The Annals of Applied Probability
2013, Vol. 23, No. 2, 722–763
DOI: 10.1214/12-AAP854

LARGE DEVIATIONS FOR THE DEGREE STRUCTURE IN PREFERENTIAL ATTACHMENT SCHEMES[1]

BY JIHYEOK CHOI AND SUNDER SETHURAMAN

Syracuse University and University of Arizona

Preferential attachment schemes, where the selection mechanism is linear and possibly time-dependent, are considered, and an infinite-dimensional large deviation principle for the sample path evolution of the empirical degree distribution is found by Dupuis–Ellis-type methods. Interestingly, the rate function, which can be evaluated, contains a term which accounts for the cost of assigning a fraction of the total degree to an "infinite" degree component, that is, when an atypical "condensation" effect occurs with respect to the degree structure.

As a consequence of the large deviation results, a sample path a.s. law of large numbers for the degree distribution is deduced in terms of a coupled system of ODEs from which power law bounds for the limiting degree distribution are given. However, by analyzing the rate function, one can see that the process can deviate to a variety of atypical nonpower law distributions with finite cost, including distributions typically associated with sub and superlinear selection models.

1. Introduction and results. Preferential attachment processes are graph networks which evolve in time by linking at each time step a new node to a vertex in the existing graph with probability based on a selection function of the vertex's connectivity. Such schemes have a long history in various guises going back to [50] and [51]; cf. surveys [40, 49]. More recently, Barabási and Albert (BA) in [4] proposed that versions of these processes, where the selection function is an increasing function of the connectivity, may serve as models for growing real-world networks such as the world wide internet web, and types of social structures.

For instance, in a "friend network," a newcomer may have a predilection to link or become friends with an individual with high connectivity, or in other words, one who already has many friends. An important property of such reinforcing networks is that when the selection function is in a linear form, asymptotically as time grows, the proportions of nodes with degrees $1, 2, \dots, k, \dots$ converge to a power-law distribution $\langle q(k) : k \geq 1\rangle$ where $0 < \lim_{k\uparrow\infty} q(k)k^{\theta} < \infty$ for some $\theta > 0$. We will say that a network with such a law of large numbers (LLN) property is "scale free." Since it has been observed that the sampled empirical degree structure in

Received January 2012.

[1]Supported in part by NSF-09-06713 and NSA Q H982301010180.

MSC2010 subject classifications. Primary 60F10; secondary 05C80.

Key words and phrases. Preferential attachment, random graphs, degree distribution, large deviations, time-dependent, law of large numbers, power law, condensation.

many real-world networks has a "scale-free" form, such preferential attachment processes have become quite popular in several ways; see [1, 3, 5, 12, 14, 16, 17, 26, 30, 40, 43–45] and references therein.

To illustrate more clearly the possible phenomena, consider the following basic example.

EXAMPLE 1.1. Initially, at time $n = 1$, the network G_1 is composed of two vertices with a single (undirected) edge between them. At time $n = 2$, a new vertex is attached to one of the two vertices in G_1 with probability proportional to a function of its degree to form the new network G_2. This scheme continues: more precisely, at time $n + 1$, a new node is linked to vertex $x \in G_n$ with probability proportional to $w(d_x(n))$, that is, chance $w(d_x(n))/\sum_{y\in G_n} w(d_y(n))$, where $d_z(n)$ is the degree at time n of vertex z, and $w = w(d) : \mathbb{N} \to \mathbb{R}_+$ is the selection function.

In this way, since the initial graph is a tree, all later networks G_n for $n \geq 1$ are also trees. Let now $\mathcal{Z}_k(n)$ be the number of vertices in G_n with k links, $\mathcal{Z}_k(n) = \sum_{y\in G_n} 1(d_y(n) = k)$. We now describe a trichotomy of growth behaviors corresponding to the strength and type of the selection function w [36].

First, when w is linear, say $w(d) = d + \alpha$ for $\alpha > -1$, the system is "scale-free." As is well understood in the literature (cf. [30], Chapter 4), the mean values $\langle M_k(n) = E[\mathcal{Z}_k(n)] : k \geq 1\rangle$ satisfy rate equations in the time index $n \geq 1$ which can be solved to show $\lim_{n\uparrow\infty} M_k(n)/n = q(k)$ for $k \geq 1$ where q is in power-law form with $\theta = 3 + \alpha$.

Later, in [10], when $\alpha = 0$, a concentration inequality was used to show convergence in probability, $\lim_{n\uparrow\infty} \mathcal{Z}_k(n)/n = q(k)$ for $k \geq 1$. We will call the $\alpha = 0$ model the "classical BA process" as it was the model originally analyzed in [4]. Also, for all $\alpha > -1$, Pólya urn/martingale ideas, and embeddings into branching processes have given alternative proofs which yield a.s. convergence; see [2, 41, 48].

However, in the sublinear case, when $w(d) = d^r$ for $0 < r < 1$, although it was shown that a.s. $\lim_{n\uparrow\infty} \mathcal{Z}_k(n)/n = q(k)$, this LLN limit q is not a power law, but in stretched exponential form [36, 48]: for $k \geq 1$,

$$q(k) = \frac{\mu}{k^r}\prod_{j=1}^{k}\left(1+\frac{\mu}{j^r}\right)^{-1} \quad \text{and}$$

$$\mu \text{ is determined by } 1 = \sum_{k=1}^{\infty}\prod_{j=1}^{k}\left(1+\frac{\mu}{j^r}\right)^{-1}. \tag{1.1}$$

Asymptotically, $\log q(k) \sim -(\mu/(1-r))k^{1-r}$ as $k \uparrow \infty$. On the other hand, when $r = 0$, the case of uniform attachment when an old vertex is selected uniformly, an a.s. LLN can also be similarly obtained where q is geometric. $q(k) = 2^{-k}$ for $k \geq 1$.

In the superlinear case, when $w(d) = d^r$ for $r > 1$, "explosion" or a sort of "condensation effect"' happens in that in the limiting graph a random single vertex dominates in accumulating connections. In particular, the limiting graph is a tree where there is a single random vertex with an infinite number of children; all other vertices have bounded degree, and of these only a finite number have degree strictly larger than $r/(r-1)$; cf., for a more precise description, [36, 46]. Moreover, a LLN limit, $\lim_{n\uparrow\infty} E\mathcal{Z}_k(n)/n = q(k)$, is argued where q is degenerate in that $q(1) = 1$ but $q(k) = 0$ for $k \geq 2$; cf. [36] and [30], Chapter 4. Such a limit implies, in the superlinear selection process, that most of the nodes at step n are leaves.

Since the work of Barabási and Albert [4], much effort has been devoted to understand the degree and other structures in generalized versions of these graphs. A partial selection of this large literature includes: more on degree structure [23, 24, 31, 32, 34, 37, 38]; growth and location of the maximum degree [2, 21, 42]; spectral gap and cover time of a random walk on the graph [19, 39]; width and diameter [9, 22, 35]; graph limits [6, 8, 11, 47].

Connection between urns and degree structure. If, however, one focuses only on the degree structure of the growing network, then it may be helpful to view the degree distribution evolution in terms of "balls-in-bins" or "Pólya urn" models. For instance, in the previous example, every new connection that a vertex gains can be represented by a new ball added to a corresponding urn in a collection of urns. More precisely, at time $n = 1$, there are two urns, each possessing one ball, in the initial collection $U(0)$. At time $j + 1$, a new urn with one ball is included in the collection, and also one ball is added to an existing urn $x \in U(j)$ with probability proportional to $w(b_x)$ where b_x is the number of balls in urn x. Then, $\mathcal{Z}_k(n)$ translates to the number of urns in $U(n)$ with k balls for $k \geq 1$.

A comprehensive form of such an urn model was formulated by Chung, Handjani and Jungreis (CHJ) in [15], motivated by the work in [27] and [36] on the organization of web tree-graph models. See also [7] and [42] for other work connecting urns to degree structure.

The CHJ model is as follows. Given an initial finite collection of urns each containing one ball, at subsequent times, with probability p, a new urn with one ball is created and added to the collection or, with probability $1 - p$, a new ball is put in one of the existing urns x with probability proportional to $(b_x)^r$ where b_x is the number of balls in x. It was proved in [15], among other results, when $r = 1$ and $p > 0$, analogous to linear selection preferential attachment graphs, that the empirical urn size distribution converges to a power law with $\theta = 1 + (1 - p)^{-1}$.

In this context, our purpose is to study a generalized preferential attachment process of urns, where at each time step a new urn is created and a new ball is added to it or an existing urn according to a time dependent linear selection function, which includes the evolving degree structure of linear selection preferential attachment model discussed above, and also a version of the $r = 1$ CHJ urn model [15]. We defer to Section 1.1 the exact description of our scheme.

As mentioned in [28], understanding preferential attachment or urn systems, where the selection function depends on time, allows for more realistic models given real world networks are time-dependent. However, it appears most of the work on time-dependent schemes consists of rate equation formulations ([25], Section E), [36] and related work, in models where at each step a random number of links or balls may be added to the structure [2, 18].

Given this background, detailing the large deviation behavior of the degree distribution in time-dependent preferential attachment schemes is a natural problem which gives much understanding of typical and, in particular, atypical evolutions. We remark, even in the usual time-homogeneous models, large deviations of the degree structure is an open question.

Previous large deviation work in preferential attachment models has focused on one-dimensional objects, for instance, the number of leaves [13], or the degree growth of a single vertex with respect to dynamics where any vertex may attach to a newly added vertex with a small chance [21]. See references cited in [13] for large deviations work with respect to other types of random trees and balls-in-bins models.

Our main work in this article includes an infinite-dimensional sample path large deviation principle (LDP) for an array of empirical urn ball size distributions $\{\langle Z_k^n(j)/n : k \geq 0\rangle : 0 \leq j \leq n\}_{n\geq 1}$, when the initial configuration, not necessarily fixed, satisfies a limit condition (Theorem 1.4). Here, $Z_k^n(j)$ stands for the count of urns with k balls at time j in the nth row of the array. Part of these results is a finite-dimensional LDP with respect to the numbers of urns with less than d balls for $d < \infty$ (Theorem 1.2).

As an application of the large deviations results, we obtain an a.s. sample path LLN for the urn counts in terms of a system of coupled ODEs (Corollary 1.7), which, for homogeneous schemes complements fixed time LLNs mentioned earlier, and gives a different way to derive them aside from the rate equation method mentioned in Example 1.1. Finally, the LLN limit trajectories are shown to have power law-type behavior in terms of bounds (Corollary 1.9), although the general behavior can interpolate between these bounds; see Figure 1.

Interestingly, the infinite-dimensional rate function I^∞ can be calculated on scaled urn ball size path distributions $\xi = \{\langle \xi_k(t) : k \geq 0\rangle : 0 \leq t \leq 1\}$. Here, since in our model, exactly one ball is added to the urn collection at each microscopic time, $\xi_k(t)/(t + c)$ is the fraction of urns with size k at macroscopic time $t \geq 0$ where $c = \sum_{k\geq 0} \xi_k(0)$ is the initial mass, that is the scaled initial number of urns. It is natural then to ask which trajectories ξ have finite cost, $I^\infty(\xi) < \infty$.

It turns out "no mass can be lost," that is, all finite cost paths ξ are such that the proportions $\{\xi_k(t)/(t + c)\}_{k\geq 0}$ form a probability distribution, $\sum_{k\geq 0} \xi_k(t)/(t + c) \equiv 1$. Also, a variety of nonpower law distributions can be achieved with finite rate at any time $0 < t \leq 1$, including geometric and stretched exponential distributions discussed in Example 1.1.

Intriguingly, on the other hand, "some of the weight may be lost" in certain finite rate trajectories, that is, the scaled mean urn ball size of the system may satisfy a "weight loss" property at a time $0 < t \leq 1$, $\sum_{k\geq 0} k\xi_k(t)/(t+\tilde{c}) < 1$, where $\tilde{c} = \sum_{k\geq 0} k\xi_k(0)$ is the scaled initial total urn ball size, even though the prelimit quantity equals 1 at all steps in the urn growth scheme. We dub a trajectory ξ with this "weight loss" property at some time $0 < t \leq 1$ as being "condensed." For instance, a "condensed" path arises when $c = \tilde{c} = 0$ and a finite number of the urns take in eventually all the balls. In this case, almost all the urns created are empty, and the associated path satisfies $\xi_0(t) = t$ for $0 \leq t \leq 1$, $\xi_k(t) \equiv 0$ for $k \geq 1$, and $\sum_k k\xi_k(t) \equiv 0$. It turns out this path, associated with superlinear selection preferential attachment models (cf. Example 1.1), has finite cost.

Moreover, the rate function I^∞ contains a term which measures the cost of "condensation" when some of the flow of urn ball size in the scaling limit escapes toward urns with "infinite" size. In addition, we point out, at any time $0 \leq t \leq 1$, LLN distributions arising from either sublinear or superlinear selection preferential attachment models may be achieved with finite cost. One might interpret that although the linear selection process is typically "scale-free," since it is between, in a sense, sublinear and superlinear selection models, its atypical degree distribution structure may include the typical behavior of its sub and superlinear relatives. See Remark 1.5 and Example 1.6 for more details and discussion.

We also remark that the large deviations and other work are, with respect to the process, starting from either "small" or "large" initial configurations, that is, when the initial urn collection has $o(n)$ balls (e.g., finite), or when the size of the collection is on order n, respectively. It appears these initial configurations, which enter into all result statements, have not been considered before, in general.

The main idea for the results is to extend a variational control problem/weak convergence approach of Dupuis and Ellis (cf. [29]) to establish finite-dimensional LDPs in the time-dependent setting. Then, a projective limit approach, and some analysis to identify the rate function, is used to obtain the infinite-dimensional LDP. For the LLN and power-law corollaries, a coupled system of ODEs, which governs the typical degree distribution evolution, is identified, and analyzed.

To be concrete, we have focused upon models where the network is incremented by one urn and one ball each time, which include basic models. However, the methods here should be of use to analyze the large deviations of the degree structure in other combinatorial models with different increment structure: for instance, the evolving graph model discussed in [16], Chapter 3, where at each time with probability p a new vertex is preferentially attached to an old one, and with probability $1-p$, an edge is added between two old nodes selected preferentially, and the BA graphs where, instead of only one vertex, $m \geq 2$ vertices are introduced and preferentially connected at each time; cf. [30], Chapter 4.

1.1. *Model.* Let $p(t):[0,1]\to[0,1]$ and $\beta(t):[0,1]\to[0,\infty)$ be given functions. We define an urn configuration U as a finite collection of urns, each urn $x\in U$ containing a nonnegative number of balls b_x. We now specify an evolving array $\{U^n(j):0\le j\le n\}_{n\ge 1}$ of urn configurations by the following time-dependent iterative scheme. In the nth row of the array:

- Start at step 0, with a given initial urn configuration $U^n(0)$.
- At step $j+1\le n$, to form a new urn configuration $U^n(j+1)$, we first create and include a new urn with no ball. Then:
  - with probability $p(j/n)$, we place a new ball in this urn;
  - with probability $1-p(j/n)$, we place a new ball in one of the other urns $x\in U^n(j)$ with probability

$$\frac{b_x+\beta(j/n)}{\sum_{y\in U^n(j)}(b_y+\beta(j/n))}.$$

We will call, for urn $x\in U^n(j)$, the term $b_x+\beta(j/n)$ as the "weight" of the urn at time j in the nth row of the process. Let now $|U^n(j)|$ and $B^n(j)=\sum_{x\in U^n(j)} b_x$ be the total number of urns and balls in $U^n(j)$, respectively. Then, the number of urns $|U^n(j)|=|U^n(0)|+j$ and the total number of balls $B^n(j)=B^n(0)+j$. Also, the total weight of the configuration at time j is

$$s^n(j):=\sum_{y\in U^n(j)}\bigl(b_y+\beta(j/n)\bigr)=\bigl(1+\beta(j/n)\bigr)j+B^n(0)+\beta(j/n)|U^n(0)|.$$

The above urn scheme, as discussed in the Introduction, may be viewed in terms of the evolving degree structure in a preferential attachment random graph process with time-dependent selection function $w(d;j,n)=d+\beta(j/n)$. Here, the step of including a new empty urn and incrementing the number of balls in an old urn corresponds to an edge being placed between a new node, with degree 1, and an old vertex in the existing graph whose degree is consequently incremented. In particular, when p and β are in particular forms, we recover the following models:

(1) "Classical" BA process. When $p(t)\equiv 0$ and $\beta(t)\equiv 1$, the scheme is time-homogeneous. When the initial urn configuration consists of two empty urns, the probability of selecting an urn x with $k\ge 0$ balls at time $j\ge 0$ is $(k+1)/(2(j+1))$, which matches the selection process in the evolution of the degree structure in the BA preferential attachment graph scheme at times $j+1\ge 1$ with selection function $w(d)=d$, as discussed in Example 1.1, where urns with $k\ge 0$ balls correspond to vertices with degree $d=k+1\ge 1$.

(2) "Offset" BA processes. When $p(t)\equiv 0$ and $\beta(t)\equiv\beta\ge 0$, again the scheme is time-homogeneous, and urns with $k\ge 0$ balls correspond to vertices with degree $k+1\ge 1$. However, now the weight of an urn with k balls is $k+\beta$, in a sense "offset" from the classical BA scheme. Correspondingly, the urn selection scheme is the same as the growth process of the degree structure in the preferential attachment model with selection function $w(d)=d+\alpha$ with $\alpha=\beta-1$ as specified in Example 1.1.

(3) CHJ model of Pólya urns. When $p(t) \equiv p$ and $\beta(t) \equiv \beta \geq 0$, the evolution of the number of urns of size $k \geq 0$ corresponds to a version of the $r = 1$ CHJ model discussed in the Introduction. However, we note, in our model, an empty urn is added at each step with probability $1 - p$, and these empty urns are kept track of in our scheme. When $\beta = 0$, the dynamics of urns of size $k \geq 1$ is the $r = 1$ CHJ model since the empty urns have no weight, and once created, they cannot be selected to fill in later steps, and do not influence the structure of urns with $k \geq 1$ balls.

For $n \geq 1$, let $Z_i^n(j)$ be the number of urns in the nth row of the urn array process with $i \geq 0$ balls at time $0 \leq j \leq n$ and, for $d \geq 0$, let $\bar{Z}_{d+1}^n(j)$ denote the number of urns with more than d balls at time $0 \leq j \leq n$. These quantities satisfy

$$\sum_{i=0}^{d} Z_i^n(j) + \bar{Z}_{d+1}^n(j) = |U^n(0)| + j,$$

$$\sum_{i=0}^{d} i Z_i^n(j) + (d+1)\bar{Z}_{d+1}^n(j) \leq B^n(0) + j.$$

Define now vectors in $\mathbb{R}^{d+2}$,

$$\mathbf{f}_0^d := \langle 0, 1, 0, \ldots, 0\rangle, \qquad \mathbf{f}_i^d := \langle 1, 0, \ldots, 0, -1, 1, 0, \ldots, 0\rangle$$
$$\text{where } -1 \text{ is at the } (i+1)\text{th position for } 1 \leq i \leq d,$$
$$\mathbf{f}_{d+1}^d := \langle 1, 0, \ldots, 0\rangle.$$

For $\mathbf{y} = \langle y_0, \ldots, y_d, y_{d+1}\rangle \in \mathbb{R}^{d+2}$ and $0 \leq i \leq d+1$, denote

$$[\mathbf{y}]_i := \sum_{l=0}^{i} y_l.$$

Note that

$$\begin{aligned} &0 \leq [\mathbf{f}^d]_i \leq 1 \qquad \text{for } 0 \leq i \leq d, \\ &[\mathbf{f}^d]_{d+1} = 1 \quad \text{and} \quad 0 \leq \sum_{i=0}^{d+1} (1 - [\mathbf{f}^d]_i) \leq 1. \end{aligned} \tag{1.2}$$

Consider now the "truncated" degree distribution

$$\{\mathbf{Z}^{n,d}(j) := \langle Z_0^n(j), \ldots, Z_d^n(j), \bar{Z}_{d+1}^n(j)\rangle | 0 \leq j \leq n\},$$

where $\bar{Z}_{d+1}^n(j) = \sum_{k \geq d+1} Z_k^n(j) = j + |U^n(0)| - \sum_{k=0}^{d} Z_k^n(j)$, which forms a discrete time Markov chain with initial state $\mathbf{Z}^{n,d}(0)$ corresponding to the initial urn configuration $U^n(0)$ and one-step transition property,

$$\mathbf{Z}^{n,d}(j+1) - \mathbf{Z}^{n,d}(j) = \begin{cases} \mathbf{f}_0^d, & \text{with prob. } p(j/n) + \bigl(1 - p(j/n)\bigr)\dfrac{\beta(j/n)Z_0^n(j)}{s^n(j)} \\ & \quad \text{for } i = 0, \\ \mathbf{f}_i^d, & \text{with prob. } \bigl(1 - p(j/n)\bigr)\dfrac{(i + \beta(j/n))Z_i^n(j)}{s^n(j)} \\ & \quad \text{for } 1 \le i \le d, \\ \mathbf{f}_{d+1}^d, & \text{with prob. } \bigl(1 - p(j/n)\bigr)\biggl(1 - \dfrac{\sum_{i=0}^d (i + \beta(j/n))Z_i^n(j)}{s^n(j)}\biggr). \end{cases}$$

We also define the "full" degree distribution

$$\{\mathbf{Z}^{n,\infty}(j) := \langle Z_0^n(j), \dots, Z_d^n(j), \dots \rangle | 0 \le j \le n\},$$

which is also a Markov chain on $\mathbb{R}^\infty$ with increments

$$\mathbf{Z}^{n,\infty}(j+1) - \mathbf{Z}^{n,\infty}(j) = \begin{cases} \mathbf{f}_0^\infty, & \text{with prob. } p(j/n) + \bigl(1 - p(j/n)\bigr)\dfrac{\beta(j/n)Z_0^n(j)}{s^n(j)} \\ & \quad \text{for } i = 0, \\ \mathbf{f}_i^\infty, & \text{with prob. } \bigl(1 - p(j/n)\bigr)\dfrac{(i + \beta(j/n))Z_i^n(j)}{s^n(j)} \\ & \quad \text{for } i \ge 1, \end{cases}$$

where $f_0^\infty = \langle 0, 1, 0, \dots, 0, \dots \rangle$ and $f_i^\infty = \langle 1, 0, \dots, 0, -1, 1, 0, \dots, 0, \dots \rangle$ with the "-1" being in the $(i+1)$th place.

We will assume throughout the following initial condition, which ensures a LLN at time $t = 0$. With respect to constants $c_i^n, c^n, \tilde{c}^n \ge 0$, for $i \ge 0$, define

$$c_i^n := \frac{1}{n} Z_i^n(0), \qquad c^n := \sum_{i \ge 0} c_i^n$$

and

$$\tilde{c}^n := \sum_{i \ge 0} i c_i^n.$$

(LIM) For constants $c_i, c, \tilde{c} \ge 0$, we have

$$c_i := \lim_{n \uparrow \infty} c_i^n \quad \text{and} \quad \tilde{c} := \lim_{n \uparrow \infty} \tilde{c}^n = \sum_{i \ge 0} i c_i < \infty.$$

Consequently, $c := \lim_{n \uparrow \infty} c^n = \sum_{i \ge 0} c_i < \infty$.

In the previous sentence, the c^n limit follows from the uniform bound, $\sum_{i\ge A} c_i^n \le A^{-1}\sum_{i\ge 0} i c_i^n \to \tilde{c}/A$. Define also

$$\bar{c}^d := \sum_{i\ge d+1} c_i \quad \text{and} \quad \mathbf{c}^d := \langle c_0, \dots, c_d, \bar{c}^d\rangle.$$

We remark one can classify the initial configurations depending on when $c_i \equiv 0$ or when $c_i > 0$ for some $i \ge 0$.

- (Small configuration) $c_i \equiv 0$ for any $i \ge 0$. Here, the initial urn configurations are small in that their size is $o(n)$. This is the case when the initial configurations do not depend on n, for instance.
- (Large configuration) $c_i > 0$ for some $i \ge 0$. Here, the initial state is already a partly-developed configuration whose size is of order n.

We also note, when the initial urn configurations correspond to initial tree configurations in the corresponding preferential attachment process, some restrictions in the values of c_i arise. One may verify that a graph with n vertices with degrees $d_1, \dots, d_n$ is a tree exactly when $\sum_{i=1}^n d_i = 2(n-1)$. Hence, since in the initial graph of the nth row, the number of vertices equals $n\sum_{k\ge 0} c_k^n$, and the sum of degrees equals $n\sum_{k\ge 0}(k+1)c_k^n$ (recall the correspondence between urn sizes and degrees discussed in the Introduction), we have $n\sum_{k\ge 0}(k+1)c_k^n = 2(n\sum_{k\ge 0} c_k^n - 1)$. By (LIM), we have then $\tilde{c} = c$.

In addition, we note (LIM) specifies an initial limiting degree distribution which has full "weight" or in other words is not "condensed," that is, $\tilde{c} = \lim_{n\uparrow\infty}\tilde{c}^n = \sum_{i\ge 0} i c_i$. See Remark 1.8, however, for comments when the initial distribution is "condensed," that is, $\tilde{c} > \sum_{i\ge 0} i c_i$.

Our results will be on the family of processes $\mathbf{X}^{n,d} = \{\mathbf{X}^{n,d}(t) : 0 \le t \le 1\}$ and $\mathbf{X}^{n,\infty} = \{\mathbf{X}^{n,\infty}(t) : 0 \le t \le 1\}$ obtained by linear interpolation of the discrete-time Markov chains $\frac{1}{n}\mathbf{Z}^{n,d}(j)$ and $\frac{1}{n}\mathbf{Z}^{n,\infty}(j)$, respectively. For $0 \le t \le 1$, let

$$\mathbf{X}^{n,d}(t) := \frac{1}{n}\mathbf{Z}^{n,d}(\lfloor nt\rfloor) + \frac{nt - \lfloor nt\rfloor}{n}\big(\mathbf{Z}^{n,d}(\lfloor nt\rfloor + 1) - \mathbf{Z}^{n,d}(\lfloor nt\rfloor)\big),$$

$$\mathbf{X}^{n,\infty}(t) := \frac{1}{n}\mathbf{Z}^{n,\infty}(\lfloor nt\rfloor) + \frac{nt - \lfloor nt\rfloor}{n}\big(\mathbf{Z}^{n,\infty}(\lfloor nt\rfloor + 1) - \mathbf{Z}^{n,\infty}(\lfloor nt\rfloor)\big).$$

The trajectories $\mathbf{X}^{n,d}$ lie in $C([0,1];\mathbb{R}^{d+2})$, and are Lipschitz with constant at most 1, satisfying $\mathbf{X}^{n,d}(0) = \frac{1}{n}\mathbf{Z}^{n,d}(0)$. On the other hand, the infinite distribution $\mathbf{X}^{n,\infty} \in \prod_{i=0}^\infty C([0,1];\mathbb{R})$, considered with the product topology, where $\mathbf{X}^{n,\infty}(0) = \frac{1}{n}\mathbf{Z}^{n,\infty}(0)$. In both cases, although $\mathbf{X}^{n,d}(t)$ and $\mathbf{X}^{n,\infty}(t)$ are not necessarily probabilities because it is possible that we do not normalize by the total mass; they are, however, finite distributions.

We now specify the assumptions on $p(t)$ and $\beta(t)$ used for the main results.

(ND) p and β are piecewise continuous and, for constants p_0, β_0 and β_1,

$$0 \le p(\cdot) \le p_0 < 1 \quad \text{and} \quad 0 < \beta_0 \le \beta(\cdot) < \beta_1 < \infty.$$

We discuss (ND) more in the remark after Theorem 1.2.

We note, throughout the article, that we use conventions

$$
\begin{aligned}
&0\log 0 = 0/0 = 0\cdot\pm\infty = 1/\infty = 0,\\
&\pm 1/0 = \pm\infty \quad \text{and} \\
&E[X;\mathbb{A}] = \int_{\mathbb{A}} X\,dP.
\end{aligned}
\tag{1.3}
$$

1.2. *Results*. We now recall the statement of a large deviation principle (LDP). A sequence $\{X^n\}$ of random variables taking values in a complete separable metric space $\mathcal{V}$ satisfies the LDP with rate n and good rate function $J:\mathcal{V}\to[0,\infty]$ if for each $M<\infty$, the level set $\{x\in\mathcal{V}|J(x)\le M\}$ is a compact subset of $\mathcal{V}$, that is, J has compact level sets, and if the following two conditions hold:

(i) Large deviation upper bound. For each closed subset F of $\mathcal{V}$,

$$\limsup_{n\to\infty}\frac{1}{n}\log P\{X^n\in F\}\le -\inf_{x\in F} J(x).$$

(ii) Large deviation lower bound. For each open subset G of $\mathcal{V}$,

$$\liminf_{n\to\infty}\frac{1}{n}\log P\{X^n\in G\}\ge -\inf_{x\in G} J(x).$$

For $d\ge 0$, we now state the LDP for $\{\mathbf{X}^{n,d}(t)|0\le t\le 1\}$. Define the function $I_d:C([0,1];\mathbb{R}^{d+2})\to[0,\infty]$ given by

$$
\begin{aligned}
I_d(\varphi) = \int_0^1 &(1-[\dot\varphi(t)]_0)\log\frac{1-[\dot\varphi(t)]_0}{p(t)+(1-p(t))\frac{\beta(t)\varphi_0(t)}{(1+\beta(t))t+\tilde c+c\beta(t)}}\\
&+\sum_{i=1}^{d}(1-[\dot\varphi(t)]_i)\log\frac{1-[\dot\varphi(t)]_i}{(1-p(t))\frac{(i+\beta(t))\varphi_i(t)}{(1+\beta(t))t+\tilde c+c\beta(t)}}\\
&+\left(1-\sum_{i=0}^{d}(1-[\dot\varphi(t)]_i)\right)\log\frac{1-\sum_{i=0}^{d}(1-[\dot\varphi(t)]_i)}{(1-p(t))(1-\frac{\sum_{i=0}^{d}(i+\beta(t))\varphi_i(t)}{(1+\beta(t))t+\tilde c+c\beta(t)})}\,dt,
\end{aligned}
$$

where $\varphi(0)=\mathbf{c}^d$, $\varphi_i\ge 0$ is Lipschitz with constant 1 such that $0\le[\dot\varphi(t)]_i\le 1$ for $0\le i\le d$, $\sum_{i=0}^{d+1}\dot\varphi_i(t)=1$, $\sum_{i=0}^{d}(1-[\dot\varphi(t)]_i)=\sum_{i=0}^{d+1} i\dot\varphi_i(t)\le 1$ for almost all t, and the integral converges; otherwise, $I_d(\varphi)=\infty$. It will turn out that I_d is convex and is a good rate function.

To explain the last condition in the definition of I_d, note that $\varphi_{d+1}(t)$ represents the fraction of urns with size at least $d+1$, so that $(d+1)\varphi_{d+1}(t)$ is the truncated fraction of balls in these urns. Since the process increments by one ball at each step, it makes sense to specify $\sum_{i=0}^{d}(1-[\dot\varphi(t)]_i)=\sum_{i=0}^{d+1} i\dot\varphi_i(t)\le 1$ or that $\sum_{i=0}^{d+1} i\varphi_i(t)\le t+\tilde c$ if $I_d(\varphi)<\infty$.

The rate function can be understood as follows: in order for $\mathbf{X}^{n,d}$ to deviate to φ, at time t, the process should behave as if the increment probabilities v_i of $\mathbf{f}_i^d$ are such that the mean $\sum_{i=0}^d v_i \mathbf{f}_i^d + v_{d+1}\mathbf{f}_{d+1}^d = \dot{\varphi}$. In the proof of Theorem 1.2, we show $v_i = 1 - [\dot{\varphi}]_i$ for $0 \le i \le d$ and $v_{d+1} = 1 - \sum_{j=0}^d (1 - [\dot{\varphi}]_j)$. But, the natural evolution increment probabilities u_i, given the process is in state $\varphi(t)$, are $u_0 = p(t) + (1 - p(t))\frac{\beta(t)\varphi_0(t)}{(1+\beta(t))t+\tilde{c}+c\beta(t)}$, $u_i = (1 - p(t))\frac{(i+\beta(t))\varphi_i(t)}{(1+\beta(t))t+\tilde{c}+c\beta(t)}$ for $1 \le i \le d$ and $u_{d+1} = (1 - p(t))(1 - \frac{\sum_{i=0}^d (i+\beta(t))\varphi_i(t)}{(1+\beta(t))t+\tilde{c}+c\beta(t)})$. Then I_d is time integral of the relative entropies of these two increment probability measures.

Recall, for probability measures μ and ν, that the relative entropy of μ with respect to ν is defined as

$$R(\mu \| \nu) := \begin{cases} \displaystyle\int \log\left(\frac{d\mu}{d\nu}\right) d\mu, & \text{if } \mu \ll \nu, \\ \infty, & \text{otherwise.} \end{cases}$$

THEOREM 1.2 (Finite-dimensional LDP). *The $C([0,1]; \mathbb{R}^{d+2})$-valued sequence $\{\mathbf{X}^{n,d}\}$ satisfies an LDP with rate n and convex, good rate function I_d.*

REMARK 1.3. We now make comments on the underlying assumption (ND) and obtain the rate function at the fixed time $t = 1$.

(A) The assumption (ND) specifies that the process considered is "nondegenerate" in some sense. (ND) does not cover some "boundary" cases, for instance, when $p(t) \equiv 1$, the process is deterministic in that at each time, one places a new ball in a new urn. Also, when $\beta(t) \equiv 0$, urns without a ball have no weight; and, if in addition $p(t) \equiv 0$, then all new balls are placed into urns in the initial configuration. Although an LDP should hold in these and other "less degenerate" cases, the form of the rate function may differ in that some increments may not be possible.

On the other hand, assumption (ND) is natural with respect to the convergence estimates needed for the proof of the lower bound in the LDP. However, we point out the LDP upper bound holds without any of the boundedness assumptions on $p(\cdot)$ and $\beta(\cdot)$ in (ND).

Formally, when $\beta(t) \equiv \infty$, this is the case of "uniform," as opposed to preferential, selection of urns. The limit $\lim_{\beta\uparrow\infty} I_d$ corresponds to the rate function for this type of dynamic.

(B) One recovers the LDP at a fixed time, say $t = 1$, by the contraction principle with respect to continuous function $F : C([0,1]; \mathbb{R}^{d+2}) \to \mathbb{R}^{d+2}$ defined by $F(\varphi) = \varphi(1)$, so that $F(\mathbf{X}^{n,d}) = \mathbf{X}^{n,d}(1) = \frac{1}{n}\mathbf{Z}^{n,d}(n)$. Then Theorem 1.2 implies the LDP for $\frac{1}{n}\mathbf{Z}^{n,d}(n)$ with rate function given by the variational expression $K(x) = \inf\{I_d(\varphi) | \varphi(0) = \mathbf{c}^d, \varphi(1) = x\}$ which might be evaluated numerically; cf. [13] for calculations when $d = 0$.

We now extend the finite-dimensional LDP results to the infinite-dimensional case ($d = \infty$). Define for $\xi \in \prod_{i=0}^{\infty} C([0,1];\mathbb{R})$ the function

$$I^{\infty}(\xi) = \int_0^1 \lim_{d\to\infty}\Bigg[(1-[\dot{\xi}(t)]_0)\log\frac{1-[\dot{\xi}(t)]_0}{p(t)+(1-p(t))\frac{\beta(t)\xi_0(t)}{(1+\beta(t))t+\tilde{c}+c\beta(t)}}$$
$$+\sum_{i=1}^{d}(1-[\dot{\xi}(t)]_i)\log\frac{1-[\dot{\xi}(t)]_i}{(1-p(t))\frac{(i+\beta(t))\xi_i(t)}{(1+\beta(t))t+\tilde{c}+c\beta(t)}}$$
$$+\left(1-\sum_{i=0}^{d}(1-[\dot{\xi}(t)]_i)\right)\log\frac{1-\sum_{i=0}^{d}(1-[\dot{\xi}(t)]_i)}{(1-p(t))(1-\frac{\sum_{i=0}^{d}(i+\beta(t))\xi_i(t)}{(1+\beta(t))t+\tilde{c}+c\beta(t)})}\Bigg]\,dt$$

where $\xi_i(0) = c_i$, $\xi_i(t) \geq 0$ is Lipschitz with constant 1, $0 \leq [\dot{\xi}(t)]_i \leq 1$ for $i \geq 0$, $\frac{d}{dt}\sum_{i=0}^{\infty}\xi_i(t) = 1$ and $\lim_d[\sum_{i=0}^{d} i\dot{\xi}_i(t) + (d+1)(1-[\dot{\xi}(t)]_d)] = \sum_{i=0}^{\infty}(1-[\dot{\xi}(t)]_i) \leq 1$ for almost all t, and the integral converges; otherwise $I^{\infty}(\xi) = \infty$. It will turn out through a projective limit approach (cf. [20], Section 4.6) that I^{∞} is well defined, convex and a good rate function, and that the integrand limit exists because the term in square brackets is increasing in d.

THEOREM 1.4 (Infinite-dimensional LDP). *The $\prod_{i=0}^{\infty} C([0,1];\mathbb{R})$-valued sequence $\{\mathbf{X}^{n,\infty}\}$ satisfies an LDP with rate n and convex, good rate function I^{∞}.*

REMARK 1.5. From the result, degree distributions, not fully supported on the nonnegative integers, that is, when $\sum_{i\geq 0}\varphi_i(t) < t + c$ or, in other words, when the distribution specifies a positive fraction of urns with an infinite number of balls, cannot be achieved with finite cost in the evolution process. This stabilization of the "mass" is understood as follows. The fraction of urns with size larger than A at time $\lfloor nt\rfloor$ is bounded in terms of the fraction of balls in the system: $\sum_{k\geq A} Z_k^n(\lfloor nt\rfloor)/n \leq A^{-1}\sum_{k\geq 0} kZ_k^n(\lfloor nt\rfloor)/n \leq A^{-1}(\lfloor nt\rfloor/n + \tilde{c}^n) \leq A^{-1}(1+2\tilde{c})$ for all large n. Hence, for all realizations of the process, the fraction of infinite sized urns vanishes.

On the other hand, it seems some fraction of the total "weight" can indeed be lost in the evolution process with finite rate, that is, it may be possible to achieve a degree distribution at a time $0 < t \leq 1$ such that $\sum_{i=0}^{d} i\xi_i(t) < t + \tilde{c}$ although prelimit $\sum_{i=0}^{\infty} iZ_i^n(\lfloor nt\rfloor)/n = \lfloor nt\rfloor/n + \tilde{c}^n$. The interpretation is that it is possible to put a positive fraction of the balls into a few very large urns with finite cost, a sort of "condensation" effect noticed in the limiting evolution when the selection function is superlinear as mentioned in Example 1.1.

The last term in the integrand of the rate function, corresponding to the increment $\mathbf{f}_{d+1}^{d}$, measures the cost of choosing urns with very large size. In the $d \uparrow \infty$ limit, this last term may be viewed as the cost of "escape" of weight from urns with bounded size, or, in other words, the cost of the increment "$\langle 1, 0, \ldots, 0, \ldots\rangle$"

which corresponds to a new empty urn being included and very large sized urns being incremented. Some "condensed" finite rate evolutions are discussed in Example 1.6.

However, on the other hand, this type of "weight" loss or "condensation" cannot happen in the typical evolution—see Corollary 1.7.

EXAMPLE 1.6. Consider the "classical" BA model which follows the evolution of a random graph with preferential attachment selection function $w(d) = d$, noted in Example 1.1 and Section 1.1, which corresponds to the urn system when $\beta(t) \equiv 1$ and $p(t) \equiv 0$. Suppose that the initial configurations satisfy $c_i = 0$ for all $i \geq 0$.

We now compute the cost of distributions in form $\xi(t) = t\gamma$ where $\gamma = \langle \gamma_i : i \geq 0 \rangle$ where constants $\gamma_i \geq 0$ are such that

$$\sum_{i \geq 0} \gamma_i = 1 \quad \text{and} \quad \sum_{i \geq 0} i\gamma_i = \sum_{i \geq 0} (1 - [\gamma]_i) \leq 1.$$

Since, $\xi(t)$ is linear in t, calculation of the rate $I^\infty(\xi)$ simplifies considerably, and one evaluates the limit of the last term in the integrand of $I^\infty(\xi)$ as the time-independent quantity,

$$\begin{aligned}\lim_{d \uparrow \infty} &\left(1 - \sum_{i=0}^{d} (1 - [\dot{\xi}(t)]_i)\right) \log \frac{1 - \sum_{i=0}^{d} (1 - [\dot{\xi}(t)]_i)}{1 - (\sum_{i=0}^{d} (i+1)\xi_i(t))/(2t)} \\ &= \left(1 - \sum_{i \geq 0} i\gamma_i\right) \log 2,\end{aligned}$$

which gives the cost of the "increment $\langle 1, 0, \ldots, 0, \ldots \rangle$" when the dynamics attaches new vertices to very large hubs or places balls into already very large urns.

This cost is positive if $\sum_{i \geq 0} i\gamma_i < 1$, and, as discussed in the remark above, corresponds to the cost of forming urns/nodes with very large size/degree in the evolution process, a "condensation" effect. It follows then

$$I^\infty(\xi) = \sum_{i \geq 0} (1 - [\gamma]_i) \log \frac{1 - [\gamma]_i}{(i+1)\gamma_i/2} + \left(1 - \sum_{i \geq 0} i\gamma_i\right) \log 2. \tag{1.4}$$

In the case $\gamma_0 = 1$ and $\gamma_i = 0$ for $i \geq 1$, one observes $I^\infty(\xi) = \log 2$, and one can associate a graph evolution to achieve this degree or size distribution. For instance, one may grow a "star" tree configuration where all new vertices connect to the same vertex, or all balls are put in the same urn. If initially, there are only two vertices with degree 1 or two empty urns, then the "star" configuration at the nth step has probability 2^{-n} of occurring. As the degree/size structure at time n consists of n leaves/empty urns and one vertex with degree n or one urn with size $n - 1$, one observes the LLN limit for the degree/size sequence is $\xi(t) = t\gamma$, from which the rate evaluation follows.

As discussed in Example 1.1, this "condensed" configuration is the limit tree with respect to superlinear selection function $w(d) = d^r$ for $r > 2$. Moreover, as noted in the Introduction, all preferential attachment evolutions with respect to superlinear selection function $w(d) = d^r$ for $r > 1$ lead to degree distribution γ, that is, $E\mathcal{Z}_i(n)/n \to \gamma_i$, where $\gamma_0 = 1$ and $\gamma_i = 0$ for $i \geq 1$.

From formula (1.4), when γ is supported only on a finite number of indices, one sees that $I^\infty(\xi) < \infty$ exactly when there exists $i^* \geq 0$ such that $\gamma_i > 0$ for $i \leq i^*$. In particular, the "straight road" evolution, leading to trees where all nodes have degree 2, except for two leaves, or urn configurations consisting of single ball urns except for two empty urns, has infinite cost: start with two vertices with degree 1 or two empty urns. At step $j+1$, connect a new vertex to one of the two leaves, or add an empty urn and place a ball in one of the two empty urns in the configuration formed at step j. This configuration at time n has probability $1/n!$ of occurring, and in the LLN limit corresponds to $\xi(t) = t\gamma$, where $\gamma_0 = 0$, $\gamma_1 = 1$ and $\gamma_i = 0$ for $i \geq 2$, for which $I^\infty(\xi) = \infty$.

Even when no weight escapes, that is, $\sum_{i\geq 0} i\gamma_i = 1$, it may be noted that deviations to nonpower law urn size paths ξ are possible with finite rate. For instance, when $\gamma_i = 2^{-(i+1)}$ for $i \geq 0$, $I^\infty(\xi) = -\sum_{i\geq 0} \frac{1}{2^{i+1}} \log \frac{i+1}{2}$. When $\gamma_i = q(i+1)$ for $i \geq 0$ and q in form of the stretched exponential in (1.1), the LLN limit for the degree distribution with respect to sublinear selection preferential attachment, a calculation verifies that $\sum_{i\geq 0} i\gamma_i = 1$ and also $I^\infty(\xi) < \infty$.

We now turn to the LLN behavior which corresponds to the "zero-cost" trajectory. Consider the system of ODEs for $\varphi^d = \varphi$, with initial condition $\varphi(0) = \mathbf{c}^d$:

$$
\begin{aligned}
\dot{\varphi}_0(t) &= 1 - p(t) - \big(1 - p(t)\big)\frac{\beta(t)\varphi_0(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)},\\
\dot{\varphi}_1(t) &= p(t) + \big(1 - p(t)\big)\frac{\beta(t)\varphi_0(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\\
&\quad - \big(1 - p(t)\big)\frac{(1+\beta(t))\varphi_1(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)},\\
\dot{\varphi}_i(t) &= \big(1 - p(t)\big)\frac{(i-1+\beta(t))\varphi_{i-1}(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\\
&\quad - \big(1 - p(t)\big)\frac{(i+\beta(t))\varphi_i(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)} \qquad \text{for } 2 \leq i \leq d,\\
\dot{\varphi}_{d+1}(t) &= 1 - \sum_{i=0}^{d} \dot{\varphi}_i(t).
\end{aligned} \tag{1.5}
$$

Recall that a "Carathéodory" solution is an absolutely continuous function satisfying the ODEs a.a. t, and the initial condition, or equivalently a function satisfying the integral equation associated to the ODEs. One can readily integrate ODEs (1.5), and find a Carathéodory solution $\zeta^d(t) = \langle \zeta_0(t), \ldots, \zeta_d(t), \bar{\zeta}_{d+1}(t)\rangle$ [see formula (4.1)], which is unique from the following theorem. One extends to "$d = \infty$" setting by defining

$$\zeta^\infty(t) := \langle \zeta_0(t), \ldots, \zeta_d(t), \ldots\rangle.$$

We now state a LLN for $\mathbf{X}^{n,d}$ and $\mathbf{X}^{n,\infty}$ as a consequence of the LDP upper bound. As remarked in the Introduction, this LLN may also be obtained by rate equation formulations as in [36] and [30], Chapter 4.

COROLLARY 1.7 (LLN). *For $d \geq 0$, ζ^d is the unique Carathéodory solution to ODEs* (1.5) *with the initial condition $\varphi(0) = \mathbf{c}^d$, and also $I_d(\zeta^d) = 0$. Then, in the sup topology on $C([0,1];\mathbb{R}^{d+2})$, $\mathbf{X}^{n,d}(\cdot) \to \zeta^d(\cdot)$ a.s.*

As a consequence, we have in the product topology on $\prod_{i=0}^\infty C([0,1];\mathbb{R})$ that $\mathbf{X}^{n,\infty}(\cdot) \to \zeta^\infty(\cdot)$. Moreover, $\sum_{i=0}^\infty \zeta_i(t) = t + c$ and $\sum_{i=0}^\infty i\zeta_i(t) = t + \tilde{c}$, and hence no "weight" is lost in the LLN limit.

REMARK 1.8. The last equality, $\sum_{i\geq 0} i\zeta_i(t) = t + \tilde{c}$, requires the condition in (LIM) that the initial scaled degree distribution is not "condensed," that is, $\tilde{c} = \lim_{n\uparrow\infty} \tilde{c}^n = \sum_{i\geq 0} ic_i$. When the initial distribution is "condensed," that is, a strict Fatou limit $\tilde{c} = \lim_{n\uparrow\infty} \tilde{c}^n > \sum_{i\geq 0} ic_i$ occurs, the large deviation results Theorems 1.2, 1.4 and Corollary 1.7 (except for the last equality) still hold with the same notation and proofs. However, one can show by similar arguments as for the proof of the last equality in Corollary 1.7 that the LLN trajectory ζ^∞ will now be "condensed," that is, $s(t) = \sum_{i\geq 0} i\zeta_i(t) < t + \tilde{c}$ for $t \geq 0$. Moreover, for a constant $C = C(c, \tilde{c}, p_0, \beta_1, \beta_0) > 0$, one can see for all large t that

$$C\biggl(\tilde{c} - \sum_{i\geq 0} ic_i\biggr)t^{(1-p_0)/(1+\beta_1)} \leq t + \tilde{c} - s(t) \leq C^{-1}\biggl(\tilde{c} - \sum_{i\geq 0} ic_i\biggr)t^{1/(1+\beta_0)}.$$

We now consider the "scale-freeness" of ζ^∞. Although it seems difficult to control each ζ_i, nevertheless ζ^∞ has "power law" behavior, in terms of bounds on $[\zeta^\infty]_i$. In general, it appears ζ^∞ can interpolate between the bounds (cf. Figure 1; as a curiosity, we note a figure with a similar "bend" is found in [33] with respect to Facebook social network data).

COROLLARY 1.9 (Power law). *Assume $0 \leq p_{\min} \leq p(\cdot) \leq p_0 =: p_{\max} < 1$, and $0 < \beta_0 =: \beta_{\min} \leq \beta(\cdot) \leq \beta_{\max} := \beta_1$. Then, ζ^∞ is bounded between two power laws:*

For small configurations, for example, $c_k \equiv 0$, we have, for $i \geq 0$ and $t \geq 0$,

$$[\eta']_i t \leq [\zeta^\infty(t)]_i \leq [\eta]_i t.$$

For large configurations, for example, $c_k > 0$ for some $k \geq 0$, we have, for $i \geq 0$,

$$[\eta']_i\bigl(t + o(1)\bigr) \leq [\zeta^\infty(t)]_i \leq [\eta]_i\bigl(t + o(1)\bigr) \qquad \text{as } t \uparrow \infty.$$

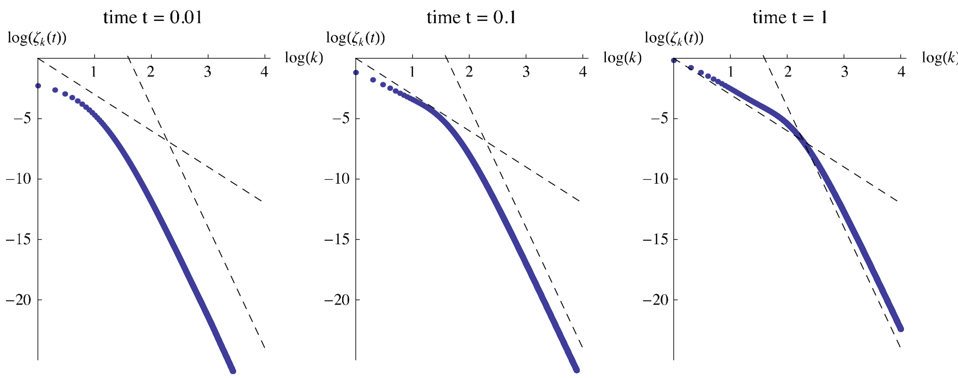

FIG. 1. *The thick curves are the* (*numerical*) *LLN ODE paths at times* $t = 0.01, 0.1, 1$ *with* $p(t) \equiv 0$, $\beta(t) = 8$ *for* $t < 0.01$, 1 *for* $t \geq 0.01$ *and* $c_k \equiv 0$. *Dashed lines are straight lines with slope* -3 *and* -10. *The plots use log–log scales*.

Here, *with respect to positive constants* C, C' *depending on* p *and* β,

$$\eta_i' := \frac{C'}{i^{1+(1+\beta_{\min})/(1-p_{\min})}}\bigl(1+o(1)\bigr)$$

and

$$\eta_i := \frac{C}{i^{1+(1+\beta_{\max})/(1-p_{\max})}}\bigl(1+o(1)\bigr).$$

The outline of the paper is that in Sections 2 and 3, we prove the finite and infinite-dimensional LDPs, Theorems 1.2 and 1.4. In Section 4, we prove the law of large numbers (Corollary 1.7). Finally, in Section 5, we discuss power-law behavior (Corollary 1.9).

2. Proof of Theorem 1.2. We follow the method and notation of Dupuis and Ellis in [29]; see also [52]. Some steps are similar to those in [13] where the "leaves" in a more simplified graph scheme are considered. However, as many things differ in our model, in the upper bound, and especially the lower bound proof, we present the full argument.

We now fix $0 \leq d < \infty$ and equip $\mathbb{R}^{d+2}$ with the L_1-norm denoted by $|\cdot|$. Recall, from assumption (LIM),

$$\mathbf{c}^{n,d} = (c_0^n, \dots, c_d^n, \bar{c}^{n,d}) := \frac{1}{n}\mathbf{Z}^{n,d}(0) \to \mathbf{c}^d,$$

where $\bar{c}^{n,d} = \sum_{i\geq d+1} c_i^n$. Denote

$$\vec{\xi}(n,t) := (p_n(t), \beta_n(t), \sigma_n(t)),$$

where

$$p_n(t) := p(\lfloor nt \rfloor/n), \qquad \beta_n(t) := \beta(\lfloor nt \rfloor/n),$$
$$\sigma_n(t) := \frac{1}{n}s^n(\lfloor nt \rfloor) = \bigl(1 + \beta_n(t)\bigr)\frac{\lfloor nt \rfloor}{n} + \tilde{c}^n + c^n\beta_n(t).$$

Let

$$\sigma(t) := \bigl(1 + \beta(t)\bigr)t + \tilde{c} + c\beta(t),$$
$$\vec{\xi}(t) := (p(t), \beta(t), \sigma(t)).$$

We note that, as $n \to \infty$, and $p(t)$ and $\beta(t)$ are piecewise continuous,

$$\vec{\xi}(n,t) \to \vec{\xi}(t) \qquad \text{for almost all } t \in [0, 1].$$

In the remainder of the section, when the context is clear, we often drop the superscript d to save on notation. Recall

$$\mathbf{X}^n(j) := \frac{1}{n}\mathbf{Z}^{n,d}(j),$$

$\mathbf{X}^n(0) = \mathbf{c}^{n,d}$ and $\mathbf{X}^n(j+1) = \mathbf{X}^n(j) + \frac{1}{n}y^n_{\mathbf{X}^n(j)}(j)$, where

$$y^n_{\mathbf{x}}(j) \text{ has distribution } \rho_{\vec{\xi}(n,j/n),\mathbf{x}}.$$

Here, for $\mathbf{x} = \langle x_0, \dots, x_d, x_{d+1}\rangle \in \mathbb{R}^{d+2}$ such that $x_i \geq 0$ for $0 \leq i \leq d+1$, numbers $p' \in [0, 1]$ and $\beta', \sigma' \geq 0$ such that $\sum_{i=0}^{d+1}(i + \beta')x_i \leq \sigma'$, and $A \subset \mathbb{R}^{d+2}$,

$$\begin{aligned}\rho_{(p',\beta',\sigma'),\mathbf{x}}(A) := {} & \biggl(p' + (1-p')\frac{\beta' x_0}{\sigma'}\biggr)\delta_{\mathbf{f}_0}(A) \\ & + \sum_{i=1}^{d}(1-p')\frac{(i+\beta')x_i}{\sigma'}\delta_{\mathbf{f}_i}(A) \\ & + (1-p')\biggl(1 - \frac{\sum_{i=0}^{d}(i+\beta')x_i}{\sigma'}\biggr)\delta_{\mathbf{f}_{d+1}}(A).\end{aligned}$$

We note when $\sigma' = 0$ and $\mathbf{x} = \langle 0, \dots, 0\rangle$, by convention $0/0 = 0$ and

$$\rho_{(p',\beta',0),\mathbf{x}}(A) := p'\delta_{\mathbf{f}_0}(A) + (1-p')\delta_{\mathbf{f}_{d+1}}(A).$$

From (1.2) and (LIM), for $A > 0$, the paths $\mathbf{X}^n(t) = \mathbf{X}^{n,d}(t)$, for all large n, belong to

$$\begin{aligned}\Gamma_{d,A} := \biggl\{ & \varphi \in C([0,1]; \mathbb{R}^{d+2}) \big| |\varphi(0) - \mathbf{c}^d| \leq A, \varphi_i \text{ is Lipschitz} \\ & \text{with bound } 1, 0 \leq [\dot{\varphi}(t)]_i \leq 1 \text{ for } 0 \leq i \leq d+1, \text{ and} \\ & \sum_{i=0}^{d+1}\dot{\varphi}_i(t) = 1, \sum_{i=0}^{d+1} i\dot{\varphi}_i(t) = \sum_{i=0}^{d}\bigl(1 - [\dot{\varphi}(t)]_i\bigr) \leq 1 \text{ for a.a. } t\biggr\}.\end{aligned} \tag{2.1}$$

Here, we equip $C([0,1];\mathbb{R}^{d+2})$ with the supremum norm.

Let $h : C([0,1];\mathbb{R}^{d+2}) \to \mathbb{R}$ be a bounded continuous function. Let also

$$W^n := -\frac{1}{n}\log E\{\exp[-nh(\mathbf{X}^n)]\}.$$

To prove Theorem 1.2, we need to establish Laplace principle upper and lower bounds (cf. [29], Section 1.2), namely upper bound

$$\liminf_{n\to\infty} W^n \geq \inf_{\varphi\in C([0,1];\mathbb{R}^{d+2})}\{I_d(\varphi)+h(\varphi)\}$$

for a good rate function I_d, and lower bound

$$\limsup_{n\to\infty} W^n \leq \inf_{\varphi\in C([0,1];\mathbb{R}^{d+2})}\{I_d(\varphi)+h(\varphi)\}.$$

Given $\mathbf{X}^n(0)=\mathbf{c}^{n,d}$, define, for $1\leq j\leq n$, that

$$\begin{aligned} &W^n(j,\{\mathbf{x}_1,\dots,\mathbf{x}_j\})\\ &\quad := -\frac{1}{n}\log E\{\exp[-nh(\mathbf{X}^n)]|\mathbf{X}^n(1)=\mathbf{x}_1,\dots,\mathbf{X}^n(j)=\mathbf{x}_j\}\end{aligned}$$

and

$$W^n := W^n(0,\varnothing) = -\frac{1}{n}\log E\{\exp[-nh(\mathbf{X}^n)]\}.$$

The Dupuis–Ellis method stems from the following discussion. From the Markov property, for $1\leq j\leq n-1$,

$$\begin{aligned} &e^{-nW^n(j,\{\mathbf{x}_1,\dots,\mathbf{x}_j\})}\\ &\quad = E\{e^{-nh(\mathbf{X}^n)}|\mathbf{X}^n(1)=\mathbf{x}_1,\dots,\mathbf{X}^n(j)=\mathbf{x}_j\}\\ &\quad = E\{E\{e^{-nh(\mathbf{X}^n)}|\mathbf{X}^n(1),\dots,\mathbf{X}^n(j+1)\}|\mathbf{X}^n(1)=\mathbf{x}_1,\dots,\mathbf{X}^n(j)=\mathbf{x}_j\}\\ &\quad = E\{e^{-nW^n(j+1,\{\mathbf{X}^n(1),\dots,\mathbf{X}^n(j),\mathbf{X}^n(j+1)\})}|\mathbf{X}^n(1)=\mathbf{x}_1,\dots,\mathbf{X}^n(j)=\mathbf{x}_j\}\\ &\quad = \int_{\mathbb{R}^{d+2}} e^{-nW^n(j+1,\{\mathbf{x}_1,\dots,\mathbf{x}_j,\mathbf{x}_j+\mathbf{y}/n\})}\rho_{\vec{\xi}(n,j/n),\mathbf{x}_j}(d\mathbf{y}).\end{aligned}$$

Recall the definition of relative entropy near Theorem 1.2. Then, by the variational formula for relative entropy (cf. [29], Proposition 1.4.2), for $1\leq j\leq n-1$,

$$\begin{aligned} &W^n(j,\{\mathbf{x}_1,\dots,\mathbf{x}_j\})\\ &\quad = -\frac{1}{n}\log\int_{\mathbb{R}^{d+2}} e^{-nW^n(j+1,\{\mathbf{x}_1,\dots,\mathbf{x}_j,\mathbf{x}_j+\mathbf{y}/n\})}\rho_{\vec{\xi}(n,j/n),\mathbf{x}_j}(d\mathbf{y})\\ &\quad = \inf_{\mu}\biggl\{\frac{1}{n}R(\mu\|\rho_{\vec{\xi}(n,j/n),\mathbf{x}_j})\\ &\qquad\qquad + \int_{\mathbb{R}^{d+2}} W^n\biggl(j+1,\biggl\{\mathbf{x}_1,\dots,\mathbf{x}_j,\mathbf{x}_j+\frac{1}{n}\mathbf{y}\biggr\}\biggr)\mu(d\mathbf{y})\biggr\}.\end{aligned}$$

We also have a terminal condition $W^n(n, \{\mathbf{x}_1, \dots, \mathbf{x}_n\}) = h(\mathbf{x}(\cdot))$, where $\mathbf{x}(\cdot)$ is the linear interpolated path connecting $\{(j/n, \mathbf{x}_j)\}_{0 \le j \le n}$.

We may understand these dynamic programming equations and terminal conditions in terms of a particular stochastic control problem. Define:

(i) $\mathcal{L}_j = (\mathbb{R}^{d+2})^j$, the state space on which $W^n(j, \cdot)$ is defined;
(ii) $\mathcal{U} = \mathcal{P}(\mathbb{R}^{d+2})$, where $\mathcal{P}(B)$ is the space of probabilities on B, is the control space on which the infimum is taken;
(iii) for $j = 0, \dots, n-1$, "control" $v_j^n(d\mathbf{y}) = v_j^n(d\mathbf{y}|\mathbf{x}_0, \dots, \mathbf{x}_j)$ which is a stochastic kernel on $\mathbb{R}^{d+2}$ given $(\mathbb{R}^{d+2})^j$;
(iv) $\{\bar{\mathbf{X}}^n(j); 0 \le j \le n\}$, the "controlled" process which is the adapted path satisfying $\bar{\mathbf{X}}^n(0) = \mathbf{c}^{n,d}$ and $\bar{\mathbf{X}}^n(j+1) = \bar{\mathbf{X}}^n(j) + \frac{1}{n}\bar{\mathbf{Y}}^n(j)$ for $0 \le j \le n-1$, where $\bar{\mathbf{Y}}^n(j)$, given $(\bar{\mathbf{X}}^n(0), \dots, \bar{\mathbf{X}}^n(j))$, has distribution $v_j^n(\cdot)$ [i.e., $\bar{P}\{\bar{\mathbf{Y}}^n(j) \in d\mathbf{y}|\bar{\mathbf{X}}^n(0), \dots, \bar{\mathbf{X}}^n(j)\} := v_j^n(d\mathbf{y}|\bar{\mathbf{X}}^n(0), \dots, \bar{\mathbf{X}}^n(j))$] and $\bar{\mathbf{X}}^n(\cdot)$ is the piecewise linear interpolation of $(j/n, \bar{\mathbf{X}}^n(j))$;
(v) "running costs" $C_j(v) = \frac{1}{n}R(v\|\rho)$ for $v \in \mathcal{P}(\mathbb{R}^{d+2})$; and
(vi) "terminal cost" equals to the function h.

Also, define, for $0 \le j \le n-1$, the minimal cost function

$$V^n(j, \{\mathbf{x}_1, \dots, \mathbf{x}_j\}) = \inf_{\{v_i^n\}} \bar{E}_{j,\mathbf{x}_1,\dots,\mathbf{x}_j} \left\{ \frac{1}{n} \sum_{i=j}^{n-1} R\big(v_i^n(\cdot) \| \rho_{\vec{\xi}(n,i/n),\bar{\mathbf{X}}^n(i)}\big) + h(\bar{\mathbf{X}}^n(\cdot)) \right\},$$

where $v_i^n(\cdot) = v_i^n(\cdot|\bar{\mathbf{X}}^n(0), \dots, \bar{\mathbf{X}}^n(i))$, and the infimum is taken over all control sequences $\{v_i^n\}$. Here, $\bar{E}_{j,\mathbf{x}_1,\dots,\mathbf{x}_j}$ denotes expectation, with respect to the adapted process $\bar{\mathbf{X}}^n(\cdot)$ associated to $\{v_i^n\}$, conditioned on $\bar{\mathbf{X}}^n(1) = \mathbf{x}_1, \dots, \bar{\mathbf{X}}^n(j) = \mathbf{x}_j$. The boundary conditions are $V^n(n, \{\mathbf{x}_1, \dots, \mathbf{x}_n\}) = h(\mathbf{x}(\cdot))$ and

$$V^n := V^n(0, \varnothing) = \inf_{\{v_j^n\}} \bar{E} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} R\big(v_j^n(\cdot) \| \rho_{\vec{\xi}(n,j/n),\bar{\mathbf{X}}^n(j)}\big) + h(\bar{\mathbf{X}}^n(\cdot)) \right\}. \tag{2.2}$$

It turns out that $\{V^n(j, \{\mathbf{x}_1, \dots, \mathbf{x}_j\}) : 0 \le j \le n\}$ also satisfies the dynamic programming equations and terminal condition, and since these equations have unique solutions (cf. [29], Section 3.2), we may conclude by [29], Corollary 5.2.1, that

$$W^n = -\frac{1}{n} \log E\{\exp[-nh(\bar{\mathbf{X}}^n(\cdot))]\} = V^n.$$

2.1. *Upper bound.* To prove the upper bound, it will be helpful to put the controls $\{v_j^n\}$ into continuous-time paths. Let $v^n(d\mathbf{y}|t) := v_j^n(d\mathbf{y})$ for $t \in [j/n, (j+1)/n)$, $j = 0, \dots, n-1$, and $v^n(d\mathbf{y}|1) := v_{n-1}^n$. Define

$$v^n(A \times B) := \int_B v^n(A|t)\, dt$$

for Borel $A \subset \mathbb{R}^{d+2}$ and $B \subset [0,1]$. Also define the piecewise constant path $\tilde{\mathbf{X}}^n(t) := \bar{\mathbf{X}}^n(j)$ for $t \in [j/n, (j+1)/n)$, $0 \le j \le n-1$, and $\tilde{\mathbf{X}}^n(1) := \bar{\mathbf{X}}^n(n-1)$. Then

$$W^n = V^n = \inf_{\{v_j^n\}} \bar{E}\left\{\int_0^1 R\big(v^n(\cdot|t)\|\rho_{\vec{\xi}(n,t),\tilde{\mathbf{X}}^n(t)}\big)\,dt + h(\bar{\mathbf{X}}^n)\right\}.$$

Given $\rho_{\vec{\xi},\mathbf{x}}$ is supported on $K := \{\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_{d+1}\}$, if $\{v_j^n\}$ is not supported on K, then $R(v^n\|\rho_{\vec{\xi},\mathbf{x}}) = \infty$. Since $|V^n| \le \|h\|_\infty < \infty$ and $K \subset \mathbb{R}^{d+2}$ is compact, for each n, there is $\{v_j^n\}$ supported on K and corresponding $v^n(d\mathbf{y} \times dt) = v^n(d\mathbf{y}|t) \times dt$ such that, for $\varepsilon > 0$,

$$W^n + \varepsilon = V^n + \varepsilon \ge \bar{E}\left\{\int_0^1 R\big(v^n(\cdot|t)\|\rho_{\vec{\xi}(n,t),\tilde{\mathbf{X}}^n(t)}\big)\,dt + h(\bar{\mathbf{X}}^n)\right\}. \tag{2.3}$$

Recall that $\bar{\mathbf{X}}^n(\cdot)$ takes values in $\Gamma_{d,A}$. Since $\Gamma_{d,A}$ is compact, by applications of the Ascoli–Arzelá theorem, and $\{v_j^n\}$ is tight, by Prokhorov's theorem, given any subsequence of $\{v^n, \bar{\mathbf{X}}^n\}$, there is a further subsubsequence, a probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$, a stochastic kernel v on $K \times [0,1]$ given $\bar{\Omega}$ and a random variable $\bar{\mathbf{X}}$ mapping $\bar{\Omega}$ into $\Gamma_{d,A}$ such that the subsubsequence converges in distribution to $(v, \bar{\mathbf{X}})$. In particular, since $\bar{\mathbf{X}}^n(0) = \mathbf{c}^{n,d} \to \mathbf{c}^d$ as $n \to \infty$, we have $\bar{\mathbf{X}}$ [cf. (2.1)] belongs to

$$\Gamma_d := \Gamma_{d,0} \qquad \text{those functions such that } \varphi(0) = \mathbf{c}^d.$$

Then, [29], Lemma 3.3.1, shows that v is a subsequential weak limit of v^n, and there exists a stochastic kernel $v(dy|t,\omega)$ on K given $[0,1] \times \bar{\Omega}$ such that $\bar{P}$-a.s. for $\omega \in \bar{\Omega}$,

$$v(A \times B|\omega) = \int_B v(A|t,\omega)\,dt.$$

Now, the same proof given for [29], Lemma 5.3.5, shows that $(v^n, \bar{\mathbf{X}}^n, \tilde{\mathbf{X}}^n)$ has a subsequential weak limit $(v, \bar{\mathbf{X}}, \bar{\mathbf{X}})$, where the last coordinate is with respect to Skorokhod space $D([0,1];\mathbb{R}^{d+2})$, and $\bar{P}$-a.s. for $t \in [0,1]$

$$\bar{\mathbf{X}}(t) = \int_{\mathbb{R}^{d+2}\times[0,t]} \mathbf{y}v(d\mathbf{y} \times ds) = \int_0^t \left(\int_K \mathbf{y}v(d\mathbf{y}|s)\right) ds,$$

$$\dot{\bar{\mathbf{X}}}(t) = \int_K \mathbf{y}v(d\mathbf{y}|t).$$

By Skorokhod's representation theorem, we may take that $(v^n, \bar{\mathbf{X}}^n, \tilde{\mathbf{X}}^n)$ converges to $(v, \bar{\mathbf{X}}, \bar{\mathbf{X}})$ a.s. In particular, $\bar{\mathbf{X}}^n \to \bar{\mathbf{X}}$ uniformly a.s., and as $\bar{\mathbf{X}}$ is continuous, it follows that also $\tilde{\mathbf{X}}^n \to \bar{\mathbf{X}}$ uniformly a.s.; cf. [29], Theorem A.6.5.

Let λ denote Lebesgue measure on $[0,1]$ and $\rho \times \lambda$ product measure on $K \times [0,1]$. Then [29], Lemma 1.4.3(f), yields

$$\int_0^1 R\big(v^n(\cdot|t)\|\rho_{\vec{\xi}(n,t),\tilde{\mathbf{X}}^n(t)}\big)\,dt = R\big(v^n(\cdot|t) \times \lambda(dt)\|\rho_{\vec{\xi}(n,t),\tilde{\mathbf{X}}^n(t)} \times \lambda(dt)\big).$$

We now evaluate the limit inferior of W^n using formula (2.3), along a subsequence as above:

$$\begin{aligned}\liminf_{n\to\infty} V^n + \varepsilon &\geq \liminf_{n\to\infty} \bar{E}\left\{\int_0^1 R\big(v^n(\cdot|t)\|\rho_{\vec{\xi}(n,t),\tilde{\mathbf{X}}^n(t)}\big)\,dt + h(\bar{\mathbf{X}}^n)\right\} \\ &= \liminf_{n\to\infty} \bar{E}\big\{R\big(v^n(\cdot|t)\times\lambda(dt)\|\rho_{\vec{\xi}(n,t),\tilde{\mathbf{X}}^n(t)}\times\lambda(dt)\big) + h(\bar{\mathbf{X}}^n)\big\} \\ &\geq \bar{E}\big\{R\big(v(\cdot|t)\times\lambda(dt)\|\rho_{\vec{\xi}(t),\bar{\mathbf{X}}(t)}\times\lambda(dt)\big) + h(\bar{\mathbf{X}})\big\} \\ &= \bar{E}\left\{\int_0^1 R\big(v(\cdot|t)\|\rho_{\vec{\xi}(t),\bar{\mathbf{X}}(t)}\big)\,dt + h(\bar{\mathbf{X}})\right\}.\end{aligned}$$

Note that we used Fatou's lemma in the second inequality, observing (i)–(iv).

(i) $v^n(d\mathbf{y}|dt)\times\lambda(dt)\to v(d\mathbf{y}|dt)\times\lambda(dt)$ a.s. as $v^n\Rightarrow v$ a.s.;

(ii) $\rho_{\vec{\xi}(n,t),\tilde{\mathbf{X}}^n(t)}\Rightarrow\rho_{\vec{\xi}(t),\bar{\mathbf{X}}(t)}$ as $\vec{\xi}(n,t)\to\vec{\xi}(t)$ a.a. $t\in[0,1]$, and $\tilde{\mathbf{X}}^n(t)\to\bar{\mathbf{X}}(t)$ uniformly on $[0,1]$ a.s.;

(iii) $\liminf_{n\to\infty} R(v^n(d\mathbf{y}|dt)\times\lambda(dt)\|\rho_{\vec{\xi}(n,t),\tilde{\mathbf{X}}^n(t)}\times\lambda(dt))\geq R(v(d\mathbf{y}|dt)\times\lambda(dt)\|\rho_{\vec{\xi}(t),\bar{\mathbf{X}}(t)}\times\lambda(dt))$ a.s. as R is lower semi-continuous;

(iv) $h(\bar{\mathbf{X}}^n)\to h(\bar{\mathbf{X}})$ a.s. as h is continuous and $\bar{\mathbf{X}}^n\to\bar{\mathbf{X}}$ uniformly on $[0,1]$ a.s.

By [29], Lemma 3.3.3(c),

$$R\big(v(\cdot|t)\|\rho_{\vec{\xi}(t),\bar{\mathbf{X}}(t)}\big)\geq L\left(\vec{\xi}(t),\bar{\mathbf{X}}(t),\int_K \mathbf{z}v(d\mathbf{z}|t)\right),$$

where

$$\begin{aligned}L(\vec{\xi}(t),\mathbf{x},\mathbf{y}) &:= \sup\left\{\langle\boldsymbol{\theta},\mathbf{y}\rangle - \log\int_K \exp\langle\boldsymbol{\theta},\mathbf{z}\rangle\rho_{\vec{\xi}(t),\mathbf{x}}(d\mathbf{z})\middle|\boldsymbol{\theta}\in\mathbb{R}^{d+2}\right\} \\ &= \inf\left\{R\big(\nu(\cdot|t)\|\rho_{\vec{\xi}(t),\mathbf{x}}\big)|\nu(\cdot|t)\in\mathcal{P}(K),\int_K \mathbf{z}\nu(d\mathbf{z}|t)=\mathbf{y}\right\}.\end{aligned}$$

We note, in this definition, the infimum is attained at some $\nu_0\in\mathcal{P}(K)$ as the relative entropy is convex and lower semicontinuous; cf. [29], Lemma 1.4.3(b). Since $\int \mathbf{z}v(d\mathbf{z}|t)=\dot{\bar{\mathbf{X}}}(t)$, we have

$$\liminf_{n\to\infty} V^n \geq \bar{E}\left\{\int_0^1 L(\vec{\xi}(t),\bar{\mathbf{X}}(t),\dot{\bar{\mathbf{X}}}(t))\,dt + h(\bar{\mathbf{X}})\right\}.$$

As $\bar{\mathbf{X}}\in\Gamma_d$, we have

$$\liminf_{n\to\infty} V^n \geq \inf_{\varphi\in\Gamma_d}\left\{\int_0^1 L(\vec{\xi}(t),\varphi(t),\dot{\varphi}(t))\,dt + h(\varphi)\right\}.$$

For $\varphi\in\Gamma_d$, we can evaluate a unique minimizer $\nu_0(\cdot|t)$ in the definition of $L(\vec{\xi}(t),\varphi(t),\dot{\varphi}(t))$: recall that $[\dot{\varphi}(t)]_i := \sum_{l=0}^i \dot{\varphi}_l(t)$. Then, as $\sum_{i=0}^{d+1}\mathbf{f}_i\nu_0(\mathbf{f}_i|t) =$

$\langle\dot\varphi_0(t),\dots,\dot\varphi_{d+1}(t)\rangle$, a calculation gives

$$\nu_0(\dot\varphi(t)|t)=\sum_{i=0}^{d}\bigl(1-[\dot\varphi(t)]_i\bigr)\delta_{\mathbf{f}_i}+\left(\sum_{i=0}^{d}[\dot\varphi(t)]_i-d\right)\delta_{\mathbf{f}_{d+1}}. \tag{2.4}$$

Hence,

$$\begin{aligned}
&L(\vec\xi(t),\varphi(t),\dot\varphi(t))\\
&\quad=R\bigl(\nu_0(\dot\varphi(t)|t)\,\|\,\rho_{\vec\xi(t),\varphi(t)}\bigr)\\
&\quad=\bigl(1-[\dot\varphi(t)]_0\bigr)\log\frac{1-[\dot\varphi(t)]_0}{p(t)+(1-p(t))\frac{\beta(t)\varphi_0(t)}{(1+\beta(t))t+\tilde c+c\beta(t)}}\\
&\qquad+\sum_{i=1}^{d}\bigl(1-[\dot\varphi(t)]_i\bigr)\log\frac{1-[\dot\varphi(t)]_i}{(1-p(t))\frac{(i+\beta(t))\varphi_i(t)}{(1+\beta(t))t+\tilde c+c\beta(t)}}\\
&\qquad+\left(1-\sum_{i=0}^{d}\bigl(1-[\dot\varphi(t)]_i\bigr)\right)\log\frac{1-\sum_{i=0}^{d}(1-[\dot\varphi(t)]_i)}{(1-p(t))(1-\frac{\sum_{i=0}^{d}(i+\beta(t))\varphi_i(t)}{(1+\beta(t))t+\tilde c+c\beta(t)})},
\end{aligned}\tag{2.5}$$

interpreted under our conventions (1.3).

Finally, define

$$I_d(\varphi):=\int_0^1 L(\vec\xi(t),\varphi(t),\dot\varphi(t))\,dt,$$

when $\varphi\in\Gamma_d$, and $I_d(\varphi)=\infty$ otherwise. Since L is convex, I_d is convex. Also I_d has compact level sets by the proof of [29], Proposition 6.2.4, and so is a good rate function. Hence, the Laplace principle upper bound holds with respect to I_d.

We will need the following result for the proof of the lower bound in the next section.

LEMMA 2.1. *Let $\ell(t)=\mathbf{e}t+\mathbf{c}^d$ be a linear function, where $\mathbf{e}=(e_0,e_1,\dots,e_{d+1})$ is such that $e_i>0$ for $i\ge 0$, $\sum_{i=0}^{d+1}e_i=1$, and $\sum_{i=0}^{d+1}ie_i\le 1$. Then, $I_d(\ell(t))<\infty$.*

PROOF. Noting $\sum_{i=0}^{d}(1-[\mathbf{e}]_i)=\sum_{i=0}^{d+1}ie_i\le 1$, explicitly

$$\begin{aligned}
I_d(\ell(t))=\int_0^1&(1-[\mathbf{e}]_0)\log\frac{1-[\mathbf{e}]_0}{p(t)+(1-p(t))\frac{\beta(t)(e_0t+c_0)}{(1+\beta(t))t+\tilde c+c\beta(t)}}\\
&+\sum_{i=1}^{d}(1-[\mathbf{e}]_i)\log\frac{1-[\mathbf{e}]_i}{(1-p(t))\frac{(i+\beta(t))(e_it+c_i)}{(1+\beta(t))t+\tilde c+c\beta(t)}}\\
&+\left(1-\sum_{i=0}^{d}(1-[\mathbf{e}]_i)\right)\log\frac{1-\sum_{i=0}^{d}(1-[\mathbf{e}]_i)}{(1-p(t))(1-\frac{\sum_{i=0}^{d}(i+\beta(t))(e_it+c_i)}{(1+\beta(t))t+\tilde c+c\beta(t)})}\,dt
\end{aligned}$$

is bounded under the bounds on p, β in assumption (ND). $\square$

2.2. *Lower bound.* Fix $h : C([0,1]; \mathbb{R}^{d+2}) \to \mathbb{R}$, a bounded, continuous function, and $\varphi^* \in \Gamma_d$ such that $I_d(\varphi^*) < \infty$. To show the lower bound, it suffices to prove, for each $\varepsilon > 0$, that

$$\limsup_{n\to\infty} V^n \le I_d(\varphi^*) + h(\varphi^*) + 8\varepsilon. \tag{2.6}$$

The main idea of the argument is to construct from φ^* a sequence of control measures suitable to evaluate formulas for V^n.

Note only in this "lower bound" subsection, to make several expressions simpler, we often take $c_{d+1} := \bar{c}^d$.

2.2.1. *Step* 1: *Convex combination and regularization.* Rather than work directly with φ^*, we consider a convex combination of paths with better regularity: for $0 \le \theta \le 1$, let

$$\varphi_\theta(t) = (1-\theta)\varphi^*(t) + \theta\ell(t),$$

where $\ell(t) = \mathbf{e}t + \mathbf{c}^d$ is a linear function such that $\mathbf{e}$ satisfies the assumptions of Lemma 2.1, say $\mathbf{e} = (\frac{1}{2}, \frac{1}{2^2}, \dots, \frac{1}{2^{d+1}}, \frac{1}{2^{d+1}})$.

LEMMA 2.2. *As* $\theta \downarrow 0$, *we have*

$$|I_d(\varphi_\theta) - I_d(\varphi^*)| \to 0 \quad \textit{and} \quad |h(\varphi_\theta) - h(\varphi^*)| \to 0.$$

PROOF. By convexity of I_d, and finiteness of $I_d(\ell(t))$ from Lemma 2.1,

$$I_d(\varphi_\theta) \le (1-\theta) I_d(\varphi^*) + \theta I_d(\ell).$$

On the other hand, since $|\varphi_\theta(t) - \varphi^*(t)| = |\int_0^t (\dot{\varphi}_\theta - \dot{\varphi}^*)(s)\, ds| \le 2t\theta(d+2)$, we have $\|\varphi_\theta - \varphi^*\|_\infty < 2\theta(d+2) \downarrow 0$, by lower semi-continuity of I_d, we have

$$\liminf_{\theta\downarrow 0} I_d(\varphi_\theta) \ge I_d(\varphi^*).$$

Also, as h is continuous, we have that $|h(\varphi_\theta) - h(\varphi^*)| \to 0$. $\square$

Now, fix $\theta > 0$ such that

$$I_d(\varphi_\theta) \le I_d(\varphi^*) + \varepsilon \quad \text{and} \quad h(\varphi_\theta) \le h(\varphi^*) + \varepsilon.$$

Next, for $\kappa \in \mathbb{N}$ and $t \in [0,1]$, define

$$\psi_\kappa(t) = \int_0^t \gamma_\kappa(s)\, ds + \mathbf{c}^d, \tag{2.7}$$

where

$$\gamma_\kappa(t) = \kappa \int_{i/\kappa}^{(i+1)/\kappa} \dot{\varphi}_\theta(s)\, ds$$

for $t \in [i/\kappa, (i+1)/\kappa)$, $0 \le i \le \kappa - 1$, and $\gamma_\kappa(1) = \gamma_\kappa(1 - 1/\kappa)$. Note that $\psi_\kappa \in \Gamma_d$, and on $[i/\kappa, (i+1)/\kappa)$ for $0 \le i \le \kappa - 1$, $\dot{\psi}_\kappa(t)$ equals the constant vector $\gamma_\kappa(i/\kappa)$. In particular, $\dot{\psi}_\kappa$ is a step function.

LEMMA 2.3. *For* $0 \le i \le d+1$ *and* $0 \le t \le 1$,

$$\psi_{\kappa,i}(t) \ge \theta(e_i t + c_i), \tag{2.8}$$

$$\sum_{i=0}^{d+1} i \dot{\psi}_{\kappa,i}(t) = \sum_{i=0}^{d} \bigl(1 - [\dot{\psi}_\kappa(t)]_i\bigr) \le 1 - \theta e_{d+1}. \tag{2.9}$$

PROOF. These are properties of φ_θ inherited from properties of $\varphi^*, \ell \in \Gamma_d$, which are preserved with respect to (2.7). Indeed, for each $0 \le i \le d+1$,

$$\begin{aligned}\psi_{\kappa,i}(t) &= \varphi_{\theta,i}(\lfloor t\kappa \rfloor/\kappa) + (t\kappa - \lfloor t\kappa \rfloor)\bigl(\varphi_{\theta,i}\bigl((\lfloor t\kappa \rfloor + 1)/\kappa\bigr) - \varphi_{\theta,i}(\lfloor t\kappa \rfloor/\kappa)\bigr) \\ &\ge \theta(e_i t + c_i).\end{aligned}$$

Last, (2.9) follows: noting that $\sum_{i=0}^{d}(1 - [\mathbf{e}]_i) = \sum_{i=0}^{d+1} i e_i = 1 - e_{d+1}$,

$$\begin{aligned}&\sum_{i=0}^{d}\bigl(1 - [\dot{\psi}_\kappa(t)]_i\bigr) \\ &\qquad = \kappa \int_{l/\kappa}^{(l+1)/\kappa} \Biggl[(1-\theta)\sum_{i=0}^{d}\bigl(1 - [\dot{\varphi}^*(s)]_i\bigr) + \theta \sum_{i=0}^{d}\bigl(1 - [\dot{\ell}(s)]_i\bigr)\Biggr] ds \\ &\qquad \le 1 - \theta + \theta \sum_{i=0}^{d}(1 - [\mathbf{e}]_i) = 1 - \theta e_{d+1}.\end{aligned}$$

□

LEMMA 2.4. *For large enough* κ, *we have*

$$h(\psi_\kappa) \le h(\varphi^*) + 2\varepsilon \quad \textit{and} \quad I_d(\psi_\kappa) \le I_d(\varphi^*) + 2\varepsilon. \tag{2.10}$$

PROOF. Since

$$\lim_{\kappa \to \infty} \sup_{t \in [0,1]} |\varphi_\theta(t) - \psi_\kappa(t)| = 0,$$

the inequality with respect to h follows from continuity of h and choosing κ in terms of θ. We also note, by absolute continuity of φ_θ, that a.s. in t,

$$\dot{\psi}_\kappa(t) = \gamma_\kappa(t) = \kappa \int_{\lfloor t\kappa \rfloor/\kappa}^{(\lfloor t\kappa \rfloor + 1)/\kappa} \dot{\varphi}_\theta(s)\, ds \to \dot{\varphi}_\theta(t) \qquad \text{as } \kappa \uparrow \infty.$$

Then, by the form of L [cf. (2.5)], bounds in Lemma 2.3 and piecewise continuity and bounds on p, β in assumption (ND), we have, as $\kappa \uparrow \infty$, that $L(\vec{\xi}(t), \psi_\kappa(t), \dot{\psi}_\kappa(t)) \to L(\vec{\xi}(t), \varphi_\theta(t), \dot{\varphi}_\theta(t))$ for almost all $t \in [0, 1]$.

Also, we can bound $L(\vec{\xi}(t), \psi_\kappa(t), \dot{\psi}_\kappa(t))$ as follows: first, using $x \log x \le 0$ for $0 \le x \le 1$, bound that

$$\begin{aligned}
&L(\vec{\xi}(t), \psi_\kappa(t), \dot{\psi}_\kappa(t)) \\
&\quad \le -(1 - [\dot{\psi}_\kappa(t)]_0) \log\biggl(p(t) + (1 - p(t))\frac{\beta(t)\psi_{\kappa,0}(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\biggr) \\
&\qquad - \sum_{i=1}^{d} (1 - [\dot{\psi}_\kappa(t)]_i) \log\biggl((1 - p(t))\frac{(i+\beta(t))\psi_{\kappa,i}(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\biggr) \\
&\qquad - \Biggl(1 - \sum_{i=0}^{d}(1 - [\dot{\psi}_\kappa(t)]_i)\Biggr) \\
&\qquad \quad \times \log\biggl((1 - p(t))\biggl(1 - \frac{\sum_{i=0}^{d}(i+\beta(t))\psi_{\kappa,i}(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\biggr)\biggr).
\end{aligned}$$

Now, as $0 \le [\dot{\psi}_\kappa]_i \le 1$ and $0 \le \sum_{i=0}^{d}(1 - [\dot{\psi}_\kappa]_i) \le 1$, we have the further upper-bound, using Lemma 2.3,

$$\begin{aligned}
&-\log\biggl(p(t) + (1 - p(t))\frac{\beta(t)\theta(e_0 t + c_0)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\biggr) \\
&\qquad - \sum_{i=1}^{d} \log\biggl((1 - p(t))\frac{(i+\beta(t))\theta(e_i t + c_i)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\biggr) \\
&\qquad - \log\biggl((1 - p(t))\frac{(d+1+\beta(t))\theta(e_{d+1} t + \bar{c}^d)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\biggr),
\end{aligned}$$

which is integrable on $[0, 1]$ given the bounds on p, β in assumption (ND).

By dominated convergence, we obtain $\lim_\kappa I_d(\psi_\kappa) = I_d(\varphi_\theta)$, and therefore the other inequality with respect to I_d. $\square$

Let now κ be such that (2.10) holds. Finally, we modify ψ_κ on the interval $[0, \delta]$, for a small enough $\delta > 0$ to be chosen later.

Define

$$t_i := \delta - \sum_{l=i}^{d} (\delta + [\mathbf{c}^d]_l - [\psi_\kappa(\delta)]_l) \tag{2.11}$$

for $0 \le i \le d$, and $t_{d+1} := \delta$; set also $t_{-1} := 0$ and $t_{d+2} = t_{d+1}$. Let also

$$\psi^*(t) = \int_0^t \gamma^*(s)\, ds + \mathbf{c}^d, \tag{2.12}$$

where

$$\gamma^*(t) = \begin{cases} \mathbf{f}_{d+1}, & \text{when } 0 \le t < t_0, \\ \mathbf{f}_i, & \text{when } t_i \le t < t_{i+1}, 0 \le i \le d, \\ \gamma_\kappa(t), & \text{when } t \ge \delta. \end{cases}$$

Note γ^* may not be defined at some endpoints as possibly $t_i = t_{i+1}$ for some i.

By inspection, $\psi^* \in \Gamma_d$. Also, $\dot{\psi}^*(t) = \mathbf{f}_{d+1}$ when $0 \le t < t_0$ and $\dot{\psi}^*(t) = \mathbf{f}_i$ when $t_i \le t < t_{i+1}$ for $0 \le i \le d$. Moreover, we have the following properties.

LEMMA 2.5. *We have* $\psi^*(\delta) = \psi_\kappa(\delta)$ *and* $t_0 \ge \theta e_{d+1}\delta$. *Also,*

$$\psi_0^*(t) = t + c_0, \qquad \psi_j^*(t) = c_j \qquad \textit{for } 1 \le j \le d+1,$$

when $0 \le t < t_0$, *and*

$$\psi_0^*(t) \ge \theta e_{d+1}\delta + c_0 \qquad \textit{when } t_0 < t < t_1,$$
$$\psi_i^*(t) \ge \theta(e_i\delta + c_i) \qquad \textit{when } t_i < t < t_{i+1} \textit{ and } 1 \le i \le d.$$

PROOF. The lower bound for t_0 follows from the integration of both sides in (2.9) and the definition of t_0. Now, we note that $\dot{\psi}_0^*(t) = 0$ if $t_0 \le t \le t_1$, and 1 otherwise. Also, note that for $1 \le i \le d+1$, $\dot{\psi}_i^*(t) = 1$ if $t_{i-1} < t < t_i$, $\dot{\psi}_i^*(t) = -1$ if $t_i < t < t_{i+1}$ and $\dot{\psi}_i^*(t) = 0$ otherwise. Thus, noting (2.11),

$$\psi_0^*(\delta) = \int_0^\delta \gamma_0^*(s)\,ds + c_0 = \delta - (t_1 - t_0) + c_0 = \psi_{\kappa,0}(\delta)$$

and, for $1 \le i \le d+1$,

$$\psi_i^*(\delta) = \int_0^\delta \gamma_i^*(s)\,ds + c_i = (t_i - t_{i-1}) - (t_{i+1} - t_i) + c_i = \psi_{\kappa,i}(\delta),$$

which proves that $\psi^*(\delta) = \psi_\kappa(\delta)$. Since $\psi_0^*(t)$ is nondecreasing, for $t \ge t_0$, $\psi_0^*(t) \ge \psi_0^*(t_0) = t_0 + c_0 \ge \theta e_{d+1}\delta + c_0$. For $1 \le i \le d$, for $t_i < t < t_{i+1}$, $\psi_i^*(t)$ decreases to its final value $\psi_{\kappa,i}(\delta) \ge \theta(e_i\delta + c_i)$ by (2.8). □

2.2.2. *Step* 2: *More properties of* ψ^*. We now show the rate of ψ^* up to time δ does not contribute too much.

LEMMA 2.6. *For small enough* $\delta > 0$,

$$\int_0^\delta L(\vec{\xi}(t), \psi^*(t), \dot{\psi}^*(t))\,dt \le \varepsilon \quad \textit{and} \quad \|\psi^* - \psi_\kappa\|_\infty < \varepsilon.$$

In particular, $h(\psi^*) \le h(\varphi^*) + 3\varepsilon$ *and* $I_d(\psi^*) \le I_d(\varphi^*) + 3\varepsilon$.

PROOF. Write, for $0 \le t \le \delta$, as $L(\vec{\xi}(t), \psi^*(t), \dot{\psi}^*(t)) = R(\delta_{\mathbf{f}_{d+1}} \| \rho_{\vec{\xi}(t),\psi^*(t)}) \times 1(0 < t < t_0) + \sum_{i=0}^{d} R(\delta_{\mathbf{f}_i} \| \rho_{\vec{\xi}(t),\psi^*(t)}) 1(t_i < t < t_{i+1})$,

$$\begin{aligned} &L(\vec{\xi}(t), \psi^*(t), \dot{\psi}^*(t)) \\ &\quad = -\log\biggl(\bigl(1 - p(t)\bigr)\biggl(1 - \frac{\sum_{l=0}^{d}(l + \beta(t))\psi_l^*(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\biggr)\biggr) 1(0 < t < t_0) \\ &\qquad - \log\biggl(p(t) + \bigl(1 - p(t)\bigr)\frac{\beta(t)\psi_0^*(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\biggr) 1(t_0 < t < t_1) \\ &\qquad - \sum_{i=1}^{d}\log\biggl(\bigl(1 - p(t)\bigr)\frac{(i + \beta(t))\psi_i^*(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\biggr) 1(t_i < t < t_{i+1}). \end{aligned}$$

By Lemma 2.5 and the bounds on p, β in assumption (ND), this expression is integrable for $0 \le t \le \delta$. (It would be bounded unless $\bar{c}^d = 0$ and $c \ne 0$, in which case the first term in the expression involves $-\log t$.) Hence, the first statement follows for small $\delta > 0$. Also, the second statement holds as $\|\psi^* - \psi_\kappa\|_\infty = \sup_{0 \le t < \delta} |\psi^* - \psi_\kappa| \le 2\delta(d+2)$. The last statement is a consequence now of (2.10). $\square$

We will take $\delta > 0$ small enough so that the bounds in the above lemma hold.

LEMMA 2.7. *We have*

$$\lim_{n\to\infty} \sup_{0 \le j \le n} \Biggl| \psi^*(j/n) - \frac{1}{n}\sum_{l=0}^{j-1} \dot{\psi}^*(l/n) - \mathbf{c}^d \Biggr| = 0. \tag{2.13}$$

Also, for $j \ge \lfloor \delta n \rfloor$ *and* $0 \le i \le d+1$,

$$\frac{1}{n}\sum_{l=0}^{j-1} \dot{\psi}_i^*(l/n) + c_i \ge \frac{\theta}{2}\biggl(\frac{e_i j}{n} + c_i\biggr). \tag{2.14}$$

PROOF. Since $\dot{\psi}^*$ is piecewise constant, when $l/n \le s \le (l+1)/n$, $|\dot{\psi}^*(s) - \dot{\psi}^*(l/n)| \ne 0$ for at most κ subintervals [cf. (2.7) and (2.12)], and is also bounded by $2(d+2)$. Hence,

$$\begin{aligned} \Biggl| \psi^*(j/n) - \frac{1}{n}\sum_{l=0}^{j-1} \dot{\psi}^*(l/n) - \mathbf{c}^d \Biggr| &= \Biggl| \sum_{l=0}^{j-1} \int_{l/n}^{(l+1)/n} \bigl(\dot{\psi}^*(s) - \dot{\psi}^*(l/n)\bigr)\, ds \Biggr| \\ &\le \frac{2(d+2)}{n}\kappa. \end{aligned}$$

The last statement follows from (2.8). $\square$

2.2.3. *Step* 3: *Admissible control measures and convergence.* We now build a sequence of controls based on ψ^*. Define $\nu_0 = \nu_0(\dot{\psi}^*(j/n)|j/n)$ using (2.4), and

$$v_j^n(d\mathbf{y}; \mathbf{x}_0, \dots, \mathbf{x}_j) = \begin{cases} \nu_0\big(\dot{\psi}^*(j/n)|j/n\big), & \text{when } 0 \le j \le \lfloor \delta n \rfloor \\ & \quad \text{or when } j \ge \lceil \delta n \rceil \\ & \quad \text{and } \mathbf{x}_{j,i} \ge \dfrac{\theta}{4}(e_i\delta + c_i) \\ & \quad \text{for } 0 \le i \le d+1, \\ \rho_{\vec{\xi}(j/n),\mathbf{x}_j}, & \text{otherwise.} \end{cases}$$

The reasoning behind this choice of controls is as follows: to bound the limit of the quantity in (2.2), using formula (2.5), by $I_d(\psi^*) + h(\psi^*)$, we would like to specify the controls in form $\nu_0(\dot{\psi}^*(j/n)|j/n)$. Such a choice, as we will see, also ensures that the adapted sequence $\bar{\mathbf{X}}^n(j)$ is close to $\psi^*(j/n)$. However, the adapted process, as it is random, may get too close to a boundary. When this happens, not often it turns out, to bound errors, we specify that the controls take the cost-free form of the natural evolution sequence. Also, to get past this boundary layer initially, ψ^* has been built as a step function so that the adapted process must follow a deterministic trajectory up to time $\lfloor \delta n \rfloor$.

Define $\bar{\mathbf{X}}^n(0) = \mathbf{c}^d$, and $\bar{\mathbf{X}}^n(j+1) = \bar{\mathbf{X}}^n(j) + \frac{1}{n}\bar{\mathbf{Y}}^n(j)$ for $j \ge 0$ where

$$\bar{P}\big(\bar{\mathbf{Y}}^n(j) \in d\mathbf{y}|\bar{\mathbf{X}}^n(0), \dots, \bar{\mathbf{X}}^n(j)\big) = v_j^n(d\mathbf{y}; \bar{\mathbf{X}}^n(0), \dots, \bar{\mathbf{X}}^n(j)).$$

Thus, for $j \ge 0$, $\bar{\mathbf{X}}^n(j) = \frac{1}{n}\sum_{l=0}^{j-1} \bar{\mathbf{Y}}^n(l) + \mathbf{c}^d$. It will be useful later to note the total weight $\sum_{i=0}^{d+1}(i + \beta(j/n))\bar{\mathbf{X}}_i^n(j) \le (j/n + \tilde{c}) + \beta(j/n)(j/n + c)$ and, for $0 \le j \le \lfloor \delta n \rfloor$, as mentioned $\bar{\mathbf{X}}^n(j)$ is deterministic and $\bar{\mathbf{X}}^n(j) = \frac{1}{n}\sum_{l=0}^{j-1} \dot{\psi}^*(l/n) + \mathbf{c}^d$.

Define now, for each $n \ge 1$, the martingale sequence for $0 \le j \le n$

$$\begin{aligned} \mathbf{M}^n(j) &:= \frac{1}{n}\sum_{l=0}^{j-1}\big(\bar{\mathbf{Y}}^n(l) - \bar{E}(\bar{\mathbf{Y}}^n(l)|\bar{\mathbf{X}}^n(l))\big) \\ &= \bar{\mathbf{X}}^n(j) - \frac{1}{n}\sum_{l=0}^{j-1}\bar{E}(\bar{\mathbf{Y}}^n(l)|\bar{\mathbf{X}}^n(l)) - \mathbf{c}^d. \end{aligned}$$

Let

$$\tau_n := n \wedge \min\Big\{\lceil \delta n \rceil \le l \le n : \bar{\mathbf{X}}_i^n(l) < \frac{\theta}{4}(e_i\delta + c_i) \text{ for some } 0 \le i \le d+1\Big\}.$$

Then, $\tau_n \geq \lceil \delta n \rceil$ is a stopping time, and the corresponding stopped process $\{\mathbf{M}^n(j \wedge \tau_n)\}$ is also a martingale for $0 \leq j \leq n$. Let now

$$\mathbb{A}_n := \left\{ \sup_{0 \leq j \leq n} |\mathbf{M}^n(j \wedge \tau_n)| > \frac{\theta e_{d+1}}{4n^{1/8}} \right\}.$$

LEMMA 2.8. *For $n \geq \delta^{-8}$, on the set $\mathbb{A}_n^c$, we have $\tau_n = n$.*

PROOF. From the definition of $\{v_j^n\}$ and τ_n, we have $\bar{E}(\bar{\mathbf{Y}}^n(l)|\bar{\mathbf{X}}^n(l)) = \dot{\psi}^*(l/n)$ for $0 \leq l \leq j \wedge \tau_n - 1$ and $j \geq \lceil \delta n \rceil$. Then, on $\mathbb{A}_n^c$, by (2.14), we have

$$\begin{aligned}
\bar{\mathbf{X}}_i^n(j \wedge \tau_n) &\geq c_i + \frac{1}{n} \sum_{l=0}^{j \wedge \tau_n - 1} \bar{E}(\bar{\mathbf{Y}}_i^n(l)|\bar{\mathbf{X}}^n(l)) - \frac{\theta e_{d+1}}{4n^{1/8}} \\
&= c_i + \frac{1}{n} \sum_{l=0}^{j \wedge \tau_n - 1} \dot{\psi}_i^*(l/n) - \frac{\theta e_{d+1}}{4n^{1/8}} \\
&\geq \frac{\theta}{2}\left(\frac{e_i(j \wedge \tau_n)}{n} + c_i\right) - \frac{\theta e_{d+1}}{4n^{1/8}} \\
&\geq \frac{\theta}{4}(e_i \delta + c_i).
\end{aligned}$$

Hence, $\tau_n = n$. $\square$

We now observe, by Doob's martingale inequality and bounds, in terms of constants $C = C_d$, that

$$\begin{aligned}
\bar{P}[\mathbb{A}_n] &\leq C n^{1/2} \bar{E}|\mathbf{M}^n(j \wedge \tau_n)|^4 \\
&= C n^{-7/2} \bar{E}\left| \sum_{l=0}^{j \wedge \tau_n - 1} (\bar{\mathbf{Y}}^n(l) - \bar{E}(\bar{\mathbf{Y}}^n(l)|\bar{\mathbf{X}}^n(l))) \right|^4 \qquad (2.15) \\
&\leq C n^{-7/2} n^2 = C n^{-3/2}.
\end{aligned}$$

We now state the following almost sure convergence.

LEMMA 2.9. *We have*

$$\lim_{n \uparrow \infty} \sup_{0 \leq j \leq n} \left| \bar{\mathbf{X}}^n(j) - \frac{1}{n} \sum_{l=0}^{j-1} \dot{\psi}^*(l/n) - \mathbf{c}^d \right| = 0 \qquad a.s. \tag{2.16}$$

PROOF. First, by (2.15) and the Borel–Cantelli lemma, $\bar{P}(\limsup \mathbb{A}_n) = 0$. On the other hand, on the full measure set $\bigcup_{m \geq 1} \bigcap_{k \geq m} \mathbb{A}_k^c$, since $\tau_n = n$ and $\bar{E}(\bar{\mathbf{Y}}^n(l)|\bar{\mathbf{X}}^n(l)) = \dot{\psi}^*(l/n)$ for $0 \leq l \leq n-1$ on $\mathbb{A}_n^c$ by Lemma 2.8, the desired convergence holds. $\square$

2.2.4. *Step* 4. We now argue the lower bound through representation (2.2). Recall the definition of $\vec{\xi}(\cdot)$ in the beginning of Section 2. The sum in (2.2) equals

$$\begin{aligned}
&\bar{E}\left[\frac{1}{n}\sum_{j=0}^{n-1} R\big(v_j^n \| \rho_{\vec{\xi}(j/n),\bar{\mathbf{X}}^n(j)}\big)\right] \\
&\quad = \bar{E}\left[\frac{1}{n}\sum_{j=0}^{\lfloor \delta n\rfloor} R\big(v_j^n \| \rho_{\vec{\xi}(j/n),\bar{\mathbf{X}}^n(j)}\big)\right] \\
&\qquad + \bar{E}\left[\frac{1}{n}\sum_{j=\lceil \delta n\rceil}^{n-1} R\big(v_j^n \| \rho_{\vec{\xi}(j/n),\bar{\mathbf{X}}^n(j)}\big); \mathbb{A}_n\right] \\
&\qquad + \bar{E}\left[\frac{1}{n}\sum_{j=\lceil \delta n\rceil}^{n-1} R\big(v_j^n \| \rho_{\vec{\xi}(j/n),\bar{\mathbf{X}}^n(j)}\big); \mathbb{A}_n^c\right] \\
&\quad = A_1 + A_2 + A_3.
\end{aligned} \tag{2.17}$$

Step 4.1. We treat the term A_2 in (2.17). Recall $\sigma(j/n) = (1+\beta(j/n))(j/n) + \tilde{c} + c\beta(j/n)$ and the "weight" bound on $\bar{\mathbf{X}}^n(j)$ in beginning of Step 3. For $\lceil \delta n\rceil \le j \le n-1$,

$$\begin{aligned}
&R\big(v_j^n \| \rho_{\vec{\xi}(j/n),\bar{\mathbf{X}}^n(j)}\big) \\
&\quad = R\big(\nu_0(\dot{\psi}^*(j/n)) \| \rho_{\vec{\xi}(j/n),\bar{\mathbf{X}}^n(j)}\big) \\
&\qquad \times 1\big(\bar{\mathbf{X}}_i^n(j) \ge (\theta/4)(e_i\delta + c_i) \text{ for } 0 \le i \le d+1\big).
\end{aligned}$$

Noting (2.5), this is bounded above, using $x \log x \le 0$ for $0 \le x \le 1$, by

$$\begin{aligned}
&\Bigg[-\bigg(1 - \Big[\dot{\psi}^*\Big(\frac{j}{n}\Big)\Big]_0\bigg)\log\bigg(p(j/n) + \big(1-p(j/n)\big)\frac{\beta(j/n)\bar{\mathbf{X}}_0^n(j)}{\sigma(j/n)}\bigg) \\
&\quad - \sum_{i=1}^{d}\bigg(1 - \Big[\dot{\psi}^*\Big(\frac{j}{n}\Big)\Big]_i\bigg)\log\bigg(\big(1-p(j/n)\big)\frac{(i+\beta(j/n))\bar{\mathbf{X}}_i^n(j)}{\sigma(j/n)}\bigg) \\
&\quad - \bigg(\sum_{i=0}^{d}\Big[\dot{\psi}^*\Big(\frac{j}{n}\Big)\Big]_i - d\bigg) \\
&\qquad\qquad \times \log\bigg(\big(1-p(j/n)\big)\bigg(1 - \frac{\sum_{i=0}^d (i+\beta(j/n))\bar{\mathbf{X}}_i^n(j)}{\sigma(j/n)}\bigg)\bigg)\Bigg] \\
&\quad \times 1\big(\bar{\mathbf{X}}_i^n(j) \ge (\theta/4)(e_i\delta + c_i) \text{ for } 0 \le i \le d+1\big).
\end{aligned}$$

Given bounds on p, β in (ND), as $0 \le [\dot{\psi}^*]_i \le 1$, we have $d \le \sum_{i=1}^d [\dot{\psi}^*]_i \le d+1$ and

$$\begin{aligned}\sum_{i=0}^{d} \big(i+\beta(j/n)\big)\bar{\mathbf{X}}_i^n(j) &\le \sigma(j/n) - \big(d+1+\beta(j/n)\big)\bar{\mathbf{X}}_{d+1}^n(j) \\ &\le \sigma(j/n) - \big(d+1+\beta(j/n)\big)\cdot(\theta/4)(e_{d+1}\delta + c_{d+1}),\end{aligned}$$

the relative entropy is further bounded by a constant C_d. Thus, for large n,

$$A_2 \le C_d \cdot \bar{P}\bigg[\sup_{0\le j\le n} |\mathbf{M}^n(j\wedge\tau_n)| > \frac{\theta e_{d+1}}{4n^{1/8}}\bigg] \le \varepsilon. \tag{2.18}$$

Step 4.2. Now, for the term A_1 in (2.17), we recall for $j \le \lfloor \delta n\rfloor$ that $\bar{\mathbf{X}}^n(j) = \frac{1}{n}\sum_{l=0}^{j-1}\dot{\psi}^*(l/n) + \mathbf{c}^d$ is deterministic. Also note, for $0 \le i \le d$, that $\dot{\psi}^*(t) = \mathbf{f}_i$ on $t_i < t < t_{i+1}$, and $\dot{\psi}^*(t) = \mathbf{f}_{d+1}$ on $0 = t_{-1} \le t \le t_0$ (cf. near Lemma 2.5). Thus, for $0 \le j \le \lfloor \delta n\rfloor$, denoting $\mathbf{f}_{-1} = \mathbf{f}_{d+1}$, we may write

$$\begin{aligned}&R\big(v_j^n \| \rho_{\vec{\xi}(j/n),\bar{\mathbf{X}}^n(j)}\big) \\ &\quad = L\bigg(\vec{\xi}\Big(\frac{j}{n}\Big), \frac{1}{n}\sum_{l=0}^{j-1}\dot{\psi}^*\Big(\frac{l}{n}\Big) + \mathbf{c}^d, \dot{\psi}^*\Big(\frac{j}{n}\Big)\bigg) \\ &\quad = \sum_{i=-1}^{d} L\bigg(\vec{\xi}\Big(\frac{j}{n}\Big), \sum_{l=-1}^{i-1}\frac{\lfloor t_{l+1}n\rfloor - \lfloor t_l n\rfloor}{n}\mathbf{f}_l + \frac{j-\lfloor t_i n\rfloor}{n}\mathbf{f}_i + \mathbf{c}^d, \mathbf{f}_i\bigg) \\ &\qquad\quad \times 1(\lfloor t_i n\rfloor \le j < \lfloor t_{i+1}n\rfloor),\end{aligned}$$

where, for $i = -1$, the empty sum in the argument for L vanishes. Comparing with the proof of Lemma 2.6, this expression, given bounds on p, β in (ND), is bounded, for $0 \le j \le \lfloor \delta n\rfloor$, except when $\bar{c}^d = 0$ and $c \ne 0$, in which case a "$-\log(j/n)$" term appears in the $i = -1$ term. But, since $-(1/n)\sum_{j=1}^{\lfloor \delta n\rfloor}\log(j/n) \le -\int_0^\delta \log(t)\,dt$, its contribution is still small. Hence,

$$A_1 \le \epsilon(\delta) \qquad \text{where } \epsilon(\delta) \to 0 \text{ as } \delta \to 0. \tag{2.19}$$

Step 4.3. We now estimate the last term A_3 in (2.17). For $n \ge \delta^{-8}$, by Lemma 2.8 and definition of L (2.5),

$$\begin{aligned}A_3 &\le \bar{E}\bigg[\frac{1}{n}\sum_{j=\lceil \delta n\rceil}^{n-1} L\bigg(\vec{\xi}\Big(\frac{j}{n}\Big), \bar{\mathbf{X}}^n(j), \dot{\psi}^*\Big(\frac{j}{n}\Big)\bigg); \mathbb{A}_n^c \cap \{\tau_n = n\}\bigg] \\ &\le \bar{E}\bigg[\int_\delta^1 L\bigg(\vec{\xi}\Big(\frac{\lfloor nt\rfloor}{n}\Big), \bar{\mathbf{X}}^n(\lfloor nt\rfloor), \dot{\psi}^*\Big(\frac{\lfloor nt\rfloor}{n}\Big)\bigg)\,dt; \mathbb{A}_n^c \cap \mathbb{B}_n\bigg],\end{aligned}$$

where $\mathbb{B}_n = \{\bar{\mathbf{X}}_i^n(j) \ge (\theta/4)(e_i\delta + c_i) \text{ for } 0 \le i \le d+1, j \ge \lceil \delta n\rceil\}$. On the event $\mathbb{A}_n^c \cap \mathbb{B}_n$, $\bar{E}(\bar{\mathbf{Y}}^n(l)|\bar{\mathbf{X}}^n(l)) = \dot{\psi}^*(l/n)$ for $l \ge 0$, and so $|\bar{\mathbf{X}}^n(\lfloor nt\rfloor) - \psi^*(\lfloor nt\rfloor/n)|1(\mathbb{A}_n^c \cap \mathbb{B}_n) \to 0$ for each realization from (2.13) and (2.16).

Also, from the form of L (2.5), as $\dot{\psi}^*$ is a step function, (2.8), and bounds and piecewise continuity of p, β in (ND), we may bound as in Step 4.1 and observe

$$2C_d \ge \left| L\left(\vec{\xi}\left(\frac{\lfloor nt\rfloor}{n}\right), \bar{\mathbf{X}}^n(\lfloor nt\rfloor), \dot{\psi}^*\left(\frac{\lfloor nt\rfloor}{n}\right)\right) - L(\vec{\xi}(t), \psi^*(t), \dot{\psi}^*(t))\right| 1(\mathbb{A}_n^c \cap \mathbb{B}_n) \to 0$$

for almost all t and each realization. Hence, by bounded convergence theorem, with respect to $d\bar{P} \times 1([\delta, t])\,dt$,

$$\limsup_{n\to\infty} A_3 \le \int_\delta^1 L(\vec{\xi}(t), \psi^*(t), \dot{\psi}^*(t))\,dt. \tag{2.20}$$

2.2.5. *Step* 5. Finally, by (2.13) and (2.16), $\lim_{n\to\infty} h(\bar{\mathbf{X}}^n(\cdot)) = h(\psi^*(\cdot))$ a.s. in the sup topology, and by bounded convergence $\lim_{n\to\infty} \bar{E}[h(\bar{\mathbf{X}}^n(\cdot))] = h(\psi^*(\cdot))$.

We now combine all bounds to conclude the proof of (2.6). By (2.2), bounds (2.18), (2.19), (2.20) and nonnegativity of L, we have

$$\begin{aligned}\limsup_{n\to\infty} V^n &\le \limsup_{n\to\infty} \bar{E}\left[\frac{1}{n}\sum_{j=0}^{n-1} R\big(v_j^n \| \rho_{\vec{\xi}(j/n), \bar{\mathbf{X}}^n(j)}\big) + h(\bar{\mathbf{X}}^n(\cdot))\right] \\ &\le 2\varepsilon + \int_0^1 L(\vec{\xi}(t), \psi^*(t), \dot{\psi}^*(t))\,dt + h(\psi^*).\end{aligned}$$

Then, by Lemma 2.6, we obtain (2.6).

3. Proof of Theorem 1.4. The proof of Theorem 1.4 follows from the following two propositions, and is given below. We first recall the projective limit approach, following notation in [20], Section 4.6. Define, for $0 \le i \le j$, $\mathcal{Y}_j = C([0,1]; \mathbb{R}^{j+2})$ and $p_{ij} : \mathcal{Y}_j \to \mathcal{Y}_i$ by $\langle \varphi_0, \dots, \varphi_{j+1}\rangle \mapsto \langle \varphi_0, \dots, \varphi_i, \sum_{l=i+1}^{j+1} \varphi_l\rangle$. Also define $\varprojlim \mathcal{Y}_j \subset \prod_{i\ge 0} \mathcal{Y}_i$ as the subset of elements $x = \langle x^0, x^1, \dots\rangle$ such that $p_{ij} x^j = x^i$, equipped with the product topology. Let also $p_j : \varprojlim \mathcal{Y}_j \to \mathcal{Y}_j$ be the canonical projection, $p_j x = x^j$.

Since I_d are convex, good rate functions on $C([0,1], \mathbb{R}^{d+2})$, by the LDPs Theorem 1.2 and [20], Theorem 4.6.1, we obtain the following proposition. Recall the notation in Theorem 1.2. For $n \ge 1$, let $\mathcal{X}^{n,\infty} = \langle \mathbf{X}^{n,0}, \mathbf{X}^{n,1}, \dots\rangle$.

PROPOSITION 3.1. *The sequence $\{\mathcal{X}^{n,\infty}\} \subset \varprojlim \mathcal{Y}_j$ satisfies an LDP with rate n and convex, good rate function*

$$J^\infty(\varphi) = \begin{cases} \sup_d \{I_d(p_d(\varphi))\}, & \text{when } \varphi \in \varprojlim \mathcal{Y}_j, \\ \infty, & \text{otherwise.}\end{cases}$$

To establish Theorem 1.4, it remains to further identify J^∞. Recall $\Gamma_d \subset C([0,1];\mathbb{R}^{d+2})$ are those elements $\varphi = \langle \varphi_0, \dots, \varphi_d, \varphi_{d+1}\rangle$ such that:

> $\varphi(0) = \mathbf{c}^d$, each $\varphi_i \geq 0$ is Lipschitz with constant 1 such that $0 \leq [\dot\varphi(t)]_i \leq 1$ for $0 \leq i \leq d$, $\sum_{i=0}^{d+1} \dot\varphi_i(t) = 1$, and $\sum_{i=0}^{d+1} i\dot\varphi_i(t) = \sum_{i=0}^{d}(1-[\dot\varphi(t)]_i) \leq 1$ for almost all t.

Let also $\Gamma^* \subset \varprojlim \mathcal{Y}_j$ be those elements $\varphi = \langle \varphi^0, \varphi^1, \dots\rangle$ such that $\varphi^d \in \Gamma_d$ for $d \geq 0$. Since $\{\Gamma_d\}_{d\geq 0}$ are compact sets, it is a straightforward exercise to see that Γ^* is compact. Define $L_d(p_d(\varphi(t)))$ equal to

$$
\begin{aligned}
&\big(1-[\dot\varphi^d(t)]_0\big)\log\frac{1-[\dot\varphi^d(t)]_0}{p(t)+(1-p(t))\frac{\beta(t)\varphi_0^d(t)}{(1+\beta(t))t+\tilde c+c\beta(t)}}\\
&\quad+\sum_{i=1}^{d}\big(1-[\dot\varphi^d(t)]_i\big)\log\frac{1-[\dot\varphi^d(t)]_i}{(1-p(t))\frac{(i+\beta(t))\varphi_i^d(t)}{(1+\beta(t))t+\tilde c+c\beta(t)}}\\
&\quad+\left(1-\sum_{i=0}^{d}(1-[\dot\varphi^d(t)]_i)\right)\log\frac{1-\sum_{i=0}^{d}(1-[\dot\varphi^d(t)]_i)}{(1-p(t))(1-\frac{\sum_{i=0}^{d}(i+\beta(t))\varphi_i^d(t)}{(1+\beta(t))t+\tilde c+c\beta(t)})}\,dt.
\end{aligned}
$$

PROPOSITION 3.2. *The rate function $J^\infty(\varphi)$ diverges when $\varphi \notin \Gamma^*$. However, for $\varphi \in \Gamma^*$, $\lim_{d\uparrow\infty} L_d(p_d(\varphi(t)))$ exists for almost all t, and we can evaluate*

$$
J^\infty(\varphi) = \int_0^1 \lim_{d\uparrow\infty} L_d(p_d(\varphi(t)))\,dt.
$$

PROOF. First, from the definition, $J^\infty(\varphi)$ diverges unless $\varphi \in \Gamma^*$. Next, for $\varphi \in \Gamma^*$ and almost all t, we argue

$$
L_r(p_r(\varphi(t))) \leq L_s(p_s(\varphi(t))) \qquad \text{when } r < s. \tag{3.1}
$$

It will be enough to show from the form of the rates the following:

$$
\begin{aligned}
&\left(1-\sum_{i=0}^{r}(1-[\dot\varphi^s(t)]_i)\right)\log\frac{1-\sum_{i=0}^{r}(1-[\dot\varphi^s(t)]_i)}{(1-p(t))(1-\frac{\sum_{i=0}^{r}(i+\beta(t))\varphi_i^s(t)}{(1+\beta(t))t+\tilde c+c\beta(t)})}\\
&\quad\leq \sum_{i=r+1}^{s}\big(1-[\dot\varphi^s(t)]_i\big)\log\frac{1-[\dot\varphi^s(t)]_i}{(1-p(t))\frac{(i+\beta(t))\varphi_i^s(t)}{(1+\beta(t))t+\tilde c+c\beta(t)}}\\
&\qquad+\left(1-\sum_{i=0}^{s}(1-[\dot\varphi^s(t)]_i)\right)\log\frac{1-\sum_{i=0}^{s}(1-[\dot\varphi^s(t)]_i)}{(1-p(t))(1-\frac{\sum_{i=0}^{s}(i+\beta(t))\varphi_i^s(t)}{(1+\beta(t))t+\tilde c+c\beta(t)})}.
\end{aligned}
$$

Consider now $h(x) = x \log x$ which is convex for $x \geq 0$. Under conventions (1.3), for nonnegative numbers, a_i and b_i, we have

$$\frac{\sum_{i=p}^q a_i}{\sum_{i=p}^q b_i} \log \frac{\sum_{i=p}^q a_i}{\sum_{i=p}^q b_i} = h\left(\frac{\sum_{i=p}^q a_i}{\sum_{i=p}^q b_i}\right) = h\left(\sum_{i=p}^q \frac{b_i}{\sum_{i=p}^q b_i} \frac{a_i}{b_i}\right)$$
$$\leq \sum_{i=p}^q \frac{b_i}{\sum_{i=p}^q b_i} h\left(\frac{a_i}{b_i}\right) = \frac{\sum_{i=p}^q a_i \log(a_i/b_i)}{\sum_{i=p}^q b_i}.$$

We now finish the proof of (3.1) by applying the last sequence, with $p = r+1$ and $q = s+1$, to

$$a_j = \begin{cases} 1 - [\dot{\varphi}^s(t)]_j, & \text{for } r+1 \leq j \leq s, \\ 1 - \sum_{i=0}^s (1 - [\dot{\varphi}^s(t)]_i), & \text{for } j = s+1, \end{cases}$$

and

$$b_j = \begin{cases} (1 - p(t)) \dfrac{(j + \beta(t))\varphi_j^s(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}, & \text{for } r+1 \leq j \leq s, \\ (1 - p(t))\left(1 - \dfrac{\sum_{i=0}^s (i + \beta(t))\varphi_i^s(t)}{(1+\beta(t))t + \tilde{c} + c\beta(t)}\right), & \text{for } j = s+1. \end{cases}$$

Finally, given $L_d(p_d(\varphi(t))) \geq 0$ is increasing in d, the identification of J^∞ in the display of the proposition follows from monotone convergence. $\square$

PROOF OF THEOREM 1.4. Let $\Gamma^\infty \subset \prod_{i \geq 0} C([0,1]; \mathbb{R})$, endowed with the product topology, be those elements $\xi = \langle \xi_0, \xi_1, \ldots \rangle$ such that:

$\xi_i(0) = c_i$, $\xi_i(t) \geq 0$ is Lipschitz with constant 1, $0 \leq [\dot{\xi}(t)]_i \leq 1$ for $i \geq 0$, and $\frac{d}{dt} \sum_{i \geq 0} \xi_i(t) = 1$ and $\lim_d [\sum_{i=0}^d i\dot{\xi}_i(t) + (d+1)(1 - [\dot{\xi}(t)]_d)] = \sum_{i \geq 0}(1 - [\dot{\xi}(t)]_i) \leq 1$ for almost all t.

We now show that Γ^∞ and Γ^* are homeomorphic. Hence, as Γ^* is compact, Γ^∞ would also be compact. (We note, one can see directly that Γ^∞ is compact.)

Define the map $F : \Gamma^\infty \to \Gamma^*$ by

$$F(\xi) = \langle \xi^0, \ldots, \xi^d, \ldots \rangle \qquad \text{where } \xi^d = \langle \xi_0, \ldots, \xi_d, t + c - [\xi]_d \rangle \in \Gamma_d.$$

In verifying the last inclusion, note $\sum_{i=0}^d i\dot{\xi}_i(t) + (d+1)(1 - [\dot{\xi}(t)]_d) = \sum_{i=0}^d (1 - [\dot{\xi}(t)]_i) \leq \sum_{i \geq 0}(1 - [\dot{\xi}(t)]_i) \leq 1$. We now argue that F is a bi-continuous bijection.

Indeed, we first note that $F^{-1} : \Gamma^* \to \Gamma^\infty$ is given by

$$F^{-1}(\varphi) = \langle \varphi_0^0, \dots, \varphi_d^d, \dots \rangle.$$

In checking $F^{-1}(\varphi) \in \Gamma^\infty$, note for $\varphi \in \Gamma^*$ that $\lim_d \sum_{i=0}^d (1 - \sum_{l=0}^i \dot{\varphi}_l^l(t)) = \lim_d \sum_{i=0}^d (1 - [\dot{\varphi}^d(t)]_i) \le 1$. Then, by bounded convergence with respect to the last term in the previous series, $\lim_d (t + \sum_{i=0}^d c_i - \sum_{i=0}^d \varphi_i^i(t)) = \lim_d (t + \sum_{i=0}^d c_i - [\varphi^d(t)]_d) = 0$, and so $\sum_{i \ge 0} \varphi_i^i(t) = t + c$. Finally, it is not difficult to see that F and F^{-1} are both continuous in the product topology.

Now, $\mathcal{X}^{n,\infty} \in \Gamma^*$, $\mathbf{X}^{n,\infty} \in \Gamma^\infty$, and $F(\mathbf{X}^{n,\infty}) = \mathcal{X}^{n,\infty}$ for $n \ge 1$. Hence, through the action of F, the LDP for $\mathcal{X}^{n,\infty}$ translates to the LDP for $\mathbf{X}^{n,\infty}$. We now identify the rate function. Given Propositions 3.1 and 3.2, for a degree distribution $\xi \in \Gamma^\infty$, we identify its rate as $I^\infty(\xi) = J^\infty(F(\xi))$. Since Γ^∞ is closed, and therefore distributions $\xi \notin \Gamma^\infty$ can never be attained by $\mathbf{X}^{n,\infty}$, we set $I^\infty(\xi) = \infty$ in this case. Last, by properties of F, as J^∞ is a convex, good rate function, one obtains readily I^∞ is also a convex, good rate function. □

4. Proof of Corollary 1.7.

We verify some properties of ζ^d in the next lemmas and conclude the proof of Corollary 1.7 at the end of the section.

LEMMA 4.1. *The ODE* (1.5) *has a unique Carathéodory solution* ζ^d.

PROOF. Any Carathéodory solution to ODE (1.5), given the assumption p, β are piecewise continuous, is piecewise continuously differentiable. Since the defining ODEs are linear, one can solve them, and so the solution is unique and given by $\zeta^d = \langle \zeta_0(t), \zeta_1(t), \dots, \bar{\zeta}_{d+1}(t) \rangle$ where, for $t \in [0, 1]$,

$$\begin{aligned}
\zeta_0(t) &:= c_0 M_0(0,t) + \int_0^t (1 - p(s)) M_0(s,t)\, ds, \\
\zeta_1(t) &:= c_1 M_1(0,t) \\
&\quad + \int_0^t \left(p(s) + (1 - p(s)) \frac{\beta(s)\zeta_0(s)}{(1+\beta(s))s + \tilde{c} + c\beta(s)} \right) M_1(s,t)\, ds, \\
\zeta_i(t) &:= c_i M_i(0,t) \\
&\quad + \int_0^t (1 - p(s)) \frac{(i - 1 + \beta(s))\zeta_{i-1}(s)}{(1+\beta(s))s + \tilde{c} + c\beta(s)} M_i(s,t)\, ds
\end{aligned} \tag{4.1}$$

for $2 \le i \le d$ and

$$\bar{\zeta}_{d+1}(t) := t + c - \sum_{i=0}^d \zeta_i(t) = \bar{c}^d + \int_0^t (1 - p(s)) \frac{(d + \beta(s))\zeta_d(s)}{(1+\beta(s))s + \tilde{c} + c\beta(s)}\, ds.$$

Here, for $0 \le i \le d$,

$$M_i(s,t) := \exp\biggl[-\int_s^t (1-p(u))\frac{i+\beta(u)}{(1+\beta(u))u+\tilde{c}+c\beta(u)}\,du\biggr]. \qquad \square$$

LEMMA 4.2. *We have $\zeta^d \in \Gamma_d$, and moreover*

$$\sum_{i=0}^{\infty} \zeta_i(t) = t + c \quad \textit{and} \quad \sum_{i=0}^{\infty} i\zeta_i(t) = t + \tilde{c}.$$

PROOF. First, from properties of the ODE system and the piecewise continuity assumption on p, β in (ND), $\zeta_i \ge 0$, ζ_i is Lipschitz with constant 1 and moreover piecewise continuously differentiable, and $0 \le [\dot{\zeta}(t)]_i \le 1$ for $i \ge 0$, and $\sum_{i=0}^d \zeta_i(t) + \bar{\zeta}_{d+1}(t) = t + c$ for $d \ge 0$ and almost all t. We postpone proving $\sum_{i=0}^d (1-[\dot{\zeta}(t)]_i) \le 1$ for $d \ge 0$ and a.a. t, which would complete the argument to show $\zeta^d \in \Gamma_d$, until the end.

We now show $\sum_{i\ge 0} \zeta_i(t) = t + c$. From the defining ODEs (1.5), for $N \ge 1$, we have $1 - \sum_{i=0}^N \dot{\zeta}_i(t) = (1-p(t))\frac{(N+\beta(t))\zeta_N(t)}{(1+\beta(t))t+\tilde{c}+c\beta(t)}$, and hence

$$t + \sum_{i=0}^N \zeta_i(0) - \sum_{i=0}^N \zeta_i(t) = \int_0^t (1-p(s))\frac{(N+\beta(s))\zeta_N(s)}{(1+\beta(s))s+\tilde{c}+c\beta(s)}\,ds. \tag{4.2}$$

We obtain, as the integrand on the right-hand side is nonnegative, that $\sum_{i=0}^N \zeta_i(t) \le t + \sum_{i=0}^N c_i \le t + c$ for all $t \ge 0$ and $N \ge 1$ where we recall from (LIM) $c = \sum_{i=0}^\infty c_i$. In particular, $\sum_{i\ge 0}\int_0^t \frac{\zeta_i(s)}{s+c}\,ds \le t$. Also, the right-hand side of (4.2), after a calculation, is bounded above by $\frac{N+1}{\min\{\beta_0,1\}}\int_0^t \frac{\zeta_N(s)}{s+c}\,ds$. Hence, since by nonnegativity and (LIM) the right-side of (4.2) has a limit, this limit must vanish and $\sum_{i\ge 0}\zeta_i(t) = t + c$.

Next, to establish $\sum_{i\ge 0} i\zeta_i(t) = t + \tilde{c}$, again from the ODEs, for $N \ge 1$,

$$\begin{aligned}\sum_{i=0}^N i\dot{\zeta}_i(t) &= p(t) + \bigl(1-p(t)\bigr)\frac{\sum_{i=0}^N (i+\beta(t))\zeta_i(t)}{(1+\beta(t))t+\tilde{c}+c\beta(t)} \\ &\quad - \bigl(1-p(t)\bigr)\frac{(N+1)(N+\beta(t))\zeta_N(t)}{(1+\beta(t))t+\tilde{c}+c\beta(t)}.\end{aligned} \tag{4.3}$$

From nonnegativity of ζ_i and $\sum_{i=0}^\infty \zeta_i = t + c$, we bound the right-hand side of (4.3) by $p(t) + (1-p(t))\frac{\sum_{i=0}^N i\zeta_i(t)+\beta(t)(t+c)}{(1+\beta(t))t+\tilde{c}+c\beta(t)}$. Let $s_N(t) := \sum_{i=0}^N i\zeta_i(t)$. Then, $\dot{s}_N(t) \le p(t) + (1-p(t))\frac{s_N(t)+\beta(t)(t+c)}{(1+\beta(t))t+\tilde{c}+c\beta(t)}$. Since, $s_N(t)$ is piecewise continuously differentiable, we have, by Lemma 4.3, that $s_N(t) \le t + \tilde{c}$ for $t \ge 0$ and $N \ge 1$. Hence, $\sum_{i=0}^\infty \int_0^t \frac{i\zeta_i(s)}{s+c}\,ds \le At$ since $\tilde{c} \le Ac$ for some $A > 0$ where $\tilde{c} = \sum_{i\ge 0} ic_i < \infty$.

Now, integrating both sides of ODE (4.3), we have

$$\sum_{i=0}^{N} i\zeta_i(t) - \sum_{i=0}^{N} ic_i$$
$$= \int_0^t p(s)\,ds + \int_0^t \bigl(1-p(s)\bigr)\frac{\sum_{i=0}^{N}(i+\beta(s))\zeta_i(s)}{(1+\beta(s))s+\tilde{c}+c\beta(s)}\,ds \tag{4.4}$$
$$- \int_0^t \bigl(1-p(s)\bigr)\frac{(N+1)(N+\beta(s))\zeta_N(s)}{(1+\beta(s))s+\tilde{c}+c\beta(s)}\,ds.$$

From nonnegativity, our estimates and (LIM), the last integral above has a limit. This last integral in (4.4) is bounded above by $\frac{(N+1)^2}{N\min\{\beta_0,1\}}\int_0^t \frac{N\zeta_N(s)}{s+c}\,ds$, and hence its limit must vanish. Then, using $\sum_{i=0}^{\infty}\zeta_i(t) = t+c$, we see $s(t) = \sum_{i\ge 0} i\zeta_i(t)$ satisfies the ODE in Lemma 4.3, and therefore $s(t) = t+\tilde{c}$.

Finally, to finish the postponed verification, noting (4.2), we have

$$\sum_{i=0}^{d}\bigl(1-[\dot{\zeta}(t)]_i\bigr) = \bigl(1-p(t)\bigr)\frac{s_d(t)+\beta(t)\sum_{i=0}^{d}\zeta_i}{(1+\beta(t))t+\tilde{c}+c\beta(t)}$$
$$\le \frac{t+\tilde{c}+\beta(t)(t+c)}{(1+\beta(t))t+\tilde{c}+c\beta(t)} = 1. \qquad \square$$

LEMMA 4.3. *The ODE*

$$\dot{f}(t) = G(t, f(t)) \qquad \text{with } G(t,x) = p(t) + \bigl(1-p(t)\bigr)\frac{x+\beta(t)(t+c)}{(1+\beta(t))t+\tilde{c}+c\beta(t)}$$

and initial condition $f(0)=\tilde{c}$ *has unique Carathéodory solution* $t+\tilde{c}$ *for* $t\ge 0$.

In addition, if $u(t)$ *is piecewise continuously differentiable*, $u(0)=u_0\le\tilde{c}$, *and* $\dot{u}(t)\le G(t,u(t))$, *then* $u(t)\le t+\tilde{c}$ *for* $t\ge 0$.

PROOF. Since the ODE is linear and, from the piecewise continuity assumption on p,β in (ND), f is piecewise continuously differentiable, we can solve uniquely

$$f(t) = \tilde{c}\exp\{B(0,t)\} + \int_0^t \biggl[p(s) + \frac{(1-p(s))\beta(s)(s+c)}{(1+\beta(s))s+\tilde{c}+c\beta(s)}\biggr]\exp\{B(s,t)\}\,ds,$$

where $B(q,r) = \int_q^r \frac{1-p(v)}{(1+\beta(v))v+\tilde{c}+c\beta(v)}\,dv$. Recall the convention $0\cdot\infty = 0$, so when $c=0$ the first term $\tilde{c}e^{B(0,t)} = 0$ vanishes. However, $t+\tilde{c}$ is a solution, and therefore $f(t)$ may be identified as desired.

The second statement is obtained similarly. $\square$

PROOF OF COROLLARY 1.7. Any root of I_d must be a Carathéodory solution to ODE (1.5). Hence, by Lemmas 4.1 and 4.2, $\zeta^d\in\Gamma_d$ is the unique minimizer of I_d. The LLNs now follow from the LDP upper bound in Theorem 1.2 and

Borel–Cantelli lemma. Statements about "mass" and "weight" of ζ^∞ are proved in Lemma 4.2. □

5. Proof of Corollary 1.9. Since $[\zeta^\infty]_i = [\zeta^d]_i$ for $i \le d$, the proof follows from the next lemma. Define, for $o_1, o_2, o_3, o_4, o_5 \ge 0$, the ODEs, $O(o_1, o_2, o_3, o_4, o_5)$: with initial condition $\varphi(0) = \mathbf{c}^d$

$$\dot{\varphi}_0(t) = 1 - o_1 - (1 - o_2)\frac{o_3}{1 + o_4} \cdot \frac{\varphi_0(t)}{t + o_5},$$

$$[\dot{\varphi}(t)]_i = 1 - (1 - o_2)\frac{i + o_3}{1 + o_4} \cdot \frac{\varphi_i(t)}{t + o_5} \qquad \text{for } 1 \le i \le d.$$

One can check that $\chi(t)$ is the solution to $O(o_1, o_2, o_3, o_4, o_5)$ above for $0 \le o_2 \le 1$, where

$$\chi_i(t) = b_i(t + o_5) + \sum_{\ell=0}^{i} a_{i,\ell}\left(\frac{o_5}{t + o_5}\right)^{(1-o_2)(\ell+o_3)/(1+o_4)} \qquad \text{for } 0 \le i \le d. \tag{5.1}$$

Here, the sequence $b_i = b_i(o_1, o_2, o_3, o_4, o_5)$ is defined by $b_0 = \frac{1-o_1}{1+(1-o_2)o_3/(1+o_4)}$, $b_1 = \frac{o_1+(1-o_2)o_3 b_0/(1+o_4)}{1+(1-o_2)(1+o_3)/(1+o_4)}$, and, for $i \ge 2$,

$$\begin{aligned} b_i &= b_1 \prod_{\ell=2}^{i} \frac{(1 - o_2)(\ell - 1 + o_3)/(1 + o_4)}{1 + (1 - o_2)(\ell + o_3)/(1 + o_4)} \\ &= b_1 \frac{\Gamma(2 + o_3 + (1 + o_4)/(1 - o_2))}{\Gamma(1 + o_3)} \frac{\Gamma(i + o_3)}{\Gamma(i + 1 + o_3 + (1 + o_4)/(1 - o_2))} \\ &\sim \frac{1}{i^{1+(1+o_4)/(1-o_2)}}. \end{aligned}$$

The sequence $a_{i,\ell} = a_{i,\ell}(o_1, o_2, o_3, o_4, o_5)$ is given by $a_{0,0} = c_0 - b_0 o_5$, and, for $i \ge 1$,

$$a_{i,\ell} = \frac{i - 1 + o_3}{i - \ell} a_{i-1,\ell} \qquad \text{where } 0 \le \ell < i$$

and

$$a_{i,i} = c_i - b_i o_5 - \sum_{\ell=0}^{i-1} a_{i,\ell}.$$

Recall now the assumption in Corollary 1.9: $0 \le p_{\min} \le p(\cdot) \le p_{\max} < 1$ and $0 < \beta_{\min} \le \beta(\cdot) \le \beta_{\max} < \infty$.

LEMMA 5.1. *The systems $O(p_{\min}, p_{\max}, \beta_{\min}, \beta_{\max}, \max\{\tilde{c}, c\})$ and $O(p_{\max}, p_{\min}, \beta_{\max}, \beta_{\min}, \min\{\tilde{c}, c\})$ have respective unique solutions $\tilde{\zeta}$ and $\hat{\zeta}$. Then, for*

$0 \le i \le d$ *and* $t \in [0, 1]$, *with respect to the zero-cost trajectory* $\zeta^d(t)$ *in Corollary* 1.7 *with initial condition* $\zeta^d(0) = \mathbf{c}^d$, *we have*

$$[\hat{\zeta}(t)]_i \le [\zeta^d(t)]_i \le [\tilde{\zeta}(t)]_i.$$

PROOF. The proof that $\tilde{\zeta}$ and $\hat{\zeta}$ are the unique solutions uses a similar argument to that in the proof of Lemma 4.1. We now establish the inequality in the display with respect to $\tilde{\zeta}$ as an analogous proof works for $\hat{\zeta}$. We use induction to see that $[\tilde{\zeta}]_i \ge [\zeta]_i$ for $0 \le i \le d$.

Since $\tilde{\zeta}(0) = \zeta(0) = \mathbf{c}^d$, from ODEs, $O(p_{\min}, p_{\max}, \beta_{\min}, \beta_{\max}, \max\{\tilde{c}, c\})$ and (1.5), we have

$$\dot{\tilde{\zeta}}_0(t) - \dot{\zeta}_0(t) \ge p(t) - p_{\min} + (1 - p_{\max}) \frac{\beta_{\min}(\zeta_0(t) - \tilde{\zeta}_0(t))}{(1 + \beta_{\max})(t + \max\{\tilde{c}, c\})}, \tag{5.2}$$

$$[\dot{\tilde{\zeta}}(t)]_i - [\dot{\zeta}(t)]_i \ge (1 - p_{\max}) \frac{(i + \beta_{\min})(\zeta_i(t) - \tilde{\zeta}_i(t))}{(1 + \beta_{\max})(t + \max\{\tilde{c}, c\})}. \tag{5.3}$$

For $i = 0$, suppose $\tilde{\zeta}_0(t) < \zeta_0(t)$ for some t. Then, by continuity, we may assume that $\tilde{\zeta}_0(t) < \zeta_0(t)$ for all $t \in (t_0, t_1]$ for some $0 \le t_0 < t_1 \le 1$, and $\tilde{\zeta}_0(t_0) = \zeta_0(t_0)$. We may further arrange t_0, t_1, from the piecewise continuity assumptions in (ND), that p, β are continuous on (t_0, t_1). From the mean value theorem, we find a $t' \in (t_0, t_1)$ such that $\dot{\tilde{\zeta}}_0(t') < \dot{\zeta}_0(t')$, which contradicts the ODE (5.2) as it gives $\dot{\tilde{\zeta}}_0(t') - \dot{\zeta}_0(t') > 0$. Therefore, $\tilde{\zeta}_0 \ge \zeta_0$.

Now, for $1 \le i \le d$, suppose $[\tilde{\zeta}(t)]_i < [\zeta(t)]_i$ for some t. By induction hypothesis ($[\tilde{\zeta}(\cdot)]_{i-1} \ge [\zeta(\cdot)]_{i-1}$), we must have $\tilde{\zeta}_i(t) < \zeta_i(t)$. Since $[\tilde{\zeta}(\cdot)]_i$, $[\zeta(\cdot)]_i$, $\tilde{\zeta}_i(\cdot)$ and $\zeta_i(\cdot)$ are continuous functions, as for the case $i = 0$, we may assume $[\tilde{\zeta}(t)]_i < [\zeta(t)]_i$ and $\tilde{\zeta}_i(t) < \zeta_i(t)$, and p, β are continuous for all $t \in (t_0, t_1)$ for some $0 \le t_0 < t_1 \le 1$, and also $\tilde{\zeta}_i(t_0) = \zeta_i(t_0)$. By the mean value theorem for $[\tilde{\zeta}(t)]_i - [\zeta(t)]_i$, there is $t' \in (t_0, t_1)$ such that $[\dot{\tilde{\zeta}}(t')]_i < [\dot{\zeta}(t')]_i$. But (5.3) gives $[\dot{\tilde{\zeta}}(t')]_i - [\dot{\zeta}(t')]_i > 0$, a contradiction. Therefore $[\tilde{\zeta}]_i \ge [\zeta]_i$. $\square$

PROOF OF COROLLARY 1.9. Given Lemma 5.1, we need only detail the solutions $\tilde{\zeta}$ and $\hat{\zeta}$ when the initial configuration is "small" and "large," respectively. To this end, when the initial configuration is "small" ($c_i \equiv 0$), $\tilde{\zeta}$, $\hat{\zeta}$ are linear, namely $\tilde{\zeta}_i(t) = \tilde{b}_i t$, and $\hat{\zeta}_i(t) = \hat{b}_i t$, where $\tilde{b}_i := b_i(p_{\min}, p_{\max}, \beta_{\min}, \beta_{\max}, 0)$ and $\hat{b}_i := b_i(p_{\max}, p_{\min}, \beta_{\max}, \beta_{\min}, 0)$ [cf. (5.1)].

On the other hand, when the initial configuration is "large" ($c_i > 0$ for some $0 \le i \le d + 1$), as $t \uparrow \infty$, $\tilde{\zeta}_i(t) = (\tilde{b}_i + o(1))t$ and $\hat{\zeta}_i(t) = (\hat{b}_i + o(1))t$. $\square$

Acknowledgments. We thank the referees/editors for constructive comments which helped improve the exposition of the paper.

REFERENCES

[1] ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys*. **74** 47–97. MR1895096

[2] ATHREYA, K. B., GHOSH, A. P. and SETHURAMAN, S. (2008). Growth of preferential attachment random graphs via continuous-time branching processes. *Proc. Indian Acad. Sci. Math. Sci*. **118** 473–494. MR2450248

[3] BARABÁSI, A.-L. (2009). Scale-free networks: A decade and beyond. *Science* **325** 412–413. MR2548299

[4] BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

[5] BARRAT, A., BARTHELEMY, M., PASTOR-SATORRAS, R. and VESPIGNANI, A. (2004). The architecture of complex weighted networks. *PNAS* **101** 3747–3752.

[6] BERGER, N., BORGS, C., CHAYES, J. and SABERI, A. (2009). A weak local limit for preferential attachment graphs. Preprint.

[7] BERGER, N., BORGS, C., CHAYES, J. T. and SABERI, A. (2005). On the spread of viruses on the internet. In *Proceedings of the Sixteenth Annual ACM–SIAM Symposium on Discrete Algorithms* 301–310 (electronic). ACM, New York. MR2298278

[8] BHAMIDI, S. (2007). Universal techniques to analyze preferential attachment trees: Global and local analysis. Preprint. Available at http://www.unc.edu/~bhamidi/preferent.pdf.

[9] BOLLOBÁS, B. and RIORDAN, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24** 5–34. MR2057681

[10] BOLLOBÁS, B., RIORDAN, O., SPENCER, J. and TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** 279–290. MR1824277

[11] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. and VESZTERGOMBI, K. (2011). Limits of randomly grown graph sequences. *European J. Combin*. **32** 985–999. MR2825531

[12] BORNHOLDT, S. and SCHUSTER, H. G., eds. (2003). *Handbook of Graphs and Networks*: *From the Genome to the Internet*. Wiley–VCH, Weinheim. MR2016116

[13] BRYC, W., MINDA, D. and SETHURAMAN, S. (2009). Large deviations for the leaves in some random trees. *Adv. in Appl. Probab.* **41** 845–873. MR2571319

[14] CALDARELLI, G. (2007). *Scale-Free Networks*: *Complex Webs in Nature and Technology*. Oxford Univ. Press, Oxford.

[15] CHUNG, F., HANDJANI, S. and JUNGREIS, D. (2003). Generalizations of Polya's urn problem. *Ann. Comb.* **7** 141–153. MR1994572

[16] CHUNG, F. and LU, L. (2006). *Complex Graphs and Networks*. *CBMS Regional Conference Series in Mathematics* **107**. Amer. Math. Soc., Providence, RI. MR2248695

[17] COHEN, R. and HAVLIN, S. (2010). *Complex Networks*: *Structure Robustness and Function*. Cambridge Univ. Press, Cambridge.

[18] COOPER, C. and FRIEZE, A. (2003). A general model of web graphs. *Random Structures Algorithms* **22** 311–335. MR1966545

[19] COOPER, C. and FRIEZE, A. (2007). The cover time of the preferential attachment graph. *J. Combin. Theory Ser. B* **97** 269–290. MR2290325

[20] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036

[21] DEREICH, S. and MÖRTERS, P. (2009). Random networks with sublinear preferential attachment: Degree evolutions. *Electron. J. Probab.* **14** 1222–1267. MR2511283

[22] DEREICH, S., MÖNCH, C. and MÖRTERS, P. (2011). Typical distances in ultrasmall random networks. Available at arXiv:1102.5680v1.

[23] DOROGOVTSEV, S. N., KRAPIVSKY, P. L. and MENDÉS, J. F. F. (2008). Transition from small to large world in growing networks. *Europhys. Lett. EPL* **81** Art. 30004, 5. MR2443959

[24] DOROGOVTSEV, S. N. and MENDES, J. F. F. (2000). Evolution of networks with aging of sites. *Phys. Rev. E* **62** 1842–1845.

[25] DOROGOVTSEV, S. N. and MENDES, J. F. F. (2001). Scaling properties of scale-free evolving networks: Continuous approach. *Phys. Rev. E* **63** 056125 19 pp.

[26] DOROGOVTSEV, S. N. and MENDES, J. F. F. (2003). *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford Univ. Press, Oxford. MR1993912

[27] DRINEA, E., ENACHESCU, M. and MITZENMACHER, M. (2001). Variations on random graph models for the web. Harvard Technical Report TR-06-01.

[28] DRINEA, E., FRIEZE, A. and MITZENMACHER, M. (2002). Balls and Bins models with feedback. In *Proc. of the 11th ACM–SIAM Symposium on Discrete Algorithms* (*SODA*) 308–315. SIAM, Philadelphia, PA.

[29] DUPUIS, P. and ELLIS, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley, New York. MR1431744

[30] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge Univ. Press, Cambridge. MR2271734

[31] FORTUNATO, S., FLAMMINI, A. and MENCZER, F. (2006). Scale-free network growth by ranking. *Phys. Rev. Lett.* **96** 218701-1–218701-4.

[32] FRIEZE, A., VERA, J. and CHAKRABARTI, S. (2006). The influence of search engines on preferential attachment. *Internet Math.* **3** 361–381. MR2372548

[33] GJOKA, M., KURANT, M., BUTTS, C. T. and MARKOPOULOU, A. (2011). Practical recommendations on crawling online social networks. *IEEE Journal on Selected Areas in Communications* **29** 1872–1892.

[34] JANSSEN, J. and PRAŁAT, P. (2010). Rank-based attachment leads to power law graphs. *SIAM J. Discrete Math.* **24** 420–440. MR2646095

[35] KATONA, Z. (2005). Width of a scale-free tree. *J. Appl. Probab.* **42** 839–850. MR2157524

[36] KRAPIVSKY, P. and REDNER, S. (2001). Organization of growing random networks. *Phys. Rev. E* **63** 066123-1–066123-14.

[37] KRAPIVSKY, P. L. and REDNER, S. (2002). Finiteness and fluctuations in growing networks. *J. Phys. A* **35** 9517–9534. MR1946936

[38] KRAPIVSKY, P. L., RODGERS, G. J. and REDNER, S. (2001). Degree distributions of growing networks. *Phys. Rev. Lett.* **86** 5401–5404.

[39] MIHAIL, M., PAPADIMITRIOU, C. and SABERI, A. (2006). On certain connectivity properties of the internet topology. *J. Comput. System Sci.* **72** 239–251. MR2205286

[40] MITZENMACHER, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Math.* **1** 226–251. MR2077227

[41] MÓRI, T. F. (2002). On random trees. *Studia Sci. Math. Hungar.* **39** 143–155. MR1909153

[42] MÓRI, T. F. (2005). The maximum degree of the Barabási–Albert random tree. *Combin. Probab. Comput.* **14** 339–348. MR2138118

[43] NEWMAN, M., BARABÁSI, A.-L. and WATTS, D. J., eds. (2006). *The Structure and Dynamics of Networks*. Princeton Univ. Press, Princeton, NJ. MR2352222

[44] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45** 167–256 (electronic). MR2010377

[45] NEWMAN, M. E. J. (2010). *Networks: An Introduction*. Oxford Univ. Press, Oxford. MR2676073

[46] OLIVEIRA, R. and SPENCER, J. (2005). Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.* **2** 121–163. MR2193157

[47] RÁTH, B. and SZAKÁCS, L. (2011). Multigraph limit of the dense configuration model and the preferential attachment graph. Preprint. Available at arXiv:1106.2058.

[48] RUDAS, A., TÓTH, B. and VALKÓ, B. (2007). Random trees and general branching processes. *Random Structures Algorithms* **31** 186–202. MR2343718

[49] SIMKIN, M. V. and ROYCHOWDHURY, V. P. (2011). Re-inventing Willis. *Phys*. *Rep*. **502** 1–35. MR2788549

[50] SIMON, H. A. (1955). On a class of skew distribution functions. *Biometrika* **42** 425–440. MR0073085

[51] YULE, G. U. (1924). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. *Philos*. *Trans*. *Roy*. *Soc*. *London Ser*. *B* **213** 21–87.

[52] ZHANG, J. X. and DUPUIS, P. (2008). Large-deviation approximations for general occupancy models. *Combin*. *Probab*. *Comput*. **17** 437–470. MR2410397

DEPARTMENT OF MATHEMATICS
SYRACUSE UNIVERSITY
SYRACUSE, NEW YORK 13244
USA
E-MAIL: jchoi46@syr.edu

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ARIZONA
617 N. SANTA RITA AVE.
TUCSON, ARIZONA 85721
USA
E-MAIL: sethuram@math.arizona.edu