GRAPHON-VALUED PROCESSES WITH VERTEX-LEVEL FLUCTUATIONS

PETER BRAUNSTEINS, FRANK DEN HOLLANDER, AND MICHEL MANDJES

Abstract. We consider a class of graph-valued stochastic processes in which each vertex has a type that fluctuates randomly over time. Collectively, the paths of the vertex types up to a given time determine the probabilities that the edges are active or inactive at that time. Our focus is on the evolution of the associated empirical graphon in the limit as the number of vertices tends to infinity, in the setting where fluctuations in the graph-valued process are more likely to be caused by fluctuations in the vertex types than in the outcomes of the edges given these types. We derive both sample-path large deviation principles and convergence of stochastic processes. Our approach is flexible because we can including a class of stochastic processes with an additional layer of dependence between the edges.

Key words. Graphs, graphons, dynamics, sample paths, process limits, large deviations, optimal paths.
MSC2010. 05C80, 60C05, 60F10.
Acknowledgment. The work in this paper was supported by the Netherlands Organisation for Scientific Research (NWO) through Gravitation-grant NETWORKS-024.002.003.

1. Introduction

1.1. **Background.** Crane [13], [14], [15] was the first to develop a mathematically comprehensive theory to characterise the evolution of a class of time-varying countably-infinite graphs and graphons. His starting point was the Aldous-Hoover theory for infinitely exchangeable arrays, augmented with the assumption of càdlàg paths and the Markov properties. His construction led to processes that are a combination of a stochastic jump process and a deterministic flow on the space of graphons. It did *not* lead to diffusion-like processes with paths of unbounded variation. In fact, such processes cannot be obtained through exchangeable càdlàg Markov processes after projecting them onto the space of graphons, and therefore require a relaxation of Crane's conditions. For an overview on this line of work we refer to [9].

The approach in [1] was to work directly with graphs and their graphon limits, with the aim to provide a *proof of concept* that diffusion-like graphon-valued processes can be built as well. In [1], a natural class of graphon-valued processes was constructed that arose from population genetics. The construction considered finite populations where individuals carry one of finitely many genetic types and change type according to Fisher-Wright resampling. At any time, each pair of individuals is linked by an edge with a probability that is given by a type-connection matrix, whose entries depend on the current types of the two individuals

Date: January 19, 2022.

and on the current empirical type distribution of the entire population via a fitness function. It was shown that, in the large-population-size limit and with an appropriate scaling of time, the evolution of the associated adjacency matrix converges to a random process in the space of graphons, driven by the type-connection matrix and the underlying Fisher-Wright diffusion on the multi-type simplex. In the limit as the number of types tends to infinity, the limiting process is driven by the type-connection kernel and the underlying Fleming-Viot diffusion.

1.2. **Motivation.** The goal of the present paper is to study a class of dense graph-valued stochastic processes where the edges turn on and off in a dependent manner. In this class, each vertex is assigned a type that changes randomly over time, and fluctuations in the types of the vertices determine how the edges interact. Specifically, the edges in the random graph at time t are independent given the paths of the types of all vertices up to time t. Collectively, these paths are called the *driving process*. Our results generalise those of [1] in a number of directions:

(i) We consider a general driving process and edge-switching dynamics, whereas [1] restricts attention to a specific driving process (the multi-type Moran model) and to a specific edge-switching dynamics.
(ii) We establish stochastic process convergence in the space of $(\mathscr{W}, d_\square)$-valued càdlàg paths, whereas [1] works in the space of $(\tilde{\mathscr{W}}, \delta_\square)$-valued Skorokhod paths. (For the definition of these two spaces, see Section 2.1 below.)
(iii) We establish sample-path large deviations, whereas [1] restricts attention to diffusion limits.

Our proofs rely on concentrations estimates, coupling arguments and continuous mapping. Along the way several examples are presented. See [30] for an example involving a multi-graphon setting built on top of the configuration model.

In [1] diffusion limits are established for the multi-type Moran model in order to derive a diffusion limit for certain graphon-valued stochastic process in $D([0,T], (\tilde{\mathscr{W}}, \delta_\square))$. In the present paper we retain a general framework, while establishing both sample-path large deviation principles and convergence of stochastic processes. For reasons given in later, we prove convergence in $D([0,T], (\mathscr{W}, d_\square))$. As in [1], we restrict attention to processes that exhibit vertex-level fluctuations. In future work we will look at processes that exhibit edge-level fluctuations.

1.3. **Outline.** In Section 2 we recall basic LDPs for graphons and present three LDPs for what we call *inhomogeneous random graphs with type dependence* (IRGwTPs) subject to a number of assumptions. In Section 3 we look at graph-valued processes and present a *sample-path LDP* in graphon space subject to a number of assumptions. We illustrate our results via a running example, and also derive *convergence* of the graph-valued process to a graphon process. In Section 4 we describe various applications and possible extensions. Section 5 contains the proofs of the various LDPs and convergence results. Appendix A identifies the rate function in the LDP of the underlying driving process.

2. Large deviations for static random graphs

In Section 2.1 we recall a few basic definitions. In Section 2.2 we introduce inhomogeneous Erdős-Rényi random graphs and recall the large deviation principle for their associated empirical graphons. In Section 2.3 we describe a generalisation of inhomogeneous Erdős-Rényi random graphs, referred to as inhomogeneous random graphs with type dependence (IRGwTP), which motivate the definition of the class of graph-valued stochastic processes introduced in Section 3. In Section 2.4 we state a number of key assumptions that are needed along the way. In Section 2.5 we establish the large deviation principle for the associated empirical graphon processes under the assumption that the driving process satisfies the LDP. The latter assumption is investigated in Appendix A.

2.1. **Graphs and graphons.** Let $\mathscr{W}$ be the space of functions $h\colon [0,1]^2 \to [0,1]$ such that $h(x,y) = h(y,x)$ for all $(x,y) \in [0,1]^2$, formed after taking the quotient with respect to the equivalence relation of almost everywhere equality. A finite simple graph G on n vertices can be represented as a graphon $h^G \in \mathscr{W}$ by setting

$$h^G(x,y) := \begin{cases} 1 & \text{if there is an edge between vertex } \lceil nx \rceil \text{ and vertex } \lceil ny \rceil, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

This object is referred to as an *empirical graphon* and has a block structure. The space of graphons $\mathscr{W}$ is endowed with the *cut distance*

$$d_\square(h_1,h_2) := \sup_{S,T \subseteq [0,1]} \left| \int_{S\times T} \mathrm{d}x\,\mathrm{d}y\,[h_1(x,y) - h_2(x,y)] \right|, \quad h_1,h_2 \in \mathscr{W}. \tag{2.2}$$ {cutdist}

The space $(\mathscr{W}, d_\square)$ is not compact.

On $\mathscr{W}$ there is a natural equivalence relation, referred to as '$\sim$'. Letting $\mathscr{M}$ denote the set of measure-preserving bijections $\sigma\colon [0,1] \to [0,1]$, we write $h_1 \sim h_2$ when there exists a $\sigma \in \mathscr{M}$ such that $h_1(x,y) = h_2(\sigma(x),\sigma(y))$ for all $(x,y) \in [0,1]^2$. This equivalence relation induces the quotient space $(\tilde{\mathscr{W}}, \delta_\square)$, where $\delta_\square$ is the *cut metric* defined by

$$\delta_\square(\tilde{h}_1,\tilde{h}_2) := \inf_{\sigma_1,\sigma_2 \in \mathscr{M}} d_\square(h_1^{\sigma_1}, h_2^{\sigma_2}), \quad \tilde{h}_1,\tilde{h}_2 \in \tilde{\mathscr{W}}. \tag{2.3}$$

The space $(\tilde{\mathscr{W}}, \delta_\square)$ *is* compact [26, Lemma 8].

2.2. **Inhomogeneous Erdos–Renyi random graph.** Let $r \in \mathscr{W}$ be a *reference graphon.* Fix $n \in \mathbb{N}$ and consider a random graph $\widehat{G}_n$ with vertex set $[n] := \{1,\dots,n\}$, where the pair of vertices $i,j \in [n]$, $i \neq j$, is connected by an edge with probability $r(\frac{i}{n},\frac{j}{n})$, independently of other pairs of vertices. Write $\mathbb{P}_n$ to denote the law of $\widehat{G}_n$. Use the same symbol to denote the law on $\mathscr{W}$ induced by the map that associates the graph $\widehat{G}_n$ with its graphon $h^{\widehat{G}_n}$. Write $\tilde{\mathbb{P}}_n$ to denote the law of $\tilde{h}^{\widehat{G}_n}$, the equivalence class associated with $h^{\widehat{G}_n}$.

The following theorem is an extension of the LDP for homogeneous Erdős-Rényi random graphs in [12]. It was first stated in [18] under additional assumptions. These assumptions were subsequently relaxed in [29], [5], [19]. The following theorem corresponds to [19, Theorem 4.1].

Theorem 2.1. *The sequence of probability measures* $(\tilde{\mathbb{P}}_n)_{n\in\mathbb{N}}$ *satisfies the LDP on* $(\tilde{\mathscr{W}}, \delta_\square)$ *with rate* $\binom{n}{2}$ *and with rate function* $\tilde{I}_r$, *i.e.,*

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\binom{n}{2}} \log \tilde{\mathbb{P}}_n(\mathcal{C}) \leqslant -\inf_{\tilde{h}\in\mathcal{C}} \tilde{I}_r(\tilde{h}) \qquad \forall\, \mathcal{C} \subseteq \tilde{\mathscr{W}} \text{ closed},\\
&\liminf_{n\to\infty} \frac{1}{\binom{n}{2}} \log \tilde{\mathbb{P}}_n(\mathcal{O}) \geqslant -\inf_{\tilde{h}\in\mathcal{O}} \tilde{I}_r(\tilde{h}) \qquad \forall\, \mathcal{O} \subseteq \tilde{\mathscr{W}} \text{ open},
\end{aligned} \tag{2.4}$$

where

$$\tilde{I}_r(\tilde{h}) = \inf_{\sigma\in\mathscr{M}} I_r(h^\sigma), \tag{2.5}$$

h *is any representative of* $\tilde{h}$, *and*

$$I_r(h) := \int_{[0,1^2]} \mathrm{d}x\, \mathrm{d}y\; \mathcal{R}(h(x,y) \mid r(x,y)), \tag{2.6}$$

with

$$\mathcal{R}(a \mid b) = a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}. \tag{2.7}$$

2.3. **Inhomogeneous random graphs with type dependence.** Consider the following generalisation of the inhomogeneous Erdős-Rényi random graph defined in Section 2.2. Suppose that each vertex $i \in [n]$ is assigned a (possibly random) type $X_i^{(n)} \in [0,1]$. Denote the *empirical type measure* by

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i^{(n)}} \tag{2.8}$$

and the *empirical type distribution* by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}[X_i^{(n)} \leqslant x], \tag{2.9}$$

where I is the indicator function. Let $\mathcal{M}([0,1])$ denote the space of measures on $[0,1]$ endowed with the topology of weak convergence. Suppose that each edge ij is active with probability $H(X_i, X_j, F_n)$ *independently* of all other edges given F_n, where $H\colon\, [0,1]^2 \times \mathcal{M}([0,1]) \to [0,1]$ is symmetric in its first two inputs. We label the resulting sequence of random graphs as $\{G_n\}_{n\in\mathbb{N}}$, and refer to them as *inhomogeneous random graphs with type dependence* (IRGwTP).

Observe that if

{eq:ERcon}

$$X_i^{(n)} = \frac{i}{n} \quad \forall\, n \in \mathbb{N},\, i \in [n], \qquad H(x,y,F) = r(x,y) \quad \forall\, x,y \in [0,1], \tag{2.10}$$

then the IRGwTP is equivalent to the inhomogeneous Erdős-Rényi random graph defined in Section 2.2. There is a further connection between IRGwTP and inhomogeneous Erdős-Rényi random graphs. Let $\bar{F}$ denote the right-continuous generalised inverse of a distribution function F with support $[0,1]$, which is defined in the usual way as

$$\bar{F}_n(u) := \inf\{x \in [0,1]\colon\, F_n(x) > u\}, \qquad u \in [0,1). \tag{2.11}$$

For $F \in \mathcal{M}([0,1])$, define the *induced reference graphon* $g^{[F]} \in \mathscr{W}$ by

{eqn:SRGd}

$$g^{[F]}(x,y) = H(\bar{F}(x), \bar{F}(y), F). \tag{2.12}$$

Given the type distribution F_n, $\tilde{h}^{\widehat{G}_n}$ has the same distribution as an inhomogeneous Erdős-Rényi random graph with reference graphon $g^{[F_n]}$. In other words, we have

{ob:SS}

$$\tilde{h}^{\widehat{G}_n} \mid F_n \overset{d}{=} \tilde{h}^{G_n} \text{ with } r = g^{[F_n]}. \tag{2.13}$$

This observation is central to the large deviation principle for IRGwTPs stated in Theorem 2.6 below.

2.4. **Key assumptions.** Before stating this theorem we make a number of assumptions.

Assumption 2.2. The sequence of type distributions $(F_n)_{n\in\mathbb{N}}$ satisfies the LDP on $\mathcal{M}([0,1])$ with rate $\ell(n)$ and with rate function K. ◊

Assumption 2.2 holds, for instance, when $(X_i^{(n)})_{n\in\mathbb{N}, i\in[n]}$ are i.i.d. random variables with distribution f, in which case $\ell(n) = n$ and $K(g) = H(g \mid f)$, the relative entropy of g with respect to f. Assumption 2.2 may also hold with $\ell(n) \neq n$. For example, if $p \geqslant 0$, $\{Y_{ij}^{(n)}\}_{n\in\mathbb{N}, i\in[n], j\in[n^p]}$ are i.i.d. random variables and

$$X_i^{(n)} = \frac{1}{n^p}\sum_{j=1}^{\lfloor n^p \rfloor} Y_{ij}^{(n)}, \tag{2.14}$$

then $\ell(n) = n^{1+p}$ and $K(g) =$ [Insert.]. See [19, Example 2.5] for an example where $\ell(n) = n^2$ and $X_i^{(n)}$, $i \in \mathbb{N}$, are dependent. When $X_i^{(n)}$, $i \in \mathbb{N}$, are fixed and $\mu_n \to \mu$ in $\mathcal{M}([0,1])$, as in (2.10), then this can be viewed as satisfying Assumption 2.2 with $\ell(n) = \infty$.

Assumption 2.3. The function $F \mapsto g^{[F]}$ defined in (2.12) is a continuous mapping from $\mathcal{M}([0,1])$ to $(\mathscr{W}, \|\cdot\|_{L_1})$. ◊

Assumption (2.3) holds, for example, when $H(x,y,F) \equiv H^*(x,y)$ (i.e., there is no dependence on the type distribution) and $H^*\colon [0,1]^2 \to [0,1]$ is a continuous function. Assumption (2.3) also holds when, in addition, $f\colon \mathcal{M}([0,1]) \to [0,1]$ and $h\colon [0,1]^2 \to [0,1]$ are continuous functions, and

$$H(x,y;F) = h(H^*(x,y), f(F)) \qquad \forall\ [x,y] \in [0,1]^2,\ F \in \mathcal{M}([0,1]). \tag{2.15}$$

In certain settings we require two further assumptions that are of a more technical nature.

Assumption 2.4. For all $F \in \mathcal{M}([0,1])$ the induced graphon $g^{[F]}$ is *away from the boundary*, i.e., there exists $\eta > 0$ such that

$$\eta \leqslant g^{[F]}(x,y) \leqslant 1-\eta \qquad \forall\ (x,y) \in [0,1]^2. \tag{2.16}$$

◊

Assumption 2.5. The rate function K has a unique zero, labelled F^*. ◊

2.5. **LDP for IRGwTPs.** We are now ready to state our large deviation principle for IRGwTPs.

Theorem 2.6. *Subjects to Assumptions 2.2 and 2.3 the following hold:*

(i) *If* $\ell(n) = o\left(\binom{n}{2}\right)$, *then* $\{\tilde{h}^{\widehat{G}_n}\}$ *satisfies the LDP with rate* $\ell(n)$ *and with rate function* $I^*(\tilde{h}) = J(\tilde{h})$.

(ii) *If* $\lim_{n\to\infty} \ell(n)/\binom{n}{2} = c$ *and Assumption 2.4 holds as well, then* $\{\tilde{h}^{\widehat{G}_n}\}$ *satisfies the LDP with rate* $\binom{n}{2}$ *and with rate function* $I^*(\tilde{h}) = \inf_{g\in\tilde{\mathscr{W}}}[cJ(\tilde{g}) + I_g(\tilde{h})]$, *where* g *is any representative and* $\tilde{g}$.

(iii) *If* $\binom{n}{2} = o(\ell(n))$ *and Assumptions 2.4 and 2.5 hold as well, then* $\{\tilde{h}^{\widehat{G}_n}\}$ *satisfies the LDP with rate* $\binom{n}{2}$ *and with rate function* $I^*(\tilde{h}) = I_{g^{[F^*]}}(\tilde{h})$.

To understand where Theorem 2.6 comes from, think of simulating outcomes of $\tilde{h}^{\widehat{G}_n}$ in two steps:

- In Step 1 simulate types of the vertices, i.e., simulate the type distribution F_n.
- In Step 2 simulate the edges given F_n, i.e., simulate $\tilde{h}^{\widehat{G}_n}$ given the induced reference graphon $g^{[F_n]}$.

Due to Assumption 2.2, large fluctuations in Step 1 are governed by the LDP with rate $\ell(n)$ and with rate function $K(\cdot)$, whereas due to (2.13) large fluctuations in Step 2 are governed by the LDP with rate $\binom{n}{2}$ and with rate function $I_{g^{[F_n]}}$. Consequently, when $\ell(n) = o\left(\binom{n}{2}\right)$ large fluctuations in $\tilde{h}^{G_n}$ are most likely to be caused by a rare event in Step 1, whereas when $\binom{n}{2} = o(\ell(n))$ large fluctuations in $\tilde{h}^{G_n}$ are most likely to be caused by a rare event in Step 2. When $\ell(n) \asymp \binom{n}{2}$ large fluctuations in $\tilde{h}^{G_n}$ are most likely to be caused by a combination of rare events in Steps 1 and 2.

When $\ell(n) = o\left(\binom{n}{2}\right)$ we say that the IRGwTP exhibits *vertex-level fluctuations*, whereas when $\binom{n}{2} = o(\ell(n))$ we say that it exhibits *edge-level fluctuations.*

The type-dependent inhomogeneous random graphs introduced in this section are of interest in their own right. However, our primary motivation for introducing IRGwTP is that it has a natural stochastic process generalisation. The rough idea behind this generalisation is that at each time the distribution of the process corresponds to a IRGwTP. We will focus primarily on processes that exhibit vertex-level fluctuations.

3. Graphon-valued processes

Section 3.1 introduces the graph-valued process of interest. Section 3.2 describes an illustrative example. Section 3.3 presents the sample-path LDP for the graph-valued process under the assumption that the driving process satisfies the LDP. The latter assumption is investigated in Appendix A.

3.1. **The model.** Let $(G_n(t))_{t\in[0,T]}$ denote our graph-valued process. We suppose that each vertex $i \in [n]$ has a type $X_i^{(n)}(t)$ that may fluctuate randomly over time. Let $(\mu_n(t))_{t\in[0,T]}$ denote the *empirical type measure process*, which is characterised by

$$\mu_n(t) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^n(t)}, \tag{3.1}$$

and let $(F_n(t;\cdot))_{t\in[0,T]}$ denote the *empirical type distribution process*, which is characterised by

$$F_n(t;x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}[X_i(t) \leqslant x], \qquad \forall x \in [0,1]. \tag{3.2}$$

The process $(\mu_n(t))_{t\in[0,T]}$ is a random variable on $D([0,T],\mathcal{M}([0,1]))$, the space of $\mathcal{M}([0,1])$-valued Cadlag paths.bWe suppose that at any time t edge ij is active with probability

$$H(t; X_i(t), X_j(t), (F_n(t;\cdot))_{t\in[0,T]}), \tag{3.3}$$

independently of all other edges given t, $X_i(t)$, $X_j(t)$, and $(F_n(t;\cdot))_{t\in[0,T]}$, where

$$H\colon\ [0,T]\times[0,1]^2\times D([0,T],\mathcal{M}([0,1])) \mapsto [0,1]. \tag{3.4}$$

The function H gives rise to the *induced reference graphon process* $g^{[F]}$, which, for $F \in D([0,T],\mathcal{M}([0,1]))$, is characterised by

$$g^{[F]}(t;x,y) = H(t;\bar{F}(t;x),\bar{F}(t;y),(F(t;\cdot)_{t\in[0,T]})). \tag{3.5}$$ {eq:Gtdef}

Observe that, for any $t \in [0,T]$, given the outcome of the empirical type distribution process F_n, the distribution of $\tilde{h}^{G_n(t)}$ corresponds to that of an inhomogeneous Erdős-Rényi random graph with reference graphon $g^{[F_n]}(t;\cdot,\cdot)$. In other words, for any $t \in [0,T]$,

$$h^{G_n(t)}|F_n \overset{d}{=} h^{\widehat{G}_n} \text{ with } r = g^{[F_n]}(t;\cdot), \tag{3.6}$$ {Ob:spIRG}

where $\widehat{G}_n$ is the inhomogeneous Erdős-Rényi random graph defined in Section 2.2. We make the following assumption on the function $F \mapsto g^{[F]}$, which due to (3.5) is an assumption on H.

Assumption 3.1. The map $F \mapsto g^{[F]}$ from $D([0,T],\mathcal{M}([0,1]))$ to $D([0,T],(\mathscr{W},\|\cdot\|_{L_1}))$ is continuous. ♢

3.2. **An illustrative example.** Suppose that $(G_n(t))_{t\in[0,T]}$ is characterised by the following dynamics:

- $G_n(0)$ is the empty graph.
- Each vertex is assigned an independent rate-γ Poisson clock, and each time the clock attached to vertex v rings all the edges that are adjacent to v become inactive.
- If edge ij is inactive, then it becomes active at rate λ.

We first describe the driving process. Let $\{\tau_k(v)\}_{k\in\mathbb{N}}$ denote the sequence of times at which the Poisson clock attached to vertex v rings, and let

$$Y_v(t) := t - \max_k\{\tau_k(v)\colon\ \tau_k(i) \leqslant t\} \tag{3.7}$$

denote the time since the clock last rung. The value of $Y_v(t)$ can be thought of as the *age of vertex v at time t*: each time the clock associated with v rings, it 'dies' and all its adjacent edges are lost. Recalling that we assumed that types take values in $[0,1]$, we let

$$X_v(t) := F^{\exp}(Y_i(t)) = 1 - \mathrm{e}^{-\gamma Y_i(t)} \tag{3.8}$$

denote the *type of vertex v at time t*, where $F^{\exp}$ can be identified as the distribution function of an exponential random variable with rate γ.

The function $H(t,v,u,F)$ can also be identified. The probability that there is an active edge between vertices of ages $\bar{u}$ and $\bar{v}$ is $1-\exp\{-(\bar{u}\wedge\bar{v})\}$. Letting $u=F^{\exp}(\bar{u})$ and $v=F^{\exp}(\bar{v})$, we obtain

$$H(t,u,v,F)=1-(1-u\wedge v)^{\lambda/\gamma}. \tag{3.9}$$

Because $H(t,u,v,F)$ is a continuous function of u and v, and is independent of t and F, it is simple to verify that Assumption 3.1 holds. A more involved example is given in Section 4.1.

3.3. **Sample-path large deviations.** Similarly as in Section 2.5, we assume that the driving process satisfies the LDP (which for the illustrative example is established in Lemma A.1).

Assumption 3.2. $\{F_n\}_{n\in\mathbb{N}}$ satisfies the LDP on $D([0,T],\mathcal{M}([0,1]))$ with rate $\ell(n)=o(\binom{n}{2})$ and with rate function K. ♢

To establish the sample-path LDP for the graphon-valued process, we need to: (I) establish the LDP in the pointwise topology; (II) strengthen the topology by establishing exponential tightness. Step (I) is settled by the following result.

Proposition 3.3. *If Assumptions 3.1 and 3.2 hold, then* $\{(\tilde{h}^{\widehat{G}_n(t)})_{t\geqslant 0}\}_{n\in\mathbb{N}}$ *satisfies the LDP in the pointwise topology with rate* $\ell(n)$ *and with rate function* $J(\tilde{h})$.

Note that Proposition 3.3 does not refer to any edge-switching dynamics. Specifically, if two process $\{G_n\}_{n\in\mathbb{N}}$ and $\{G_n^*\}_{n\in\mathbb{N}}$ have a common sequence of types $\{(X_i(t))_{t\geqslant 0}\}_{i\in[n]}$ and a common edge connection function H, then the marginal distributions are equivalent, i.e.,

$$\tilde{h}^{G_n(t)}\overset{d}{=}\tilde{h}^{G_n^*(t)},\qquad \text{for any } t\in[0,T]. \tag{3.10}$$

However, this does *not* necessarily mean that the the joint distributions are equivalent, i.e., we may have

$$(\tilde{h}^{\widehat{G}_n(t)})_{t\in[0,T]}\overset{d}{\neq}(\tilde{h}^{\widehat{G}_n^*(t)})_{t\geqslant[0,T]}, \tag{3.11}$$

as these joint distributions depend on the specific edge-switching dynamics. Nonetheless, Proposition 3.3 implies that both $\{\widehat{G}_n\}_{n\in\mathbb{N}}$ and $\{\widehat{G}_n^*\}_{n\in\mathbb{N}}$ satisfy equivalent LDPs in the pointwise topology, i.e., the rate function depends only on the marginal distributions of the process and not on the specific edge-switching dynamics. In Sections 4.1 and 4.2 we provide examples of processes with equivalent marginals and different edge-switching dynamics.

The specific edge-switching dynamics *do* need to be taken into consideration when we want to strengthen the topology of the LDP in Proposition 3.3 by establishing exponential tightness. We next provide a condition that can be used to verify that $\{\tilde{h}^{G_n}\}_{n\geqslant 0}$ are exponentially tight. Let

$$E_{ij}^{(n)}(t)=\begin{cases}1, & \text{if edge } ij \text{ is active at time } t,\\ 0, & \text{otherwise,}\end{cases} \tag{3.12}$$

and define

$$C^{(n)}(t,\delta)=\sum_{1\leqslant i<j\leqslant n}\sup_{t\leqslant u\leqslant v\leqslant t+\delta}|E_{ij}^{(n)}(u)-E_{ij}^{(n)}(v)|. \tag{3.13}$$

In other words, $C_n(t,\delta)$ is the number of edges that change (i.e., go from active to inactive or from inactive to active) at some time between t and $t+\delta$.

Proposition 3.4. *If, for all $t \in [0,T]$ and $\varepsilon > 0$,*

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\frac{1}{\ell(n)}\log\mathbb{P}\left(C_n(t,\delta) > \varepsilon\binom{n}{2}\right) = -\infty, \tag{3.14}$$ {eq:Tcon}

then $\{\tilde{h}^{G_n}\}_{n\geqslant 0}$ is exponentially tight.

Combining the above results, we obtain the following.

Theorem 3.5. *If the conditions of Propositions 3.3 and 3.4 hold, then the sequence of processes $\{\tilde{h}^{\widehat{G}_n}\}_{n\in\mathbb{N}}$ satisfies the LDP on $D([0,T],\tilde{\mathscr{W}})$ with rate $\ell(n)$ and with rate function J.*

In view of Lemma A.1, the conditions of Theorem 3.5 can be readily verified for the illustrative example. In Theorem 3.10 we establish a sample-path LDP for a class of processes that includes the illustrative example.

3.4. **Stochastic process convergence.** Let $\Rightarrow$ denote convergence in distribution, and $\overset{f.d.d.}{\Rightarrow}$ denote convergence of the finite-dimensional distributions. We assume that the empirical type distribution satisfies a stochastic process limit.

Assumption 3.6. Suppose that $F_n \Rightarrow F$ on $D(\mathcal{M}([0,1]),[0,T])$. ♢

We establish the stochastic process limit of $(h^{G_n(t)})_{t\in[0,T]}$ on $D((\mathscr{W},d_\square),[0,T])$, i.e., we no longer take the quotient with respect to the equivalence relation $\sim$. To establish a stochastic process limit in this finer topology, we need to ensure that the labels of the vertices update dynamically.

Assumption 3.7. *At any time $t\in[0,1]$ the labels of the vertices are such that*

$$X_1(t) \leqslant X_2(t) \leqslant \cdots \leqslant X_n(t). \tag{3.15}$$

♢

Given the dynamic labelling above and the illustrative example, the motivation behind establishing our stochastic process limits on $D((\mathscr{W},d_\square),[0,T])$ rather than on $D((\tilde{\mathscr{W}},\delta_\square),[0,T])$ is clear: Are the older vertices more connected than the younger vertices? If we establish a limit on $D((\mathscr{W},d_\square),[0,T])$, then we have a definitive answer, whereas if we establish a limit in $D((\tilde{\mathscr{W}},\delta_\square),[0,T])$, then we do not gain any insight.

Proposition 3.8. *If Assumptions 3.1, 3.6 and 3.7 hold, then $h^{G_n}\overset{f.d.d.}{\Rightarrow} r[F]$.*

Strengthening the topology to obtain convergence in distribution on $D((\mathscr{W},d_\square),[0,T])$ is more difficult than in Section 3.3. This is because, unlike $(\tilde{\mathscr{W}},\delta_\square)$, the space $(\mathscr{W},d_\square)$ is not Polish, and hence we cannot directly apply established sufficient conditions for tightness, such as those stated in [20, Sections 3.6–3.9]. Nonetheless, we are able to establish convergence directly by using [20, Corollary 3.3].

Assumption 3.9. *For any* $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \sup_{t\in[0,T]} \frac{T}{\delta} \mathbb{P}\left(C_n(t,\delta) > \varepsilon \binom{n}{2} \right) = 0. \tag{3.16}$$

Theorem 3.10. *Subject to Assumptions 3.1, 3.6–3.7, 3.9,* $h^{G_n} \Rightarrow g^{[F]}$ *in* $D((\mathscr{W}, d_\square), [0,T])$.

In view of Lemma A.1, the conditions of Theorem 3.10 can again be readily verified for the illustrative example. We establish a more general result in Proposition 4.2.

4. Applications and extensions

In this section we consider a class of processes that generalise the illustrative example. We use this class of processes to make three points that we believe apply more generally:

(I) An additional layer of dependence between the edges can be introduced that cannot be captured by the types of the vertices (so that (3.6) no longer holds) but still allows to establish limiting results in the spirit of Section 3. Roughly speaking, this is the case when the additional layer of dependence between the edges is of mean-field type (see Section 4.1).

(II) The specific edge-switching dynamics rarely affects the limiting path of the process (see Section 4.2).

(III) The dependence between edges in inhomogeneous random graphs with type dependence leads to new behaviour in the corresponding variational problems, even in relatively simple settings (see Section 4.3).

4.1. **Beyond conditional independence of edges.**

4.1.1. *Model and LDP.* Suppose that $(G_n(t))_{t\in[0,T]}$ is characterised by the following dynamics:

- $G_n(0)$ is the empty graph.
- Each vertex is assigned an independent rate-λ Poisson clock, and each time the clock associated with vertex v rings all the edges that are adjacent to v become inactive.
- If edge ij is inactive, then it becomes active at rate $\lambda(t, X_i(t), X_j(t), F_n(t;\cdot), \tilde{h}^{G_n(t)})$.
- If edge ij is active, then it becomes inactive at rate $\mu(t, X_i(t), X_j(t), F_n(t;\cdot), \tilde{h}^{G_n(t)})$.

Here, we assume that λ and μ are Lipshitz-continuous functions on $[0,T]\times[0,1]^2\times([0,1])\times \tilde{\mathscr{W}}$.

If $\lambda(\cdot)$ and $\mu(\cdot)$ do not depend on the current state of the unlabelled graph ($\tilde{h}^{G_n(t)}$), i.e., if

{eq:FF} $$\lambda(t,u,v,F,\tilde{h}) \equiv \lambda(t,u,v,F), \qquad \mu(t,u,v,F,\tilde{h}) \equiv \mu(t,u,v,F), \tag{4.1}$$

then the process fits into the framework of Section 3, otherwise it does not. To understand why, we compute the probability that edge ij is active under (4.1) at time t given $X_i(t) = x_i$,

$X_j(t) = x_j$ and $F_n = F$. We have

$$H(t; x_i, x_j, F) = \int_{t - x_i \wedge x_j}^{t} \mathrm{d}s \, \lambda(s, x_i - t + s, x_j - t + s, F(t; \cdot))$$
$$\times \exp\left\{ - \int_s^t \mathrm{d}a \, [\mu(s, x_i - t + a, x_j - t + a, F(t; \cdot)) + \lambda(s, x_i - t + a, x_j - t + a, F(t; \cdot))] \right\}. \tag{4.2}$$

{eq:HFF}

It is also easy to see that two edges ij and $k\ell$ are independent given t, $X_i(t)$, $X_j(t)$, $X_k(t)$, $X_\ell(t)$ and $F_n = F$, and hence the process indeed falls into the framework of Section 3. To recover the illustrative example, take $\lambda(t, u, v, F) = \lambda \in \mathbb{R}_+$ and $\mu(t, u, v, F) = 0$.

An example of a choice for $\lambda(\cdot)$ that *does* depend on the unlabelled graph ($\tilde{h}^{G_n(t)}$) is

$$\lambda(t, u, v, F, \tilde{h}) = 1 + s(G, \tilde{h}), \tag{4.3}$$

{eq:lex}

where G is a simple graph (e.g. a triangle) and $s(G, \tilde{h})$ denotes the homomorphism density of G in $\tilde{h}$. Note that, by the counting lemma (see, for instance, [10, Proposition 2.2]), this particular choice of $\lambda(\cdot)$ is Lipshitz-continuous. In this case, the two edges ij and $k\ell$ are *not* independent given t, $X_i(t)$, $X_j(t)$, $X_k(t)$, $X_\ell(t)$ and $F_n = F$. Indeed, if edge ij is active, then it may participate in additional copies of G, which means that edge $k\ell$ is more likely to be active. This dependence is inherent to the model, in that it cannot be removed by changing the definition of the types $X_i(t)$.

The next theorem demonstrates that, despite the above observation, we can still establish a sample-path LDP for the process. To show why, we define a mapping $F \mapsto g^{(F)}$ via a differential equation. We first state this differential equation and then explain the intuition behind it. Let $g^{(F)}(0; x, y) = 0$ for $(x, y) \in [0, 1]^2$, and

$$g^{(F)}(t + \mathrm{d}t; x, y) = g^{(F)}(t; x', y') + \mathrm{d}t \lambda(t, \bar{F}(t; x'), \bar{F}(t; y'), F(t; \cdot), \tilde{g}^{(F)})(1 - g^{(F)}(t; x', y'))$$
$$- \mu(t, \bar{F}(t; x'), \bar{F}(t; y'), F(t; \cdot), \tilde{g}^{(F)}) g^{(F)}(t; x', y'), \tag{4.4}$$

{eq:GFdef}

where $u' = F(t; \bar{F}(t + \mathrm{d}t; u) - \mathrm{d}t)$ for all $u \in [0, 1]$. Note that, under (4.1),

$$g^{(F)}(t; x, y) = H(t, \bar{F}(t, x), \bar{F}(t, y), F) \tag{4.5}$$

with H given in (4.2), and hence the more complicated expression in (4.4) is only required when (4.1) does not hold. The differential equation in (4.4) can be understood as follows. Consider the process h^{G_n} under Assumption 3.7, so that the vertices are labelled in order of increasing age. Roughly speaking, given $F_n = F$, $g^{(F)}(t; x, y)$ can be thought of as the probability that edge $(\lceil nx \rceil, \lceil ny \rceil)$ is active. Indeed, as time $t - \mathrm{d}t$ these vertices had the labels $(\lceil nx' \rceil, \lceil ny' \rceil)$, respectively (with x', y' given below (4.4)). The first term in the right-hand side of (4.4), $g^{(F)}(t; x', y')$, is the probability that the edge was active at time $t - \mathrm{d}t$, the second term accounts for the event that the edge turned on during the time interval $[t - \mathrm{d}t, t]$, while the third term accounts for the event that it turned off during the time interval $[t - \mathrm{d}t, t]$.

For $A \subseteq \mathbb{R}_+$ (such that if $A = [a, b]$ then $A + c = [a + c, b + c]$), let

$$D_{\mathbf{1}} \mu_t(A) = \lim_{h \to 0} \frac{\mu_{t+h}(A + h) - \mu_t(A)}{h}. \tag{4.6}$$

Theorem 4.1. *The sequence of processes* $\{(\tilde{h}^{G_n(t)})_{t\geqslant 0}\}_{n\in\mathbb{N}}$ *satisfies the LDP with rate* n *and with rate function*

{eq:RFex}
$$J(\tilde{h}) = \inf_{F\in D([0,t]\times\mathcal{M}([0,1])):\ \tilde{g}^{(F)}=\tilde{h}} K(F), \tag{4.7}$$

where

{eq:rateex1alt}
$$\begin{aligned} K(\mu) = &\int_0^T \mathrm{d}t \int_0^\infty [\gamma\mu_i(\mathrm{d}x) - D_{\mathbf{1}}\mu_t(\mathrm{d}x)] + \int_0^T \mathrm{d}t \int_0^\infty D_{\mathbf{1}}\mu_t(\mathrm{d}x) \log\left(\frac{D_{\mathbf{1}}\mu_t(\mathrm{d}x)}{\mu_i(\mathrm{d}x) f_t(0)}\right) \\ &+ \int_0^T \mathrm{d}t f_t(0)\log(f_t(0)/\gamma) \end{aligned} \tag{4.8}$$

with $f_t(0) := \lim_{h\to 0}\mu_t([0,h])/h$.

Proposition 4.2. **[To be written: A functional law of large numbers in the Skorohod space $D((\mathscr{W}, d_\square), [0,T])$. The proof uses the same arguments as in the proof of Theorem 4.1.**

Theorem 4.1 is interesting because it shows that we can add a 'mean-field type' interaction between the edges and still obtain equivalent results.

4.1.2. *Numerical illustration.* **[To be added.]**

4.2. **Different edge-switching dynamics, equivalent sample-path LDP.** Next suppose that the process $\tilde{h}^{G_n(\cdot)}$ has the same underlying driving process, but different edge-switching dynamics. In particular, consider the edge switching dynamics introduced in [1]. Let $(U_{ij})_{1\leqslant i<j\leqslant n}$ be a sequence of independent uniform variables on $[0,1]$, and suppose that edge ij is active if $U_{ij} \leqslant H(t; X_i(t), X_j(t), F_n(t;))$, where H is given by (4.1). **[Check if this can be generalised.]** Note that this process is not Markov. Nonetheless, by Proposition 3.3 we immediately have that, in the pointwise topology, the sequence of processes satisfies the same LDP as the processes described in Section 4.2. To strengthen the topology it remains to verify establish exponential tightness. This can be done by using Proposition 3.4, which leads to the following result. **[Proof still needs to be done.]**

Proposition 4.3. *The sequence of processes* $\{(\tilde{h}^{\widehat{G}_n(t)})_{t\geqslant 0}\}_{n\in\mathbb{N}}$ *satisfies the LDP with rate* n *and with rate function* (4.7).

4.3. **The most likely path to an unusually small edge density.** Consider the simple setting with $\mu(t,u,v,F,\tilde{h}) = 0$, $\lambda(t,u,v,F,\tilde{h}) \equiv \lambda \in \mathbb{R}_+$, and $\lambda = \gamma$. Observe that in this case

{eq:43S}
$$H(u,v,F) = u \wedge v. \tag{4.9}$$

Our goal is to find the most likely path the process takes to a prescribed edge density e^* at time T. We do this in two stages: we first compute the most likely state of the process at time T given this edge density, and afterwards use this computation to obtain the most likely trajectory of the process. Note that, given (4.9), the results in this section apply to the models introduced in Sections 4.1 and 4.2 (with the above simplification).

4.3.1. *Step 1: most likely state of the process at time T.* Let Q denote the distribution of $X_v(t)$. Note that

{eq:Qex}
$$Q(\mathrm{d}x) = \begin{cases} \mathrm{d}x, & \text{if } x < F^{\exp}(T), \\ 1 - F^{\exp}(T), & \text{if } x = F^{\exp}(T), \\ 0, & \text{otherwise.} \end{cases} \tag{4.10}$$

However, for the moment we will assume that Q is a general measure on $[0,1]$. We first consider the event that G_n has an unusually small edge density e^* at time T. By Theorem 2.6, the corresponding variational problem is

$$\begin{array}{ll} \text{minimize} & \int_0^1 \log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}P \\ \text{subject to} & 2\int_0^1 \int_0^y xP(\mathrm{d}x)P(\mathrm{d}y) \leqslant \mathrm{e}^* \\ \text{over} & P \in \mathcal{M}([0,1]). \end{array} \tag{4.11}$$
{eq:OP}

Proposition 4.4. *The feasible region of the variational problem described in* (4.11) *is convex.*

Because the objective function in (4.11) is strictly convex, Proposition 4.4 implies that the variational problem has a unique *global* minimum. Consequently, there are a number of numerical methods that we can apply to obtain this minimum.

If we consider the probability that G_n has an unusually high edge density at time T, then we must solve the same variational problem as described in (4.11) with "$\leqslant$" replaced by "$\geqslant$". Now the feasible region is no longer convex. Consequently, as we illustrate with a numerical example, the corresponding variational problem may have multiple local maxima and multiple local minima.

Let

$$Q(x) = \begin{cases} \frac{4}{5} - \frac{1}{1000}, & \text{if } x = 0, \\ \frac{1}{5}, & \text{if } x = \frac{1}{10}, \\ \frac{1}{1000}, & \text{if } x = 1. \end{cases} \tag{4.12}$$
{eq:NumQ}

In Figure 1 we plot the rate (the value of the objective function evaluated at the maxima) against the edge density e^*. When $e^* = 0.085$ there are two distinct optimal solutions corresponding to

$$P_1^*(x) = \begin{cases} 0.0782, & \text{if } x = 0, \\ 0.9159, & \text{if } x = \frac{1}{10}, \\ 0.0059, & \text{if } x = 1, \end{cases} \qquad P_2^*(x) = \begin{cases} 0.3728, & \text{if } x = 0, \\ 0.4020, & \text{if } x = \frac{1}{10}, \\ 0.2252, & \text{if } x = 1. \end{cases}$$

For $e^* \approx 0.085$, solutions near P_1^* and P_2^* are local minima. These local minima are illustrated by the dotted curve in Figure 1: values above 0.085 correspond to solutions that are close to P_1^* and values below 0.085 correspond to solutions that are close to P_2^*. Observe that we can restrict our search of an optimal measure P to measures that are absolutely continuous with respect to Q. For Q given by (4.12), these measures live on the 2-dimensional simplex. Consequently, there can be at most two local minima. If Q has a continuous component (as it does in (4.10)), then there is, in principle, no bound on the number of local minima.

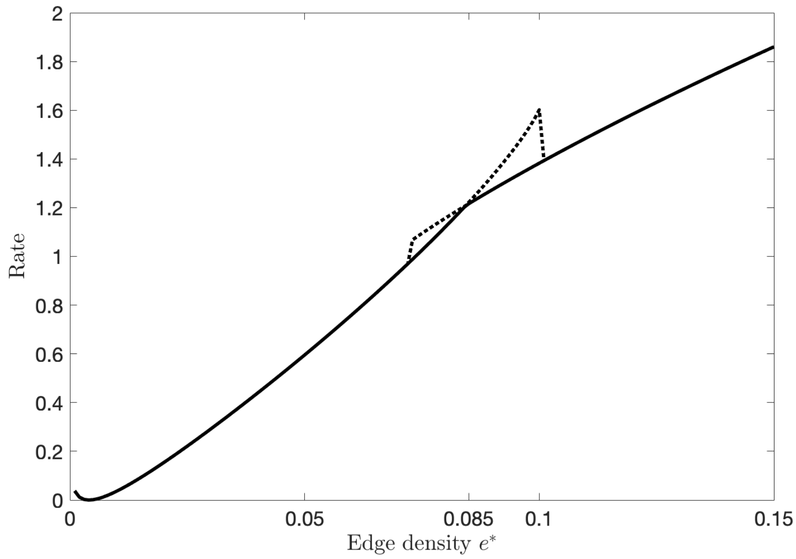

Figure 1.

4.3.2. *Step 2: Computing the optimal path.* **[Not sure whether this is worth doing. The interesting part is Step 1.]**

5. Proofs

Sections 5.1–5.3 contain the proofs of the theorems in Sections 2–4.

5.1. **Proofs of the results in Section 2.** We prove Theorem 2.6 via a sequence of lemmas. Recall that we use $\widehat{G}_n$ to denote an inhomogeneous Erdős–Rényi random graph (IRG) and G_n to denote a inhomogeneous random graph with type dependence (IRGwTP).

5.1.1. *Inhomogeneous Erdős–Rényi random graphs.* The first lemma is similar to [18] and [19, Theorem 4.1], the primary difference being that in these papers there is a single reference graphon r, i.e., $r_n = r$ for all $n \geqslant 0$. The generalisation comes at the cost of the addition of Assumption 2.4 in the lower bound (which is not made in [19, Theorem 4.1], but is in [18]).

Lemma 5.1. *Let r_n denote the reference graphon for $\widehat{G}_n$ and suppose that $r_n \to r$ in L^1. Then*

$$\limsup_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}(\tilde{h}^{\widehat{G}_n} \in \mathcal{C}) \leqslant - \inf_{\tilde{h}\in\mathcal{C}} \tilde{I}_r(\tilde{h}), \quad \forall\, \mathcal{C} \text{ closed}, \tag{5.1}$$

and, subject to Assumption 2.4,

$$\liminf_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}(\tilde{h}^{\widehat{G}_n} \in \mathcal{O}) \geqslant - \inf_{\tilde{h}\in\mathcal{O}} \tilde{I}_r(\tilde{h}), \quad \forall\, \mathcal{O} \text{ open}. \tag{5.2}$$

Proof. The upper bound follows from the same arguments as used in [19, Theorem 4.1] (by noting that the specific requirement that $r_n = r$ is not used there). The lower bound again follows from similar arguments. However, now instead of applying Jensen's inequality we apply the dominated convergence theorem, which is possible because of Assumption 2.4. □

The previous lemma can be used to obtain the following concentration type result for inhomogeneous Erdős–Rényi random graphs.

Lemma 5.2. *Let $\widehat{G}_n$ be an IRG with reference graphon $r_n \in \mathscr{W}_n$. If $\|r_n - r\|_{L_1} \to 0$, then, for any $r \in \mathscr{W}$,*

$$\limsup_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}(\tilde{h}^{\widehat{G}_n} \notin \mathbb{B}_\square(\tilde{r}, \varepsilon)) \leqslant -\varepsilon^2. \tag{5.3}$$

Proof. Suppose that $\tilde{h} \notin \mathbb{B}_\square(r, \varepsilon)$, and let h be any member of the equivalence class $\tilde{h}$. Using a Taylor expansion in the first step and Jensen's inequality in the second step, we have

$$\begin{aligned} I_r(h) &= \frac{1}{2}\int_{[0,1]^2} \mathrm{d}x\,\mathrm{d}y \left[h(x,y)\log\left(\frac{h(x,y)}{r(x,y)}\right) + (1-h(x,y))\log\left(\frac{1-g_n(x,y)}{1-r(x,y)}\right)\right] \\ &\geqslant \int_{[0,1]^2} \mathrm{d}x\,\mathrm{d}y\,(h(x,y)-r(x,y))^2 \geqslant \|h-r\|_{L_1}^2 \geqslant d_\square(h,r)^2 \geqslant \varepsilon^2. \end{aligned} \tag{5.4}$$

Since $\mathbb{B}_\square(\tilde{r}, \varepsilon)$ is open, its complement is closed, which implies that we can apply the upper bound in Lemma 5.1, from which the result follows. □

5.1.2. *Inhomogeneous random graphs with type dependence.* We next turn our attention to inhomogeneous random graphs with type dependence G_n. We first use the previous lemma to show that G_n is close to the induced reference graphon $g^{[F_n]}$ with high probability.

Lemma 5.3. *If Assumption 2.3 holds, then*

$$\limsup_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}(\tilde{h}^{G_n} \notin \tilde{\mathbb{B}}_\square(\tilde{g}^{[F_n]}, \varepsilon)) \leqslant -\varepsilon^2. \tag{5.5}$$

{eq:LemEF}

Proof. Suppose that Assumption 2.3 holds. We proceed by contradiction. Suppose that (5.5) does not hold. Then there necessarily exist sequences $(n_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$ and $(F^\star_{n_k})_{k\in\mathbb{N}} \subset \mathcal{M}([0,1])$ such that

$$\liminf_{k\to\infty} \frac{1}{\binom{n_k}{2}} \log \mathbb{P}\left(\tilde{h}^{G_{n_k}} \notin \tilde{\mathbb{B}}_\square(\tilde{g}^{[F^\star_{n_k}]}, \varepsilon) \mid F_{n_k} = F^\star_{n_k}\right) > -\varepsilon^2, \tag{5.6}$$

where, for each $k \in \mathbb{N}$, $F^\star_{n_k}$ is an empirical distribution function with n_k data points. Since $\mathcal{M}([0,T])$ is compact, there exists a convergent subsequence of $(F^\star_{n_k})_{k\in\mathbb{N}}$. Consequently, w.l.o.g. we may assume that there exists $F^\star$ such that $F^\star_{n_k} \to F^\star$ in $\mathcal{M}([0,T])$ as $k \to \infty$. Under Assumption 2.3 we therefore have

$$\|r^{[F^\star_{n_k}]} - r^{[F^\star]}\|_{L_1} \to 0, \qquad \text{as } k \to \infty. \tag{5.7}$$

Recalling that, due to 2.13 (i.e., conditional on the induced reference graphon the graph has the distribution of an inhomogeneous random graph), we can apply Lemma 5.2 to obtain a contradiction. □

The next lemma establishes a large deviation principle for the sequence of induced reference graphons $g^{[F_n]}$.

Lemma 5.4. *Subject to Assumptions 2.2 and 2.3, $\{\tilde{g}^{[F_n]}\}_{n\in\mathbb{N}}$ satisfies the LDP on $(\tilde{\mathscr{W}}, \delta_\square)$ with rate $\ell(n)$ and with rate function*

$$J(\tilde{h}) = \inf_{F\in\mathcal{M}([0,1])\,:\,\tilde{g}^{[F]}=\tilde{h}} K(F). \tag{5.8}$$ {RateLDPT}

Proof. For any $g, f \in \mathscr{W}$, $\|f-g\|_{L_1} \geqslant \delta_\square(\tilde{g},\tilde{f})$. Thus, under Assumption 2.2, the map $F \mapsto \tilde{g}^{[F]}$ is continuous. The result therefore follows by using Assumption 2.3 and the contraction principle [24, Theorem III.20]. □

Proof of Theorem 2.6: We prove (i), (ii), (iii) separately.

<u>(i) lower bound:</u> Let $\mathcal{O}$ be an open subset of $\tilde{\mathscr{W}}$, and let $\mathcal{O}^{(-\varepsilon)}$ denote the largest open set whose ε-neighbourhood is contained in $\mathcal{O}$. We have

{eqn:LBi}
$$\mathbb{P}(\tilde{h}^{G_n}\in\mathcal{O}) \geqslant \mathcal{P}(\tilde{g}^{[F_n]}\in\mathcal{O}^{(-\varepsilon)})(1-\mathbb{P}(\delta_\square(\tilde{h}^{G_n},\tilde{g}^{[F_n]})>\varepsilon)). \tag{5.9}$$

Applying Lemmas 5.4 and 5.3 to the first and second terms on the right-hand-side of (5.9), respectively, we obtain

$$\limsup_{n\to\infty}\frac{1}{\ell(n)}\log\mathbb{P}(\tilde{h}^{G_n}\in\mathcal{O}) \geqslant -\lim_{\varepsilon\downarrow 0}\inf_{\tilde{h}\in\mathcal{O}^{(-\varepsilon)}} J(\tilde{h}) = -\inf_{\tilde{h}\in\mathcal{O}} J(\tilde{h}). \tag{5.10}$$

<u>(i) upper bound:</u> Let $\mathcal{C}$ be a closed subset of $\tilde{\mathscr{W}}$, and let $\mathcal{C}^{(+\varepsilon)}$ denote the largest closed set that contains the ε-neighbourhoods of all the points in $\mathcal{C}$. We have

{eqn:UBi}
$$\mathbb{P}(\tilde{h}^{G_n}\in\mathcal{C}) \geqslant \mathcal{P}(\tilde{g}^{[F_n]}\in\mathcal{C}^{(+\varepsilon)}) + \mathbb{P}(\delta_\square(\tilde{h}^{G_n},\tilde{g}^{[F_n]})>\varepsilon). \tag{5.11}$$

Again applying Lemmas 5.4 and 5.3 to the first and second terms on the right-hand-side of (5.9), respectively, we obtain

$$\limsup_{n\to\infty}\frac{1}{\ell(n)}\log\mathbb{P}(\tilde{h}^{G_n}\in\mathcal{C}) \leqslant -\lim_{\varepsilon\downarrow 0}\inf_{\tilde{h}\in\mathcal{C}^{(+\varepsilon)}} J(\tilde{h}) = -\inf_{\tilde{h}\in\mathcal{C}} J(\tilde{h}), \tag{5.12}$$

where in the first step we use the fact that $\ell(n) = o(\binom{n}{2})$.

<u>(ii) lower bound:</u> For $r \in \mathscr{W}$, let

$$F(r,\varepsilon) = \{F\in\mathcal{M}([0,1]) : \|g^{[F]}-r\|_{L_1}<\varepsilon)\}, \tag{5.13}$$

and observe that, by Assumption 2.3, the set $F(r,\varepsilon)$ is measurable. Under Assumption 2.4 we therefore have

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{1}{\binom{n}{2}}\log\mathbb{P}(\tilde{h}^{G_n}\in\mathcal{O})\\
&\geqslant \lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty}\frac{1}{\binom{n}{2}}\Big[\log\mathbb{P}(F_n\in F(r,\varepsilon)) + \log\mathbb{P}(\tilde{h}^{\hat{G}_n}\in\mathcal{O}|F_n\in F(r,\varepsilon))\Big]\\
&\geqslant -\lim_{\varepsilon\downarrow 0}\Big[cK(F(r,\varepsilon)) + \sup_{F\in F(r,\varepsilon)}\inf_{\tilde{h}\in\mathcal{O}} I_{r^{[F]}}(\tilde{h})\Big]\\
&= -[cJ(\tilde{r}) + \inf_{\tilde{h}\in\mathcal{O}} I_r(\tilde{h})].
\end{aligned} \tag{5.14}$$

In the second step we use the fact that, since $\mathcal{M}([0,T])$ is compact, for any sequence $(F_n)_{n\in\mathbb{N}}$ in $F(r,\varepsilon)$ there exists a convergent subsequence, which allows us to apply Lemma 5.1. In the final step we use the lower semi-continuity of K and the assumption that $F \mapsto r^{[F]}$ is a continuous mapping from $\mathcal{M}([0,1])$ to $(\mathscr{W}, L_1)$, in combination with the fact that, under Assumption 2.4, if $\|r_n - r\|_{L_1} \to 0$, then $I_{r_n}(\tilde{h}) \to I_r(\tilde{h})$ uniformly over $h \in \mathscr{W}$ as $n \to \infty$ (see [18, Lemma 2.3]). Because these arguments hold for any $r \in \mathscr{W}$ we have

$$\liminf_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}(\tilde{h}^{G_n} \in \mathcal{O}) \geqslant -\inf_{\tilde{r}\in\tilde{\mathscr{W}}}[cJ(\tilde{r}) + \inf_{\tilde{h}\in\mathcal{O}} I_r(\tilde{h})] = -\inf_{\tilde{h}\in\mathcal{O}}\{\inf_{\tilde{r}\in\tilde{\mathscr{W}}}[cJ(\tilde{r}) + I_r(\tilde{h})]\}. \quad (5.15)$$

<u>(ii) upper bound:</u> Let $L(\cdot,\cdot)$ be the Lévy metric, let $B_L(F,\varepsilon) = \{H \in \mathcal{M}([0,1])\colon\ L(H,F) \leqslant \varepsilon\}$, and recall that $L(\cdot,\cdot)$ metrises the weak topology. Since $\mathcal{M}([0,1])$ is a compact space, for any $\varepsilon > 0$ we can construct a finite set $F[\varepsilon]$ with the property that for any $H \in \mathcal{M}([0,1])$ there exists $F \in F[\varepsilon]$ such that $L(F,H) \leqslant \varepsilon$. We therefore have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}(\tilde{h}^{\hat{G}_n} \in \mathcal{C}) \\
&\quad\leqslant \lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty} \frac{1}{\binom{n}{2}} \log\Bigg[\sum_{F\in F[\varepsilon]} \mathbb{P}(F_n \in B_L(F,\varepsilon))\mathbb{P}(\tilde{h}^{\hat{G}_n} \in \mathcal{C} \mid F_n \in B_L(F,\varepsilon))\Bigg] \\
&\quad\leqslant -\lim_{\varepsilon\downarrow 0}\min_{F\in F[\varepsilon]}[cK(B_L(F,\varepsilon)) + \inf_{F^\star\in B_L(F,\varepsilon)}\inf_{\tilde{h}\in\mathcal{C}} I_{g^{[F^\star]}}(\tilde{h})]. \\
&\quad\leqslant -\lim_{\varepsilon\to 0}\min_{F\in\mathcal{M}([0,1])}[cK(B_L(F,\varepsilon)) + \inf_{F^\star\in B_L(F,\varepsilon)}\inf_{\tilde{h}\in\mathcal{C}} I_{g^{[F^\star]}}(\tilde{h})] \\
&\quad= -\min_{F\in\mathcal{M}([0,1])}[cK(F) + \inf_{\tilde{h}\in\mathcal{C}} \tilde{I}_r^{[F]}(\tilde{h})] \\
&\quad= \inf_{\tilde{h}\in\mathcal{C}}\{\min_{F\in\mathcal{M}([0,T])}[cK(F) + \tilde{I}_r^{[F]}(\tilde{h})]\}.
\end{aligned} \quad (5.16)$$

In the second step we apply Assumption 2.2, Lemma 5.1 and Laplace's method, using a similar justification as in the lower bound. In the fourth step we use lower semi-continuity of K (Assumption 2.2) and apply [18, Lemma 2.3]).

<u>(iii):</u> In this case we can apply similar (albeit simpler) arguments as in case (i). □

5.2. **Proofs of the results in Section 3.**

5.2.1. *Large deviations.* The next lemma will be used to prove Proposition 3.3.

Lemma 5.5. *Subject to Assumption 3.1, $\tilde{H}(\cdot;\cdot,\cdot,F_n)$ satisfies the LDP with rate $\ell(n)$ and with rate function*

$$J(\tilde{h}) = \inf_{F\in\mathcal{M}\times[0,T]:\tilde{H}(\cdot;\cdot,\cdot,F)=\tilde{h}} K(F). \quad (5.17)$$ {eq:Jdef}

Proof. The claim follows from the contraction principle (cf. Lemma 5.4). □

Proof of Proposition 3.3. To establish a multi-point LDP, we can follow similar arguments as in the proof of Theorem 2.6 *(i)*. For example, to establish the lower bound, pick

$0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_k \leqslant T$, let $\mathcal{O}_i$ be an open subset of $\tilde{\mathscr{W}}$, and let $\mathcal{O}_i^{(-\varepsilon)}$ be as in the proof of Theorem 2.6. We have

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{\ell(n)} \log \mathbb{P}\left(\tilde{h}^{G_n(t_i)} \in \mathcal{O}_i,\ \forall i = 1,\ldots,k\right) \\
&\geqslant \lim_{\varepsilon\downarrow 0} \liminf_{n\to\infty} \frac{1}{\ell(n)} \log \Bigg[\mathbb{P}\left(\tilde{g}^{[F_n]}(t_i) \in \mathcal{O}_i^{(-\varepsilon)},\ \forall i = 1,\ldots,k\right) \\
&\qquad\qquad + \left(1 - \sum_{i=1}^{k} \mathbb{P}(\delta_\square(\tilde{h}^{G_n(t_i)}, \tilde{g}^{[F_n]}(t_i)) > \varepsilon\right)\Bigg] \\
&\geqslant \inf_{\tilde{h}:\tilde{h}(t_i)\in\mathcal{O}_i,\,\forall i=1,\ldots,k} J(\tilde{h}),
\end{aligned} \tag{5.18}$$

where $J(\tilde{h})$ is given by (5.17), and we use a similar justification as in the lower bound of Theorem 2.6 *(i)* (applying Lemma 5.5 in place of Lemma 5.4). The upper bound is again similar and is therefore omitted.vWith the multi-point LDP established, we can apply the Dawson-Gärtner ([17, Theorem 4.6.1]) projective limit theorem to establish an LDP in the pointwise topology, and [17, Lemma 4.6.5] to obtain the specific form of the rate function in Proposition 3.3. □

Proof of Proposition 3.4. For $\delta > 0$ and $T > 0$, define the modulus of continuity in $D([0,T], \tilde{\mathscr{W}})$ by

$$w'(\tilde{h}^{G_n}, \delta, T) = \inf_{t_i} \max_{i} \sup_{s,t\in[t_i,t_{i+1})} \delta_\square\left(\tilde{h}^{G_n(s)}, \tilde{h}^{G_n(t)}\right), \tag{5.19}$$

where the infimum is over $\{t_i\}$ satisfying

$$0 = t_0 < t_1 < \cdots < t_{m-1} < T \leqslant t_m \tag{5.20}$$

and $\min_{1\leqslant i\leqslant n}(t_i - t_{i-1}) \geqslant \delta$. By [21, Theorem 4.1] (and the compactness of $\tilde{\mathscr{W}}$), the sequence of processes $\{(h^{G_n(t)})_{t\in[0,T]}\}_{n\in\mathbb{N}}$ are exponentially tight if

{eq:Tgoal}
$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{\ell(n)} \log \mathbb{P}\left(w'(\tilde{h}^{G_n}, \delta, T) > \varepsilon\right) = -\infty \tag{5.21}$$

for all $\varepsilon > 0$. Suppose that (3.14) holds. We will show that this entails that (5.21) holds with $t_i = i\delta$ for $i \in \{0,\ldots,\lceil T/\delta\rceil\}$. Indeed, observe that

$$\sup_{s,t\in[t_i,t_{i+1})} \delta_\square\left(\tilde{h}^{G_n(s)}, \tilde{h}^{G_n(t)}\right) \leqslant \sup_{s,t\in[t_i,t_{i+1})} \|h^{G_n(s)} - h^{G_n(t)}\|_{L_1} \leqslant \binom{n}{2}^{-1} C_n(t_i,\delta). \tag{5.22}$$

Consequently,

{eq:wdid}
$$\mathbb{P}\left(w'(\tilde{h}^{G_n}, \delta, T) > \varepsilon\right) \leqslant \lceil T/\delta\rceil \sup_{t\in[0,T]} \mathbb{P}\left(C_n(t,\delta) > \varepsilon\binom{n}{2}\right). \tag{5.23}$$

Hence

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\frac{1}{\ell(n)}\log\mathbb{P}\left(w'(\tilde{h}^{G_n},\delta,T)>\varepsilon\right)\\
&\quad\leqslant \lim_{\delta\downarrow 0}\limsup_{n\to\infty}\left[\frac{1}{\ell(n)}\log\lceil T/\delta\rceil+\frac{1}{\ell(n)}\log\left(\sup_{t\in[0,T]}\mathbb{P}\left(C_n(t,\delta)>\varepsilon\binom{n}{2}\right)\right)\right] \qquad (5.24)\\
&\quad=-\infty,
\end{aligned}$$

where in the final step we apply (3.14) and use the fact that $\ell(n)\to\infty$. □

5.2.2. *Weak convergence.* The next lemma is needed in the proofs of Proposition 3.8 and Theorem 3.10.

Lemma 5.6. *Subject to Assumptions 3.1 and 3.6,* $g^{[F_n]}\Rightarrow g^{[F]}$ *on* $D((\mathscr{W},d_\square),[0,T])$.

Proof. By Assumptions 3.1 and 3.6, we can apply the continuous mapping theorem to establish that

$$g^{[F_n]}\Rightarrow g^{[F]}\quad\text{on } D((\mathscr{W},\|\cdot\|_{L_1}),[0,T]). \tag{5.25}$$

Because $(\mathscr{W},\|\cdot\|_{L_1})$ is a stronger topology than $(\mathscr{W},d_\square)$, this implies the claim. □

Proof of Proposition 3.8. By Lemma 5.6, we have $g^{[F_n]}\overset{f.d.d.}{\Rightarrow}g^{[F]}$ on $(\mathscr{W},d_\square)$. From (3.6), Assumption 3.7, the uniform bound and [10, Lemma 5.11] we know that, for any $t\in[0,T]$,

$$\lim_{n\to\infty}\mathbb{P}(\delta_\square(h^{G_n(t)},g^{[F_n]}(t))\leqslant\varepsilon)=1, \tag{5.26}$$

which implies that

$$\lim_{n\to\infty}\mathbb{P}(h^{G_n(t_i)}\in\mathbb{B}_\square(g^{[F_n]}(t_i),\varepsilon),\,\forall i)=1. \tag{5.27}$$

The claim therefore follows from [20, Corollary 3.3]. □

Proof of Theorem 3.10. By Lemma 5.6 and [20, Corollary 3.3], it suffices to prove that

$$\|g^{[F_n]}-h^{G_n}\|\to 0,\qquad\text{with probability 1,} \tag{5.28}$$

where $\|\cdot\|$ denotes the uniform norm. Define $g_\delta^{[F_n]}$ such that

$$g_n^{[F_n]}(t)=g(\delta i),\qquad\text{for } t\in[\delta i,\delta(i+1)). \tag{5.29}$$

It suffices to show that

$$\lim_{\delta\downarrow 0}\lim_{n\to\infty}\left[\|g^{[F_n]}-g_\delta^{[F_n]}\|+\|g_\delta^{[F_n]}-h^{G_n}\|\right]=0\qquad\text{with probability 1.} \tag{5.30}$$ {eqn:NGl}

We first deal with the second term on the left-hand-side of (5.30).vBy the same arguments as in the proof of Proposition 3.8, for any $\delta,\varepsilon>0$,

$$\lim_{n\to\infty}\mathbb{P}(h^{G_n(\delta i)}\in\mathbb{B}_\square(g^{[F_n]}(\delta i),\varepsilon)\ \forall\ i=0,1,\ldots,\lfloor T/\delta\rfloor)=0. \tag{5.31}$$ {eqn:CLLN}

Thus,

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\lim_{n\to\infty} \mathbb{P}(\|g_\delta^{[F_n]} - h^{G_n}\| > 2\varepsilon) \\
&\leqslant \lim_{\delta\downarrow 0}\lim_{n\to\infty} \sum_{i=1}^{T/\delta} \left\{ \mathbb{P}(d_\square(g_\delta^{[F_n]}(\delta i), h^{G_n(\delta i)} > \varepsilon) + \mathbb{P}\left(C_n(t,\delta) > \varepsilon \binom{n}{2} \right) \right\} \qquad (5.32)\\
&= 0,
\end{aligned}$$

where in the final step we apply (5.31) and Assumption 3.9. To deal with the first term on the right-hand-side, we use the fact that $g^{[F]}$ takes values in $D((\mathscr{W}, d_\square), [0,T])$. **[Need something slightly more here.]** □

5.3. **Proofs of the results in Section 4.**

5.3.1. *Proofs of the results in Section 4.1.* To prove Theorem 3.10 we construct a graphon-valued process that mimics the behaviour of $(\tilde{h}^{\hat{G}_n(t)})_{t\geqslant 0}$ while still falling into the framework of Section 3. We couple the two processes and demonstrate that, under the coupling, the probability that the two processes deviate from each other significantly is on the same scale as the edge-level fluctuations, i.e., of order $\mathrm{e}^{-\binom{n}{2}+o(1)}$.

Constructing a mimicking process: Suppose that the process $(G_n^*(t))_{t\geqslant 0}$ is characterised by the following dynamics:

- $G_n^*(0)$ is the empty graph.
- Each vertex v is assigned an independent rate-γ Poisson clock. Each time the clock rings, all the edges that are adjacent to v become inactive.
- If edge ij is inactive, then it becomes active at rate
$$\lambda(t, Y_i(t), Y_j(t), F_n(t;\cdot), \tilde{g}^{[F_n]}(t;\cdot,\cdot)).$$
- If edge ij is active, then it becomes inactive at rate
$$\mu(t, Y_i(t), Y_j(t), F_n(t;\cdot), \tilde{g}^{[F_n]}(t;\cdot,\cdot)).$$

Here, $g^{[\cdot]}(t;\cdot,\cdot)$ is defined in (4.4). We point out that the induced reference graphon process of $(G^*(t))_{t\geqslant 0}$ is indeed $g^{[F]}$. Note that the only difference between the transition rates of $(G_n(t))_{t\geqslant 0}$ and $(G_n^*(t))_{t\geqslant 0}$ is that in the transition rate functions $\lambda(\cdot)$ and $\mu(\cdot)$ we have replaced $\tilde{h}^{G_n(t)}$ by $\tilde{g}^{[F_n]}(t;\cdot,\cdot))$.

Theorem 3.10 follows by verifying that we can apply Theorem 3.5 to $\{(G_n(t))_{t\geqslant 0}\}_{n\in\mathbb{N}}$, and using the following lemma.

Lemma 5.7. *There exists a coupling of $(G_n(t))_{t\geqslant 0}$ and $(G_n^*(t))_{t\geqslant 0}$ such that, for any $\eta > 0$,*

$$\limsup_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}\left(\|\tilde{h}^{G_n(t)} - \tilde{h}^{G_n^*(t)}\|_{L_1} > \eta, \text{ for some } t\in[0,T] \right) \geqslant -C(\eta) > 0. \qquad (5.33)$$

Proof. The claim is proved in three steps.

Step 1: describe the coupling. Let C_{max} be the maximal value that $\lambda(\cdot)$ and $\mu(\cdot)$ can take, i.e.,

$$C_{max} = \max_{t\in[0,T],u,v\in[0,1],F\in\mathcal{M}([0,1]),\tilde{h}\in\tilde{\mathscr{W}}} \lambda(t,u,v,F,\tilde{h}) \vee \mu(t,u,v,F,\tilde{h}), \qquad (5.34)$$

and observe that $C_{max} < \infty$ because $\lambda(\cdot)$ and $\mu(\cdot)$ are Lipshitz continuous functions with a compact domain. Suppose that outcomes of $(G_n(t))_{t\geqslant 0}$ and $(G_n^*(t))_{t\geqslant 0}$ are generated in the following manner.

- For each $i \in [n]$, vertex i is assigned the same Poisson clock in both processes, so that if the clock associated with vertex i rings in $(G_n(t))_{t\geqslant 0}$ at time s, then the clock associated with vertex i also rings in $(G_n^*(t))_{t\geqslant 0}$ at time s (and vice-versa).
- Assign each edge the same (coupled) Poisson rate-C_{max} clock. When the Poisson clock associated with edge ij rings, generate an outcome u of a $U([0,1])$ distribution.
  - $u \leqslant \lambda(t, Y_i(t), Y_j(t), F_n, \tilde{h}^{G_n(t)}))/C_{max}$:
    edge ij becomes active in $(\widehat{G}_n(t))_{t\geqslant 0}$.
  - $u \leqslant \lambda(t, Y_i(t), Y_j(t), F_n, \tilde{g}^{[F_n]}(t;\cdot,\cdot))/C_{max}$:
    edge ij becomes active in $(\widehat{G}_n^*(t))_{t\geqslant 0}$.

  (If it was already active, then it remains active.)
- Assign each edge a second (coupled) Poisson rate-C_{max} clock. When the Poisson clock associated with edge ij rings, generate an outcome u of a $U([0,1])$ distribution.
  - $u \leqslant \mu(t, Y_i(t), Y_j(t), F_n, \tilde{h}^{G_n(t)})/C_{max}$:
    edge ij becomes inactive in $(\widehat{G}_n(t))_{t\geqslant 0}$.
  - $u \leqslant \mu(t, Y_i(t), Y_j(t), F_n, \tilde{g}^{[F_n]}(t;\cdot,\cdot))/C_{max}$:
    edge ij becomes inactive in $(\widehat{G}_n^*(t))_{t\geqslant 0}$.

  (If it was already inactive, then it remains inactive.)

Step 2: dominate the L_1 distance. Observe that if edge ij is inactive in both models and the clock associated with an edge ij rings, then a *difference* is formed (i.e., edge ij is active in one process and inactive in the other) with probability

$$\frac{|\lambda(t, Y_i(t), Y_j(t), F_n, \tilde{h}^{G_n(t)})) - \lambda(t, Y_i(t), Y_j(t), F_n, \tilde{g}^{[F_n]}(t;\cdot,\cdot))|}{C_{max}}. \tag{5.35}$$

At any time t, by the Lipshitz continuity of $\lambda(\cdot)$,

$$\begin{aligned} &|\lambda(t, Y_i(t), Y_j(t), F_n, \tilde{h}^{G_n(t)})) - \lambda(t, Y_i(t), Y_j(t), F_n, \tilde{g}^{[F_n]}(t;\cdot,\cdot))| \\ &\quad \leqslant c\left[\delta_{\square}(\tilde{g}^{(F_n)}(t;\cdot,\cdot), \tilde{h}^{G_n^*(t)}) + \delta_{\square}(\tilde{h}^{G_n^*}(t), \tilde{h}^{G_n(t)})\right] \\ &\quad \leqslant c\left[\delta_{\square}(\tilde{g}^{(F_n)}(t;\cdot,\cdot), \tilde{h}^{G_n^*(t)}) + \|h^{G_n^*(t)} - h^{G_n(t)}\|_{L_1}\right], \end{aligned} \tag{5.36}$$ {eqn:rBg}

where c is the Lipshitz constant. Observe that an equivalent bound holds when $\lambda(\cdot)$ is replaced by $\mu(\cdot)$.

To bound the first term on the right-hand-side of (5.36) we verify that, for any $\beta > 0$,

$$-\limsup_{n\to\infty} \frac{1}{\binom{n}{2}} \mathbb{P}\left(\delta_{\square}(\tilde{g}^{(F_n)}(t;\cdot,\cdot), \tilde{h}^{G_n^*(t)}) > \beta, \text{ for some } t \in [0,T]\right) \geqslant C_1(\beta) > 0. \tag{5.37}$$

This can be done by using similar arguments as in the proof of Theorem 3.10. **[Explain.]** Hence we may assume that $\delta_{\square}(\tilde{g}^{(F_n)}(t;\cdot,\cdot), \tilde{h}^{G_n^*(t)}) > \beta$ for all $t \in [0,T]$ and any $\beta > 0$.

To bound the second term on the right-hand-side of (5.36), first note that when the clock associated with a vertex rings it can only eliminate differences because in both processes all edges adjacent to this vertex are then inactive. Consequently, using superposition Poisson processes, we see that the number of differences at time t is dominated by a Markov chain $(Z_n(t))_{t\geqslant 0}$ with $Z_n(0) = 0$ and transition probabilities

{eq:dp}
$$i \mapsto i+1 \quad \text{at rate} \quad 2c\,C_{max}\left[\beta\binom{n}{2} + i\right]. \tag{5.38}$$

In other words, on an event with probability $1 - \mathrm{e}^{-C_1(\beta)\binom{n}{2}+o(1)}$, we have

$$\|h^{G_n^*(t)} - h^{G_n(t)}\|_{L_1} \overset{s.t.}{\leqslant} \binom{n}{2}^{-1} Z_n(t). \tag{5.39}$$

Note that the first term of the right-hand-side of (5.38) corresponds to the first term on the right-hand-side of (5.36), while the second term of the right-hand-side of (5.38) corresponds to the second term on the right-hand-side of (5.36).

Step 3: bound the dominating process. By observing that $(Z_n(t))_{t\in[0,T]}$ is a pure birth process, we see that it remains to show that, for any $\eta > 0$,

$$\limsup_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}\left(Z_n(T) > \eta\binom{n}{2}\right) \geqslant -C(\eta) > 0 \tag{5.40}$$

for some $C(\eta) > 0$. To bound $Z_n(T)$, let $C^* := 2c\,C_{max}$. Observe that the Markov chain $(Z_n(t))_{t\geqslant 0}$ described by (5.38) is a continuous-time branching process with immigration. The initial population size is 0, immigrants arrive at rate $2c\,C_{max}\beta\binom{n}{2}$, while individuals in the population give birth at rate $2cC_{max}$ and die at rate 0. Let $X(t)$ denote of the number of descendants that are alive at time T of an individual that immigrated to the population at time $t < T$. It was shown by Yule (cf. [23, Chapter V.8]) that

$$\mathbb{P}(X_t = i) = \mathrm{e}^{-C^*(T-t)}(1 - \mathrm{e}^{-C^*(T-t)})^{i-1}, \quad i \in \mathbb{N}, \tag{5.41}$$

and 0 otherwise, i.e., $X - 1$ has a geometric distribution with success probability $\mathrm{e}^{-C^*(T-t)}$. Note that (since the death rate of individuals is zero) $X_0 \overset{s.t.}{\geqslant} X_t$ for all $t \geqslant 0$. In addition, the total number of immigrants has a Poisson distribution with mean $C^*\beta\binom{n}{2}T$. Thus, if $\{X_0^{(k,\ell)}\}_{k,\ell\in\mathbb{N}}$ are i.i.d. copies of X_0 and $Y = \sum_{k+1}^{\binom{n}{2}} Y^{(k)}$, where $Y^{(k)} \sim Poi(C^*\beta T)$ is independent of everything else, then

$$Z_n(T) \overset{s.t.}{\leqslant} Z_n^*(T) := \sum_{i=1}^{Y} X_0^{(i,1)} \overset{d}{=} \sum_{i=1}^{\binom{n}{2}} \sum_{k=1}^{Y^{(i)}} X_0^{(i,k)}. \tag{5.42}$$

We have

$$\begin{aligned}\varphi(s) &:= \mathbb{E}\left(\mathrm{e}^{s\sum_{k=1}^{Y^{(i)}} X_0^{(i,k)}}\right) = \mathbb{E}\left(\mathbb{E}\left(\mathrm{e}^{sX_0^{(i,k)}}\right)^{Y^{(i)}}\right) \\ &= \exp\left\{C^*\beta T\left(\frac{\mathrm{e}^{-C^*T+s}}{1-(1-\mathrm{e}^{-C^*T})\mathrm{e}^s} - 1\right)\right\}.\end{aligned} \tag{5.43}$$

Letting $I(z) = \sup_{s\in\mathbb{R}}[zs - \varphi(s)]$, and applying Cramer's theorem, we therefore have

$$\mathbb{P}\left(Z_n^*(T) \geqslant \eta\binom{n}{2}\right) \leqslant \mathrm{e}^{-\binom{n}{2}(I(\eta)+o(1))} \qquad \forall\, z \geqslant \mathbb{E}(Y) = c\beta\binom{n}{2}T\mathrm{e}^{cT}, \tag{5.44}$$

The claim now follows by observing that for any $\eta > 0$ we can select $\beta > 0$ sufficiently small so that $I(\eta) > 0$. □

5.3.2. *Proofs of the results in Section 4.2.* **[To be written.]**

5.3.3. *Proofs of the results in Section 4.3.* .

Proof of Proposition 4.4. Pick $P_1, P_2 \in \mathcal{M}([0,T])$ and suppose that

$$2\int_0^1\int_0^y xP_i(\mathrm{d}x)P_i(\mathrm{d}y) \leqslant e^*, \qquad i=1,2. \tag{5.45}$$ {eq:EDA}

Observe that if $X_i^{(k)}$ are independent random variables with distribution P_i, then

$$2\int_0^1\int_0^y xP_i(\mathrm{d}x)P_i(\mathrm{d}y) = \mathbb{E}(X_i^{(1)} \wedge X_i^{(2)}). \tag{5.46}$$

Let $P_3 = cP_1 + (1-c)P_2$ with $c \in [0,1]$. We have

$$\begin{aligned}
2\int_0^1\int_0^y xP_3(\mathrm{d}x)P_3(\mathrm{d}y) &= \mathbb{E}(X_3^{(1)} \wedge X_3^{(2)})\\
&= c^2\mathbb{E}(X_1^{(1)} \wedge X_1^{(2)}) + (1-c)^2\mathbb{E}(X_2^{(1)} \wedge X_2^{(2)}) + 2c(1-c)\mathbb{E}(X_1^{(\cdot)} \wedge X_2^{(\cdot)})\\
&\leqslant e^*(c^2+(1-c)^2) + 2c(1-c)\mathbb{E}(X_1^{(1)} \wedge X_2^{(1)}).
\end{aligned} \tag{5.47}$$

Hence it remains to show that $\mathbb{E}(X_1^{(1)} \wedge X_2^{(1)}) \leqslant e^*$. We have

$$\mathbb{E}(X_1^{(\cdot)} \wedge X_2^{(\cdot)}) = \int_0^1 \mathrm{d}x\,\mathbb{P}(X_1 \geqslant x)\,\mathbb{P}(X_2 \geqslant x) \tag{5.48}$$

$$\leqslant \left(\int_0^1 \mathrm{d}x\,\mathbb{P}(X_1 \geqslant x)^2 \int_0^1 \mathrm{d}x\,\mathbb{P}(X_2 \geqslant x)^2\right)^{1/2} \leqslant e^*, \tag{5.49}$$

where in the second step we apply the Cauchy-Schwarz inequality, and in the final step use (5.45). □

Appendix A. Appendix: rate function for the driving process

To establish an LDP for $\{(\tilde{h}^{\widehat{G}_n(t)})_{t\in[0,T]}\}_{n\in\mathbb{N}}$, we first need to establish an LDP for the driving process, and thus verify that Assumption 3.2 holds. Note that Lemma A.1 gives an LDP for $\{(f_n^*(t))_{t\in[0,T]}\}_{n\geqslant 0}$, where

$$f_n^*(t) = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i(t)}. \tag{A.1}$$

In preparation, for $A \subseteq \mathbb{R}_+$ (such that $A=[a,b]$ implies $A+c=[a+c,b+c]$) we let

$$D_{\mathbf{1}}\mu_t(A) = \lim_{h\to 0}\frac{\mu_{t+h}(A+h)-\mu_t(A)}{h}. \tag{A.2}$$

Lemma A.1. *The sequence of processes $(f_n^*)_{n\in\mathbb{N}}$ satisfies the LDP on $D([0,T],\mathcal{M}([0,1]))$ with rate n and with rate function*

$$\begin{aligned}
K(\mu) = &\int_0^T \mathrm{d}t\int_0^\infty[\gamma\mu_i(\mathrm{d}x) - D_{\mathbf{1}}\mu_t(\mathrm{d}x)] + \int_0^T \mathrm{d}t\int_0^\infty D_{\mathbf{1}}\mu_t(\mathrm{d}x)\log\left(\frac{D_{\mathbf{1}}\mu_t(\mathrm{d}x)}{\mu_i(\mathrm{d}x)f_t(0)}\right)\\
&+\int_0^T \mathrm{d}t\, f_t(0)\log(f_t(0)/\gamma),
\end{aligned} \tag{A.3}$$ {eq:rateex1}

where $f_t(0) := \lim_{h\to 0}\mu_t([0,h])/h$.

Proof. **[Some arguments are in the notes document.]** □

We apply the standard method of proving sample-path LDPs: establish a finite-dimensional LDP and prove tightness. **[Write out the tightness arguments.]**

For $t \in [0,T]$, let

$$L_n(t) = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i(t)), \tag{A.4}$$

where $L_n(t)$ is a random variable on $\mathcal{M}(\mathbb{R}_+)$. We assume that

{DGa} $$L_n(0) = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i(0)) \to v \tag{A.5}$$

in $\mathcal{M}(\mathbb{R}_+)$.

<u>Guessing the rate function:</u> Before deriving the finite-dimensional LDP, let us first guess the one-step rate function (the finite-dimensional LDP is an easy consequence). First note that if there exists $x > 0$ such that $\mu(x+t) > v(x)$, then $I_v^{(t)}(\mu) = \infty$. This is because the only vertices that can be of age $x+t$ at time t are those whose age at time 0 was x. When $I_v^{(t)}(\mu) < \infty$, we expect that

{Gu1} $$I_v^{(t)}(\mu) = \int_0^\infty v(\mathrm{d}x)\Bigg[\frac{\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)}\log\frac{\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)\mathrm{e}^{-\gamma t}} \tag{A.6}$$

{Gu2} $$+\left(\frac{v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)}\right)\log\left(\frac{v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)(1-\mathrm{e}^{-\gamma t})}\right)\Bigg] \tag{A.7}$$

{Gu3} $$+\int_0^t \mu(\mathrm{d}y)\log\left(\frac{\mu(\mathrm{d}y)}{\mathrm{d}y\gamma\mathrm{e}^{-\gamma y}}\right). \tag{A.8}$$

This is because we can effectively think of simulating the age distribution at time t in two steps. First, for each initial age window $\mathrm{d}x$ we determine what proportion of the vertices have not turned off. Second, conditional on a vertex turning off in $[0,t]$, its age is distributed according to a truncated exponential with rate γ (see explanation below). Thus, we obtain (A.6) and (A.7) from the LDP for sums of Bernoulli random variables, and (A.8) from Sanov's theorem.

Below we will find that this expression is not correct.

<u>Derivation of the rate function:</u> To establish a finite-dimensional LDP we apply [16, Theorem 3.5], which is stated below. In preparation, let

$$P_x^{(\boldsymbol{t})}(\mathrm{d}y^{(1)},\dots,\mathrm{d}y^{(r)}) = \mathbb{P}(X_i(t_1)\in \mathrm{d}y^{(1)},\dots,X_i(t_r)\in\mathrm{d}y^{(r)}), \tag{A.9}$$

where $\boldsymbol{t} = (t_1,\dots,t_r)$ with $0 \leqslant t_1 < t_2 < \cdots < t_r \leqslant T$.

Theorem A.2. *If* (A.5) *holds, then the measures* $\mathbb{P}((L_n(t_1),\dots,L_n(t_r))\in\cdot)$ *satisfy the LDP with speed* n *and rate function*

$$I_v^{(\boldsymbol{t})}(\mu_1,\dots,\mu_r)$$

$$= \sup_{f_1,\dots,f_r\in C_b(\mathbb{R}_+)^r}\Bigg[\sum_{i=1}^{r}\int_{\mathbb{R}_+}\mu_i(\mathrm{d}z)f_i(z)$$

$$-\int_{\mathbb{R}_+} v(dx)\log\int_{\mathbb{R}_+^r} P_x^{(\boldsymbol{t})}(\mathrm{d}y^{(1)},\ldots,\mathrm{d}y^{(r)})\exp\left(\sum_{i=1}^r f_i(y^{(i)})\right)\Bigg],$$

where $(\mu_1,\ldots,\mu_r)\in\mathcal{M}(\mathbb{R}_+)^r$.

To apply Theorem A.2, we need to be able to write down a formula for $P_x^{(\boldsymbol{t})}(\mathrm{d}y^{(1)},\ldots,\mathrm{d}y^{(r)})$. By the Markov property, this is essentially equivalent to writing down an expression for $P_x^{(t)}$ (i.e., for a single time step). If $X_i(0)=x$, then the probability that $X_i(t)=x+t$ is $\mathrm{e}^{-\gamma t}$ (i.e., the probability that the Poisson clock associated with vertex i does not ring in the time interval $[0,t]$). On the other hand, if $y\leqslant t$, then the probability that $X_i(t)\in\mathrm{d}y$ is the probability that the Poisson clock associated with vertex i rings in the time interval $t-\mathrm{d}y$ (which occurs with probability $\gamma\mathrm{d}y$) and afterwards does not ring again (which occurs with probability $\mathrm{e}^{-\gamma y}$). We thus have

$$P_x^{(t)}(\mathrm{d}y)=\begin{cases}\mathrm{e}^{-\gamma t} & \text{if } y=x+t,\\ \gamma\mathrm{d}y\mathrm{e}^{-\gamma y} & \text{if } y\leqslant t,\\ 0 & \text{otherwise.}\end{cases}\tag{A.10}$$

If we apply Theorem A.2 for a single time step, then we obtain

$$\begin{aligned}I_v^{(t)}(\mu)&=\sup_{f\in C_b([0,\infty))}\left[\int_0^\infty\mu(\mathrm{d}z)f(z)-\int_0^\infty v(\mathrm{d}x)\log\left(\int_0^\infty P_x^{(t)}(\mathrm{d}y)\mathrm{e}^{f(y)}\right)\right]\\&=\sup_{f\in C_b[0,\infty)}\left[\int_0^\infty\mu(\mathrm{d}z)f(z)-\int_0^\infty v(\mathrm{d}x)\log\left(\mathrm{e}^{-\gamma t+f(x+t)}+\int_0^t\mathrm{d}y\gamma\mathrm{e}^{-\gamma y+f(y)}\right)\right].\end{aligned}\tag{A.11}$$ {DZc}

We would like to derive a closed form expression for $I^{(t)}(\mu)$. To do this, we first take the derivative of

$$\int_0^\infty\mu(\mathrm{d}z)f(z)-\int_0^\infty v(\mathrm{d}x)\log\left(\mathrm{e}^{-\gamma t+f(x+t)}+\int_0^t\mathrm{d}y\,\gamma\mathrm{e}^{-\gamma y+f(y)}\right)\tag{A.12}$$ {JEq}

with respect to $f(x+t)$ when $x\geqslant 0$, and set this to 0. This gives

$$\mu(\mathrm{d}(x+t))-v(\mathrm{d}x)\,\frac{\mathrm{e}^{-\gamma t+f(x+t)}}{\mathrm{e}^{-\gamma t+f(x+t)}+\int_0^t\mathrm{d}y\gamma\mathrm{e}^{-\gamma y+f(y)}}=0,\tag{A.13}$$

which implies

$$\mu(\mathrm{d}(x+t))\int_0^t\mathrm{d}y\,\gamma\mathrm{e}^{-\gamma y+f(y)}=[v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))]\,\mathrm{e}^{-\gamma t+f(x+t)}.\tag{A.14}$$

Thus,

$$\mathrm{e}^{f(x+t)}=\frac{\mu(\mathrm{d}(x+t))\int_0^t\mathrm{d}y\,\gamma\mathrm{e}^{-\gamma y+f(y)}}{[v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))]\,\mathrm{e}^{-\gamma t}}\tag{A.15}$$

and

$$f(x+t)=\log\left(\frac{\mu(\mathrm{d}(x+t))\int_0^t\mathrm{d}y\,\gamma\mathrm{e}^{-\gamma y+f(y)}}{[v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))]\,\mathrm{e}^{-\gamma t}}\right).\tag{A.16}$$ {fxtE}

Substituting (A.16) into (A.12), we obtain

$$
\begin{aligned}
\text{(A.12)} &= \int_0^{t-} \mu(\mathrm{d}z)\, f(z) + \int_0^{\infty} \mu(\mathrm{d}(x+t)) \log\left(\frac{\mu(\mathrm{d}(x+t)) \int_0^t \mathrm{d}y\, \gamma \mathrm{e}^{-\gamma y + f(y)}}{[v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))]\, \mathrm{e}^{-\gamma t}} \right) \\
&\quad - \int_0^{\infty} v(\mathrm{d}x)\, \log\left(\mathrm{e}^{-\gamma t} \frac{\mu(\mathrm{d}(x+t)) \int_0^t \mathrm{d}y\, \gamma \mathrm{e}^{-\gamma y + f(y)}}{[v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))]\, \mathrm{e}^{-\gamma t}} + \int_0^t \mathrm{d}y\, \gamma \mathrm{e}^{-\gamma y + f(y)} \right) \\
&= \int_0^{t-} \mu(\mathrm{d}z) f(z) + \int_0^{\infty} \mu(\mathrm{d}(x+t)) \log\left(\frac{\mu(\mathrm{d}(x+t)) \int_0^t \mathrm{d}y\, \gamma \mathrm{e}^{-\gamma y + f(y)}}{[v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))]\, \mathrm{e}^{-\gamma t}} \right) \\
&\quad - \int_0^{\infty} v(\mathrm{d}x) \log\left(\frac{v(\mathrm{d}x)}{v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))} \int_0^t \mathrm{d}y\, \gamma \mathrm{e}^{-\gamma y + f(y)} \right) \\
&= \int_0^{\infty} \mu(\mathrm{d}(x+t)) \log\left(\frac{\mu(\mathrm{d}(x+t))}{[v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))]\, \mathrm{e}^{-\gamma t}} \right) \\
&\quad - \int_0^{\infty} v(\mathrm{d}x)\, \log\left(\frac{v(\mathrm{d}x)}{v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))} \right) \\
&\quad + \int_0^{t-} \mu(\mathrm{d}z)\, f(z) - \int_0^{\infty} [v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))] \log\left(\int_0^t \mathrm{d}y\, \gamma \mathrm{e}^{-\gamma y + f(y)} \right).
\end{aligned}
\tag{A.17}
$$

We now optimise over $f(z)$, $0 \leqslant z \leqslant t$. To do this, we take the derivative of

{fuwrE}
$$
\int_0^{t-} \mu(\mathrm{d}z)\, f(z) - \int_0^{\infty} [v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))] \log\left(\int_0^t \mathrm{d}z\, \gamma \mathrm{e}^{-\gamma z + f(z)} \right) \tag{A.18}
$$

{flwrE}
$$
= \int_0^{t-} \mu(\mathrm{d}z)\, f(z) - \left(1 - \int_{t+}^{\infty} \mu(\mathrm{d}x)\right) \log\left(\int_0^t \mathrm{d}z\, \gamma \mathrm{e}^{-\gamma z + f(z)} \right) \tag{A.19}
$$

with respect to $f(z)$ for a fixed $z \in [0,t]$, and set this to 0. We obtain

$$
\mu(\mathrm{d}z) - \left(1 - \int_{t+}^{\infty} \mu(\mathrm{d}x)\right) \frac{\mathrm{d}z\, \gamma \mathrm{e}^{-\gamma z + f(z)}}{\int_0^t \mathrm{d}z\, \gamma \mathrm{e}^{\gamma z + f(z)}} = 0, \tag{A.20}
$$

which implies that

$$
\mu(\mathrm{d}z) \int_0^t \mathrm{d}y\, \gamma \mathrm{e}^{-\gamma y + f(y)} = \left(1 - \int_{t+}^{\infty} \mu(\mathrm{d}x)\right) \mathrm{d}z\, \gamma \mathrm{e}^{-\gamma z + f(z)}. \tag{A.21}
$$

Thus,

$$
\mathrm{e}^{f(z)} = \frac{\mu(\mathrm{d}z) \int_0^t \mathrm{d}y\, \gamma \mathrm{e}^{-\gamma y + f(y)}}{\left(1 - \int_{t+}^{\infty} \mu(\mathrm{d}x)\right) \mathrm{d}z\, \gamma \mathrm{e}^{-\gamma z}} \tag{A.22}
$$

and

{fzLwr}
$$
f(z) = \log\left(\frac{\mu(\mathrm{d}z) \int_0^t \mathrm{d}y\, \gamma \mathrm{e}^{-\gamma y + f(y)}}{\left(1 - \int_{t+}^{\infty} \mu(\mathrm{d}x)\right) \mathrm{d}z\, \gamma \mathrm{e}^{-\gamma z}} \right). \tag{A.23}
$$

Substituting (A.23) into (A.19), we obtain

$$\int_0^{t-} \mu(\mathrm{d}z)\,\log\left(\frac{\mu(\mathrm{d}z)\int_0^t \mathrm{d}y\,\gamma\mathrm{e}^{-\gamma y+f(y)}}{\left(1-\int_{t+}^\infty \mu(\mathrm{d}x)\right)\mathrm{d}z\,\gamma\mathrm{e}^{-\gamma z}}\right)-\left(1-\int_{t+}^\infty \mu(\mathrm{d}x)\right)\log\left(\int_0^t \mathrm{d}z\,\gamma\mathrm{e}^{-\gamma z+f(z)}\right)$$
$$=\int_0^{t-} \mu(\mathrm{d}z)\,\log\left(\frac{\mu(\mathrm{d}z)}{dz\gamma\mathrm{e}^{-\gamma z}}\right)-\int_0^{t-} \mu(\mathrm{d}z)\,\log\left(\int_0^{t-} \mu(\mathrm{d}z)\right). \tag{A.24}$$

Combining this with (A.18), we obtain

$$\begin{aligned}
&(\mathrm{A}.12)\\
&=\int_0^\infty \mu(\mathrm{d}(x+t))\log\left(\frac{\mu(\mathrm{d}(x+t))}{[v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))]\mathrm{e}^{-\gamma t}}\right)\\
&\quad-\int_0^\infty v(\mathrm{d}x)\log\left(\frac{v(\mathrm{d}x)}{v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))}\right)\\
&\quad+\int_0^{t-} \mu(\mathrm{d}z)\log\left(\frac{\mu(\mathrm{d}z)}{\mathrm{d}z\,\gamma\mathrm{e}^{-\gamma z}}\right)-\int_0^{t-} \mu(\mathrm{d}z)\,\log\left(\int_0^{t-} \mu(\mathrm{d}z)\right)\\
&=\int_0^\infty [v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))]\log\left(v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))\right)\\
&\quad+\int_0^\infty \mu(\mathrm{d}(x+t))\log\left(\frac{\mu(\mathrm{d}(x+t))}{\mathrm{e}^{-\gamma t}}\right)\\
&\quad-\int_0^\infty v(\mathrm{d}x)\log(v(\mathrm{d}x))-\int_0^{t-} \mu(\mathrm{d}z)\,\log\left(\int_0^{t-} \mu(\mathrm{d}z)\right)+\int_0^{t-} \mu(\mathrm{d}z)\,\log\left(\frac{\mu(\mathrm{d}z)}{\mathrm{d}x\,\gamma\mathrm{e}^{-\gamma z}}\right)\\
&=\int_0^\infty v(\mathrm{d}x)\left[\frac{v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)}\log\left(\frac{v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)}\right)\right.\\
&\qquad\left.+\frac{\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)}\log\left(\frac{\mu(\mathrm{d}(x+t))}{\mathrm{e}^{-\gamma t}v(\mathrm{d}x)}\right)\right]\\
&\quad-\left(\int_0^{t-} \mu(\mathrm{d}z)\right)\log\left(\int_0^{t-} \mu(\mathrm{d}z)\right)+\int_0^{t-} \mu(\mathrm{d}z)\log\left(\frac{\mu(\mathrm{d}z)}{\mathrm{d}x\,\gamma\mathrm{e}^{-\gamma z}}\right).
\end{aligned}$$

Rearranging further, we obtain

$$\begin{aligned}
I_v^{(t)}(\mu)=&\int_0^\infty \mu(\mathrm{d}(x+t))\log\left(\frac{\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)\,\mathrm{e}^{-\gamma t}}\right)\\
&+\int_0^\infty [v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))]\log\left(\frac{v(\mathrm{d}x)-\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)\int_0^{t-}\mu(\mathrm{d}z)}\right)\\
&+\int_0^{t-} \mu(\mathrm{d}z)\log\left(\frac{\mu(\mathrm{d}z)}{\mathrm{d}x\,\gamma\mathrm{e}^{-\gamma z}}\right).
\end{aligned} \tag{A.25}$$

{FDRF}

Is this expression correct?: If we rewrite the expression that we guessed for the rate function in (A.6)–(A.8), then we have

$$I_v^{(t)*}(\mu) = \int_0^\infty \mu(\mathrm{d}(x+t)) \log \frac{\mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)\,\mathrm{e}^{-\gamma t}} + \int_0^\infty [v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))] \log \left(\frac{v(\mathrm{d}x) - \mu(\mathrm{d}(x+t))}{v(\mathrm{d}x)\,(1-\mathrm{e}^{-\gamma t})} \right) + \int_0^t \mu(\mathrm{d}y) \log \left(\frac{\mu(\mathrm{d}y)}{\mathrm{d}y\,\gamma \mathrm{e}^{-\gamma y}} \right). \tag{A.26}$$ {Gu4}

We then need to determine which of (A.25) or (A.26) gives the correct expression for the rate function.vWe know that if all individuals initially have the same age, then $I_v^{(t)}(\mu)$ is the relative entropy from $P_v^{(t)}$ to μ. Suppose that

$$v(\mathrm{d}x) = \begin{cases} 1, & \text{if } x = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{A.27}$$

From (A.25) we have

$$\begin{aligned} I_v^{(t)}(\mu) &= \mu(t) \log \left(\frac{\mu(t)}{\mathrm{e}^{-\gamma t}} \right) + (1-\mu(t)) \log \left(\frac{1-\mu(t)}{1-\mu(t)} \right) + \int_0^{t-} \mu(\mathrm{d}y) \log \left(\frac{\mu(\mathrm{d}y)}{\mathrm{d}y\,\gamma \mathrm{e}^{-\gamma y}} \right) \\ &= \mu(t) \log \left(\frac{\mu(t)}{\mathrm{e}^{-\gamma t}} \right) + \int_0^{t-} \mu(\mathrm{d}y) \log \left(\frac{\mu(\mathrm{d}y)}{\mathrm{d}y\,\gamma \mathrm{e}^{-\gamma y}} \right), \end{aligned} \tag{A.28}$$

which is the relative entropy from $P_v^{(t)}$ to μ. Thus, we expect (A.25) to be the correct expression for the rate function.

[References need to be made complete and need to be checked.]

References

[1] Athreya, S., Hollander, F. D., and Röllin, A. (2019). Graphon-valued stochastic processes from population genetics. arXiv:1908.06241.

[2] Athreya, S., and Rollin, A. (2016). Dense graph limits under respondent-driven sampling. The Annals of Applied Probability 26(4), 2193–2210.

[3] Billingsley, P. (1979). *Probability and Measure*. Wiley, New York.

[4] Budhiraja, A., and Dupuis, P. (2019). *Analysis and Approximation of Rare Events: Representations and Weak Convergence Methods*. Springer, New York.

[5] Borgs, C., Chayes, J., Gaudio, J., Petti, S., and Sen, S. (2020). A large deviation principle for block models. arXiv preprint arXiv:2007.14508.

[6] Bovier, A., and Gayrard, V. (1999). Sample path large deviations for a class of Markov chains related to disordered mean field models. arXiv math/9905022.

[7] Bovier, A., and den Hollander, F. (2016). *Metastability: A Potential-Theoretic Approach*. Grundlehren der mathematischen Wissenschaften 351. Springer.

[8] Braunsteins, P., den Hollander, F., and Mandjes, M. (2020). A sample-path large deviation principle for dynamic Erdős–Rényi random graphs. arXiv:2009.12848.

[9] Černý, J., and Klimovsky, A. (2020). Markovian dynamics of exchangeable arrays. In: *Genealogies of Interacting Particle Systems*. Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore, pp. 209–228.

[10] Chatterjee, S. (2017). *Large Deviations for Random Graphs* Lecture Notes in Mathematics 2197. Springer.

[11] Chatterjee, S., and Dembo, A. (2016). Nonlinear large deviations. Advances in Mathematics 299, 396–450.

[12] Chatterjee, S., and Varadhan, S.R.S. (2011). The large deviation principle for the Erdős-Rényi random graph. European Journal of Combinatorics 32 (7), 1000–1017.

[13] Crane, H. (2015). Time-varying network models. Bernoulli 21(3), 1670–1696.

[14] Crane, H. (2016). Dynamic random networks and their graph limits. The Annals of Applied Probability 26(2), 691–721.

[15] Crane, H. (2017). Exchangeable graph-valued Feller processes. Probability Theory and Related Fields 168(3), 849–899.

[16] Dawson, D A., and Gärtner, J. (1987). Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. Stochastics: An International Journal of Probability and Stochastic Processes 20(4), 247–308.

[17] Dembo, A., and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability 38. Springer, New York.

[18] Dhara, S., and Sen, S. (2019). Large deviation for uniform graphs with given degrees. arXiv:1904.07666.

[19] Dupuis, P., and Medvedev, G. (2020). The large deviation principle for interacting dynamical systems on random graphs. arXiv:2007.13899.

[20] Ethier, S. N., and Kurtz, T.G. (2009). *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics 282. John Wiley & Sons.

[21] Feng, J., and Kurtz, T.G. (2006). *Large Deviations for Stochastic Processes*. Mathematical Surveys and Monographs 131. American Mathematical Society.

[22] Grebík, J., and Pikhurko, O. (2021). Large deviation principles for block and step graphon random graph models. arXiv:2101.07025.

[23] Harris, T.E. (1963). *The Theory of Branching Processes*. Grundlehren der mathematischen Wissenschaften 196. Springer, Berlin.

[24] den Hollander, F. (2000). *Large Deviations*. Fields Institute Monograph 14. American Mathematical Society, Providence RI, USA.

[25] Lovász, L. (2012). *Large Networks and Graph Limits*. American Mathematical Society, Providence RI, USA.

[26] Lovász, L., and Szegedy, B. (2006). Limits of dense graph sequences. Journal of Combinatorial Theory, Series B 96, 933–957.

[27] Lubetzky, E., and Zhao, Y. (2015). On replica symmetry of large deviations in random graphs. Random Structures & Algorithms 47(1), 109–146.

[28] Méliot, P.-L. (2021) A central limit theorem for singular graphons. arXiv:2103.15741.

[29] Markering, M. (2020). The large deviation principle for inhomogeneous Erdő sRényi random graphs. arXiv:2010.03504.

[30] Röllin, A., and Zhang, Z.S. (2021). Dense multigraphon-valued stochastic processes and edge-changing dynamics in the configuration model. arXiv:2104.13024.

Korteweg-de-Vries Instituut, Universiteit van Amsterdam, PO Box 94248, 1090 GE Amsterdam, The Netherlands
Email address: pbraunsteins@gmail.com

Mathematisch Instituut, Universiteit Leiden, PO Box 9512, 2300 RA Leiden, The Netherlands
Email address: denholla@math.leidenuniv.nl

Korteweg-de Vries Instituut, Universiteit van Amsterdam, PO Box 94248, 1090 GE Amsterdam, The Netherlands
Email address: M.H.R.Mandjes@uva.nl