

EURANDOM PREPRINT SERIES

2019-010

August 15, 2019

Graphon-valued stochastic processes from population genetics

S. Athreya, F. den Hollander, Adrian Röllin
ISSN 1389-2355

Graphon-valued stochastic processes from population genetics

Siva Athreya, Frank den Hollander and Adrian Röllin

*Indian Statistical Institute Bangalore, Leiden University
and National University of Singapore*

Abstract

The goal of this paper is to develop a theory of graphon-valued stochastic processes, and to construct and analyse a natural class of such processes arising from population genetics. We consider finite populations where individuals change type according to Wright-Fisher 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 empirical type distribution of the entire population via a fitness function. We show 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 Wright-Fisher 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.

MSC 2010: 05C80, 60J68, 60K35.

Keywords: Graphons, graphon dynamics, Moran model, Wright-Fisher diffusion, Fleming-Viot diffusion, Skorohod topology.

1 Introduction

In this paper we construct a class of graphon-valued Markov processes from a sequence of dynamically evolving dense graphs. First, we characterise weak convergence for processes that take values in the metric space of graphons (see Theorem 3.1). Afterwards, using this characterisation, we find interesting scaling limits with the help of models from population genetics and use these to construct *graphon-valued diffusions* (see Theorem 3.3 and Theorem 3.4).

1.1 Background

Various fields of research – including physics, computer science, sociology and epidemiology – have produced a considerable literature on *dynamics of real-world networks*. The first example of a time-changing network was proposed by Holland and Leinhardt

(1977), for the evolution of social networks. Accounts of subsequent developments are given in Snijders (2001) and Snijders et. al (2010), including a description of statistical procedures to monitor the effects of the dynamics. In epidemiology, there is a long tradition of modelling the spread of pathogens on social contact networks. In particular, the study of sexually transmitted infections needs to take into account that partnership networks change at about the same speed as the infections spread. Most of the mathematical analysis in this area is done by means of compartmental models, which goes back to Kermack and McKendrick (1927). These models lead to systems of ordinary differential equations, and consequently stochastic effects are lost. Agent-based models are also frequently used, but these are typically intractable mathematically (see, for example, Morris and Kretzschmar (1997)). In physics and computer science, much of the efforts are driven by simulations or mathematical non-rigorous techniques. See Holme and Saramäki (2012) for a survey with many real-world applications, and an attempt to unify various sub-disciplines that have emerged.

In contrast to these efforts, the mathematical treatment of the topic is still in its early stages. Some interacting particle systems, such as oriented percolation, can be interpreted as the spread of an infection on time-varying networks, but these networks are intrinsically highly structured (typically taking the form of a lattice), and therefore are very different from the networks we have in mind in the present paper. Also, most of the work deals with *sparse* networks, where the degrees typically remain bounded. Quantities of interest are the mixing times and the cover times of random walks on these networks, under different types of dynamics. These enable the description of propagation of information through the network (see, for example, Leskovec (2008), Levin and Peres (2017), and references therein).

Crane (2016) probably contains the first mathematically rigorous attempt to capture the limiting dynamics of time-varying networks as the number of edges per vertex grows to infinity. In the context of dense graph limit theory, initiated by Lovász and Szegedy (2006), these limits can be understood by means of *graphons*, which can be turned into a compact metric space. While it is intuitively (and mathematically) easy to construct random dynamics of graphs on n vertices for each fixed n , it is non-trivial to realise them in such a way that the dynamics remain visible in the limit as $n \rightarrow \infty$. This is because, as the time-evolving dense graph sequence approaches the appropriate evolving graphon, a lot of averaging takes place that typically results in a *deterministic* flow.

Crane's starting point is the Aldous-Hoover theory for infinitely exchangeable arrays. Let \mathcal{G}_∞ be the space of infinite arrays equipped with the product topology. Define a modulus map $|\cdot|$ that takes an array $\Gamma \in \mathcal{G}_\infty$ to a graphon $|\Gamma|$ (which is well-defined with probability 1 when the array is exchangeable). One of the main results of Crane (2016) is that a \mathcal{G}_∞ -valued exchangeable Markov process $(\Gamma(s))_{s \geq 0}$ induces a graphon-valued Markov process $(|\Gamma(s)|)_{s \geq 0}$, and that the latter has locally bounded variation. Consequently, this route only leads to jump processes and deterministic flows on the space of graphons, *not* to diffusion-like processes (see Crane (2016), Černý and Klimovsky (2018)).

However, this does *not* imply that there *are* no diffusion-like processes on graphons.

In fact, as we show in what follows, the limitation is imposed by the theory of infinitely exchangeable arrays, not by the theory of graphons. More precisely, we start with the Aldous-Hoover theory and represent an $[0, 1]$ -valued infinitely exchangeable array $(X_{ij})_{i,j \in \mathbb{N}}$ as

$$X_{ij} = f(U, U_i, U_j, U_{ij}), \quad i, j \in \mathbb{N} := \{1, 2, \dots\}, \quad (1.1)$$

for some function $f: [0, 1]^4 \rightarrow [0, 1]$, where U , $(U_i)_{i \in \mathbb{N}}$ and $(U_{ij})_{i,j \in \mathbb{N}}$ are independent and identically distributed uniform random variables. If we assume that $(U(s))_{s \geq 0}$, $(U_i(s))_{s \geq 0}$ and $(U_{ij}(s))_{s \geq 0}$ are Markov processes having the uniform distribution as their equilibrium distribution, then we can construct from (1.1) a \mathcal{G}_∞ -valued process

$$X_{ij}(s) = f(U(s), U_i(s), U_j(s), U_{ij}(s)), \quad i, j \in \mathbb{N}.$$

It is clear that, in general, $(X_{ij}(s))_{s \geq 0}$ is not Markov, since functions of Markov processes need not be Markov under the filtration of the mapped process. Moreover, even if it is Markov, then once the process is projected to the graphon space many of its properties are lost. Below we will illustrate that interesting Markov processes on the graphon space need not be Markov on \mathcal{G}_∞ , and thus are *not captured through the lens of the Aldous-Hoover theory*.

1.2 A short primer on graphon-valued stochastic processes

The approach we take in this paper is to work directly with networks and their graphon limits, and we illustrate this by means of a classical model from population genetics. Concretely, we assign types to each individual and define connection probabilities based on the types. Then we impose the dynamics on the types from population genetic models to enable a scaling limit that results in diffusive dynamics on the space of graphons. The type space is allowed to be continuous or discrete. This allows us to also observe the dynamics as a rescaled limit of a time-evolving finite dense graph sequence, thus providing a natural justification for the dynamics. We present this approach first in the case where each individual is one of two types and all individuals within each type are connected with each other via the below example.

The following example shows that graphon-valued diffusions can arise from discrete models.

Example 1.1. Consider n individuals, each carrying Type 0 or Type 1. Suppose that each individual, independently and at rate 1, randomly draws an individual from the population (possibly itself) and adopts its type. Let $X^n(s)$ be the number of individuals of type 0 at time s . For each $s \geq 0$, think of the n individuals as the vertices of a random graph $G^n(s)$ in which individuals i and j are connected by an edge with probability 1 if they are of the same type and remain disconnected if their types are different. Using (2.1) we see that, for any connected graph F on k vertices, the subgraph density of F in $G^n(s)$

0	1
1	0

$Y(s)$

Figure 1: Graphical representation of the limiting graphon-valued stochastic process arising from a simple dynamical graph model, with $(Y(s))_{s \geq 0}$ the Wright-Fisher diffusion.

is

$$\begin{aligned}
 t_F(G^n(s)) &:= \frac{\# \text{ of copies of } F \text{ in } G^n(s)}{\# \text{ of copies of } F \text{ in the complete graph}} \\
 &= \frac{X^n(s)^k + (n - X^n(s))^k}{n^k}
 \end{aligned} \tag{1.2}$$

(we will give a rigorous definition of t_F later). If F consists of multiple components, then t_F is just the product of t_{F_i} with F_i the individual connected components of F . Let $Y^n(s) = \frac{1}{n} X^n(ns)$ represent the fraction of individuals of Type 1 in the population at time s on time scale n . It is well known that if $Y^n(0)$ converges weakly to $Y(0)$, then $Y^n = (Y^n(s))_{s \geq 0}$ converges weakly to $Y = (Y(s))_{s \geq 0}$ in path space with respect to the Skorohod topology as $n \rightarrow \infty$, where the limiting process is the Wright-Fisher diffusion on $[0, 1]$, given by the SDE

$$Y(s) = Y(0) + \int_0^s \sqrt{Y(u)(1 - Y(u))} dW(u)$$

with initial condition $Y(0)$ and with $W = (W(s))_{s \geq 0}$ being standard Brownian motion. Clearly, (1.2) implies that

$$\lim_{n \rightarrow \infty} t_F(G^n(ns)) = Y(s)^k + (1 - Y(s))^k. \tag{1.3}$$

We observe that the right-hand side of (1.3) equals $t_F(\tilde{h}(s))$, the subgraph density corresponding to the graphon $\tilde{h}(s)$ drawn in Figure 1. It is also easily seen that $t_F(\tilde{h}_s)$ is adapted to the filtration generated by Y_s , and is a Markov process. Furthermore, we can calculate the modulus of continuity of $t_F(\tilde{h})$ and conclude in the subgraph distance \tilde{h} is diffusive. Consequently, we have constructed a graphon-valued diffusion and a sequence of finite graph-valued processes that converges to it (see Section 2.1 for precise definitions). \square

The next example shows that interesting and natural processes can be constructed in such a way that they are not Markov when seen through the Aldous-Hoover theory of infinitely exchangeable arrays, but are Markov when projected to graphons.

Example 1.2. Recall the notation $\{X_{ij} : i, j \in \mathbb{N}\}$ used for exchangeable arrays earlier. Consider the Wright-Fisher diffusion $(Y(s))_{s \geq 0}$ on $[0, 1]$, and put

$$X_{ij}(s) = \mathbb{I}[U_{ij} \leq Y(s)], \quad i, j \in \mathbb{N}, s \geq 0,$$

where the U_{ij} are independent and identically distributed uniform random variables. This process is non-Markov in \mathcal{G}_∞ with respect to its filtration

$$\mathcal{F}(s) = \sigma(X_{ij}(u) : 0 \leq u \leq s, i, j \in \mathbb{N}).$$

Although $Y(s)$ is measurable with respect to $\mathcal{F}(s)$, because

$$\lim_{n \rightarrow \infty} \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} X_{ij}(s) = Y(s) \quad \text{almost surely,}$$

the individual U_{ij} are *not*. Still, the projected graphon-valued process is $(Y(s))_{s \geq 0}$, which *is* Markov with respect to its filtration $\mathcal{F}(s) = \sigma(Y_u : 0 \leq u \leq s)$. \square

In view of the above, it seems that the Aldous-Hoover theory, while being a natural starting point, is ultimately not the right way to develop a theory of graphon processes. Being Markov on the space of infinitely exchangeable arrays is too strong a condition, since it restricts the possible dynamics in the graphon space to those with locally bounded variation.

1.3 Outline

The remainder of the paper is organised as follows. In Section 2, we present the preliminaries required. We provide a brief introduction into the theory of graphons in Section 2.1 and a quick review of population models in Section 2.2. In Section 3, we present our main results. We begin in Section 3.1 by discussing the Skorohod topology on the graphon space and provide a framework for understanding weak convergence in graphon space by means of sub-graph densities (Theorem 3.1 and Corollary 3.2). In Section 3.2, we show that the models discussed in Section 2.2 provide a natural class of graphon dynamics: the graphon is obtained by connecting pairs of individuals with a probability that is given by a type-connection matrix, whose entries depend on the empirical type distribution of the entire population via a fitness function (Theorem 3.3). After that we let the number of types tend to infinity and arrive at a graphon dynamics governed by the Fleming-Viot diffusion (Theorem 3.4). In Section 4, we give concrete examples to illustrate the abstract results in Section 3 (Examples 4.1–4.3) and offer some remarks on possible generalisations. In Section 5, we give the proofs of the theorems.

2 Preliminaries

We begin this section by stating the minimally required preliminaries on dense graphs, graphons, their equivalence classes, and the metric space they belong to.

2.1 Graphons

We provide a brief introduction of the preliminaries of graphons and construct an example of a simple diffusion on the space of graphons. Let \mathcal{G}_n be the set of all graphs on n vertices. Graphs on n vertices with order n^2 edges are referred to as *dense graphs*. For any two graphs there is a natural definition of distance between them, given by the subgraph distance. More precisely, if F is a simple graph on k vertices and G is a graph on n vertices, then the *subgraph density* is defined as

$$t_F(G) := \frac{|\text{hom}(F, G)|}{n^k} \in [0, 1], \quad (2.1)$$

where $\text{hom}(F, G)$ denotes the set of homomorphisms from F to G .¹

Let \mathcal{F} denote the set of isomorphism classes of finite graphs given by $\mathcal{F} = \{F_i\}_{i \in \mathbb{N}}$, with each F_i being a representative of an isomorphism class. We can then define the *subgraph distance* of two graphs G_1 and G_2 as

$$d_{\text{sub}}(G_1, G_2) := \sum_{i \in \mathbb{N}} 2^{-i} |t_{F_i}(G_1) - t_{F_i}(G_2)|.$$

This metric has some nice properties. For instance, it is known that $(\mathcal{F}, d_{\text{sub}})$ is a discrete metric space and that the completion of \mathcal{F} with respect to this metric is given by the space \mathcal{W} , which is the space of measurable functions $h: [0, 1]^2 \rightarrow [0, 1]$ satisfying $h(x, y) = h(y, x)$ for all $(x, y) \in [0, 1]^2$. The elements of \mathcal{W} are called *graphons*. The definition of d_{sub} can be extended to graphons. For $h \in \mathcal{W}$ and F a simple finite graph on k vertices, we let

$$t_F(h) := \int_{[0, 1]^k} \prod_{\{i, j\} \in E(F)} h(x_i, x_j) dx_1 \cdots dx_k. \quad (2.2)$$

One of the key results in dense graph theory the above definition is based on is the following theorem.

Theorem 2.1. *Let $(G_n)_{n \geq 1}$ be a dense-graph sequence that is Cauchy with respect to d_{sub} . Then there exists an $h \in \mathcal{W}$ such that*

$$d_{\text{sub}}(G_n, h) \rightarrow 0 \quad (n \rightarrow \infty).$$

For a proof of the above see Lovász and Szegedy (2006), which uses Szemerédi partitions and the Martingale Convergence Theorem, or Diaconis and Janson (2008), who show that it can be proved by using results from Hoover (1979) and Aldous (1981). Note that h above is in general not unique, but this will not be of importance for what follows; we refer to Borgs et al. (2008) for a discussion of this and related questions. We refer to Lovász (2012, Chapter 11) for a detailed discussion of convergence of dense graph sequences.

¹Recall that a homomorphism from a graph F to a graph G is a function that maps the vertices of F to the vertices of G in such a way that edges are mapped to edges.

A convenient way of “creating” finite (random) graphs on n vertices from a standard kernel h is the following model, which we will denote by $G(n, h)$. Firstly, let U_1, \dots, U_n be i.i.d. with uniform distribution on $[0, 1]$. Secondly, for each two vertices i and j , connect them with probability $h(U_i, U_j)$, independently of all the other edges. It is not difficult to prove that

$$d_{\text{sub}}(G(n, h), h) \rightarrow 0 \quad \text{almost surely} \quad (n \rightarrow \infty). \quad (2.3)$$

This is, in some sense, the basic law of large numbers in dense graph theory. In this paper, instead of sampling the labels i.i.d. and uniformly from $[0, 1]$, we will allow the labels to be sampled in a more general way.

We can interpret (2.2) as the normalized number of homomorphisms of F into a weighted graph with the uncountable vertex set $[0, 1]$, with edge weights given by h . Furthermore, any finite simple graph G on n vertices can be represented canonically by a graphon via

$$h^G(x, y) := \begin{cases} 1 & \text{if there is an edge between vertex } \lceil nx \rceil \text{ and vertex } \lceil ny \rceil, \\ 0 & \text{otherwise.} \end{cases} \quad (2.4)$$

We easily verify that

$$t_F(G) = t_F(h^G),$$

so that all definitions are consistent. The representation of a graph via a graphon is not unique. Indeed, there are graphs that have the same graphon representation (for example, all complete graphs have the same representation $h^G \equiv 1$). In the context of graph limit theory this is not an issue, as long as we assume that the number of vertices of the graph sequence tends to infinity. Moreover, the representation in (2.4) depends on the ordering of the vertices. Since typically we are interested only in graph properties that are independent of vertex labels (for example subgraph densities), it is natural to consider equivalence classes of graphons obtained by letting two graphons be equivalent when they are identical “up to vertex labels”. To make this rigorous, let Σ be the space of measure-preserving bijections $\sigma: [0, 1] \rightarrow [0, 1]$. Then $h_1 \equiv h_2$ if there is a $\sigma \in \Sigma$ such that

$$h_1(x, y) = h_2(\sigma x, \sigma y), \quad x, y \in [0, 1].$$

The equivalence relation yields the quotient space $\widetilde{\mathcal{W}}$. It is known that $(\widetilde{\mathcal{W}}, d_{\text{sub}})$ is a compact metric space, and therefore complete and separable (see, for example, Crane (2016, p. 695) or Bollobás and Riordan (2009, Section 2), although it is not difficult to deduce this fact from the fundamental results of Lovász (2012, Theorems 9.23 and 11.5)).

The reader is referred to Bollobás and Riordan (2009, Section 2), Borgs et al. (2008), Lovász (2012) for a structured and more detailed exposition on dense graphs and graphons. Below is a simple example where we see a diffusion arising naturally on graphon space.

2.2 Population dynamics

For obtaining dynamics on graphons we will use models from population biology. We provide a very brief and quick review of the literature. The multi-type Moran model has $m + 1$ types, which are labelled $0, \dots, m$. Consider n individuals, each carrying one of the types from $\{0, \dots, m\}$. Suppose that each individual, independently and at rate 1, randomly draws an individual from the population (possibly itself) and adopts its type. Let $X_\ell^{m,n}(s)$ denote the number of individuals of type ℓ at time s , where $0 \leq \ell \leq m - 1$, and let

$$X^{m,n}(s) = (X_0^{m,n}(s), \dots, X_{m-1}^{m,n}(s))$$

be the corresponding vector of type counts. For convenience, we also define the number of individuals of type m at time s to be $X_m^{m,n}(s) := n - \sum_{\ell=0}^{m-1} X_\ell^{m,n}(s)$. Whenever we consider the entire process, we will drop time and write $X_\ell^{m,n}$ for the individual counting processes, or $X^{m,n}$ for the multivariate counting process.

Consider the space-time rescaling

$$Y^{m,n}(s) := \frac{1}{n} X^{m,n}(ns), \quad s \geq 0,$$

which consists of m components

$$Y^{m,n}(s) = (Y_0^{m,n}(s), \dots, Y_{m-1}^{m,n}(s)), \quad (2.5)$$

representing the fractions of individuals of types $0, \dots, m - 1$ at time ns . Analogously to before, the fraction of individuals of type m at time ns is denoted by $Y_m^{m,n}(s) := 1 - \sum_{\ell=0}^{m-1} Y_\ell^{m,n}(s)$. It is known (see Dawson (1993, Section 2)) that if

$$Y^{m,n}(0) \implies Y^m(0) \quad (n \rightarrow \infty), \quad (2.6)$$

then

$$Y^{m,n} \implies Y^m \quad (n \rightarrow \infty), \quad (2.7)$$

(keep in mind that (2.7) states weak convergence at the process level). The limiting process consists of m components

$$Y^m(s) = (Y_0^m(s), \dots, Y_{m-1}^m(s)),$$

taking values in the m -dimensional simplex

$$\mathcal{S}^m = \left\{ x = (x_0, \dots, x_{m-1}) \in \mathbb{R}^m : x_\ell \geq 0 \text{ for all } 0 \leq \ell \leq m - 1, \sum_{\ell=0}^{m-1} x_\ell \leq 1 \right\},$$

and is referred to as the *Wright-Fisher diffusion*. For convenience, we let $Y_m^m(s) = 1 - \sum_{\ell=0}^{m-1} Y_\ell^m(s)$. We will also need the cumulative distribution function of the type distribution, defined as

$$F^m(s; x) := \sum_{\ell=0}^{\lfloor (m+1)x \rfloor} Y_\ell^m(s), \quad x \in [0, 1), \quad (2.8)$$

and $F^m(s; 1) = 1$. Note that $x \mapsto F^m(s; x)$ can have jumps at $x \in \{0, \frac{1}{m+1}, \dots, \frac{m}{m+1}\}$.

Recalling the definition of F^m at (2.8), we can define the corresponding empirical type distribution

$$Z^m(s) = \sum_{\ell=0}^m Y_\ell^m(s) \delta_{\ell/(m+1)}.$$

Note that

$$Z^m(s; [0, x]) = F^m(s; x), \quad x \in [0, 1].$$

It is known (see Dawson (1993, Section 2)) that if

$$Z^m(0) \Longrightarrow Z(0) \quad (m \rightarrow \infty), \tag{2.9}$$

then

$$Z^m \Longrightarrow Z \quad (m \rightarrow \infty).$$

The limiting process Z takes values in $\mathcal{P}([0, 1])$, the set of probability measures on $[0, 1]$ endowed with the topology of weak convergence, and is referred to as the *Fleming-Viot diffusion*. We will also need the process of cumulative distribution function of Z , defined by

$$F(s; x) := Z(s; [0, x]), \quad x \in [0, 1].$$

3 Main results

On the metric space $(\widetilde{\mathcal{W}}, d_{\text{sub}})$, we can define the Skorohod topology on $\widetilde{\mathcal{W}}$ -valued paths in the usual way; see for example or Billingsley (1999); Ethier and Kurtz (1986). Denote by $D = D([0, \infty), \widetilde{\mathcal{W}})$ the set of càdlàg paths in $\widetilde{\mathcal{W}}$, which can be equipped with a metric d° , turning D into a complete and separable metric space since $(\widetilde{\mathcal{W}}, d_{\text{sub}})$, being compact, is complete and separable. We use “ \longrightarrow ” to denote convergence with respect to the underlying metric space, and we use “ \Longrightarrow ” to denote weak convergence with respect to the Borel-sigma-algebra induced by that metric. Note that we will use “ \longrightarrow ” and “ \Longrightarrow ” also for convergence, respectively, weak convergence in $(\widetilde{\mathcal{W}}, d_{\text{sub}})$ itself. Let \tilde{h} be a $\widetilde{\mathcal{W}}$ -valued stochastic process. We write $\tilde{h}(s)$ to denote the value of the process a time $s \geq 0$, which is an equivalence class of graphons. If $h(s)$ is a representative graphon of the equivalence class $\tilde{h}(s)$, we write $h(s; x, y)$ to denote the value of that graphon evaluated at coordinates $(x, y) \in [0, 1]^2$. Note that, for a given $\tilde{h} \in D$ and a given simple finite graph F , we can consider the real-valued process $t_F(\tilde{h})$ as an element of $D([0, \infty), [0, 1])$. We write $t_F(\tilde{h}(s))$ to denote the value of the process a time $s \geq 0$.

3.1 Weak convergence of graphons

Let $(\tilde{h}^n)_{n \in \mathbb{N}}$ be a sequence of $\widetilde{\mathcal{W}}$ -valued stochastic processes. In order to prove weak convergence in the graphon space, that is, $\tilde{h}^n \Longrightarrow \tilde{h}$ in D , we need to establish (as in the case of processes on any metric space):

- (i) Convergence of finite-dimensional distributions $(\tilde{h}^n(s_i))_{1 \leq i \leq d} \implies (\tilde{h}(s_i))_{1 \leq i \leq d}$ for all points $s_1, \dots, s_d \geq 0$ at which \tilde{h} is continuous almost surely.
- (ii) Tightness of the sequence $(\tilde{h}^n)_{n \in \mathbb{N}}$.

Our first theorem provides equivalent criteria for establishing weak convergence by means of the corresponding subgraph density processes. In the statement and proof, \mathbb{P} will denote the probability measure under which the expectation \mathbb{E} is taken.

Theorem 3.1. *Let \tilde{h} and $(\tilde{h}^n)_{n \in \mathbb{N}}$ be random elements in $D([0, \infty), \widetilde{\mathcal{W}})$. Then the following are equivalent:*

- (i) $\tilde{h}^n \implies \tilde{h}$ as $n \rightarrow \infty$.
- (ii) For all $d \geq 1$ and all graphs $F_1, \dots, F_d \in \mathcal{F}$,

$$(t_{F_1}(\tilde{h}^n), \dots, t_{F_d}(\tilde{h}^n)) \implies (t_{F_1}(\tilde{h}), \dots, t_{F_d}(\tilde{h})) \quad (n \rightarrow \infty). \quad (3.1)$$

- (iii) For every graph $F \in \mathcal{F}$, the sequence $(t_F(\tilde{h}^n))_{n \in \mathbb{N}}$ is tight and, for all $k \geq 1$, all real numbers $0 \leq s_1 < \dots < s_d < \infty$ where \tilde{h} is continuous almost surely, and all graphs $F_1, \dots, F_d \in \mathcal{F}$,

$$\lim_{n \rightarrow \infty} \mathbb{E}\{t_{F_1}(\tilde{h}^n(s_1)) \cdots t_{F_d}(\tilde{h}^n(s_d))\} = \mathbb{E}\{t_{F_1}(\tilde{h}(s_1)) \cdots t_{F_d}(\tilde{h}(s_d))\}. \quad (3.2)$$

For finite graphs it is typically easier to work with *injective* homomorphisms, the set of which we denote by $\text{inj}(F, G)$. If F has k vertices, then for $n \geq k$ there are at most $n_{(k)} := n(n-1) \dots (n-k+1)$ such mappings, which provides a standard normalisation to the count of injective homomorphisms. Thus, if F is a graph on k vertices and G a graph on n vertices, then we define

$$t_F^{\text{inj}}(G) := \frac{|\text{inj}(F, G)|}{n_{(k)}} \in [0, 1]$$

if $k \leq n$ and $t_F^{\text{inj}}(G) := 0$ otherwise. It is easy to see that

$$|t_F^{\text{inj}}(G) - t_F(G)| \leq \frac{C_F}{n} \quad (3.3)$$

for some constant C_F that only depends on F . So, for most purposes the two objects are equivalent in the limit $n \rightarrow \infty$.

We say that $G := (G(s))_{s \geq 0}$ is a *graph process* if the induced graphon process $h^G := (h^{G(s)})_{s \geq 0}$ is a random element of $D([0, \infty), \widetilde{\mathcal{W}})$. Similarly as before, we can consider the process $t_F^{\text{inj}}(G)$ as an element of $D([0, \infty), [0, 1])$. In order to simplify notation, for a sequence of graph processes $(G_n)_{n \in \mathbb{N}}$ we will write $G_n \implies \tilde{h}$ instead of $h^{G_n} \implies \tilde{h}$. We can use (3.3) to prove the following corollary of Theorem 3.1.

Corollary 3.2. *Let $(G_n)_{n \in \mathbb{N}}$ be a sequence of graph processes such that*

$$\inf_{s \geq 0} |V(G_n(s))| \rightarrow \infty \quad (n \rightarrow \infty),$$

and let \tilde{h} be a random element in $D([0, \infty), \tilde{\mathcal{W}})$. Then the following are equivalent:

(i) $G_n \Longrightarrow \tilde{h}$ as $n \rightarrow \infty$.

(ii) For all $d \geq 1$ and all graphs $F_1, \dots, F_d \in \mathcal{F}$,

$$(t_{F_1}^{\text{inj}}(G_n), \dots, t_{F_d}^{\text{inj}}(G_n)) \Longrightarrow (t_{F_1}(\tilde{h}), \dots, t_{F_d}(\tilde{h})) \quad (n \rightarrow \infty). \quad (3.4)$$

(iii) For every graph $F \in \mathcal{F}$, the sequence $(t_F^{\text{inj}}(G_n))_{n \in \mathbb{N}}$ is tight and, for all $d \geq 1$, all real numbers $0 \leq s_1 < \dots < s_d < \infty$ where \tilde{h} is continuous almost surely, and all graphs $F_1, \dots, F_d \in \mathcal{F}$,

$$\lim_{n \rightarrow \infty} \mathbb{E}\{t_{F_1}^{\text{inj}}(G_n(s_1)) \cdots t_{F_d}^{\text{inj}}(G_n(s_d))\} = \mathbb{E}\{t_{F_1}(\tilde{h}(s_1)) \cdots t_{F_d}(\tilde{h}(s_d))\}. \quad (3.5)$$

Proof. Equivalence of (3.1) and (3.4) follows from (3.3) and Slutsky's Theorem. Equivalence of (3.2) and (3.5) is immediate from (3.3) and the bounded convergence theorem, whereas equivalence of tightness in (iii) of Theorem 3.1 and (iii) of Corollary 3.2 is a consequence of (3.3), Ethier and Kurtz (1986, Inequality (6.3), p. 122) and Ethier and Kurtz (1986, Corollary 7.4, p. 129). \square

3.2 Graphon dynamics

We will use the population models described in Section 2.2 to construct a sequence of random graphs that evolve in time and converge weakly in the space of graphons.

Consider the interval $[0, 1]$ with the Euclidean metric, and let $\mathcal{L} = C([0, 1], [0, 1])$, the space of continuous functions from $[0, 1]$ to $[0, 1]$ endowed with the uniform topology, which we will call the space of *fitness landscapes*. Note that, in the definition of \mathcal{L} , the first appearance of $[0, 1]$ represents a space of types, and the second appearance of $[0, 1]$ represents some sort of *fitness* associated with each type through a function from \mathcal{L} .

Here are some specifications that we will assume throughout this section.

- (R1) The connection probabilities between individuals depend on the types and fitness. They will be given by a continuous function $r: [0, 1] \times [0, 1] \rightarrow [0, 1]$.
- (H1) The fitness landscape changes dynamically over time. To represent this, for $m, n \in \mathbb{N}$, we consider $H^{m,n}, H^m, H$ to be random elements in $D([0, \infty), \mathcal{L})$, which will be referred to as *fitness landscape processes*. We write $H^m(s)$ for the fitness landscape at time s , and $H^m(s; x)$ for the fitness of type x at time s .

In what follows, we denote by \bar{F} the right-continuous generalised inverse of a distribution function F with support $[0, 1]$, defined in the usual way as

$$\bar{F}(u) := \inf\{x \in [0, 1]: F(x) > u\}, \quad u \in [0, 1),$$

and, for convenience, we set $\bar{F}(1) := \lim_{x \uparrow 1} \bar{F}(x)$.

Wright-Fisher dynamics. Our first construction of a graphon dynamics comes from the scaling limit of the multi-type Moran model to the multi-dimensional Wright-Fisher diffusion. As discussed in Section 2.2, the multi-type Moran model has $m + 1$ types, which are labelled $0, \dots, m$. Consider n individuals, each carrying one of the types from $\{0, \dots, m\}$. Let $\tau_i^{m,n}(s)$ denote the type of individual i divided by $m + 1$ at time ns ; that is, if at time ns the type of individual i equals k , where $0 \leq k \leq m$, then $\tau_i^{m,n}(s) = k/(m + 1)$.

Discrete graphon dynamics: We can construct the graph $G^{m,n}(s) \in \mathcal{G}_n$ at time $s \geq 0$ by connecting i and j if

$$U_{ij}^n < r(H^{m,n}(s; \tau_i^{m,n}(s)), H^{m,n}(s; \tau_j^{m,n}(s))), \quad (3.6)$$

with $\{U_{ij}^n: n \in \mathbb{N}, 1 \leq i < j \leq n\}$ being a collection of independent uniform random variables on $[0, 1]$, independent of everything else. Thus, $(G^{m,n})_{n \in \mathbb{N}}$ is a sequence of graph processes, which evolve in time due to the induced dynamics of the Moran model on $m + 1$ types. Note that we have scaled time in such a way that the graph $G^{m,n}(s)$ represents the situation of the underlying Moran model at time ns .

Theorem 3.3. *Let $G^{m,n}$ be constructed as above. Assume **(R1)**, **(H1)**, and suppose that $(Y^{m,n}, H^{m,n}) \implies (Y^m, H^m)$ as $n \rightarrow \infty$. Then*

$$G^{m,n} \implies \tilde{h}^m \quad (n \rightarrow \infty),$$

where, for each $s \geq 0$, the equivalence class $\tilde{h}^m(s)$ has a representative $h^m(s)$ of the form

$$h^m(s; x, y) = r(H^m(s; \bar{F}^m(s; x)), H^m(s; \bar{F}^m(s; y))), \quad (x, y) \in [0, 1]^2. \quad (3.7)$$

While the space of types $[0, 1]$ allows for uncountably many types for convenience, in the example of the $(m + 1)$ -Moran model, we will embed the $m + 1$ types into $[0, 1]$ in a canonical way, using the functions $\tau_i^{m,n}(s)$. For both the type space and the fitness space, the interval $[0, 1]$ could be replaced by any other Polish space, as is the case for graphons.

In $(G^{m,n})_{n \in \mathbb{N}}$, the probability for two vertices to be connected depends on their respective fitness, which in the case of (3.6) depends on their types. The limiting graphon is an $(m + 1) \times (m + 1)$ -block graphon whose block boundaries move along the diagonal according to the Wright-Fisher diffusion and whose block heights are controlled by the fitness landscape. As observed for the case $m = 1$ in Section 1, it is not too hard to see that indeed the limiting graphon is a diffusion on the graphon space.

Fleming-Viot dynamics. In the above dynamics the resulting graphon process falls in the class of stochastic block models. Our second theorem provides a graphon dynamics that arises from the Fleming-Viot diffusion and has no block structure in the limit.

Theorem 3.4. *Assume **(R1)**, **(H1)**, and suppose that $(Z^m, H^m) \implies (Z, H)$ as $m \rightarrow \infty$. Let \tilde{h}^m be defined through its representative h^m as given in (3.7). Then*

$$\tilde{h}^m \implies \tilde{h} \quad (m \rightarrow \infty),$$

where, for each $t \geq 0$, the equivalence class $\tilde{h}(s)$ has a representative $h(s)$ of the form

$$h(s; x, y) = r(H(s; \bar{F}(s; x)), H(s; \bar{F}(s; y))), \quad (x, y) \in [0, 1]^2. \quad (3.8)$$

Our H can be any fitness landscape process in **(H1)** and r any graphon in **(R1)**. We will show in Example 4.3 below that these allow for a general class of graphons beyond the stochastic block model process observed earlier.

Also note that the appearance of the inverse distribution function in (3.7), respectively (3.8), can be understood as a change of reference measure of the vertex space from uniform to Z^m , respectively Z . For example, with h^m as in (3.7), subgraph densities with respect to this graphon can be written as

$$\begin{aligned} t_F(h^m(s)) &= \int_{[0,1]^k} \prod_{\{i,j\} \in E(F)} h^m(s; x_i, x_j) dx_1 \dots dx_k \\ &= \int_{[0,1]^k} \prod_{\{i,j\} \in E(F)} r(H^m(s, x_i), H^m(s, x_j)) Z^m(s; dx_1) \dots Z^m(s; dx_k), \end{aligned}$$

and similarly for $t_F(h(s))$ with h as in (3.8). This effect was already observed by Athreya and Röllin (2016) in the context of Respondent Driven Sampling.

4 Examples and Remarks

In this section we illustrate the abstract results in Section 3 via concrete examples of type-connection graphons and fitness landscapes.

4.1 Three examples

Example 4.1. Let

$$H^{m,n}(s; u) = H^m(s; u) = H(s; u) = u.$$

It follows that $G^{m,n}(s) \in \mathcal{G}_n$ is the random graph in which, at time s , individuals i and j are connected by an edge with probability $r(\tau_i^{m,n}(s), \tau_j^{m,n}(s))$, where again $\tau_i^{m,n}(s)$ denotes the type of individual i divided by m at time ns . By Theorem 3.3, as $n \rightarrow \infty$, the graph process $G^{m,n}$ converges to $\tilde{h}^m(s)$ given by (3.7). Thus, individuals i and j are connected with a probability that *does not* depend on the rest of the population. In this case, r will only ever be evaluated at the points $\{0, \frac{1}{m+1}, \dots, \frac{m}{m+1}\} \times \{0, \frac{1}{m+1}, \dots, \frac{m}{m+1}\}$, and so r can be viewed as an $(m+1) \times (m+1)$ type-connection matrix $(r_{ij})_{0 \leq i, j \leq m}$ (the matrix can be extended to a function on $[0, 1]^2$ by an arbitrary continuous interpolation). This leads to the representation

$$h^m(s; x, y) = r_{(m+1)\bar{F}^m(s;x), (m+1)\bar{F}^m(s;y)}.$$

In particular, for the case $m = 1$ (two types), the graphon r can, without loss of generality, be assumed to be given by a (2×2) -matrix, namely,

$$r = \begin{bmatrix} \alpha & \delta \\ \delta & \beta \end{bmatrix},$$

δ	β
α	δ

$Y_0^1(s)$

Figure 2: Graphical representation h^1 , with $Y_0^1(s)$ the fraction of Type 0 in the population at time s .

which yields the representation

$$h^1(s; x, y) = \begin{cases} \alpha & \text{if } (x, y) = [0, Y_0^1(s)] \times [0, Y_0^1(s)], \\ \beta & \text{if } (x, y) = [Y_0^1(s), 1] \times [Y_0^1(s), 1], \\ \delta & \text{if } (x, y) = [0, Y_0^1(s)] \times [Y_0^1(s), 1], \\ \delta & \text{if } (x, y) = [Y_0^1(s), 1] \times [0, Y_0^1(s)]. \end{cases}$$

We have convergence to a graphon-valued Markov process that is driven by Y_0^1 , as illustrated in Figure 2. \square

In the above example we clearly see that, while the proportions of types are diffusive, the connection probabilities within and between types are constant. We next present an example where the connection probabilities diffuse as well.

Example 4.2. Let $r(u, v) = uv$, and define

$$H^{m,n}(s; \frac{\ell}{m+1}) = Y_\ell^{m,n}(s), \quad H^m(s; \frac{\ell}{m+1}) = Y_\ell^m(s), \quad 0 \leq \ell \leq m,$$

and, for $x \neq \frac{\ell}{m+1}$, $0 \leq \ell \leq m$, define $H^{m,n}(s; x)$ and $H^m(s; x)$ via linear interpolation. By Theorem 3.3, as $n \rightarrow \infty$ the graph process $G^{m,n}$ converges to \tilde{h}^m given by its representative

$$h^m(s; x, y) = Y_j^m(s)Y_\ell^m(s),$$

where j and ℓ are such that

$$\sum_{i=0}^{j-1} Y_i^m(s) \leq x < \sum_{i=0}^j Y_i^m(s), \quad \sum_{i=0}^{\ell-1} Y_i^m(s) \leq y < \sum_{i=0}^{\ell} Y_i^m(s)$$

(for $x = 1$, respectively, $y = 1$, we just set $j = m$, respectively, $\ell = m$). \square

If we let $m \rightarrow \infty$ in the above example, then $h^m(s)$ converges to the zero graphon. Consequently, we need to do adapt the fitness landscape in order to obtain a non-trivial limit.

Example 4.3. Let us consider r as above, namely,

$$r(u, v) = uv.$$

Consider a suitable *mutual fitness function*, given by a continuous function $f \in \mathcal{W}$, and let $c \in (0, 1)$ be the *fitness threshold*. For each $m \in \mathbb{N}$ and $0 \leq \ell \leq m$, define the mutual fitness partner sets as

$$A_\ell^m = \left\{ \frac{j}{m+1} : f\left(\frac{\ell}{m+1}, \frac{j}{m+1}\right) \geq c \right\}$$

and define the fitness landscape process as

$$H^m\left(s; \frac{\ell}{m+1}\right) = \sum_{k \in A_\ell^m} Y_k^m(s), \quad 0 \leq \ell \leq m.$$

Thus, individuals i and j are connected with probability

$$r\left(H^m\left(s; \frac{i}{m+1}\right), H^m\left(s; \frac{j}{m+1}\right)\right) = H^m\left(s; \frac{i}{m+1}\right) H^m\left(s; \frac{j}{m+1}\right).$$

We note that, unlike in the earlier example, this probability depends on the rest of the population via the *total fractions* of all individuals whose type has a *sufficiently large mutual fitness* with respect to the type of individual i , respectively, j . For the graphon process we may define

$$H^m(s; u) = \int_{[0,1]} \mathbb{I}[f(\bar{F}^m(s; u), \bar{F}^m(s; v)) \geq c] Z^m(s; dv), \quad u \in (0, 1),$$

and define the corresponding graphon process \tilde{h}^m to be given by its representative

$$h^m(s; x, y) = H^m(s, x) H^m(s, y), \quad x, y, \in (0, 1).$$

It is easy to see that H^m satisfy (H1) with

$$H(s; u) = \int_{[0,1]} \mathbb{I}[f(\bar{F}(s; u), \bar{F}(s; v)) \geq c] Z(s; dv).$$

By Theorem 3.4, we obtain that graphon process \tilde{h}^m in the limit as $m \rightarrow \infty$ converges to \tilde{h} given by its representative

$$h(s; x, y) = H(s, x)H(s, y), \quad x, y, \in (0, 1).$$

Here, r can be viewed as a type-connection matrix or type connection kernel that is controlled by the empirical type distribution of the entire population.

Note that r need not take the product form, as illustrated by the fact that any $r \in \mathcal{W}$ satisfying (R1) will suffice. Thus, the example's framework provides a rich class of non-trivial diffusions in the space of graphons, well beyond the stochastic block model framework. \square

4.2 Remarks

The Moran model, the Wright-Fisher diffusion and the Fleming-Viot diffusion were used to create the graphon dynamics constructed in Theorems 3.3 and 3.4. We mention three generalisations of these models where our results carry over.

- (a) In Section 1, we discussed the Moran model with n individuals, each carrying Type 0 or Type 1, and we identified the limiting process as the Wright-Fisher diffusion on $[0, 1]$ given by the SDE

$$dY(s) = \sqrt{Y(s)(1 - Y(s))} dW(s)$$

with initial condition $Y(0)$, where W is standard Brownian motion. This SDE has a unique strong solution, that is, there is a unique path $s \mapsto Y(s)$ that is measurable with respect to the canonical filtration associated with the Brownian motion. The generator L of the Wright-Fisher diffusion is

$$(L\varphi)(x) = x(1 - x) \frac{\partial^2 \varphi(x)}{\partial x^2}, \quad x \in [0, 1],$$

for test functions $\varphi: [0, 1] \rightarrow \mathbb{R}$ that are twice continuously differentiable. It is possible to generalise the Wright-Fisher diffusion by allowing for a *state-dependent resampling rate*. Indeed, if individuals resample at rate $s(x)$ when the fraction of individuals of type 0 is x , then the generator becomes

$$(L^s \varphi)(x) = s(x) x(1 - x) \frac{\partial^2 \varphi(x)}{\partial x^2}, \quad x \in [0, 1].$$

In order for the SDE to be well-defined, some restrictions need to be imposed on the function s , for instance, $s(x) > 0$ for $x \in (0, 1)$ and $x \mapsto s(x) x(1 - x)$ is Lipschitz on $[0, 1]$ (see Dawson (1993, Section 2)). An example is $s(x) = x(1 - x)$, which corresponds to Ohta-Kimura resampling at a rate that is proportional to the genetic diversity of the population.

- (b) The Moran model with $m + 1$ types is the Markov process on the simplex \mathcal{S}^m with generator

$$(L^m \varphi)(x) = \sum_{0 \leq k < \ell \leq m} x_k x_\ell \frac{\partial^2 \varphi(x)}{\partial x_k \partial x_\ell}, \quad x \in \mathcal{S}^m,$$

for test functions $\varphi: \mathcal{S}^m \rightarrow \mathbb{R}$ that are twice continuously differentiable in each coordinate. Again, it is possible to modulate the resampling rate by a function $s: \mathcal{S}^m \rightarrow [0, \infty)$, but *severe restrictions* need to be imposed on the behaviour of s near the boundary of \mathcal{S}^m (for example, s is close to 1; see Bass and Perkins (2008)).

- (c) The Fleming-Viot diffusion with infinitely many types is the Markov process on the set $\mathcal{P}([0, 1])$ with generator

$$(L\varphi)(x) = \int_{[0, 1]^2} [x(du)\delta_u(dv) - x(du)x(dv)] \frac{\partial^2 \varphi(x)}{\partial x^2} [\delta_u, \delta_v],$$

where

$$\frac{\partial^2 \varphi(x)}{\partial x^2}[\delta_u, \delta_v] = \frac{\partial}{\partial x} \left(\frac{\partial \varphi(x)}{\partial x}[\delta_u] \right)[\delta_v], \quad u, v \in [0, 1],$$

for test functions $\varphi: \mathcal{P}([0, 1]) \rightarrow \mathbb{R}$ of the form

$$\varphi(x) = \int_{[0, 1]^n} x(du_1) \times \cdots \times x(du_n) \psi(u_1, \dots, u_n),$$

where $x \in \mathcal{P}([0, 1])$, $n \in \mathbb{N}$, and $\psi: [0, 1]^n \rightarrow \mathbb{R}$ is continuous. Again, it is possible to modulate the resampling rate by a function $s: \mathcal{P}([0, 1]) \rightarrow [0, \infty)$, but *severe restrictions* need to be imposed on s in order to make sure that the diffusion is well-defined (for example, s is close to 1; see Dawson and March (1995)).

The limiting graphons dynamics in Theorems 3.3 and 3.4 are Markov processes on the space of graphons $(\widetilde{\mathcal{W}}, \delta_{\text{sub}})$. Indeed, because

$$\tilde{h}(s) = G(Y(s)), \quad s \geq 0,$$

for some invertible map $G: \mathcal{S}^m \rightarrow \widetilde{\mathcal{W}}$ or $G: \mathcal{P}([0, 1]) \rightarrow \widetilde{\mathcal{W}}$, the generator \tilde{L} of the graphon dynamics can be *formally* computed as

$$\begin{aligned} (\tilde{L}\varphi)(\tilde{h}(0)) &= \lim_{t \downarrow 0} \frac{\mathbb{E}\{\varphi(\tilde{h}(s)) - \varphi(\tilde{h}(0)) \mid \tilde{h}(0)\}}{s} \\ &= \lim_{t \downarrow 0} \frac{\mathbb{E}\{(\varphi \circ G)(Y(s)) - (\varphi \circ G)(Y(0)) \mid Y(0)\}}{y} \\ &= L(\varphi \circ G)(Y(0)) = L(\varphi \circ G)(G^{-1}(\tilde{h}(0))), \end{aligned}$$

for appropriate test functions $\varphi: \widetilde{\mathcal{W}} \rightarrow \mathbb{R}$. Hence we have the representation

$$\tilde{L}\varphi = L(\varphi \circ G) \circ G^{-1}.$$

An explicit form of the generator may be possible for specific choices of G .

5 Proof of main results

5.1 Proof of Theorem 3.1

Proof. “(i) \implies (ii)”. Since t_F is a continuous function from $(\widetilde{\mathcal{W}}, d_{\text{sub}})$ to $[0, 1]$, and thus continuous from $D([0, \infty), \widetilde{\mathcal{W}})$ to $D([0, \infty), [0, 1])$ by Ethier and Kurtz (1986, Problem 13, p. 151), it is clear that (i) implies (ii) by the continuous mapping theorem.

“(ii) \implies (iii)”. Now, observe that (ii) for $k = 1$ implies tightness of $(t_F(\tilde{h}_n))_{n \in \mathbb{N}}$, which is the first part of (iii). For points of continuity of $(t_{F_1}(\tilde{h}), \dots, t_{F_k}(\tilde{h}))$, (ii) implies that the finite dimensional distributions converge which implies the second part of (iii).

“(iii) \implies (i)”. Note first that functions of the form

$$f = \sum_{i=1}^m a_i t_{F_i}, \quad m \in \mathbb{N}, a_1, \dots, a_m \in \mathbb{R}, F_1, \dots, F_m \in \mathcal{F}, \quad (5.1)$$

form an algebra because, for any two graphs F_1 and F_2 ,

$$t_{F_1}(h)t_{F_2}(h) = t_{F_1 \uplus F_2}(h), \quad h \in \widetilde{\mathcal{W}}, F_1, F_2 \in \mathcal{F},$$

where $F_1 \uplus F_2$ denotes the disjoint union of the two graphs (that is, no merging of vertices), which is again in \mathcal{F} . By Diao et al. (2015, Theorem 2.2), this set of functions is dense in the space of continuous functions on $\widetilde{\mathcal{W}}$ with respect to the topology of uniform convergence (the cited theorem is with respect to a metric that is equivalent to d_{sub} with respect to convergence). Since, by assumption, all $(t_{F_i}(\tilde{h}_n))_{n \in \mathbb{N}}$ are relatively compact, so are finite collections of them, and therefore so are $(f(\tilde{h}_n))_{n \in \mathbb{N}}$ for all f of the form (5.1), which implies relative compactness of $(\tilde{h}_n)_{n \in \mathbb{N}}$ by Ethier and Kurtz (1986, Theorem 9.1, p. 142) (the *compact containment condition* in that theorem trivially holds because $\widetilde{\mathcal{W}}$ itself is compact).

In order to establish convergence of the finite-dimensional distributions of \tilde{h}_n to \tilde{h} , we next show that the family of functions $\{t_F: F \in \mathcal{F}\}$ *strongly separates points* in $(\widetilde{\mathcal{W}}, d_{\text{sub}})$, that is, we show that for each $h \in \widetilde{\mathcal{W}}$ and each $\delta > 0$ there exist $F_1, \dots, F_k \in \mathcal{F}$ such that

$$\inf_{h': d_{\text{sub}}(h, h') \geq \delta} \max_{1 \leq i \leq k} |t_{F_i}(h) - t_{F_i}(h')| > 0$$

(c.f. Ethier and Kurtz (1986, p. 113)). So, let $h \in \widetilde{\mathcal{W}}$ and $\delta > 0$ be given. Let $(F_i)_{i \in \mathbb{N}}$ be the enumeration of \mathcal{F} giving rise to d_{sub} , and choose m such that

$$\sum_{i > m} 2^{-i} \leq \delta/2.$$

If h' is such that $d_{\text{sub}}(h, h') \geq \delta$, then we must have

$$\sum_{i=1}^m 2^{-i} |t_{F_i}(h) - t_{F_i}(h')| \geq \frac{\delta}{2}. \quad (5.2)$$

If all differences $|t_{F_i}(h) - t_{F_i}(h')|$ were strictly smaller than $\delta/2m$, then we would have

$$\sum_{i=1}^m 2^{-i} |t_{F_i}(h) - t_{F_i}(h')| \leq \sum_{i=1}^m |t_{F_i}(h) - t_{F_i}(h')| < \delta/2,$$

which contradicts (5.2), and so at least one of the differences must be larger than $\delta/(2m)$. Hence

$$\inf_{h': d_{\text{sub}}(h, h') \geq \delta} \max_{1 \leq i \leq m} |t_{F_i}(h) - t_{F_i}(h')| \geq \delta/2m,$$

which establishes that $\{t_F: F \in \mathcal{F}\}$ strongly separate points, and therefore, by Ethier and Kurtz (1986, Theorem 4.5, p.113), is convergence determining. By Ethier and Kurtz (1986, Proposition 4.6(b), p.115), functions of the form $t_{F_1} \cdots t_{F_k}$ are convergence determining on the product space $\widetilde{\mathcal{W}}^k$ equipped with the canonical metric induced by d_{sub} , and so convergence of the finite-dimensional distributions follows from (3.2). This establishes (i). \square

5.2 Proof of Theorem 3.3

Proof. Using the Skorokhod embedding, we may assume that the processes $X^{m,n}$, $Y^{m,n}$, $H^{m,n}$, Y^m and H^m are constructed in such a way that

$$(Y^{m,n}, H^{m,n}) \longrightarrow (Y^m, H^m) \quad (n \rightarrow \infty) \quad \mathbb{P}\text{-almost surely.} \quad (5.3)$$

We write \mathbb{P}' and \mathbb{E}' to denote conditional probability and conditional expectation given $X^{m,n}$, $Y^{m,n}$, $H^{m,n}$, Y^m and H^m . We will establish (iii) of Theorem 3.1 to prove our result. We do this in two steps.

Step 1: Proof of (3.2). We first prove a univariate concentration argument of the subgraph densities around their mean. Fix $s \geq 0$ and a finite graph F . Recall McDiarmid's concentration inequality from McDiarmid (1989). If $W = (W_1, \dots, W_N)$ is a collection of N independent random variables and f is a function of N variables such that an arbitrary change of the i -th variable changes the value of f by at most c_i , then

$$\mathbb{P}[|f(W) - \mathbb{E}\{f(W)\}| \geq \varepsilon] \leq 2 \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^N c_i^2}\right). \quad (5.4)$$

Since F has k vertices and $G^{m,n}(s)$ has n vertices, it follows that $t_F^{\text{inj}}(G^{m,n}(s))$ changes by at most $\binom{n-2}{k-2} / \binom{n}{k} = \binom{k}{2} / \binom{n}{2}$ when one edge is changed. Applying (5.4) with $N = \binom{n}{2}$ we have, for any $\varepsilon > 0$,

$$\mathbb{P}'\left[|t_F^{\text{inj}}(G^{m,n}(s)) - \mathbb{E}'\{t_F^{\text{inj}}(G^{m,n}(s))\}| > \varepsilon\right] \leq 2 \exp\left(-2\varepsilon^2 \binom{n}{2} \binom{k}{2}^{-2}\right). \quad (5.5)$$

The mean of $t_F^{\text{inj}}(G^{m,n}(s))$ with respect to the connection probabilities can be expressed as

$$\begin{aligned} & \mathbb{E}'\{t_F^{\text{inj}}(G^{m,n}(s))\} \\ &= \sum_{\ell_1, \dots, \ell_k=0}^m p_{X_0^{m,n}(ns), \dots, X_m^{m,n}(ns)}(\ell_1, \dots, \ell_k) \\ & \quad \times \prod_{\{i,j\} \in E(F)} r\left(H^{m,n}\left(s; \frac{\ell_i}{m+1}\right), H^{m,n}\left(s; \frac{\ell_j}{m+1}\right)\right), \end{aligned} \quad (5.6)$$

where $p_{K_0, \dots, K_m}(\ell_1, \dots, \ell_k)$ is the probability that an ordered sample of k balls drawn without replacement from an urn with K_i balls of colour i , where $0 \leq i \leq m$, is such

that the i -th ball has colour ℓ_i , $0 \leq i \leq m$. We note that there is a constant $C > 0$ (which may depend on m and k but not on n) such that

$$\left| p_{X_0^{m,n}(ns), \dots, X_m^{m,n}(ns)}(\ell_1, \dots, \ell_k) - \prod_{i=1}^k Y_{\ell_i}^{m,n}(s) \right| \leq \frac{C}{n} \quad \text{for all } n \in \mathbb{N}. \quad (5.7)$$

Using continuity of r , (5.3) and (5.7), we conclude that, \mathbb{P} -almost surely,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{E}' \{ t_F^{\text{inj}}(G^{m,n}(s)) \} \\ &= \sum_{\ell_1, \dots, \ell_k=0}^m \prod_{i=1}^k Y_{\ell_i}^m(s) \prod_{\{i,j\} \in E(F)} r\left(H^m\left(s; \frac{\ell_i}{m+1}\right), H^m\left(s; \frac{\ell_j}{m+1}\right)\right) \\ &= t_F(\tilde{h}^m(s)) \end{aligned} \quad (5.8)$$

whenever s is a point at which (Y^m, H^m) is continuous almost surely; here we also need the fact that $H^{m,n}(s) \in C([0, 1], [0, 1])$, so that, for x fixed, $H^{m,n}(s; x)$ is a continuous function of $H^{m,n}(s)$ and converges to $H^m(s; x)$. Using Borel-Cantelli with (5.5), followed by (5.8), we obtain

$$t_F^{\text{inj}}(G^{m,n}(s)) \longrightarrow t_F(\tilde{h}^m(s)) \quad (n \rightarrow \infty) \quad \mathbb{P}\text{-almost surely.}$$

Let $d \geq 1$, let $0 < s_1 < s_2 < \dots < s_d$ be points at which (Y^m, H^m) is continuous almost surely, and let F_1, F_2, \dots, F_d be finite sub-graphs. Then, by the above, we have

$$\prod_{i=1}^d t_{F_i}^{\text{inj}}(G^{m,n}(s_i)) \longrightarrow \prod_{i=1}^d t_{F_i}(\tilde{h}^m(s_i)) \quad (n \rightarrow \infty) \quad \mathbb{P}\text{-almost surely.}$$

Using the bounded convergence theorem, we conclude that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \prod_{i=1}^d t_{F_i}^{\text{inj}}(G^{m,n}(s_i)) \right\} = \mathbb{E} \left\{ \prod_{i=1}^d t_{F_i}(\tilde{h}^m(s_i)) \right\},$$

which establishes (3.2).

Step 2: Tightness. Fix a finite graph F on k vertices. We need to show that the family of processes $(t_F^{\text{inj}}(G^{m,n}))_{n \in \mathbb{N}}$ is tight. We show this via tightness of some intermediate processes. First, for each $n \geq 1$, let

$$W^{m,n}(s) := \sum_{\ell_1, \dots, \ell_k=0}^m \prod_{i=1}^k Y_{\ell_i}^{m,n}(s) \prod_{\{i,j\} \in E(F)} r\left(H^{m,n}\left(s; \frac{\ell_i}{m+1}\right), H^{m,n}\left(s; \frac{\ell_j}{m+1}\right)\right), \quad s \geq 0.$$

Note that $W^{m,n}$ is a continuous function of $Y^{m,n}$ and $H^{m,n}$. Since $(Y^{m,n}, H^{m,n})_{n \geq 1}$ converges weakly and is therefore tight, and since compact sets remain compact under continuous mappings, it follows that $(W^{m,n})_{n \geq 1}$ is tight. Second, for each $n \geq 1$, let

$$V^{m,n}(s) := \mathbb{E}' \{ t_F^{\text{inj}}(G^{m,n}(s)) \}, \quad s \geq 0.$$

Recalling (5.6) and applying (5.7), it follows that

$$\sup_{s \geq 0} |V^{m,n}(s) - W^{m,n}(s)| \leq \frac{C}{n},$$

and so, by Ethier and Kurtz (1986, Inequality (6.3), p. 122) and Ethier and Kurtz (1986, Corollary 7.4, p. 129) and tightness of $(W^{m,n})_{n \geq 1}$, it follows that $(V^{m,n})_{n \geq 1}$ is also tight.

Third, observe that the sequence of events when vertices resample their type follows a Poisson point process with rate n (since there are n vertices and each vertex resamples at rate 1), so that the number of resampling events in the time interval $[0, Tn]$, where $T > 0$ is fixed, follows a Poisson distribution with mean $n \times Tn$. Denote by $A_{n,m}$ the event that the process $Y^{m,n}$ saw no more than n^3 such resampling events in the interval $[0, T]$, and so, letting $Z \sim \text{Po}(n^2T)$, we have

$$\mathbb{P}[A_{m,n}^c] = \mathbb{P}[Z > n^3] \leq \frac{\mathbb{E}Z^2}{n^6} \leq \frac{2T^2}{n^2}, \quad n^2 \geq \frac{1}{T}. \quad (5.9)$$

On the event $A_{n,m}$, we now apply a union bound over all resampling events and then apply (5.5), which yields

$$\mathbb{P}' \left[\sup_{0 \leq s \leq T} |t_F^{\text{inj}}(G^{m,n}(s)) - V^{m,n}(s)| > \varepsilon \mid A_{n,m} \right] \leq 2n^3 \exp \left(-2\varepsilon^2 \binom{n}{2} \binom{k}{2}^{-2} \right).$$

Choosing

$$\varepsilon^2 = 2 \log(n) \binom{k}{2}^2 \binom{n}{2}^{-1},$$

we conclude that

$$\mathbb{P}' \left[\sup_{0 \leq s \leq T} |t_F^{\text{inj}}(G^{m,n}(s)) - V^{m,n}(s)| > \frac{k \log n}{n} \mid A_{n,m} \right] \leq \frac{2}{n}. \quad (5.10)$$

Using (5.9) and (5.10) and tightness of $(V^{m,n})_{n \geq 1}$, it is now straightforward to check Condition (b) of Ethier and Kurtz (1986, Corollary 7.4, p. 129) by also using Ethier and Kurtz (1986, Inequality (6.3), p. 122), and it follows that $(t_F^{\text{inj}}(G^{m,n}))_{n \geq 1}$ is tight. \square

5.3 Proof of Theorem 3.4

Proof. Using the Skorokhod embedding, we may assume that the processes Z^m, H^m, Z and Z^m are constructed in such a way that

$$(Z^m, H^m) \longrightarrow (Z, H) \quad (m \rightarrow \infty) \quad \mathbb{P}\text{-almost surely.} \quad (5.11)$$

We first establish the following fact. As before, denote by $\mathcal{P}([0, 1])$ the set of probability measures on $[0, 1]$ endowed with the topology of weak convergence, and let $\lambda : \mathcal{P}[0, 1] \times C([0, 1], [0, 1]) \rightarrow \mathbb{R}$ be defined as

$$\lambda(\mu, f) = \int_{[0,1]^k} \prod_{\{i,j\} \in E(F)} r(f(x_i), f(x_j)) \mu(dx_1) \cdots \mu(dx_k).$$

Then λ is a continuous function. To see this, take a sequence (μ_n, f_n) that converges to (μ, f) . Write

$$|\lambda(\mu_n, f_n) - \lambda(\mu, f)| \leq |\lambda(\mu_n, f_n) - \lambda(\mu_n, f)| + |\lambda(\mu_n, f) - \lambda(\mu, f)| =: r_{1,n} + r_{2,n}.$$

In order to show that $r_{1,n}$ converges to zero, it suffices to show that the quantity $r'_{1,n} := \int r(f_n(x), f_n(y))\pi(dx)\pi(dy)$ converges to zero uniformly in $\pi \in \mathcal{P}([0, 1])$ and then use a telescoping sum and the fact that r is bounded by 1. But since r and f are defined on compact sets and are therefore uniformly continuous, it is easy to establish that $r'_{1,n}$ converges to zero uniformly in π . The fact that $r_{2,n}$ converges to zero follows from the fact that μ_n converges weakly to μ and that therefore the k -fold convolution converges weakly, too.

We now proceed to establish (iii) of Theorem 3.1.

Step 1: Proof of (3.2). For every simple finite graph F on k vertices, (2.2) yields the representations

$$t_F(\tilde{h}^m(s)) = \lambda(Z^m, H^m(s)), \quad t_F(\tilde{h}(s)) = \lambda(Z, H(s)).$$

Let s be a point at which (Z, H) is continuous almost surely. By continuity of λ , we conclude that

$$\lim_{m \rightarrow \infty} t_F(\tilde{h}^m(s)) = t_F(\tilde{h}(s)) \quad \text{almost surely.}$$

For $d \geq 1$ and successive times $0 \leq s_1 < \dots < s_d < \infty$ at which (Z, H) is continuous almost surely, (3.2) can now be established in the same way as in the previous proof.

Step 2: Tightness. Fix a finite graph F on k vertices; we need to show that the family of processes $(t_F(\tilde{h}^m))_{m \geq 1}$ is tight. Noting that the function that maps the process (Z^m, H^m) to $\lambda \circ (Z^m, H^m)$ is continuous since λ is continuous (c.f. Ethier and Kurtz (1986, Problem 13, p. 151)), tightness is immediate because it is preserved under continuous mappings. \square

Acknowledgements

SA was supported in part through an ISF-UGC grant and a CPDA from the Indian Statistical Institute. FdH was supported through NWO Gravitation Grant NETWORKS-024.002.003. AR was supported through Singapore Ministry of Education Academic Research Fund Tier 2 grant MOE2018-T2-2-076. Part of this work was done while SA was visiting the Indian Institute of Science, Bangalore, and SA and FdH were visiting the International Center for Theoretical Sciences, Bangalore. The hospitality that was provided at these institutes is greatly appreciated.

References

D.J. Aldous (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11**, 581–598.

- S. Athreya and A. Röllin (2016). Dense graph limits under respondent-driven sampling. *Ann. Appl. Probab.* **44**, 2193–2210.
- R.F. Bass and E.A. Perkins (2008). Degenerate stochastic differential equations arising from catalytic branching networks. *Electron. J. Probab.* **13**, 1808–1885.
- P. Billingsley (1999). *Convergence of Probability Measures* (second edition). John Wiley & Sons, Inc., New York.
- B. Bollobás and O. Riordan (2009). Metrics for sparse graphs. In *Surveys in combinatorics 2009*, volume 365 of *London Math. Soc. Lecture Note Ser.*, pages 211–287. Cambridge University Press, Cambridge.
- C. Borgs, J. Chayes, L. Lovász, V. Sós and K. Vesztegombi (2008). Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics* **219**, 1801–1851.
- J. Černý and A. Klimovsky (2018). Markovian dynamics of exchangeable arrays. *arXiv preprint arXiv:1810.1316a5*.
- H. Crane (2016). Dynamic random networks and their graph limits. *Ann. Appl. Probab.* **26**, 691–721.
- D.A. Dawson (1993). Measure-valued Markov processes. In *École d’été de Probabilités de Saint-Flour XXI-1991*. pages 7–249. Springer.
- D.A. Dawson and P. March (1995). Resolvent estimates for Fleming-Viot operators and uniqueness of solutions to related martingale problems. *J. Funct. Anal.* **132**, 417–472.
- P. Diaconis and S. Janson (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)* **28**, 33–61.
- P. Diao, D. Guillot, A. Khare and B. Rajaratnam (2015). Differential calculus on graphon space. *J. Combin. Theory Ser. A* **133**, 183–227.
- S.N. Ethier and T.G. Kurtz (1986). *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York. Characterization and convergence.
- P.W. Holland and S. Leinhardt (1977). A dynamic model for social networks. *Journal of Mathematical Sociology* **5**, 5–20.
- P. Holme and J. Saramäki (2012). Temporal networks. *Physics Reports* **519**, 97–125.
- D.N. Hoover (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ 2*.
- S. Karlin and H.E. Taylor (1981). *A Second Course in Stochastic Processes* Elsevier.
- W.O. Kermack and A.G. McKendrick (1927). Contributions to the mathematical theory of epidemics. In *Proc. R. Soc. Lond. A*, volume 115, pages 700–721.
- J. Leskovec (2008). Dynamics of large networks *Doctoral dissertation*, Carnegie Mellon University, School of Computer Science, Machine Learning Department.
- D. Levin and Y. Peres (2017). *Markov Chains and Mixing Times* (second edition). American Mathematical Society.
- L. Lovász (2012). *Large Networks and Graph Limits*. American Mathematical Society.
- L. Lovász and B. Szegedy (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96**, 933–957.
- C. McDiarmid (1989). On the method of bounded differences. In *Surveys in Combinatorics*, London Mathematical Society Lecture Note Series, pages 148–188. Cambridge University Press.
- M. Morris and M. Kretzschmar (1997). Concurrent partnerships and the spread of HIV. *Aids*

11, 641–648.

T.A.B. Snijders (2001). The statistical evaluation of social network dynamics. *Sociological Methodology* **31**, 361–395.

T.A.B. Snijders, J. Koskinen, and M. Schweinberger (2010). Maximum likelihood estimation for social network dynamics. *Ann. Appl. Statist.* **4**, 567–588.