

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/46585345>

Large deviation results for Critical Multitype Galton–Watson trees

Article in *Far East Journal of Mathematical Sciences* · August 2017

DOI: 10.17654/MS102030463 · Source: arXiv

CITATIONS

3

READS

60

1 author:



[Kwabena Doku-Amponsah](#)

University of Ghana

42 PUBLICATIONS 59 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Joint Large deviation Principles for d -regular random graphs [View project](#)

LARGE DEVIATION RESULTS FOR MULTITYPE GALTON-WATSON TREES

BY KWABENA DOKU-AMPONSAH
University of Ghana

Draft of September 15, 2010

Abstract

Using the notion of *consistency* of empirical measures, under fairly general assumption we prove a joint large deviation principles in n for the *empirical pair measure* and *empirical offspring measure* of multitype Galton-Watson tree conditioned to have exactly n vertices in the weak topology. From these results we obtain large deviation principle for empirical pair measure of Markov chains indexed by simply generated trees obtain by conditioning Galton-Watson trees on the total number of vertices. For the case where the offspring law of the tree is geometric distribution with parameter $\frac{1}{2}$, we get an exact rate function. All our rate functions are expressed in terms of relative entropies.

Keywords: Tree-indexed Markov chain, Tree-indexed process, random tree, Galton-Watson tree, multitype Galton-Watson tree, typed tree, joint large deviation principle, empirical pair measure, empirical offspring measure.

MSC 2000: Primary 60F10. Secondary 60J80, 05C05.

1. INTRODUCTION

Our main motivation for revisiting this model is the study of *Markov chain indexed by geometric $\frac{1}{2}$ offspring law* defined as follows: First we sample a tree from geometric distribution with parameter $\frac{1}{2}$, and then, given this tree, we run a Markov chain on the vertices of the tree in such a way that the state of a vertex depends only on the state of its parent. The result of this two-step experiment can also be interpreted as a *typed tree*. Large deviation principle for empirical pair measure of Markov chains indexed by trees and the limit points find many application in information theory, statistical physics and the theory of Gibbs measures. See e.g. [3] or [4] for basic information theorem for hierarchical structures.

Large deviation study of Galton-Watson trees conditioned on the total size was first study by Dembo, Mörters and Sheffield [6]. In this paper, notions of shift-invariance and specific relative entropy-as typically understood for Markov fields on deterministic graphs such as \mathbb{Z}^d was extended to Markov fields on random trees. With these concepts, large deviation principles for empirical measures of a class of random trees including multitype Galton-Watson trees conditioned to have exactly n vertices were proved in a topology *stronger* than the weak topology. Their analysis have shown that large deviation results, which are well-known for classical Markov chains, can be extended to Markov chains indexed by random trees with offspring laws which have superexponential decay at infinity.i.e. Offspring law $p(\cdot)$ with $\ell^{-1} \log p(\ell) \rightarrow -\infty$ as $\ell \rightarrow \infty$. In the course of the proof of their main result, large deviation

principle for the empirical offspring measure of multitype Galton-Watson trees whose exponential moments are all finite was established in their topology, see [6].

The aim of this paper is to prove joint large deviation principle for the empirical pair measure and empirical offspring measure of multitype Galton-Watson trees with offspring laws which have finite second moments. This includes offspring laws considered in [6]. We extend the concept of *consistency* as understood for empirical measures of coloured random graphs, see Doku and Mörters [5] or Doku [3], to multitype Galton-Watson trees. The proof of our main results use the technique of (exponential) *change of measure* and three large deviations results, see [6, Lemmas 3.1 and 3.6] and [6, Theorem 2.2].

Using the contraction principle, see Dembo [7], we derive from our main results large deviation principle for empirical pair measure of Markov chain indexed by random trees, see Benjamini and Peres[2]. This result is similar to the one in [6]. We remark here that the process level large deviation principles for the empirical subtree measure and single-generation empirical measure, see [6], can be developed from our main results.

Specifically, we consider random tree models where trees and types are chosen *simultaneously* according to a *multitype Galton-Watson tree*. We recall from [6] the model of multitype Galton-Watson tree. To begin, we write $\mathcal{X}^* = \bigcup_{n=0}^{\infty} \{n\} \times \mathcal{X}^n$ and equip it with the discrete topology. We denote by \mathcal{T} the set of all finite rooted planar trees T , by $V = V(T)$ the set of all vertices and by $E = E(T)$ the set of all edges oriented away from the root, which is always denoted by ρ . We write $|T|$ for the number of vertices in the tree T . We note that the offspring of any vertex $v \in T$ is characterized by an element of \mathcal{X}^* and that there is an element $(0, \emptyset)$ in \mathcal{X}^* symbolizing absence of offspring. For each typed tree X and each vertex v we denote by

$$C(v) = (N(v), X_1(v), \dots, X_{N(v)}(v)) \in \mathcal{X}^*$$

the number and types of the children of v , ordered from left to right.

Given a probability measure μ on \mathcal{X} , serving as the initial distribution, and an offspring transition kernel \mathbb{Q} from \mathcal{X} to \mathcal{X}^* , we define the law \mathbb{P} of a tree-indexed process X , see Pemantle [9], by the following rules:

- The root ρ carries a random type $X(\rho)$ chosen according to the probability measure μ on \mathcal{X} .
- For each vertex with type $a \in \mathcal{X}$ the offspring number and types are given independently of everything else, by the offspring law $\mathbb{Q}\{\cdot | a\}$ on \mathcal{X}^* . We write

$$\mathbb{Q}\{\cdot | a\} = \mathbb{Q}\{(N, X_1, \dots, X_N) \in \cdot | a\},$$

i.e. we have a random number N of offspring particles with types X_1, \dots, X_N .

For every $c = (n, a_1, \dots, a_n) \in \mathcal{X}^*$ and $a \in \mathcal{X}$, the *multiplicity* of the symbol a in c is given by $m(a, c) = \sum_{i=1}^n 1_{\{a_i=a\}}$, and the matrix A with index set $\mathcal{X} \times \mathcal{X}$ and nonnegative entries is given by

$$A(a, b) = \sum_{c \in \mathcal{X}^*} \mathbb{Q}\{c | b\} m(a, c), \text{ for } a, b \in \mathcal{X}. \text{ i.e. } A(a, b) \text{ are the expected number of offspring of type } a$$

of a vertex of type b . We also recall from [6] the weak form of irreducibility concept. With $A^*(a, b) = \sum_{k=1}^{\infty} A^k(a, b) \in [0, \infty]$ we say that the matrix A is *weakly irreducible* if \mathcal{X} can be partitioned into a non empty set \mathcal{X}_r of *recurrent states* and a disjoint set \mathcal{X}_t of *transient states* such that

- $A^*(a, b) > 0$ whenever $b \in \mathcal{X}_r$, while
- $A^*(a, b) = 0$ whenever $b \in \mathcal{X}_t$ and either $a = b$ or $a \in \mathcal{X}_r$.

For example, any *irreducible* matrix A has A^* strictly positive, hence is also weakly irreducible with $\mathcal{X}_r = \mathcal{X}$. The multitype Galton-Watson tree is called weakly irreducible (or irreducible) if the matrix A

is weakly irreducible (or irreducible, respectively) and the number $\sum_{a \in \mathcal{X}_t} m(a, c)$ of transient offspring is uniformly bounded under \mathbb{Q} .

Recall that, by the Perron-Frobenius theorem, see e.g. [7, Theorem 3.1.1], the largest eigenvalue of an irreducible matrix is real and positive. Obviously, the same applies to weakly irreducible matrices. The multitype Galton-Watson tree is called *critical* if this eigenvalue is 1 for the matrix A .

The remaining part of the paper is organized in the following manner: The complete statement of our results is given in Section 2, we begin with joint LDP for empirical pair measures and empirical offspring measures of multitype Galton-Watson trees, followed by a corollary of LDP for the empirical offspring measure of multitype Galton-Watson trees in subsection 2.1. In subsection 2.2 we state the LDP for empirical pair measures of Markov chains indexed by tree. The proofs of our main results are then given in Section 3. All corollaries and Theorem 2.4 are proved in Section 4.

2. STATEMENT OF THE RESULTS

2.1 Joint large deviation principle for empirical pair measure and empirical offspring measure of multitype Galton-Watson trees.

For every sample chain X , we associate the *empirical offspring measure* M_X on $\mathcal{X} \times \mathcal{X}^*$, by

$$M_X(a, c) = \frac{1}{|T|} \sum_{v \in V} \delta_{(X(v), C(v))}(a, c), \quad (2.1)$$

and the empirical pair measure on $\mathcal{X} \times \mathcal{X}$, by

$$\tilde{L}_X(a, b) = \frac{1}{|T|} \sum_{e \in E} \delta_{(X(e_1), X(e_2))}(a, b), \quad \text{for } a, b \in \mathcal{X}, \quad (2.2)$$

where e_1, e_2 are the beginning and end vertex of the edge $e \in E$ (so e_1 is closer to ρ than e_2). We note that

$$\tilde{L}_X(a, b) = \sum_{c \in \mathcal{X}^*} m(b, c) M_X(a, c).$$

By definition, we notice that M_X is a probability vector and that total mass $\|L_X\|$ of L_X is $\frac{|T|-1}{|T|} \leq 1$.

Our main result is a large deviation principle for (\tilde{L}_X, M_X) if X is a multitype Galton-Watson tree. For its formulation denote, for every probability measure ν on $\mathcal{X} \times \mathcal{X}^*$, by ν_1 the \mathcal{X} -marginal of ν .

We call (ϖ, ν) *sub-consistent* if

$$\varpi(a, b) \geq \sum_{c \in \mathcal{X}^*} m(b, c) \nu(a, c) \quad \text{for all } a, b \in \mathcal{X}. \quad (2.3)$$

It call *consistent* if equality hold in (2.3). Observe that, if (ϖ, ν) is empirical pair measure and empirical offspring measure of multitype Galton-Watson tree then both sides of (2.3) is

$$\frac{1}{n} \times \#\{\text{number of edges with beginning vertex of type } a \text{ and end vertex of type } b\}.$$

We denote by $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ the space of probability measures ν on $\mathcal{X} \times \mathcal{X}^*$ with $\int n \nu(da, dc) < \infty$, using the convention $c = (n, a_1, \dots, a_n)$. Denote by $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ the space of finite measures on $\mathcal{X} \times \mathcal{X}$ and endow the space $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with the weak topology.

We call an offspring distribution \mathbb{Q} *bounded* if for some $k < \infty$, we have

$$\mathbb{Q}\{N > k | a\} = 0, \quad \text{for all } a \in \mathcal{X}.$$

Otherwise we call it *unbounded*.

Theorem 2.1. *Suppose that X is a weakly irreducible, critical multitype Galton-Watson tree with an unbounded offspring law \mathbb{Q} whose second moment is finite, conditioned to have exactly n vertices. Then, for $n \rightarrow \infty$, (\tilde{L}_X, M_X) satisfies a large deviation principle in $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with speed n and the convex, good rate function*

$$J(\varpi, \nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}) & \text{if } (\varpi, \nu) \text{ is sub-consistent and } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (2.4)$$

For $\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X})$, we write

$$\langle m(\cdot, c), \nu(a, c) \rangle(b) := \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} m(b, c) \nu(a, c), \text{ for } b \in \mathcal{X},$$

and state a corollary of Theorem 2.1.

Corollary 2.2. *Suppose that X is a weakly irreducible, critical multitype Galton-Watson tree with an unbounded offspring law \mathbb{Q} whose second moment is finite, conditioned to have exactly n vertices. Then, for $n \rightarrow \infty$, the empirical offspring measure M_X satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with speed n and the convex, good rate function*

$$K(\nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}) & \text{if } \langle m(\cdot, c), \nu(a, c) \rangle \leq \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (2.5)$$

We write $\mathcal{X}_k^* = \bigcup_{n=0}^k \{n\} \times \mathcal{X}^n$ and denote by \mathbb{Q}_k offspring transition kernel from \mathcal{X} to \mathcal{X}_k^* . The next large deviation principle is the main ingredient in the proof of the lower bound of Theorem 2.1.

Theorem 2.3. *Suppose that X is a weakly irreducible, critical multitype Galton-Watson tree with an offspring law \mathbb{Q}_k , conditioned to have exactly n vertices. Then, for $n \rightarrow \infty$, (\tilde{L}_X, M_X) satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ with speed n and the convex, good rate function*

$$J_k(\varpi, \nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}_k) & \text{if } (\varpi, \nu) \text{ is consistent and } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (2.6)$$

2.2 LDP for empirical pair measure of Markov chains indexed by trees. In this subsection, we look at the situation where the tree is generated independently of the types.

Suppose that T is any finite tree and we are given an initial probability measure μ on a finite alphabet \mathcal{X} and a Markovian transition kernel $Q: \mathcal{X} \times \mathcal{X} \geq 0$. We can obtain a *Markov chain indexed by tree* $T, X: V \rightarrow \mathcal{X}$ as follows: Choose $X(\rho)$ according to μ and choose $X(v)$, for each vertex $v \neq \rho$, using the transition kernel given the value of its parent, independently of everything else. If the tree is chosen randomly, we always consider $X = \{X(v) : v \in T\}$ under the *joint law* of tree and chain. It is sometimes convenient to interpret X as a *typed tree*, considering $X(v)$ as the *type* of the vertex v .

We consider the class of *simply generated trees*, see [8] or [1], obtained by conditioning a critical Galton-Watson on its total number of vertices. To be specific, we look at the class of Galton-Watson trees, where the number of children $N(v)$ of each $v \in T$ is chosen independently according to the same law $p(\cdot) = \mathbb{P}\{N(v) = \cdot\}$ for all $v \in T$, while $0 < p(0) < 1$. We assume that p is critical. That is, the mean offspring number $\sum_{\ell=0}^{\infty} \ell p(\ell)$ is one, but this assumption is not restrictive: Note that the distribution of T conditioned on $\{|T| = n\}$ is exactly the same as when the offspring law is $p_{\theta}(\ell) = p(\ell)e^{\theta\ell} / \sum_j p(j)e^{\theta j}$, regardless of the value of $\theta \in \mathbb{R}$. With $0 < p(0) < 1 - p(1)$ there exists a unique θ_* such that $\sum_{\ell} \ell p_{\theta_*}(\ell) = 1$. Hence all our results hold in the noncritical cases with p_{θ_*} in place of p . We allow offspring laws p with unbounded support, but we relax the assumption

$$\ell^{-1} \log p(\ell) \rightarrow -\infty.$$

Throughout the paper statements conditioned on the event $\{|T| = n\}$ are made only for those values of n where $\mathbb{P}\{|T| = n\} > 0$.

We associate with each finite tree and sample chain X a probability measure on $\mathcal{X} \times \mathcal{X}$, the (normalized) *empirical pair measure* L_X , by

$$L_X(a, b) = \frac{1}{|E|} \sum_{e \in E} \delta_{(X(e_1), X(e_2))}(a, b), \text{ for } a, b \in \mathcal{X}, \quad (2.7)$$

where e_1, e_2 are the beginning and end vertex of the edge $e \in E$ (so e_1 is closer to ρ than e_2). Note that, $L_X = \frac{n}{n-1} \tilde{L}_X$ on the set $\{|T| = n\}$ and hence the LDP for \tilde{L}_X implies L_X by exponential equivalent Theorem, see Dembo [7].

Our first result in this subsection is a large deviation principle for L_X , conditional upon the event $\{|T| = n\}$ with n chosen such that the latter has positive probability. For its formulation recall the definition of the relative entropy $H(\cdot \| \cdot)$ from [7, (2.1.5)] and the Cramér's rate function, see e.g. [7, (2.1.26)],

$$I_p(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - \log \left[\sum_{n=0}^{\infty} p(n) e^{\lambda n} \right] \right\}. \quad (2.8)$$

Theorem 2.4. *Suppose that T is a Galton-Watson tree, with offspring law $p(\cdot)$ such that $0 < p(0) < 1 - p(1)$, $\sum_{\ell} \ell p(\ell) = 1$ and $\sum_{\ell} \ell^2 p(\ell) < \infty$. Let X be a Markov chain indexed by T with arbitrary initial distribution and an irreducible Markovian transition kernel Q . Then, for $n \rightarrow \infty$, the empirical pair measure L_X , conditioned on $\{|T| = n\}$ satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X})$ with speed n and the convex, good rate function*

$$I(\mu) = \begin{cases} H(\mu \| \mu_1 \otimes Q) + \sum_{a \in \mathcal{X}} \mu_2(a) I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right) & \text{if } \mu_1 \ll \mu_2, \\ \infty & \text{otherwise,} \end{cases} \quad (2.9)$$

where μ_1 and μ_2 are the first and second marginal of μ and $\mu_1 \otimes Q(a, b) = Q\{b | a\} \mu_1(a)$.

From Theorem 2.4 we obtain LDP for empirical pair measures of Galton-Watson trees with geometric distribution with parameter $\frac{1}{2}$.

Corollary 2.5. *Suppose that T is a Galton-Watson tree, with offspring law $p(\ell) = 2^{-(\ell+1)}$, $\ell = 0, 1, \dots$. Let X be a Markov chain indexed by T with arbitrary initial distribution and an irreducible Markovian transition kernel Q . Then, for $n \rightarrow \infty$, the empirical pair measure L_X , conditioned on $\{|T| = n\}$ satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X})$ with speed n and the convex, good rate function*

$$I(\mu) = \begin{cases} H(\mu \| \mu_1 \otimes Q) + H(\mu_1 \| (\mu_1 + \mu_2)/2) + H(\mu_2 \| (\mu_1 + \mu_2)/2) & \text{if } \mu_1 \ll \mu_2, \\ \infty & \text{otherwise,} \end{cases} \quad (2.10)$$

where μ_1 and μ_2 are the first and second marginal of μ and $\mu_1 \otimes Q(a, b) = Q\{b | a\} \mu_1(a)$.

3. PROOF OF MAIN RESULTS

3.1 Change of Measure, Exponential Tightness and Some General Principles.

Given a bounded function $\tilde{g} : \mathcal{X} \times \mathcal{X}^* \rightarrow \mathbb{R}$ we define the function

$$U_{\tilde{g}}(a) = \log \sum_{c \in \mathcal{X}^*} \mathbb{Q}\{c | a\} e^{\tilde{g}(a, c)}, \quad (3.1)$$

for $a \in \mathcal{X}$. We use \tilde{g} to define a new multitype Galton-Watson tree as follows:

- The type of the root ρ is $a \in \mathcal{X}$ with probability

$$\mu_{\tilde{g}}(a) = \frac{e^{U_{\tilde{g}}(a)} \mu(a)}{\int e^{U_{\tilde{g}}(b)} \mu(db)}. \quad (3.2)$$

- for each vertex with type $a \in \mathcal{X}$ the offspring number and types are given independently of everything else, by the offspring law $\tilde{\mathbb{Q}}\{\cdot | a\}$ given by

$$\tilde{\mathbb{Q}}\{c | a\} = \exp(\tilde{g}(a, c) - U_{\tilde{g}}(a)) \mathbb{Q}\{c | a\}. \quad (3.3)$$

We denote the transformed law by $\tilde{\mathbb{P}}$ and make the simple observation that $\tilde{\mathbb{P}}$ is absolutely continuous with respect to \mathbb{P} , as for each finite $X \in \tilde{\mathcal{X}}$,

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(X) = \frac{e^{U_{\tilde{g}}(X(\rho))}}{\int e^{U_{\tilde{g}}(b)} \mu(db)} \prod_{v \in V} \exp[\tilde{g}(X(v), C(v)) - U_{\tilde{g}}(X(v))] \quad (3.4)$$

$$= \frac{1}{\int e^{U_{\tilde{g}}(a)} \mu(da)} \prod_{v \in V} \exp\left[\tilde{g}(X(v), C(v)) - \sum_{b \in \mathcal{X}} m(b, C(v)) U_{\tilde{g}}(b)\right], \quad (3.5)$$

recalling that $C(v) = (N(v), X_1(v), \dots, X_N(v))$.

We begin by establishing exponential tightness of the family of laws of M_X on the space $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$.

Lemma 3.1. *For every $\alpha > 0$ there exists a compact $K_\alpha \subset \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{M_X \notin K_\alpha \mid |T| = n\} \leq -\alpha.$$

Proof. Given $l \in \mathbb{N}$, we may choose $k(l) \in \mathbb{N}$ so large that

$$\mathbb{Q}\{\exp(l^2 \mathbf{1}_{\{N > k(l)\}}) \mid a\} \leq 2, \text{ for all } a \in \mathcal{X}.$$

Using the exponential Chebyshev inequality,

$$\begin{aligned} \mathbb{P}\left\{\int_{\{N > k(l)\}} dM_X \geq \frac{1}{l}, |T| = n\right\} &\leq e^{-ln} \mathbb{E}\left\{\exp\left(l^2 n \int_{\{N > k(l)\}} dM_X\right), |T| = n\right\} \\ &= e^{-ln} \mathbb{E}\left\{\prod_{v \in T} \exp(l^2 \mathbf{1}_{\{N(v) > k(l)\}}), |T| = n\right\} \\ &\leq e^{-ln} \left(\sup_{a \in \mathcal{X}} \mathbb{Q}\{\exp(l^2 \mathbf{1}_{\{N > k(l)\}}) \mid a\}\right)^n \leq e^{-n(l - \log 2)}. \end{aligned}$$

Now choose $M > \alpha + \log 2$. Define the set

$$\Gamma_M = \left\{\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) : \int_{\{N > k(l)\}} d\nu < \frac{1}{l}, \text{ for all } l \geq M\right\}.$$

As $\{N \leq k(l)\} \subset \mathcal{X} \times \mathcal{X}^*$ is compact, the set Γ_M is pre-compact in the weak topology, by Prohorov's criterion. As

$$\mathbb{P}\{M_X \notin \Gamma_M \mid |T| = n\} \leq \frac{1}{\mathbb{P}\{|T| = n\}} \frac{1}{1 - e^{-1}} \exp(-n(M - \log 2)),$$

we can use [6, Lemma 3.1] to infer that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{M_X \notin K_\alpha \mid |T| = n\} \leq -\alpha,$$

for the closure K_α of Γ_M as required for the proof. \square

We denote by \mathcal{M}_s the set of all sub-consistent measures, and by \mathcal{M}_c the set of all consistent measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ and notice that $\mathcal{M}_c \subseteq \mathcal{M}_s$. For k a natural number, we denote by $\mathcal{M}_{c,k}$ the set of consistent measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. Then, \mathcal{M}_s is a closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ and $\mathcal{M}_{c,k}$ is a closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. The next two large deviation principles will help us extend LDP in $\mathcal{M}_{c,k}$, \mathcal{M}_s to $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ and $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ respectively.

Lemma 3.2. *Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q} . Assume (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in \mathcal{M}_s with convex, good rate function*

$$\tilde{J}(\varpi, \nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}) & \text{if } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.6)$$

Then, (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^)$ with convex, good rate function J .*

Proof. Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q} and that an LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds in \mathcal{M}_s , with convex, good rate function \tilde{J} . Then, we have $\{|T| = n\} := \{\omega : |T|(\omega) = n\} \subseteq \{\omega : (\tilde{L}_X, M_X)(\omega) \in \mathcal{M}_c\} =: \{(\tilde{L}_X, M_X) \in \mathcal{M}_c\}$. Hence, for all n , we have

$$\begin{aligned} \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_s \mid |T| = n\} &= \frac{1}{\mathbb{P}\{|T| = n\}} \times \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_s, |T| = n\} \\ &\geq \frac{1}{\mathbb{P}\{|T| = n\}} \times \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_c, |T| = n\} \\ &= \frac{1}{\mathbb{P}\{|T| = n\}} \times \mathbb{P}\{|T| = n\} \\ &= 1. \end{aligned}$$

Also, if $(\varpi_n, \nu_n) \in \mathcal{M}_s$ converges (ϖ, ν) then we have that

$$\varpi(a, b) = \lim_{n \rightarrow \infty} \varpi_n(a, b) \geq \lim_{n \rightarrow \infty} \sum_{c \in \mathcal{X}^*} m(b, c) \nu_n(a, c) \geq \liminf_{n \rightarrow \infty} \sum_{c \in \mathcal{X}^*} m(b, c) \nu_n(a, c) \geq \sum_{c \in \mathcal{X}^*} m(b, c) \nu(a, c),$$

which implies (ϖ, ν) is sub-consistent. This means \mathcal{M}_s is closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$. Therefore, by [7, Lemma 4.1.5], the LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds with convex, good rate function J . \square

Lemma 3.3. *Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q}_k . Assume (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in $\mathcal{M}_{c,k}$ with convex, good rate function*

$$\tilde{J}_k(\varpi, \nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}_k) & \text{if } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.7)$$

Then, (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^)$ with convex, good rate function J_k .*

Proof. Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q}_k and that an LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds in $\mathcal{M}_{c,k}$, with convex, good rate function \tilde{J}_k . Then, we have, $\{|T| = n\} \subseteq \{\omega : (\tilde{L}_X, M_X)(\omega) \in \mathcal{M}_{c,k}\} =: \{(\tilde{L}_X, M_X) \in \mathcal{M}_{c,k}\}$. Hence, for all n ,

$$\begin{aligned} \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_{c,k} \mid |T| = n\} &= \frac{1}{\mathbb{P}\{|T| = n\}} \times \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_{c,k}, |T| = n\} \\ &= \frac{1}{\mathbb{P}\{|T| = n\}} \times \mathbb{P}\{|T| = n\} \\ &= 1. \end{aligned}$$

Also, if $(\varpi_n, \nu_n) \in \mathcal{M}_{c,k}$ converges (ϖ, ν) then we have

$$\varpi(a, b) = \lim_{n \rightarrow \infty} \varpi_n(a, b) = \lim_{n \rightarrow \infty} \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu_n(a, c) = \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu(a, c),$$

which implies (ϖ, ν) is consistent. This means $\mathcal{M}_{c,k}$ is closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. Hence, by [7, Lemma 4.1.5], the LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds with convex, good rate function J_k which completes the proof of the Lemma. \square

In view of Lemmas 3.3 and 3.2, we establish large deviation principles in the spaces $\mathcal{M}_{c,k}$ and \mathcal{M}_s .

3.2 Proof of the upper bound in Theorem 2.1. Next we derive an upper bound in a variational formulation. Denote by \mathcal{C} the space of bounded functions on $\mathcal{X} \times \mathcal{X}^*$ and define for each (ϖ, ν) sub-consistent element in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$, the function \hat{J} by

$$\hat{J}(\varpi, \nu) = \sup_{g \in \mathcal{C}} \left\{ \int g(b, c) \nu(db, dc) - \int U_g(b) \varpi(da, db) \right\}, \quad (3.8)$$

where $c = (n, a_1, \dots, a_n)$. We recall that \mathcal{M}_s is the set of all sub-consistent measures.

Lemma 3.4. *For each closed set $F \subset \mathcal{M}_s$, we have*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in F \mid |T| = n\} \leq - \inf_{(\varpi, \nu) \in F} \hat{J}(\varpi, \nu).$$

Proof. Fix $\tilde{g} \in \mathcal{C}$ bounded by some $M > 0$, then also $\int e^{U_{\tilde{g}}(a)} \mu(da) \leq e^M$. We observe that, by (3.5),

$$\begin{aligned} e^M &\geq \tilde{\mathbb{P}}\{|T| = n\} \int e^{U_{\tilde{g}}(a)} \mu(da) = \mathbb{E}\left\{ \prod_{v \in V} \exp\left[\tilde{g}(X(v), C(v)) - \sum_{b \in \mathcal{X}} m(b, C(v)) U_{\tilde{g}}(b)\right] \mathbf{1}_{\{|T|=n\}} \right\} \\ &= \mathbb{E}\left\{ e^{n\langle \tilde{g}, M_X \rangle - n\langle U_{\tilde{g}}, \tilde{L}_X \rangle} \mathbf{1}_{\{|T|=n\}} \right\}. \end{aligned}$$

Together with [6, Lemma 3.1] this shows that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}\left\{ e^{n\langle \tilde{g}, M_X \rangle - n\langle U_{\tilde{g}}, \tilde{L}_X \rangle} \mid |T| = n \right\} \leq 0. \quad (3.9)$$

In view of (3.4) the same bound (3.9) holds for only M_X

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}\left\{ e^{n\langle \tilde{g}, M_X \rangle - n\langle U_{\tilde{g}}, M_X \rangle} \mid |T| = n \right\} \leq 0. \quad (3.10)$$

Now fix $\varepsilon > 0$, and let $\hat{J}_\varepsilon(\varpi, \nu) = \min\{\hat{J}(\varpi, \nu), \varepsilon^{-1}\} - \varepsilon$. Suppose first that $(\varpi, \nu) \in F$ is such that $\varpi_2 = \nu_1$.

Choose $\tilde{g}_{(\varpi, \nu)} \in \mathcal{C}$ such that

$$\int \tilde{g}_{(\varpi, \nu)}(a, c) \nu(a, c) - \int U_{\tilde{g}_{(\varpi, \nu)}}(b) \varpi(da, db) \geq \hat{J}_\varepsilon(\varpi, \nu). \quad (3.11)$$

Since $\tilde{g}_{(\varpi, \nu)}$ is bounded, the mapping $(\tilde{\varpi}, \tilde{\nu}) \mapsto \langle \tilde{g}_{(\varpi, \nu)}, \tilde{\nu} \rangle - \langle U_{\tilde{g}_{(\varpi, \nu)}}, \tilde{\varpi} \rangle$ is continuous in \mathcal{M}_s .

Hence there exists an open neighbourhood $B_{(\varpi, \nu)}$ of (ϖ, ν) such that

$$\inf_{(\tilde{\varpi}, \tilde{\nu}) \in B_{(\varpi, \nu)}} \left\{ \langle \tilde{g}_{(\varpi, \nu)}, \tilde{\nu} \rangle - \langle U_{\tilde{g}_{(\varpi, \nu)}}, \tilde{\varpi} \rangle \right\} \geq \left\{ \langle \tilde{g}_{(\varpi, \nu)}, \nu \rangle - \langle U_{\tilde{g}_{(\varpi, \nu)}}, \varpi \rangle \right\} - \varepsilon \geq \hat{J}_\varepsilon(\varpi, \nu) - \varepsilon.$$

Using the exponential Chebyshev inequality and the remark following (3.9) we obtain that,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in B_{(\varpi, \nu)} \mid |T| = n\} \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}\{e^{n\langle g, M_X \rangle - n\langle U_g, \tilde{L}_X \rangle} \mid |T| = n\} - \hat{J}_\varepsilon(\varpi, \nu) + \varepsilon \leq - \inf_{\nu \in F} \hat{J}_\varepsilon(\varpi, \nu) + \varepsilon. \end{aligned} \quad (3.12)$$

Now suppose that (ϖ, ν) is such that $\varpi_2 \neq \nu_1$. Assume first that there exists $a \in \mathcal{X}$ such that

$$\nu_1(a) < \varpi_2(a). \quad (3.13)$$

As the mappings $(\varpi, \nu) \mapsto \nu_1 - \varpi_2$ are continuous in the weak topology, there exist $\delta > 0$ and a small open neighbourhood $B_{(\varpi, \nu)} \subset \mathcal{M}_s$ such that

$$\tilde{\nu}_1(a) < \tilde{\varpi}(a) - \delta, \text{ for all } (\tilde{\varpi}, \tilde{\nu}) \in B_{(\varpi, \nu)}. \quad (3.14)$$

Let $\tilde{g} \in \mathcal{C}$ be defined by $\tilde{g}(b, c) = -(\delta\varepsilon)^{-1} \mathbf{1}_a(b)$. Note that, by the definition (3.1), we have $U_{\tilde{g}}(b) = \tilde{g}(b, c)$ for all b and this vanishes unless $b = a$. Hence, by (3.14), for every $(\tilde{\varpi}, \tilde{\nu}) \in B_{(\varpi, \nu)}$ we have

$$\int \tilde{g}(a, c) \nu(da, dc) - \int U_{\tilde{g}}(b) \varpi_2(db) > \varepsilon^{-1}.$$

Then, using the exponential Chebyshev inequality and (3.9), we have that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in B_{(\varpi, \nu)} \mid |T| = n\} \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}\{e^{n\langle g, M_X \rangle - n\langle U_g, \tilde{L}_X \rangle} \mid |T| = n\} - \varepsilon^{-1} \leq -\varepsilon^{-1} \leq - \inf_{(\varpi, \nu) \in F} \hat{J}_\varepsilon(\varpi, \nu). \end{aligned} \quad (3.15)$$

In case the opposite inequality holds in (3.13) the same argument leads to (3.15) if \tilde{g} is defined as

$$\tilde{g}(b, c) = (\delta\varepsilon)^{-1} \mathbf{1}_a(b).$$

Now we use Lemma 3.1 to choose a compact set K_α (for $\alpha = \varepsilon^{-1}$) with

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{M_X \notin K_\alpha \mid |T| = n\} \leq -\varepsilon^{-1}. \quad (3.16)$$

We write $F_\alpha := \{(\varpi, \nu) : (\varpi, \nu) \in \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times K_\alpha\}$ and notice that, the set $F_\alpha \cap F$ is compact and hence it may be covered by finitely many of the sets $B_{(\varpi_1, \nu_1)}, \dots, B_{(\varpi_m, \nu_m)}$, with $(\varpi_i, \nu_i) \in F$ for $i = 1, \dots, m$. Hence,

$$\mathbb{P}\{M_X \in F \mid |T| = n\} \leq \sum_{i=1}^m \mathbb{P}\{(\tilde{L}_X, M_X) \in B_{(\varpi_i, \nu_i)} \mid |T| = n\} + \mathbb{P}\{(\tilde{L}_X, M_X) \notin F_\alpha \mid |T| = n\}.$$

Using (3.12) and (3.15) we obtain, for small enough $\varepsilon > 0$, that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in F \mid |T| = n\} & \leq \max_{i=1}^m \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in B_{\nu_i} \mid |T| = n\} - \varepsilon^{-1} \\ & \leq - \inf_{(\varpi, \nu) \in F} \hat{J}_\varepsilon(\varpi, \nu) + \varepsilon. \end{aligned}$$

Taking $\varepsilon \downarrow 0$ gives the required statement. \square

Recall that $\tilde{J} : \mathcal{M}_s \rightarrow [0, \infty]$ is given by

$$\tilde{J}(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}) & \text{if } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.17)$$

We show that the convex rate function \tilde{J} may replace the function \hat{J} of (3.8) in the upper bound of Lemma 3.4.

Lemma 3.5. *The function \tilde{J} is convex and lower semicontinuous on \mathcal{M}_s . Moreover, $\tilde{J}(\varpi, \nu) \leq \hat{J}(\varpi, \nu)$, for any $(\varpi, \nu) \in \mathcal{M}_s$.*

Proof. We start by proving the inequality $\tilde{J}(\varpi, \nu) \leq \hat{J}(\varpi, \nu)$. To begin, suppose that $\varpi_2 \neq \nu_1$. Then, there exists $a_0 \in \mathcal{X}$ such that $\nu_1(a_0) \neq \varpi_2(a_0)$. For a_0 we define the function $\tilde{g}(b, c) = K\mathbf{1}_{a_0}(b)$ and observe that $U_{\tilde{g}}(b) = g(b, c)$. Using this \tilde{g} in (3.8) we obtain

$$\int \tilde{g}(a, c)\nu(da, dc) - \int U_{\tilde{g}}(b)\varpi_2(db) = K(\nu_1(a_0) - \varpi_2(a_0)) \longrightarrow \infty,$$

for $|K| \uparrow \infty$, with the sign of K chosen so that the right hand side is positive.

Next, suppose that $\varpi_2 = \nu_1$ but $\nu \not\ll \nu_1 \otimes \mathbb{Q}$. Then, there exists $(a', c') \in \mathcal{X} \times \mathcal{X}^*$ with $\nu(a', c') > 0$ and $\mathbb{Q}\{c' | a'\} = 0$. Consequently, recalling (3.1), we have $U_{\tilde{g}} = 0$ for $\tilde{g}(b, c) = R\mathbf{1}_{(a', c')}(b, c)$ and any R . Considering such \tilde{g} in (3.8) with $R \uparrow \infty$ we see that $\hat{J}(\varpi, \nu) = \infty$ in this case.

Finally suppose $\varpi_2 = \nu_1$ and $\nu \ll \nu_1 \otimes \mathbb{Q}$. By the variational characterisation of the relative entropy, see e.g. [7, Lemma 6.2.13], the definition of U_g , Jensen's inequality, and (??),

$$\begin{aligned} H(\nu \| \nu_1 \otimes \mathbb{Q}) &= \sup_{g \in \mathcal{C}} \left\{ \int g d\nu - \log \iint e^{g(a, c)} \mathbb{Q}\{dc | a\} \nu_1(da) \right\} \\ &= \sup_{g \in \mathcal{C}} \left\{ \int g d\nu - \log \int e^{U_g(a)} \nu_1(da) \right\} \\ &\leq \sup_{g \in \mathcal{C}} \left\{ \int g d\nu - \int U_g(a) \nu_1(da) \right\} = \hat{J}(\varpi, \nu). \end{aligned} \tag{3.18}$$

Now consider the convex, good rate function $\phi : \mathbb{R} \rightarrow [0, \infty]$ given by $\phi(x) = x \log x - x + 1$. Then, we can represent the left side of (3.18) in the form

$$H(\nu \| \nu_1 \otimes \mathbb{Q}) = \begin{cases} \int \phi \circ f d(\nu_1 \otimes \mathbb{Q}) & \text{if } f := \frac{d\nu}{d(\nu_1 \otimes \mathbb{Q})} \text{ exists} \\ \infty & \text{otherwise.} \end{cases} \tag{3.19}$$

Consequently, by [7, Lemma 6.2.16], \tilde{J} is a convex, good rate function. \square

By Lemma 3.2 the large deviation upper bound Lemma 3.4 holds with rate function \tilde{J} replaced by J .

3.3 Proof of Theorem 2.3. We begin by recalling from [6] that $\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ is shift-invariant if

$$\nu_1(a) = \sum_{(b, c) \in \mathcal{X} \times \mathcal{X}^*} m(a, c)\nu(b, c), \text{ for all } a \in \mathcal{X}.$$

This Theorem is derived from [6, Theorem 2.2] by applying the contraction principle to the linear mapping $G : \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*) \mapsto \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ given by $G(\nu) = (\varpi, \nu)$ where (ϖ, ν) is consistent. Specifically, [6, Theorem 2.2] implies the large deviation for $G(M_X) = (\tilde{L}, M_X)$ with convex, good rate function $\tilde{J}_k(\varpi, \nu) = \inf \{K_k(\nu) : G(\nu) = (\varpi, \nu), (\varpi, \nu) \text{ is consistent}\}$, where

$$K_k(\nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}_k) & \text{if } \langle m(\cdot, c), \nu(a, c) \rangle = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \tag{3.20}$$

Using shift-invariance and consistency we have

$$\nu_1(a) = \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}_k^*} m(a,c) \nu(b,c) = \sum_{b \in \mathcal{X}} \varpi(b,a) = \varpi_2(a), \text{ for all } a \in \mathcal{X}.$$

Therefore, by Lemma 3.3, the LDP for (\tilde{L}, M_X) conditional on the event $\{|T| = n\}$ holds in $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ with convex, good rate function J_k .

3.4 Proof of the Lower Bound in Theorem 2.1

We define for every weakly irreducible, critical offspring kernel $\mathbb{Q}\{c|b\}$ the conditional offspring law

$$\mathbb{Q}_k\{c|a\} = \begin{cases} \frac{1}{\mathbb{Q}\{\mathcal{X}_k^*|a\}} \mathbb{Q}\{c|a\} & \text{if } c \in \mathcal{X}_k^* \\ 0 & \text{otherwise,} \end{cases} \quad (3.21)$$

where $\mathbb{Q}\{\mathcal{X}_k^*|a\} = \sum_{c \in \mathcal{X}_k^*} \mathbb{Q}\{c|a\}$. We write $\|\nu\|_k = \nu(\mathcal{X} \times \mathcal{X}_k^*)$ and observe that

$$\lim_{k \rightarrow \infty} \nu(\mathcal{X} \times \mathcal{X}_k^*) = \lim_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \mathbb{1}_{\{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*\}} \nu(a,c) = 1,$$

by the dominated convergence. We define ν_k a probability measure on $\mathcal{X} \times \mathcal{X}^*$ by

$$\nu_k(a,c) = \begin{cases} \frac{\nu(a,c)}{\|\nu\|_k} & \text{if } (a,c) \in \mathcal{X} \times \mathcal{X}_k^* \\ 0 & \text{otherwise,} \end{cases} \quad (3.22)$$

and denote by $(\nu_k)_1$ the \mathcal{X} -marginal of the probability measure ν_k . Write

$$\varpi_k(a,b) := \sum_{c \in \mathcal{X}_k^*} m(b,c) \nu_k(a,c).$$

Lemma 3.6. *Let $\nu \ll \nu_1 \otimes \mathbb{Q}$. Then, we have*

$$\lim_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \frac{\nu(a,c)}{\|\nu\|_k} \log \frac{\frac{\nu(a,c)}{\|\nu\|_k}}{\left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k(a,c)} = \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a,c) \log \frac{\nu(a,c)}{\nu_1 \otimes \mathbb{Q}(a,c)}. \quad (3.23)$$

Proof. Recall that we have assumed $\nu \ll \nu_1 \otimes \mathbb{Q}$. Then we have $\frac{\nu}{\|\nu\|_k} \ll \left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k$ and so, we may write

$$\sum_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \frac{\nu(a,c)}{\|\nu\|_k} \log \frac{\frac{\nu(a,c)}{\|\nu\|_k}}{\left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k(a,c)} = \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \mathbb{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a,c)}{\|\nu\|_k} \mathbb{1}_{\{\nu(a,c) > 0\}} \log \frac{\frac{\nu(a,c)}{\|\nu\|_k}}{\left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k(a,c)}.$$

We write $f_k(a,c) := \mathbb{1}_{\{\nu(a,c) > 0\}} \log \frac{\frac{\nu(a,c)}{\|\nu\|_k}}{\left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k(a,c)}$, for $(a,c) \in \mathcal{X} \times \mathcal{X}^*$ and notice that

$$\lim_{n \rightarrow \infty} f_k(a,c) = \mathbb{1}_{\{\nu(a,c) > 0\}} \log \frac{\nu(a,c)}{\nu_1 \otimes \mathbb{Q}(a,c)} := f(a,c)$$

is bounded function. Hence, for any $\delta > 0$ and sufficiently large k we have that

$$\sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \mathbb{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a,c)}{\|\nu\|_k} (f(a,c) - \delta) \leq \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \mathbb{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a,c)}{\|\nu\|_k} f_k(a,c) \leq \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \mathbb{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a,c)}{\|\nu\|_k} (f(a,c) + \delta). \quad (3.24)$$

Now we take limit as k approaches ∞ of all sides of (3.24) to obtain

$$\sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a,c)(f(a,c) - \delta) \leq \lim_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \mathbf{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a,c)}{\|\nu\|_k} f_k(a,c) \leq \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a,c)(f(a,c) + \delta)$$

Taking $\delta \downarrow 0$, yields

$$\begin{aligned} \lim_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \mathbf{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a,c)}{\|\nu\|_k} f_k(a,c) &= \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a,c) f(a,c) = \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \mathbf{1}_{\{\nu(a,c) > 0\}} \nu(a,c) \log \frac{\nu(a,c)}{\nu_1 \otimes \mathbb{Q}(a,c)} \\ &= \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a,c) \log \frac{\nu(a,c)}{\nu_1 \otimes \mathbb{Q}(a,c)} \end{aligned}$$

by the sandwich theorem of limits in the first equality, and the conventions $0 \log \frac{0}{0} = 0$, $0 \log 0 = 0$. \square

We define the total variation metric d by

$$d(\nu, \tilde{\nu}) = \frac{1}{2} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} |\nu(a,c) - \tilde{\nu}(a,c)| \quad (3.25)$$

and observe that it generates the weak topology.

By $\mathcal{N}(\mathcal{X})$ we denote the space of counting measures on \mathcal{X} . For \mathbb{Q}_k we recall the definition of the rate function $J_k : \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*) \rightarrow [0, \infty]$ from Theorem 2.3 as

$$J_k(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}_k) & \text{if } (\varpi, \nu) \text{ is consistent and } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases}$$

The next lemma which is a key ingredient in our proof will be proved in two approximation steps.

Lemma 3.7. *Suppose (ϖ, ν) is sub-consistent, $\varpi_2 = \nu_1$ and $\nu \ll \nu_1 \otimes \mathbb{Q}$. Then, for every $\varepsilon > 0$, there exists $(\hat{\varpi}, \hat{\nu}) \in \mathcal{M}_{c,k}$ such that $|\varpi(a,b) - \hat{\varpi}(a,b)| < \varepsilon$, for all $a, b \in \mathcal{X}$, $d(\nu, \hat{\nu}) \leq \varepsilon$, $\hat{\varpi}_2 = \hat{\nu}_1$ and*

$$|J_k(\hat{\varpi}, \hat{\nu}) - J(\varpi, \nu)| \leq \varepsilon.$$

Proof. Step 1: Approximating sub-consistent (ϖ, ν) by a sequence of consistent $(\hat{\varpi}_n, \nu_n)$. Recall our assumption that (ϖ, ν) is sub-consistent, $\varpi_2 = \nu_1$ and $\nu \ll \nu_1 \otimes \mathbb{Q}$. For any $b \in \mathcal{X}$ we define $e^{(b)} \in \mathcal{N}(\mathcal{X})$ by $e^{(b)}(a) = 0$ if $a \neq b$, and $e^{(b)}(a) = 1$ if $a = b$. We write $n(c) = (m(a,c), a \in \mathcal{X})$ and for large n define $\hat{\nu}_n \in \mathcal{M}_s$ by

$$\hat{\nu}_n(a,c) = \nu(a,c) \left(1 - \frac{\|\varpi\| - \|\langle m(\cdot,c), \nu(\cdot,c) \rangle\|}{n} \right) + \sum_{b \in \mathcal{X}} \mathbf{1}\{n(c) = ne^{(b)}\} \frac{\varpi(a,b) - \langle m(\cdot,c), \nu(\cdot,c) \rangle(a,b)}{n}.$$

We note that $\hat{\nu}_n \rightarrow \nu$ and that, for all $a, b \in \mathcal{X}$,

$$\begin{aligned} \sum_{c \in \mathcal{X}^*} m(a,c) \hat{\nu}_n(b,c) &= \left(1 - \frac{\|\varpi\| - \|\langle m(\cdot,c), \nu(\cdot,c) \rangle\|}{n} \right) \sum_{c \in \mathcal{X}^*} m(a,c) \nu(b,c) + \varpi(a,b) - \langle m(\cdot,c), \nu(\cdot,c) \rangle(a,b) \\ &= \varpi(a,b) - \frac{\|\varpi\| - \|\langle m(\cdot,c), \nu(\cdot,c) \rangle\|}{n} \langle m(\cdot,c), \nu(\cdot,c) \rangle(a,b) \xrightarrow{n \uparrow \infty} \varpi(a,b). \end{aligned}$$

Defining $\hat{\varpi}_n$ by $\hat{\varpi}_n(a,b) = \sum_{c \in \mathcal{X}^*} m(a,c) \hat{\nu}_n(b,c)$, we have a sequence of consistent pairs $(\hat{\varpi}_n, \nu_n)$ converging to (ϖ, ν) , $\nu_n \ll (\nu_n)_1 \otimes \mathbb{Q}$ and $(\hat{\varpi}_n)_2 = (\hat{\nu}_n)_1$, as if $(\hat{\varpi}_n)_2(a) \neq (\hat{\nu}_n)_1(a)$ for some $a \in \mathcal{X}$, then we have

$$\hat{\varpi}_2(a) = \lim_{n \rightarrow \infty} (\hat{\varpi}_n)_2(a) \neq \lim_{n \rightarrow \infty} (\hat{\nu}_n)_1(a) = \nu_1(a).$$

Further, by the continuity of relative entropy, we have that

$$\lim_{n \rightarrow \infty} J(\hat{\varpi}_n, \hat{\nu}_n) = \sum_{a \in \mathcal{X}} \hat{\nu}_n(a) 1_{\{\hat{\nu}_n(a) > 0\}} H(\hat{\nu}_n(\cdot|a) \| \mathbb{Q}) = J(\varpi, \nu), \quad (3.26)$$

where $0 \log \frac{0}{0} = 0$ and $0 \log 0 = 0$ by convection.

Step 2: Approximating sub-consistent (ϖ_n, ν_n) by a sequence $(\varpi_{k,n}, \nu_{k,n}) \in \mathcal{M}_{c,k}$. Let $\varepsilon > 0$ and set $\delta = 2\varepsilon$. Choose $k(\varepsilon)$ large enough such that

$$\sup_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \left| \frac{1_{\{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*\}}}{\|\hat{\nu}_n\|_k} - 1 \right| \leq \delta.$$

For $k \geq k(\varepsilon)$, define $\nu_{k,n}$ a probability measure on $\mathcal{X} \times \mathcal{X}^*$ by

$$\nu_{k,n}(a, c) = \begin{cases} \frac{\hat{\nu}_n(a, c)}{\|\hat{\nu}_n\|_k} & \text{if } (a, c) \in \mathcal{X} \times \mathcal{X}_k^* \\ 0 & \text{otherwise,} \end{cases} \quad (3.27)$$

and denote by $(\nu_{k,n})_1$ the \mathcal{X} -marginal of the probability measure $\nu_{k,n}$. Write

$$\varpi_{k,n}(a, b) := \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu_{k,n}(a, c).$$

We note that

$$\begin{aligned} d(\nu_{k,n}, \hat{\nu}_n) &= \frac{1}{2} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} |\nu_{k,n}(a, c) - \hat{\nu}_n(a, c)| = \frac{1}{2} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \left| \frac{1_{\{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*\}}}{\|\hat{\nu}_n\|_k} \hat{\nu}_n(a, c) - \hat{\nu}_n(a, c) \right| \leq \frac{1}{2} \delta \\ &= \varepsilon \end{aligned}$$

Define for each $n > 1$ the measure $\tilde{\varpi}_n$ by $\tilde{\varpi}_n(a, b) = (1 + \frac{1}{n})\varpi(a, b)$, for $a, b \in \mathcal{X}$. Then, it is not hard to see that

$$\lim_{n \rightarrow \infty} \tilde{\varpi}_n(a, b) = \varpi(a, b), \quad \text{for all } a, b \in \mathcal{X}.$$

Also, observe that

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \varpi_{k,n}(a, b) \geq \lim_{n \rightarrow \infty} \liminf_{k \rightarrow \infty} \varpi_{k,n}(a, b) \geq \lim_{n \rightarrow \infty} \varpi_n(a, b) = \varpi(a, b), \quad \text{for all } a, b \in \mathcal{X}.$$

Hence, there exists $n(\varepsilon)$ such that $n \geq n(\varepsilon)$ implies $\varpi_{k,n}(a, b) \geq \varpi(a, b) - \frac{\varepsilon}{2}$. Hence, for large $k \geq k(\varepsilon)$ and large $n \geq n(\varepsilon)$ we have that

$$|\tilde{\varpi}_n(a, b) - \varpi_{k,n}(a, b)| \leq |\tilde{\varpi}_n(a, b) - \varpi(a, b)| + |\varpi(a, b) - \varpi_{k,n}(a, b)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

by the triangle inequality. This gives

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \varpi_{k,n}(a, b) = \lim_{n \rightarrow \infty} \tilde{\varpi}_n(a, b) = \varpi(a, b), \quad \text{for all } a, b \in \mathcal{X}.$$

Using similar argument as above we obtain

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} (\varpi_{k,n})_2(b) = \lim_{n \rightarrow \infty} (\tilde{\varpi}_n)_2(b) = (\varpi)_2(b), \quad \text{for all } b \in \mathcal{X}.$$

Observe that, by construction $(\varpi_{k,n}, \nu_{k,n})$ is consistent, $\nu_{k,n} \ll (\nu_{k,n})_1 \otimes \mathbb{Q}_k$ and $(\hat{\varpi}_{k,n})_2 = (\hat{\nu}_{k,n})_1$, as if $(\nu_{k,n})_1(a) \neq (\varpi_{k,n})_2(a)$ for some $a \in \mathcal{X}$, then we have that

$$\varpi_2(a) = \lim_{n \rightarrow \infty} \varpi_n)_2(a) = \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} (\varpi_{k,n})_2(a) \neq \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} (\nu_{k,n})_1(a) = \lim_{n \rightarrow \infty} (\nu_n)_1(a) = \nu_1(a).$$

Now we recall from (3.27) the definition of $\nu_{k,n}$ and notice that we have,

$$\sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu_{k,n}(a,c) \log \frac{\nu_{k,n}(a,c)}{(\nu_{k,n})_1 \otimes \mathbb{Q}_k(a,c)} = \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \frac{\hat{\nu}_n(a,c)}{\|\hat{\nu}_n\|_k} \log \frac{\hat{\nu}_n(a,c)}{\left(\frac{\hat{\nu}_n}{\|\hat{\nu}_n\|_k}\right)_1 \otimes \mathbb{Q}_k(a,c)},$$

where $0 \log \frac{0}{0} = 0$ and $0 \log 0 = 0$ by convection. Using Lemma 3.6 and (3.26) we obtain

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} J_k(\varpi_{k,n}, \nu_{k,n}) = \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu_{k,n}(a,c) \log \frac{\nu_{k,n}(a,c)}{(\nu_{k,n})_1 \otimes \mathbb{Q}_k(a,c)} = \lim_{n \rightarrow \infty} J(\hat{\varpi}_n, \hat{\nu}_n) = J(\varpi, \nu),$$

which completes the proof of the lemma. \square

We recall that $C(v) = (N(v), X_1(v), \dots, X_{N(v)}(v))$ and note that, for every k such that $\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} > 0$ and any tree-indexed process x , we have that

$$\begin{aligned} \mathbb{P}\{X = x \mid |T| = n\} &\geq \mathbb{P}\{(X = x, C(v) \in \mathcal{X}_k^*, v \in V \mid |T| = n)\} \\ &= \prod_{v \in V, |T|=n} \mathbb{Q}\{\mathcal{X}_k^* | x(v)\} \times \mathbb{P}_k\{X = x \mid |T| = n\} \\ &\geq \left(\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\}\right)^n \times \mathbb{P}_k\{X = x \mid |T| = n\}, \end{aligned} \quad (3.28)$$

where \mathbb{P}_k denote the law of tree-indexed process with initial distribution μ and offspring kernel \mathbb{Q}_k , and

$$\lim_{k \rightarrow \infty} \min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} = \lim_{k \rightarrow \infty} \min_{a \in \mathcal{X}} \sum_{c \in \mathcal{X}_k} \mathbb{Q}\{c | a\} = \lim_{k \rightarrow \infty} \sum_{c \in \mathcal{X}^*} 1_{\{c \in \mathcal{X}_k\}} \min_{a \in \mathcal{X}} \mathbb{Q}\{c | a\} = 1.$$

To complete the proof of the lower bound, we take $O \subset \mathcal{M}_g$. Then, for any $(\varpi, \nu) \in O$ sub-consistent with $\varpi_2 = \nu_1$, $\nu \ll \nu_1 \otimes \mathbb{Q}$ we may find $\varepsilon > 0$ with ball around (ϖ, ν) of radius 2ε contained in O . By our approximation Lemma 3.7, we may find $(\varpi_{k,n}, \nu_{k,n}) \in O \cap \mathcal{M}_{c,k}$ with $|\varpi_{k,n}(a,b) - \varpi(a,b)| \downarrow 0$, $d(\nu_{k,n}, \nu) \downarrow 0$, $(\varpi_{k,n})_2 = (\nu_{k,n})_1$ and

$$|J_k(\varpi_{k,n}, \nu_{k,n}) - J(\varpi, \nu)| \downarrow 0.$$

Hence, using the lower bound of Theorem 2.3 for offspring kernel \mathbb{Q}_k given by (3.21), (3.28) for large $k \geq k(\varepsilon)$ (with $\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} > 0$) and for large $n \geq n(\varepsilon)$, we obtain

$$\begin{aligned} \mathbb{P}\{(\tilde{L}_X, M_X) \in O \mid |T| = n\} &\geq \mathbb{P}\{|\varpi_{k,n}(a,b) - \tilde{L}_X(a,b)| < \varepsilon, \forall a, b \in \mathcal{X}, d(\nu_{k,n}, M_X) < \varepsilon \mid |T| = n\} \\ &\geq e^{n\alpha_k} \times \mathbb{P}_k\{|\varpi_{k,n}(a,b) - \tilde{L}_X(a,b)| < \varepsilon, \forall a, b \in \mathcal{X}, \\ &\hspace{15em} d(\nu_{k,n}, M_X) < \varepsilon \mid |T| = n\} \\ &\geq \exp(-n(J_k(\varpi_{k,n}, \nu_{k,n}) + \varepsilon - \alpha_k)) \end{aligned}$$

where $\alpha_k = \log(\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\})$. Taking limits we have that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in O \mid |T| = n\} \geq -\limsup_{n \rightarrow \infty} \lim_{k \rightarrow \infty} J_k(\varpi_{k,n}, \nu_{k,n}) - \varepsilon \geq -J(\varpi, \nu) - \varepsilon.$$

Taking $\varepsilon \downarrow 0$ we have have the desired result which completes the proof of the lower bound.

4. PROOF OF COROLLARIES 2.2, 2.5 AND THEOREM 2.4

4.1 Proof of Corollary 2.2. We derive this corollary from Theorem 2.1 by applying the contraction principle to the linear mapping $W : \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) \mapsto \mathbb{R}^{\mathcal{X} \times \mathcal{X}^*}$ defined by

$$W(\varpi, \nu)(a, c) = \nu(a, c), \quad \text{for all } (a, c) \in \mathcal{X} \times \mathcal{X}^*.$$

In fact Theorem 2.1 implies the large deviation principle for $W(\tilde{L}_X, M_X)$ with convex, good rate function $\hat{J}(\nu) = \inf \{J(\varpi, \nu) : W(\varpi, \nu) = \nu\}$. Now, using sub-consistency and $\varpi_2 = \nu_1$ we obtain the form $\hat{J}(\nu) = H(\nu \| \nu_1 \otimes \mathbb{Q})$, for ν satisfying $\langle m(\cdot, c), \nu(a, c) \rangle \leq \nu_1$. Recall the definition of shift-invariant and denote by \mathcal{M}_1 set of shift-invariant measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$. Write

$$\mathcal{M}_2 = \left\{ \nu : \nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X}^*), m(\cdot, c), \nu(a, c) \leq \nu_1 \right\}$$

and note that $\mathcal{M}_1 \subseteq \mathcal{M}_2$. Also, for all (values of) n where $\mathbb{P}\{|T| = n\} > 0$, we have

$$\mathbb{P}\{M_X \in \mathcal{M}_2 \mid |T| = n\} = 1.$$

Moreover, if $\nu_n \in \mathcal{M}_2$ converges to ν then

$$\nu_1(a) = \lim_{n \rightarrow \infty} (\nu_n)_1(a) \geq \liminf_{n \rightarrow \infty} \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}^*} m(a, c) \nu_n(b, c) \geq \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}^*} m(a, c) \nu(b, c),$$

which implies ν is sub-consistent. This means \mathcal{M}_2 is closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$. Therefore, by [7, Lemma 4.1.5], the LDP for M_X conditional on the event $\{|T| = n\}$ holds with convex, good rate function K , which completes the proof of the corollary.

4.2 Proof of Theorem 2.4. We begin the proof of the theorem by stating the following Lemma.

Lemma 4.1. *Suppose that $q(c) = p(n) \prod_{i=1}^n \hat{q}(a_i)$ for all $c = (n, a_1, \dots, a_n)$, where $\hat{q}(\cdot)$ is a probability vector on \mathcal{X} and $p(\cdot)$ a probability measure with mean one on the nonnegative integers. Then, we have*

$$\inf \left\{ H(\tilde{\nu} \| q) : \tilde{\nu} \in \mathcal{M}(\mathcal{X}^*), \phi(b) = \sum_{c \in \mathcal{X}^*} m(b, c) \tilde{\nu}(c) \text{ for all } b \in \mathcal{X} \right\} = zH(\phi/z \| \hat{q}) + I_p(z), \quad (4.1)$$

where $z = \sum_b \phi(b)$.

Proof. For $\tilde{\nu} \in \mathcal{M}(\mathcal{X}^*)$, we let $\phi(b) = \sum_{c \in \mathcal{X}^*} m(b, c) \tilde{\nu}(c)$, for all $b \in \mathcal{X}$ and suppose first that $z = 0$, i.e. $\phi(b) = 0$ for all $b \in \mathcal{X}$. Then, $\tilde{\nu}((0, \emptyset)) = 1$ is the only possible measure in left side of (4.1), leading to $\tilde{I}(\phi, q) = -\log q((0, \emptyset)) = -\log p(0)$. It follows from (2.8) that $I_p(0) = -\log p(0)$ establishing (4.1) for such $\phi(\cdot)$. We assume hereafter that $z > 0$. Now the possible measures $\tilde{\nu}(\cdot)$ in the left side of (4.1) are of the form $\tilde{\nu}(c) = s(n)v_n(a_1, \dots, a_n)$ for $c = (n, a_1, \dots, a_n)$, with $v_0 = 1$, where $s(\cdot)$ is a probability measure on the nonnegative integers whose mean is z , and $v_n(\cdot)$, $n \geq 1$, are probability measures on \mathcal{X}^n with marginals $v_{n,i}(\cdot)$ such that

$$\phi(b) = \sum_{n=1}^{\infty} s(n) \sum_{i=1}^n v_{n,i}(b) \quad \text{for all } b \in \mathcal{X}. \quad (4.2)$$

By the assumed structure of $q(\cdot)$ we have for such $\tilde{\nu}(\cdot)$ that

$$H(\tilde{\nu} \| q) = \sum_{n=1}^{\infty} s(n) H(v_n \| \hat{q}^n) + H(s \| p), \quad (4.3)$$

where \widehat{q}^n denotes the product measure on \mathcal{X}^n with equal marginals \widehat{q} . Recall that

$$\sum_{n=1}^{\infty} s(n) H(v_n \parallel \widehat{q}^n) \geq \sum_{n=1}^{\infty} s(n) \sum_{i=1}^n H(v_{n,i} \parallel \widehat{q}) \geq z H\left(z^{-1} \sum_{n=1}^{\infty} s(n) \sum_{i=1}^n v_{n,i} \parallel \widehat{q}\right),$$

with equality whenever $v_n = \prod_{i=1}^n v_{n,i}$ and $v_{n,i}$ are independent of n and i (see [7, Lemma 7.3.25] for the first inequality, with the second inequality following by convexity of $H(\cdot \parallel \widehat{q})$ and the fact that $\sum_n s(n)n = z$). So, in view of (4.2),

$$H(\widetilde{\nu} \parallel q) \geq z H(\phi/z \parallel \widehat{q}) + H(s \parallel p), \quad (4.4)$$

with equality when

$v_n = (z^{-1}\phi)^n$ for all $n \geq 1$. Now, write $\Lambda_p(\lambda) := \log \sum_n e^{\lambda n} p(n)$ and notice that Λ convex function and $\Lambda(0) = 0 < \infty$, and so, we have, for every $\lambda \in \mathbb{R}$, $\Lambda_p(\lambda) > -\infty$. Using Jensen's inequality, for every $s \in \mathcal{M}(\mathbb{N} \cup \{0\})$ and every $\lambda \in \mathbb{R}$, we have

$$\Lambda_p(\lambda) = \log \sum_n s(n) \left(\frac{e^{\lambda n} p(n)}{s(n)} \right) \geq \sum_n s(n) \log \left(\frac{e^{\lambda n} p(n)}{s(n)} \right) = \lambda \sum_n n s(n) - H(s \parallel p),$$

with equality if $s_\lambda(n) = p(n)e^{\lambda n - \Lambda(\lambda)}$. Thus, for all λ and all z , we have

$$\lambda z - \Lambda(\lambda) \leq \inf \{ H(s \parallel p) : s \in \mathcal{M}(\mathbb{N} \cup \{0\}) \text{ and } \sum_n s(n)n = z \} := \Lambda^*(z), \quad (4.5)$$

with equality when $\sum_n s(n)n = z$. Elementary calculus also shows that

$$\Lambda^*(z) = \lambda_* z - \Lambda_p(\lambda_*), \quad (4.6)$$

where λ_* is the solution of $\Lambda'_p(\lambda_*) = z$ and $\frac{d\Lambda_p}{d\lambda} := \Lambda'_p(\lambda)$. Combining (4.6) and (4.5) we obtain

$$\sup_{\lambda \in \mathbb{R}} \{ \lambda z - \Lambda_p(\lambda) \} \leq \Lambda^*(z) \leq \sup_{\lambda \in \mathbb{R}, \Lambda'_p(\lambda)=z} \{ \lambda z - \Lambda(\lambda) \} \leq \sup_{\lambda \in \mathbb{R}} \{ \lambda z - \Lambda_p(\lambda) \}.$$

This yields $\Lambda^*(z) = I_p(z)$, which ends the proof of the Lemma. \square

Next, note that X is an irreducible, critical multitype Galton-Watson tree with offspring law

$$\mathbb{Q}\{c | b\} = p(n) \prod_{i=1}^n Q\{a_i | b\}, \text{ for } c = (n, a_1, \dots, a_n). \quad (4.7)$$

We derive Theorem 2.4 from Theorems 2.1 and 2.3 by applying the contraction principle to the continuous linear mapping $F : \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) \rightarrow \mathbb{R}^{\mathcal{X} \times \mathcal{X}}$, defined by

$$F(\varpi, \nu)(a, b) = \varpi(a, b), \text{ for all } (\varpi, \nu) \in \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) \text{ and } a, b \in \mathcal{X}. \quad (4.8)$$

It is easy to see that on $\{|T| = n\}$ we have $L_X = \frac{n}{n-1} F(\widetilde{L}_X, M_X) = \frac{n}{n-1} \widetilde{L}_X$. It follows that conditioned on $\{|T| = n\}$ the random variables L_X are exponentially equivalent to \widetilde{L}_X , hence L_X satisfy the same large deviation principle as $F(\widetilde{L}_X, M_X)$, see [7, Theorem 4.2.13]. Without loss of generality we restrict the space for the large deviation principle of L_X to the set of all probability vectors on $\mathcal{X} \times \mathcal{X}$, see [7, Lemma 4.1.5(b)]. Specifically, we consider the cases where $p(n)$ has (i) Bounded support and (ii) Unbounded support and finite second moment, separately.

Bounded support. Suppose $k < \infty$ and p has support $\{1, 2, \dots, k\}$. Then, Theorem 2.3 implies the large deviation principle for $F(\widetilde{L}_X, M_X)$ conditioned on $\{|T| = n\}$ with the good rate function $I(\mu) = \inf \{ J_k(\mu, \nu) : F(\mu, \nu) = \mu \}$, see for example [7, Theorem 4.2.1]. Convexity of I follows easily from the linearity of F and convexity of J_k .

To prove (2.10), we recall that ν is consistent if and only if $F(\mu, \nu)(a, b) = \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu(a, c)$ for all $a, b \in \mathcal{X}$. Hence, we have that

$$I(\mu) = \inf \left\{ H(\nu \| \nu_1 \otimes \mathbb{Q}) : F(\varpi, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}_k^*} m(\cdot, c) \nu(\cdot, c), \nu_1 = \mu_2 \right\}. \quad (4.9)$$

Note that $\nu_1(a) = 0$ yields $\sum_b F(\mu, \nu)(a, b) = 0$. Hence if $\mu_1(a) > 0 = \mu_2(a)$ for some $a \in \mathcal{X}$ then $\left\{ \nu : F(\mu, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}_k^*} m(\cdot, c) \nu(\cdot, c), \nu_1 = \mu_2 \right\}$ is an empty set, and therefore $I(\mu) = \infty$. We assume hereafter that $\mu_1 \ll \mu_2$. Then, it is not uneasy to verify that

$$I(\mu) = \sum_{a \in \mathcal{X}} \mu_2(a) \widehat{I} \left(\frac{\mu(a, \cdot)}{\mu_2(a)}, \mathbb{Q}\{\cdot | a\} \right), \quad (4.10)$$

where for $q \in \mathcal{M}(\mathcal{X}_k^*)$, $\widehat{I} \left(\frac{\mu(a, \cdot)}{\mu_2(a)}, q \right) = \inf \left\{ \widetilde{I}(\phi, q) : \phi : \mathcal{X} \rightarrow \mathbb{R}_+, \phi(a) = \frac{\mu(a, \cdot)}{\mu_2(a)}, \text{ for all } a \in \mathcal{X} \right\}$ and

$$\widetilde{I}(\phi, q) := \inf \left\{ H(\tilde{\nu} \| q) : \tilde{\nu} \in \mathcal{M}(\mathcal{X}_k^*), \phi(b) = \sum_{c \in \mathcal{X}_k^*} m(b, c) \tilde{\nu}(c) \text{ for all } b \in \mathcal{X} \right\}. \quad (4.11)$$

Suppose now that $q(c) = p(n) \prod_{i=1}^n \widehat{q}(a_i)$ for all $c = (n, a_1, \dots, a_n)$, where $\widehat{q}(\cdot)$ is a probability vector on \mathcal{X} and $p(\cdot)$ a probability measure with mean one on $\{0, 1, 2, \dots, k\}$. Then, by Lemma 4.1, we have the representation

$$\widetilde{I}(\phi, q) = \|\phi\| H(\phi / \|\phi\| \| \widehat{q}) + I_p(\|\phi\|), \quad (4.12)$$

where $\|\phi\| := \sum_{b \in \mathcal{X}} \phi(b)$. Writing $\phi(b) = \frac{\mu(a, b)}{\mu_2(a)}$ in 4.12 we obtain (2.10) which proves the theorem in the case of p with bounded support.

Unbounded support and finite second moment. Suppose p has unbounded support and finite second moment. Then, Theorem 2.1 implies the large deviation principle for $F(\tilde{L}_X, M_X)$ conditioned on $\{|T| = n\}$ with the good rate function $I(\mu) = \inf \{ J(\mu, \nu) : F(\mu, \nu) = \mu \}$, see for example [7, Theorem 4.2.1]. Convexity of I follows easily from the linearity of F and convexity of J .

Turning to the proof of (2.10), recall that ν is sub-consistent if and only if $F(\varpi, \nu)(a, b) \geq \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu(a, c)$ for all $a, b \in \mathcal{X}$. Hence, we have that

$$I(\mu, \nu) = \inf \left\{ H(\nu \| \nu_1 \otimes \mathbb{Q}) : F(\mu, \nu)(\cdot, \cdot) \geq \sum_{c \in \mathcal{X}_k^*} m(\cdot, c) \nu(\cdot, c), \nu_1 = \mu_2 \right\}. \quad (4.13)$$

Note that $\nu_1(a) = 0$ yields $\sum_b F(\nu)(a, b) = 0$ if $F(\mu, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}_k^*} m(\cdot, c) \nu(\cdot, c)$ and

$$\sum_{(b, c) \in \mathcal{X}^*} m(b, c) \nu(a, c) < 0 \text{ if } F(\mu, \nu)(\cdot, \cdot) > \sum_{c \in \mathcal{X}_k^*} m(\cdot, c) \nu(\cdot, c).$$

Hence if $\mu_1(a) > 0 = \mu_2(a)$ for some $a \in \mathcal{X}$ then $\left\{ \nu : F(\mu, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}_k^*} m(\cdot, c) \nu(\cdot, c), \nu_1 = \mu_2 \right\} \cup \left\{ \nu : \right.$

$F(\varpi, \nu)(\cdot, \cdot) > \sum_{c \in \mathcal{X}_k^*} m(\cdot, c) \nu(\cdot, c), \nu_1 = \mu_2 \left. \right\}$ is an empty set, and therefore $I(\mu) = \infty$. Assuming,

throughout the rest of the proof that $\mu_1 \ll \mu_2$, it is not uneasy to verify that

$$I(\mu) = \sum_{a \in \mathcal{X}} \mu_2(a) \widehat{I} \left(\frac{\mu(a, \cdot)}{\mu_2(a)}, \mathbb{Q}\{\cdot | a\} \right), \quad (4.14)$$

where for $q \in \mathcal{M}(\mathcal{X}^*)$, $\widehat{I}\left(\frac{\mu(a, \cdot)}{\mu_2(a)}, q\right) = \inf \left\{ \widetilde{I}(\phi, q) : \phi : \mathcal{X} \rightarrow \mathbb{R}_+, \phi(a) \leq \frac{\mu(a, \cdot)}{\mu_2(a)}, \text{ for all } a \in \mathcal{X} \right\}$ and

$$\widetilde{I}(\phi, q) := \inf \left\{ H(\widetilde{\nu} \| q) : \widetilde{\nu} \in \mathcal{M}(\mathcal{X}^*), \phi(b) = \sum_{c \in \mathcal{X}^*} m(b, c) \widetilde{\nu}(c) \text{ for all } b \in \mathcal{X} \right\}. \quad (4.15)$$

Suppose now that $q(c) = p(n) \prod_{i=1}^n \widehat{q}(a_i)$ for all $c = (n, a_1, \dots, a_n)$, where $\widehat{q}(\cdot)$ is a probability vector on \mathcal{X} and $p(\cdot)$ a probability measure with mean one on the nonnegative integers, whose second moment is finite. Then, by Lemma 4.1, we have the representation

$$\widetilde{I}(\phi, q) = \|\phi\| H(\phi / \|\phi\| \| \widehat{q}) + I_p(\|\phi\|), \quad (4.16)$$

where $\|\phi\| := \sum_{b \in \mathcal{X}} \phi(b)$. Therefore, it suffice for us to show that

$$\inf \left\{ \widetilde{I}(\phi, \widehat{q}) : \phi : \mathcal{X} \rightarrow \mathbb{R}_+, \phi(b) \leq \frac{\mu(a, b)}{\mu_2(a)}, \text{ for all } a \in \mathcal{X} \right\} = \frac{\mu_1(a)}{\mu_2(a)} H\left(\frac{\mu(a, \cdot)}{\mu_1(a)} \| \widehat{q}\right) + I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right). \quad (4.17)$$

To do this, we write

$$h(\phi(b)) := \widetilde{I}(\phi, \widehat{q}) + \alpha(b) \left(\phi(b) - \frac{\mu(a, b)}{\mu_2(a)} \right), \text{ for } b \in \mathcal{X},$$

where α is a lagrange multiplier. Then, elementary calculus shows that $\alpha(b)$ is the solution of the equation

$$I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right) - \frac{\mu(a, b)}{\mu_2(a)} \sum_{a \in \mathcal{X}} e^{-\alpha(a)} \widehat{q}(a) = 0$$

and that $\phi(b) = \frac{\mu(a, b)}{\mu_2(a)}$ is the minimizer of our constraint optimization problem. Writing $\phi(b) = \frac{\mu(a, b)}{\mu_2(a)}$ in 4.16 we obtain left side of (4.17) which proves the theorem in case of p with unbounded support and finite second moments.

4.3 Proof of Corollary 2.5. Recall that T is Galton-Watson tree with offspring law $p(\ell) = 2^{-(\ell+1)}$, $\ell = 0, \dots$. Also, we recall that X is markov chain indexed by T with arbitrary initial distribution and transition kernel Q . Then, X satisfies all assumptions of Theorem 2.4, in particular we have $\sum_{\ell=0}^{\infty} \ell^2 p(\ell) = 3 < \infty$. Therefore, by Theorem 2.4, L_X conditioned on the events $\{|T| = n\}$ satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X})$ with good, convex rate function

$$I(\mu) = \begin{cases} H(\mu \| \mu_1 \otimes Q) + \sum_{a \in \mathcal{X}} \mu_2(a) I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right) & \text{if } \mu_1 \ll \mu_2, \\ \infty & \text{otherwise,} \end{cases} \quad (4.18)$$

where $I_p(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x + \log(2 - e^\lambda)\}$. Elementary Calculus shows that

$$\sup_{\lambda \in \mathbb{R}} \{\lambda x + \log(2 - e^\lambda)\} = x \log x - (x + 1) \log \frac{(x+1)}{2}. \quad (4.19)$$

Therefore, writing (4.19) in (4.18) and rearranging terms we obtain the form of the rate function in the corollary which completes the proof.

BIBLIOGRAPHY

- [1] D. ALDOUS. The continuum random tree II: An overview. In: *Stochastic analysis*, Proc. Symp., Durham/UK 1990, Lond. Math. Soc. Lect. Note Ser. 167 (1991), 23–70.
- [2] I. BENJAMINI AND Y. PERES. Markov chains indexed by trees. *Ann. Probab.* 22 (1994), 219–243.

- [3] K. DOKU-AMPONSAH. Large deviations and basic information theory for hierarchical and networked data structures. PhD Thesis, Bath 2006.
- [4] K. DOKU-AMPONSAH. Asymptotic equipartition properties for hierarchical and networked structures. (to appear) *Esiam: Probability and Statistics*.
- [5] K. DOKU-AMPONSAH and P. MÖRTERS. Large deviation principle for empirical measures of coloured random graphs. (to appear) *Annals of applied probability*.
- [6] A. DEMBO, P. MÖRTERS and S. SHEFFIELD. Large deviations of Markov chains indexed by random trees. *Ann. Inst. Henri Poincaré: Probab.et Stat.*41, (2005) 971-996.
- [7] A. DEMBO AND O. ZEITOUNI. Large deviations techniques and applications. Springer, New York, (1998).
- [8] A. MEIR AND J.W. MOON. On the altitude of nodes in random trees. *Canad. J. Math.* 30 (1978), 997–1015.
- [9] R. PEMANTLE. Tree-indexed processes. *Statist. Sci.* 10 (1995), no. 2, 200–231.

LARGE DEVIATION RESULTS FOR MULTITYPE GALTON-WATSON TREES

BY KWABENA DOKU-AMPONSAH
University of Ghana

Draft of September 18, 2010

Abstract

Using the notion of *consistency* of empirical measures, under fairly general assumption we prove a joint large deviation principles in n for the *empirical pair measure* and *empirical offspring measure* of multitype Galton-Watson tree conditioned to have exactly n vertices in the weak topology. From these results we obtain large deviation principle for empirical pair measure of Markov chains indexed by simply generated trees obtain by conditioning Galton-Watson trees on the total number of vertices. For the case where the offspring law of the tree is geometric distribution with parameter $\frac{1}{2}$, we get an exact rate function. All our rate functions are expressed in terms of relative entropies.

Keywords: Tree-indexed Markov chain, Tree-indexed process, random tree, Galton-Watson tree, multitype Galton-Watson tree, typed tree, joint large deviation principle, empirical pair measure, empirical offspring measure.

MSC 2000: Primary 60F10. Secondary 60J80, 05C05.

1. INTRODUCTION

Our main motivation for revisiting this model is the study of *Markov chain indexed by geometric $\frac{1}{2}$ offspring law* defined as follows: First we sample a tree from geometric distribution with parameter $\frac{1}{2}$, and then, given this tree, we run a Markov chain on the vertices of the tree in such a way that the state of a vertex depends only on the state of its parent. The result of this two-step experiment can also be interpreted as a *typed tree*. Large deviation principle for empirical pair measure of markov chains indexed by trees and the limit points find many application in information theory, statistical physics and the theory of Gibbs measures. See [3] or [4] for basic information theorem for hierarchical structures.

Large deviation study of Galton-Watson trees conditioned on the total size was first study by Dembo, Mörters and Sheffield [6]. In this paper, notions of shift-invariance and specific relative entropy-as typically understood for Markov fields on deterministic graphs such as \mathbb{Z}^d was extended to Markov fields on random trees. With these concepts, large deviation principles for empirical measures of a class of random trees including multitype Galton-Watson trees conditioned to have exactly n vertices were proved in a topology *stronger* than the weak topology. Their analysis have shown that large deviation results, which are well-known for classical Markov chains, can be extended to Markov chains indexed by random trees with offspring laws which have superexponential decay at infinity.i.e. Offspring law $p(\cdot)$ with $\ell^{-1} \log p(\ell) \rightarrow -\infty$ as $\ell \rightarrow \infty$. In the course of the proof of their main result, large deviation

principle for the empirical offspring measure of multitype Galton-Watson trees whose exponential moments are all finite was established in their topology, see [6].

The aim of this paper is to prove joint large deviation principle for the empirical offspring measure and empirical pair measure of multitype Galton-Watson trees with offspring laws which have finite second moments. This includes offspring laws considered in [6]. We extend the concept of *consistency* as understood for empirical measures of coloured random graphs, see Doku and Mörters [5] or Doku [3], to multitype Galton-Watson trees. The proof of our main results use the technique of (exponential) *change of measure* and three large deviations results, see [6, Lemmas 3.1 and 3.6] and [6, Theorem 2.2].

Using the contraction principle, see Dembo [7], we derive from our main results large deviation principle for empirical pair measure of Markov chain indexed by random trees, see Benjamini and Peres[2]. This result is similar to the one in [6]. We remark here that the process level large deviation principles for the empirical subtree measure and single-generation empirical measure, see [6], can be developed from our main results.

Specifically, we consider random tree models where trees and types are chosen *simultaneously* according to a *multitype Galton-Watson tree*. We recall from [6] the model of multitype Galton-Watson tree. To begin, we write $\mathcal{X}^* = \bigcup_{n=0}^{\infty} \{n\} \times \mathcal{X}^n$ and equip it with the discrete topology. We denote by \mathcal{T} the set of all finite rooted planar trees T , by $V = V(T)$ the set of all vertices and by $E = E(T)$ the set of all edges oriented away from the root, which is always denoted by ρ . We write $|T|$ for the number of vertices in the tree T . We note that the offspring of any vertex $v \in T$ is characterized by an element of \mathcal{X}^* and that there is an element $(0, \emptyset)$ in \mathcal{X}^* symbolizing absence of offspring. For each typed tree X and each vertex v we denote by

$$C(v) = (N(v), X_1(v), \dots, X_{N(v)}(v)) \in \mathcal{X}^*$$

the number and types of the children of v , ordered from left to right.

Given a probability measure μ on \mathcal{X} , serving as the initial distribution, and an offspring transition kernel \mathbb{Q} from \mathcal{X} to \mathcal{X}^* , we define the law \mathbb{P} of a tree-indexed process X , see Pemantle [9], by the following rules:

- The root ρ carries a random type $X(\rho)$ chosen according to the probability measure μ on \mathcal{X} .
- For each vertex with type $a \in \mathcal{X}$ the offspring number and types are given independently of everything else, by the offspring law $\mathbb{Q}\{\cdot | a\}$ on \mathcal{X}^* . We write

$$\mathbb{Q}\{\cdot | a\} = \mathbb{Q}\{(N, X_1, \dots, X_N) \in \cdot | a\},$$

i.e. we have a random number N of offspring particles with types X_1, \dots, X_N .

For every $c = (n, a_1, \dots, a_n) \in \mathcal{X}^*$ and $a \in \mathcal{X}$, the *multiplicity* of the symbol a in c is given by $m(a, c) = \sum_{i=1}^n \mathbf{1}_{\{a_i=a\}}$, and the matrix A with index set $\mathcal{X} \times \mathcal{X}$ and nonnegative entries is given by

$$A(a, b) = \sum_{c \in \mathcal{X}^*} \mathbb{Q}\{c | b\} m(a, c), \text{ for } a, b \in \mathcal{X}. \text{ i.e. } A(a, b) \text{ are the expected number of offspring of type } a$$

of a vertex of type b . We also recall from [6] the weak form of irreducibility concept. With $A^*(a, b) = \sum_{k=1}^{\infty} A^k(a, b) \in [0, \infty]$ we say that the matrix A is *weakly irreducible* if \mathcal{X} can be partitioned into a non empty set \mathcal{X}_r of *recurrent states* and a disjoint set \mathcal{X}_t of *transient states* such that

- $A^*(a, b) > 0$ whenever $b \in \mathcal{X}_r$, while
- $A^*(a, b) = 0$ whenever $b \in \mathcal{X}_t$ and either $a = b$ or $a \in \mathcal{X}_r$.

For example, any *irreducible* matrix A has A^* strictly positive, hence is also weakly irreducible with $\mathcal{X}_r = \mathcal{X}$. The multitype Galton-Watson tree is called weakly irreducible (or irreducible) if the matrix A

is weakly irreducible (or irreducible, respectively) and the number $\sum_{a \in \mathcal{X}_t} m(a, c)$ of transient offspring is uniformly bounded under \mathbb{Q} .

Recall that, by the Perron-Frobenius theorem, see e.g. [7, Theorem 3.1.1], the largest eigenvalue of an irreducible matrix is real and positive. Obviously, the same applies to weakly irreducible matrices. The multitype Galton-Watson tree is called *critical* if this eigenvalue is 1 for the matrix A .

The remaining part of the paper is organized in the following manner: The complete statement of our results is given in Section 2, we begin with joint LDP for empirical pair measures and empirical offspring measures of multitype Galton-Watson trees, followed by a corollary of LDP for the empirical offspring measure of multitype Galton-Watson trees in subsection 2.1. In subsection 2.2 we state the LDP for empirical pair measures of Markov chains indexed by tree. The proofs of our main results are then given in Section 3. All corollaries and Theorem 2.4 are proved in Section 4.

2. STATEMENT OF THE RESULTS

2.1 Joint large deviation principle for empirical pair measure and empirical offspring measure of multitype Galton-Watson trees.

For every sample chain X , we associate the *empirical offspring measure* M_X on $\mathcal{X} \times \mathcal{X}^*$, by

$$M_X(a, c) = \frac{1}{|T|} \sum_{v \in V} \delta_{(X(v), C(v))}(a, c), \quad (2.1)$$

and the empirical pair measure on $\mathcal{X} \times \mathcal{X}$, by

$$\tilde{L}_X(a, b) = \frac{1}{|T|} \sum_{e \in E} \delta_{(X(e_1), X(e_2))}(a, b), \quad \text{for } a, b \in \mathcal{X}, \quad (2.2)$$

where e_1, e_2 are the beginning and end vertex of the edge $e \in E$ (so e_1 is closer to ρ than e_2). We note that

$$\tilde{L}_X(a, b) = \sum_{c \in \mathcal{X}^*} m(b, c) M_X(a, c).$$

By definition, we notice that M_X is a probability vector and that total mass $\|L_X\|$ of L_X is $\frac{|T|-1}{|T|} \leq 1$.

Our main result is a large deviation principle for (\tilde{L}_X, M_X) if X is a multitype Galton-Watson tree. For its formulation denote, for every probability measure ν on $\mathcal{X} \times \mathcal{X}^*$, by ν_1 the \mathcal{X} -marginal of ν .

We call (ϖ, ν) *sub-consistent* if

$$\varpi(a, b) \geq \sum_{c \in \mathcal{X}^*} m(b, c) \nu(a, c) \quad \text{for all } a, b \in \mathcal{X}. \quad (2.3)$$

It call *consistent* if equality hold in (2.3). Observe that, if (ϖ, ν) is empirical pair measure and empirical offspring measure of multitype Galton-Watson tree then both sides of (2.3) is

$$\frac{1}{n} \times \#\{\text{number of edges with beginning vertex of type } a \text{ and end vertex of type } b\}.$$

We denote by $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ the space of probability measures ν on $\mathcal{X} \times \mathcal{X}^*$ with $\int n \nu(da, dc) < \infty$, using the convention $c = (n, a_1, \dots, a_n)$. Denote by $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ the space of finite measures on $\mathcal{X} \times \mathcal{X}$ and endow the space $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with the weak topology.

We call an offspring distribution \mathbb{Q} *bounded* if for some $k < \infty$, we have

$$\mathbb{Q}\{N > k | a\} = 0, \quad \text{for all } a \in \mathcal{X}.$$

Otherwise we call it *unbounded*.

Theorem 2.1. *Suppose that X is a weakly irreducible, critical multitype Galton-Watson tree with an unbounded offspring law \mathbb{Q} whose second moment is finite, conditioned to have exactly n vertices. Then, for $n \rightarrow \infty$, (\tilde{L}_X, M_X) satisfies a large deviation principle in $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with speed n and the convex, good rate function*

$$J(\varpi, \nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}) & \text{if } (\varpi, \nu) \text{ is sub-consistent and } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (2.4)$$

For $\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X})$, we write

$$\langle m(\cdot, c), \nu(a, c) \rangle(b) := \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} m(b, c) \nu(a, c), \text{ for } a \in \mathcal{X}$$

and state a corollary of Theorem 2.1:

Corollary 2.2. *Suppose that X is a weakly irreducible, critical multitype Galton-Watson tree with an unbounded offspring law \mathbb{Q} whose second moment is finite, conditioned to have exactly n vertices. Then, for $n \rightarrow \infty$, the empirical offspring measure M_X satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with speed n and the convex, good rate function*

$$K(\nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}) & \text{if } \langle m(\cdot, c), \nu(a, c) \rangle \leq \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (2.5)$$

We write $\mathcal{X}_k^* = \bigcup_{n=0}^k \{n\} \times \mathcal{X}^n$ and denote by \mathbb{Q}_k offspring transition kernel from \mathcal{X} to \mathcal{X}_k^* . The next large deviation principle is the main ingredient in the proof of the lower bound of Theorem 2.1.

Theorem 2.3. *Suppose that X is a weakly irreducible, critical multitype Galton-Watson tree with an offspring law \mathbb{Q}_k , conditioned to have exactly n vertices. Then, for $n \rightarrow \infty$, (\tilde{L}_X, M_X) satisfies a large deviation principle in $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ with speed n and the convex, good rate function*

$$J_k(\varpi, \nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}_k) & \text{if } (\varpi, \nu) \text{ is consistent and } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (2.6)$$

2.2 LDP for empirical pair measure of Markov chains indexed by trees. In this subsection, we look at the situation where the tree is generated independently of the types.

Suppose that T is any finite tree and we are given an initial probability measure μ on a finite alphabet \mathcal{X} and a Markovian transition kernel $\mathbb{Q}: \mathcal{X} \times \mathcal{X} \geq 0$. We can obtain a *Markov chain indexed by tree* $T, X: V \rightarrow \mathcal{X}$ as follows: Choose $X(\rho)$ according to μ and choose $X(v)$, for each vertex $v \neq \rho$, using the transition kernel given the value of its parent, independently of everything else. If the tree is chosen randomly, we always consider $X = \{X(v) : v \in T\}$ under the *joint law* of tree and chain. It is sometimes convenient to interpret X as a *typed tree*, considering $X(v)$ as the *type* of the vertex v .

We consider the class of *simply generated trees*, see [8] or [1], obtained by conditioning a critical Galton-Watson on its total number of vertices. To be specific, we look at the class of Galton-Watson trees, where the number of children $N(v)$ of each $v \in T$ is chosen independently according to the same law $p(\cdot) = \mathbb{P}\{N(v) = \cdot\}$ for all $v \in T$, while $0 < p(0) < 1$. We assume that p is critical. That is, the mean offspring number $\sum_{\ell=0}^{\infty} \ell p(\ell)$ is one, but this assumption is not restrictive: Note that the distribution of T conditioned on $\{|T| = n\}$ is exactly the same as when the offspring law is $p_{\theta}(\ell) = p(\ell)e^{\theta\ell} / \sum_j p(j)e^{\theta j}$, regardless of the value of $\theta \in \mathbb{R}$. With $0 < p(0) < 1 - p(1)$ there exists a unique θ_* such that $\sum_{\ell} \ell p_{\theta_*}(\ell) = 1$. Hence all our results hold in the noncritical cases with

p_{θ_*} in place of p . We allow offspring laws p with unbounded support, but we relax the assumption $\ell^{-1} \log p(\ell) \rightarrow -\infty$.

Throughout the paper statements conditioned on the event $\{|T| = n\}$ are made only for those values of n where $\mathbb{P}\{|T| = n\} > 0$.

We associate with each finite tree and sample chain X a probability measure on $\mathcal{X} \times \mathcal{X}$, the (normalized) *empirical pair measure* L_X , by

$$L_X(a, b) = \frac{1}{|E|} \sum_{e \in E} \delta_{(X(e_1), X(e_2))}(a, b), \quad \text{for } a, b \in \mathcal{X}, \quad (2.7)$$

where e_1, e_2 are the beginning and end vertex of the edge $e \in E$ (so e_1 is closer to ρ than e_2). Note that, $L_X = \frac{n}{n-1} \tilde{L}_X$ on the set $\{|T| = n\}$ and hence the LDP for \tilde{L}_X implies L_X by exponential equivalent Theorem, see Dembo [7].

Our first result in this subsection is a large deviation principle for L_X , conditional upon the event $\{|T| = n\}$ with n chosen such that the latter has positive probability. For its formulation recall the definition of the relative entropy $H(\cdot \| \cdot)$ from [7, (2.1.5)] and the Cramér's rate function, see e.g. [7, (2.1.26)],

$$I_p(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - \log \left[\sum_{n=0}^{\infty} p(n) e^{\lambda n} \right] \right\}. \quad (2.8)$$

Theorem 2.4. *Suppose that T is a Galton-Watson tree, with offspring law $p(\cdot)$ such that $0 < p(0) < 1 - p(1)$, $\sum_{\ell} \ell p(\ell) = 1$ and $\sum_{\ell} \ell^2 p(\ell) < \infty$. Let X be a Markov chain indexed by T with arbitrary initial distribution and an irreducible Markovian transition kernel Q . Then, for $n \rightarrow \infty$, the empirical pair measure L_X , conditioned on $\{|T| = n\}$ satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X})$ with speed n and the convex, good rate function*

$$I(\mu) = \begin{cases} H(\mu \| \mu_1 \otimes Q) + \sum_{a \in \mathcal{X}} \mu_2(a) I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right) & \text{if } \mu_1 \ll \mu_2, \\ \infty & \text{otherwise,} \end{cases} \quad (2.9)$$

where μ_1 and μ_2 are the first and second marginal of μ and $\mu_1 \otimes Q(a, b) = Q\{b | a\} \mu_1(a)$.

From Theorem 2.4 we obtain LDP for empirical pair measures of Galton-Watson trees with geometric distribution with parameter $\frac{1}{2}$.

Corollary 2.5. *Suppose that T is a Galton-Watson tree, with offspring law $p(\ell) = 2^{-(\ell+1)}$, $\ell = 0, 1, \dots$. Let X be a Markov chain indexed by T with arbitrary initial distribution and an irreducible Markovian transition kernel Q . Then, for $n \rightarrow \infty$, the empirical pair measure L_X , conditioned on $\{|T| = n\}$ satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X})$ with speed n and the convex, good rate function*

$$I(\mu) = \begin{cases} H(\mu \| \mu_1 \otimes Q) + H(\mu_1 \| (\mu_1 + \mu_2)/2) + H(\mu_2 \| (\mu_1 + \mu_2)/2) & \text{if } \mu_1 \ll \mu_2, \\ \infty & \text{otherwise,} \end{cases} \quad (2.10)$$

where μ_1 and μ_2 are the first and second marginal of μ and $\mu_1 \otimes Q(a, b) = Q\{b | a\} \mu_1(a)$.

3. PROOF OF MAIN RESULTS

3.1 Change of Measure, Exponential Tightness and Some General Principles.

Given a bounded function $\tilde{g} : \mathcal{X} \times \mathcal{X}^* \rightarrow \mathbb{R}$ we define the function

$$U_{\tilde{g}}(a) = \log \sum_{c \in \mathcal{X}^*} \mathbb{Q}\{c | a\} e^{\tilde{g}(a,c)}, \quad (3.1)$$

for $a \in \mathcal{X}$. We use \tilde{g} to define a new multitype Galton-Watson tree as follows:

- The type of the root ρ is $a \in \mathcal{X}$ with probability

$$\mu_{\tilde{g}}(a) = \frac{e^{U_{\tilde{g}}(a)} \mu(a)}{\int e^{U_{\tilde{g}}(b)} \mu(db)}. \quad (3.2)$$

- for each vertex with type $a \in \mathcal{X}$ the offspring number and types are given independently of everything else, by the offspring law $\tilde{\mathbb{Q}}\{\cdot | a\}$ given by

$$\tilde{\mathbb{Q}}\{c | a\} = \exp(\tilde{g}(a,c) - U_{\tilde{g}}(a)) \mathbb{Q}\{c | a\}. \quad (3.3)$$

We denote the transformed law by $\tilde{\mathbb{P}}$ and make the simple observation that $\tilde{\mathbb{P}}$ is absolutely continuous with respect to \mathbb{P} , as for each finite $X \in \mathcal{X}$,

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(X) = \frac{e^{U_{\tilde{g}}(X(\rho))}}{\int e^{U_{\tilde{g}}(b)} \mu(db)} \prod_{v \in V} \exp[\tilde{g}(X(v), C(v)) - U_{\tilde{g}}(X(v))] \quad (3.4)$$

$$= \frac{1}{\int e^{U_{\tilde{g}}(a)} \mu(da)} \prod_{v \in V} \exp[\tilde{g}(X(v), C(v)) - \sum_{b \in \mathcal{X}} m(b, C(v)) U_{\tilde{g}}(b)], \quad (3.5)$$

recalling that $C(v) = (N(v), X_1(v), \dots, X_N(v))$.

We begin by establishing exponential tightness of the family of laws of M_X on the space $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$.

Lemma 3.1. *For every $\alpha > 0$ there exists a compact $K_\alpha \subset \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{M_X \notin K_\alpha \mid |T| = n\} \leq -\alpha.$$

Proof. Given $l \in \mathbb{N}$, we may choose $k(l) \in \mathbb{N}$ so large that

$$\mathbb{Q}\{\exp(l^2 \mathbf{1}_{\{N > k(l)\}}) \mid a\} \leq 2, \text{ for all } a \in \mathcal{X}.$$

Using the exponential Chebyshev inequality,

$$\begin{aligned} \mathbb{P}\left\{\int_{\{N > k(l)\}} dM_X \geq \frac{1}{l}, |T| = n\right\} &\leq e^{-ln} \mathbb{E}\left\{\exp\left(l^2 n \int_{\{N > k(l)\}} dM_X\right), |T| = n\right\} \\ &= e^{-ln} \mathbb{E}\left\{\prod_{v \in T} \exp\left(l^2 \mathbf{1}_{\{N(v) > k(l)\}}\right), |T| = n\right\} \\ &\leq e^{-ln} \left(\sup_{a \in \mathcal{X}} \mathbb{Q}\{\exp(l^2 \mathbf{1}_{\{N > k(l)\}}) \mid a\}\right)^n \leq e^{-n(l - \log 2)}. \end{aligned}$$

Now choose $M > \alpha + \log 2$. Define the set

$$\Gamma_M = \left\{\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) : \int_{\{N > k(l)\}} d\nu < \frac{1}{l}, \text{ for all } l \geq M\right\}.$$

As $\{N \leq k(l)\} \subset \mathcal{X} \times \mathcal{X}^*$ is compact, the set Γ_M is pre-compact in the weak topology, by Prohorov's criterion. As

$$\mathbb{P}\{M_X \notin \Gamma_M \mid |T| = n\} \leq \frac{1}{\mathbb{P}\{|T| = n\}} \frac{1}{1 - e^{-1}} \exp(-n(M - \log 2)),$$

we can use [6, Lemma 3.1] to infer that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{M_X \notin K_\alpha \mid |T| = n\} \leq -\alpha,$$

for the closure K_α of Γ_M as required for the proof. \square

We denote by \mathcal{M}_s the set of all sub-consistent measures, and by \mathcal{M}_c the set of all consistent measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ and notice that $\mathcal{M}_c \subseteq \mathcal{M}_s$. For k a natural number, we denote by $\mathcal{M}_{c,k}$ the set of consistent measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. Then, \mathcal{M}_s is a closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ and $\mathcal{M}_{c,k}$ is a closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. The next two large deviation principles will help us extend LDP in $\mathcal{M}_{c,k}$, \mathcal{M}_s to $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ and $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ respectively.

Lemma 3.2. *Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q} . Assume (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in \mathcal{M}_s with convex, good rate function*

$$\tilde{J}(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}) & \text{if } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.6)$$

Then, (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^)$ with convex, good rate function J .*

Proof. Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q} and that an LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds in \mathcal{M}_s , with convex, good rate function \tilde{J} . Then, we have $\{|T| = n\} := \{\omega : |T|(\omega) = n\} \subseteq \{\omega : (\tilde{L}_X, M_X)(\omega) \in \mathcal{M}_c\} =: \{(\tilde{L}_X, M_X) \in \mathcal{M}_c\}$. Hence, for all n , we have

$$\begin{aligned} \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_s \mid |T| = n\} &= \frac{1}{\mathbb{P}\{|T| = n\}} \times \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_s, |T| = n\} \\ &\geq \frac{1}{\mathbb{P}\{|T| = n\}} \times \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_c, |T| = n\} \\ &= \frac{1}{\mathbb{P}\{|T| = n\}} \times \mathbb{P}\{|T| = n\} \\ &= 1. \end{aligned}$$

Also, if $(\varpi_n, \nu_n) \in \mathcal{M}_s$ converges (ϖ, ν) then we have that

$$\varpi(a, b) = \lim_{n \rightarrow \infty} \varpi_n(a, b) \geq \lim_{n \rightarrow \infty} \sum_{c \in \mathcal{X}^*} m(b, c) \nu_n(a, c) \geq \liminf_{n \rightarrow \infty} \sum_{c \in \mathcal{X}^*} m(b, c) \nu_n(a, c) \geq \sum_{c \in \mathcal{X}^*} m(b, c) \nu(a, c),$$

which implies (ϖ, ν) is sub-consistent. This means \mathcal{M}_s is closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$. Therefore, by [7, Lemma 4.1.5], the LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds with convex, good rate function J . \square

Lemma 3.3. *Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q}_k . Assume (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in $\mathcal{M}_{c,k}$ with convex, good rate function*

$$\tilde{J}_k(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}_k) & \text{if } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.7)$$

Then, (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ satisfies the LDP in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^)$ with convex, good rate function J_k .*

Proof. Suppose X is a multitype Galton-Watson tree with offspring law \mathbb{Q}_k and that an LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds in $\mathcal{M}_{c,k}$, with convex, good rate function \tilde{J}_k . Then, we have, $\{|T| = n\} \subseteq \{\omega : (\tilde{L}_X, M_X)(\omega) \in \mathcal{M}_{c,k}\} = \{(\tilde{L}_X, M_X) \in \mathcal{M}_{c,k}\}$. Hence, for all n ,

$$\begin{aligned} \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_{c,k} \mid |T| = n\} &= \frac{1}{\mathbb{P}\{|T|=n\}} \times \mathbb{P}\{(\tilde{L}_X, M_X) \in \mathcal{M}_{c,k}, |T| = n\} \\ &= \frac{1}{\mathbb{P}\{|T|=n\}} \times \mathbb{P}\{|T| = n\} \\ &= 1. \end{aligned}$$

Also, if $(\varpi_n, \nu_n) \in \mathcal{M}_{c,k}$ converges (ϖ, ν) then we have

$$\varpi(a, b) = \lim_{n \rightarrow \infty} \varpi_n(a, b) = \lim_{n \rightarrow \infty} \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu_n(a, c) = \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu(a, c),$$

which implies (ϖ, ν) is consistent. This means $\mathcal{M}_{c,k}$ is closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$. Hence, by [7, Lemma 4.1.5], the LDP for (\tilde{L}_X, M_X) conditioned on the event $\{|T| = n\}$ holds with convex, good rate function J_k which completes the proof of the Lemma. \square

In view of Lemmas 3.3 and 3.2, we establish large deviation principles in the spaces $\mathcal{M}_{c,k}$ and \mathcal{M}_s .

3.2 Proof of the upper bound in Theorem 2.1. Next we derive an upper bound in a variational formulation. Denote by \mathcal{C} the space of bounded functions on $\mathcal{X} \times \mathcal{X}^*$ and define for each (ϖ, ν) sub-consistent element in $\mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$, the function \hat{J} by

$$\hat{J}(\varpi, \nu) = \sup_{g \in \mathcal{C}} \left\{ \int g(b, c) \nu(db, dc) - \int U_g(b) \varpi(da, db) \right\}, \quad (3.8)$$

where $c = (n, a_1, \dots, a_n)$. We recall that \mathcal{M}_s is the set of all sub-consistent measures.

Lemma 3.4. *For each closed set $F \subset \mathcal{M}_s$, we have*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in F \mid |T| = n\} \leq - \inf_{(\varpi, \nu) \in F} \hat{J}(\varpi, \nu).$$

Proof. Fix $\tilde{g} \in \mathcal{C}$ bounded by some $M > 0$, then also $\int e^{U_{\tilde{g}}(a)} \mu(da) \leq e^M$. We observe that, by (3.5),

$$\begin{aligned} e^M &\geq \tilde{\mathbb{P}}\{|T| = n\} \int e^{U_{\tilde{g}}(a)} \mu(da) = \mathbb{E}\left\{ \prod_{v \in V} \exp\left[\tilde{g}(X(v), C(v)) - \sum_{b \in \mathcal{X}} m(b, C(v)) U_{\tilde{g}}(b)\right] \mathbf{1}_{\{|T|=n\}} \right\} \\ &= \mathbb{E}\left\{ e^{n\langle \tilde{g}, M_X \rangle - n\langle U_{\tilde{g}}, \tilde{L}_X \rangle} \mathbf{1}_{\{|T|=n\}} \right\}. \end{aligned}$$

Together with [6, Lemma 3.1] this shows that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}\left\{ e^{n\langle \tilde{g}, M_X \rangle - n\langle U_{\tilde{g}}, \tilde{L}_X \rangle} \mid |T| = n \right\} \leq 0. \quad (3.9)$$

In view of (3.4) the same bound (3.9) holds for only M_X

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}\left\{ e^{n\langle \tilde{g}, M_X \rangle - n\langle U_{\tilde{g}}, M_X \rangle} \mid |T| = n \right\} \leq 0. \quad (3.10)$$

Now fix $\varepsilon > 0$, and let $\hat{J}_\varepsilon(\varpi, \nu) = \min\{\hat{J}(\varpi, \nu), \varepsilon^{-1}\} - \varepsilon$. Suppose first that $(\varpi, \nu) \in F$ is such that $\varpi_2 = \nu_1$.

Choose $\tilde{g}_{(\varpi, \nu)} \in \mathcal{C}$ such that

$$\int \tilde{g}_{(\varpi, \nu)}(a, c) \nu(a, c) - \int U_{\tilde{g}_{(\varpi, \nu)}}(b) \varpi(da, db) \geq \widehat{J}_\varepsilon(\varpi, \nu). \quad (3.11)$$

Since $\tilde{g}_{(\varpi, \nu)}$ is bounded, the mapping $(\tilde{\varpi}, \tilde{\nu}) \mapsto \langle \tilde{g}_{(\varpi, \nu)}, \tilde{\nu} \rangle - \langle U_{\tilde{g}_{(\varpi, \nu)}}, \tilde{\varpi} \rangle$ is continuous in \mathcal{M}_s .

Hence there exists an open neighbourhood $B_{(\varpi, \nu)}$ of (ϖ, ν) such that

$$\inf_{(\tilde{\varpi}, \tilde{\nu}) \in B_{(\varpi, \nu)}} \left\{ \langle \tilde{g}_{(\varpi, \nu)}, \tilde{\nu} \rangle - \langle U_{\tilde{g}_{(\varpi, \nu)}}, \tilde{\varpi} \rangle \right\} \geq \left\{ \langle \tilde{g}_{(\varpi, \nu)}, \nu \rangle - \langle U_{\tilde{g}_{(\varpi, \nu)}}, \varpi \rangle \right\} - \varepsilon \geq \widehat{J}_\varepsilon(\varpi, \nu) - \varepsilon.$$

Using the exponential Chebyshev inequality and the remark following (??) we obtain that,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in B_{(\varpi, \nu)} \mid |T| = n\} \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}\{e^{n\langle \tilde{g}, M_X \rangle - n\langle U_{\tilde{g}}, \tilde{L}_X \rangle} \mid |T| = n\} - \widehat{J}_\varepsilon(\varpi, \nu) + \varepsilon \leq - \inf_{\nu \in F} \widehat{J}_\varepsilon(\varpi, \nu) + \varepsilon. \end{aligned} \quad (3.12)$$

Now suppose that (ϖ, ν) is such that $\varpi_2 \neq \nu_1$. Assume first that there exists $a \in \mathcal{X}$ such that

$$\nu_1(a) < \varpi_2(a). \quad (3.13)$$

As the mappings $(\varpi, \nu) \mapsto \nu_1 - \varpi_2$ are continuous in the weak topology, there exist $\delta > 0$ and a small open neighbourhood $B_{(\varpi, \nu)} \subset \mathcal{M}_s$ such that

$$\tilde{\nu}_1(a) < \tilde{\varpi}(a) - \delta, \text{ for all } (\tilde{\varpi}, \tilde{\nu}) \in B_{(\varpi, \nu)}. \quad (3.14)$$

Let $\tilde{g} \in \mathcal{C}$ be defined by $\tilde{g}(b, c) = -(\delta\varepsilon)^{-1} \mathbf{1}_a(b)$. Note that, by the definition (3.1), we have $U_{\tilde{g}}(b) = \tilde{g}(b, c)$ for all b and this vanishes unless $b = a$. Hence, by (3.14), for every $(\tilde{\varpi}, \tilde{\nu}) \in B_{(\varpi, \nu)}$ we have that

$$\int \tilde{g}(a, c) \nu(da, dc) - \int U_{\tilde{g}}(b) \varpi_2(db) > \varepsilon^{-1}.$$

Then, using the exponential Chebyshev inequality and (3.9),

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in B_{(\varpi, \nu)} \mid |T| = n\} \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}\{e^{n\langle \tilde{g}, M_X \rangle - n\langle U_{\tilde{g}}, \tilde{L}_X \rangle} \mid |T| = n\} - \varepsilon^{-1} \leq -\varepsilon^{-1} \leq - \inf_{(\varpi, \nu) \in F} \widehat{J}_\varepsilon(\varpi, \nu). \end{aligned} \quad (3.15)$$

In case the opposite inequality holds in (3.13) the same argument leads to (3.15) if \tilde{g} is defined as

$$\tilde{g}(b, c) = (\delta\varepsilon)^{-1} \mathbf{1}_a(b).$$

Now we use Lemma 3.1 to choose a compact set K_α (for $\alpha = \varepsilon^{-1}$) with

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{M_X \notin K_\alpha \mid |T| = n\} \leq -\varepsilon^{-1}. \quad (3.16)$$

We write $F_\alpha := \{(\varpi, \nu) : (\varpi, \nu) \in \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times K_\alpha\}$ and notice that, the set $F_\alpha \cap F$ is compact and hence it may be covered by finitely many of the sets $B_{(\varpi_1, \nu_1)}, \dots, B_{(\varpi_m, \nu_m)}$, with $(\varpi_i, \nu_i) \in F$ for $i = 1, \dots, m$. Hence,

$$\mathbb{P}\{M_X \in F \mid |T| = n\} \leq \sum_{i=1}^m \mathbb{P}\{(\tilde{L}_X, M_X) \in B_{(\varpi_i, \nu_i)} \mid |T| = n\} + \mathbb{P}\{(\tilde{L}_X, M_X) \notin F_\alpha \mid |T| = n\}.$$

Using (3.12) and (3.15) we obtain, for small enough $\varepsilon > 0$, that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in F \mid |T| = n\} &\leq \max_{i=1}^m \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in B_{\nu_i} \mid |T| = n\} - \varepsilon^{-1} \\ &\leq - \inf_{(\varpi, \nu) \in F} \hat{J}_\varepsilon(\varpi, \nu) + \varepsilon. \end{aligned}$$

Taking $\varepsilon \downarrow 0$ gives the required statement. \square

Recall that $\tilde{J} : \mathcal{M}_s \rightarrow [0, \infty]$ is given by

$$\tilde{J}(\varpi, \nu) = \begin{cases} H(\nu \parallel \nu_1 \otimes \mathbb{Q}) & \text{if } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.17)$$

We show that the convex rate function \tilde{J} may replace the function \hat{J} of (3.8) in the upper bound of Lemma 3.4.

Lemma 3.5. *The function \tilde{J} is convex and lower semicontinuous on \mathcal{M}_s . Moreover, $\tilde{J}(\varpi, \nu) \leq \hat{J}(\varpi, \nu)$, for any $(\varpi, \nu) \in \mathcal{M}_s$.*

Proof. We start by proving the inequality $\tilde{J}(\varpi, \nu) \leq \hat{J}(\varpi, \nu)$. To begin, suppose that $\varpi_2 \neq \nu_1$. Then, there exists $a_0 \in \mathcal{X}$ such that $\nu_1(a_0) \neq \varpi_2(a_0)$. For a_0 we define the function $\tilde{g}(b, c) = K \mathbf{1}_{a_0}(b)$ and observe that $U_{\tilde{g}}(b) = g(b, c)$. Using this \tilde{g} in (3.8) we obtain

$$\int \tilde{g}(a, c) \nu(da, dc) - \int U_{\tilde{g}}(b) \varpi_2(db) = K(\nu_1(a_0) - \varpi_2(a_0)) \longrightarrow \infty,$$

for $|K| \uparrow \infty$, with the sign of K chosen so that the right hand side is positive.

Next, suppose that $\varpi_2 = \nu_1$ but $\nu \not\ll \nu_1 \otimes \mathbb{Q}$. Then, there exists $(a', c') \in \mathcal{X} \times \mathcal{X}^*$ with $\nu(a', c') > 0$ and $\mathbb{Q}\{c' \mid a'\} = 0$. Consequently, recalling (3.1), we have $U_{\tilde{g}} = 0$ for $\tilde{g}(b, c) = R \mathbf{1}_{(a', c')}(b, c)$ and any R . Considering such \tilde{g} in (3.8) with $R \uparrow \infty$ we see that $\hat{J}(\varpi, \nu) = \infty$ in this case.

Finally suppose $\varpi_2 = \nu_1$ and $\nu \ll \nu_1 \otimes \mathbb{Q}$. By the variational characterisation of the relative entropy, see e.g. [7, Lemma 6.2.13], the definition of $U_{\tilde{g}}$ and Jensen's inequality,

$$\begin{aligned} H(\nu \parallel \nu_1 \otimes \mathbb{Q}) &= \sup_{g \in \mathcal{C}} \left\{ \int g d\nu - \log \iint e^{g(a, c)} \mathbb{Q}\{dc \mid a\} \nu_1(da) \right\} \\ &= \sup_{g \in \mathcal{C}} \left\{ \int g d\nu - \log \int e^{U_g(a)} \nu_1(da) \right\} \\ &\leq \sup_{g \in \mathcal{C}} \left\{ \int g d\nu - \int U_g(a) \nu_1(da) \right\} = \hat{J}(\varpi, \nu). \end{aligned} \quad (3.18)$$

Now consider the convex, good rate function $\phi : \mathbb{R} \rightarrow [0, \infty]$ given by $\phi(x) = x \log x - x + 1$. Then, we can represent the left side of (3.18) in the form

$$H(\nu \parallel \nu_1 \otimes \mathbb{Q}) = \begin{cases} \int \phi \circ f d(\nu_1 \otimes \mathbb{Q}) & \text{if } f := \frac{d\nu}{d(\nu_1 \otimes \mathbb{Q})} \text{ exists} \\ \infty & \text{otherwise.} \end{cases} \quad (3.19)$$

Consequently, by [7, Lemma 6.2.16], \tilde{J} is a convex, good rate function. \square

By Lemma 3.2 the large deviation upper bound Lemma 3.4 holds with rate function \tilde{J} replaced by J .

3.3 Proof of Theorem 2.3. We begin by recalling from [6] that $\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ is shift-invariant if

$$\nu_1(a) = \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}^*} m(a,c) \nu(b,c), \text{ for all } a \in \mathcal{X}.$$

This Theorem is derived from [6, Theorem 2.2] by applying the contraction principle to the linear mapping $G : \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*) \mapsto \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*)$ given by $G(\nu) = (\varpi, \nu)$ where (ϖ, ν) is consistent. Specifically, [6, Theorem 2.2] implies the large deviation for $G(M_X)$ with convex, good rate function $J_k(\varpi, \nu) = \inf \{K_k(\nu) : G(\nu) = (\varpi, \nu), (\varpi, \nu) \text{ is consistent}\}$, where

$$K_k(\nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}_k) & \text{if } \langle m(\cdot, c), \nu(a, c) \rangle = \nu_1, \\ \infty & \text{otherwise.} \end{cases} \quad (3.20)$$

Using shift-invariance and consistency we have

$$\nu_1(a) = \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}_k^*} m(a,c) \nu(b,c) = \sum_{b \in \mathcal{X}} \varpi(b, a) = \varpi_2(b), \text{ for all } a \in \mathcal{X},$$

which completes the proof of the theorem.

3.4 Proof of the Lower Bound in Theorem 2.1

We define for every weakly irreducible, critical offspring kernel $\mathbb{Q}\{c|b\}$ the conditional offspring law

$$\mathbb{Q}_k\{c|a\} = \begin{cases} \frac{1}{\mathbb{Q}\{\mathcal{X}_k^*|a\}} \mathbb{Q}\{c|a\} & \text{if } c \in \mathcal{X}_k^* \\ 0 & \text{otherwise,} \end{cases} \quad (3.21)$$

where $\mathbb{Q}\{\mathcal{X}_k^*|a\} = \sum_{c \in \mathcal{X}_k^*} \mathbb{Q}\{c|a\}$. We write $\|\nu\|_k = \nu(\mathcal{X} \times \mathcal{X}_k^*)$ and observe that

$$\lim_{k \rightarrow \infty} \nu(\mathcal{X} \times \mathcal{X}_k^*) = \lim_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \mathbb{1}_{\{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*\}} \nu(a, c) = 1,$$

by the dominated convergence. We define ν_k a probability measure on $\mathcal{X} \times \mathcal{X}^*$ by

$$\nu_k(a, c) = \begin{cases} \frac{\nu(a, c)}{\|\nu\|_k} & \text{if } (a, c) \in \mathcal{X} \times \mathcal{X}_k^* \\ 0 & \text{otherwise,} \end{cases} \quad (3.22)$$

and denote by $(\nu_k)_1$ the \mathcal{X} -marginal of the probability measure ν_k . Write

$$\varpi_k(a, b) := \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu_k(a, c).$$

Lemma 3.6. *Let $\nu \ll \nu_1 \otimes \mathbb{Q}$. Then, we have*

$$\lim_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \frac{\nu(a, c)}{\|\nu\|_k} \log \frac{\frac{\nu(a, c)}{\|\nu\|_k}}{\left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k(a, c)} = \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a, c) \log \frac{\nu(a, c)}{\nu_1 \otimes \mathbb{Q}(a, c)}. \quad (3.23)$$

Proof. Recall that we have assumed $\nu \ll \nu_1 \otimes \mathbb{Q}$. Then we have $\frac{\nu}{\|\nu\|_k} \ll \left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k$ and so, we may write

$$\sum_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \frac{\nu(a, c)}{\|\nu\|_k} \log \frac{\frac{\nu(a, c)}{\|\nu\|_k}}{\left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k(a, c)} = \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \mathbb{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a, c)}{\|\nu\|_k} \mathbb{1}_{\{\nu(a, c) > 0\}} \log \frac{\frac{\nu(a, c)}{\|\nu\|_k}}{\left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k(a, c)}.$$

We write $f_k(a, c) := \mathbf{1}_{\{\nu(a, c) > 0\}} \log \frac{\frac{\nu(a, c)}{\|\nu\|_k}}{\left(\frac{\nu}{\|\nu\|_k}\right)_1 \otimes \mathbb{Q}_k(a, c)}$, for $(a, c) \in \mathcal{X} \times \mathcal{X}^*$ and notice that

$$\lim_{n \rightarrow \infty} f_k(a, c) = \mathbf{1}_{\{\nu(a, c) > 0\}} \log \frac{\nu(a, c)}{\nu_1 \otimes \mathbb{Q}(a, c)} := f(a, c)$$

is bounded function. Hence, for any $\delta > 0$ and sufficiently large k we have that

$$\sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \mathbf{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a, c)}{\|\nu\|_k} (f(a, c) - \delta) \leq \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \mathbf{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a, c)}{\|\nu\|_k} f_k(a, c) \leq \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \mathbf{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a, c)}{\|\nu\|_k} (f(a, c) + \delta). \quad (3.24)$$

Now we take limit as k approaches ∞ of all sides of (3.24) to obtain

$$\sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a, c) (f(a, c) - \delta) \leq \lim_{k \rightarrow \infty} \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \mathbf{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a, c)}{\|\nu\|_k} f_k(a, c) \leq \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a, c) (f(a, c) + \delta)$$

Taking $\delta \downarrow 0$, yields

$$\begin{aligned} \lim_{k \rightarrow \infty} \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \mathbf{1}_{\{c \in \mathcal{X}_k^*\}} \frac{\nu(a, c)}{\|\nu\|_k} f_k(a, c) &= \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a, c) f(a, c) = \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \mathbf{1}_{\{\nu(a, c) > 0\}} \nu(a, c) \log \frac{\nu(a, c)}{\nu_1 \otimes \mathbb{Q}(a, c)} \\ &= \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} \nu(a, c) \log \frac{\nu(a, c)}{\nu_1 \otimes \mathbb{Q}(a, c)} \end{aligned}$$

by the sandwich theorem of limits in the first equality, and the conventions $0 \log 0 = 0$, $0 \log 0 = 0$. \square

We define the total variation metric d by

$$d(\nu, \tilde{\nu}) = \frac{1}{2} \sum_{(a, c) \in \mathcal{X} \times \mathcal{X}^*} |\nu(a, c) - \tilde{\nu}(a, c)| \quad (3.25)$$

and observe that it generates the weak topology. For \mathbb{Q}_k we recall the definition of the rate function $J_k : \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}_k^*) \rightarrow [0, \infty]$ from Theorem 2.3 as

$$J_k(\varpi, \nu) = \begin{cases} H(\nu \| \nu_1 \otimes \mathbb{Q}_k) & \text{if } (\varpi, \nu) \text{ is consistent and } \varpi_2 = \nu_1, \\ \infty & \text{otherwise.} \end{cases}$$

The next lemma which is a key ingredient in our proof will be proved in two approximation steps.

Lemma 3.7. *Suppose (ϖ, ν) is sub-consistent, $\varpi_2 = \nu_1$ and $\nu \ll \nu_1 \otimes \mathbb{Q}$. Then, for every $\varepsilon > 0$, there exists $(\hat{\varpi}, \hat{\nu}) \in \mathcal{M}_{c, k}$ such that $|\varpi(a, b) - \hat{\varpi}(a, b)| < \varepsilon$, for all $a, b \in \mathcal{X}$, $d(\nu, \hat{\nu}) \leq \varepsilon$, $\hat{\varpi}_2 = \hat{\nu}_1$ and*

$$|J_k(\hat{\varpi}, \hat{\nu}) - J(\varpi, \nu)| \leq \varepsilon.$$

Proof. Step 1: Approximating sub-consistent (ϖ, ν) by a sequence of $(\hat{\varpi}_n, \nu_n)$. Recall our assumption that (ϖ, ν) is sub-consistent, $\varpi_2 = \nu_1$ and $\nu \ll \nu_1 \otimes \mathbb{Q}$. For any $b \in \mathcal{X}$ we define $e^{(b)} \in \mathcal{N}(\mathcal{X})$ by $e^{(b)}(a) = 0$ if $a \neq b$, and $e^{(b)}(a) = 1$ if $a = b$. We write $n(c) = (m(a, c), a \in \mathcal{X})$ and for large n define $\hat{\nu}_n \in \mathcal{M}_s$ by

$$\hat{\nu}_n(a, c) = \nu(a, c) \left(1 - \frac{\|\varpi\| - \|\langle \nu(\cdot, c), m(\cdot, c) \rangle\|}{n} \right) + \sum_{b \in \mathcal{X}} \mathbf{1}_{\{n(c) = ne^{(b)}\}} \frac{\varpi(a, b) - \langle \nu(\cdot, c), m(\cdot, c) \rangle(a, b)}{n}.$$

We note that $\hat{\nu}_n \rightarrow \nu$ and that, for all $a, b \in \mathcal{X}$,

$$\begin{aligned} \sum_{c \in \mathcal{X}^*} m(a, c) \hat{\nu}_n(b, c) &= \left(1 - \frac{\|\varpi\| - \|\langle \nu(\cdot, c), m(\cdot, c) \rangle\|}{n}\right) \sum_{c \in \mathcal{X}^*} m(a, c) \nu(b, c) + \varpi(a, b) - \langle \nu(\cdot, c), m(\cdot, c) \rangle(a, b) \\ &= \varpi(a, b) - \frac{\|\varpi\| - \|\langle \nu(\cdot, c), m(\cdot, c) \rangle\|}{n} \langle \nu(\cdot, c), m(\cdot, c) \rangle(a, b) \xrightarrow{n \uparrow \infty} \varpi(a, b). \end{aligned}$$

Defining $\hat{\omega}_n$ by $\hat{\omega}_n(a, b) = \sum_{c \in \mathcal{X}^*} m(a, c) \hat{\nu}_n(b, c)$, we have a sequence of consistent pairs $(\hat{\omega}_n, \nu_n)$ converging to (ϖ, ν) , $\nu_n \ll (\nu_n)_1 \otimes \mathbb{Q}$ and $(\hat{\omega}_n)_2 = (\hat{\nu}_n)_1$, as if $(\hat{\omega}_n)_2(a) \neq (\hat{\nu}_n)_1(a)$ for some $a \in \mathcal{X}$, then we have

$$\hat{\omega}_2(a) = \lim_{n \rightarrow \infty} (\hat{\omega}_n)_2(a) \neq \lim_{n \rightarrow \infty} (\hat{\nu}_n)_1(a) = \nu_1(a).$$

Further, by the continuity of relative entropy, we have that

$$\lim_{n \rightarrow \infty} J(\hat{\omega}_n, \hat{\nu}_n) = \sum_{a \in \mathcal{X}} \hat{\nu}_n(a) 1_{\{\hat{\nu}_n(a) > 0\}} H(\hat{\nu}_n(\cdot|a) \| \mathbb{Q}) = J(\varpi, \nu), \quad (3.26)$$

where $0 \log \frac{0}{0} = 0$ and $0 \log 0 = 0$ by convection.

Step 2: Approximating consistent $(\hat{\omega}_n, \hat{\nu}_n)$ by a sequence $(\varpi_{k,n}, \nu_{k,n}) \in \mathcal{M}_{c,k}$. Let $\varepsilon > 0$ and set $\delta = 2\varepsilon$. Choose $k(\varepsilon)$ large enough such that

$$\sup_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \left| \frac{1_{\{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*\}}}{\|\hat{\nu}_n\|_k} - 1 \right| \leq \delta.$$

For $k \geq k(\varepsilon)$, define $\nu_{k,n}$ a probability measure on $\mathcal{X} \times \mathcal{X}^*$ by

$$\nu_{k,n}(a, c) = \begin{cases} \frac{\hat{\nu}_n(a, c)}{\|\hat{\nu}_n\|_k} & \text{if } (a, c) \in \mathcal{X} \times \mathcal{X}_k^* \\ 0 & \text{otherwise,} \end{cases} \quad (3.27)$$

and denote by $(\nu_{k,n})_1$ the \mathcal{X} -marginal of the probability measure $\nu_{k,n}$. Write

$$\varpi_{k,n}(a, b) := \sum_{c \in \mathcal{X}_k^*} m(b, c) \nu_{k,n}(a, c).$$

We note that

$$\begin{aligned} d(\nu_{k,n}, \hat{\nu}_n) &= \frac{1}{2} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} |\nu_{k,n}(a, c) - \hat{\nu}_n(a, c)| = \frac{1}{2} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \left| \frac{1_{\{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*\}}}{\|\hat{\nu}_n\|_k} \hat{\nu}_n(a, c) - \hat{\nu}_n(a, c) \right| \leq \frac{1}{2} \delta \\ &= \varepsilon \end{aligned}$$

Define for each $n > 1$ the measure $\tilde{\omega}_n$ by $\tilde{\omega}_n(a, b) = (1 + \frac{1}{n})\varpi(a, b)$, for $a, b \in \mathcal{X}$. Then, it is not hard to see that

$$\lim_{n \rightarrow \infty} \tilde{\omega}_n(a, b) = \varpi(a, b) \text{ for all } a, b \in \mathcal{X}.$$

Observe that

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \varpi_{k,n}(a, b) \geq \lim_{n \rightarrow \infty} \liminf_{k \rightarrow \infty} \varpi_{k,n}(a, b) \geq \lim_{n \rightarrow \infty} \hat{\omega}_n(a, b) = \varpi(a, b), \text{ for all } a, b \in \mathcal{X}.$$

Hence, there $n(\varepsilon)$ such that $n \geq n(\varepsilon)$ implies $\varpi_{k,n}(a, b) \geq \varpi(a, b) - \frac{\varepsilon}{2}$. Hence, for large $k \geq k(\varepsilon)$ and large $n \geq n(\varepsilon)$ we have that

$$|\tilde{\omega}_n(a, b) - \varpi_{k,n}(a, b)| \leq |\tilde{\omega}_n(a, b) - \varpi(a, b)| + |\varpi(a, b) - \varpi_{k,n}(a, b)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

by the triangle inequality. This gives

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \varpi_{k,n}(a, b) = \lim_{n \rightarrow \infty} \tilde{\varpi}_n(a, b) = \varpi(a, b), \text{ for all } a, b \in \mathcal{X}.$$

Using similar argument as above we obtain

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} (\varpi_{k,n})_2(b) = \lim_{n \rightarrow \infty} (\tilde{\varpi}_n)_2(b) = (\varpi)_2(b), \text{ for all } b \in \mathcal{X}.$$

Observe that, by construction $(\varpi_{k,n}, \nu_{k,n})$ is consistent, $\nu_{k,n} \ll (\nu_{k,n})_1 \otimes \mathbb{Q}_k$ and $(\varpi_{k,n})_2 = (\nu_{k,n})_1$, as if $(\nu_{k,n})_1(a) \neq (\varpi_{k,n})_2(a)$ for some $a \in \mathcal{X}$, then we have that

$$\varpi_2(a) = \lim_{n \rightarrow \infty} (\varpi_n)_2(a) = \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} (\varpi_{k,n})_2(a) \neq \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} (\nu_{k,n})_1(a) = \lim_{n \rightarrow \infty} (\nu_n)_1(a) = \nu_1(a).$$

Now we recall from (3.27) the definition of $\nu_{k,n}$ and notice that we have,

$$\sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu_{k,n}(a, c) \log \frac{\nu_{k,n}(a,c)}{(\nu_{k,n})_1 \otimes \mathbb{Q}_k(a, c)} = \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}_k^*} \frac{\hat{\nu}_n(a,c)}{\|\hat{\nu}_n\|_k} \log \frac{\frac{\hat{\nu}_n(a,c)}{\|\hat{\nu}_n\|_k}}{\left(\frac{\hat{\nu}_n}{\|\hat{\nu}_n\|_k}\right)_1 \otimes \mathbb{Q}_k(a, c)},$$

where $0 \log \frac{0}{0} = 0$ and $0 \log 0 = 0$ by convection. Using Lemma 3.6 and (3.26) we obtain

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} J_k(\varpi_{k,n}, \nu_{k,n}) = \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \sum_{(a,c) \in \mathcal{X} \times \mathcal{X}^*} \nu_{k,n}(a, c) \log \frac{\nu_{k,n}(a,c)}{(\nu_{k,n})_1 \otimes \mathbb{Q}_k(a, c)} = \lim_{n \rightarrow \infty} J(\hat{\varpi}_n, \hat{\nu}_n) = J(\varpi, \nu),$$

which completes the proof of the lemma. \square

We recall that $C(v) = (N(v), X_1(v), \dots, X_{N(v)}(v))$ and note that, for every k such that $\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} > 0$ and any tree-indexed process x , we have that

$$\begin{aligned} \mathbb{P}\{X = x \mid |T| = n\} &\geq \mathbb{P}\{(X = x, C(v) \in \mathcal{X}_k^*, v \in V \mid |T| = n)\} \\ &= \prod_{v \in V, |T|=n} \mathbb{Q}\{\mathcal{X}_k^* | x(v)\} \times \mathbb{P}_k\{X = x \mid |T| = n\} \\ &\geq \left(\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\}\right)^n \times \mathbb{P}_k\{X = x \mid |T| = n\}, \end{aligned} \quad (3.28)$$

where \mathbb{P}_k denote the law of tree-indexed process with initial distribution μ and offspring kernel \mathbb{Q}_k , and

$$\lim_{k \rightarrow \infty} \min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} = \lim_{k \rightarrow \infty} \min_{a \in \mathcal{X}} \sum_{c \in \mathcal{X}_k} \mathbb{Q}\{c | a\} = \lim_{k \rightarrow \infty} \sum_{c \in \mathcal{X}^*} 1_{\{c \in \mathcal{X}_k\}} \min_{a \in \mathcal{X}} \mathbb{Q}\{c | a\} = 1.$$

To complete the proof of the lower bound, we take $O \subset \mathcal{M}_s$. Then, for any $(\varpi, \nu) \in O$ sub-consistent with $\varpi_2 = \nu_1$, $\nu \ll \nu_1 \otimes \mathbb{Q}$ we may find $\varepsilon > 0$ with ball around (ϖ, ν) of radius 2ε contained in O . By our approximation Lemma 3.7, we may find $(\varpi_{k,n}, \nu_{k,n}) \in O \cap \mathcal{M}_{c,k}$ with $|\varpi_{k,n}(a, b) - \varpi(a, b)| \downarrow 0$, $d(\nu_{k,n}, \nu) \downarrow 0$ and $|J_k(\varpi_{k,n}, \nu_{k,n}) - J(\varpi, \nu)| \downarrow 0$. Hence, using the lower bound of Theorem 2.3 for offspring kernel \mathbb{Q}_k given by (3.21), (3.28) for large $k \geq k(\varepsilon)$ (with $\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k^* | a\} > 0$) and for large $n \geq n(\varepsilon)$, we obtain

$$\begin{aligned} \mathbb{P}\{(\tilde{L}_X, M_X) \in O \mid |T| = n\} &\geq \mathbb{P}\{|\varpi_{k,n}(a, b) - \tilde{L}_X(a, b)| < \varepsilon, \ a, b \in \mathcal{X}, d(\nu_{k,n}, M_X) < \varepsilon \mid |T| = n\} \\ &\geq e^{n\alpha_k} \times \mathbb{P}_k\{|\varpi_{k,n}(a, b) - \tilde{L}_X(a, b)| < \varepsilon, \ a, b \in \mathcal{X}, d(\nu_{k,n}, M_X) < \varepsilon \mid |T| = n\} \\ &\geq \exp(-n(J_k(\varpi_{k,n}, \nu_{k,n}) + \varepsilon - \alpha_k)) \end{aligned}$$

where $\alpha_k = \log(\min_{a \in \mathcal{X}} \mathbb{Q}\{\mathcal{X}_k|a\})$. Taking limits we have that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\{(\tilde{L}_X, M_X) \in O \mid |T| = n\} \geq -\limsup_{n \rightarrow \infty} \lim_{k \rightarrow \infty} J_k(\varpi_{k,n}, \nu_{k,n}) - \varepsilon \geq -J(\varpi, \nu) - \varepsilon.$$

Taking $\varepsilon \downarrow 0$ we have the desired result which completes the proof of the lower bound.

4. PROOF OF COROLLARIES 2.2, 2.5 AND THEOREM 2.4

4.1 Proof of Corollary 2.2. We derive this corollary from Theorem 2.1 by applying the contraction principle to the linear mapping $W : \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) \mapsto \mathbb{R}^{\mathcal{X} \times \mathcal{X}^*}$ defined by

$$W(\varpi, \nu)(a, c) = \nu(a, c), \quad \text{for all } (a, c) \in \mathcal{X} \times \mathcal{X}^*.$$

In fact Theorem 2.1 implies the large deviation principle for $W(\tilde{L}_X, M_X)$ with convex, good rate function $\hat{J}(\nu) = \inf \{J(\varpi, \nu) : W(\varpi, \nu) = \nu\}$. Now, using sub-consistency and $\varpi_2 = \nu_1$ we obtain the form $\hat{J}(\nu) = H(\nu \parallel \nu_1 \otimes \mathbb{Q})$, for ν satisfying $\langle m(\cdot, c), \nu(a, c) \rangle \leq \nu_1$. Recall the definition of shift-invariant and denote by \mathcal{M}_1 set of shift-invariant measures in $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$. Write

$$\mathcal{M}_2 = \left\{ \nu : \nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X}^*), m(\cdot, c), \nu(a, c) \leq \nu_1 \right\}$$

and note that $\mathcal{M}_1 \subseteq \mathcal{M}_2$. Also, for all (values of) n where $\mathbb{P}\{|T| = n\} > 0$, we have

$$\mathbb{P}\{M_X \in \mathcal{M}_2 \mid |T| = n\} = 1.$$

Moreover, if $\nu_n \in \mathcal{M}_2$ converges to ν then

$$\nu_1(a) = \lim_{n \rightarrow \infty} (\nu_n)_1(a) \geq \liminf_{n \rightarrow \infty} \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}^*} m(a, c) \nu_n(b, c) \geq \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}^*} m(a, c) \nu(b, c),$$

which implies ν is sub-consistent. This means \mathcal{M}_2 is closed subset of $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$. Therefore, by [7, Lemma 4.1.5], the LDP for M_X conditional on the event $\{|T| = n\}$ holds with convex, good rate function K , which completes the proof of the corollary.

4.2 Proof of Theorem 2.4. We begin the proof of the theorem by stating the following Lemma.

Lemma 4.1. *Suppose that $q(c) = p(n) \prod_{i=1}^n \hat{q}(a_i)$ for all $c = (n, a_1, \dots, a_n)$, where $\hat{q}(\cdot)$ is a probability vector on \mathcal{X} and $p(\cdot)$ a probability measure with mean one on the nonnegative integers. Then, we have*

$$\inf \left\{ H(\tilde{\nu} \parallel q) : \tilde{\nu} \in \mathcal{M}(\mathcal{X}^*), \phi(b) = \sum_{c \in \mathcal{X}^*} m(b, c) \tilde{\nu}(c) \text{ for all } b \in \mathcal{X} \right\} = zH(\phi/z \parallel \hat{q}) + I_p(z), \quad (4.1)$$

where $z = \sum_b \phi(b)$.

Proof. For $\tilde{\nu} \in \mathcal{M}(\mathcal{X}^*)$, we let $\phi(b) = \sum_{c \in \mathcal{X}^*} m(b, c) \tilde{\nu}(c)$, for all $b \in \mathcal{X}$ and suppose first that $z = 0$, i.e. $\phi(b) = 0$ for all $b \in \mathcal{X}$. Then, $\tilde{\nu}((0, \emptyset)) = 1$ is the only possible measure in left side of (4.1), leading to $\tilde{I}(\phi, q) = -\log q((0, \emptyset)) = -\log p(0)$. It follows from (2.8) that $I_p(0) = -\log p(0)$ establishing (4.1) for such $\phi(\cdot)$. We assume hereafter that $z > 0$. Now the possible measures $\tilde{\nu}(\cdot)$ in the left side of (4.1) are of the form $\tilde{\nu}(c) = s(n)v_n(a_1, \dots, a_n)$ for $c = (n, a_1, \dots, a_n)$, with $v_0 = 1$, where $s(\cdot)$ is a probability measure on the nonnegative integers whose mean is z , and $v_n(\cdot)$, $n \geq 1$, are probability measures on \mathcal{X}^n with marginals $v_{n,i}(\cdot)$ such that

$$\phi(b) = \sum_{n=1}^{\infty} s(n) \sum_{i=1}^n v_{n,i}(b) \quad \text{for all } b \in \mathcal{X}. \quad (4.2)$$

By the assumed structure of $q(\cdot)$ we have for such $\tilde{\nu}(\cdot)$ that

$$H(\tilde{\nu} \| q) = \sum_{n=1}^{\infty} s(n) H(v_n \| \hat{q}^n) + H(s \| p), \quad (4.3)$$

where \hat{q}^n denotes the product measure on \mathcal{X}^n with equal marginals \hat{q} . Recall that

$$\sum_{n=1}^{\infty} s(n) H(v_n \| \hat{q}^n) \geq \sum_{n=1}^{\infty} s(n) \sum_{i=1}^n H(v_{n,i} \| \hat{q}) \geq z H\left(z^{-1} \sum_{n=1}^{\infty} s(n) \sum_{i=1}^n v_{n,i} \| \hat{q}\right),$$

with equality whenever $v_n = \prod_{i=1}^n v_{n,i}$ and $v_{n,i}$ are independent of n and i (see [7, Lemma 7.3.25] for the first inequality, with the second inequality following by convexity of $H(\cdot \| \hat{q})$ and the fact that $\sum_n s(n)n = z$). So, in view of (4.2),

$$H(\tilde{\nu} \| q) \geq z H(\phi/z \| \hat{q}) + H(s \| p), \quad (4.4)$$

with equality when

$v_n = (z^{-1}\phi)^n$ for all $n \geq 1$. Now, write $\Lambda_p(\lambda) := \log \sum_n e^{\lambda n} p(n)$ and notice that Λ convex function and $\Lambda(0) = 0 < \infty$, and so, we have, for every $\lambda \in \mathbb{R}$, $\Lambda_p(\lambda) > -\infty$. Using Jensen's inequality, for every $s \in \mathcal{M}(\mathbb{N} \cup \{0\})$ and every $\lambda \in \mathbb{R}$, we have

$$\Lambda_p(\lambda) = \log \sum_n s(n) \left(\frac{e^{\lambda n} p(n)}{s(n)}\right) \geq \sum_n s(n) \log \left(\frac{e^{\lambda n} p(n)}{s(n)}\right) = \lambda \sum_n n s(n) - H(s \| p),$$

with equality if $s_\lambda(n) = p(n)e^{\lambda n - \Lambda(\lambda)}$. Thus, for all λ and all z , we have

$$\lambda z - \Lambda(\lambda) \leq \inf \{H(s \| p) : s \in \mathcal{M}(\mathbb{N} \cup \{0\}) \text{ and } \sum_n s(n)n = z\} := \Lambda^*(z), \quad (4.5)$$

with equality when $\sum_n s(n)n = z$. Elementary calculus also shows that

$$\Lambda^*(z) = \lambda_* z - \Lambda_p(\lambda_*), \quad (4.6)$$

where λ_* is the solution of $\Lambda_p'(\lambda_*) = z$ and $\frac{d\Lambda_p}{d\lambda} := \Lambda_p'(\lambda)$. Combining (4.6) and (4.5) we obtain

$$\sup_{\lambda \in \mathbb{R}} \{\lambda z - \Lambda_p(\lambda)\} \leq \Lambda^*(z) \leq \sup_{\lambda \in \mathbb{R}, \Lambda_p'(\lambda)=z} \{\lambda z - \Lambda(\lambda)\} \leq \sup_{\lambda \in \mathbb{R}} \{\lambda z - \Lambda_p(\lambda)\}.$$

This yields $\Lambda^*(z) = I_p(z)$, which ends the proof of the Lemma. \square

Next, note that X is an irreducible, critical multitype Galton-Watson tree with offspring law

$$\mathbb{Q}\{c | b\} = p(n) \prod_{i=1}^n Q\{a_i | b\}, \text{ for } c = (n, a_1, \dots, a_n). \quad (4.7)$$

We derive Theorem 2.4 from Theorems 2.1 and 2.3 by applying the contraction principle to the continuous linear mapping $F : \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) \rightarrow \mathbb{R}^{\mathcal{X} \times \mathcal{X}}$, defined by

$$F(\varpi, \nu)(a, b) = \varpi(a, b), \text{ for all } (\varpi, \nu) \in \mathcal{M}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{X}^*) \text{ and } a, b \in \mathcal{X}. \quad (4.8)$$

It is easy to see that on $\{|T| = n\}$ we have $L_X = \frac{n}{n-1} F(\tilde{L}_X, M_X) = \frac{n}{n-1} \tilde{L}_X$. It follows that conditioned on $\{|T| = n\}$ the random variables L_X are exponentially equivalent to \tilde{L}_X , hence L_X satisfy the same large deviation principle as $F(\tilde{L}_X, M_X)$, see [7, Theorem 4.2.13]. Without loss of generality we restrict the space for the large deviation principle of L_X to the set of all probability vectors on $\mathcal{X} \times \mathcal{X}$, see [7, Lemma 4.1.5(b)].

Specifically, we consider the cases where $p(n)$ has (i) Bounded support and (ii) Unbounded support and finite second moment, separately.

Bounded support. Suppose $k < \infty$ and p has support $\{1, 2, \dots, k\}$. Then, Theorem 2.3 implies the large deviation principle for $F(\tilde{L}_X, M_X)$ conditioned on $\{|T| = n\}$ with the good rate function $I(\mu) = \inf\{J_k(\mu, \nu) : F(\mu, \nu) = \mu\}$, see for example [7, Theorem 4.2.1]. Convexity of I follows easily from the linearity of F and convexity of J_k .

To prove (2.10), we recall that ν is consistent if and only if $F(\mu, \nu)(a, b) = \sum_{c \in \mathcal{X}_k^*} m(b, c)\nu(a, c)$ for all $a, b \in \mathcal{X}$. Hence, we have that

$$I(\mu) = \inf \left\{ H(\nu \parallel \nu_1 \otimes \mathbb{Q}) : F(\varpi, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}_k^*} m(\cdot, c)\nu(\cdot, c), \nu_1 = \mu_2 \right\}. \quad (4.9)$$

Note that $\nu_1(a) = 0$ yields $\sum_b F(\mu, \nu)(a, b) = 0$. Hence if $\mu_1(a) > 0 = \mu_2(a)$ for some $a \in \mathcal{X}$ then $\left\{ \nu : F(\mu, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}_k^*} m(\cdot, c)\nu(\cdot, c), \nu_1 = \mu_2 \right\}$ is an empty set, and therefore $I(\mu) = \infty$. We assume

hereafter that $\mu_1 \ll \mu_2$. Then, it is not uneasy to verify that

$$I(\mu) = \sum_{a \in \mathcal{X}} \mu_2(a) \widehat{I}\left(\frac{\mu(a, \cdot)}{\mu_2(a)}, \mathbb{Q}\{\cdot | a\}\right), \quad (4.10)$$

where for $q \in \mathcal{M}(\mathcal{X}_k^*)$, $\widehat{I}\left(\frac{\mu(a, \cdot)}{\mu_2(a)}, q\right) = \inf \left\{ \tilde{I}(\phi, q) : \phi : \mathcal{X} \rightarrow \mathbb{R}_+, \phi(a) = \frac{\mu(a, \cdot)}{\mu_2(a)}, \text{ for all } a \in \mathcal{X} \right\}$ and

$$\tilde{I}(\phi, q) := \inf \left\{ H(\tilde{\nu} \parallel q) : \tilde{\nu} \in \mathcal{M}(\mathcal{X}_k^*), \phi(b) = \sum_{c \in \mathcal{X}_k^*} m(b, c)\tilde{\nu}(c) \text{ for all } b \in \mathcal{X} \right\}. \quad (4.11)$$

Suppose now that $q(c) = p(n) \prod_{i=1}^n \widehat{q}(a_i)$ for all $c = (n, a_1, \dots, a_n)$, where $\widehat{q}(\cdot)$ is a probability vector on \mathcal{X} and $p(\cdot)$ a probability measure with mean one on $\{0, 1, 2, \dots, k\}$. Then, by Lemma 4.1, we have the representation

$$\tilde{I}(\phi, q) = \|\phi\| H(\phi / \|\phi\| \parallel \widehat{q}) + I_p(\|\phi\|), \quad (4.12)$$

where $\|\phi\| := \sum_{b \in \mathcal{X}} \phi(b)$. Writing $\phi(b) = \frac{\mu(a, b)}{\mu_2(a)}$ in 4.12 we obtain (2.10) which proves the theorem in the case of p with bounded support.

Unbounded support and finite second moment. Suppose p has unbounded support and finite second moment. Then, Theorem 2.1 implies the large deviation principle for $F(\tilde{L}_X, M_X)$ conditioned on $\{|T| = n\}$ with the good rate function $I(\mu) = \inf\{J(\mu, \nu) : F(\mu, \nu) = \mu\}$, see for example [7, Theorem 4.2.1]. Convexity of I follows easily from the linearity of F and convexity of J .

Turning to the proof of (2.10), recall that ν is sub-consistent if and only if $F(\varpi, \nu)(a, b) \geq \sum_{c \in \mathcal{X}_k^*} m(b, c)\nu(a, c)$ for all $a, b \in \mathcal{X}$. Hence, we have that

$$I(\mu, \nu) = \inf \left\{ H(\nu \parallel \nu_1 \otimes \mathbb{Q}) : F(\mu, \nu)(\cdot, \cdot) \geq \sum_{c \in \mathcal{X}_k^*} m(\cdot, c)\nu(\cdot, c), \nu_1 = \mu_2 \right\}. \quad (4.13)$$

Note that $\nu_1(a) = 0$ yields $\sum_b F(\nu)(a, b) = 0$ if $F(\mu, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}_k^*} m(\cdot, c)\nu(\cdot, c)$ and

$$\sum_{(b, c) \in \mathcal{X}^*} m(b, c)\nu(a, c) < 0 \text{ if } F(\mu, \nu)(\cdot, \cdot) > \sum_{c \in \mathcal{X}_k^*} m(\cdot, c)\nu(\cdot, c).$$

Hence if $\mu_1(a) > 0 = \mu_2(a)$ for some $a \in \mathcal{X}$ then $\left\{ \nu : F(\mu, \nu)(\cdot, \cdot) = \sum_{c \in \mathcal{X}_k^*} m(\cdot, c)\nu(\cdot, c), \nu_1 = \mu_2 \right\} \cup \left\{ \nu : F(\varpi, \nu)(\cdot, \cdot) > \sum_{c \in \mathcal{X}_k^*} m(\cdot, c)\nu(\cdot, c), \nu_1 = \mu_2 \right\}$ is an empty set, and therefore $I(\mu) = \infty$. Assuming,

throughout the rest of the proof that $\mu_1 \ll \mu_2$, it is not uneasy to verify that

$$I(\mu) = \sum_{a \in \mathcal{X}} \mu_2(a) \widehat{I}\left(\frac{\mu(a, \cdot)}{\mu_2(a)}, \mathbb{Q}\{\cdot | a\}\right), \quad (4.14)$$

where for $q \in \mathcal{M}(\mathcal{X}^*)$, $\widehat{I}\left(\frac{\mu(a, \cdot)}{\mu_2(a)}, q\right) = \inf \left\{ \widetilde{I}(\phi, q) : \phi : \mathcal{X} \rightarrow \mathbb{R}_+, \phi(a) \leq \frac{\mu(a, \cdot)}{\mu_2(a)}, \text{ for all } a \in \mathcal{X} \right\}$ and

$$\widetilde{I}(\phi, q) := \inf \left\{ H(\tilde{\nu} \| q) : \tilde{\nu} \in \mathcal{M}(\mathcal{X}^*), \phi(b) = \sum_{c \in \mathcal{X}^*} m(b, c) \tilde{\nu}(c) \text{ for all } b \in \mathcal{X} \right\}. \quad (4.15)$$

Suppose now that $q(c) = p(n) \prod_{i=1}^n \widehat{q}(a_i)$ for all $c = (n, a_1, \dots, a_n)$, where $\widehat{q}(\cdot)$ is a probability vector on \mathcal{X} and $p(\cdot)$ a probability measure with mean one on the nonnegative integers, whose second moment is finite. Then, by Lemma 4.1, we have the representation

$$\widetilde{I}(\phi, q) = \|\phi\| H(\phi / \|\phi\| \| \widehat{q}) + I_p(\|\phi\|), \quad (4.16)$$

where $\|\phi\| := \sum_{b \in \mathcal{X}} \phi(b)$. Therefore, it suffice for us to show that

$$\inf \left\{ \widetilde{I}(\phi, \widehat{q}) : \phi : \mathcal{X} \rightarrow \mathbb{R}_+, \phi(b) \leq \frac{\mu(a, b)}{\mu_2(a)}, \text{ for all } a \in \mathcal{X} \right\} = \frac{\mu_1(a)}{\mu_2(a)} H\left(\frac{\mu(a, \cdot)}{\mu_1(a)} \| \widehat{q}\right) + I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right). \quad (4.17)$$

To do this, we write

$$h(\phi(b)) := \widetilde{I}(\phi, \widehat{q}) + \alpha(b) \left(\phi(b) - \frac{\mu(a, b)}{\mu_2(a)} \right), \text{ for } b \in \mathcal{X},$$

where α is a lagrange multiplier. Then, elementary calculus shows that $\alpha(b)$ is the solution of the equation

$$I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right) - \frac{\mu(a, b)}{\mu_2(a)} \sum_{a \in \mathcal{X}} e^{-\alpha(a)} \widehat{q}(a) = 0$$

and that $\phi(b) = \frac{\mu(a, b)}{\mu_2(a)}$ is the minimizer of our constraint optimization problem. Writing $\phi(b) = \frac{\mu(a, b)}{\mu_2(a)}$ in 4.16 we obtain left side of (4.17) which proves the theorem in case of p with unbounded support and finite second moments.

4.3 Proof of Corollary 2.5. Recall that T is Galton-Watson tree with offspring law $p(\ell) = 2^{-(\ell+1)}$, $\ell = 0, \dots$. Also, we recall that X is markov chain indexed by T with arbitrary initial distribution and transition kernel Q . Then, X satisfies all assumptions of Theorem 2.4, in particular we have $\sum_{\ell=0}^{\infty} \ell^2 p(\ell) = 3 < \infty$. Therefore, by Theorem 2.4, L_X conditioned on the events $\{|T| = n\}$ satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X})$ with good, convex rate function

$$I(\mu) = \begin{cases} H(\mu \| \mu_1 \otimes Q) + \sum_{a \in \mathcal{X}} \mu_2(a) I_p\left(\frac{\mu_1(a)}{\mu_2(a)}\right) & \text{if } \mu_1 \ll \mu_2, \\ \infty & \text{otherwise,} \end{cases} \quad (4.18)$$

where $I_p(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x + \log(2 - e^\lambda)\}$. Elementary Calculus shows that

$$\sup_{\lambda \in \mathbb{R}} \{\lambda x + \log(2 - e^\lambda)\} = x \log x - (x + 1) \log \frac{(x+1)}{2}. \quad (4.19)$$

Therefore, writing (4.19) in (4.18) and rearranging terms we obtain the form of the rate function in the corollary which completes the proof.

BIBLIOGRAPHY

- [1] D. ALDOUS. The continuum random tree II: An overview. In: *Stochastic analysis*, Proc. Symp., Durham/UK 1990, Lond. Math. Soc. Lect. Note Ser. 167 (1991), 23–70.
- [2] I. BENJAMINI AND Y. PERES. Markov chains indexed by trees. *Ann. Probab.* 22 (1994), 219–243.
- [3] K. DOKU-AMPONSAH. Large deviations and basic information theory for hierarchical and networked data structures. PhD Thesis, Bath 2006.
- [4] K. DOKU-AMPONSAH. Asymptotic equipartition properties for hierarchical and networked structures. (to appear) *Esiam: Probability and Statistics*.
- [5] K. DOKU-AMPONSAH AND P. MÖRTERS. Large deviation principle for empirical measures of coloured random graphs. (to appear) *Annals of applied probability*.
- [6] A. DEMBO, P. MÖRTERS AND S. SHEFFIELD. Large deviations of Markov chains indexed by random trees. *Ann. Inst. Henri Poincaré: Probab.et Stat.*41, (2005) 971-996.
- [7] A. DEMBO AND O. ZEITOUNI. Large deviations techniques and applications. Springer, New York, (1998).
- [8] A. MEIR AND J.W. MOON. On the altitude of nodes in random trees. *Canad. J. Math.* 30 (1978), 997–1015.
- [9] R. PEMANTLE. Tree-indexed processes. *Statist. Sci.* 10 (1995), no. 2, 200–231.