$ , and let  $v \in \mathcal{S}$ . Since adding an edge  $(a, b)$  to  $\mathcal{G}(n, \hat{\rho})$  is equivalent to setting  $p_{ab} = 1$ , and since  $F(\mathcal{V} \setminus \{v\}, \{v\}, \vec{\mathcal{E}} \cup \mathcal{E}^+(\mathcal{S})) = F(\mathcal{V} \setminus \mathcal{S}, \mathcal{S}, \vec{\mathcal{E}})$  – where  $\vec{\mathcal{E}}^+$  is the full set of directed edges on  $\mathcal{S}$  – Lemma 3.4 shows that

$$P_{n,\hat{\rho}}(\mathcal{G}(n, \hat{\rho}) \cup \mathcal{E}^+(\mathcal{S}) \text{ is connected}) = \mathbb{P}_{n,\hat{\rho}}[F(\mathcal{V} \setminus \mathcal{S}, \mathcal{S}, \vec{\mathcal{E}})]. \tag{3.20}$$

Then we note

$$F(\mathcal{V}, \mathcal{S}, \vec{\mathcal{E}}) = F(\mathcal{V} \setminus \mathcal{S}, \mathcal{S}, \vec{\mathcal{E}}) \bigcap F(\mathcal{S}, \mathcal{V}, \vec{\mathcal{E}}). \tag{3.21}$$

and point out that events on the right hand side are independent. Since the rightmost event occurs with (uniform) probability of order unity and the event on the left hand side is bounded between  $F(\mathcal{V}, \{v_1\}, \vec{\mathcal{E}})$  and  $F(\mathcal{V}, \mathcal{V}, \vec{\mathcal{E}})$ , we can use Lemma 3.3 to conclude

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_{n,\hat{\rho}}[F(\mathcal{V} \setminus \mathcal{S}, \mathcal{S}, \vec{\mathcal{E}})] = \Xi(\hat{\rho}). \tag{3.22}$$

Since this holds for all choices of  $\mathcal{S}$ , and the convergence is uniform, we use (3.20) and (3.19) to finish the proof. ■

**Corollary 3.5.** *Let  $m \geq 1$  and  $\hat{\rho}^{(0)}$  be given. Then,*

$$\frac{1}{n} \log P_{n,\hat{\rho}}(K_m) \leq \Xi(\hat{\rho}) + o(1), \tag{3.23}$$

where the  $o(1)$  term is uniformly bounded for  $(1/n, \dots, 1/n) \leq \hat{\rho} \leq \hat{\rho}^{(0)}$ . Furthermore, if  $m = 1$  then the  $o(1)$  term is uniformly bounded for  $0 \leq \hat{\rho} \leq \hat{\rho}^{(0)}$ .

*Proof.* We first note that the lower bound on  $\hat{\rho}$  is chosen to guarantee that each layer has at least one vertex. For  $m > 1$ , this guarantees that no vertices are strictly isolated. For  $m = 1$  such a restriction is unnecessary – indeed, the existence of vertices which are isolated would cause the left hand side of (3.23) to be negative infinity. Thus for the remainder of the proof we assume that each vertex has at least one potential neighbor.

Following the proof of Theorem 3.1 up to around (3.19), we get

$$\frac{1}{n} \log P_{n,\hat{\rho}}(K_m) = o(1) + \sup_{\substack{\mathcal{W} \subset \mathcal{V} \\ |\mathcal{W}|=m}} \frac{1}{n} \log P_{n,\hat{\rho}}(\mathcal{G}(n, \hat{\rho}) \cup \mathcal{E}^+(\mathcal{W}) \text{ is connected}), \tag{3.24}$$

where  $o(1)$  is independent of  $\hat{\rho}$ . Then taking  $\mathcal{S}$  as in the proof of Theorem 3.1, we have by (3.21)

$$\mathbb{P}_{n,\hat{\rho}}[F(\mathcal{V} \setminus \mathcal{S}, \mathcal{S}, \vec{\mathcal{E}})] \leq O(n)^m \prod_{k=1}^L \left( 1 - \prod_{i=1}^L \left( 1 - \frac{\alpha_{ik}}{n} \right)^{\lfloor \hat{\rho}_i n \rfloor} \right)^{\lfloor \hat{\rho}_k n \rfloor}, \tag{3.25}$$

where the  $O(n)^m$  term comes from  $\mathbb{P}_{n,\hat{\rho}}[F(\mathcal{S}, \mathcal{V}, \vec{\mathcal{E}})]$  – here we have used the assumption that each vertex has at least one potential neighbor – and the remainder of the right hand side comes from bounding the event  $F(\mathcal{V}, \mathcal{S}, \vec{\mathcal{E}})$  by  $F(\mathcal{V}, \mathcal{V}, \vec{\mathcal{E}})$  and using (3.10). The result then follows by applying (3.20) and taking limits. ■

*Proof of Theorem 2.5.* We note that for all  $\hat{\eta}$  with  $0 \leq \hat{\eta} \leq \hat{\rho}$  we have

$$P_{n,\hat{\rho}}(\langle C(v) \rangle = \lfloor \hat{\eta} n \rfloor) = \prod_{i=1}^L \binom{\lfloor \hat{\rho}_i n \rfloor}{\lfloor \hat{\eta}_i n \rfloor} P_{n,\hat{\eta}}(K_1) \prod_{i,j=1}^L \left( 1 - \frac{\alpha_{ij}}{n} \right)^{\lfloor \hat{\eta}_i n \rfloor (\lfloor \hat{\rho}_j n \rfloor - \lfloor \hat{\eta}_j n \rfloor)}. \tag{3.26}$$

Thus for a lower bound we note that for each  $\hat{\eta} \in \mathcal{S}$  we have  $\frac{1}{n} \lfloor \hat{\eta} n \rfloor \in \mathcal{S}$  for sufficiently large  $n$  – following from  $\mathcal{S}$  open. Thus for each  $\hat{\eta} \in \mathcal{S}$  we use (3.26) and Theorem 3.1 to get

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{n,\hat{\rho}} \left( \frac{1}{n} \langle C(v) \rangle \in \mathcal{S} \right) \geq S(\hat{\eta}, \hat{\rho}) + \Xi(\hat{\eta}) - \hat{\eta}^T \mathbf{A}(\hat{\rho} - \hat{\eta}). \tag{3.27}$$

For an upper bound, note that  $\langle C(v) \rangle$  can take only polynomially many values, and thus

$$P_{n,\hat{\rho}} \left( \frac{1}{n} \langle C(v) \rangle \in \mathcal{S} \right) = e^{o(n)} \sup_{\hat{\eta} \in \mathcal{S}} P_{n,\hat{\rho}}(\langle C(v) \rangle = \lfloor \hat{\eta} n \rfloor). \tag{3.28}$$

Now using (3.26) and Corollary 3.5, we get an upper bound which matches (3.27). This proves the Theorem. ■

## 4. NON-PERCOLATING SUPERCRITICALITY

### 4.1. Overview

In order to prove Theorems 2.6 & 2.7, (as well as Theorem 2.1) we will have to consider certain paths in density parameter space; roughly speaking, these represent progress in the *reduction* of the density by the extraction of existing components. In this subsection, we will first define the relevant sorts of paths and introduce a cost function for motion along these paths. Our primary result of this section – Theorem 4.2 – relates the large deviation rate of density reduction to the optimal cost among all paths which connect the densities. We shall turn to some preliminary definitions that will culminate in a statement of Theorem 4.2 – after which we can conclude this overview.

Let  $\hat{\Gamma} : [0, \infty) \mapsto [0, \infty)^L$  denote a path in density parameter space. Of exclusive interest will be *ascents*, which are Lipschitz continuous paths emanating from the origin and which are nondecreasing in all components and increasing in at least one component. For convenience, we shall parameterize the ascents in such a way that  $|\hat{\Gamma}(t)| = t$  for all  $t > 0$ . Finally, we denote by  $\mathcal{P}(\hat{\eta})$  the set of ascents that pass through the point  $\hat{\eta}$ .

For  $\hat{\gamma}$  a vector with  $|\hat{\gamma}| = 1$  and  $t \geq 0$  let us define

$$\frac{1}{\chi(\hat{\gamma}, t)} = \min \left\{ x \in [0, 1] \left| \sum_{i=1}^L \frac{\hat{\gamma}_i}{x + t \sum_{j=1}^L \alpha_{ij} \hat{\gamma}_j} \leq 1 \right. \right\} \tag{4.1}$$

where, for historical reasons the above is denoted by a reciprocal quantity. If the minimum is achieved at  $x = 0$  we will, somewhat informally, declare  $\chi$  to be infinite.

Next, we define  $\hat{\psi}$  by

$$\hat{\psi}_i(\hat{\gamma}, t) = \frac{\hat{\gamma}_i t}{1/\chi(\hat{\gamma}, t) + t \sum_{j=1}^L \alpha_{ij} \hat{\gamma}_j} \tag{4.2}$$

It is noted that at  $t = 0$ , we have  $\chi = 1$  (and  $\hat{\psi} = 0$ ). Thus by continuity, for  $t$  small enough,  $[\chi]^{-1} > 0$  (with  $\hat{\psi}$  non-trivial). However, under most circumstances, it will be the case that for  $t \gg 1$ , the only choice is  $[\chi]^{-1} = 0$ . Here we let  $t^*$  denote the supremum of  $t$ 's for which  $\chi < \infty$  – where it is noted, by obvious monotonicity considerations, that  $t^*$  is well defined although possibly infinite. Assuming otherwise (since  $t^* = \infty$  represents a trivial problem) we have for  $t > t^*$ , the resultant  $\hat{\psi}$  is independent of  $t$ . We claim that for  $t < t^*$  the quantities  $(\hat{\psi}, \chi \hat{\gamma})$  represent a (density, average cluster size) pairing in the subcritical regime while for  $t \geq t^*$  the model with density  $\hat{\psi}$  is critical with a corresponding interpretation. This is the subject of our next Lemma.

**Lemma 4.1.** *If  $v$  is a vertex chosen uniformly at random from  $\mathcal{V}$  then*

$$\chi(\hat{\gamma}, t) = \lim_{n \rightarrow \infty} E_{n, \hat{\psi}(\hat{\gamma}, t)}[|C(v)|] \tag{4.3}$$

with both sides infinity if  $t \geq t^*$ . Further, when  $t < t^*$ ,

$$\lim_{n \rightarrow \infty} \frac{E_{n, \hat{\psi}(\hat{\gamma}, t)}[|C(v)|]}{E_{n, \hat{\psi}(\hat{\gamma}, t)}[|C(v)|]} = \hat{\gamma}. \tag{4.4}$$

Finally, for  $t \geq t^*$ , the model defined by  $B_{\hat{\psi}}$  is critical and  $\hat{\gamma}$  is precisely the maximum eigenvector of  $B_{\hat{\psi}}^T$  (with eigenvalue unity).

*Proof.* Let us first consider the  $t < t^*$  – subcritical – cases. The goal is to show that with  $\hat{\rho} = \hat{\psi}$  and  $\hat{v} = \chi \hat{\gamma}$ , equation (2.10) is satisfied. For ease of exposition, let us temporarily denote  $\hat{w}_i := \sum_j \alpha_{ij} \hat{\gamma}_j$ . Then

$$[B_{\hat{\psi}}^T \chi \hat{\gamma}]_i = \sum_{j=1}^L \hat{\psi}_i \alpha_{ij} \chi \hat{\gamma}_j = \frac{t \gamma_i}{[\chi]^{-1} + t \hat{w}_i} \chi \hat{w}_i.$$

On the other hand,  $\chi \hat{\gamma} - |\hat{\psi}|^{-1} \hat{\psi}$  is given, componentwise, by

$$\chi \hat{\gamma}_i - |\hat{\psi}|^{-1} \hat{\psi}_i = \frac{\chi \gamma_i ([\chi]^{-1} + t \hat{w}_i) - \gamma_i}{[\chi]^{-1} + t \hat{w}_i}$$

which is manifestly the same. Thus, as far as subcritical cases are concerned, via Proposition 2.4, all claims have been vindicated.

For  $t > t^*$  the above is formally true with  $\chi = \infty$ . Notwithstanding, let us first pause to note that if there is any  $\hat{w}_i = 0$ , with  $\hat{\gamma}_i \neq 0$  we can *never* get to criticality while if both  $\gamma_i$  and  $\hat{w}_i$  are zero, the correct interpretation of the ratio (from the subcritical limit) is zero. Proceeding, we have

$$[\mathbf{B}_{\hat{\psi}}^T \hat{\gamma}]_i = \sum_{j=1}^L \hat{\psi}_i \alpha_{ij} \hat{\gamma}_j = \frac{\hat{\gamma}_i}{\hat{w}_i} \cdot \hat{w}_i = \hat{\gamma}_i$$

In light of the above, we may associate with (almost) every point on an ascending path the critical or subcritical density  $\hat{\psi}(\hat{\Gamma}'(t)/|\hat{\Gamma}'(t)|, |\hat{\Gamma}(t)|)$ ; although due to our normalization this simplifies, a.e., to  $\hat{\psi}(\hat{\Gamma}'(t), t)$ . We next set

$$\xi(\hat{\gamma}, \hat{\eta}) = \sum_{i=1}^L \hat{\gamma}_i \log \hat{\eta}_i - \hat{\gamma}^T \mathbf{A} \hat{\eta}. \tag{4.5}$$

and on the basis of  $\xi$  and  $\hat{\psi}$  we define a free-energy like object for ascents:

$$H(\hat{\Gamma}, a, b) = \int_a^b \left( \xi(\hat{\Gamma}'(t), \hat{\Gamma}(t)) - \xi(\hat{\Gamma}'(t), \hat{\psi}(\hat{\Gamma}'(t), t)) \right) dt. \tag{4.6}$$

Most often, we will be interested in maximizing  $H$  over paths with fixed endpoints  $\hat{\Gamma}(a), \hat{\Gamma}(b)$ , which will be clear from context and suppressed from our notation.

Finally, let  $C^{(1)}, C^{(2)}, \dots$  be the components of  $\mathcal{G}(n, \hat{\rho})$  randomly ordered according to size bias: I.e.  $C^{(1)}$  is sampled by choosing a vertex uniformly and removing its entire component and, in general,  $C^{(k)}$  is sampled by removing the component of a vertex chosen uniformly from the remaining graph after  $C^{(1)}, \dots, C^{(k-1)}$  has been removed. Then for  $0 \leq \hat{\eta} \leq \hat{\rho}$ , let

$$\Upsilon(\hat{\rho}, \hat{\eta}, r, n) = P_{n, \hat{\rho}} \left( \exists k : \sum_{i=1}^k |C^{(i)}| = \lfloor \hat{\rho} n \rfloor - \lfloor \hat{\eta} n \rfloor, |C^{(i)}| \leq r \forall i \leq k \right). \tag{4.7}$$

Put simply,  $\Upsilon(\hat{\rho}, \hat{\eta}, r, n)$  is the probability that, starting with the system at parameter  $\hat{\rho}$  if we pluck out components at random we arrive at the system with parameter  $\hat{\eta}$  (and do so without ever having selected a component of size larger than  $r$ ).

We are now in a position to state the following, which is the central object of this section:

**Theorem 4.2.** *Let  $\hat{\rho}$  and  $\hat{\eta}$  be given with  $\hat{\rho} \geq \hat{\eta} \geq 0$ . Then*

$$\lim_{r \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \log \Upsilon(\hat{\rho}, \hat{\eta}, r, n) = \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log \Upsilon(\hat{\rho}, \hat{\eta}, \epsilon n, n) = \sup_{\hat{\Gamma} \in \mathcal{P}(\hat{\eta}) \cap \mathcal{P}(\hat{\rho})} H(\hat{\Gamma}, |\hat{\eta}|, |\hat{\rho}|). \tag{4.8}$$

*Moreover, the convergence is uniform for  $\hat{\rho}$  bounded above.*

Theorem 4.2 has an interpretation that is not without appeal. Assume that  $|\hat{\Gamma}(a) - \hat{\Gamma}(b)| \ll 1$  so that we may envision the integrand for a minimizing path of  $H(\hat{\Gamma}, a, b)$  as essentially linear. The cost in (4.6) is now seen as  $[\hat{\Gamma}(b) - \hat{\Gamma}(a)]$  times the probability – measured on the exponential scale – of observing the system at parameter  $\hat{\psi}$  if the actual system has parameter  $\hat{\Gamma}$ . Then,  $\hat{\psi}$  is chosen so that in a typical selection of its clusters, the

size ratio is such that, typically, when these clusters are selected, the density decreases in the (desired) direction of  $\hat{\Gamma}(a) - \hat{\Gamma}(b)$ .

We now proceed with the overview: In the next subsection, we will present a variety of results which concern the distribution of cluster sizes in these systems and in Subsection 4.3 these will be assembled into a proof of Theorem 4.2. In Subsection 4.4, we will study the  $H$ -functional in its own right. In particular, we will define paths called *natural ascents* which characterize the maximizers of the cost-functional. Subsection 4.5, noted for its brevity, will contain the proofs of Theorems 2.6 & 2.7 as well as Theorem 2.1.

### 4.2. Cluster Distributions

**Lemma 4.3.** *Let  $A, L$ , and  $\hat{\rho}^{(0)}$  be given, and for each  $\hat{\rho}$ , let  $p(\hat{k}, \hat{\rho}, n) = P_{n, \hat{\rho}}(\langle C(v) \rangle = \hat{k})$ , where  $v$  is a vertex chosen uniformly at random from  $\mathcal{G}(n, \hat{\rho})$ . Then there exists a  $c = c(\hat{\rho}^{(0)})$  and a  $p(\hat{k}, \hat{\rho})$  – which may be acquired as the  $n \rightarrow \infty$  limit of  $p(\hat{k}, \hat{\rho}, n)$  – such that for all  $\hat{\rho} \leq \hat{\rho}^{(0)}$ ,*

$$\exp\left(-c \sum_{i=1}^L \frac{\hat{k}_i^2}{(1 \wedge \hat{\rho}_i)n}\right) \leq \frac{p(\hat{k}, \hat{\rho}, n)}{p(\hat{k}, \hat{\rho})} \leq \exp\left(c(|\hat{k}|^2/n) + \frac{c}{|\hat{\rho}|n}\right), \tag{4.9}$$

where the lower bound holds only for  $\hat{k} \leq \hat{\rho}n$ . In this bound,  $\hat{k}_i^2/\hat{\rho}_i$  is considered to be zero if both  $\hat{k}_i$  and  $\hat{\rho}_i$  are zero, and  $p(\hat{k}, \hat{\rho}, n)/p(\hat{k}, \hat{\rho})$  is considered to be one if both  $p(\hat{k}, \hat{\rho}, n)$  and  $p(\hat{k}, \hat{\rho})$  are zero. Furthermore, for all  $\hat{\rho}, \hat{\eta} > 0$ ,

$$\frac{p(\hat{k}, \hat{\rho})}{p(\hat{k}, \hat{\eta})} = \frac{|\hat{\eta}|}{|\hat{\rho}|} \prod_{i=1}^L \left(\frac{\hat{\rho}_i}{\hat{\eta}_i} e^{-\sum_{j=1}^L \alpha_{ij}(\hat{\rho}_j - \hat{\eta}_j)}\right)^{\hat{k}_i}. \tag{4.10}$$

*Proof.* Recall that  $\mathcal{V}$  is the vertex set of  $\mathcal{G}(n, \hat{\rho})$  and let  $\mathcal{W} \subset \mathcal{V}$  with  $|\mathcal{W}| = \hat{k}$  so that

$$g(\hat{k}, \hat{\rho}, n) = \prod_{i,j=1}^L \left(1 - \frac{\alpha_{ij}}{n}\right)^{\hat{k}_i(\lfloor \hat{\rho}_j n \rfloor - \hat{k}_j)} = e^{-\hat{\rho}^T A \hat{k} + O(|\hat{k}|^2/n)} \tag{4.11}$$

is the probability that  $\mathcal{W}$  is disconnected from  $\mathcal{V} \setminus \mathcal{W}$  in  $\mathcal{G}(n, \hat{\rho})$ . Note that the error here is uniformly bounded for  $\hat{\rho}$  bounded above. Let  $(s_{ij} \in \mathbb{N})_{1 \leq i < j \leq L}$  be given and let  $T$  be a tree on  $\mathcal{W}$  which has  $s_{ij}$  edges between  $\mathcal{W}_i$  and  $\mathcal{W}_j$  for each  $i$  and  $j$ . Then if  $v$  is a vertex uniformly chosen from  $\mathcal{V}$ , we have

$$P(T \text{ spans } C(v)) = \frac{|\hat{k}|}{\sum_{i=1}^L \lfloor \hat{\rho}_i n \rfloor} g(\hat{k}, \hat{\rho}, n) \prod_{1 \leq i < j \leq L} \left(\frac{\alpha_{ij}}{n}\right)^{s_{ij}}, \tag{4.12}$$

where by ‘spans’ we means that  $T$  and  $C(v)$  have the same vertex set, and the edge set of  $T$  is contained in that of  $C(v)$ . Additionally, we have  $\sum_{1 \leq i < j \leq L} s_{ij} = |\hat{k}| - 1$ , so

$$P(C(v) = T | T \text{ spans } C(v)) = \left[ \prod_{i=1}^L \left(1 - \frac{\alpha_{ii}}{n}\right)^{\binom{\hat{k}_i}{2} - s_{ii}} \right] \left[ \prod_{1 \leq i < j \leq L} \left(1 - \frac{\alpha_{ij}}{n}\right)^{\hat{k}_i \hat{k}_j - s_{ij}} \right] \tag{4.13}$$

$$= \exp[O(|\hat{k}|^2/n)] \tag{4.14}$$

whenever  $P(T \text{ spans } C(v)) > 0$ . We note that this error term is independent of  $\hat{\rho}$ .

Now let  $N(\hat{k}, (s_{ij}))$  be the number of trees on  $\hat{k}$  layered vertices that have  $s_{ij}$  edges between layers  $i$  and  $j$  for each  $i$  and  $j$ . We define

$$f(\hat{k}) = \frac{|\hat{k}|}{\prod_{i=1}^L \hat{k}_i!} \sum_{(s_{ij}): \sum_{i \leq j} s_{ij} = |\hat{k}| - 1} N(\hat{k}, (s_{ij})) \prod_{1 \leq i \leq j \leq L} \alpha_{ij}^{s_{ij}}. \tag{4.15}$$

Thus we have

$$p(\hat{k}, \hat{\rho}, n) = \exp[O(|\hat{k}|^2/n) + O(1/(|\hat{\rho}n|))] \left( \prod_{i=1}^L \hat{k}_i! \right) f(\hat{k}) \frac{g(\hat{k}, \hat{\rho}, n)}{|\hat{\rho}n^{|\hat{k}|}} \prod_{i=1}^L \binom{\lfloor \hat{\rho}_i n \rfloor}{\hat{k}_i}, \tag{4.16}$$

where the  $O(1/(|\hat{\rho}n|))$  is non-negative, and comes from replacing  $\sum_i \lfloor \hat{\rho}_i n \rfloor$  with  $|\hat{\rho}n$  in the denominator. For an upper bound, we expand the binomials and bound  $\frac{\lfloor \hat{\rho}_i n \rfloor!}{|\hat{\rho}_i n - \hat{k}_i|!}$  above by  $(\hat{\rho}_i n)^{\hat{k}_i}$ . For a lower bound, we bound this same term below by  $(\hat{\rho}_i n)^{\hat{k}_i} \exp(-O(\hat{k}_i^2/(\hat{\rho}_i n)))$ . Now defining  $p(\hat{k}, \hat{\rho}) = \lim_{n \rightarrow \infty} p(\hat{k}, \hat{\rho}, n)$ , we get both (4.9) and (4.10) (as well as the fact that the limit exists). Note that without knowing  $N(\hat{k}, (s_{ij}))$ , we cannot be more specific about  $p(\hat{k}, \hat{\rho})$  however, for subcritical  $\hat{\rho}$ 's the quantity  $p(\hat{k}, \hat{\rho})$  is a full probability mass function. ■

We now define

$$\mathcal{M}(\hat{\eta}) = \left\{ \mu : \mathbb{N}^L \mapsto [0, \infty); \mu(0, \dots, 0) = 0 \mid \sum_{\hat{k}} \hat{k} \mu(\hat{k}) = \hat{\eta} \right\} \tag{4.17}$$

$$\mathcal{M}(\hat{\eta}, r) = \{ \mu \in \mathcal{M}(\hat{\eta}) \mid \mu(\hat{k}) = 0 \text{ for } |\hat{k}| > r \}. \tag{4.18}$$

We will use, repeatedly, the primitive bound

$$\sum_{\hat{k}} \mu(\hat{k}) \leq |\hat{\eta}|$$

for  $\mu \in \mathcal{M}(\hat{\eta})$ .

This brings us to

**Lemma 4.4.** *Given  $\delta > 0$ , there exists a  $c > 0$  so that for all  $\hat{\eta}^{(1)}, \hat{\eta}^{(2)}$  with  $\hat{\eta}^{(1)} \leq \hat{\eta}^{(2)}$ ,  $\delta \leq |\hat{\eta}^{(1)}| \leq |\hat{\eta}^{(2)}| \leq 1/\delta$ , and  $|\hat{\eta}^{(2)} - \hat{\eta}^{(1)}|$  sufficiently small, all  $\mu \in \mathcal{M}(\hat{\eta}^{(2)} - \hat{\eta}^{(1)}, \epsilon n)$  for some  $\epsilon$ , and any function  $\hat{\eta}(\hat{k})$  with  $\hat{\eta}^{(1)} \leq \hat{\eta}(\hat{k}) \leq \hat{\eta}^{(2)}$ ,*

$$\sum_{\hat{k}} \mu(\hat{k}) \log \frac{p(\hat{k}, \hat{\eta}(\hat{k}), n)}{p(\hat{k}, \hat{\eta}^{(2)})} \leq c |\hat{\eta}^{(2)} - \hat{\eta}^{(1)}| \left( \epsilon + |\hat{\eta}^{(2)} - \hat{\eta}^{(1)}| + \frac{1}{n} \right). \tag{4.19}$$

*Also, given  $\hat{\rho}^{(2)} \geq \hat{\rho}^{(1)} > 0$ , there exists a  $c$  so that for all  $\hat{\eta}^{(1)}, \hat{\eta}^{(2)}$  with  $\hat{\rho}^{(1)} \leq \hat{\eta}^{(1)} \leq \hat{\eta}^{(2)} \leq \hat{\rho}^{(2)}$  and  $|\hat{\eta}^{(2)} - \hat{\eta}^{(1)}|$  sufficiently small, all  $\mu \in \mathcal{M}(\hat{\eta}^{(2)} - \hat{\eta}^{(1)}, \epsilon n)$  for some  $\epsilon$ , and any function  $\hat{\eta}(\hat{k})$  with  $\hat{\eta}^{(1)} \leq \hat{\eta}(\hat{k}) \leq \hat{\eta}^{(2)}$ ,*

$$\sum_{\hat{k}} \mu(\hat{k}) \log \frac{p(\hat{k}, \hat{\eta}(\hat{k}), n)}{p(\hat{k}, \hat{\eta}^{(2)})} \geq -c |\hat{\eta}^{(2)} - \hat{\eta}^{(1)}| (\epsilon + |\hat{\eta}^{(2)} - \hat{\eta}^{(1)}|). \tag{4.20}$$

Note that the difference in the requirements of the first and second half of the lemma is that in the first half we require that  $\hat{\eta}$  be bonded away from zero in magnitude, whereas in the second half we require that  $\hat{\eta}$  be bounded away from zero in all components.

*Proof.* Using (4.9) and the fact that  $\mu(\hat{k}) = 0$  for  $|\hat{k}| > \epsilon n$  – and the “primitive” bound – we have that

$$-c'\epsilon \sum_{i=1}^L \frac{\hat{\eta}_i^{(2)} - \hat{\eta}_i^{(1)}}{1 \wedge \hat{\eta}_i^{(1)}} \leq \sum_{\hat{k}} \mu(\hat{k}) \log \frac{p(\hat{k}, \hat{\eta}(\hat{k}), n)}{p(\hat{k}, \hat{\eta}(\hat{k}))} \leq c' \left( \epsilon + \frac{1}{|\hat{\eta}^{(1)}|n} \right) |\hat{\eta}^{(2)} - \hat{\eta}^{(1)}|, \tag{4.21}$$

where  $c'$  is the  $c$  in (4.9). Then using (4.10) we have

$$\begin{aligned} \sum_i (\hat{\eta}_i^{(2)} - \hat{\eta}_i^{(1)}) \log \left( \frac{\hat{\eta}_i^{(1)}}{\hat{\eta}_i^{(2)}} \right) &\leq \sum_{\hat{k}} \mu(\hat{k}) \log \frac{p(\hat{k}, \hat{\eta}(\hat{k}))}{p(\hat{k}, \hat{\eta}^{(2)})} \\ &= \sum_{\hat{k}} \mu(\hat{k}) \left[ \log \frac{|\hat{\eta}^{(2)}|}{|\hat{\eta}(\hat{k})|} + \sum_{i=1}^L \hat{k}_i \log \frac{\hat{\eta}_i(k)}{\hat{\eta}_i^{(2)}} + \sum_{i=1}^L \hat{k}_i \sum_{j=1}^L \alpha_{ij} (\hat{\eta}_j^{(2)} - \hat{\eta}_j(k)) \right] \\ &\leq |\hat{\eta}^2 - \hat{\eta}^{(1)}| \log \left( \frac{|\hat{\eta}^{(2)}|}{|\hat{\eta}^{(1)}|} \right) + (\hat{\eta}^{(2)} - \hat{\eta}^{(1)})^T \mathbf{A} (\hat{\eta}^{(2)} - \hat{\eta}^{(1)}) : \end{aligned} \tag{4.22}$$

In the first and third term in the square brackets, we have replaced  $\hat{\eta}(k)$  with  $\hat{\eta}^{(1)}$ , componentwise and pointwise; then this third term evolves into the quadratic form due  $\mu \in \mathcal{M}(\hat{\eta}^{(2)} - \eta^{(1)})$  and in the first term, we use the primitive bound. The middle term(s) can be neglected since each member is non-positive.

Combining (4.21) and (4.22) and using the fact that  $|\hat{\eta}^{(1)}|$  is bounded below gives the stated result. ■

**Lemma 4.5.** *Let  $\hat{\rho}, \hat{\gamma}$  be given with  $\hat{\rho}, \hat{\gamma} \geq 0$  componentwise and  $|\hat{\gamma}| = 1$ . Then we have*

$$\sup_{\mu \in \mathcal{M}(\hat{\gamma})} \sum_{\hat{k}} \mu(\hat{k}) \log \frac{p(\hat{k}, \hat{\rho}) \sum_{\hat{\ell}} \mu(\hat{\ell})}{\mu(\hat{k})} = \xi(\hat{\gamma}, \hat{\rho}) - \xi(\hat{\gamma}, \hat{\psi}(\hat{\gamma}, |\hat{\rho}|)). \tag{4.23}$$

*Proof.* To reduce clutter, let us write  $\hat{\psi}$  in place of  $\hat{\psi}(\hat{\gamma}, |\hat{\rho}|)$ . Then using (4.10) to rewrite  $p(\hat{k}, \hat{\rho})$  in terms of  $p(\hat{k}, \hat{\psi})$ , we have

$$\begin{aligned} \sum_{\hat{k}} \mu(\hat{k}) \log \frac{p(\hat{k}, \hat{\rho}) \sum_{\hat{\ell}} \mu(\hat{\ell})}{\mu(\hat{k})} &= \left[ \sum_{\hat{k}} \mu(\hat{k}) \log \frac{|\hat{\psi}|}{|\hat{\rho}|} p(\hat{k}, \hat{\psi}) \sum_{\hat{\ell}} \mu(\hat{\ell}) \right. \\ &\quad \left. + \sum_{\hat{k}} \mu(\hat{k}) \sum_{i=1}^L \hat{k}_i \left( \log \frac{\hat{\rho}_i}{\hat{\psi}_i} - \sum_{j=1}^L \alpha_{ij} (\hat{\rho}_j - \hat{\psi}_j) \right) \right]. \end{aligned} \tag{4.24}$$

Due to the requirement that  $\sum_{\hat{k}} \hat{k} \mu(\hat{k}) = \hat{\gamma}$  for each  $\mu \in \mathcal{M}(\hat{\gamma})$ , the second sum on the right hand side evaluates to  $\xi(\hat{\gamma}, \hat{\rho}) - \xi(\hat{\gamma}, \hat{\psi})$ . Thus it remains to be proved that



$$\sup_{\mu \in \mathcal{M}(\hat{\gamma})} \sum_{\hat{k}} \mu(\hat{k}) \log \frac{|\hat{\psi}| p(\hat{k}, \hat{\psi}) \sum_{\hat{\ell}} \mu(\hat{\ell})}{\mu(\hat{k})} = 0, \tag{4.25}$$

which we shall do in two parts. We start by showing that the supremum is at least zero.

We recall that  $\hat{\psi}$  is either subcritical or critical. In the former case let  $\chi(\hat{\psi})$  denote, as in (4.1) the limiting average size of the appropriate connected cluster (see, also, Lemma 4.1) and let us take  $\mu(\hat{k}) = [\chi]^{-1} p(\hat{k}, \hat{\psi})$ . Then, since we are subcritical,  $\sum_{\hat{\ell}} \mu(\hat{\ell}) = [\chi]^{-1}$  while (4.4) – and the fact that  $\mu(0, \dots, 0) = 0$  – shows that indeed  $\mu \in \mathcal{M}(\hat{\gamma})$ . Then using  $|\hat{\psi}| = |\hat{\rho}|$ , we have, for all  $\hat{k}$ ,

$$\log \frac{|\hat{\psi}| p(\hat{k}, \hat{\psi}) \sum_{\hat{\ell}} \mu(\hat{\ell})}{\mu(\hat{k})} \equiv 0. \tag{4.26}$$

The critical cases are handled with a similar subcritical strategy. Let  $\hat{\psi}(\hat{\gamma}, t)$  be as in (4.2) and suppose  $t < t^*$ . We are therefore assuming  $|\hat{\rho}| \geq t^*$  so  $\hat{\psi} = \hat{\psi}(\hat{\gamma}, t^*)$ . We utilize, for  $t < t^*$

$$\mu_t(\hat{k}) = \frac{1}{\chi(t)} p(\hat{k}, \hat{\psi}(\hat{\gamma}, t)).$$

Then, again by (4.4), etc.,  $\mu_t$  is in  $\mathcal{M}(\hat{\gamma})$ . The quantity of interest in (4.25) becomes

$$\sum_{\hat{k}} \mu_t(\hat{k}) \log \frac{|\hat{\psi}|}{|\hat{\rho}|} + \sum_{\hat{k}} \mu_t(\hat{k}) \log \frac{p(\hat{k}, \hat{\psi})}{p(\hat{k}, \hat{\psi}(\hat{\gamma}, t))}$$

and we will let  $t \uparrow t^*$ . The first term is manifestly of order  $[\chi]^{-1}$  and vanishes in this limit. As for the second, we will use (4.10) inside the log. We obtain

$$\begin{aligned} \left| \sum_{\hat{k}} \mu_t(\hat{k}) \log \frac{p(\hat{k}, \hat{\psi})}{p(\hat{k}, \hat{\psi}(\hat{\gamma}, t))} \right| &\leq \frac{1}{\chi(t)} \left[ \left| \log \frac{|\hat{\psi}(\hat{\gamma}, t)|}{|\hat{\psi}|} \right| + \sum_{i=1}^L \hat{k}_i \left| \log \frac{|\hat{\psi}_i(\hat{\gamma}, t)|}{|\hat{\psi}_i|} \right| \right] \\ &+ \sum_{\hat{k}} \mu_t(\hat{k}) \sum_{i=1}^L \hat{k}_i \sum_{j=1}^L \alpha_{ij} |\hat{\psi}_j(\hat{\gamma}, t) - \hat{\psi}_j|. \end{aligned} \tag{4.27}$$

Again, the first term vanishes with  $[\chi]^{-1}$ . As for the last two terms, since  $|\hat{\psi}_j(\hat{\gamma}, t) - \hat{\psi}_j|$  will vanish as  $t \uparrow t^*$  and, meanwhile,  $\mu_t(\hat{k}) \hat{k}_i$  sums to something finite which is independent of  $t$  (namely  $\hat{\gamma}_i$ ) these terms go to zero as well.

To show that the supremum is at most zero, we drop the  $\frac{|\hat{\psi}|}{|\hat{\rho}|} \leq 1$  term and the rest follows from elementary convexity considerations. Indeed, let  $q(\hat{k}) \propto \mu(\hat{k})$  be a probability measure on  $\mathbb{N}^L$  we have (regardless of whether  $\mu \in \mathcal{M}(\hat{\gamma})$ )

$$\mu(\hat{k}) \log \frac{p(\hat{k}, \hat{\psi})}{q(\hat{k})} \propto q(\hat{k}) \log \frac{p(\hat{k}, \hat{\psi})}{q(\hat{k})} = p(\hat{k}, \hat{\psi}) \left[ \frac{q(\hat{k})}{p(\hat{k}, \hat{\psi})} \log \frac{p(\hat{k}, \hat{\psi})}{q(\hat{k})} \right].$$

Letting  $G(x) := -x \log x$  – which is concave – then by Jensen’s inequality,

$$\sum_{\hat{k}} p(\hat{k}, \hat{\psi}) G\left(\frac{q(\hat{k})}{p(\hat{k}, \hat{\psi})}\right) \leq G\left(\sum_{\hat{k}} q(\hat{k})\right) = G(1) = 0$$

and we have finished the proof of the lemma. ■

For the proof of Theorem 4.2, we will need to introduce a discrete version of  $\mathcal{M}(\hat{\gamma}, r)$ . Let  $M(\hat{b}, r)$  be the set of sequences indexed by elements of  $\mathbb{N}^L \setminus \{0\}$  given by

$$M(\hat{b}, r) = \left\{ m_{\hat{k}} : \mathbb{N}^L \setminus \{\hat{0}\} \rightarrow \mathbb{N} : \sum_{\hat{k}} \hat{k} m_{\hat{k}} = \hat{b}; m_{\hat{k}} = 0 \text{ for } |\hat{k}| > r \right\}. \tag{4.28}$$

We are then in need of a bound on the size of  $M$ :

**Lemma 4.6.**  $|M(\hat{b}, r)| = \exp\left(O\left(n^{\frac{L+1}{L+2}} (\log n)^2\right)\right)$ , where  $n = |\hat{b}|$ , uniformly in  $r$ . In particular,  $\frac{1}{n} \log |M(\hat{b}, r)| = o(1)$ .

*Proof.* We will prove this by instead bounding the size of the larger set

$$Q(n) = \left\{ m_{\hat{k}} : \mathbb{N}^L \setminus \{\hat{0}\} \rightarrow \mathbb{N} : \sum_{\hat{k}} |\hat{k}| m_{\hat{k}} \leq n \right\}, \tag{4.29}$$

which does not depend on  $r$  at all. Let  $J_j = \{\hat{k} \in \mathbb{N}^L : 2^j \leq |\hat{k}| < 2^{j+1}\}$ , and let  $w(a, b) = |\{\hat{k} \in \mathbb{N}^b : |\hat{k}| = a\}| = \binom{a+b-1}{b-1}$ . We then note that for each  $(m_{\hat{k}}) \in Q(n)$ , we must have  $\sum_{\hat{k} \in J_j} m_{\hat{k}} \leq n2^{-j}$  for all  $j$ . Thus we have

$$|Q(n)| \leq \prod_{j=0}^{\infty} \sum_{a=0}^{\lfloor n2^{-j} \rfloor} w(a, |J_j|) \tag{4.30}$$

$$\leq \prod_{j=0}^{\lfloor \log_2 n \rfloor} nw(\lfloor n2^{-j} \rfloor, |J_j|). \tag{4.31}$$

We claim that

$$w(a, b) \leq \left(1 + \frac{a}{b}\right)^b \left(1 + \frac{b}{a}\right)^a$$

This is verified as follows: First, to simplify matters a bit, note that by monotonicity in  $b$ , we may replace  $(b - 1)$ 's by  $b$ 's in the expression for  $w(a, b)$ . If  $a = N$  and  $b = 1$ , the claim is obviously true. Now we can increment  $b$  and the full claim follows an inductive argument which uses the monotonicity of  $\left(1 + \frac{a}{b+\vartheta}\right)^b \left(1 + \frac{b+\vartheta}{a}\right)^a$  as a function of  $\vartheta$  for  $\vartheta > 0$ .

Next we claim that, e.g., if  $a \geq b$  that

$$\left(1 + \frac{a}{b}\right)^b \left(1 + \frac{b}{a}\right)^a \leq \left(4\frac{a}{b}\right)^b.$$

Indeed, this is manifestly an equality if  $a = b$  and logarithmic differentiation of both sides verifies the result due to the fact that  $\log(1 + b/a) \leq b/a$ .

All this allows us to estimate  $|J_j| \leq c_L 2^{j(L+1)}$ , and so

$$w(\lfloor n2^{-j} \rfloor, |J_j|) \leq n^{O(n^{\frac{L+1}{L+2}})} \tag{4.32}$$

for  $j = 0, \dots, \lfloor \log_2 n \rfloor$ . Combined with (4.31), this gives us the result. ■

### 4.3. The Rate for $\Upsilon$

*Proof of Theorem 4.2.* We will prove this by getting an upper and lower bound, although we will only show the upper bound explicitly. The proof of the upper bound requires a bit of boot strapping, and the first half of the process is to show that for any  $\delta > 0$  there is a  $c = c(\hat{\eta}^{(2)}, \hat{\eta}^{(1)})$  so that

$$\begin{aligned} & \frac{1}{n} \log \Upsilon(\hat{\eta}^{(2)}, \hat{\eta}^{(1)}, \epsilon n, n) \\ & \leq |\hat{\gamma}|[o(1) + c|\hat{\gamma}|] + \left( \xi(\hat{\gamma}/|\hat{\gamma}|, \hat{\eta}^{(2)}) - \xi(\hat{\gamma}/\hat{\gamma}, \hat{\psi}(\hat{\gamma}/|\hat{\gamma}|, |\hat{\eta}^{(2)}|)) \right), \end{aligned} \tag{4.33}$$

for all  $\hat{\rho} \geq \hat{\eta}^{(2)} \geq \hat{\eta}^{(1)}$  with  $|\hat{\eta}^{(1)}|, |\hat{\eta}^{(2)}|^{-1} \geq \delta$  and sufficiently small  $\hat{\gamma} = \hat{\eta}^{(2)} - \hat{\eta}^{(1)}$ , where the  $o(1)$  term is uniformly bounded for  $\hat{\eta}^{(1)}, \hat{\eta}^{(2)}$  (which are fixed independent of  $n$ ) in this domain and this term tends to zero as  $n$  tends to infinity and  $\epsilon$  tends to zero.

We begin by setting  $\hat{b} = \lfloor \hat{\eta}^{(2)} n \rfloor - \lfloor \hat{\eta}^{(1)} n \rfloor$ , and noting that the event upon which  $\Upsilon$  is based occurs if and only if there is some  $(m_{\hat{k}}) \in M(\hat{b}, \epsilon n)$  so that exactly  $m_{\hat{k}}$  of  $C^{(1)}, C^{(2)}, \dots, C^{(\sum m_{\hat{\ell}})}$  have  $\langle C^{(\cdot)} \rangle = \hat{k}$ . Thus we have

$$\Upsilon(\hat{\eta}^{(2)}, \hat{\eta}^{(1)}, \epsilon n, n) = \sum_{m_{\hat{k}} \in M(\hat{b}, \epsilon n)} \frac{(\sum m_{\hat{k}})!}{\prod m_{\hat{k}}!} \left[ \prod p(\hat{k}, \hat{\eta}^{(2)} - O(|\hat{\gamma}|), n)^{m_{\hat{k}}} \right], \tag{4.34}$$

where the  $O(|\hat{\gamma}|)$  term is strictly bounded in magnitude by  $|\hat{\gamma}|$ . Now let  $m_{\hat{k}} \in M(\hat{b}, \epsilon n)$  be chosen to maximize the summand. Since Lemma 4.6 gives us  $|M(\hat{b}, \epsilon n)| \leq e^{|\hat{\gamma}|o(n)}$ , we have

$$\log \Upsilon(\hat{\eta}^{(2)}, \hat{\eta}^{(1)}, \epsilon n, n) \leq |\hat{\gamma}|o(n) + \log \frac{(\sum m_{\hat{k}})!}{\prod m_{\hat{k}}!} \left[ \prod p(\hat{k}, \hat{\eta}^{(2)} - O(|\hat{\gamma}|), n)^{m_{\hat{k}}} \right], \tag{4.35}$$

where the  $o(n)$  term is uniformly bounded in  $\hat{\eta}^{(1)}, \hat{\eta}^{(2)}$ , and  $\epsilon$  as  $n \rightarrow \infty$ . Now dividing by  $n$  and using Stirling's approximation, we get that there exists a  $\mu \in \mathcal{M}(\frac{1}{n}\hat{b}, \epsilon n)$  with

$$\frac{1}{n} \log \Upsilon(\hat{\eta}^{(2)}, \hat{\eta}^{(1)}, \epsilon n, n) \leq |\hat{\gamma}|o(1) + \sum_{\hat{k}} \mu(\hat{k}) \log \frac{p(\hat{k}, \hat{\eta}^{(2)} - O(|\hat{\gamma}|), n)^{\sum_{\hat{\ell}} \mu(\hat{\ell})}}{\mu(\hat{k})}. \tag{4.36}$$

Note that the dangerous looking  $\prod m_{\hat{k}}!$  in (4.35) does not cause us trouble – because it is in the denominator – and the  $o(1)$  term is still uniformly bounded (e.g., by an inverse power of  $n$ ). For the equivalent step in the lower bound, one would use the fact that  $(m_{\hat{k}}) \in M(\hat{b}, r)$

to bound the error. We now apply (4.19) from Lemma 4.4 in order to rid ourselves of the  $n$ -argument in the  $p$ -term with the result:

$$\frac{1}{n} \log \Upsilon(\hat{\eta}^{(2)}, \hat{\eta}^{(1)}, \epsilon n, n) \leq |\hat{\gamma}|o(1) + c|\hat{\gamma}|(\epsilon + |\hat{\gamma}| + 1/n) + \sum_{\hat{k}} \mu(\hat{k}) \log \frac{p(\hat{k}, \hat{\eta}^{(2)}) \sum_{\hat{\ell}} \mu(\hat{\ell})}{\mu(\hat{k})} \tag{4.37}$$

for some  $c$  depending only on  $\delta$ . By scaling  $\mu$  by a factor of  $n/|\hat{b}|$  we can apply Lemma 4.5 and use the continuity of  $\xi$  – which covers the  $O(1/n)$  gap between  $\hat{\gamma}$  and  $\hat{b}/n$  – to get (4.33).

We next note that if all components (discovered in the  $\Upsilon$  process) in  $\mathcal{G}(n, \hat{\rho})$  are of size smaller than  $\epsilon n$ , then for every  $x \in (|\hat{\eta}|, |\hat{\rho}|)$  there must be at least one  $k$  so that

$$(|\hat{\rho}| - x)n \leq \sum_{i=1}^k |C^{(i)}| \leq (|\hat{\rho}| - x)n + \epsilon n. \tag{4.38}$$

Next it is remarked that while the size bias, of course, skews the distribution of cluster sizes, we recollect that this distribution is based on uniform selection of a vertex and then “pulling out” that which is attached. As such, the process can be implemented dynamically by stochastically growing the cluster of the each vertex selected *after* the selection has taken place. Now let us suppose that  $\hat{\rho}$  is bounded above and that each component is bounded strictly away from zero. And further, let us suppose that not too much of the overall sample has been processed. Then, at any point, any particular cluster – with component sizes independent of  $n$  – can be selected with probability uniformly bounded below independent of  $n$  as is seen by the consideration of a sequence of pertinent events leading to that cluster.

Thus we have that for any given collection of integers  $m_1, \dots, m_j \in [|\hat{\eta}n|, |\hat{\rho}n|]^1$  there exist a corresponding  $k_1, \dots, k_j$  so that  $\sum_{i=1}^{k_\ell} |C^{(i)}| = m_\ell$  for each  $\ell$  with probability  $e^{jO(\epsilon n)}$ . Hence for all  $j$  the independence of edges in  $\mathcal{G}(n, \hat{\rho})$  lets us write

$$\Upsilon(\hat{\rho}, \hat{\eta}, \epsilon n, n) \leq e^{j(O(\log n) + O(\epsilon n))} \sup_{\substack{\hat{\rho} = \hat{\rho}^{(j)} \geq \dots \geq \hat{\rho}^{(0)} = \hat{\eta} \\ |\hat{\rho}^{(i)}| = |\hat{\eta}| + \frac{1}{j} |\hat{\rho} - \hat{\eta}|}} \prod_{i=0}^{j-1} \Upsilon(\hat{\rho}^{(i+1)}, \hat{\rho}^{(i)}, \epsilon n, n), \tag{4.39}$$

where the  $e^{O(\log n)}$  term – which comes from the number of choices for each  $\hat{\rho}^{(i)}$  – is uniformly bounded for  $\hat{\rho}$  bounded above. We note at this point that each  $\Upsilon$  term has a trivial upper bound of 1, and so – in preparation for the application of (4.33) – we will discard terms on the right hand side of (4.39) for which  $|\hat{\rho}^{(i)}| < \delta$ . Furthermore, to simplify notation, let us consider only  $\hat{\rho}$  for which  $|\hat{\rho}| \geq \delta$  at the moment. Then, because  $\Upsilon$  discretizes the  $\hat{\rho}^{(i)}$  terms, the supremum can be considered to be taken over a finite set, so there is a specific choice of  $(\hat{\rho}^{(i)})$  which achieves the supremum. Using, temporarily,

$$f(\hat{\rho}, \hat{\eta}, r, n) := \frac{1}{n} \log \Upsilon(\hat{\rho}, \hat{\eta}, r, n), \tag{4.40}$$

---

<sup>1</sup>The range we are actually interested in is  $[|\hat{\eta}n|, |\hat{\rho}n|]$ , but the “floors” will distract from the presented idea and are omitted.

we have for all  $\epsilon, n, j$  the existence of a partition  $\hat{\eta} = \hat{\rho}^{(0)} \leq \dots \leq \hat{\rho}^{(j)} = \hat{\rho}$  with  $|\hat{\rho}^{(i)}| = |\hat{\eta}| \vee \delta + \frac{i}{j}(|\hat{\rho}| - |\hat{\eta}| \vee \delta)$  such that

$$f(\hat{\rho}, \hat{\eta}, \epsilon n, n) \leq j o(1) + \sum_{i=0}^{j-1} f(\hat{\rho}^{(i+1)}, \hat{\rho}^{(i)}, \epsilon n, n). \tag{4.41}$$

Interpolating by straight lines, it is clear that  $(\hat{\rho}^{(i)} \mid i = 0, \dots, j)$  defines an ascent restricted to the interval  $[|\hat{\eta}| \vee \delta, |\hat{\rho}|]$  which we denote by  $\hat{\Gamma}$  (remembering that  $\hat{\Gamma}$  depends on  $\epsilon, n,$  and  $j$ ). Let  $t^{(j,i)} = |\hat{\eta}| \vee \delta + \frac{i}{j}(|\hat{\rho}| - |\hat{\eta}| \vee \delta)$ . We now apply (4.33) to the right hand side of (4.41) to get

$$f(\hat{\rho}, \hat{\eta}, \epsilon, n) \leq j o(1) + O(1/j) + (|\hat{\rho}| - |\hat{\eta}| \vee \delta) \frac{1}{j} \sum_{i=0}^{j-1} \left[ \xi \left( \frac{\hat{\Gamma}(t^{(j,i+1)}) - \hat{\Gamma}(t^{(j,i)})}{j^{-1}(|\hat{\rho}| - |\hat{\eta}| \vee \delta)}, \hat{\Gamma}(t^{(j,i)}) \right) - \xi \left( \frac{\hat{\Gamma}(t^{(j,i+1)}) - \hat{\Gamma}(t^{(j,i)})}{j^{-1}(|\hat{\rho}| - |\hat{\eta}| \vee \delta)}, \hat{\psi} \left( \frac{\hat{\Gamma}(t^{(j,i+1)}) - \hat{\Gamma}(t^{(j,i)})}{j^{-1}(|\hat{\rho}| - |\hat{\eta}| \vee \delta)}, t^{(j,i)} \right) \right) \right], \tag{4.42}$$

where both error terms are uniformly bounded for  $\hat{\rho}$  bounded above. It is noted that the substantive term in (4.42) is perfectly well defined, e.g.  $\hat{\Gamma}_k(t^{(j,i+1)}) - \hat{\Gamma}_k(t^{(j,i)})$  vanishes if and only if the  $k^{\text{th}}$  component of the corresponding  $\hat{\psi}$  vanishes – thence no difficulty with interpretation of logs.

We next claim that, for all intents and purposes, the “substantive” term is just  $H(\hat{\Gamma}, |\hat{\eta}| \vee \delta, |\hat{\rho}|)$ . This is not quite exact since if  $\hat{\psi}(\hat{\Gamma}', t)$  is subcritical, at time  $t^{(j,i)}$  it will change a bit over the course of  $[t^{(j,i)}, t^{(j,i+1)}]$ . However, in these circumstance, the change in  $\hat{\psi}$  is (at most) proportional to the change in  $t$  itself and/or the change in  $\chi^{-1}$  – which in turn is bounded by the change in  $t$ . Moreover, the object of proportion for the  $k^{\text{th}}$  component is the  $k^{\text{th}}$  component itself which again alleviates any concerns about the singularities associated with logarithms. On this basis it is seen that, in the course of each increment, the error incurred by replacing the appropriate term in (4.42) by the integration of  $\xi(\hat{\Gamma}', \hat{\Gamma}) - \xi(\hat{\Gamma}', \hat{\psi})$  along the corresponding portion of the path is, in fact, bounded by a uniform constant times  $t^{(j,i+1)} - t^{(j,i)}$ . We thus have

$$f(\hat{\rho}, \hat{\eta}, \epsilon n, n) \leq j o(1) + O(1/j) + H(\hat{\Gamma}, |\hat{\eta}| \vee \delta, |\hat{\rho}|) \tag{4.43}$$

where the additional small error terms have been incorporated into the  $O(1/j)$  term. Obviously we may replace  $H(\hat{\Gamma}, |\hat{\eta}| \vee \delta, |\hat{\rho}|)$  by the supremum over available  $\hat{\Gamma}$  and, allowing  $j$  to be considered as a function of  $\epsilon$  and  $n$ , we can replace both error terms with a single  $o(1)$  term which tends to zero uniformly – for  $\hat{\rho}$  bounded above – as  $\epsilon$  tends to zero and  $n$  tends to infinity. Furthermore, it is straightforward to see that the maximal difference between  $H(\hat{\Gamma}, |\hat{\eta}| \vee \delta, |\hat{\rho}|)$  and  $H(\hat{\Gamma}, |\hat{\eta}|, |\hat{\rho}|)$  is uniformly bounded – that is, for all  $\hat{\eta}, \hat{\rho}$ , and ascents  $\hat{\Gamma}$  – by an  $O(\delta)$  correction, for small  $\delta$ . Since we may retroactively declare  $\delta$  to have been picked as small as desired – after which we may take  $\epsilon$  to be small and  $n$  large, we find

$$\frac{1}{n} \log \Upsilon(\hat{\rho}, \hat{\eta}, \epsilon n, n) \leq o(1) + \sup_{\hat{\Gamma} \in \mathcal{P}(\hat{\eta}) \cap \mathcal{P}(\hat{\rho})} H(\hat{\Gamma}, |\hat{\eta}|, |\hat{\rho}|), \tag{4.44}$$

where the  $o(1)$  term converges uniformly to zero as  $\epsilon \rightarrow 0$  and  $n \rightarrow \infty$ .

The opposite bound, namely

$$\frac{1}{n} \log \Upsilon(\hat{\rho}, \hat{\eta}, r, n) \geq o(1) + \sup_{\hat{\Gamma} \in \mathcal{P}(\hat{\eta}) \cap \mathcal{P}(\hat{\rho})} H(\hat{\Gamma}, |\hat{\eta}|, |\hat{\rho}|), \tag{4.45}$$

where the  $o(1)$  term tends to zero uniformly as  $r, n \rightarrow \infty$ , is derived by the similar methods. First, an analog opposite bound of the form in (4.33) is derived for small  $|\hat{\gamma}|$  – although due to the weaker lower bound of Lemma 4.4 the bound only holds uniformly for  $\hat{\eta}^{(1)}$  larger than  $\delta$  in all components. Then, at the point of (4.39) we can simply pick any particular ascent from  $\hat{\eta} \vee (\delta, \dots, \delta)$  to  $\hat{\rho}$  and place the points  $\hat{\rho}^{(i)}$  along this path. The rest of the argument is identical with the final step being an optimization over ascents, and using brute force to traverse the distance from  $\hat{\eta} \vee (\delta, \dots, \delta)$  to  $\hat{\eta}$ . ■

#### 4.4. Natural Assents

Consider an ascent,  $\hat{\Gamma}(t)$  which starts at the origin and goes at least some distance into the supercritical region. Then (since at least one component increases) there is a unique  $t^*$  such that for  $t < t^*$ ,  $\hat{\Gamma}(t)$  is subcritical while for  $t > t^*$ , it is supercritical. To define a *natural ascent*, we shall treat separately the subcritical and super critical behaviors. Starting with the former, it is stipulated that

$$\hat{\Gamma}'(t) = \frac{[\mathbf{I} - \mathbf{B}_{\hat{\Gamma}(t)}^T]^{-1} \hat{\Gamma}(t)}{||[\mathbf{I} - \mathbf{B}_{\hat{\Gamma}(t)}^T]^{-1} \hat{\Gamma}(t)||} \tag{4.46}$$

at least for  $\hat{\Gamma} \neq 0$ . (As is not hard to see, if we actually wish to start the ascent at the origin, an initial direction must also be specified.) Notice that (4.46) implies (c.f. (4.1) – (4.4)) that  $\hat{\psi}(\hat{\Gamma}', |\hat{\Gamma}'|) = \hat{\Gamma}$  (while, of course, the associated average cluster size ratios, is proportional to  $\hat{\Gamma}'$ ). In the limit as  $t \uparrow t^*$  we find that  $\hat{\Gamma}'$  tends to the limiting size ratio associated with the critical density. Denoting these objects by  $\hat{\gamma}_c$  and  $\hat{\psi}_c$  respectively (so that  $\hat{\gamma}_c$  is the maximum eigenvector of  $\mathbf{B}_{\hat{\psi}_c}^T$ ) we define, for  $t > t^*$   $\hat{\Gamma}' \equiv \hat{\gamma}_c$  – with  $\hat{\Gamma}'' \equiv 0$  – i.e.

$$\hat{\Gamma}(t) = \hat{\psi}_c + (t - t^*)\hat{\gamma}_c. \tag{4.47}$$

**Lemma 4.7.** *Let  $\hat{\rho}^{(2)} > \hat{\rho}^{(1)} \geq 0$  be given, and suppose there is a  $\hat{\Gamma} \in \mathcal{P}(\hat{\rho}^{(1)}) \cap \mathcal{P}(\hat{\rho}^{(2)})$  which is an ascent such that*

$$H(\hat{\Gamma}, |\hat{\rho}^{(1)}|, |\hat{\rho}^{(2)}|) = \sup_{\hat{\Phi} \in \mathcal{P}(\hat{\rho}^{(1)}) \cap \mathcal{P}(\hat{\rho}^{(2)})} H(\hat{\Phi}, |\hat{\rho}^{(1)}|, |\hat{\rho}^{(2)}|). \tag{4.48}$$

*Then there is some  $\hat{c}$  with  $\sum_{i=1}^L \hat{c}_i = 0$  such that  $\hat{\Gamma} + \hat{c}$  is a natural ascent restricted to  $[|\hat{\rho}^{(1)}|, |\hat{\rho}^{(2)}|]$ .*

*Proof.* Let us abbreviate  $a = |\hat{\rho}^{(1)}|$  and  $b = |\hat{\rho}^{(2)}|$ . It is first noted that for any ascent  $\hat{\Gamma}$  connecting  $\hat{\rho}^{(1)}$  to  $\hat{\rho}^{(2)}$  the portion of  $H$  consisting of  $\int_a^b \xi(\hat{\Gamma}, \Gamma)$  is a constant  $c$  which depends only on  $\hat{\rho}^{(1)}$  and  $\hat{\rho}^{(2)}$  and hence need not be further discussed. Our first goal is to establish lower bounds on the speed. Of course it may be the case that  $\hat{\rho}_i^{(2)} = \hat{\rho}_i^{(1)}$  for one or more values of  $i$  in which case, in all available choices of ascents,  $\hat{\Gamma}_i = \text{const}$ . We

claim that in all other circumstances, the speeds are, componentwise, uniformly bounded below (recalling once more that our parameterization gives ascents whose total speed is always one). We adopt the notation  $\hat{\Gamma}' = \hat{\gamma}$  and, subtracting and adding  $\sum_i \hat{\gamma}_i \log \hat{\gamma}_i$ , it is seen that the remains of the integrand (after the  $-\hat{\gamma}_i \log \hat{\gamma}_i$  term) are non-singular, with non-singular derivatives as any particular  $\hat{\gamma}_i \rightarrow 0$ . However the “principal” term is concave with a singular derivative ( $\propto -\log \hat{\gamma}_i$ ) which, as we shall see, does not permit any component of  $\hat{\gamma}_i$  to get too small. Indeed, let  $\gamma_i$  be non-trivial and suppose there is a set of size  $\Delta t$ , on which  $\hat{\gamma}_i$  does not exceed some  $\varepsilon$  with  $\varepsilon \ll 1$ . Let us find another portion of the path where  $\gamma_i$  exceeds, half the total required rate of ascent:

$$\hat{\gamma}_i > \frac{1}{2} \frac{\hat{\rho}_i^{(2)} - \hat{\rho}_i^{(1)}}{b - a}.$$

Here we may have to assume that  $\Delta t$  is not too large which, obviously, we may do without loss of generality. Denoting the two sets by  $\underline{b}$  and  $\underline{s}$  (big and small) we may consider the canonical map from  $\underline{b}$  to  $\underline{s}$  and by this means, replacing  $\hat{\gamma}_i|_{\underline{b}}$  with  $(1 - \varepsilon)\hat{\gamma}_i|_{\underline{b}}$  we can increase  $\hat{\gamma}_i|_{\underline{s}}$  by an amount of order  $\varepsilon$ . The gain from this transfer, by consideration of the principal part along the set  $\underline{s}$ , is of the order  $|\varepsilon \log \varepsilon|$ . Meanwhile the “losses” from the other parts of the functional on both  $\underline{s}$  and  $\underline{b}$  are bounded above by a constant times  $\varepsilon$  for all  $\varepsilon$  sufficiently small. Of course in addition, we must now recalibrate to unit speed but this causes changes in the above effects which are also only of the order of  $\varepsilon$ . Thus, along a minimizer, for any  $i$  in which  $\hat{\rho}_i^{(2)} - \hat{\rho}_i^{(1)} > 0$ , we may conclude that there is an  $\varepsilon_0 > 0$  – which will depend on  $\hat{\rho}^{(2)}$  and  $\hat{\rho}^{(1)}$  – such that  $\hat{\gamma}_i > \varepsilon_0$  for a.e.  $t$ .

With the above in hand, we may now add a perturbation to the minimizing ascent secure in the knowledge that, for sufficiently small perturbation, the resultant function is indeed an ascent. The natural procedure is to derive Euler–Lagrange equations but, unfortunately, there is no *a priori* guarantee that all required partial derivatives exist. For this reason we deviate from the usual methods at the point where an integration by parts would normally be performed.

Let  $\hat{g} : (0, \infty)^L \mapsto \mathbb{R}^L$  be a Lipschitz continuous function with  $\hat{g}(a) = \hat{g}(b) = 0$  and  $\sum_{i=1}^L g(t) = 0$  for  $t \in (a, b)$ . Then for all sufficiently small  $\delta$ , we may write

$$H(\hat{\Gamma} + \delta\hat{g}, a, b) = c - \int_a^b \xi \left( (\hat{\Gamma} + \delta\hat{g})'(t), \hat{\psi}((\hat{\Gamma} + \delta\hat{g})'(t), t) \right) dt. \tag{4.49}$$

Now suppose that  $\hat{\gamma}, t > 0$  are given with  $|\hat{\gamma}| = 1$ . We recall that  $\chi^{-1}(\hat{\gamma}, t)$  and  $\hat{\psi}(\hat{\gamma}, t)$  are defined in such a way that either  $|\hat{\psi}(\hat{\gamma}, t)| = t$  or  $\hat{\chi}^{-1}(\hat{\gamma}, t) = 0$ . Using this, (see also (4.60)) we have

$$\xi(\hat{\gamma}, \hat{\psi}(\hat{\gamma}, t)) = -1 + \chi^{-1}(\hat{\gamma}, t) + \sum_{i=1}^L \hat{\gamma}_i \log \hat{\psi}_i(\hat{\gamma}, t), \tag{4.50}$$

but we here we encounter our first problem with differentiability. Let  $\hat{\eta}$  be given with  $\sum_{i=1}^L \hat{\eta}_i = 0$ . Then while  $\frac{\partial}{\partial \delta} \hat{\psi}(\hat{\gamma} + \delta\hat{\eta}, t)|_{\delta=0}$  and  $\frac{\partial}{\partial \delta} \chi^{-1}(\hat{\gamma} + \delta\hat{\eta}, t)|_{\delta=0}$  are well defined in the region where  $|\hat{\psi}| \neq t$  and the region where  $\chi^{-1} > 0$ , the limiting values may not agree on the boundary. Nevertheless, a careful calculation shows that the discontinuities in the derivative arising from the last two terms in (4.50) cancel each other exactly, yielding

$$\frac{\partial}{\partial \delta} \xi(\hat{\gamma} + \delta\hat{\eta}, \hat{\psi}(\hat{\gamma} + \delta\hat{\eta}, t)) \Big|_{\delta=0} = \xi(\hat{\eta}, \hat{\psi}(\hat{\gamma}, t)). \tag{4.51}$$

For brevity, let us for the remainder of the proof write  $\hat{\psi}(t)$  for  $\hat{\psi}(\hat{\Gamma}'(t), t)$ . Then by combining (4.49) and (4.51), we see that any  $\hat{\Gamma}$  which maximizes  $H(\hat{\Gamma}, a, b)$  must have

$$\int_a^b \xi \left( \hat{g}'(t), \hat{\psi}(t) \right) dt = 0. \tag{4.52}$$

Since this must hold for all Lipschitz  $\hat{g}$  with  $\hat{g}(a) = \hat{g}(b) = 0$  and  $\sum_{i=1}^L \hat{g}_i(t) = 0$ , we can for any pair  $i, j$  take  $\hat{g}_i(t) = -\hat{g}_j(t) = g(t)$  with all other components zero and conclude

$$\int_a^b g'(t) \left[ \left( \log \hat{\psi}_i(t) - \sum_{k=1}^L \alpha_{ik} \hat{\psi}_k(t) \right) - \left( \log \hat{\psi}_j(t) - \sum_{k=1}^L \alpha_{jk} \hat{\psi}_k(t) \right) \right] dt = 0 \tag{4.53}$$

for every Lipschitz function  $g$  with  $g(a) = g(b) = 0$ . Thus the difference in the integrand must be constant almost everywhere, meaning that there must exist some  $\hat{d}$  and  $f(t)$  such that

$$\log \hat{\psi}_i(t) - \sum_{k=1}^L \alpha_{ik} \hat{\psi}_k(t) = \hat{d}_i + f(t) \tag{4.54}$$

for almost every  $t$ , for each  $i$ .

Since it will clean up the proof without making a substantial difference, let us suppose that (4.54) holds for every  $t$ , instead of merely almost every  $t$ . We then claim (the proof of which shall be postponed) that if  $\hat{T}(\hat{x})$  is defined for  $\hat{x}$  critical or subcritical with

$$\hat{T}_i(\hat{x}) = \log \hat{x}_i - \sum_{k=1}^L \alpha_{ik} \hat{x}_k, \tag{4.55}$$

then  $\hat{T}$  is invertible and  $\hat{T}^{-1}$  is increasing. This property necessarily implies that if  $t_0$  is a time such that  $\hat{\psi}(t_0)$  is critical then for all  $t \in [a, b]$ , we have  $f(t) \leq f(t_0)$ . Indeed, assuming the contrary at  $t = t'$  then, from (4.54) we would have  $\hat{\psi}(t')$  supercritical and yet there is no mechanism in the construction of  $\psi$ 's for anything except critical and subcritical. Furthermore, if  $\hat{\psi}(t)$  is subcritical, we have  $t = |\hat{\psi}(t)| < |\hat{\psi}(t_0)| \leq t_0$ . From this we conclude that if  $t^* = \inf\{t : \hat{\psi}(t) \text{ is critical}\}$ , then  $\hat{\psi}(t)$  is critical and constant for  $t > t^*$ . Furthermore, since  $|\hat{\psi}(t)| = t$  for  $t < t^*$ , we must have that  $f$  is increasing, and thus  $\hat{\psi}(t)$  is increasing in all components. Thus  $\hat{\psi}$  is Lipschitz continuous.

Since  $\hat{\psi}$  is Lipschitz continuous, it is almost everywhere differentiable. Then differentiating both sides of (4.54) and multiplying by  $\hat{\psi}_i(t)$ , we get

$$[\mathbf{I} - \mathbf{B}_{\hat{\psi}(t)}^T] \hat{\psi}'(t) = f'(t) \hat{\psi}(t). \tag{4.56}$$

Multiplying both sides by  $[\mathbf{I} - \mathbf{B}_{\hat{\psi}(t)}^T]^{-1}$  and recalling that  $\frac{d}{dt} |\hat{\psi}(t)| = 1$  for  $t < t^*$ , we get

$$\frac{d}{dt} \hat{\psi}(t) = \frac{[\mathbf{I} - \mathbf{B}_{\hat{\psi}(t)}^T]^{-1} \hat{\psi}(t)}{|\mathbf{I} - \mathbf{B}_{\hat{\psi}(t)}^T]^{-1} \hat{\psi}(t)|} \tag{4.57}$$

for  $t < t^*$ . From (4.4) and (4.46) we see that  $\hat{\psi}$  obeys the differential equation for a natural ascent which we may call  $\hat{\Xi}$  (but, of course, with the wrong initial condition). Integrating



gives that, if  $t \leq t^*$ , then  $\hat{\Xi}(t) + \hat{c} = \hat{\psi}(t)$  for some  $\hat{c}$ . From (4.54) we can see that  $\hat{\Xi}'(t)$  is continuous, and since  $\hat{\psi}(t)$  is constant for  $t > t^*$ , we have that  $\hat{\Xi}'(t)$  is the maximum eigenvector of  $B_{\hat{\psi}(t^*)}^T$  for  $t > t^*$ . Combined with the fact that (4.57) shows that  $\hat{\psi}(t)$  restricted to  $t < t^*$  is indeed a natural ascent, this gives us the desired result. ■

*Proof that  $\hat{T}$  given by (4.55) is invertible and  $\hat{T}^{-1}$  is increasing.* We first show that  $\hat{T}$  is invertible. Let  $\hat{a}$  and  $\hat{b}$  be given with  $\hat{a}$  subcritical or critical and  $T(\hat{a}) = \hat{b}$ . Then for any  $\hat{x}$ , we can write  $\hat{x}_i = (1 - \hat{s}_i)\hat{a}_i$ . Doing this, we find that  $\hat{T}(\hat{x}) = \hat{b}$  if and only if

$$\hat{s}_i = 1 - e^{-\sum_{j=1}^L a_{ij}\hat{a}_j\hat{s}_j}. \tag{4.58}$$

Since  $\hat{a}$  is not supercritical, Lemma 2.2 tells us that all solutions to (4.58) are bounded by zero; and thus  $\hat{T}(\hat{x}) = \hat{b}$  implies  $\hat{x} \geq \hat{a}$ . Similarly, if  $\hat{x}$  is a subcritical or critical solution to  $\hat{T}(\hat{x}) = \hat{b}$ , we must also have  $\hat{a} \geq \hat{x}$ . Thus  $\hat{T}$  is injective from the set of subcritical and critical densities, and is thus invertible over this domain.

To see that  $\hat{T}^{-1}$  is increasing, let us define, for fixed  $\hat{y}$  the function  $\hat{F}$  by

$$\hat{F}_i(\hat{x}) = e^{\hat{y}_i + \sum_{j=1}^L a_{ij}\hat{x}_j}. \tag{4.59}$$

Now consider the iterative map

$$\hat{x}_i^{(m+1)} = F_i(\hat{x}^{(m)}) \equiv F_i^{(m)}(\hat{0})$$

and note that  $[\hat{T}^{-1}(\hat{y})]_i = \hat{F}_i^{(\infty)}(\hat{0})$ . (Indeed,  $\hat{T}^{-1}(\hat{y})$  must be the minimal fixed point of  $\hat{F}$ . Since  $\hat{F}$  is order preserving and bounded below by zero, this minimal fixed point must be the one starting from  $\hat{x} = \hat{0}$ .) Since  $\hat{F}(\cdot)$  is increasing as a function of  $\hat{y}$ , we see that  $\hat{T}^{-1}$  is increasing. ■

We conclude this section with a proof of *existence* of natural ascents

**Proposition 4.8.** *Let  $\hat{\rho}$  denote any non-zero density. Then there is a unique natural ascent passing through  $\hat{\rho}$ .*

*Proof.* To prove the above, we will use the  $H$ -functional and a certain modification to be explained below. The usual strategy in situations of this sort would be to maximize  $H$  via a maximizing sequence and compare (favorably) the functional evaluated at some limit of the sequence with the maximized  $H$ . We will do something along these lines after we implement our modifications. In any case, by Lemma 4.7, if we arrive at an object which maximizes  $H$  it must be a natural ascent; and we will tend to uniqueness in the final portion of the proof. First some standing notation: For a general ascent  $\hat{\Gamma}$  we will denote the derivative by  $\hat{\gamma}$  and the inverse of the associated  $\chi$  by  $\mathbb{X}$ .

Let us start by examining the final piece of  $H$  namely  $\hat{\gamma}^T A \hat{\psi}$ . At fixed  $t > 0$ , reiterating the arguments of Lemma 4.1 – and already discussed in the context of (4.50) – we have

$$\hat{\psi}_i \sum_{j=1}^L a_{ij}\hat{\gamma}_j = \frac{t\hat{\gamma}_i \sum_{j=1}^L a_{ij}\hat{\gamma}_j}{\mathbb{X} + t \sum_{j=1}^L a_{ij}\hat{\gamma}_j} = \hat{\gamma}_i - \frac{\mathbb{X}}{t} \hat{\psi}_i \tag{4.60}$$

Now if  $\hat{\psi}$  is subcritical then when we sum the result we get  $1 - \mathbb{X}$  and, similarly, in the critical cases since here  $\mathbb{X}$  anyway vanishes. Thus, provided the correct relation between

$\mathbb{X}$  and the norm of  $\hat{\psi}$  is enforced, the final term in the functional can be replaced by the integral of  $1 - \mathbb{X}$ . We (re)make one more observation: Consider the first two pieces, the integral of  $\xi(\hat{\gamma}, \hat{\Gamma})$ . It is noted that the contributors to the integrand satisfy

$$\hat{\gamma}_i \log \hat{\Gamma}_i = \frac{d}{dt} [\hat{\Gamma}_i \log \hat{\Gamma}_i - \hat{\Gamma}_i] \tag{4.61}$$

and

$$\hat{\gamma}^T A \hat{\Gamma} = \frac{1}{2} \frac{d}{dt} (\hat{\Gamma}^T A \hat{\Gamma}). \tag{4.62}$$

Therefore the integral of these objects depends only on  $\hat{\rho}$  and not the ascent itself. Thus, for all intents and purposes, these term may be omitted from our considerations.

Thus (assuming the appropriate relation between  $\hat{\psi}$  and  $\mathbb{X}$ ) the  $H$ -functional is furnished by

$$I(\hat{\Gamma}, \mathbb{X}) := \int_0^{|\hat{\rho}|} [\xi(\hat{\gamma}, \hat{\Gamma}) - \sum_{i=1}^L \hat{\gamma}_i \log \hat{\psi}_i + 1 - \mathbb{X}] dt \tag{4.63}$$

Now for given  $\mathbb{X} : [0, |\hat{\rho}|] \rightarrow [0, 1]$  (with  $\mathbb{X}(0) = 1$ ) and  $\hat{\psi}(\hat{\gamma}, t; \mathbb{X})$  that is defined by the formula in (4.2) *without regards to normalization* we shall consider  $I(\hat{\Gamma}, \mathbb{X})$ . Obviously

$$\sup_{\hat{\Gamma} \in \mathcal{P}(\hat{\rho}, \mathbb{X})} I(\hat{\Gamma}, \mathbb{X}) \geq \sup_{\hat{\Gamma} \in \mathcal{P}(\hat{\rho})} H(\hat{\Gamma}). \tag{4.64}$$

Eventually, we will establish that these are equal, for now, we consider the auxiliary problem of maximizing  $I(\cdot, \cdot)$

To this end, let  $(\hat{\Gamma}^{(n)}, \mathbb{X}^{(n)})$  denote a maximizing sequence for  $I(\cdot, \cdot)$  i.e.,

$$\lim_{n \rightarrow \infty} I(\hat{\Gamma}^{(n)}, \mathbb{X}^{(n)}) = \sup_{\hat{\Gamma} \in \mathcal{P}(\hat{\rho}, \mathbb{X})} I(\hat{\Gamma}, \mathbb{X}).$$

In accord with the above convention, we use  $\hat{\gamma}^{(n)}$  as notation for the derivative of  $\hat{\Gamma}^{(n)}$ .

Since all quantities are bounded, we may extract convergent subsequences e.g., weak  $L^2$  convergence for the  $\hat{\gamma}$ 's and  $\mathbb{X}$ 's and e.g., uniform convergence for the  $\Gamma$ 's. We assume, for ease of notation, that the subsequence is in fact, the original sequence. We let the unadorned  $\hat{\gamma}$ ,  $\mathbb{X}$  and  $\hat{\Gamma}$  denote the sequential limit; it is noted that  $|\hat{\gamma}| = 1$  (since this is linear). Furthermore since the convergence of  $\hat{\Gamma}^{(n)}$  is rather strong it is clear that  $\hat{\Gamma} \in \mathcal{P}(\hat{\rho})$ . Thus in the first, second and also the fourth term the limit of the maximizing sequence agrees with the corresponding quantities evaluated at their limits.

Let us turn to the third item which is the most serious of the pieces. Consider the individual components:  $-\int_0^{|\hat{\rho}|} \hat{\gamma}_i \log \hat{\psi}_i dt$ . Let us, for simplicity, dispense with the minus sign – so the goal is minimization. We claim that as a functional of  $\hat{\Gamma}$ 's and  $\mathbb{X}$ 's this object (without the minus sign) is convex.

In particular, focusing on  $i = 1$  we claim that

$$\hat{x}_1 \log \hat{\psi}_1(\hat{x}, t; X) = \hat{x}_1 \log t \hat{x}_1 - \hat{x}_1 \log \left[ X + t \sum_{j=1}^L a_{i,j} \hat{x}_j \right]$$

is convex as a function of  $(\hat{x}, X)$  (provided all quantities are all in an appropriate range). Since the principal objects of interest concern differentiable functions and we are in finite dimensions, the following is sufficient: Let  $(\hat{x}(s), X(s))$  denote affine functions (which for  $s$  in some range keep all quantities positive). Then defining

$$F(s) := \hat{x}_1(s) \log t\hat{x}_1(s) - \hat{x}_1(s) \log[X(s) + t \sum_{j=1}^L a_{ij}\hat{x}_j(s)],$$

convexity of the full object is equivalent to convexity of  $F(s)$ . (This is easily seen by appeal to the definition of convexity according to *supporting planes*.) For brevity, let  $\hat{x}_1(s) = y_0 + sy' := y$  and  $X(s) + t \sum_j a_{ij}\hat{x}_j(s) = Z_0 + sZ' := Z$ . Then

$$F''(s) = \frac{1}{yZ^2}(y'Z - yZ')^2 \geq 0. \tag{4.65}$$

Having established convexity of  $\hat{\gamma}_i \log \hat{\psi}_i$  the remainder our treatment is straightforward. We may use a theorem of Mazur to construct convex combinations of the  $(\hat{\gamma}^{(n)}, \mathbb{X}^{(n)})$ , which we shall denote by  $(\hat{\gamma}^{(m)}, \mathbb{X}^{(m)})$ , that converge strongly to  $(\hat{\gamma}, \mathbb{X})$ . Then, e.g.,

$$\begin{aligned} \lim_{m \rightarrow \infty} - \int_0^{|\hat{\rho}|} \hat{\gamma}_1^{(m)} \log \hat{\psi}_1(\hat{\gamma}^{(m)}, \mathbb{X}^{(m)}) dt &\equiv - \int_0^{|\hat{\rho}|} \hat{\gamma}_1 \log \hat{\psi}_1 dt \\ &\geq \lim_{n \rightarrow \infty} - \int_0^{|\hat{\rho}|} \hat{\gamma}_1^{(n)} \log \hat{\psi}_1(\hat{\gamma}^{(n)}, \mathbb{X}^{(n)}) dt. \end{aligned} \tag{4.66}$$

And finally, manifestly,  $\lim_{n \rightarrow \infty} \int_0^{|\hat{\rho}|} \mathbb{X}^{(n)} dt = \int_0^{|\hat{\rho}|} \mathbb{X} dt$ .

We thus conclude that our limit  $(\hat{\Gamma}, \mathbb{X})$  is a legitimate maximizer of  $I(\cdot, \cdot)$ . It remains to show that  $\mathbb{X}$  and  $\hat{\psi}$  satisfy the appropriate relationship. To this end, let us start with the consideration of a generic  $t$  where  $\hat{\psi}$  is supposed to be critical but, nevertheless,  $\mathbb{X}$  is positive. Under these conditions (for two reasons) the  $t^{-1}\hat{\psi}$  would have norm less than 1. Let us consider the pointwise change if  $\mathbb{X} \rightarrow \mathbb{X} - \delta\mathbb{X}$  with  $0 < \delta\mathbb{X} \ll 1$ . Then to within  $o(\delta\mathbb{X})$ ,

$$-\hat{\gamma}_i \log \hat{\psi}_i \rightarrow -\hat{\gamma}_i \log \hat{\psi}_i - \frac{\hat{\gamma}_i \delta\mathbb{X}}{\mathbb{X} + t \sum_{j=1}^L a_{ij}\hat{\gamma}_j} = -\hat{\gamma}_i \log \hat{\psi}_i - t^{-1}\hat{\psi}_i \delta\mathbb{X}.$$

Upon summing, there is a loss of  $t^{-1}|\hat{\psi}|\delta\mathbb{X}$  which (due to the final term in  $I$  that is linear in  $\mathbb{X}$ ) is offset by a gain of  $\delta\mathbb{X}$ . Clearly, in the region that is supposed to be critical, the value of the functional can be improved unless  $\mathbb{X}$  is actually zero. A similar argument shows that in the subcritical regions, the functional can be improved unless  $\mathbb{X}$  is such that  $t^{-1}|\hat{\psi}| = 1$ .

It is now indeed evident that (4.64) holds as an equality and that there is a tangible maximizer for  $H(\cdot, 0, |\hat{\rho})$  in  $\mathcal{P}(\hat{\rho})$ . By Lemma 4.7, this maximizer is a natural ascent. We turn to uniqueness.

Suppose, then, that there are two maximizing ascents. We shall denote these ascents by  $(\hat{\gamma}, \mathbb{X})$  and  $(\hat{\eta}, \mathbb{Y})$  although we may now assume (for a.e.  $t$ ) the proper normalization of the corresponding  $\hat{\psi}$ -functions. We shall work in the language of the  $I$ -functional. Non-uniqueness for the maximizers implies that strict concavity of the  $-\hat{\gamma}_i \log \hat{\psi}_i$  term must fail, a.e., for any convex combination of the solutions (since otherwise the combination would

do better). Let  $t > 0$  denote a generic time and let us assume linearity of the  $\hat{\gamma}_i \log \hat{\psi}_i$  – term under convex combinations of minimizers. We denote by  $(\hat{\gamma}, \mathbb{X})$  and  $(\hat{\eta}, \mathbb{Y})$  the extreme solutions and  $(\hat{\zeta}(s), \mathbb{W}(s))$  the linear interpolation with interpolation parameter  $s$ . Then (4.65) implies

$$\frac{1}{t} \hat{\psi}_i(\hat{\zeta}(s)) \equiv \frac{\hat{\gamma}_j - \hat{\eta}_j}{\mathbb{X} - \mathbb{Y} + t \sum_{j=0}^L a_{ij}(\hat{\gamma}_j - \hat{\eta}_j)}$$

i.e., is independent of the parameter  $s$ . Applying (4.50) to  $\hat{\gamma}$ , then  $\hat{\eta}$  then  $\hat{\zeta}(s)$  itself – using the constancy of  $\hat{\psi}$  – yields that  $\mathbb{W}(s)$  is given by the corresponding convex combination of  $\mathbb{X}$  and  $\mathbb{Y}$ . With this in hand, the constancy of  $\hat{\psi}$  implies that the numerators and denominators in the expressions for  $\hat{\psi}(\hat{\gamma})$  and  $\hat{\psi}(\hat{\eta})$  are separately equal so, in particular  $\hat{\gamma} = \hat{\eta}$  and we conclude there is only one maximizing ascent. ■

### 4.5. Main Proofs

*Proof of Theorem 2.6.* Consider any (non-zero)  $\hat{\rho}$ . By the considerations of Theorem 4.2 and the definition of  $\mathbb{S}_r$ , in order to establish (2.15) it is sufficient to show

$$\sup_{\hat{\Gamma} \in \mathcal{P}(\hat{\rho})} H(\hat{\Gamma}, 0, |\hat{\rho}|) = \Psi(\hat{\rho}). \tag{4.67}$$

By Proposition 4.8, we can evaluate this supremum along natural ascents and, as will be demonstrated this indeed yields  $\Psi(\hat{\rho})$ .

It is observed that in the subcritical region of a natural ascent,  $\hat{\Gamma} \equiv \hat{\psi}$  and the integrand of  $H$  vanishes identically. We are done with the cases of subcritical (and critical)  $\hat{\rho}$ 's and turn to the supercritical cases.

Here we let  $\hat{\rho}^*$  denote the (unique) point of entry of the ascent into the critical region. Then, for  $t \geq |\hat{\rho}^*|$ , we have  $\hat{\psi}(t) \equiv \hat{\rho}^*$  and  $\hat{\Gamma}'(t) \equiv (\hat{\rho} - \hat{\rho}^*)/(|\hat{\rho}| - |\hat{\rho}^*|)$ . Then, from the definitions in (4.6) and (4.5), all quantities can be integrated directly which yields

$$\int_{|\hat{\rho}^*|}^{|\hat{\rho}|} \xi(\hat{\Gamma}', \hat{\Gamma}) - \xi(\hat{\Gamma}', \psi) dt = \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)}$$

where

$$\begin{aligned} \text{(I)} &= \sum_{i=1}^L [(\hat{\rho}_i \log \hat{\rho}_i - \hat{\rho}_i) - (\hat{\rho}_i^* \log \hat{\rho}_i^* - \hat{\rho}_i^*)], \\ \text{(II)} &= -\frac{1}{2}(\hat{\rho}^T A \hat{\rho} - [\hat{\rho}^*]^T A \hat{\rho}^*), \\ \text{(III)} &= -[|\hat{\rho}| - |\hat{\rho}^*|] \sum_{i=1}^L \frac{\hat{\rho}_i - \hat{\rho}_i^*}{|\hat{\rho}| - |\hat{\rho}^*|} \log \hat{\rho}_i^*, \\ \text{(IV)} &= |\hat{\rho}| - |\hat{\rho}^*|. \end{aligned} \tag{4.68}$$

(Item (IV) may seem a bit mysterious but it is recalled from earlier discussions that in the critical region, the final piece of the  $H$ -functional reduces to the integral of  $1 - \chi^{-1}$  and here  $\chi^{-1}$  is zero.)

The combination of (I) + (III) + (IV) yields  $\sum_{i=1}^L \hat{\rho}_i \log \hat{\rho}_i / \hat{\rho}_i^*$ . As for (II), we write

$$\frac{1}{2}(\hat{\rho}^T A \hat{\rho} - [\hat{\rho}^*]^T A \hat{\rho}^*) = \frac{1}{2}(\hat{\rho}^T A \hat{\rho} + [\hat{\rho}^*]^T A (\hat{\rho} - \hat{\rho}^*) - [\hat{\rho}^*]^T A \hat{\rho})$$

and use the fact that  $[\hat{\rho}^*]^T A (\hat{\rho} - \hat{\rho}^*) = |\hat{\rho}| - |\hat{\rho}^*|$ . The proof of (4.67) is complete and the uniform convergence clause of this theorem follows from Theorem 4.2. ■

**Corollary 4.9.** For  $\hat{\rho}$  supercritical,  $\Psi(\hat{\rho}) < 0$ .

*Proof.* While, in principal this follows directly from (2.14), we instead shall look at the integrand of  $H$  in the supercritical region. For simplicity, let us revert to the  $\hat{\psi}, \hat{\gamma}$  language i.e.,  $\hat{\Gamma}' = \hat{\gamma}, \hat{\Gamma} = \hat{\psi} + \tau \hat{\gamma}$  for  $\tau := t - t^* \geq 0$ . We have

$$[\xi(\hat{\gamma}, \hat{\Gamma})]_i := \hat{\gamma}_i \log(\hat{\psi}_i + \tau \hat{\gamma}_i) - \hat{\gamma}_i \sum_{j=1}^L a_{ij}(\hat{\psi}_j + \tau \hat{\gamma}_j) \tag{4.69}$$

so, of course  $[\xi(\hat{\gamma}, \hat{\Gamma}) - \xi(\hat{\gamma}, \hat{\psi})]_i$  vanishes at  $\tau = 0$ . Differentiating with respect to  $\tau$ :

$$\begin{aligned} \frac{d}{d\tau} [\xi(\hat{\gamma}, \hat{\Gamma}) - \xi(\hat{\gamma}, \hat{\psi})]_i &= \frac{\hat{\gamma}_i^2}{\hat{\psi}_i + \tau \hat{\gamma}_i} - \hat{\gamma}_i \sum_{j=1}^L a_{ij} \hat{\gamma}_j \\ &\leq \frac{\hat{\gamma}_i}{\hat{\psi}_i} \left[ \hat{\gamma}_i - \sum_{j=1}^L \hat{\psi}_i a_{ij} \hat{\gamma}_j \right] = 0 \end{aligned} \tag{4.70}$$

where the last identity follows from the fact that  $\hat{\gamma}$  is the critical (and maximal) eigenvector of  $B_{\hat{\psi}}^T$ . It is noted that, at least for some  $i$ , the inequality is strict for  $\tau > 0$ . ■

*Proof of Theorem 2.7.* Let  $\mathbb{B}_r$  be the event that all components in  $\mathcal{G}(n, \hat{\rho})$  are of size  $r$  or bigger, and recall that  $\mathbb{S}_r$  is the event that all components in  $\mathcal{G}(n, \hat{\rho})$  are of size smaller than  $r$ . Then for all  $\hat{\eta}$  with  $0 \leq \hat{\eta} \leq \hat{\rho}$ ,

$$\begin{aligned} P \left( \hat{\theta}(\epsilon n, n, \hat{\rho}) = \frac{1}{n} \lfloor \hat{\eta} n \rfloor \right) \\ = e^{O(1)} \left[ \prod_{i=1}^L \binom{\lfloor \hat{\rho}_i n \rfloor}{\lfloor \hat{\eta}_i n \rfloor} \right] P_{n, \hat{\eta}}(\mathbb{B}_{\epsilon n}) P_{n, \hat{\rho} - \hat{\eta}}(\mathbb{S}_{\epsilon n}) \prod_{i,j=1}^L \left( 1 - \frac{\alpha_{ij}}{n} \right)^{\hat{\eta}_i n (\hat{\rho}_j n - \hat{\eta}_j n)}, \end{aligned} \tag{4.71}$$

where the error term comes from rounding (and is uniformly bounded). Following the proof of Theorem 2.5, we note that for each  $\hat{\eta} \in \mathcal{S}$  we have  $\frac{1}{n} \lfloor \hat{\eta} n \rfloor \in \mathcal{S}$  for sufficiently large  $n$ . We will also use that for  $m > |\hat{\rho}|/\epsilon$ , we have  $K_1 \subset \mathbb{B}_{\epsilon n} \subset K_m$ . Now for a lower bound we fix a  $\hat{\eta} \in \mathcal{S}$  with  $\hat{\eta} > 0$ , combining Theorem 3.1, Theorem 2.6, and direct calculation gives

$$\begin{aligned} \liminf_{\epsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P \left( \hat{\theta}(\epsilon n, n, \hat{\rho}) = \frac{1}{n} \lfloor \hat{\eta} n \rfloor \right) \\ \geq S(\hat{\eta}, \hat{\rho}) + \Xi(\hat{\eta}) + \Psi(\hat{\rho} - \hat{\eta}) - \hat{\eta}^T A (\hat{\rho} - \hat{\eta}). \end{aligned} \tag{4.72}$$

For an upper bound, since  $\hat{\theta}(\epsilon n, n, \hat{\rho})$  can take only polynomially many values, we have

$$P(\hat{\theta}(\epsilon n, n, \hat{\rho}) \in \mathcal{S}) = e^{o(n)} \sup_{\hat{\eta} \in \mathcal{S}} P\left(\hat{\theta}(\epsilon n, n, \hat{\rho}) = \frac{1}{n} \lfloor \hat{\eta} n \rfloor\right). \tag{4.73}$$

Note also that for any nonzero  $\hat{\eta}$ , and particular (finite)  $n$  we have

$$P(\hat{\theta}(\epsilon n, n, \hat{\rho}) = \hat{\eta} \vee (1/n, \dots, 1/n)) \geq e^{o(1)} P(\hat{\theta}(\epsilon n, n, \hat{\rho}) = \hat{\eta}), \tag{4.74}$$

(although we would be willing to accept any sub-exponential correction). Thus using Corollary 3.5 we may replace the  $P_{n, \hat{\eta}}(\mathbb{B}_{\epsilon n})$  term in (4.71) with a uniform upper bound of  $\exp(n \Xi(\hat{\eta}) + o(n))$ . Furthermore, using Theorem 2.6, we may replace the  $P_{n, \hat{\rho} - \hat{\eta}}(\mathbb{S}_{\epsilon n})$  term with a uniform bound of  $\exp(o(n) + n \Psi(\hat{\rho} - \hat{\eta}))$ , in which the  $o(n)$  term requires us to first take  $n \rightarrow \infty$  followed by  $\epsilon \rightarrow 0$ .

Thus from (4.71) and (4.73) we have

$$\frac{1}{n} \log P(\hat{\theta}(\epsilon n, n, \hat{\rho}) \in \mathcal{S}) \leq o(1) + \sup_{\hat{\eta} \in \mathcal{S}} [S(\hat{\eta}, \hat{\rho}) + \Xi(\hat{\eta}) + \Psi(\hat{\rho} - \hat{\eta}) - \hat{\eta}^T A(\hat{\rho} - \hat{\eta})]. \tag{4.75}$$

Taking  $n \rightarrow \infty$ , followed by  $\epsilon \rightarrow 0$ , we get the desired upper bound, and prove the result. ■

*Proof of Theorem 2.1.* By Theorem 2.7, we will have proved Theorem 2.1 if we can show that

$$S(\hat{\eta}, \hat{\rho}) + \Xi(\hat{\eta}) + \Psi(\hat{\rho} - \hat{\eta}) - \hat{\eta}^T A(\hat{\rho} - \hat{\eta}) \tag{4.76}$$

is maximized only when  $\hat{\eta}_i = \hat{\theta}_i^*(\hat{\rho})$  for all  $i$ . The easiest way to see this is to use an idea mentioned in [4] following Theorem 2.1 therein: Dropping the  $\Psi$  term and taking exponentials, we find

$$e^{n[S(\hat{\eta}, \hat{\rho}) + \Xi(\hat{\eta}) - \hat{\eta}^T A(\hat{\rho} - \hat{\eta})]} = e^{o(n)} \prod_{i=1}^L \left[ \binom{\lfloor \hat{\rho}_i n \rfloor}{\lfloor \hat{\eta}_i n \rfloor} (1 - e^{-\sum_{j=1}^L \alpha_{ij} \hat{\eta}_j})^{\hat{\eta}_i n} (e^{-\sum_{j=1}^L \alpha_{ij} \hat{\eta}_j})^{(\hat{\rho}_i - \hat{\eta}_i) n} \right]. \tag{4.77}$$

Well known results regarding binomials give us that the right hand side is exponentially small unless  $\frac{\hat{\eta}_i}{\hat{\rho}_i} \approx 1 - e^{-\sum_{j=1}^L \alpha_{ij} \hat{\eta}_j}$ , which is to say that  $\hat{\eta}_i$  must satisfy (2.2) for all  $i$ . We can also get this result by maximizing  $S(\hat{\eta}, \hat{\rho}) + \Xi(\hat{\eta}, \hat{\rho}) - \hat{\eta}^T A(\hat{\rho} - \hat{\eta})$  directly, and we note that the maximum is zero (as, indeed, it must be). If  $\hat{\rho}$  is critical or subcritical, this finishes the proof, since  $\hat{\Psi}(\hat{\rho}) = 0$  (and Lemma 2.2 shows that  $\hat{\theta}^*(\hat{\rho}) = 0$ ).

If  $\hat{\rho}$  is supercritical, the  $\Psi$ -term may be negative, i.e., if  $\hat{\rho} - \hat{\eta}$ , is still supercritical. However, the contribution from the remaining terms certainly does not exceed the value acquired at  $\eta = \theta^*$  (which happens to be zero). Thus we are finished if we can show that  $\hat{\rho} - \hat{\theta}^*$  is *not* supercritical – again forcing  $\Psi = 0$ . Supposing, to the contrary that  $\hat{\rho} - \hat{\theta}^*$  is supercritical. Let  $\hat{\kappa}$  denote the solution to the appropriate mean-field equation:

$$\hat{\kappa}_j = (\hat{\rho}_j - \hat{\theta}_j^*) (1 - e^{-\sum_{i=1}^L \alpha_{ij} \hat{\kappa}_i})$$

purported to be not identically zero. Then, consider  $\hat{\tau}$  which is given, componentwise, by the sum  $\hat{\tau}_i = \hat{\theta}_i^* + \hat{\kappa}_i$ . Then it is not difficult to show that  $\hat{\tau}$  satisfies the original mean-field equation, 2.2 i.e.,  $\hat{\tau}_j = \hat{\rho}_j(1 - e^{-\sum_{i=1}^L \alpha_{ij}\hat{\tau}_i})$ . Thus, the conclusion that  $\hat{\kappa}$  is not identically contradicts the maximality – and uniqueness – of the  $\hat{\theta}^*$  solution to the mean-field equation, (2.2).

We may thus conclude that,  $\hat{\rho} - \hat{\theta}^*$  is indeed *not* supercritical and, therefore, in the cases where  $\hat{\rho}$  is supercritical, the maximum of  $S(\hat{\eta}, \hat{\rho}) + \Xi(\hat{\eta}) + \Psi(\hat{\rho} - \hat{\eta}) - \hat{\eta}^T A(\hat{\rho} - \hat{\eta})$  is only achieved at  $\hat{\eta} = \hat{\theta}^*$ . ■

### 5. THE FINAL STAGE

We finish this note with a proof of Theorem 2.8 which requires one additional preliminary lemma:

**Lemma 5.1.** *Let  $\mathcal{K}_\epsilon$  (AKA  $\mathbb{B}_{\epsilon n}$ ) be the event that all components of  $\mathcal{G}(n, \hat{\rho})$  are of size at least  $\epsilon n$ , and recall that  $K_1$  is the event that  $\mathcal{G}(n, \hat{\rho})$  is connected. Then for all  $\hat{\rho}^{(1)}$  and  $\hat{\rho}^{(2)}$  with  $\hat{\rho}^{(2)} \geq \hat{\rho}^{(1)} > 0$  and sufficiently small  $\epsilon > 0$ ,*

$$\limsup_{n \rightarrow \infty} \sup_{\hat{\rho}: \hat{\rho}^{(1)} \leq \hat{\rho} \leq \hat{\rho}^{(2)}} \frac{1}{n} \log P_{n, \hat{\rho}}(K_1^c \mid \mathcal{K}_\epsilon) < 0. \tag{5.1}$$

*Proof.* Given vertex sets  $\mathcal{X}, \mathcal{Y} \subset \mathcal{V}$ , let  $K_m|_{\mathcal{X}}$  represent the event that  $\mathcal{G}(n, \hat{\rho})$  has  $m$  or fewer components when restricted to  $\mathcal{X}$ , and let  $\mathcal{X} \leftrightarrow \mathcal{Y}$  represent the event that there are no edges between vertices of  $\mathcal{X}$  and  $\mathcal{Y}$  in  $\mathcal{G}(n, \hat{\rho})$ . Then with  $m = \lfloor \hat{\rho} \rfloor / \epsilon$ , we have

$$\mathcal{K}_\epsilon \cap K_1^c \subset \bigcup_{\mathcal{A} \subset \mathcal{V}: |\mathcal{A}| \geq \epsilon n, |\mathcal{V} \setminus \mathcal{A}| \geq \epsilon n} \left[ (\mathcal{A} \leftrightarrow \mathcal{V} \setminus \mathcal{A}) \& K_m|_{\mathcal{A}} \& K_m|_{\mathcal{V} \setminus \mathcal{A}} \right], \tag{5.2}$$

and so

$$P_{n, \hat{\rho}}(\mathcal{K}_\epsilon \cap K_1^c) \leq e^{o(n)} \sup_{\hat{\eta}: \lfloor \hat{\eta} \rfloor \geq \epsilon n, \lfloor \hat{\rho} - \hat{\eta} \rfloor \geq \epsilon n} \binom{\lfloor \hat{\rho} \rfloor n}{\lfloor \hat{\eta} \rfloor n} P_{n, \hat{\rho}}(K_m|_{\mathcal{A}}) P_{n, \hat{\rho}}(K_m|_{\mathcal{V} \setminus \mathcal{A}}) e^{-n \hat{\eta}^T A(\hat{\rho} - \hat{\eta})}, \tag{5.3}$$

where  $\mathcal{A} \subset \mathcal{V}$  is a set with  $|\mathcal{A}| = \lfloor \hat{\eta} \rfloor n$ . Since the ‘‘cost’’ of moving a vertex from  $\mathcal{A}$  to  $\mathcal{V} \setminus \mathcal{A}$  (or the reverse direction) is at most polynomial, we can add the requirement that both  $\hat{\eta} n$  and  $\lfloor \hat{\rho} \rfloor n - \lfloor \hat{\eta} \rfloor n$  are bounded below by  $(1, \dots, 1)$  without altering the error term. By applying Corollary 3.5 (and immediately dropping the added requirement on  $\hat{\eta}$ ), we have

$$P_{n, \hat{\rho}}(\mathcal{K}_\epsilon \cap K_1^c) \leq e^{o(n)} \sup_{\hat{\eta}: \lfloor \hat{\eta} \rfloor \geq \epsilon, \lfloor \hat{\rho} - \hat{\eta} \rfloor \geq \epsilon} \prod_{i=1}^L \binom{\lfloor \hat{\rho}_i \rfloor n}{\lfloor \hat{\eta}_i \rfloor n} e^{n(\Xi(\hat{\eta}) + \Xi(\hat{\rho} - \hat{\eta}) - (\hat{\rho} - \hat{\eta})^T A \hat{\eta})}. \tag{5.4}$$

From this and Theorem 3.1, if we let

$$h(\hat{\gamma}) = \sum_{i=1}^L \hat{\gamma}_i \log \frac{1 - e^{-\sum_{j=1}^L \alpha_{ij} \hat{\gamma}_j}}{\hat{\gamma}_i} + \frac{1}{2} \hat{\gamma}^T A \hat{\gamma}, \tag{5.5}$$

we get (using uniform convergence)

$$\limsup_{n \rightarrow \infty} \sup_{\hat{\rho}: \hat{\rho}^{(1)} \leq \hat{\rho} \leq \hat{\rho}^{(2)}} \frac{1}{n} \log P_{n, \hat{\rho}}(K_1^c | \mathcal{K}_{\epsilon, 2}) \leq \sup_{\substack{\hat{\rho}: \hat{\rho}^{(1)} \leq \hat{\rho} \leq \hat{\rho}^{(2)} \\ \hat{\eta}: |\hat{\eta}| \geq \epsilon n, |\hat{\rho} - \hat{\eta}| \geq \epsilon n}} h(\hat{\eta}) + h(\hat{\rho} - \hat{\eta}) - h(\hat{\rho}). \tag{5.6}$$

By writing  $\hat{x}_i = t \sum_{j=1}^L \alpha_{ij} \hat{\eta}_j$  and  $\hat{y}_i = t \sum_{j=1}^L \alpha_{ij} \hat{\rho}_j$ , we find, after some calculations,

$$\begin{aligned} & \frac{d}{dt} \left[ \frac{1}{t} (h(t\hat{\eta}) + h(t(\hat{\rho} - \hat{\eta})) - h(t\hat{\rho})) \right] \\ &= \frac{1}{2t} \sum_{i=1}^L \left[ \hat{\eta}_i \hat{x}_i \frac{1 + e^{-\hat{x}_i}}{1 - e^{-\hat{x}_i}} + (\hat{\rho}_i - \hat{\eta}_i)(\hat{x}_i - \hat{y}_i) \frac{1 + e^{-(\hat{x}_i - \hat{y}_i)}}{1 - e^{-(\hat{x}_i - \hat{y}_i)}} - \hat{\rho}_i \hat{y}_i \frac{1 + e^{-\hat{y}_i}}{1 - e^{-\hat{y}_i}} \right]. \end{aligned} \tag{5.7}$$

It is not hard to show that  $x \frac{1+e^{-x}}{1-e^{-x}} = x \coth(x/2)$  is strictly increasing on  $(0, \infty)$ , and so (5.7) shows that  $t^{-1}[h(t\hat{\eta}) + h(t(\hat{\rho} - \hat{\eta})) - h(t\hat{\rho})]$  is decreasing in  $t$  for any applicable choice of  $\hat{\eta}$ . Then by taking  $t \rightarrow 0$ , we find

$$h(\hat{\eta}) + h(\hat{\rho} - \hat{\eta}) - h(\hat{\rho}) < g(\hat{\eta}) + g(\hat{\rho} - \hat{\eta}) - g(\hat{\rho}). \tag{5.8}$$

where  $g(\hat{\gamma}) = \sum_{i=1}^L \hat{\gamma}_i \log \frac{\sum_{j=1}^L \alpha_{ij} \hat{\gamma}_j}{\hat{\gamma}_i}$ . As a final step, we differentiate with respect to  $\alpha_{kk}$ , which yields

$$\frac{d}{d\alpha_{kk}} (g(\hat{\eta}) + g(\hat{\rho} - \hat{\eta}) - g(\hat{\rho})) = \frac{(\hat{\eta}_k \hat{y}_k - \hat{\rho}_k \hat{x}_k)^2}{\hat{x}_k (\hat{y}_k - \hat{x}_k) \hat{y}_k}. \tag{5.9}$$

Thus  $g(\hat{\eta}) + g(\hat{\rho} - \hat{\eta}) - g(\hat{\rho})$  is increasing as  $\alpha_{kk} \rightarrow \infty$ , but a brief examination shows

$$\lim_{\alpha_{11}, \alpha_{22}, \dots, \alpha_{LL} \rightarrow \infty} g(\hat{\eta}) + g(\hat{\rho} - \hat{\eta}) - g(\hat{\rho}) = 0. \tag{5.10}$$

Combined with (5.8), this means that  $h(\hat{\eta}) + h(\hat{\rho} - \hat{\eta}) - h(\hat{\rho}) < 0$  for all applicable  $\hat{\rho}$  and  $\hat{\eta}$ . Since  $h$  is continuous and we are working on a compact set, this proves the lemma. ■

*Proof of Theorem 2.8.* From the independence of edges in  $\mathcal{G}(n, \hat{\rho})$ , we have

$$P_{n, \hat{\rho}} \left( Q_\epsilon \mid \hat{\theta}(\epsilon n, n, \hat{\rho}) = \frac{1}{n} \lfloor \hat{\eta} n \rfloor \right) = P_{n, \hat{\eta}}(K_1^c \mid \mathcal{K}_\epsilon). \tag{5.11}$$

The result then follows from Lemma 5.1. ■

**REFERENCES**

[1] K. B. Athreya and A. N. Vidyashankar, *Branching Processes*, 1999. Available at <http://citeseer.ist.psu.edu/211428.html>.  
 [2] B. Bollobás, G. Grimmett, and S. Janson, The random-cluster model on the complete graph, *Probab Theory Relat Fields* 104 (1996), 283–317.  
 [3] B. Bollobás, S. Janson, and O. Riordan, The phase transition in inhomogeneous random graphs, *Random Structures Algorithms* 31 (2007), 3–122.



- [4] M. Biskup, L. Chayes, and S. A. Smith, Large-deviations/thermodynamic approach to percolation on the complete graph, *Random Structures Algorithms* 31 (2007), 354–370.
- [5] H. Nakanishi and M. E. Fisher, Critical point shifts in films, *J Chem Phys* 78 (1983).
- [6] B. Söderberg, General Formalism for Inhomogeneous Random Graphs, *Phys Rev E* 66 (2002), 066121.
- [7] R. Zandi, A. Shackell, J. Rudnick, M. Kardar, and L. P. Chayes, Thinning of superfluid films below the critical point, *Phys Rev E* 76 (2007), 030601(R).