

# From flip processes to dynamical systems on graphons

Frederik Garbe<sup>1,a</sup>, Jan Hladký<sup>2,b</sup>, Matas Šileikis<sup>2,c</sup> and Fiona Skerman<sup>3,d</sup>

<sup>1</sup>Faculty of Informatics, Masaryk University, Brno, Czechia, <sup>a</sup>[garbe@fi.muni.cz](mailto:garbe@fi.muni.cz)

<sup>2</sup>Institute of Computer Science of the Czech Academy of Sciences, Prague, Czechia, <sup>b</sup>[hladky@cs.cas.cz](mailto:hladky@cs.cas.cz), <sup>c</sup>[matas@cs.cas.cz](mailto:matas@cs.cas.cz)

<sup>3</sup>Department of Mathematics, Uppsala University, Uppsala, Sweden, <sup>d</sup>[fiona.skerman@math.uu.se](mailto:fiona.skerman@math.uu.se)

Received 9 June 2022; revised 30 March 2023; accepted 14 May 2023

**Abstract.** We introduce a class of random graph processes, which we call *flip processes*. Each such process is given by a function  $\mathcal{R} : \mathcal{H}_k \rightarrow \mathcal{H}_k$  from all labeled  $k$ -vertex graphs into itself ( $k$  is fixed). The process starts with a given  $n$ -vertex graph  $G_0$ . In each step, the graph  $G_i$  is obtained by sampling  $k$  random vertices  $v_1, \dots, v_k$  of  $G_{i-1}$  and replacing the induced graph  $F := G_{i-1}[v_1, \dots, v_k]$  by  $\mathcal{R}(F)$ . Actually, our definition is more general, in that  $\mathcal{R}(F)$  is a probability distribution on  $\mathcal{H}_k$ , thus allowing randomised replacements.

Given a flip process  $\mathcal{R}$ , we construct time-indexed trajectories  $\Phi : \mathcal{W}_0 \times [0, \infty) \rightarrow \mathcal{W}_0$  in the space of graphons. We prove that for any  $T > 0$  starting with a large  $n$ -vertex graph  $G_0$  which is close to a graphon  $W_0$ , with high probability the flip process stays in a thin sausage around  $(\Phi(W_0, t))_{t=0}^T$  (after rescaling the time by  $n^2$ ).

Among others, we study continuity properties of these trajectories with respect to time and initial graphon, existence and stability of fixed points and speed of convergence. We give an example of a flip process with a periodic trajectory.

**Résumé.** Nous introduisons une classe de processus aléatoires sur graphes, que nous appelons *processus de retournement* (*flip processes*). Chacun de ces processus est donné par une règle qui est une fonction  $\mathcal{R} : \mathcal{H}_k \rightarrow \mathcal{H}_k$  de tous les graphes à  $k$  sommets étiquetés vers lui-même ( $k$  est fixe). Le processus commence avec un graphe  $G_0$  à  $n$  sommets. À chaque étape, le graphe  $G_i$  est obtenu en échantillonnant  $k$  sommets aléatoires  $v_1, \dots, v_k$  de  $G_{i-1}$  et en remplaçant le graphe induit  $F := G_{i-1}[v_1, \dots, v_k]$  par  $\mathcal{R}(F)$ . Cette classe contient plusieurs processus connus comme le graphe aléatoire Erdős–Rényi et le processus de suppression de triangles. En fait, notre définition est plus générale, dans le sens que  $\mathcal{R}(F)$  est une distribution de probabilité sur  $\mathcal{H}_k$ , permettant ainsi des remplacements aléatoires.

Étant donné un processus de retournement de règle  $\mathcal{R}$ , on construit des trajectoires indexées en temps  $\Phi : \mathcal{W}_0 \times [0, \infty) \rightarrow \mathcal{W}_0$  dans l'espace des graphons. Nous prouvons que pour chaque  $T > 0$  et chaque (grand) graphe fini  $G_0$  qui est proche d'un graphon  $W_0$  dans la norme de coupe, avec une forte probabilité le processus de retournement initié à  $G_0$  restera dans une fine région autour de la trajectoire  $(\Phi(W_0, t))_{t=0}^T$  (après remise à l'échelle du temps par le carré de l'ordre du graphe).

Ces trajectoires de graphons sont ensuite étudiées du point de vue des systèmes dynamiques. Entre autres sujets, nous étudions (i) les propriétés de continuité de ces trajectoires par rapport au temps et au graphon initial, (ii) l'existence et la stabilité des points fixes et (iii) la vitesse de convergence (lorsque la limite de temps infinie existe). Nous donnons un exemple de processus de retournement avec une trajectoire périodique.

*MSC2020 subject classifications:* 05C80

*Keywords:* Random graph processes; Evolving networks; Graph limits

## 1. Introduction

### 1.1. The Erdős–Rényi random graph process

The uniform Erdős–Rényi random graph model  $\mathbb{G}(n, M)$  is 63 years old and now the theory of random graphs is an active field at the border between graph theory and probability. While studying this model provides insights in many situations, sometimes one is in an evolving setting, of which  $\mathbb{G}(n, M)$  is just one snapshot. More precisely, the *Erdős–Rényi random graph process*, for a given integer  $n$ , is a random sequence of graphs  $G_0, G_1, \dots, G_{\binom{n}{2}}$ , all on vertex set  $[n]$ , in which  $G_0$

is edgeless and each  $G_{\ell+1}$  is created from  $G_\ell$  by flipping a uniformly random non-edge into an edge. Lastly,  $G_{\binom{n}{2}}$  is a complete graph and so there is no meaningful way to continue the process.

It is easy to prove that as  $n$  goes to infinity, all the graphs  $G_0, G_1, \dots, G_{\binom{n}{2}}$  are  $o(1)$ -quasirandom, asymptotically almost surely. Here and throughout the paper, we will use the concept of quasirandomness which was developed in the 1980s by Chung, Graham and Wilson [10] and by Thomason [25] and reflected the then emerging theory surrounding the Szemerédi regularity lemma [22]. Namely, we say that an  $n$  vertex graph is  $\varepsilon$ -quasirandom of density  $d$  if it has  $d\binom{n}{2}$  edges and each set of vertices  $U$  induces  $\binom{|U|}{2}d \pm \varepsilon n^2$  edges.

Let us consider a small modification of this process, which we call the *Erdős–Rényi flip process*. We again start with the edgeless graph on  $[n]$ . To create  $G_{\ell+1}$  from  $G_\ell$ , we pick a uniformly random pair  $xy$  of vertices and make that pair an edge. In particular, if  $xy$  was an edge, then  $G_{\ell+1} = G_\ell$ . Let us make some remarks about the Erdős–Rényi flip process. Firstly, this process is defined for any number of steps, and eventually stabilises at the complete graph (around time  $\tau_{ER} = n^2 \ln n$ ). Secondly, the probability that a fixed pair  $xy$  forms an edge in a graph  $G_\ell$  is equal to

$$1 - \left(1 - \frac{1}{\binom{n}{2}}\right)^\ell \approx 1 - \exp\left(-\frac{2\ell}{n^2}\right),$$

and indeed it is easy to prove that throughout the evolution the graphs  $G_\ell$  stay quasirandom, of density around

$$(1) \quad 1 - \exp(-2t)$$

for each  $t > 0$  and  $\ell = tn^2$ , asymptotically almost surely.

Note the number of edges in the Erdős–Rényi flip process after  $k$  steps is not concentrated on one number (unlike in the uniform Erdős–Rényi random graph) and is not binomial either (unlike in the binomial Erdős–Rényi random graph). However, both these models are good proxies since our view and results will be insensitive to changes of  $o(n^2)$  edges, as usual in the theory of dense graph limits.

### 1.2. The triangle removal process

The *triangle removal process* was introduced by Bollobás and Erdős around 1990. We start with the complete graph  $G_0$ . Each  $G_{\ell+1}$  is created from  $G_\ell$  by firstly selecting a uniformly random triangle in  $G_\ell$  and then deleting all its three edges. Thus, the number of edges in each step of the process decreases by three and the process cannot be continued when we arrive at a triangle-free graph. Observe this stopping time is random. A result of Bohman, Frieze, and Lubetzky [6] says that there are  $n^{3/2+o(1)}$  edges left in this final graph, asymptotically almost surely. To show this, Bohman, Frieze, and Lubetzky set up complicated machinery which allows them to control quasirandomness of the sequence during the evolution. In particular, throughout the evolution, the graphs stay  $o(1)$ -quasirandom asymptotically almost surely.

As above, we can modify the triangle removal process into the *triangle removal flip process*. In this modification, in each step we sample a triple of vertices uniformly at random and if that triple induces a triangle, then we remove all the three edges. Note that the triangle removal flip process is exactly the triangle removal process, where between each pair of consecutive steps of the triangle removal process we insert a random number of copies of the first graph of the pair, the random number being geometrically distributed with a parameter corresponding to the density of triangles of this first graph. It can be proven that this process reaches a triangle-free graph in  $\tau_{TR} = n^{4+o(1)}$  steps asymptotically almost surely, at which point it stabilises (it is not trivial to prove this, however we will only need it for informal comments).

Let us now try to understand the evolution of the density in the triangle removal flip process. Suppose that the density of a graph  $G_\ell$  is  $d$ . Then, because  $G_\ell$  is quasirandom, the probability that a random triple of vertices induces a triangle is roughly  $d^3$ . In other words, if we fix  $\varepsilon > 0$ , then the number of triangle removals between rounds  $\ell$  and  $\ell + \varepsilon n^2$  is around  $\varepsilon d^3 n^2$  and hence the density drops to around  $d - 6\varepsilon d^3$ . Note that one source of imprecision is the assumption that within this window, a triangle is still sampled with probability  $d^3$  which is true only at the beginning of the window. This imprecision, however, is negligible if  $\varepsilon$  is small. So, if for  $t \geq 0$  we denote by  $d(t)$  the density of the graph  $G_{\lfloor tn^2 \rfloor}$  and we treat  $d(t)$  as a number (actually, it is a random variable, though highly concentrated), then the heuristics given above suggest a differential equation

$$(2) \quad \begin{aligned} d(0) &= 1, \\ d'(t) &= -6d(t)^3, \end{aligned}$$

which has a unique solution on  $[0, \infty)$ ,

$$(3) \quad d(t) = (1 + 12t)^{-1/2}.$$

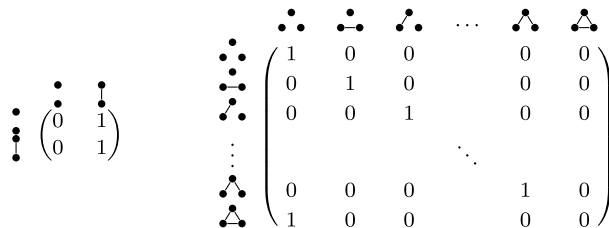


Fig. 1. Rules of the Erdős–Rényi and triangle removal flip processes.

1.3. Flip processes

We can now generalise the idea of local replacements that defined the Erdős–Rényi flip process and the triangle removal flip process.

We first define a subclass of flip processes which we call *definite flip processes of order k*. Let  $\mathcal{H}_k$  be the family of graphs on vertex set  $[k]$ . A *rule* is an arbitrary function  $\mathcal{R} : \mathcal{H}_k \rightarrow \mathcal{H}_k$ . We then consider a random discrete time process  $(G_\ell : \ell = 0, 1, \dots)$  of graphs on the vertex set  $[n]$ , where  $n \geq k$  is arbitrary. We are free to choose the initial graph  $G_0$  (deterministically or randomly). In each step, the graph  $G_{\ell+1}$  is obtained by sampling  $k$  random vertices  $v_1, \dots, v_k$  of  $G_\ell$  (conditioned to be distinct) and replacing the induced graph  $G_\ell[v_1, \dots, v_k]$  – which we call the *drawn graph* – by  $\mathcal{R}(G_\ell[v_1, \dots, v_k])$  – which we call the *replacement graph*.

Clearly, the Erdős–Rényi flip process and the triangle removal flip process are definite flip processes of order 2 and 3, respectively, see Figure 1. Also, note that there is no need to start the Erdős–Rényi flip process with the edgeless graph and the triangle removal flip process with the complete graph. For example, if we start the triangle removal flip process with a triangle-free graph, then it stabilises immediately.

To generalise these definite flip processes we allow additional randomisation in the replacement procedure. A *rule of a flip process of order k* is a matrix  $\mathcal{R} \in [0, 1]^{\mathcal{H}_k \times \mathcal{H}_k}$  such that for every  $H \in \mathcal{H}_k$  we have

$$(4) \quad \sum_{J \in \mathcal{H}_k} \mathcal{R}_{H,J} = 1.$$

So, given a rule  $\mathcal{R}$  and integer  $n \geq k$ , the *flip process with rule  $\mathcal{R}$*  is a discrete time process  $(G_\ell : \ell = 0, \dots)$  of graphs on the vertex set  $[n]$ . Again, we start with a given graph  $G_0$ , and for  $\ell \geq 0$  we obtain graph  $G_{\ell+1}$  from graph  $G_\ell$  as follows. Sample, uniformly at random, an ordered tuple  $\mathbf{v} = (v_1, \dots, v_k)$  of distinct vertices. Let  $H \in \mathcal{H}_k$  be the unique graph, called the *drawn graph*, such that  $v_i v_j$  is an edge in  $G_\ell$  if and only if  $ij \in E(H)$ . Sample a graph  $J$  from the *replacement distribution (for the given drawn graph  $H$ )*  $(\mathcal{R}_{H,J})_J$  and replace  $G_\ell[\mathbf{v}]$  by  $J$ , that is, turn  $v_i v_j$  into an edge if  $ij \in J$  and  $v_i v_j$  into a non-edge if  $ij \notin J$ . We call the resulting graph  $G_{\ell+1}$ . Note that the order of the elements in  $\mathbf{v}$  matters.

For brevity we will sometimes say “flip process  $\mathcal{R}$ ” meaning “flip process with rule  $\mathcal{R}$ ”, having in mind, precisely speaking, the class of processes (given by all possible initial graphs  $G_0$ ).

Flip processes are defined for all times. However, we shall only be looking at typical behaviours at times  $x \cdot n^2$ ,  $x \in [0, +\infty)$ , which we call *bounded horizon*. While there are certainly interesting features at later times, as we saw with  $\tau_{ER}$  and  $\tau_{TR}$  in Sections 1.1 and 1.2, these features cannot be captured by our theory.

Given a process of order  $k$  with rule  $\mathcal{R}$ , we call a graph  $H \in \mathcal{H}_k$  *idle* if  $\mathcal{R}_{H,H} = 1$ . We say that the process is *trivial* if each graph is idle.

Observe that each flip process is a time-homogeneous Markov process.

Let us give a sample questions we address in this paper (of course, with respect to the asymptotics  $n \rightarrow \infty$ ; all statements hold asymptotically almost surely; the number 99 below is given only for the sake of concreteness).

- (Q1) Is there a flip process which would start with a quasirandom graph of order  $n$  and density 0.5 and in  $99n^2$  steps produces a complete balanced bipartite graph? *No for two reasons. Firstly, flip processes cannot create additional macroscopic structure in a bounded horizon. Secondly, they cannot achieve densities 0 or 1 (unless the corresponding part of the graph had these densities initially) in a bounded horizon.*
- (Q2) Is there a flip process which would start with a complete balanced bipartite graph and in  $99n^2$  steps produces a quasirandom graph of density 0.5? *No, flip processes cannot entirely wipe out the macroscopic structure in a bounded horizon.*
- (Q3) Is there a flip process and an initial graph  $G_0$  so that the evolution is  $99n^2$ -periodic? That is, we want that  $G_\ell$  is similar to  $G_{\ell+99n^2}$  for all  $\ell$ , while it is different from, say,  $G_{\ell+40n^2}$ . *Yes, we can construct an example of such a periodic process.*

- (Q4) Suppose our partner runs a flip process with a known rule  $\mathcal{R}$  but a hidden initial graph  $G_0$  for  $99n^2$  steps and shows us the final graph  $G_{99n^2}$ . Is there a way to approximately reconstruct  $G_0$ ? *Yes.*
- (Q5) Is there a flip process where a small change in the initial graph  $G_0$  would result in a large change in  $G_{99n^2}$ ? *No.*

### 1.4. Passing to graphons

In this section, we assume basic familiarity with the theory of graphons. We give more details in Section 3.2. In particular,  $\mathcal{W}_0$  is the set of all graphons, that is, symmetric measurable functions defined on the square of a probability space  $\Omega^2$  and with values in the interval  $[0, 1]$ . We write  $d_{\square}(U, W) = \|U - W\|_{\square}$  for the cut norm distance between  $U$  and  $W$ , and regard  $\mathcal{W}_0$  as a metric space with metric  $d_{\square}$ .

As we noted earlier, the Erdős–Rényi flip process, when started from the edgeless graph on a vertex set  $[n]$ , is typically quasirandom, with edge density around  $1 - e^{-2t}$  in step  $tn^2$ . We can thus say that after rescaling the time by  $n^2$ , the Erdős–Rényi flip process typically evolves like the function  $t \mapsto \Phi_{\text{ER}}^t$ , where  $\Phi_{\text{ER}}^t \equiv 1 - e^{-2t}$  is the constant graphon with density  $(1 - e^{-2t})$ .

Similarly, we saw that in the triangle removal flip process, when started from the complete graph, the graphs are typically quasirandom with edge density around  $(1 + 12t)^{-1/2}$  at step  $tn^2$ . We can thus say that after rescaling the time by  $n^2$ , the triangle removal flip process typically evolves like the function  $t \mapsto \Phi_{\text{TR}}^t$ , where  $\Phi_{\text{TR}}^t \equiv U_{(1+12t)^{-1/2}}$  is the constant graphon with density  $(1 + 12t)^{-1/2}$ .

In general, given a fixed rule  $\mathcal{R}$  of order  $k$ , we shall construct *trajectories*, that is, a we shall define the evolution function  $\Phi : \mathcal{W}_0 \times [0, +\infty) \rightarrow \mathcal{W}_0$ , where we treat the first coordinate as the *initial graphon*, and the second coordinate as the *time*. Instead of writing  $\Phi(W, t)$  we put the time in the superscript as  $\Phi^t(W)$ . It will turn out that the trajectories correspond to a dynamical system generalising the differential equation (2). However, (2) was a differential equation with a single numerical parameter (the edge density). In the general setting, the differential equation will involve a graphon-valued trajectory. So, the key objects here are Banach-space-valued differential equations (the main differential equation is given in (27) below). The main statement about the existence of trajectories is Theorem 4.5.

As a preview of our general results in Section 4, in Figure 2 we illustrate some trajectories that take values among non-constant graphons.

The *Transference Theorem* (Theorem 5.1) says that if we start with a large initial graph  $G_0$  of order  $n$ , fix a representation of its vertices on  $\Omega$  and hence obtain a graphon representation  $W_0$  of  $G_0$ , then, with high probability, each graph  $G_{tn^2}$  (where  $t \in [0, \infty)$ ) when represented as a graphon, is close to  $\Phi^t(W_0)$  in the cut norm. Obviously, we then must have  $\Phi^0(U) = U$  for each graphon  $U$ . Furthermore, the trajectories are continuous with respect to time, and with respect to the initial graphon. The last fact means that  $G_{tn^2}$  is also close to  $\Phi^t(U)$ , if  $W_0$  is close to  $U$  in the cut norm. Let us state an informal version of this key result here.

**Theorem (Informal and simplified consequence of Theorems 4.5, 4.16 and 5.1).** *Suppose that  $\mathcal{R}$  is a rule. Then there exists  $\Phi : \mathcal{W}_0 \times [0, +\infty) \rightarrow \mathcal{W}_0$  with the following property. Suppose that  $T > 0$ . Suppose that  $G_0$  is an  $n$ -vertex graph and  $U$  is a graphon such that the graphon representation  $W_0$  of  $G_0$  satisfies  $d_{\square}(W_0, U) = o(1)$ .*

*Consider the flip process  $(G_i)_{i \geq 0}$  with rule  $\mathcal{R}$  started from  $G_0$ . Write  $W_i$  for the graphon representation of  $G_i$ . Then with probability  $1 - o(1)$  we have  $\max_{i \in \mathbb{N} \cap [0, Tn^2]} d_{\square}(W_i, \Phi^{i/n^2}U) = o(1)$ . (The asymptotics  $o(1)$  is for fixed  $\mathcal{R}$ ,  $T$  and as  $n \rightarrow \infty$ .)*

Note that we equipped the space of graphons with the cut norm distance  $d_{\square}$  rather than the cut distance  $\delta_{\square}$  (which factors out isomorphic graphons). This allows us not only to track how an initial graph evolves up to isomorphism, but also tells us how individual parts of the graph evolve. We give a specific example in Figure 3.

So, Theorem 5.1 transfers problems about random discrete trajectories on graphs to problems about deterministic continuous trajectories on graphons. With this latter view, combinatorial problems such as those we asked in Section 1.3 translate to problems studied in dynamical systems such as the existence and stability of fixed points or periodic trajectories.

This opens a new perspective and a new set of problems: we answer many of them, yet many more are left unexplored.

### 1.5. Related work

In addition to the Erdős–Rényi and triangle removal processes there are many other graph processes considered in the literature, the  $H$ -free process (see [7] and references therein), the Achlioptas process (see [20] and references therein), or the preferential attachment process (see [5]), to name a few. These processes do not fall into our framework, and in

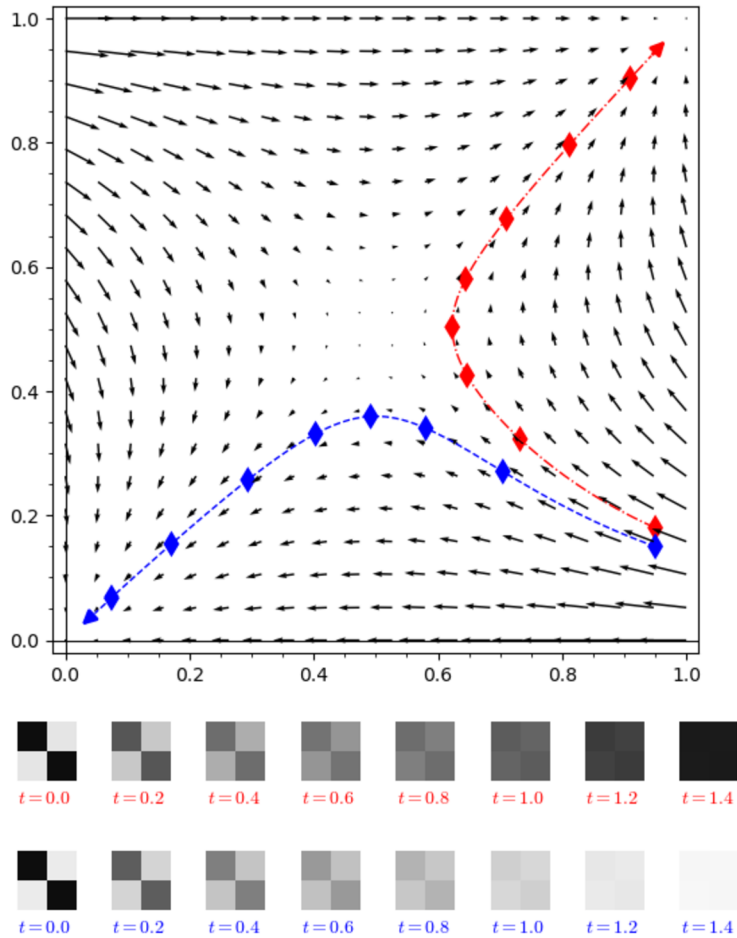


Fig. 2. Extremist flip process of order three works as follows: if the drawn graph has two or three edges we replace it with the triangle and otherwise we replace it by the edgeless graph. Consider a class of step graphons  $\{U_{x,y} : x, y \in [0, 1]\}$  with steps  $\Omega_1, \Omega_2$  of measure  $1/2$ , where  $U_{x,y}$  takes value  $x$  on  $\Omega_1^2 \cup \Omega_2^2$  and value  $y$  elsewhere. We represent  $U_{x,y}$  by a point  $(x, y) \in [0, 1]^2$ . It turns out that trajectories starting in this class do not leave it. The arrows in the diagram depict the tangent vectors of such trajectories. The red line is the trajectory from initial graphon  $U_{0.95,0.18}$  with diamond markers showing the positions along the trajectory at times  $t = 0, t = 0.2, \dots, t = 1.4$ ; the graphon represented by each marker is plotted in grayscale underneath the diagram. The blue line is similar, from initial graphon  $U_{0.95,0.15}$ .

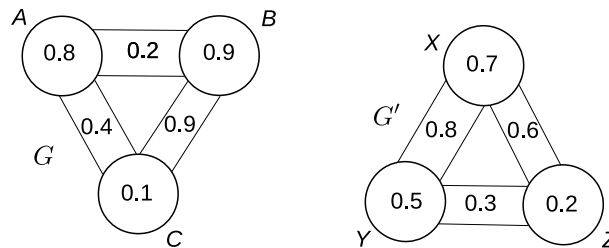


Fig. 3. Two graphs  $G$  and  $G'$ , consisting of three blobs  $A, B, C$  and  $X, Y, Z$ , respectively, each of order  $n/3$ . Inside the blobs and between the pairs of blobs we have quasirandom and quasirandom bipartite graphs, with densities as indicated. Suppose that we start a flip process with  $G$ . Then our results not only say whether after a certain number of steps we typically arrive to a graph with a structure as in  $G'$ , but they also give, in the positive case, the correspondence of the sets  $(A, B, C)$  and of  $(X, Y, Z)$ .

particular are typically studied in the sparse regime. An interesting open direction, suggested to us by Mihyun Kang, is to study modifications of flip processes which would incorporate some features of these other standard models.

For example one could model preferential attachment by introducing ‘types’ of vertices. The initial graph would consist of a single edge between two ‘active’ vertices (or loop at one ‘active’ vertex) and very many isolated ‘unused’ vertices. The rule would then be to pick two vertices at random and if one of them is ‘active’ and one ‘unused’ add an edge between them and change the type of the ‘unused’ vertex to ‘active’; otherwise do nothing. Note the rate of growth

would depend on the number of ‘unused’ vertices in the construction. Further generalisations with rules of higher order are also possible.

Another related work is that of Keliger [15]. In that paper, vertices have a finite number of types, and an update rule at each vertex takes into account the type of that vertex and distribution of the types among the neighborhood. The paper focuses on dense graphs and uses a somewhat similar theory of ODEs.

A different line of research looks at stochastic processes on graphs which satisfy abstract properties and studies their projection into the space of graphons. This direction was started by Crane [11], who studied exchangeable<sup>1</sup> stochastic processes on graphs with countably many vertices. Roughly speaking, leaving out various regularity conditions, they proved that these correspond to Markov processes in the space of graphons. In the follow-up work of Athreya, den Hollander, and Röllin [3], more complicated processes on graphs are studied, to which the corresponding graphon processes exhibit diffusive behavior.

Last, let us draw a connection between our work and *Wormald’s differential equation method* (DEM). This method was introduced in [26], unifying many previous ad-hoc analyses, mostly of randomized algorithms on random graphs. In such a setting one needs to keep track of random numerical parameters (say, the number of matched vertices in some sequential matching algorithm), and there is a heuristic suggesting that these parameters in expectation evolve according to a certain (deterministic) differential equation. DEM then gives general conditions which guarantee that with high probability the parameters indeed do follow the evolution predicted by the differential equation. In that sense there is a clear similarity to our Transference Theorem (Theorem 5.1). The main difference is that DEM links typical behavior of numerical parameters, whereas Theorem 5.1 links typical behaviour of the entire structure of the evolving graph (described by the language of graphons) to a graphon-valued differential equation.

## 1.6. Organisation of the paper

We start by giving examples of several interesting classes of flip processes in Section 2. It uses only a moderate amount of references to later parts of the paper and is therefore a good starting point for reading the paper. Section 3 gives necessary preliminaries, mostly on graphons. In Section 4 we construct trajectories  $\Phi : \mathcal{W}_0 \times [0, +\infty) \rightarrow \mathcal{W}_0$  for an arbitrary flip process as advertised in Section 1.4. In fact, the construction extends also to negative times, as was already hinted by (Q4) in Section 1.3. In Section 4, we also state basic properties of trajectories, such as continuity mentioned in (Q5). In Section 5 we state and prove the Transference Theorem, which tells us that a typical evolution of a flip process on a finite graph stays close to the graphon trajectory. In Section 6 we establish many general features of trajectories, including those that settle questions (Q1)–(Q3).

## 2. Examples

We believe that our framework provides countless interesting specific examples. Below we give seven classes of flip processes which we found interesting. See [2] for details. In this section,  $k$  is the order of the flip process in question.

### 2.1. Ignorant flip processes

Ignorant flip processes are processes where the replacement graph does not depend on the drawn graph. More formally, we require that the rule  $\mathcal{R}$  of an ignorant flip process satisfies for each  $H, H', J \in \mathcal{H}_k$  that  $\mathcal{R}_{H,J} = \mathcal{R}_{H',J}$ . We then have the *ignorant replacement distribution*  $\mathcal{D} = \{\mathcal{R}_{H,J}\}_J$  (here,  $H$  is arbitrary), and its *average density*  $d := \frac{1}{\binom{k}{2}} \cdot \mathbb{E}_{J \in \mathcal{D}}[e(J)]$ .

Observe that the Erdős–Rényi flip process is an ignorant flip process of order  $k = 2$  with ignorant replacement distribution which is the Dirac measure on the edge  $K_2$ .

A crucial (and easy to establish) property is that each ignorant process of order  $k \geq 2$  has all trajectories converging to the constant graphon corresponding to the average density of its replacement distribution.

### 2.2. Monotone flip processes

Roughly speaking, a monotone flip process should be one, in which the replacement graph is ‘monotone’ with respect to the drawn graph. Besides an obvious choice whether ‘monotone’ should mean non-decreasing or non-increasing, and whether we require non-strict monotonicity or strict monotonicity,<sup>2</sup> there are three further natural definitions: (we now stick to the choice of non-decreasing flip processes):

<sup>1</sup>A processes is *exchangeable* if it is invariant under arbitrary relabeling of vertices by finite permutations. The concept of exchangeability is connected to graph limits as explained in [4,12].

<sup>2</sup>Of course, this strictness has to be exempted when the drawn graph is complete (for increasing flip processes) or edgeless (for decreasing flip processes).



- *Pointwise non-decreasing flip processes* in which we require that  $\mathcal{R}_{H,J} = 0$  unless  $J$  is a supergraph of  $H$  (as a labelled graph).
- *Edge non-decreasing flip processes* in which we require that  $\mathcal{R}_{H,J} = 0$  unless  $e(J) \geq e(H)$ .
- *Averaged-edge non-decreasing flip processes* in which we require that  $\sum_J \mathcal{R}_{H,J} \cdot e(J) \geq e(H)$  for every  $H$ .

Obviously, the ‘averaged-edge’ version is weaker than the ‘edge’ version which is weaker than the pointwise version. Edge non-decreasing flip processes (and thus pointwise non-decreasing flip processes) are always convergent whereas averaged-edge non-decreasing flip processes need not be.

### 2.3. Removal flip processes

These flip processes generalise the triangle removal flip process. That is, we have a graph  $H \in \mathcal{H}_k$ . When the drawn graph  $H$  is a supergraph of  $F$ , the replacement graph is  $H \setminus F$ . Other drawn graphs are idle.

Obviously, each removal flip process is a pointwise non-increasing flip process. The set of destinations of all trajectories is the set  $\mathcal{W}_F$  of  $F$ -free graphons, that is, the graphons  $W$  with  $t(F, W) = 0$ .

The question of speed of convergence in removal flip processes is a non-trivial one with a connection to the removal lemma.

### 2.4. The complementing flip process

The complementing flip process is a process in which the drawn graph is replaced by its complement.

Trajectories of this flip process converge very quickly to the constant- $\frac{1}{2}$  graphon.

### 2.5. The component completion flip process

The component completion flip process is a process in which the drawn graph is replaced by its *component closure*, that is the graph is a disjoint union of cliques which has the same component structure.

For  $k \geq 3$ , the trajectories of a component completion process converge to a component closure of the original graphon.

### 2.6. The stirring flip process

The stirring process is defined in two variants, *firm* or *loose*, as follows. Suppose that the drawn graph is  $H \in \mathcal{H}_k$ . Set  $M := e(H)$  and  $p := M / \binom{k}{2}$ . The replacement graph is the uniform Erdős–Rényi random graph  $\mathbb{G}(k, M)$  (the firm variant) or the binomial Erdős–Rényi random graph  $\mathbb{G}(k, p)$  (the loose variant).

So, in the firm stirring process we always preserve the number of edges of the underlying graph, and in the loose stirring process we preserve it in expectation.

Each trajectory converges to a constant graphon of the same density as the initial graph. For initial graphons that are regular (i.e., all the degrees are equal to a constant  $d$ ) these trajectories are the same as for an ignorant process of replacement distribution with average density  $d$ . However, for non-regular initial graphons, the trajectories are different than trajectories of ignorant processes.

### 2.7. The extremist process

The extremist process is defined as follows. Suppose  $H \in \mathcal{H}_k$  is the sampled graph. Set  $\ell := e(H)$ . If  $\ell > \binom{k}{2}/2$ , then the replacement graph is the complete graph. If  $\ell < \binom{k}{2}/2$ , then the replacement graph is the edgeless graph. If  $\ell = \binom{k}{2}/2$ , then  $\mathcal{R}$  is idle.

If  $W$  is a constant- $d$  graphon, then the trajectory converges to constant-0 for  $d < \frac{1}{2}$  and to constant-1 for  $d > \frac{1}{2}$ . Trajectories starting at non-constant graphons are more subtle. For  $k \geq 5$  there exist a *density threshold*  $d_k > 0$  such that the trajectory of each graphon of edge density less than  $d_k$  converges to constant-0, and symmetrically, the trajectory of each graphon of edge density more than  $1 - d_k$  converges to constant-1. Further,  $d_k \rightarrow \frac{1}{2}$  as  $k \rightarrow \infty$ . No positive density thresholds exist for  $k = 3, 4$ .

## 3. Preliminaries

### 3.1. Basic notation and simple facts

#### 3.1.1. Basic notation

We write  $(n)_k$  for the *falling factorial*,  $(n)_k := n \cdot (n - 1) \cdot \dots \cdot (n - k + 1)$ . Given a set  $A$ , we write  $\mathbb{1}_A$  for the indicator function of  $A$ . The symmetric difference is denoted by  $\ominus$ .

### 3.1.2. Graphs

All graphs in this paper are simple, loopless and undirected. Given a graph  $G = (V, E) = (V(G), E(G))$ , we write  $v(G) = |V|$  and  $e(G) = |E|$ . Given two sets  $A, B \subset V$ , we write  $e_G(A, B)$  for the number of pairs  $(a, b) \in A \times B$  such that  $ab \in E$ . In particular, edges in  $A \cap B$  are counted twice.

Let  $k \in \mathbb{N}$ . We denote the family of all labelled graphs on the vertex set  $[k]$  by  $\mathcal{H}_k$ . Given a labeled graph  $F = (V(F), E(F))$  and an ordered subset  $R \subseteq V(F)$  of vertices, we call  $(R, V(F), E(F))$  a *rooted graph*. Whenever  $R = (a, b)$  is an ordered pair, we simply denote  $(R, V(F), E(F))$  by  $F^{a,b}$ .

### 3.2. Graphons

Our notation concerning graphons mostly follows [16], which is recommended for further reading.

All our graphons will be defined on an atomless measure space  $(\Omega, \pi)$ . The sigma-algebra is given implicitly and is assumed to be *separable*, that is, there is a countable set  $\mathcal{C}$  of measurable sets such that for every measurable set  $S$  and every  $\delta > 0$  there is  $T \in \mathcal{C}$  such that  $\pi(S \ominus T) < \delta$ . We will often also make the measure implicit and, in particular, for  $k \in \mathbb{N}$  we write  $\Omega^k$  for the product probability space. By  $\mathcal{W}$  we denote the space of measurable functions  $W : \Omega^2 \rightarrow \mathbb{R}$  which are symmetric (i.e.  $W(x, y) = W(y, x)$  almost everywhere<sup>3</sup>) and bounded (i.e.,  $\text{ess sup } |W| < \infty$ ). We identify two functions that are equal a.e. and tacitly assume  $\mathcal{W}$  consists of equivalence classes of such functions. Hence  $\mathcal{W}$  is a closed linear subspace of the space  $L^\infty(\Omega^2)$  and thus Banach when endowed with the norm  $\|\cdot\|_\infty$ . We, however, will also consider other norms on  $\mathcal{W}$ . We call elements of  $\mathcal{W}$  *kernels* and elements of  $\mathcal{W}_0 := \{W \in \mathcal{W} : W \in [0, 1]\}$  *graphons*. Here and elsewhere, we write  $W \in S$  to denote that up to a null set the  $W$  takes values in  $S$ , that is, the set  $\{(x, y) \in \Omega^2 : W(x, y) \notin S\}$  is null. We will also write, for two measurable functions defined on the same measure space,  $f = g$  and  $f \leq g$  meaning that equality or inequality holds a.e. With this notation, we recall that the *essential supremum*  $\text{ess sup } f$  is the least constant  $c$  such that  $f \leq c$  and the *essential infimum*  $\text{ess inf } f$  is defined analogously.

The set  $\mathcal{W}_0$  is a closed  $\|\cdot\|_\infty$ -ball of radius  $1/2$  centered at the constant- $\frac{1}{2}$  graphon. For  $\varepsilon \geq -1/2$ , we define a ball with the same center but radius  $\frac{1}{2} + \varepsilon$ ,

$$(5) \quad \mathcal{W}_0^\varepsilon := \{W \in \mathcal{W} : W \in [-\varepsilon, 1 + \varepsilon]\}.$$

#### 3.2.1. Graphon representation

Suppose that  $G = (V, E)$  is a graph. Suppose that we have a partition  $(\Omega_v : v \in V)$  of  $\Omega$  into sets of measure  $1/|V|$  each. Then we define the *graphon representation of  $G$* , which we will denote by  $W_G$ , as the graphon  $W : \Omega^2 \rightarrow \{0, 1\}$  defined to be 1 on each rectangle  $\Omega_u \times \Omega_v$  for which  $uv \in E$ , and 0 otherwise. Note that this graphon representation may depend on the choice of the partition  $(\Omega_v : v \in V)$ . However, we will frequently keep the partition implicit. The important thing is that we will assume that graphs on the same vertex set are represented using the same partition of  $\Omega$ .

#### 3.2.2. Cut norm, cut distance and densities

The *cut norm* on  $\mathcal{W}$  is defined to be

$$(6) \quad \|W\|_\square := \sup_{A, B \subseteq \Omega} \left| \int_{A \times B} W \, d\pi^2 \right|,$$

where the supremum is taken over measurable sets.

We will occasionally use the obvious inequalities

$$(7) \quad \|W\|_\square \leq \|W\|_1 \leq \|W\|_\infty, \quad W \in \mathcal{W}.$$

The metric induced by the cut norm is called the *cut norm distance*,

$$d_\square(U, W) := \|U - W\|_\square.$$

The *cut distance* is defined for  $U, W \in \mathcal{W}$  as

$$(8) \quad \delta_\square(U, W) := \inf_{\varphi \in \mathbb{S}_\Omega} d_\square(U, W^\varphi),$$

where  $W^\varphi(x, y) = W(\varphi(x), \varphi(y))$  and  $\mathbb{S}_\Omega$  is the set of all measure preserving bijections from  $\Omega$  to  $\Omega$ .

<sup>3</sup>Here and below we say *almost everywhere* or *a.e.* when we mean  $\mu$ -almost everywhere, whenever the measure  $\mu$  is clear from the context; say, in this particular instance  $\mu = \pi^2$ .



**Remark 3.1.** It is known that  $\delta_{\square}(U, W) = 0$  if and only if  $U$  and  $W$  are *weakly isomorphic*, that is, when for every finite graph  $F$ , we have  $t(F, U) = t(F, W)$  (using the notion of densities defined in Definition 3.3 below). Since weak isomorphism is an equivalence relation, we define  $\widetilde{\mathcal{W}}_0$  to be  $\mathcal{W}_0$  with equivalent graphons identified, which, together with  $\delta_{\square}$  turns  $\widetilde{\mathcal{W}}_0$  into a metric space.

The following fact will also be useful.

**Lemma 3.2.** *The metric space  $(\mathcal{W}_0, d_{\square})$  is complete.*

A proof of a substantially more general result can be found in [19, Proposition 1]. We give a self-contained proof of Lemma 3.2 in the Appendix.

For a function  $f : \Omega^{[k]} \rightarrow \mathbb{R}$  and a subset  $S \subset [k]$  we write

$$\int f \, d\pi^S = \int_{\Omega^S} f(x_i : i \in [k]) \prod_{i \in S} \pi(dx_i).$$

We use this notation to introduce the key concept of density of a finite graph in a kernel.

**Definition 3.3 (Densities in kernels).** Given a kernel  $W \in \mathcal{W}$  and a graph  $F = (V, E)$ , we define the *density* of  $F$  in  $W$  as

$$t(F, W) := \int \prod_{ij \in E} W(x_i, x_j) \, d\pi^V,$$

and the *induced density* as

$$t_{\text{ind}}(F, W) := \int \prod_{ij \in E} W(x_i, x_j) \prod_{ij \notin E} (1 - W(x_i, x_j)) \, d\pi^V.$$

Furthermore, we need to work with a notion of a rooted density.

**Definition 3.4 (Rooted densities in kernels).** Given a kernel  $W \in \mathcal{W}$ , a rooted graph  $F = (R, V, E)$ , and  $\mathbf{x} \in \Omega^R$  we define

$$(9) \quad t^{\mathbf{x}}(F, W) := \int \prod_{ij \in E} W(x_i, x_j) \, d\pi^{V \setminus R}, \quad \text{and}$$

$$(10) \quad t_{\text{ind}}^{\mathbf{x}}(F, W) := \int \prod_{ij \in E} W(x_i, x_j) \prod_{ij \notin E} [1 - W(x_i, x_j)] \, d\pi^{V \setminus R}.$$

Note that  $\mathbf{x} \mapsto t^{\mathbf{x}}(F, W)$  and  $\mathbf{x} \mapsto t_{\text{ind}}^{\mathbf{x}}(F, W)$  are functions in  $L^{\infty}(\Omega^R)$  and are not necessarily symmetric.

For convenience we define the operator  $T_F^{a,b} : \mathcal{W} \rightarrow L^{\infty}(\Omega^2)$ , by setting

$$(11) \quad (T_F^{a,b} W)(x, y) = t_{\text{ind}}^{(x,y)}(F^{a,b}, W).$$

For future reference we note that for any rooted graph  $F$  and any  $k \geq v(F)$ ,

$$(12) \quad t^{\mathbf{x}}(F, W) = \sum_{H \in \mathcal{H}_k : E(F) \subseteq E(H)} t_{\text{ind}}^{\mathbf{x}}(H, W),$$

where in the summation  $F$  is treated as a fixed rooted labeled graph, say, with root  $R = [R]$  and vertex set  $V(F) = [v(F)]$ , and each  $H$  is treated as the rooted graph  $(R, [k], E(H))$ . Moreover, for any two graphs  $F, G$  with  $v(F) \leq v(G)$  and for a graphon representation  $W_G$  of  $G$  we have (see [16, Exercise 7.7]) that

$$(13) \quad |t_{\text{ind}}(F, G) - t_{\text{ind}}(F, W_G)| \leq \frac{\binom{v(F)}{2}}{v(G)}.$$

Similarly, if  $W_G$  is a graphon representation of  $G$  with respect to a partition  $(\Omega_v : v \in V(G))$ , then for every  $u, v \in V(G)$  and  $x \in \Omega_u, y \in \Omega_v$ , we have

$$(14) \quad |t_{\text{ind}}^{(u,v)}(F, G) - t_{\text{ind}}^{(x,y)}(F, W_G)| \leq \frac{\binom{v(F)}{2}}{v(G)}.$$

(The proof is also an easy exercise.)

Next, we need a bound on expansion of the operator  $T_F^{a,b}$  with respect to the  $L^\infty$ -metric.

**Lemma 3.5.** *Let  $F \in \mathcal{H}_k$  and  $a, b \in [k]$  be two distinct vertices. Let  $\varepsilon \geq 0$  be a constant. If  $U, W \in \mathcal{W}_0^\varepsilon$ , then*

$$\|T_F^{a,b}U - T_F^{a,b}W\|_\infty \leq (1 + \varepsilon) \binom{k}{2}^{-1} \binom{k}{2} \|U - W\|_\infty.$$

We postpone the proof of the lemma to Appendix B.

### 3.2.3. Decorated homomorphisms

We will also need to consider homomorphism densities of graphs in  $m$ -tuples of graphons which can be viewed as a density of edge-coloured subgraphs, where each graphon describes a colour class.

**Definition 3.6 (Densities in vectors of kernels).** Given a graph  $F = (V, E)$  and a tuple  $\mathbf{W} = (W_e : e \in E)$  of kernels, we set

$$t(F, \mathbf{W}) := \int \prod_{ij \in E} W_{ij}(x_i, x_j) d\pi^V.$$

We can define similarly the induced density. This time, suppose that we have a graph  $F$ , a complete graph  $K$  on  $V(F)$ , and a tuple  $\mathbf{W} = (W_e : e \in E(K))$  of kernels. Set

$$t_{\text{ind}}(F, \mathbf{W}) := \int \prod_{ij \in E(K) \setminus E(F)} (1 - W_{ij}(x_i, x_j)) \prod_{ij \in E(F)} W_{ij}(x_i, x_j) d\pi^{V(F)}.$$

We will make use of a counting lemma similar to [16, Lemma 10.24]. It is convenient to extend this to kernels. A proof can be found in Appendix B.

**Lemma 3.7.** *Given  $\varepsilon \geq 0$ , a simple graph  $F = (V, E)$ , and two tuples  $\mathbf{U} = (U_e \in \mathcal{W}_0^\varepsilon : e \in E)$  and  $\mathbf{W} = (W_e \in \mathcal{W}_0^\varepsilon : e \in E)$  we have*

$$|t(F, \mathbf{U}) - t(F, \mathbf{W})| \leq (1 + 2\varepsilon)^{|E|-1} \sum_{e \in E} \|U_e - W_e\|_\square.$$

### 3.2.4. Step functions and twins

We say that  $f \in L^\infty(\Omega^2)$  is a *step function* if there is a partition (finite or countable)  $S_1 \cup S_2 \cup \dots$  of  $\Omega$  into sets (steps) of positive measure so that  $f$  is constant almost everywhere on each set  $S_i \times S_j$ . We say that a partition into steps is *minimal* for  $f$  if  $f$  is not a step function with respect to any coarser partition of  $\Omega$ . A *step graphon* is a graphon which is a step function.

We say that  $A \subset \Omega$  is a *twin-set* of a kernel  $W$  if there exists a conull set  $A' \subset A$  such that for all  $x, x' \in A'$ , the (single variable) functions  $W(x, \cdot)$  and  $W(x', \cdot)$  are equal a.e. A variant of this concept is studied in detail in [16, Section 13.1.1]. Later we shall need the following easy result, a proof of which can be found in Appendix B.

**Lemma 3.8.** *Suppose that  $U, U_1, U_2, \dots$  are kernels and  $\lim_{n \rightarrow \infty} \|U_n - U\|_\square = 0$ . If a set  $A \subset \Omega$  has positive measure and is a twin-set for each  $U_n$ , then  $A$  is a twin-set for  $U$ .*

### 3.2.5. Sampling from a graphon

Each graphon  $W$  defines a probability distribution on  $\mathcal{H}_k$  as follows. Sample  $k$  independent  $\pi$ -random elements  $\mathbf{U}_v, v \in [k]$  of  $\Omega$  and then, for each pair of distinct vertices  $u, v \in [k]$ , make  $uv$  an edge independently with probability  $W(\mathbf{U}_u, \mathbf{U}_v)$ . For a random graph  $G$  obtained this way we write  $G \sim \mathbb{G}(k, W)$  or even use  $\mathbb{G}(k, W)$  to denote the random graph itself.

Sometimes it will be convenient to write the rooted density defined in (10) in a probabilistic form as

$$(15) \quad t_{\text{ind}}^{\mathbf{x}}(F, W) = \mathbb{P}(\mathbb{G}(k, W) = F \mid \mathbf{U}_v = x_v : v \in R).$$

### 3.3. Differential equations

To develop our framework we will need some theory of differential equations, both for real-valued functions and for Banach-space-valued functions. This is because our trajectories on graphons will be implicitly defined as solutions to differential equations with the derivative which is a function of the flip rule and current position/graphon.

#### 3.3.1. Grönwall's inequality

Among real-valued functions, the differential equation

$$(16) \quad f'(t) = Cf(t)$$

with the initial condition  $f(0) = z$  has a unique solution  $f(t) = ze^{Ct}$ . We will often be able to assume only a differential *inequality*, which implies that a function satisfying it is bounded by the solution of a corresponding equation. A statement of this kind are known as Grönwall's inequality. There are many versions of it with different assumptions on signs of constants and continuity/differentiability properties of the function in question. We shall need two versions; the first is borrowed from the literature and the second one (which we could not find a reference for) is derived using standard tools.

**Lemma 3.9 (Integral Grönwall's inequality, [23, Theorem 1.10]).** *Let  $f : [t_0, t_1] \rightarrow [0, \infty)$  be a continuous non-negative function. Suppose that for some constants  $A \geq 0, C > 0$  we have*

$$(17) \quad f(t) \leq A + C \int_{t_0}^t f(x) \, dx$$

for every  $t \in [t_0, t_1]$ . Then we have

$$(18) \quad f(t) \leq Ae^{C(t-t_0)} \quad \text{for all } t \in [t_0, t_1].$$

**Lemma 3.10 (Differential Grönwall's inequality).** *Let  $F : [t_0, t_1] \rightarrow \mathbb{R}$  be an absolutely continuous function, that is, there exists an integrable function  $f : [t_0, t_1] \rightarrow \mathbb{R}$  such that*

$$F(t) = F(t_0) + \int_{t_0}^t f(x) \, dx$$

for all  $t \in [t_0, t_1]$ . Suppose that for some constant  $C \in \mathbb{R}$

$$(19) \quad f(t) \leq CF(t)$$

for almost every  $t \in [t_0, t_1]$ . Then we have

$$(20) \quad F(t) \leq F(t_0)e^{C(t-t_0)} \quad \text{for all } t \in [t_0, t_1].$$

Moreover if (19) holds with  $\geq$  instead of  $\leq$ , then (20) holds with  $\geq$  instead of  $\leq$ .

**Proof.** We revisit the proof of [23, Theorem 1.12]. Define a function  $G : [t_0, t_1] \rightarrow \mathbb{R}$  by

$$G(t) = F(t)e^{-C(t-t_0)}.$$

Since  $G(t)$  is a product of two absolutely continuous functions, it is absolutely continuous (see for example [21, §6.4, Exercise 42]). Hence, we can use the product rule (see for example [21, §6.5, Exercise 52]),

$$G(t) = G(t_0) + \int_{t_0}^t (f(x)e^{-C(x-t_0)} + F(x)(-Ce^{-C(x-t_0)})) \, dx.$$

By (19) the integrand is non-positive almost everywhere, whence  $G(t) \leq G(t_0)$ , which is equivalent to (20). If we assume the opposite inequality in (19), then we get the opposite inequality in (20).  $\square$

### 3.3.2. Banach-space-valued differential equations

A key step in our construction involves a solution to a first-order autonomous differential equation. This solution is however not real-valued but rather function-valued. This falls within the more general setting of differential equations in Banach spaces. Recall Definition C.1 of the *derivative*.

The following theorem is a basic statement about local existence and uniqueness of solutions to ordinary differential equations in Banach spaces. For a reference, see, e.g., [1], Lemma 4.1.6 (existence and uniqueness) and Lemma 4.1.8 (Lipschitz continuity on the initial condition<sup>4</sup>).

**Theorem 3.11.** *Consider a Banach space  $\mathcal{E}$  with norm  $\|\cdot\|$ . Let  $\mathcal{U} \subset \mathcal{E}$  be an open set, and  $K, M > 0$  be constants. Let  $X : \mathcal{U} \rightarrow \mathcal{E}$  be a  $K$ -Lipschitz function, that is*

$$\|X(x) - X(y)\| \leq K\|x - y\| \quad \text{for all } x, y \in \mathcal{U}.$$

Suppose that

$$(21) \quad \sup\{\|X(x)\| : x \in \mathcal{U}\} \leq M.$$

Let  $x_0 \in \mathcal{U}$ . If  $b$  is a positive number such that the closed ball  $\mathcal{B}(x_0, b) := \{x \in \mathcal{E} : \|x - x_0\| \leq b\}$  is contained in  $\mathcal{U}$ , then for  $\alpha := \min\{1/K, b/M\}$  there is a unique continuously differentiable function  $f = f_{x_0} : [-\alpha, \alpha] \rightarrow \mathcal{U}$  satisfying the equation

$$f'(t) = X(f(t))$$

and the initial value condition  $f(0) = x_0$ .

Furthermore, if  $x_0, y_0 \in \mathcal{U}$  are such that  $\mathcal{B}(x_0, b), \mathcal{B}(y_0, b)$  are contained in  $\mathcal{U}$  and  $f_{x_0}$  and  $f_{y_0}$  correspond, respectively, to the initial values  $x_0$  and  $y_0$ , then for any  $t \in [-\alpha, \alpha]$  we have

$$(22) \quad \|f_{x_0}(t) - f_{y_0}(t)\| \leq e^{K|t|}\|x_0 - y_0\|.$$

## 4. Trajectories

### 4.1. Heuristics for the trajectories

Let us explain the main idea behind constructing trajectories  $\Phi : \mathcal{W}_0 \times [0, +\infty) \rightarrow \mathcal{W}_0$  for a given flip process. As we explained in Section 1.4, the main motivation for trajectories is that we want to have a Transference Theorem, which says that if  $G_0$  is an  $n$ -vertex graph which is close in cut norm to a graphon  $W_0$ , then performing  $tn^2$  steps of the flip process we get a graph  $G_{tn^2}$ , which is, with high probability, close to  $\Phi^t(W_0)$ . The Markov property of the flip process can be informally rephrased as “my nearest future depends only on where I am now”. Vaguely transferring this to the graphon realm, “the nearest future” corresponds to the time derivative  $\frac{d}{dt}\Phi^t(W_0)$ , and “depends only on where I am now” corresponds to the equation

$$\frac{d}{dt}\Phi^t(W_0) = \mathfrak{V}(\Phi^t(W_0)),$$

where  $\mathfrak{V} : \mathcal{W} \rightarrow \mathcal{W}$  is some operator mapping kernels (and thus graphons) to kernels. In this section we will intuitively describe this (so far) hypothetical operator in the case of the triangle removal flip process.

To interpret the time derivative, we want to see how a typical graph  $G_{\delta n^2}$  evolved in  $\delta n^2$  steps from an  $n$ -vertex graph  $G_0$  which is close to  $U$ . Consider any points  $x, y \in \Omega$ . The key idea behind the cut norm is that the points  $x$  and  $y$  correspond to certain sets  $X \subset V(G_0)$  and  $Y \subset V(G_0)$ , where  $|X| = |Y| = \gamma n$  (and  $\gamma$  should be thought of as infinitesimally small<sup>5</sup>), and that

$$(23) \quad U(x, y) \approx \frac{e_{G_0}(X, Y)}{\gamma^2 n^2}.$$

Since in the triangle removal process edges are only removed,  $U - \Phi^\delta U$  at  $(x, y)$  corresponds to the number of edges removed from  $G_0[X, Y]$  during the first  $\delta n^2$  steps. How can we remove edges in one step, say the first one? Firstly, we

<sup>4</sup>In [1] the claim is actually weaker, but by examining the proof one easily shows (22).

<sup>5</sup>Instead of points  $x, y$  one can also approximate  $U$  with a step graphon of step size  $\gamma$  and take sets  $I_x, I_y$  containing  $x$  and  $y$  of measure  $\gamma$ .

need one of the three sampled vertices to go in  $X$  and one in  $Y$ , which happens with probability  $\approx 6\gamma^2$  (the approximation justified by the smallness of  $\gamma$ ). Conditioning on this, we need all three vertices to form a triangle, in which case we remove *one* edge from  $G_0[X, Y]$ . This conditional event happens with a probability approximately reflected in the 2-rooted triangle density  $t^{(x,y)}(\triangle^{1,2}, U)$  (cf. (15)). Hence,

$$\mathbb{E}[e_{G_0}(X, Y) - e_{G_1}(X, Y)] \approx 6\gamma^2 t^{(x,y)}(\triangle^{1,2}, U).$$

Since the overall changes of the host graph are fairly minor for the first  $\delta n^2$  steps, the calculations are almost unchanged for  $e_{G_1}(X, Y) - e_{G_2}(X, Y), \dots, e_{G_{\delta n^2}}(X, Y) - e_{G_{\delta n^2-1}}(X, Y)$ . Hence,

$$\mathbb{E}[e_{G_0}(X, Y) - e_{G_{\delta n^2}}(X, Y)] \approx 6\gamma^2 t^{(x,y)}(\triangle^{1,2}, U) \cdot \delta n^2.$$

Recalling the scaling (23), we therefore have  $(U - \Phi^\delta U)(x, y) \approx 6\gamma^2 t^{(x,y)}(\triangle^{1,2}, U) \cdot \delta$ , and hence we expect that the time derivative  $\lim_{\delta \rightarrow 0} [\Phi^\delta U(x, y) - U(x, y)]/\delta$  equals

$$(24) \quad \mathfrak{V} U(x, y) := -6t^{(x,y)}(\triangle^{1,2}, U).$$

To generalise to more complicated flip processes, we just need to consider edge erasures and edge additions coming from different subgraph replacements of the host graph.

#### 4.2. The velocity operator and definition of trajectories

In this section, we first define the velocity operator  $\mathfrak{V}_{\mathcal{R}} : \mathcal{W} \rightarrow \mathcal{W}$  for a flip process with a rule  $\mathcal{R}$  of order  $k$ . As we saw in the motivating Section 4.1, 2-rooted densities of graphs in  $\mathcal{H}_k$  will be key parameters in that definition. Having defined the velocity, we then in Theorem 4.5 construct trajectories  $\Phi : \mathcal{W}_0 \times [0, +\infty) \rightarrow \mathcal{W}_0$  using Theorem 3.11. In fact it will be more convenient to define trajectories starting at arbitrary kernels, justifying the domain of the operator  $\mathfrak{V}_{\mathcal{R}}$ .

**Definition 4.1.** Recall the operator  $T_F^{a,b} : \mathcal{W} \rightarrow L^\infty(\Omega^2)$ , defined in (11), which, when applied to a graphon, gives the rooted induced density. Given a rule  $\mathcal{R}$  of order  $k$ , we define an operator  $\mathfrak{V}_{\mathcal{R}} : \mathcal{W} \rightarrow \mathcal{W}$  by

$$(25) \quad \mathfrak{V}_{\mathcal{R}} W := \sum_{F, H \in \mathcal{H}_k} \mathcal{R}_{F,H} \sum_{1 \leq a \neq b \leq k} T_F^{a,b} W \cdot (\mathbb{I}_{\{ab \in H \setminus F\}} - \mathbb{I}_{\{ab \in F \setminus H\}}),$$

where  $ab$  is a shortening of  $\{a, b\}$ . We call  $\mathfrak{V}_{\mathcal{R}} W$  the *velocity at  $W$  for rule  $\mathcal{R}$* . We will often omit the rule that is clear from the context, denoting  $\mathfrak{V} = \mathfrak{V}_{\mathcal{R}}$ .

Note that  $T_F^{a,b} W$  is not necessarily symmetric in  $(x, y)$ , but  $T_F^{a,b} W$  and  $T_F^{b,a} W$  appear with the same coefficient in  $\mathfrak{V}_{\mathcal{R}} W$ , making it a symmetric function.

With this definition, we may formally calculate the velocity at  $W$  for the rule  $\mathcal{R}^{\text{TR}}$  which defines the triangle removal flip process (see Figure 1, right). We have  $k = 3$ ,  $\mathcal{R}_{\triangle^{1,2}}^{\text{TR}} = 1$  and  $\mathcal{R}_{F,F}^{\text{TR}} = 1$  for all  $F \in \mathcal{H}_3 \setminus \{\triangle^{1,2}\}$ . This gives

$$\mathfrak{V}_{\mathcal{R}^{\text{TR}}} W = - \sum_{1 \leq a \neq b \leq 3} T_{\triangle^{1,2}}^{a,b} W = -6T_{\triangle^{1,2}}^{1,2} W,$$

which matches our heuristic in (24).

Calculation of the velocity will often be simplified by the following reformulation using notation from Section 3.2.5.

**Lemma 4.2.** *Suppose that a graphon  $W$  and a rule  $\mathcal{R}$  of order  $k$  are given. Let  $(\mathbf{F}, \mathbf{H})$  be a random pair of graphs in  $\mathcal{H}_k$  such that  $\mathbf{F} \sim \mathbb{G}(k, W)$  and  $\mathbb{P}(\mathbf{H} = H \mid \mathbf{F} = F) = \mathcal{R}_{F,H}$  for every  $F, H \in \mathcal{H}_k$ . We have*

$$(26) \quad (\mathfrak{V}_{\mathcal{R}} W)(x, y) = \sum_{a \neq b} [\mathbb{P}(ab \in \mathbf{H} \mid \mathbf{U}_a = x, \mathbf{U}_b = y) - W(x, y)].$$

**Proof.** By (15),

$$\begin{aligned} \mathfrak{A}_{\mathcal{R}} W(x, y) &= \sum_{a \neq b} [\mathbb{P}(ab \in \mathbf{H} \setminus \mathbf{F} \mid \mathbf{U}_a = x, \mathbf{U}_b = y) - \mathbb{P}(ab \in \mathbf{F} \setminus \mathbf{H} \mid \mathbf{U}_a = x, \mathbf{U}_b = y)] \\ &= \sum_{a \neq b} [\mathbb{P}(ab \in \mathbf{H} \mid \mathbf{U}_a = x, \mathbf{U}_b = y) - \mathbb{P}(ab \in \mathbf{F} \mid \mathbf{U}_a = x, \mathbf{U}_b = y)] \\ &= \sum_{a \neq b} [\mathbb{P}(ab \in \mathbf{H} \mid \mathbf{U}_a = x, \mathbf{U}_b = y) - W(x, y)], \end{aligned}$$

where the second equality follows by adding and subtracting, in each term, the number  $\mathbb{P}(ab \in \mathbf{F} \cap \mathbf{H} \mid \mathbf{U}_a = x, \mathbf{U}_b = y)$ .  $\square$

Let us explain how Definition 4.1 relates to a flip process with rule  $\mathcal{R}$  of order  $k$ . Suppose that  $(\mathbf{G}_i)_{i \geq 0}$  is the flip process on  $n$  vertices. Fix distinct  $u, v \in [n]$ . The probability that the  $a$ th sampled vertex is  $u$  and the  $b$ th is  $v$  is  $1/(n)_2$ , whence

$$\begin{aligned} &\mathbb{E}(\mathbb{I}_{\{uv \in \mathbf{G}_{i+1}\}} - \mathbb{I}_{\{uv \in \mathbf{G}_i\}} \mid \mathbf{G}_i) \\ &= \frac{1}{(n)_2} \sum_{1 \leq a \neq b \leq k} \sum_{F, H \in \mathcal{H}_k} \mathcal{R}_{F, H} \cdot t_{\text{ind}}^{(u, v)}(F^{a, b}, \mathbf{G}_i) (\mathbb{I}_{\{ab \in H \setminus F\}} - \mathbb{I}_{\{ab \in F \setminus H\}}). \end{aligned}$$

Furthermore let  $W_i$  be the graphon representation of  $\mathbf{G}_i$  with respect to the same partition. Since, by (14),

$$|t_{\text{ind}}^{(x, y)}(F^{a, b}, \mathbf{G}_i) - t_{\text{ind}}^{(x, y)}(F^{a, b}, W_i)| \leq \binom{k-2}{2} \frac{1}{n},$$

we get the following formula linking the flip process with the velocity operator.

**Remark 4.3.** Consider the flip process  $(\mathbf{G}_i)_{i \geq 0}$  with rule  $\mathcal{R}$  of order  $k$ . If  $W_i$  is the graphon representation of  $\mathbf{G}_i$ , then for a.e.  $x, y \in \Omega$ ,

$$\mathbb{E}[W_{i+1}(x, y)] = W_i(x, y) + \frac{1}{(n)_2} \mathfrak{A}_{\mathcal{R}} W_i(x, y) (1 + O(1/n)).$$

We can now define the key object of our paper.

**Definition 4.4.** Given a rule  $\mathcal{R}$  and a kernel  $W \in \mathcal{W}$ , a *trajectory starting at  $W$*  is a differentiable function  $\Phi \cdot W : I \rightarrow (\mathcal{W}, \|\cdot\|_{\infty})$  defined on an open interval  $I \subseteq \mathbb{R}$  containing zero, that satisfies the autonomous differential equation

$$(27) \quad \frac{d}{dt} \Phi \cdot W = \mathfrak{A} \Phi \cdot W$$

with the initial condition

$$(28) \quad \Phi^0 W = W.$$

The ultimate goal of this section is the following theorem (proved in Section 4.4), which asserts that trajectories exist, are unique, and whenever they start at a graphon (rather than at a general kernel), they stay among graphons for every positive time.

**Theorem 4.5.** *Suppose that  $\mathcal{R}$  is a rule.*

- (i) *For each kernel  $W \in \mathcal{W}$ , there is an open interval  $\mathcal{D}_W \subseteq \mathbb{R}$  containing 0 and a trajectory  $\Phi \cdot W : \mathcal{D}_W \rightarrow \mathcal{W}$  starting at  $W$  such that any other trajectory starting at  $W$  is a restriction of  $\Phi \cdot W$  to a subinterval of  $\mathcal{D}_W$ .*
- (ii) *For any  $u \in \mathcal{D}_W$  we have  $\mathcal{D}_{\Phi^u W} = \{t \in \mathbb{R} : t + u \in \mathcal{D}_W\}$  and for every  $t \in \mathcal{D}_{\Phi^u W}$  we have*

$$(29) \quad \Phi^t \Phi^u W = \Phi^{t+u} W.$$

- (iii) *Whenever  $W \in \mathcal{W}_0$ , the set  $I_W := \{t \in \mathcal{D}_W : \Phi^t W \in \mathcal{W}_0\}$  is a closed interval containing  $[0, \infty)$ .*



**Definition 4.6.** For  $W \in \mathcal{W}_0$  we write  $\text{age}(W) := -\inf I_W \in [0, \infty]$ .

We will sometimes (as in the following remark) abuse the notation by saying *trajectory*, when we actually mean the orbit (image) of a trajectory, that is the set  $\{\Phi^t W : t \in I\}$ .

**Remark 4.7.** Note that part (ii) also implies that trajectories from two different kernels  $W, U$  are disjoint, unless  $W = \Phi^t U$  for some  $t$ . In other words, trajectories partition  $\mathcal{W}$ .

### 4.3. Bounds on the velocity

The first claim of the next lemma, (31), gives bounds on the velocity in terms of the corresponding graphon. Roughly speaking, it says that the  $\mathfrak{V}_{\mathcal{R}} W$  cannot be too positive for those  $(x, y)$  for which  $W(x, y)$  is large, and cannot be too negative for those  $(x, y)$  for which  $W(x, y)$  is small. Combinatorially, this is quite intuitive: if, for example,  $W(x, y)$  is close to 0, then this corresponds to a graph with a sparse bipartite spot  $(X, Y)$ , where the vertex sets  $X$  and  $Y$  correspond to the points  $x$  and  $y$ . The lower bound in (31) corresponds to the fact that deletions of  $(X, Y)$ -edges are indeed rare, since they may only happen when we actually sample an  $(X, Y)$ -edge.

The following lemma also claims two more facts about the velocity operator on the Banach space  $(\mathcal{W}, \|\cdot\|_{\infty})$  that are crucial in application of Theorem 3.11: (32) says it is locally Lipschitz, while (33) gives a bound on the norm on a ball around the constant-1/2 graphon.

**Lemma 4.8.** *Suppose that  $\mathcal{R}$  is a rule of order  $k$ . Let  $\varepsilon \geq 0$  be a constant. Define*

$$(30) \quad C_k := (k)_2 2^{\binom{k}{2}-1}, \quad C_{\infty}(k, \varepsilon) := C_k(1 + \varepsilon)^{\binom{k}{2}-1}.$$

For every graphon  $W \in \mathcal{W}_0$  we have

$$(31) \quad -(k)_2 \cdot W \leq \mathfrak{V}_{\mathcal{R}} W \leq (k)_2 \cdot (1 - W).$$

For every  $U, W \in \mathcal{W}_0^{\varepsilon}$  we have

$$(32) \quad \|\mathfrak{V}_{\mathcal{R}} U - \mathfrak{V}_{\mathcal{R}} W\|_{\infty} \leq C_{\infty}(k, \varepsilon) \|U - W\|_{\infty}.$$

For every  $W \in \mathcal{W}_0^{\varepsilon}$  we have

$$(33) \quad \|\mathfrak{V}_{\mathcal{R}} W\|_{\infty} \leq (k)_2 + C_{\infty}(k, \varepsilon)\varepsilon.$$

**Proof.** Using the probabilistic form of the velocity (26), and noting that the conditional probability takes values in  $[0, 1]$ , inequalities (31) follow.

To obtain (32), using Definition 4.1 we have

$$\begin{aligned} \|\mathfrak{V}_{\mathcal{R}} U - \mathfrak{V}_{\mathcal{R}} W\|_{\infty} &= \left\| \sum_{a \neq b} \sum_{F \in \mathcal{H}_k} (T_F^{a,b} U - T_F^{a,b} W) \sum_{H \in \mathcal{H}_k} \mathcal{R}_{F,H} (\mathbb{I}_{\{ab \in H \setminus F\}} - \mathbb{I}_{\{ab \in F \setminus H\}}) \right\|_{\infty} \\ &\leq \sum_{a \neq b} \sum_{F \in \mathcal{H}_k} \|T_F^{a,b} U - T_F^{a,b} W\|_{\infty} \sum_{H \in \mathcal{H}_k} \mathcal{R}_{F,H} \\ &\stackrel{\text{by (4)}}{=} \sum_{a \neq b} \sum_{F \in \mathcal{H}_k} \|T_F^{a,b} U - T_F^{a,b} W\|_{\infty} \\ &\stackrel{\text{Lemma 3.5}}{\leq} (k)_2 |\mathcal{H}_k| \cdot (1 + \varepsilon)^{\binom{k}{2}-1} \binom{k}{2} \|U - W\|_{\infty} = C_{\infty}(k, \varepsilon) \|U - W\|_{\infty}. \end{aligned}$$

Finally, to see (33), given  $W$ , pick the nearest  $U \in \mathcal{W}_0$  (so that  $\|W - U\|_{\infty} \leq \varepsilon$ ). Using the triangle inequality, (31) and (32), we infer

$$\|\mathfrak{V} W\|_{\infty} \leq \|\mathfrak{V} U\|_{\infty} + C_{\infty}(k, \varepsilon) \|W - U\|_{\infty} \leq (k)_2 + C_{\infty}(k, \varepsilon)\varepsilon. \quad \square$$

The continuity property (32) allows us to rewrite the differential equation (27) as an integral equation. This is formulated in the remark below.

**Remark 4.9.** Since  $\mathfrak{V}$  is continuous on  $(\mathcal{W}, \|\cdot\|_\infty)$  by (32), Theorem C.2 tells us that the trajectory  $\Phi \cdot W : I \rightarrow \mathcal{W}$  satisfies

$$\Phi^t W = W + \int_0^t \mathfrak{V}_{\mathcal{R}} \Phi^\tau W \, d\tau, \quad t \in I,$$

where the integral is defined with respect to the norm  $\|\cdot\|_\infty$ .

We will also need that the velocity operator is locally Lipschitz with respect to the cut norm.

**Lemma 4.10.** *For any rule of order  $k$  and every  $\varepsilon \geq 0$  and  $U, W \in \mathcal{W}_0^\varepsilon$  we have*

$$\|\mathfrak{V}_{\mathcal{R}} U - \mathfrak{V}_{\mathcal{R}} W\|_\square \leq C_\square \|U - W\|_\square,$$

where  $C_\square = C_\square(k, \varepsilon) = (k)_2^2(2 + 4\varepsilon) \binom{k}{2}$ .

**Proof.** First, we want to show that for every  $a, b \in [k]$ ,  $a \neq b$ , and  $G \in \mathcal{H}_k$  the following holds.

**Claim 4.11.** *If  $U, W \in \mathcal{W}_0^\varepsilon$ , then*

$$\|T_G^{a,b} U - T_G^{a,b} W\|_\square \leq (k)_2(1 + 2\varepsilon) \binom{k}{2} \|U - W\|_\square.$$

**Proof of Claim 4.11.** Fix an arbitrary set  $A \subseteq \Omega$  and let  $\mathbb{I}_{A \times A} : \Omega^2 \rightarrow \{0, 1\}$  be the indicator function of the set  $A \times A$ . Given  $W, G, a$  and  $b$ , we define a tuple  $\mathbf{W} = (W_e : e \in E(K_k))$  by

$$W_e = \begin{cases} W, & \text{if } e \neq ab, e \in E(G), \\ 1 - W, & \text{if } e \neq ab, e \notin E(G), \\ W \cdot \mathbb{I}_{A \times A}, & \text{if } e = ab \in E(G), \\ (1 - W) \cdot \mathbb{I}_{A \times A}, & \text{if } e = ab \notin E(G), \end{cases}$$

and similarly  $\mathbf{U} = (U_e : e \in E(K_k))$ , with  $W$  replaced by  $U$ . We claim that

$$(34) \quad \|U_e - W_e\|_\square \leq \|U - W\|_\square$$

for every  $e \in K_k$ . When  $e \neq ab$ , regardless whether  $e \in E(G)$  or not, we have  $U_e - W_e = U - W$  and therefore  $\|U_e - W_e\|_\square = \|U - W\|_\square$ .

When  $e = ab$ , we have  $U_e - W_e = (U - W)\mathbb{I}_{A \times A}$  and therefore

$$\begin{aligned} \|U_e - W_e\|_\square &= \|(U - W)\mathbb{I}_{A \times A}\|_\square = \sup_{S, T \subseteq \Omega} \left| \int_{S \times T} (U - W) \cdot \mathbb{I}_{A \times A} \, d\pi^2 \right| \\ &= \sup_{S, T \subseteq \Omega} \left| \int_{(S \cap A) \times (T \cap A)} (U - W) \, d\pi^2 \right| \leq \|U - W\|_\square. \end{aligned}$$

Note that

$$\begin{aligned} \int_{A \times A} T_G^{a,b} U \, d\pi^2 &= \int \mathbb{I}_{A \times A}(x_a, x_b) \int \prod_{ij \in E(G)} U(x_i, x_j) \prod_{ij \notin E(G)} (1 - U(x_i, x_j)) \, d\tau^{[k] \setminus \{a,b\}} \, d\tau^{\{a,b\}} \\ &= t(K_k, \mathbf{U}). \end{aligned}$$

And similarly  $\int_{A \times A} T_G^{a,b} W \, d\pi^2 = t(K_k, \mathbf{W})$ . Hence

$$\begin{aligned} \left| \int_{A \times A} (T_G^{a,b} U - T_G^{a,b} W) \, d\pi^2 \right| &= |t(K_k, \mathbf{U}) - t(K_k, \mathbf{W})| \\ &\stackrel{\text{Lemma 3.7}}{\leq} (1 + 2\varepsilon) \binom{k}{2} \sum_{e \in K_k} \|U_e - W_e\|_{\square} \\ &\stackrel{\text{by (34)}}{\leq} (1 + 2\varepsilon) \binom{k}{2} \binom{k}{2} \|U - W\|_{\square}. \end{aligned}$$

Since this was verified for each set  $A$ , the result follows from Fact B.1. □

Now, using the triangle inequality and Claim 4.11, we get

$$\begin{aligned} \|\mathfrak{V}_{\mathcal{R}} U - \mathfrak{V}_{\mathcal{R}} W\|_{\square} &\leq \sum_{1 \leq a \neq b \leq k} \sum_{(G,F) \in \mathcal{H}_k \times \mathcal{H}_k} \mathcal{R}_{G,F} \cdot \|T_G^{a,b} U - T_G^{a,b} W\|_{\square} \\ &\stackrel{\text{by (4)}}{\leq} \sum_{1 \leq a \neq b \leq k} \sum_{G \in \mathcal{H}_k} \|T_G^{a,b} U - T_G^{a,b} W\|_{\square} \\ &\leq (k)_2 2 \binom{k}{2} \cdot (k)_2 (1 + 2\varepsilon) \binom{k}{2} \|U - W\|_{\square} < C_{\square} \|U - W\|_{\square}, \end{aligned}$$

which proves the lemma. □

#### 4.4. Proof of Theorem 4.5

Fix a rule  $\mathcal{R}$  of order  $k$  and further write  $\mathfrak{V} = \mathfrak{V}_{\mathcal{R}}$ .

We start with the existence of a time-local solution for every  $W \in \mathcal{W}$ .

**Claim 4.12.** *For every  $\rho \geq 0$  there is  $T_{\rho} > 0$ , so that for every  $W \in \mathcal{W}_0^{\rho}$  there is a unique trajectory  $\Phi \cdot W : J_{\rho} = (-T_{\rho}, T_{\rho}) \rightarrow \mathcal{W}_0^{1+\rho}$ . Moreover, for every  $t \in J_{\rho}$  the function  $W \mapsto \Phi^t W$  is continuous on  $(\mathcal{W}_0^{\rho}, \|\cdot\|_{\infty})$ .*

**Proof.** We apply Theorem 3.11 for the Banach space  $\mathcal{E} = (\mathcal{W}, \|\cdot\|_{\infty})$  with function  $X = \mathfrak{V}$ . Using the notation of Theorem 3.11, we set  $\mathcal{U}$  to be the interior of  $\mathcal{W}_0^{1+\rho}$  (equivalently, an open ball around constant- $\frac{1}{2}$  of radius  $\frac{3}{2} + \rho$ ), that is,

$$\mathcal{U} = \bigcup_{\beta > 0} \mathcal{W}_0^{1+\rho-\beta}.$$

Let function  $C_{\infty}$  be as in (30). By (32), operator  $\mathfrak{V}$  is  $K$ -Lipschitz on  $\mathcal{U}$  with  $K = C_{\infty}(k, 1 + \rho)$ . Setting  $b := 0.9$ , say, note that for every  $x_0 = W \in \mathcal{W}_0^{\rho}$ , we have that the closed ball  $\mathcal{B}(x_0, b)$  lies in  $\mathcal{U}$ . By (33), condition (21) is satisfied by  $M := (k)_2 + (1 + \rho)C_{\infty}(k, 1 + \rho)$ . Theorem 3.11 implies the claim with  $T_{\rho} = \alpha = \min\{b/M, 1/K\}$ , with the continuity of  $W \mapsto \Phi^t W$  following from (22). □

For a given rule, consider all trajectories starting at  $W$  that are defined on an open interval containing 0 and let  $\mathcal{D}_W$  be the union of their domains. Using uniqueness given by Claim 4.12 it is easy to see (cf. [24, Section 2.6]) that any two such trajectories agree on the intersection of their domains and therefore determine a unique trajectory  $\Phi \cdot W : \mathcal{D}_W \rightarrow \mathcal{W}$  with a maximal domain. This proves (i).

Uniqueness also implies (ii), by noting that  $f_1(t) = \Phi^{t+u} W$  and  $f_2(t) = \Phi^t \Phi^u W$  are two solutions for the same initial condition  $f(0) = \Phi^u W$ .

It remains to prove (iii). Despite being “heuristically obvious”, its analytic proof is somewhat tedious. The first step is to show this for graphons that lie in the interior of  $\mathcal{W}_0$  with respect to  $\|\cdot\|_{\infty}$ .

**Claim 4.13.** *For every  $\varepsilon > 0$  and every  $W \in \mathcal{W}_0$  such that  $\varepsilon \leq W \leq 1 - \varepsilon$  and every  $t \in \mathcal{D}_W \cap [0, \infty)$  we have  $\Phi^t W \in \mathcal{W}_0$ .*

**Proof.** Let  $T := \sup\{t \in \mathcal{D}_W : \Phi^u W \in \mathcal{W}_0 \forall u \in [0, t]\}$  and suppose for contradiction that  $T < \infty$ .

The continuity<sup>6</sup> of the function  $t \mapsto \Phi^t W$  implies the following three facts:  $T > 0$  (since  $W$  belongs to the interior of  $\mathcal{W}_0$  with respect to  $\|\cdot\|_\infty$ );  $\Phi^T W \in \mathcal{W}_0$  (since  $\mathcal{W}_0$  is a closed set); by maximality of  $T$ ,  $\Phi^T W$  lies on the boundary of  $\mathcal{W}_0$ , that is,

$$(35) \quad \text{ess inf } \Phi^T W = 0 \quad \text{or} \quad \text{ess sup } \Phi^T W = 1.$$

We will obtain a contradiction by showing that  $\Phi^T W$  lies in the interior of  $\mathcal{W}_0$ .

Define continuous kernel-valued functions  $f, g$ , with domain  $[0, T]$  by  $f_t := \mathfrak{V} \Phi^t W$  and  $g_t = -(k)_2 \Phi^t W$ . By (31) we have that  $f_t \geq g_t$  for  $t \in [0, T]$  and by Remark 4.9  $\Phi^t W = W + \int_0^t f_\tau d\tau$ . In view of Proposition C.4, we can assume that for every  $(x, y) \in \Omega^2$ ,

- the function  $\tau \mapsto \mathfrak{V} \Phi^\tau W(x, y)$  is continuous on  $[0, T]$ ;
- $\mathfrak{V} \Phi^\tau W(x, y) \geq -(k)_2 \Phi^\tau W(x, y)$ ,  $\tau \in [0, T]$ ;
- and

$$\Phi^t W(x, y) = W(x, y) + \int_0^t \mathfrak{V} \Phi^\tau W(x, y) d\tau, \quad t \in [0, T].$$

Let  $h(t) = \mathfrak{V} \Phi^t W(x, y)$  so that  $H(t) := \Phi^t W(x, y)$  satisfies  $H(t) = H(0) + \int_0^t h(x) dx$  and  $h(t) \geq -(k)_2 H(t)$ . By Lemma 3.10,  $H(T) \geq H(0)e^{-(k)_2 T}$ . Since  $W \geq \varepsilon$ , we conclude that  $\Phi^T W \geq \varepsilon e^{-(k)_2 T} =: \varepsilon' > 0$ . By an analogous argument applied to graphon  $1 - W$ , we get that  $1 - W \geq \varepsilon'$ . This is a contradiction to (35).  $\square$

We now extend Claim 4.13 to all graphons.

**Claim 4.14.** *For every  $W \in \mathcal{W}_0$  and every  $t \in \mathcal{D}_W \cap [0, \infty)$  we have  $\Phi^t W \in \mathcal{W}_0$ .*

**Proof.** Let  $T_0$  be given by Claim 4.12 for  $\rho = 0$ . Observe that it is enough to prove  $\Phi^t W \in \mathcal{W}_0$  only for times  $t \in \mathcal{D}_W \cap [0, T_0)$ . Indeed, with such a weaker statement we can write any time  $t$  as  $t = t_1 + \dots + t_\ell$  where each  $t_s \in (0, T_0)$  and then inductively for  $s = 1, \dots, \ell$  deduce that  $\Phi^{\sum_{i=1}^s t_i} W \in \mathcal{W}_0$ . The advantage for working on this smaller time interval is that Claim 4.12 tells us that for each  $\tau \in (-T_0, T_0)$ , the map  $U \mapsto \Phi^\tau U$  is  $\|\cdot\|_\infty$ -continuous on  $\mathcal{W}_0^0 = \mathcal{W}_0$ .

Let  $W \in \mathcal{W}_0$  and  $t \in \mathcal{D}_W \cap [0, T_0)$  be arbitrary. Then for every  $\varepsilon > 0$  let  $U_\varepsilon := \varepsilon + (1 - 2\varepsilon)W$ . Noting that  $\varepsilon \leq U_\varepsilon \leq 1 - \varepsilon$  we conclude by Claim 4.13 that  $\Phi^t U_\varepsilon \in \mathcal{W}_0$ . Since  $\|W - U_\varepsilon\|_\infty \rightarrow 0$  as  $\varepsilon \rightarrow 0$ , continuity of the map  $U \mapsto \Phi^t U$  implies that  $\Phi^t W \in \mathcal{W}_0$ , as was needed.  $\square$

Claim 4.14 and (ii) imply that  $I_W \subseteq \mathbb{R}$  is an (open or closed) interval containing  $[0, \infty)$ . Since  $\Phi \cdot W : \mathcal{D}_W \rightarrow (\mathcal{W}, \|\cdot\|_\infty)$  is continuous and  $\mathcal{W}_0$  is closed, interval  $I_W$  is closed with respect to  $\mathcal{D}_W$ . Hence, recalling the notation  $\text{age}(W) = -\inf I_W$ , we have that either  $I_W = \mathcal{D}_W$  or  $I_W = [-\text{age}(W), \infty)$ . It remains to check that in the former case  $\text{age}(W) = \infty$ , so that  $I_W = \mathcal{D}_W = \mathbb{R}$ . Suppose for contradiction that we have  $\text{age}(W) < \infty$ , so that  $\mathcal{D}_W = (-\text{age}(W), \infty)$ . Let  $T_0$  be given by Claim 4.12 for  $\rho = 0$ . Let  $t = -\text{age}(W) + T_0/4 \in I_W$ . Since  $\Phi^t W \in \mathcal{W}_0$ , we can use Claim 4.12 to extend the trajectory from the graphon  $\Phi^t W$  to the interval  $(-T_0, T_0)$ . In particular, we have extended the trajectory of the graphon  $W$  by at least  $3T_0/4$  before the  $-\text{age}(W)$ , contradicting the maximality of  $\mathcal{D}_W$ .

#### 4.5. Smoothness of trajectories in time

The following lemma says that for any graphon the trajectory is Lipschitz in time. Inequality (37) will be used in the proof of concentration of flip processes (Theorem 5.1).

**Lemma 4.15.** *For every rule  $\mathcal{R}$ , every graphon  $W \in \mathcal{W}_0$  and every  $t, u \in I_W$ ,*

$$(36) \quad \|\Phi^t W - \Phi^u W\|_\infty \leq (k)_2 |t - u|,$$

and for every  $\delta > 0$ ,

$$(37) \quad \left\| \frac{\Phi^\delta W - W}{\delta} - \mathfrak{V} W \right\|_\infty \leq C_k (k)_2 \delta,$$

where  $C_k$  is defined in (30).

<sup>6</sup>The continuity with respect to  $\|\cdot\|_\infty$  is implied by the fact that Theorem 3.11 outputs a function which is even differentiable with respect to  $\|\cdot\|_\infty$ .

**Proof.** By Remark 4.9 we have, for every  $t, u \geq -\text{age}(W)$ ,

$$(38) \quad \Phi^t W - \Phi^u W = \int_u^t \mathfrak{V} \Phi^\tau W \, d\tau.$$

Combining (86) and the inequality  $\|\mathfrak{V} U\|_\infty \leq (k)_2$  for  $U \in \mathcal{W}_0$  (see (33)), we obtain

$$\|\Phi^t W - \Phi^u W\|_\infty \leq |t - u| \sup_{\tau \in [u, t]} \|\mathfrak{V} \Phi^\tau W\|_\infty \leq (k)_2 |t - u|,$$

thus proving (36).

To prove (37), setting  $t = \delta, u = 0$  in (38) becomes

$$(39) \quad \Phi^\delta W - W = \int_0^\delta \mathfrak{V} \Phi^\tau W \, d\tau.$$

By (36) we have, for every  $\tau \geq 0$ ,

$$\|\Phi^\tau W - W\|_\infty = \|\Phi^\tau W - \Phi^0 W\|_\infty \leq (k)_2 \tau,$$

and since  $\Phi^\tau W \in \mathcal{W}_0$ , by (32),

$$\|\mathfrak{V} \Phi^\tau W - \mathfrak{V} W\|_\infty \leq C_\infty(k, 0)(k)_2 \tau = C_k(k)_2 \tau.$$

Hence, using (39), and elementary properties of the integral (see Section C),

$$\begin{aligned} \|\Phi^\delta W - W - \delta \mathfrak{V} W\|_\infty &= \left\| \int_0^\delta (\mathfrak{V} \Phi^\tau W - \mathfrak{V} W) \, d\tau \right\|_\infty \\ &= (\delta - 0) \sup_{\tau \in [0, \delta]} \|\mathfrak{V} \Phi^\tau W - \mathfrak{V} W\|_\infty \leq C_k(k)_2 \delta^2. \end{aligned}$$

Dividing by  $\delta$  we obtain (37). □

#### 4.6. Dependence on the initial conditions

The following theorem implies that on any bounded time interval the trajectories depend continuously on the initial conditions. We have already used such a fact in the proof of Claim 4.14, where it relied on the well-known machinery of Banach-space-valued differential equations and on (22) in particular. However, there the metric was induced by the  $L^\infty$ -norm. For many combinatorial problems the cut norm is more relevant, and to deduce Lipschitz continuity we need to be more careful. Even though the velocity is also Lipschitz with respect to the cut norm (see Lemma 4.10), we are not in the setting of Banach spaces.<sup>7</sup>

**Theorem 4.16.** *Given  $k \in \mathbb{N}$ , let  $C_\square = C_\square(k, 0)$  as in Lemma 4.10. Let  $\mathcal{R}$  be a rule of order  $k$ , and  $U, W$  be any graphons. If  $t \in I_U \cap I_W$ , then*

$$\|\Phi^t U - \Phi^t W\|_\square \leq e^{C_\square |t|} \|U - W\|_\square.$$

**Proof.** For  $t \in I_U \cap I_W = [-\min(\text{age}(U), \text{age}(W)), +\infty)$ , let  $W_t := \Phi^t W$  and  $U_t := \Phi^t U$ . Let us first consider the case  $t \geq 0$  and set  $F(t) := \|U_t - W_t\|_\square$ . By Remark 4.9, for any  $s$  and  $t$  such that  $-\min(\text{age}(U), \text{age}(W)) \leq s \leq t < \infty$  we have

$$(40) \quad W_t - W_s = \int_s^t \mathfrak{V} W_x \, dx,$$

where the integral is understood in terms of the Banach space  $(\mathcal{W}, \|\cdot\|_\infty)$ . Since  $x \mapsto \mathfrak{V} W_x$  is also continuous with respect to the cut norm, integral can be treated with respect of the completion  $\mathcal{E}$  of the normed space  $(\mathcal{W}, \|\cdot\|_\square)$ . Note

<sup>7</sup>That is, we do not know if  $(\mathcal{W}, \|\cdot\|_\square)$  is complete. Note that the subset of uniformly bounded functions is complete by Lemma 3.2.

that  $\mathcal{W}_0$  is a closed subset of  $\mathcal{E}$  by Lemma 3.2. Hence, using the triangle inequality, property of the integral (86), and Lemma 4.10, we obtain, for  $t \geq s \geq 0$ ,

$$\begin{aligned}
 \boxed{\text{triangle inequality}} \quad & F(t) - F(s) \leq \|U_t - U_s - (W_t - W_s)\|_{\square} \\
 \boxed{\text{by (40)}} \quad & \leq \left\| \int_s^t (\mathfrak{A} U_x - \mathfrak{A} W_x) dx \right\|_{\square} \\
 \boxed{\text{by (86)}} \quad & \leq \int_s^t \|\mathfrak{A} U_x - \mathfrak{A} W_x\|_{\square} dx \\
 \boxed{\text{Lemma 4.10}} \quad & \leq C_{\square} \int_s^t \|U_x - W_x\|_{\square} = C_{\square} \int_s^t F(x) dx.
 \end{aligned}$$

In particular this implies that  $F$  is  $C_{\square}$ -Lipschitz and so continuous. Moreover, it implies that  $F$  satisfies (17) with  $t_0 = 0$ ,  $A = F(0)$  and  $C = C_{\square}$ . Hence Lemma 3.9 implies Theorem for  $t \geq 0$ .

Now considering  $t < 0$ , we define

$$F(x) = \|U_{-x} - W_{-x}\|_{\square}, \quad x \in [0, \min\{\text{age}(U), \text{age}(W)\}]$$

and similarly as above obtain that, for  $0 \leq s \leq t \leq \min\{\text{age}(U), \text{age}(W)\}$

$$F(t) - F(s) \leq C_{\square} \int_s^t \|U_{-x} - W_{-x}\|_{\square} dx = C_{\square} \int_s^t F(x) dx,$$

whence by Lemma 3.9,

$$F(t) \leq F(0)e^{C_{\square}t}, \quad t \in [0, \min\{\text{age}(U), \text{age}(W)\}]. \quad \square$$

### 5. The transference theorem

In this section we prove that for a quadratic number of steps the flip process starting with some large graph  $G$ , viewed as a graphon-valued process, is likely to stay close (with respect to the cut norm) to the deterministic trajectory  $t \mapsto \Phi^t W_G$ .

Let  $W_G$  be the graphon representation of a graph  $G$  with respect to some partition  $(\Omega_v : v \in [n])$  of  $\Omega$  (see Section 3.2.1).

**Theorem 5.1.** *For every  $k \in \mathbb{N}$  there is a constant  $C > 0$  so that the following holds. Given a rule  $\mathcal{R}$  of order  $k$  and a graph  $G$  on the vertex set  $[n]$ , let  $(G_i)_{i \geq 0}$  be the flip process starting with  $G_0 = G$ .*

*For any  $T > 0$  and  $\varepsilon > 0$  have*

$$(41) \quad \max_{i \in (0, Tn^2] \cap \mathbb{Z}} \|W_{G_i} - \Phi^{i/n^2} W_G\|_{\square} < \varepsilon$$

*with probability at least*

$$(42) \quad 1 - \frac{CTe^{CT}}{\varepsilon} \exp\left((2 \ln 2)n - \frac{C\varepsilon^3 n^2}{e^{CT}}\right).$$

**Remark 5.2.** A typical application of Theorem 5.1 will also invoke Theorem 4.16. That is, instead of relating  $W_{G_i}$  to a trajectory started at  $W_G$  (at time  $i/n^2$ ), it seems more useful to relate it to a cleaned version of that trajectory started at some (simple) graphon  $W$  which is close to  $W_G$ . For example, if  $G$  is a quasirandom graph of density  $\frac{1}{2}$ , then taking  $W \equiv \frac{1}{2}$  seems to be a good choice.

**Remark 5.3.** While throughout the paper we are motivated by the idea that  $T$  is an arbitrarily large constant (corresponding to ‘bounded horizons’), the probability bound (42) is useful even for  $T = O(\log n)$ . We are not sure if this bound can be pushed further.

This exact limit of applicability does not seem that important because in practice we will also be limited by Theorem 4.16 as we explained in Remark 5.2. For example, suppose that we have a rule of order 3 in which the edgeless graph stays edgeless and other graphs are replaced by a triangle. Let  $G$  be a graph on  $n$  vertices and  $o(n^2)$  edges. This is a



sparse graph and so  $W \equiv 0$  seems like a good choice of a starting point for the trajectory. Obviously, we have  $\Phi^t W \equiv 0$  but since  $\|W_G - 0\|_{\square} \geq 1/n^2$ , we cannot conclude anything useful about  $G_i$  for  $i \gg n^2 \log n$  due to the Lipschitz constant in Theorem 4.16.

For convenience, we also include an easy corollary of Theorem 5.1 phrased in terms of the cut distance.

**Corollary 5.4.** *For every  $k \in \mathbb{N}$  the following holds with the same constant  $C$  as in Theorem 5.1. Given a rule  $\mathcal{R}$  of order  $k$  and an  $n$ -vertex graph  $G$ , let  $(G_i)_{i \geq 0}$  be the flip process starting with  $G_0 = G$ . Let  $W_0, W_1, \dots$  be arbitrary graphon representations of the graphs  $G_0, G_1, \dots$ . For any  $T > 0$  and  $\varepsilon > 0$  have  $\max\{\delta_{\square}(W_i, \Phi^{i/n^2} W_0) : i \in (0, Tn^2] \cap \mathbb{Z}\} < \varepsilon$  with probability at least (42).*

5.1. Proof of Theorem 5.1

The idea is to subdivide the time domain into intervals of size  $\delta n^2$  for some small enough  $\delta$  (depending on  $k, \varepsilon$  and  $T$ ) and approximate, in each of the intervals, the expected change of the graphon  $W_{G_i}$  linearly in terms of the velocity at the start of the interval. The following lemma quantifies the error of this approximation.

**Lemma 5.5.** *Let  $k, t, n \in \mathbb{N}$  be such that  $n \geq 2$  and  $\delta := t/n^2 \geq 1/n$ . Furthermore, let  $\mathcal{R}$  be a rule of order  $k$  and  $G$  be a graph with vertex set  $[n]$ . Let  $(G_i)_{i \geq 0}$  be the flip process with initial graph  $G_0 = G$ . With probability at least*

$$(43) \quad 1 - 2^{2n+1} \exp\left(-\frac{\delta^3 n^2}{2(k)_2^2}\right)$$

we have

$$(44) \quad \|W_{G_t} - (W_G + \delta \mathfrak{V}_{\mathcal{R}} W_G)\|_{\square} \leq C''' \delta^2,$$

with a constant

$$(45) \quad C''' := 2 \binom{k}{2} \left( 2 \binom{k}{2} |\mathcal{H}_k| + 6 + C_{\square} \right) + 1,$$

where  $C_{\square} = C_{\square}(k, 0) = (k)_2^2 2^{\binom{k}{2}}$ , as in Lemma 4.10.

**Proof.** In addition to  $C_{\square}$  and  $C'''$  we need to define two further constants,

$$C' := 2 \left( \binom{k}{2}^2 |\mathcal{H}_k| + (k)_2 \right) \quad \text{and} \quad C'' := C_{\square} \binom{k}{2}.$$

For simplicity write  $W_i = W_{G_i}$ . For  $S \subseteq [n]$  write  $\Omega_S = \cup\{\Omega_u : u \in S\}$ . Since all graphons and kernels involved in the proof are constant on each of the sets  $\Omega_u \times \Omega_v$ , in view of Fact B.2 we will bound the cut norm only on sets of the form  $\Omega_S \times \Omega_T$ .

We plan to use the Hoeffding–Azuma inequality (see Section A) to show concentration of

$$e_{G_t}(S, T) - e_{G_0}(S, T)$$

for fixed  $S, T \subseteq [n]$  and then apply the union bound over the  $2^{2n}$  choices of  $S, T$ . For  $i \in [t]$ , let  $\mathbf{v}_i$  be the  $k$ -tuple of vertices chosen at step  $i$  from graph  $G_{i-1}$  and  $\mathbf{H}_i \in \mathcal{H}_k$  be the graph by which we replace  $G_{i-1}[\mathbf{v}_i]$ . We consider a filtration  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_t$ , where  $\mathcal{F}_i$  is generated by  $(\mathbf{v}_j, \mathbf{H}_j)_{1 \leq j \leq i}$ .

For each  $i \in [t]$ , define a random variable  $X_i := e_{G_i}(S, T) - e_{G_{i-1}}(S, T)$ . We have  $e_{G_t}(S, T) - e_{G_0}(S, T) = \sum_{i=1}^t X_i$ . Note that  $X_i$  is determined by  $(\mathbf{v}_j, \mathbf{H}_j)_{j \in [i]}$ , and so is  $\mathcal{F}_i$ -measurable. Defining  $Y_i := X_i - \mathbb{E}(X_i | \mathcal{F}_{i-1})$ , we have that  $Y_1, \dots, Y_t$  is a sequence of martingale differences.

We will show that, with probability one,

$$(46) \quad \left| \sum_{i=1}^t \mathbb{E}(X_i | \mathcal{F}_{i-1}) - t \int_{\Omega_S \times \Omega_T} \mathfrak{V}_{\mathcal{R}} W_{G_0} \right| \leq (C''' - 1) \delta^2 n^2,$$

and

$$(47) \quad \mathbb{P}\left(\left|\sum_{i=1}^t Y_i\right| \geq \delta^2 n^2\right) \leq 2 \exp\left(-\frac{\delta^4 n^4}{2t(k)_2^2}\right) = 2 \exp\left(-\frac{\delta^3 n^2}{2(k)_2^2}\right).$$

The triangle inequality and inequalities (46) and (47) imply that, with probability at least  $1 - 2 \exp(-\delta^3 n^2 / (2(k)_2^2))$ ,

$$\left|e_{G_t}(S, T) - e_{G_0}(S, T) - t \int_{\Omega_S \times \Omega_T} \mathfrak{W}_{\mathcal{R}} W_{G_0}\right| \leq C''' \delta^2 n^2.$$

Taking the union bound over  $2^{2n}$  choices of  $S, T$ , we obtain that

$$\|W_{G_t} - W_{G_0} - \delta \mathfrak{W}_{\mathcal{R}} W_{G_0}\|_{\square} \leq C''' \delta^2$$

holds with probability (43).

To prove (46), we fix  $i \in [t]$ . Then for any two distinct vertices  $u$  and  $v$  we define random variables

$$Y_{uv} = Y_{vu} := \mathbb{E}(\mathbb{I}_{uv \in G_i} - \mathbb{I}_{uv \in G_{i-1}} \mid \mathcal{F}_{i-1}),$$

and write

$$\mathbb{E}(X_i \mid \mathcal{F}_{i-1}) = \sum_{u \in S, v \in T: u \neq v} Y_{uv}.$$

Note that  $Y_{uv}$  is a function of  $G_{i-1}$ . Hence we can fix (“condition on”)  $G_{i-1}$  and calculate the expectation of  $\mathbb{I}_{uv \in G_i} - \mathbb{I}_{uv \in G_{i-1}}$  with respect to the random pair  $(\mathbf{v}_i, \mathbf{H}_i) \in [n]^k \times \mathcal{H}_k$ .

Writing  $\mathbf{v}_i = (v_1, \dots, v_k)$ , note that events  $E_{a,b} := \{v_a = u, v_b = v\}$ ,  $1 \leq a < b \leq k$ , are mutually exclusive and equally likely, each of probability  $1/(n)_2$ . If we condition on  $E_{a,b}$ , then for any  $F, H \in \mathcal{H}_k$  the probability that the function  $j \mapsto v_j$ ,  $j \in [k]$  defines an isomorphism between  $F$  and  $G_{i-1}[\mathbf{v}_i]$  is exactly  $t_{\text{ind}}^{(u,v)}(F^{a,b}, G_{i-1})$ . The probability that then  $G_{i-1}[\mathbf{v}_i]$  is replaced by  $H$  is  $\mathcal{R}_{F,H}$ . Hence

$$(48) \quad Y_{uv} = \sum_{1 \leq a \neq b \leq k} \frac{1}{(n)_2} \sum_{F, H \in \mathcal{H}_k} t_{\text{ind}}^{(u,v)}(F^{a,b}, G_{i-1}) \cdot \mathcal{R}_{F,H} \cdot (\mathbb{I}_{\{ab \in H \setminus F\}} - \mathbb{I}_{\{ab \in F \setminus H\}}).$$

Now we want to relate (48) to the function  $\mathfrak{W}_{\mathcal{R}} W_{i-1}$ , which is constant on the set  $\Omega_u \times \Omega_v$ . By (14), for  $x \in \Omega_u$ ,  $y \in \Omega_v$  we have

$$\left|t_{\text{ind}}^{(u,v)}(F^{a,b}, G_{i-1}) - t_{\text{ind}}^{(x,y)}(F^{a,b}, W_{i-1})\right| \leq \frac{\binom{k}{2}}{n},$$

which allows us to rewrite (48) as

$$Y_{uv} = \sum_{1 \leq a \neq b \leq k} \frac{1}{(n)_2} \sum_{F, H \in \mathcal{H}_k} \left(t_{\text{ind}}^{(x,y)}(F^{a,b}, W_{i-1}) \pm \frac{\binom{k}{2}}{n}\right) \cdot \mathcal{R}_{F,H} \cdot (\mathbb{I}_{\{ab \in H \setminus F\}} - \mathbb{I}_{\{ab \in F \setminus H\}}).$$

To bound the aggregate contribution of the error term  $\pm \frac{\binom{k}{2}}{n}$ , we need to include a factor of  $\binom{k}{2}$  because of the summation over  $1 \leq a \neq b \leq k$ , a factor of  $|\mathcal{H}_k|$  because of the summation of  $H \in \mathcal{H}_k$ , and a factor of 1 because of the summation  $F \in \mathcal{H}_k$  (here, we make use of the term  $\mathcal{R}_{F,H}$  and (4)). Hence,

$$Y_{uv} = \sum_{1 \leq a \neq b \leq k} \frac{1}{(n)_2} \sum_{F, H \in \mathcal{H}_k} t_{\text{ind}}^{(x,y)}(F^{a,b}, W_{i-1}) \cdot \mathcal{R}_{F,H} \cdot (\mathbb{I}_{\{ab \in H \setminus F\}} - \mathbb{I}_{\{ab \in F \setminus H\}}) \pm \frac{\binom{k}{2}^2 |\mathcal{H}_k|}{n},$$

and we recover (25). Thus, for any  $x \in \Omega_u$ ,  $y \in \Omega_v$ ,

$$(49) \quad \left|Y_{uv} - \frac{1}{(n)_2} \mathfrak{W}_{\mathcal{R}} W_{i-1}(x, y)\right| \leq \frac{\binom{k}{2}^2 |\mathcal{H}_k|}{n^2(n-1)}.$$

Since  $\mathfrak{Y}_{\mathcal{R}} W_{i-1}(x, y) = n^2 \int_{\Omega_u \times \Omega_v} \mathfrak{Y}_{\mathcal{R}} W_{i-1}$  for any  $x \in \Omega_u, y \in \Omega_v$ , and (33) implies

$$\left| \int_{\Omega_u \times \Omega_v} \mathfrak{Y}_{\mathcal{R}} W_{i-1} \right| \leq \frac{(k)_2}{n^2},$$

we obtain

$$(50) \quad \left| Y_{uv} - \int_{\Omega_u \times \Omega_v} \mathfrak{Y}_{\mathcal{R}} W_{i-1} \right| \leq \frac{\binom{k}{2}^2 |\mathcal{H}_k| + (k)_2}{n^2(n-1)} \leq \frac{C'}{n^3},$$

where the extra term  $(k)_2$  is due to replacement of  $(n)_2$  by  $n^2$  in (49). Again using the bound  $\|\mathfrak{Y} W_{i-1}\|_{\infty} \leq (k)_2$  from (33),

$$(51) \quad \begin{aligned} & \left| \int_{\Omega_S \times \Omega_T} \mathfrak{Y}_{\mathcal{R}} W_{i-1} - \sum_{u \in S, v \in T: u \neq v} \int_{\Omega_u \times \Omega_v} \mathfrak{Y}_{\mathcal{R}} W_{i-1} \right| \\ &= \left| \sum_{u \in S \cap T} \int_{\Omega_u \times \Omega_v} \mathfrak{Y}_{\mathcal{R}} W_{i-1} \right| \leq \frac{(k)_2 |S \cap T|}{n^2}. \end{aligned}$$

Summing (50) over  $u \neq v$  and using (51), we obtain

$$(52) \quad \left| \mathbb{E}(X_i | \mathcal{F}_{i-1}) - \int_{\Omega_S \times \Omega_T} \mathfrak{Y}_{\mathcal{R}} W_{i-1} \right| \leq \frac{C' |S| |T|}{n^3} + \frac{(k)_2 |S \cap T|}{n^2} \leq \frac{C' + (k)_2}{n}.$$

Now Lemma 4.10 implies that, for  $i = 1, \dots, t$ ,

$$(53) \quad \begin{aligned} & \left| \int_{\Omega_S \times \Omega_T} \mathfrak{Y}_{\mathcal{R}} W_i - \int_{\Omega_S \times \Omega_T} \mathfrak{Y}_{\mathcal{R}} W_0 \right| \leq C_{\square} \|W_i - W_0\|_{\square} \\ & \leq C_{\square} \int_{\Omega^2} |W_i - W_0| = \frac{C_{\square} |E(G_i) \ominus E(G_0)|}{n^2} \\ & \leq C_{\square} \binom{k}{2} \frac{i}{n^2} \leq C_{\square} \binom{k}{2} \delta = C'' \delta. \end{aligned}$$

Recalling the assumption  $\delta \geq 1/n$ , from (52) and (53) we infer that the left-hand side of (46) is at most

$$(C' + (k)_2) \delta n + C'' \delta^2 n^2 \leq (C' + (k)_2 + C'') \delta^2 n^2 = (C''' - 1) \delta^2 n^2.$$

For every  $i$  we have  $X_i \in [-(\binom{k}{2}), (\binom{k}{2})]$  and hence  $\mathbb{E}(X_i | \mathcal{F}_{i-1}) \in [-(\binom{k}{2}), (\binom{k}{2})]$ . Consequently  $|Y_i| \leq (k)_2$ . Hence, (47) follows by applying Lemma A.1 to both  $\sum_{i=1}^t Y_i$  and  $\sum_{i=1}^t (-Y_i)$ .  $\square$

We can now prove Theorem 5.1.

**Proof of Theorem 5.1.** For simplicity we write  $W_i = W_{G_i}$  and  $\mathfrak{Y} W_i = \mathfrak{Y}_{\mathcal{R}} W_{G_i}$ .

We first show that it is enough to fix a small number  $\delta > 0$  such that  $t := \delta n^2$  is an integer and to prove (41) only for  $i$  being multiples of  $t$  that fall in the interval  $(0, Tn^2]$ . For any integer  $i \in [Tn^2]$  let  $i'$  be the largest multiple of  $t$  not larger than  $i$ , namely  $i' = t \lfloor i/t \rfloor$ . By the triangle inequality,

$$\left| \|W_i - \Phi^{i/n^2} W_0\|_{\square} - \|W_{i'} - \Phi^{i'/n^2} W_0\|_{\square} \right| \leq \|W_i - W_{i'}\|_{\square} + \|\Phi^{i'/n^2} W_0 - \Phi^{i/n^2} W_0\|_{\square}.$$

We have  $\|W_i - W_{i'}\|_{\square} \stackrel{(7)}{\leq} \|W_i - W_{i'}\|_1 = 2|E(G_i) \ominus E(G_{i'})|/n^2 \leq 2(i - i') \binom{k}{2} / n^2$ . Moreover, by (7) and (36),  $\|\Phi^{i'/n^2} W_0 - \Phi^{i/n^2} W_0\|_{\square} \leq (k)_2 (i - i') / n^2$ . Hence

$$\begin{aligned} & \left| \|W_i - \Phi^{i/n^2} W_0\|_{\square} - \|W_{i'} - \Phi^{i'/n^2} W_0\|_{\square} \right| \\ & \stackrel{\text{triangle inequality}}{\leq} \|W_i - W_{i'}\|_{\square} + \|\Phi^{i'/n^2} W_0 - \Phi^{i/n^2} W_0\|_{\square} \\ & \leq \frac{2(i - i')}{n^2} \binom{k}{2} + (k)_2 \frac{i - i'}{n^2} \leq 4 \binom{k}{2} \delta. \end{aligned}$$

Hence if we choose  $\delta$  small enough to make sure that, say,  $4\binom{k}{2}\delta \leq \varepsilon/2$ , it suffices to prove that with high probability

$$(54) \quad \max_{i \in (0, Tn^2] \cap t\mathbb{Z}} \|W_i - \Phi^{i/n^2} W_0\|_{\square} \leq \frac{\varepsilon}{2}.$$

With foresight we define a constant (depending on  $k$ )

$$C^* := \frac{C_{\square}}{2(C''' + C_k(k)_2 + 1)},$$

where  $C_{\square} = C_{\square}(k, 0)$  is the constant defined in Lemma 4.10,  $C_k$  is defined in (30), and  $C'''$  is defined in (45). Now we set

$$(55) \quad \delta = \frac{C^* \varepsilon}{e^{C_{\square} T}}.$$

We further set

$$r := |\{i \in (0, Tn^2] \cap t\mathbb{Z}\}| = \left\lfloor \frac{Tn^2}{t} \right\rfloor = \left\lfloor \frac{T}{\delta} \right\rfloor.$$

By Lemma 5.5 and union bound we have

$$(56) \quad \max_{i \in [r]} \|W_{it} - W_{(i-1)t} - \delta \mathfrak{Y} W_{(i-1)t}\|_{\square} \leq (C''' + 1)\delta^2$$

with probability at least

$$\begin{aligned} & 1 - r2^{2n+1} \exp\left(-\frac{\delta^3 n^2}{2(k)_2^2}\right) \\ & \geq 1 - \frac{T e^{C_{\square} T}}{C^* \varepsilon} \exp\left((2 \ln 2)n - \frac{C^* \varepsilon^3 n^2}{2(k)_2^2 e^{3C_{\square} T}}\right). \end{aligned}$$

For the following fix an outcome of the flip process such that (56) holds. For every  $i \in [0, r]$  we have that

$$(57) \quad \|\Phi^{\delta} W_i - W_i - \delta \mathfrak{Y} W_i\|_{\square} \stackrel{(7)}{\leq} \|\Phi^{\delta} W_i - W_i - \delta \mathfrak{Y} W_i\|_{\infty} \stackrel{(37)}{\leq} C_k(k)_2 \delta^2.$$

From (56) and (57) we get that for every  $i \in [r]$  that

$$(58) \quad \begin{aligned} \|W_{it} - \Phi^{\delta} W_{(i-1)t}\|_{\square} & \leq \|W_{it} - W_{(i-1)t} - \delta \mathfrak{Y} W_{(i-1)t}\|_{\square} \\ & \quad + \|W_{(i-1)t} + \delta \mathfrak{Y} W_{(i-1)t} - \Phi^{\delta} W_{(i-1)t}\|_{\square} \\ & \leq (C''' + 1 + C_k(k)_2)\delta^2 =: \varepsilon_2. \end{aligned}$$

Theorem 4.16 implies that for every  $i \in [r]$

$$(59) \quad \|\Phi^{\delta} W_{(i-1)t} - \Phi^{\delta} \Phi^{(i-1)\delta} W_0\|_{\square} \leq e^{C_{\square} \delta} \|W_{(i-1)t} - \Phi^{(i-1)\delta} W_0\|_{\square}.$$

We further prove by induction that for  $i = 0, 1, \dots, r$  we have

$$(60) \quad \|W_{it} - \Phi^{i\delta} W_0\|_{\square} \leq \frac{(e^{C_{\square} \delta i} - 1)\varepsilon_2}{e^{C_{\square} \delta} - 1}.$$

The base  $i = 0$  is trivial and for the induction step assuming  $i \geq 1$  we have

$$\begin{aligned} \|W_{it} - \Phi^{i\delta} W_0\|_{\square} & \stackrel{(29)}{\leq} \|W_{it} - \Phi^{\delta} W_{(i-1)t}\|_{\square} + \|\Phi^{\delta} W_{(i-1)t} - \Phi^{\delta} \Phi^{(i-1)\delta} W_0\|_{\square} \\ & \stackrel{\text{by (58), (59)}}{\leq} \varepsilon_2 + e^{C_{\square} \delta} \|W_{(i-1)t} - \Phi^{(i-1)\delta} W_0\|_{\square} \\ & \stackrel{\text{induction hypothesis}}{\leq} \varepsilon_2 + \frac{e^{C_{\square} \delta} (e^{C_{\square} \delta (i-1)} - 1)\varepsilon_2}{e^{C_{\square} \delta} - 1} = \frac{(e^{C_{\square} \delta i} - 1)\varepsilon_2}{e^{C_{\square} \delta} - 1}, \end{aligned}$$

as was needed.

This puts us in a position to prove (54), which was needed to finish the proof. Indeed, for every  $i \in [r]$  we have

$$\begin{aligned} \|W_{it} - \Phi^{i\delta} W_0\|_{\square} &\stackrel{(60)}{\leq} \frac{e^{C_{\square}\delta r} - 1}{e^{C_{\square}\delta} - 1} \varepsilon_2 \\ \boxed{r \leq T/\delta, e^x \geq 1+x} &\leq e^{C_{\square}T} \varepsilon_2 / C_{\square}\delta \\ \boxed{\text{by (58)}} &= e^{C_{\square}T} \cdot (C''' + C_k(k)_2 + 1)\delta / C_{\square} \\ \boxed{\text{by (55)}} &= \varepsilon/2. \end{aligned} \quad \square$$

### 5.2. Updates along a Poisson point process

In the model of flip processes we discussed so far, the flips occur at discrete times. The following modification may be more suitable for applications. Starting with a graph  $G = G_0^*$  we consider a continuous time Markov process  $(G_t^*)_{t \geq 0}$  in which we update the graph along a Poisson point process with intensity 1 according to the same procedure as in the discrete case. We then have a counterpart of Theorem 5.1. Indeed, at any given time  $Tn^2$ , by the law of large numbers, the number of arrivals is  $(1 + o(1))Tn^2$ , with exponentially high probability.

## 6. Properties of trajectories

In this part we collect various useful properties of trajectories with respect to a fixed flip process. Some of these properties are interesting per se, other are more of an auxiliary nature and used later on. Let us quickly summarise these results. In Section 6.1 we study backwards evolution of trajectories. Recall that in Definition 4.6, we defined the age of each graphon  $W$  as the maximum time by which the backward trajectory stays in the space of graphons. We define the *origin* of  $W$  as  $\Phi^{-\text{age}(W)} W$ . We also prove that *age* is upper semicontinuous. For many natural flip processes and many initial graphons  $W$ , the trajectory  $\Phi^t W$  converges as  $t \rightarrow \infty$ . In Section 6.2 we define these limits as *destinations*. Section 6.4 treats graphons which have a step structure, and in particular proves that this step structure is preserved throughout the evolution. In Section 6.5 we prove that values 0 and 1 cannot emerge in a graphon throughout its evolution. Section 6.6 proves the plausible fact that a zero velocity on a part  $\Omega_1 \times \Omega_2$  of  $\Omega^2$  is a sufficient condition for a graphon to stay constant on  $\Omega_1 \times \Omega_2$  throughout its evolution. Section 6.7 treats evolution of constant graphons. In particular, we get that each flip process has at least one *fixed point*, that is, a graphon where the trajectory stays constant. In Section 6.8 we prove that any nontrivial rule has at least one nontrivial trajectory. In Section 6.9 we introduce stable destinations as counterparts to asymptotically stable fixed points in the theory of dynamical systems.

The most interesting results about trajectories are given in Sections 6.10 and 6.11. We try to understand the situation in which a given trajectory does not converge, thus complementing Section 6.2. In Section 6.10 we prove the existence (in a non-constructive way) of a periodic trajectory, using the Poincaré–Bendixson theorem as the main tool. In Section 6.11 we study how complicated a single trajectory can be. For example, we prove that a single trajectory cannot be dense in the space of graphons.

Section 6.12 studies the speed of convergence, that is, how fast convergent trajectories approach their destination. The last Section 6.13 is short, but contains an important uniqueness question: what can we say about rules which yield the same trajectories?

### 6.1. Going back in time

In this section, we look at origins of each graphon with respect to our trajectories. Recall that, given a graphon  $W$ ,  $\text{age}(W)$  is the time for which the trajectory can be extended from  $W$  backwards within the space  $\mathcal{W}_0$  of graphons. For many flip processes, the age of many graphons is finite. For example, in the Erdős–Rényi flip process, when going backwards, the values at each point of the graphon decrease faster and faster, eventually becoming negative.

If  $\text{age}(W) < \infty$ , we define the *origin* of  $W$

$$\text{orig}_{\mathcal{R}}(W) := \Phi^{-\text{age}(W)} W,$$

which, by Theorem 4.5(iii), is well-defined and is indeed a graphon.

**Remark 6.1.** Given a rule  $\mathcal{R}$ , let  $W \in \mathcal{W}_0$  be a graphon such that  $\text{age}(W) < \infty$ . Writing  $U := \text{orig}(W)$ , we have  $\text{ess inf } U = 0$  or  $\text{ess sup } U = 1$ , since otherwise  $U$  belongs to the interior of  $\mathcal{W}_0$  with respect to the norm  $\|\cdot\|_{\infty}$ , and the trajectory starting at  $U$  stays, by continuity (cf. Lemma 4.15), inside  $\mathcal{W}_0$  for some negative time contradicting the fact that  $t = -\text{age}(W)$  was the earliest time  $\Phi^t W$  was a graphon.

Let us now look at continuity properties of the map  $W \mapsto \text{age}(W)$  in the space of graphons. As usual, the topology of interest is that of the cut norm. Unfortunately, the map  $\text{age}$  need not be continuous. Consider the triangle removal flip process. The age of the constant-0 graphon is  $\infty$ . However, for an arbitrarily small  $\alpha > 0$  we can take a set  $A \subset \Omega$  of measure  $\alpha$  and a graphon that is zero everywhere except on a set  $A \times A$  where it is 1. The age of that graphon is 0. To be more exact, this example shows that the map  $\text{age}$  need not be lower semicontinuous. However, upper semicontinuity is guaranteed by the following theorem.

**Theorem 6.2.** *For any rule  $\mathcal{R}$ , if  $(W_n)_{n=1}^\infty$  is a sequence of graphons that converges to a graphon  $W$  in the cut norm, then  $\limsup_n \text{age}(W_n) \leq \text{age}(W)$ .*

**Proof.** The statement is vacuous when  $\limsup_n \text{age}(W_n) = 0$ , so assume the contrary. We need to prove that for any (finite)  $\gamma \in (0, \limsup_n \text{age}(W_n))$  the trajectory of  $W$  is well-defined in backward time until  $-\gamma$ , that is, there exists a graphon  $U$  so that  $\Phi^\gamma U = W$ . By passing to a subsequence, we can assume that for every  $n$  we have  $\text{age}(W_n) > \gamma$ , and hence graphons  $U_n := \Phi^{-\gamma} W_n$  exist. We apply Theorem 4.16 with  $t = -\gamma$  and graphons  $(W_n)_n$ , and see that the sequence  $(U_n)_n$  is Cauchy with respect to the cut norm distance. Hence, by Lemma 3.2 the sequence  $(U_n)_n$  converges to some graphon  $U$  in the cut norm. Applying Theorem 4.16, this time with  $t = \gamma$  and to graphons  $U, U_1, U_2, \dots$ , we see that  $(\Phi^\gamma U_n = W_n)_n$  converges to  $\Phi^\gamma U$ , implying  $\Phi^\gamma U = W$ .  $\square$

6.1.1. *Going back in time even more*

- If  $\text{age}(W) < \infty$ , by Theorem 4.5 the trajectory  $\Phi^t W$  is defined on an open interval  $\mathcal{D}_W$  containing  $[-\text{age}(W), \infty)$ . In particular it is defined for some  $t < -\text{age}(W)$ . We may have  $\mathcal{D}_W \neq (-\infty, \infty)$ . For example, we saw that the trajectory of the triangle removal flip process along constant graphons given by (3) satisfies  $\lim_{t \searrow -\frac{1}{12}} \Phi^t(1) = +\infty$ , so  $\mathcal{D}_1 = (-\frac{1}{12}, \infty)$ .
- Theorem 4.5(iii) implies that if we follow the trajectory from a graphon backwards, once we leave the graphon space, we never come back.
- We do not have a combinatorially meaningful interpretation of the trajectories outside the space of graphons. For example the velocity of the constant-(-1) kernel with respect to the triangle removal flip process is positive, contrary to the intuition that the triangle removal flip process decreases values.

6.2. *Destinations*

Suppose that we have a flip process with a rule  $\mathcal{R}$ . Given a graphon  $U$ , we say that *the trajectory of  $U$  is convergent* if there exists a *destination of  $U$* , denoted  $\text{dest}_{\mathcal{R}}(U)$  such that  $\Phi^t U \xrightarrow{\square} \text{dest}_{\mathcal{R}}(U)$  as  $t \rightarrow \infty$ . We say that a the flip process with rule  $\mathcal{R}$  is *convergent* if all trajectories converge.

Trivially each fixed point  $W$  is a destination (of the trajectory  $\Phi \cdot W \equiv W$ ). We observe that a destination is necessarily a fixed point.

**Proposition 6.3.** *Suppose that  $\mathcal{R}$  is a flip process. If  $U$  is a graphon with a destination  $W$ , then  $\mathfrak{V}_{\mathcal{R}} W = 0$ .*

**Proof.** For every  $t \geq 0$ , by the cut norm continuity of  $W \mapsto \Phi^t W$  (see Lemma 4.10) we have that (will all limits with respect to the cut norm)

$$W = \lim_{u \rightarrow \infty} \Phi^{t+u} U = \lim_{u \rightarrow \infty} \Phi^t \Phi^u U = \Phi^t \left( \lim_{u \rightarrow \infty} \Phi^u U \right) = \Phi^t W$$

whence  $\mathfrak{V} W = \lim_{t \searrow 0} (\Phi^t W - W)/t \equiv 0$ .  $\square$

The next conjecture asserts that destinations are not only cut norm limits but also  $L^1$ -norm limits. It could be that they are even  $L^\infty$ -norm limits.

**Conjecture 6.4.** *Suppose that  $\mathcal{R}$  is a flip process. If  $U$  is a graphon with a destination  $W$ , then  $(\Phi^t U)_{t \rightarrow \infty}$  converges to  $W$  also in  $L^1$ .*

6.3. *Permuting the ground space*

Recall that  $\mathbb{S}_\Omega$  is the set of all measure preserving bijections from  $\Omega$  to  $\Omega$ . Given a rule  $\mathcal{R}$  and  $\varphi \in \mathbb{S}_\Omega$ , it is easy to check that for every graphon  $W$  we have  $\mathfrak{V}(W^\varphi) = (\mathfrak{V} W)^\varphi$ . Since the operator  $W \mapsto W^\varphi$  is continuous on  $L^\infty(\Omega^2)$ , it easily



follows from the definition that for every graphon  $W$  the function  $t \mapsto (\Phi^t W)^\varphi$  is a trajectory starting at  $U := W^\varphi$ . By uniqueness,  $I_U = I_W$  and  $\Phi \cdot U = (\Phi \cdot W)^\varphi$ . Continuity also implies that  $\text{dest } U = (\text{dest } W)^\varphi$ , should at least one of the destinations exist.

### 6.4. Step structure

One of the structurally simplest classes of graphons are step graphons. Related random graph models called *stochastic block models* were actually studied prior to the notion of graphons (recall the terminology from Section 3.2.4). One can check that the velocity of a step graphon  $W$  (for any fixed flip process) is a kernel with the same step structure, which in turn yields that  $\Phi^t(W)$  still has the same step structure (cf. Remark 4.9). We derive this claim from the following slightly stronger proposition.

**Proposition 6.5.** *Consider an arbitrary flip process and a graphon  $W$ . If  $A \subset \Omega$  is a twin-set of  $W$ , then for any  $t \in I_W$ , the set  $A$  is also a twin-set for  $\Phi^t W$ .*

**Proof.** The set  $\mathcal{W}_A := \{W \in \mathcal{W} : A \text{ is a set of twins in } W\}$  is clearly a linear subspace of the space of kernels  $\mathcal{W}$ . Lemma 3.8 tells us that  $\mathcal{W}_A$  is a closed subspace with respect to the  $\|\cdot\|_\infty$ -norm. Crucially, observe that whenever  $U \in \mathcal{W}_A$ , then it follows from Definition 4.1 that  $\mathfrak{V} U \in \mathcal{W}_A$ . Applying Theorem 3.11, along the same lines as in Theorem 4.5, to the Banach space  $(\mathcal{W}_A, \|\cdot\|_\infty)$  (rather than to the whole  $\mathcal{W}$ ) it follows that  $\Phi^t W \in \mathcal{W}_A$ .  $\square$

The following corollary says that the step structure of a graphon stays the same along the trajectory. A special case of the corollary is that if  $W$  is not a constant graphon, then  $\Phi^t(W)$  is not a constant graphon for every  $t \in I_W$ .

**Corollary 6.6.** *If  $W$  is a step graphon with a minimal step partition  $S_1 \cup S_2 \cup \dots$ , then for every  $t \in I_W$  the graphon  $\Phi^t W$  is a step graphon with a minimal step partition  $S_1 \cup S_2 \cup \dots$ .*

**Proof.** It follows immediately from Proposition 6.5 that  $S_1 \cup S_2 \cup \dots$  is a step partition for any  $\Phi^t W$ . It remains to argue that it stays minimal.

Suppose  $\Phi^t W =: U$  is a step graphon with respect to a coarser step partition  $T_1 \cup T_2 \cup \dots$ . By Proposition 6.5, applied to graphon  $U$ , graphon  $W = \Phi^{-t} U \in \mathcal{W}_0$  is a step graphon with partition  $T_1 \cup T_2 \cup \dots$ , contradicting the minimality of partition  $S_1 \cup S_2 \cup \dots$  for  $W$ .  $\square$

### 6.5. Values 0 and 1

Here, we give a quick proof of the fact that we cannot arrive to values 0 and 1. This result will be useful in Section 6.11.

**Proposition 6.7.** *For any flip process, any graphon  $W$ , and every  $t > 0$  the sets*

$$A_t := \{(x, y) \in \Omega^2 : W(x, y) > 0, \Phi^t W(x, y) = 0\},$$

$$B_t := \{(x, y) \in \Omega^2 : W(x, y) < 1, \Phi^t W(x, y) = 1\}$$

*have measure zero.*

**Proof.** The proof is similar to the proof of Claim 4.13, but things are now simpler since we know  $\Phi^t W \in \mathcal{W}_0$  for every  $t \geq 0$ . By (31), we have  $\mathfrak{V} \Phi^t W \geq -(k)_2 \Phi^t W$ , so using Remark 4.9 and Proposition C.4 for justification, we can fix  $x, y \in \Omega$  and assume that  $F(t) = \Phi^t W(x, y)$  and  $f(t) = \mathfrak{V} \Phi^t W(x, y)$  satisfy  $F(t) - F(0) = \int_0^t f(\tau) \, d\tau$  and  $f(t) \geq -(k)_2 F(t)$  for  $t \geq 0$ . Lemma 3.10 implies that

$$\Phi^t W(x, y) = F(t) \geq F(0)e^{-(k)_2 t} = e^{-(k)_2 t} W(x, y).$$

It follows that  $\Phi^t W > 0$  wherever  $W > 0$ . Since, when applying Proposition C.4, we modified  $W$  and  $\Phi^t W$  only on sets of measure zero, the proof is complete for the set  $A_t$ . The proof for  $B_t$  is analogous.  $\square$

### 6.6. Zero velocity on a section

Consider a subspace of graphons which agree on a fixed subset  $S$ . Our next proposition states an intuitive and sometimes “obvious” thing: if the velocity of every graphon in the subspace equals zero on  $S$ , then any trajectory started in the subspace stays inside it.

**Proposition 6.8.** *Consider a flip process of order  $k$ . Given two sets  $\Omega_1, \Omega_2 \subset \Omega$ , let  $S := \Omega_1 \times \Omega_2 \cup \Omega_2 \times \Omega_1$  and let  $\alpha : S \rightarrow [0, 1]$  be a symmetric function. Let  $\mathcal{A} = \{W \in \mathcal{W}_0 : W|_S = \alpha\}$ . Suppose that for each  $W \in \mathcal{A}$  we have  $(\mathfrak{V} W)|_S = 0$ . If  $U \in \mathcal{A}$  and  $t \geq 0$ , then  $\Phi^t U \in \mathcal{A}$ .*

**Proof.** We denote the  $L^\infty$ -norm on the space of functions with domain  $S$  as  $\|\cdot\|_*$ . That is to make a distinction from the  $L^\infty$ -space on  $\Omega^2$  for which we use the symbol  $\|\cdot\|_\infty$ . Given an arbitrary graphon  $W \in \mathcal{W}_0$ , let its  $\alpha$ -surgery be a graphon  $W^{[\alpha]}$  which is equal to  $\alpha$  on  $S$  and to  $W$  on  $\Omega^2 \setminus S$ . Observe that we always have

$$(61) \quad \|W - W^{[\alpha]}\|_\infty = \|W|_S - \alpha\|_*,$$

and also

$$(62) \quad \|W - U\|_\infty \geq \|W|_S - U|_S\|_*.$$

Suppose that  $U \in \mathcal{A}$  is given. For each  $t \geq 0$  we define  $U_t := \Phi^t U$ , and write  $f(t) := \|(U_t)|_S - \alpha\|_*$ . In particular,  $f(0) = 0$ . Using that  $U|_S = \alpha$  and  $(\mathfrak{V}(U_t^{[\alpha]}))|_S \equiv 0$ , we can write

$$\alpha = \left( U + \int_{\tau=0}^t \mathfrak{V}((U_\tau)^{[\alpha]}) \right) |_S.$$

We have

$$\begin{aligned} f(t) &= \left\| \left( U + \int_{\tau=0}^t \mathfrak{V} U_\tau \right) |_S - \left( U + \int_{\tau=0}^t \mathfrak{V}((U_\tau)^{[\alpha]}) \right) |_S \right\|_* \\ &\stackrel{\text{by (62)}}{\leq} \left\| \left( U + \int_{\tau=0}^t \mathfrak{V} U_\tau \right) - \left( U + \int_{\tau=0}^t \mathfrak{V}((U_\tau)^{[\alpha]}) \right) \right\|_\infty \\ &\stackrel{\text{triangle inequality}}{\leq} \int_{\tau=0}^t \|\mathfrak{V} U_\tau - \mathfrak{V}((U_\tau)^{[\alpha]})\|_\infty \\ &\stackrel{\text{by (32)}}{\leq} C_k \int_{\tau=0}^t \|U_\tau - ((U_\tau)^{[\alpha]})\|_\infty \\ &\stackrel{\text{by (61)}}{=} C_k \int_{\tau=0}^t \|(U_\tau)|_S - \alpha\|_* = C_k \int_{\tau=0}^t f(\tau). \end{aligned}$$

Applying Lemma 3.9 (with  $A = 0$ ), we obtain, for every  $t \geq 0$ , that  $f(t) \leq A \exp(C_k t) = 0$ , which implies that  $f(t) = 0$  and therefore  $\Phi^t U = U_t \in \mathcal{A}$ , as was needed.  $\square$

### 6.7. Constant graphons

Suppose that  $\mathcal{R}$  is a flip process of order  $k$ . We shall study the trajectories when restricted to the space of constant graphons (recall that by Corollary 6.6 trajectories started from constants stay constant and that no trajectory started at a non-constant reaches a constant). Moreover, the velocity of a constant graphon is also constant. Given a constant- $d$  graphon  $W$ , let  $\mathbf{F}$  be the drawn graph and  $\mathbf{H}$  be the replacement graph. By integrating (26) over  $\Omega^2$ , it follows that

$$(63) \quad \mathfrak{V} d = \int_{\Omega^2} \mathfrak{V} d = 2\mathbb{E}(e(\mathbf{H}) - e(\mathbf{F})).$$

For each  $\ell \in \{0, \dots, \binom{k}{2}\}$ , write

$$(64) \quad \Delta_\ell := \mathbb{E}(e(\mathbf{H}) - e(\mathbf{F}) \mid e(\mathbf{F}) = \ell) = \binom{k}{\ell}^{-1} \sum_{H \in \mathcal{H}_k : e(H) = \ell} \sum_{H' \in \mathcal{H}_k} \mathcal{R}_{H, H'} \cdot (e(H') - \ell).$$

That is,  $\Delta_\ell$  is the expected change of the number of edges in the replacement graph in one step of the flip process, if the drawn graph is a uniformly random graph with  $\ell$  edges. Note that

$$(65) \quad \Delta_0 \geq 0 \quad \text{and} \quad \Delta_{\binom{k}{2}} \leq 0.$$

Since  $\mathbf{F}$  is a binomial random graph  $\mathbb{G}(k, d)$ , for each constant  $d \in [0, 1]$  the velocity satisfies

$$(66) \quad \mathfrak{V}d = 2 \sum_{\ell=0}^{\binom{k}{2}} \binom{\binom{k}{2}}{\ell} d^\ell (1-d)^{\binom{k}{2}-\ell} \Delta_\ell.$$

Viewed as a function in  $d$ , this is a polynomial of degree at most  $\binom{k}{2}$ . Firstly, let us characterise the case when this polynomial is trivial.

**Fact 6.9.**  $\mathfrak{V}d \equiv 0$  if and only if  $\Delta_\ell = 0$  for all  $\ell \in \{0, \dots, \binom{k}{2}\}$ .

**Proof.** If  $\Delta_\ell = 0$  for all  $\ell \in \{0, \dots, \binom{k}{2}\}$ , then obviously  $\mathfrak{V}d \equiv 0$ . Suppose, on the other hand, that there exists  $\Delta_h \neq 0$ , and let  $h$  be minimum such. Then looking at (66), we have  $\lim_{d \rightarrow 0} \frac{\mathfrak{V}d}{d^h} = \binom{\binom{k}{2}}{h} \Delta_h$  which implies that  $\mathfrak{V}d$  is indeed non-zero.  $\square$

So, let us proceed under the assumption that  $\mathfrak{V}d$  is nontrivial. The roots of  $\mathfrak{V}d$  on the interval  $[0, 1]$  are those constant graphons which are fixed points of  $\mathcal{R}$ . By (65) and the intermediate value theorem, there is at least one such fixed point. Hence we get the following.

**Proposition 6.10.** *Any flip process has at least one fixed point.*

A trajectory started at a value  $d \in (r, r')$  between two consecutive roots  $r$  and  $r'$  converges either to  $r$  or  $r'$  depending on the signum of  $\mathfrak{V}$  on that interval. Further, polynomial speed of convergence (see Section 6.12 for general treatment of non-constant graphons) can be easily deduced.

So, we have reduced the problem of studying trajectories on constants to questions about roots of polynomials of the form

$$(67) \quad f(x) = \sum_{\ell=0}^{\binom{k}{2}} c_\ell x^\ell (1-x)^{\binom{k}{2}-\ell},$$

where  $c_\ell \in 2 \binom{\binom{k}{2}}{\ell} \cdot [-\ell, \binom{k}{2} - \ell]$ . (It is clear from the second form of  $\Delta_\ell$  in (64) that one can choose a rule  $\mathcal{R}$  so that  $\mathfrak{V}_{\mathcal{R}}d = f(d)$ .) This coefficient restriction can be relaxed by multiplying by a small constant. After this the only restrictions that remain are  $c_0 \geq 0$  and  $c_{\binom{k}{2}} \leq 0$ , which translates as

$$(68) \quad f(0) \geq 0 \quad \text{and} \quad f(1) \leq 0.$$

The space of polynomials of the form (67) (with arbitrary real coefficients) is a linear space of dimension  $\binom{k}{2} + 1$ . This can be seen from the fact that each of the  $\binom{k}{2} + 1$  coefficients  $c_\ell$  can be chosen, and the only constant-0 polynomial is the one with all coefficients zero. Hence this space is the space of all polynomials of degree at most  $\binom{k}{2}$ .

Recall that  $f$  was the velocity. From (68) we conclude, for a given  $k$  and arbitrary points  $0 \leq x_1 \leq \dots \leq x_m \leq 1$ , where  $m \leq \binom{k}{2}$ , there exists a rule of order  $\mathcal{R}$ , which, when restricted to constant graphons, has fixed points  $\{x_1, \dots, x_m\}$  and non-fixed trajectories are monotone and take values in the intervals between the roots.

### 6.8. Examples of graphons not fixed for any (nontrivial) rule

In this section, we shall find graphons  $W$  for which we cannot have  $\mathfrak{V}_{\mathcal{R}}W \equiv 0$  for any nontrivial rule  $\mathcal{R}$ .

We shall work with step graphons with a small number of blocks. Hence, throughout this section  $\Omega$  is finite and  $\pi$  is a measure such that  $\pi(\{\omega\}) > 0$  for every  $\omega \in \Omega$ . Suppose there are steps  $x, y, z \in \Omega$  (possibly  $x = y$ ) such that  $W(x, y) = 0$

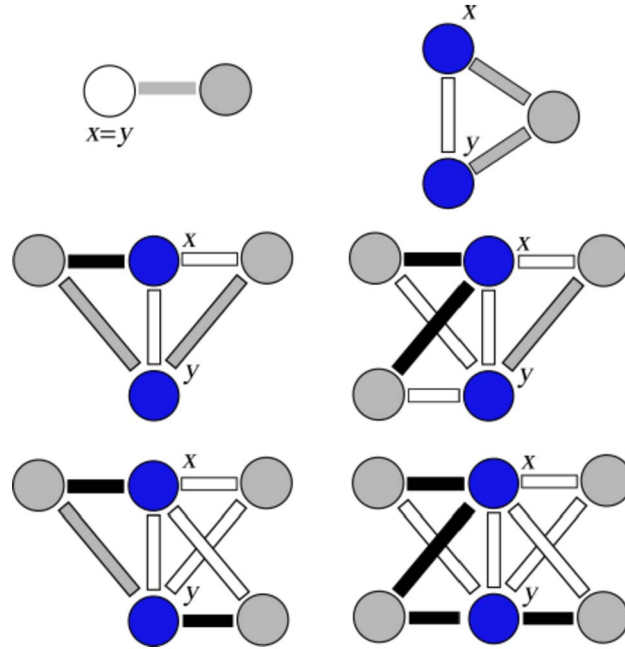


Fig. 4. Examples of configurations in a step graphon preventing  $\mathfrak{W}_{\mathcal{R}} W \equiv 0$  for any nontrivial rule  $\mathcal{R}$ . White has density 0, black has density 1. Different grey objects allow different densities (but necessarily in  $(0, 1)$ ). Blue clusters have unrestricted densities. Changing density of  $W(x, y)$  from 0 to 1 gives six more configurations.

and  $W(x, z), W(y, z), W(z, z) \in (0, 1)$ , see Figure 4, top two examples. The value  $\mathfrak{W}_{\mathcal{R}} W(x, y)$  is a linear combination of numbers  $\mathcal{R}_{G,H}$ , where the coefficient of each  $\mathcal{R}_{G,H}$  is proportional to

$$\sum_{ab \in H \setminus G} t_{\text{ind}}^{(x,y)}(G^{ab}, W) - \sum_{ab \in G \setminus H} t_{\text{ind}}^{(x,y)}(G^{ab}, W) = \sum_{ab \in H \setminus G} t_{\text{ind}}^{(x,y)}(G^{ab}, W)$$

because of  $W(x, y) = 0$ . Note that the first sum is positive whenever  $H \setminus G \neq \emptyset$ , since if we fix  $ab \in H \setminus G$ , then by mapping  $a$  to  $x$ ,  $b$  to  $y$  and the remaining vertices of  $G$  to  $z$ , we see that  $t_{\text{ind}}^{(x,y)}(G^{ab}, W) > 0$ . So in order to have  $\mathfrak{W}_{\mathcal{R}} W(x, y) = 0$  we need  $\mathcal{R}_{G,H} = 0$  whenever  $|H \setminus G| \geq 1$ . In other words, we can have  $\mathcal{R}_{H,G} > 0$  only if  $H \subseteq G$ . Suppose that  $\mathcal{R}_{G,H} > 0$  for some  $G, H$  such that  $H \subsetneq G$ . Since in  $\mathfrak{W}_{\mathcal{R}} W(z, z)$  the coefficient of such  $\mathcal{R}_{G,H}$  is, because of  $W(z, z) \in (0, 1)$ ,

$$- \sum_{ab \in G \setminus H} t_{\text{ind}}^{(z,z)}(G^{ab}, W) < 0,$$

implying that  $\mathcal{R}_{G,H} = 0$  for  $H \subsetneq G$  as well. This leaves only the possibility  $\mathcal{R}_{G,H} > 0$  when  $G = H$ , hence the rule  $\mathcal{R}$  is trivial.

By a symmetric argument  $\mathfrak{W}_{\mathcal{R}} W \equiv 0$  implies  $\mathcal{R}$  is trivial if value 0 for  $W(x, y)$  is replaced by value 1. There are a few more configurations which prevent a step graphon from being a fixed point in a nontrivial rule, see Figure 4.

The most important corollary is the following.

**Corollary 6.11.** *If a rule  $\mathcal{R}$  is such that every graphon is a fixed point, then  $\mathcal{R}$  is trivial.*

### 6.9. Stable destinations

We say that a graphon  $S \in \mathcal{W}_0$  is a *stable destination* for a given flip process if there exists  $\delta > 0$  such that for each  $W \in \mathcal{W}_0$  with  $d_{\square}(W, S) < \delta$  we have  $\text{dest}(W) = S$ . We call the supremum of such  $\delta$  the *radius of the stable destination*. In the dynamical systems terminology a stable destination is called an *asymptotically stable fixed point*.

Many nontrivial flip processes have at least one stable destination. However, this probably is not always the case. In [2] it is shown that the “extremist flip process” (see Section 2.7) for  $k = 3, 4$  does not have stable destinations at its most likely locations (though it does not rule out the existence of a stable destination at some other locations).

It is plausible that Theorem 5.1 (see also Remarks 5.2 and 5.3) can be strengthened for trajectories converging to a stable destination. That is, it is likely that a flip process started with a finite  $n$ -vertex graph close to such a trajectory typically stays in the neighbourhood of that destination for  $\exp(\Theta(n^2))$  steps, until a really rare event kicks in.

6.10. A periodic trajectory

In this section we prove existence of a flip process which has a periodic trajectory in  $\mathcal{W}_0$  (which is not a fixed point).

Let us first present the main idea, introducing necessary notation on the way. Fix a partition  $\Omega = \Omega_1 \cup \Omega_2$  with  $\pi(\Omega_1) = \pi(\Omega_2) = 1/2$ . We define a map  $\Gamma : \mathbb{R}^2 \rightarrow \mathcal{W}$  which associates to each  $(p_1, p_2) \in \mathbb{R}^2$  a step function

$$(69) \quad \Gamma(p_1, p_2) = p_1 \mathbb{1}_{\Omega_1^2} + p_2 \mathbb{1}_{\Omega_2^2},$$

that is,  $\Gamma(p_1, p_2)$  is constant  $p_i$  on  $\Omega_i^2$  for  $i = 1, 2$ , and zero elsewhere. Define a class of graphons  $\mathcal{U}_0 := \Gamma([0, 1]^2)$  and a class of kernels  $\mathcal{U} := \Gamma(\mathbb{R}^2)$ . We will show that for some  $\mathcal{R}$  there is a periodic trajectory taking values in  $\mathcal{U}_0$ . We will sometimes omit the symbol  $\Gamma$  and make no distinction between points in  $\mathbb{R}^2$  and elements of  $\mathcal{U}$ . In particular, we write  $\mathfrak{V}_{\mathcal{R}}(p_1, p_2) := \mathfrak{V}_{\mathcal{R}}(\Gamma(p_1, p_2))$ . It is immediate from the definition that  $\mathfrak{V}_{\mathcal{R}}(p_1, p_2)$  is a step function with respect to the partition  $\Omega = \Omega_1 \cup \Omega_2$ . Furthermore, the trajectory will be confined to  $\mathcal{U}_0$ , which is equivalent (with the help of Proposition 6.8) to

$$(70) \quad \mathfrak{V}_{\mathcal{R}}(p_1, p_2) \text{ is zero on } \Omega_1 \times \Omega_2.$$

Hence, we can treat  $\mathfrak{V}_{\mathcal{R}}(p_1, p_2)$  as a point in  $\mathbb{R}^2$  via  $\Gamma^{-1}$ .

For  $r > 0$  and  $p \in \mathbb{R}^2$ , we denote by  $B(r, p)$  the open ball of radius  $r$  around a point  $p$ , by  $\overline{B}(r, p)$  the corresponding closed ball and we set  $\text{circle}(r, p) = \overline{B}(r, p) \setminus B(r, p)$ .

(The following description uses tacitly the map  $\Gamma$ , as discussed above.) Our first attempt is to have the trajectory constrained to  $\text{circle}(r, q)$  with  $r = 0.1$  and  $q = (a, b) := (0.2, 0.8)$ . If we want to achieve that, then for each point  $(x, y) \in \text{circle}(r, q)$ , the velocity  $\mathfrak{V}(x, y)$  has to be a constant multiple of the tangent field  $g : \mathbb{R}^2 \setminus \{q\} \rightarrow \mathbb{R}^2$ ,

$$(71) \quad g(x, y) = (-y + b, x - a) / \sqrt{(-y + b)^2 + (x - a)^2}.$$

See Figure 5(left). So let us aim at  $\mathfrak{V}_{\mathcal{R}}(x, y) = g(x, y)$ . Assume for a while an idealised setting where we also have that

- (A) in the drawn graph  $F$ , each vertex carries the information whether it was sampled from  $\Omega_1$  or from  $\Omega_2$ , and also that it carries the information about the density  $x$  on  $\Omega_1 \times \Omega_1$  and the density  $y$  on  $\Omega_2 \times \Omega_2$ , and
- (B) the replacement distribution of the rule  $\mathcal{R}$  is allowed to depend on these additional input parameters.

In such a setting, we could easily add or remove edges on the vertices sampled from  $\Omega_1$  according to the first coordinate of  $g(x, y)$  (even cases where, for example, the first coordinate of  $g(x, y)$  is irrational could be handled by randomised outputs), and similarly for the vertices sampled from  $\Omega_2$ . We shall leave the pairs of vertices corresponding to  $\Omega_1 \times \Omega_2$  intact because of (70).

In reality we do not have (A). But when the order  $k$  of the flip process is large, then we can, with high probability, get a fairly precise estimate of all this data. To this end we use basic large deviation tools and further standard calculations to see that with probability  $1 - o(1)$  for the drawn graph  $F$  for the rule,  $F \sim \mathbb{G}(n, \Gamma(x, y))$  satisfies the following properties:

- (♣)  $F$  consists of two components, each of size  $(1 + o(1))\frac{k}{2}$  (henceforth until the end of section asymptotics are for  $k \rightarrow \infty$ ). Further, one component forms a graph of density  $d_1 = x + o(1)$  and all its vertices were sampled from  $\Omega_1$ . The other component forms a graph of density  $d_2 = y + o(1)$  and all its vertices were sampled from  $\Omega_2$ .

Recall that for each  $(x, y) \in \text{circle}(r, q)$  we have  $x \leq 0.3$  and  $y \geq 0.7$ . This means that with high probability the vertices in the sparser component were sampled from  $\Omega_1$  and the vertices from the denser component from  $\Omega_2$ . So, to circumvent the unavailability of (B) we might define the rule  $\mathcal{R}$  to be nontrivial on graphs of the form (♣) and for each such drawn graph the rule would change the density in the first and the second component according to the first and the second coordinate of  $g(d_1, d_2)$ , respectively.

Unfortunately, such a rule alone cannot guarantee that  $\text{circle}(r, q)$  will be the trajectory, since we do not have any control over the  $o(1)$ -errors. To give a specific example, even when the current graphon is  $(x, y)$ , with a small but positive probability it may happen that the drawn graph  $F$  will consist of two components of respective densities  $d_1 = \hat{x} \pm o(1)$  and  $d_2 = \hat{y} \pm o(1)$  for some very different  $(\hat{x}, \hat{y}) \in \text{circle}(r, q)$ . Hence, unwanted changes to the densities of the components of  $F$  will be applied.

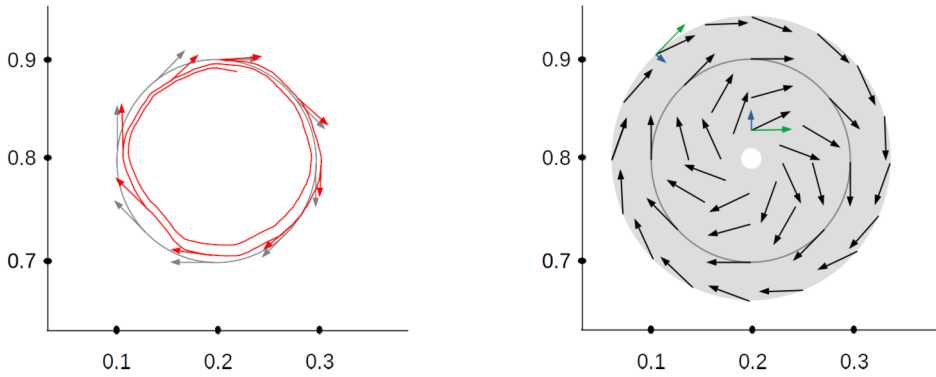


Fig. 5. *Left:* Ideal periodic trajectory given by the field  $g$  shown in grey arrows. Red is the actual trajectory (if the rule comes only from  $g$ ) which is perturbed by small inaccuracies and rare events. *Right:* Vector field  $g + h$  in black defines  $\mathcal{R}$  on an annulus containing  $\text{circle}(r, q)$ . Vectors of  $g$  and  $h$  appear in green and blue, respectively.

To amend the situation, rather than focusing solely on the circle, we work in an annulus  $C$  containing  $\text{circle}(r, q)$  in its interior, and  $\mathcal{R}$  will involve all possible graphs  $F$  as in (♣) where  $(d_1, d_2) \in C$ . On  $\text{circle}(r, q)$ , the rule  $\mathcal{R}$  is as before. On the rest of the annulus, the rule is given by the tangent vector field  $g$  composed with a vector field  $h : \mathbb{R}^2 \setminus \{q\} \rightarrow \mathbb{R}^2$  oriented towards  $\text{circle}(r, q)$ ,

$$(72) \quad h(x, y) = \left( \frac{r}{\sqrt{(x-a)^2 + (y-b)^2}} - 1 \right) (x-a, y-b).$$

See Figure 5(right). For drawn graphs  $F$  which do not consist of exactly two components whose joint densities lie in  $C$ , we define the replacement graph  $H$  arbitrarily, subject to the condition that

$$(73) \quad \text{no edges are introduced between any two components of } F.$$

The point is that even though we still know that  $\mathfrak{A}_{\mathcal{R}}(x, y)$  is only approximately equal to  $(h + g)(x, y)$  on  $C$ , we can prove that the undesired perturbation from rare events cannot cause that the actual velocity vector escapes from  $C$ . The classical Poincaré–Bendixson theorem then implies that there must exist a periodic trajectory. Note that (73) is needed to preserve (70); here even rare violations of (73) would kill (70).

Lemma 6.12 below allows us to produce a rule whose velocities approximate a given vector field; in fact such a rule is produced in the way we described above. To motivate some of the conditions of the lemma, we recall that we relied on the fact that the two components of the graph  $F$  are distinguishable by their densities (see (ii)), and also that in the planned change, either component cannot have density less than 0 or more than 1 (see (iv)).

**Lemma 6.12.** *Let  $\Gamma$  be defined by (69). For every  $\varepsilon, \delta > 0$  exists  $k \in \mathbb{N}$  such that the following holds. If  $S_1, S_2 \subseteq [0, 1]^2$  and  $f : S_1 \rightarrow \mathbb{R}^2$  are such that*

- (i)  $B(\delta, p) \subseteq S_1$  for every  $p \in S_2$ ,
- (ii)  $0 < \delta \leq p_1 \leq p_2 - \delta \leq 1 - 2\delta$  for every  $(p_1, p_2) \in S_1$ ,
- (iii)  $f$  is uniformly continuous on  $S_1$ , and
- (iv)  $0 \leq p_i + (k)_2^{-1} f_i(p_1, p_2) \leq 1$  for every  $i = 1, 2$  and  $(p_1, p_2) \in S_1$ ,

then there exists a rule  $\mathcal{R}$  of order  $k$  such that (70) holds for every  $p = (p_1, p_2) \in S_2$  and such that

$$(74) \quad \sup_{p \in S_2} |f(p) - \Gamma^{-1}(\mathfrak{A}_{\mathcal{R}}\Gamma(p))| \leq \varepsilon.$$

**Proof.** Within the proof the asymptotic notation is with respect to  $k \rightarrow \infty$  and whenever we say “for  $k$  large enough”, we mean  $k$  larger than some constant depending on  $\varepsilon, \delta$ . We fix some sequences  $\alpha_1 = \alpha_1(k), \alpha_2 = \alpha_2(k)$  such that

$$\frac{1}{\sqrt{k}} \ll \alpha_1 \ll \alpha_2 \ll 1.$$

We define a rule  $\mathcal{R}$  as follows. Assume that  $F \in \mathcal{H}_k$  is the drawn graph. We define the probabilities  $\mathcal{R}_{F,\cdot}$  by describing a random replacement graph  $\mathbf{H}$ . If  $F$  has an isolated vertex or the number of connected components in  $F$  is not two,

we do nothing, that is  $\mathbf{H} := F$ . If  $F$  has two components and each of them consists of at least 2 vertices, let  $F_1$  be the component of  $F$  with smaller edge density and let  $F_2$  be the other one (the ties can be broken arbitrarily). For  $i = 1, 2$ , we write  $t_i := e(F_i)/\binom{v(F_i)}{2}$  for the edge density of  $F_i$ . If  $t_i + f_i(t_1, t_2) < 0$  or  $t_i + f_i(t_1, t_2) > 1$  for some  $i \in [2]$ , then we again set  $\mathbf{H} := F$ . Otherwise we replace  $F_i$  by a random binomial graph with edge probability  $t_i + \binom{k}{2}^{-1} f_i(t_1, t_2)$ . Hence  $\mathbf{H}$  is the vertex-disjoint union of these two (independently drawn) random graphs.

Write  $p_{1,1} := p_1, p_{2,2} := p_2$  and  $p_{1,2} = p_{2,1} := 0$ . Let  $\mathbf{F} \sim \mathbb{G}(k, \Gamma(p))$  be a random  $k$ -vertex graph sampled from  $\Gamma(p_1, p_2)$ . We get the same random graph by sampling independent uniform elements  $\mathbf{U}_v \in [2]$  for each  $v \in [k]$  and then connecting each pair  $u, v$  independently with probability  $p_{\mathbf{U}_u, \mathbf{U}_v}$ . By Lemma 4.2,

$$(75) \quad (\mathfrak{A}_{\mathcal{R}}\Gamma(p))_{|\Omega_i \times \Omega_j} = \sum_{a \neq b, a, b \in [k]} (\mathbb{P}(ab \in \mathbf{H} \mid \mathbf{U}_a = i, \mathbf{U}_b = j) - p_{i,j}), \quad i, j \in [2].$$

Whenever  $i \neq j$ , the condition  $\mathbf{U}_a = i, \mathbf{U}_b = j$  implies that pair  $ab$  is a non-edge in  $\mathbf{H}$  and thus each term in (75) is zero. This implies that (70) holds for every  $p \in [0, 1]^2$ .

We further assume that  $p = (p_1, p_2) \in \mathcal{S}_2$  and prove (74). Let  $D_i$  be the edge density of the subgraph  $\mathbf{F}[i]$  of  $\mathbf{F}$  induced by  $\{v : \mathbf{U}_v = i\}$ . A standard Chernoff bound implies that with probability at least  $1 - 2 \exp(-\alpha_1^2 k/6)$  we have  $v(\mathbf{F}[i]) = (1 \pm \alpha_1)k/2$  for  $i \in [2]$ . We denote this event by  $\mathcal{A}$ . Furthermore,  $\mathbb{E}(e(\mathbf{F}[i]) \mid \mathcal{A}) = \binom{(1 \pm \alpha_1)k/2}{2} p_i \sim k^2 p_i/8$  and, assuming that  $\mathcal{A}$  holds, again a standard Chernoff bound implies that with probability at least  $1 - 2 \exp(-\alpha_1^2 (1 - \alpha_1)^2 k^2 p_i/24)$  we have  $e(\mathbf{F}[i]) = (1 \pm \alpha_1) \binom{(1 \pm \alpha_1)k/2}{2} p_i$  and therefore  $D_i = (1 \pm \alpha_2) p_i$ . Lastly, assuming  $\mathcal{A}$  holds, with probability at least

$$1 - \binom{v(\mathbf{F}[i])}{2} (1 - p_i^2)^{v(\mathbf{F}[i])-2} \geq 1 - \binom{k}{2} (1 - p_i^2)^{(1 - \alpha_1)k/2-2},$$

we have that in  $\mathbf{F}[i]$  every two vertices have a common neighbour and hence  $\mathbf{F}[i]$  is connected. We denote the event such that all of the three events above hold by  $\mathcal{C}$  and note that  $\mathbb{P}(\mathcal{C}) = 1 - \exp(-\Omega(\alpha_1^2 k)) = 1 - o(1)$ . In particular  $\mathcal{C}$  together with (ii) implies that  $D_2 \geq p_2 - \alpha_2 > p_2 - \delta/2 \geq p_1 + \delta/2 > p_1 + \alpha_2 \geq D_1$  and also that  $(D_1, D_2) \in B(\delta, (p_1, p_2))$  and therefore by (i) and (iv) we have that  $0 \leq D_i + \binom{k}{2}^{-1} f_i(D_1, D_2) \leq 1$ . In view of (i), for large enough  $k$  we have  $p(1 \pm \alpha_2) \in \mathcal{S}_1$ , so that condition (iii) implies  $|f_i(1 \pm \alpha_2)p - f_i(p)| = o(1)$ , uniformly over all choices of  $p \in \mathcal{S}_2$ .

Hence for  $i \in [2]$  and  $a, b \in [k]$  (note that by symmetry of the rule the following does not depend on the actual choice of  $ab$  as long as  $\mathbf{U}_a = \mathbf{U}_b = i$ ),

$$\begin{aligned} \mathbb{P}(ab \in \mathbf{H} \mid \mathbf{U}_a = \mathbf{U}_b = i) &= \mathbb{P}(ab \in \mathbf{H} \mid \mathbf{U}_a = \mathbf{U}_b = i, \mathcal{C}) \cdot \mathbb{P}(\mathcal{C}) \pm \mathbb{P}(\bar{\mathcal{C}}) \\ &= ((1 \pm \alpha_2) p_i + \binom{k}{2}^{-1} f_i((1 \pm \alpha_2) p_1, (1 \pm \alpha_2) p_2)) \cdot \mathbb{P}(\mathcal{C}) \pm \mathbb{P}(\bar{\mathcal{C}}) \\ &\stackrel{\text{by (iii)}}{=} ((1 \pm \alpha_2) p_i + \binom{k}{2}^{-1} f_i(p_1, p_2) + o(1)) \cdot (1 - o(1)) + o(1) \\ &\stackrel{\text{for large enough } k}{=} p_i + \binom{k}{2}^{-1} f_i(p_1, p_2) \pm \varepsilon/(2(k)_2). \end{aligned}$$

Therefore, recalling (75), we finally get that

$$(\mathfrak{A}_{\mathcal{R}}\Gamma(p))_{|\Omega_i \times \Omega_i} = f_i(p_1, p_2) \pm \varepsilon/2, \quad i \in [2],$$

and in particular

$$\sup_{p \in \mathcal{S}_2} |f(p) - \Gamma^{-1}(\mathfrak{A}_{\mathcal{R}}\Gamma(p))| \leq \varepsilon. \quad \square$$

We now construct the set where we will find a periodic trajectory. Fix values  $r = 0.1, r_s = 0.09, r_w = 0.11, \theta < 0.0001, a = 0.2, b = 0.8$  and set  $q = (a, b)$ . Recall that functions  $g, h : B(0.15, q) \setminus q \rightarrow \mathbb{R}^2$  were defined by (71) and (72). Define a function  $f : B(0.15, q) \setminus q \rightarrow \mathbb{R}^2$  by

$$f(x, y) = \theta \cdot (g(x, y) + h(x, y)).$$

Note that the dynamical system over  $B(0.15, q) \setminus q$  defined by  $f$  has a periodic trajectory  $\{(x, y) \in \mathbb{R}^2 \mid (x - a)^2 + (y - b)^2 = r^2\}$ . To finish our construction we apply Lemma 6.12 to the function  $f$ , with the constants

$$\varepsilon < \min\{|r - r_s|, |r - r_w|, \theta\} \quad \text{and} \quad \delta < 0.005,$$

and the sets  $S_1 = B(0.15, q) \setminus \overline{B(0.005, q)}$  and  $S_2 = B(0.145, q) \setminus \overline{B(0.01, q)}$ . Note that  $f$  is continuous on the closure of  $S_1$ , which is compact, implying that condition (iii) holds. Verifying the other conditions is straightforward. Hence by



Lemma 6.12 there is a rule  $\mathcal{R}$  of order  $k$  so that for the vector field  $f_{\mathcal{R}} : S_2 \rightarrow \mathbb{R}^2$  defined by  $f_{\mathcal{R}}(p) = \Gamma^{-1}(\mathfrak{A}_{\mathcal{R}}\Gamma(p))$  we have

$$(76) \quad |f(p) - f_{\mathcal{R}}(p)| \leq \varepsilon \quad \text{for all } p \in S_2.$$

Using the continuity and linearity of  $\Gamma$  and  $\Gamma^{-1}$  we see that the trajectory  $\Phi_{\mathcal{R}}\Gamma(p)$  (bijectively) corresponds to the solution  $x(t) = \Gamma^{-1}(\Phi^t\Gamma(p))$  of the dynamical system defined by  $F = f_{\mathcal{R}}$  with initial condition  $F(0) = p$ . Of course, one is periodic if and only if the other is periodic. We will show that the dynamical system defined by  $F$  has a periodic trajectory in a closed annulus using the Poincaré–Bendixson theorem, see [18, Theorem 1, p. 245].

**Theorem 6.13 (Poincaré–Bendixson).** *Suppose  $S \subseteq \mathbb{R}^2$  is an open set and a continuously differentiable function  $F : S \rightarrow \mathbb{R}^2$  defines the dynamical system*

$$(*) \quad x'(t) = F(x(t)).$$

If  $C \subseteq S$  is a compact set and

- (i) *there exists  $p \in C$  and a trajectory with  $x(0) = p$  such that  $x(t) \in C$  for every  $t \geq 0$ , and*
- (ii) *there are no points  $p \in C$  with  $F(p) = 0$ ,*

*then the system (\*) has at least one periodic trajectory inside  $C$ .*

We apply Theorem 6.13 to the function  $F = f_{\mathcal{R}}$  and the sets  $S = S_2$  and  $C = \overline{B(r_w, q)} \setminus B(r_s, q)$ . The following claim shows that our choices fulfil the conditions of Theorem 6.13, implying the existence of a periodic trajectory.

**Claim 6.14.** *The following is true for  $F = f_{\mathcal{R}}$ .*

- (i)  *$F$  points inwards at every point  $p$  on the boundary of  $C$ , that is,  $F(p)$  is not collinear with the tangent of the boundary at  $p$  and  $p + \delta F(p) \in C$  for small enough  $\delta > 0$ .*
- (ii) *If  $p$  lies on the boundary of  $C$ , that is,  $\text{circle}(r_w, q) \cup \text{circle}(r_s, q)$ , and  $x(T) = p$  for some trajectory defined by (\*) and some  $T > 0$ , then for some  $\tau > 0$  we have  $x([T, T + \tau]) \in C$ .*
- (iii) *Condition (i) of Theorem 6.13 holds (in fact, for every  $p \in C$ ).*
- (iv) *Condition (ii) of Theorem 6.13 holds.*

**Proof.** For (i), assume first that  $(x, y) = p \in \text{circle}(r_s, q)$ . The component  $g(p)$  of  $f(p)$  is tangent to  $\text{circle}(r_s, q)$  and the component  $h(p)$  points away from the point  $q$ . Since  $p \in \text{circle}(r_s, q)$ , we have  $\sqrt{(x-a)^2 + (y-b)^2} = r_s$ , therefore  $|h(p)| = |r - r_s| > \varepsilon$  and hence, in view of (76), vector  $F(p)$  points inwards. An analogous argument shows that  $F(p)$  also points inwards for every  $p \in \text{circle}(r_w, q)$ .

We now prove (ii). By (i), vector  $F(p)$  forms a non-zero angle  $\gamma \leq \pi/2$  with the tangent at  $p$  to a boundary circle and by adding to  $p$  a small multiple of  $F(p)$  we end up in  $C$ . Let  $D_\delta$  be a wedge-shaped set consisting of vectors of length at most  $\delta$  at angle at most  $\gamma/2 (\leq \pi/4)$  with  $F(p)$  (see Figure 6). For sufficiently small  $\delta$  we have that  $p + y \in C$  for every  $y \in D_\delta$ . Since the trajectory  $x$  has the derivative  $x'(t) = F(p)$ , we have

$$|x(t + \tau) - (p + \tau F(p))| = o(\tau).$$

Note that for small enough  $\tau$  (depending on  $\delta$ ) the vector  $\tau F(p)$  is  $\Omega(\tau)$ -far from the edge of  $D_\delta$ . This implies that  $x(t + \tau) - p \in D_\delta$ , for small enough  $\tau$ , and hence  $x(t + \tau) \in C$ .

Now we prove (iii). Pick an arbitrary  $p \in C$  and let  $x(t)$  be the solution of (\*) with the initial condition  $x(0) = p$ . Writing  $T = \sup\{t \geq 0 : x([0, t]) \subset C\}$ , we need to show  $T = \infty$ . Suppose the contrary, that is,  $T < \infty$ . Clearly,  $x([0, T]) \subset C$ , and  $x(T)$  lies on the boundary of  $C$  (as otherwise the trajectory would stay inside  $C$  a bit longer). By (ii) there exists  $\tau > 0$  such that  $x([0, T + \tau]) \subseteq C$  contradicting the maximality of  $T$ .

Finally, to prove (iv), fix  $p \in C$ . Since  $C \subset S_2$ ,  $g(p)$  is perpendicular to  $h(p)$ , and  $|g(p)| = 1$ , we get that

$$|F(p)| \stackrel{(76)}{\geq} |f(p)| - \varepsilon = \theta |g(p) + h(p)| - \varepsilon \geq \theta |g(p)| - \varepsilon = \theta - \varepsilon > 0,$$

which proves (iv). □

**Remark 6.15.** We stated the Poincaré–Bendixson theorem in an abridged form, sufficient for our purposes. Using the original version of the Poincaré–Bendixson theorem (see [18, Theorem 1, p. 245]) and Claim 6.14(iii), it actually follows that every trajectory starting in  $C$  is periodic or spirals towards a periodic orbit.

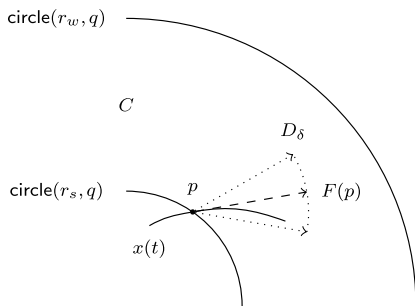


Fig. 6. The trajectory  $x(t)$  entering the set  $C$  at the point  $p$ , is approximated by the tangent  $F(p)$  and thus stays in the wedge-shaped set  $D_\delta$ .

We note that our proof of the existence of a periodic trajectory is not constructive, since the Poincaré–Bendixson theorem is not. We leave it as an open problem to find an explicit construction.

6.11. *Complicated trajectories*

Given a rule, a natural question about a rule and an initial graphon  $W$  is to understand which graphons a trajectory approaches infinitely often, namely to determine the set

$$\omega_+(W) := \{U \in \mathcal{W}_0 : \|\Phi^{t_n} W - U\|_\square \rightarrow 0 \text{ for some sequence } t_n \rightarrow \infty\}.$$

Convergent trajectories are simplest in the sense that for them  $\omega_+(W)$  contains only  $\text{dest}(W)$ . Even though the compactness of  $\widetilde{\mathcal{W}}_0$  implies that every trajectory approximates some graphon  $U$  in the cut distance  $\delta_\square$ , we still do not have an answer to the following question.

**Question 6.16.** Is it true that, for every rule  $\mathcal{R}$ , and every graphon  $W$ , set  $\omega_+(W)$  is nonempty?

We have seen examples where  $\omega_+(W)$  is a continuum set, namely a periodic trajectory from Section 6.10 and a spiraling trajectory in Remark 6.15.

The following question has been suggested to us by Joel Spencer.

**Question 6.17.** Does there exist a flip process and an initial graphon  $W$  so that the edge density oscillates between constant graphons 0 and 1, that is  $\{0, 1\} \in \omega_+(W)$ ?

What we find the most intriguing open question in this project is to find flip processes and graphons with really complicated trajectories. There is a topological way to express this (and it would be interesting to see alternative notions of complexity of trajectories). Recall that the metric  $d_\square(U, W) = \|U - W\|_\square$  induces a non-compact topology on  $\mathcal{W}_0$ . We say that the trajectory starting at  $W$  is *complicated* if  $\{\Phi^t W : t \geq 0\}$  is not totally bounded, or equivalently, the closure of its orbit, that is,  $\overline{\{\Phi^t W : t \geq 0\}}$  is not compact with respect to the metric  $d_\square$ . Then the main question is this.

**Question 6.18.** Is there a flip process and a graphon  $W$  such that  $\overline{\{\Phi^t W : t \geq 0\}}$  is not compact?

A negative answer to Question 6.16 (which would be rather surprising) would give a positive answer to Question 6.18. The proposition below excludes convergent trajectories from the search.

**Proposition 6.19.** *If a rule and a graphon  $W$  are such that  $\text{dest}(W)$  exists, then  $\overline{\{\Phi^t W : t \geq 0\}}$  is compact.*

**Proof.** Let  $\mathcal{R}$  be a rule of order  $k$  and  $W$  be the initial graphon with  $U := \text{dest}(W)$ . We need to prove total boundedness, that is, for any given  $\varepsilon > 0$  there exist finitely many  $d_\square$ -balls  $B_1, \dots, B_h$  in  $\mathcal{W}_0$  of radius  $\varepsilon$  so that  $\bigcup B_i \supset \{\Phi^t W : t \geq 0\}$ . Since  $U = \text{dest}(W)$ , there exists  $t_0 \geq 0$  such that for each  $t \geq t_0$ ,  $d_\square(\Phi^t W, U) < \varepsilon$ . Now, the appropriate balls have centers  $U$  and  $\Phi^\tau W$  where  $\tau$  ranges over multiples of  $\varepsilon/k^2$  from 0 until (it exceeds)  $t_0$ . To prove the statement of the proposition, we need to argue that, for any  $t \geq 0$ , the graphon  $\Phi^t W$  is covered by one of the balls above. If  $t \geq t_0$ , then  $\Phi^t W$  is covered by the ball around  $U$ . If  $t \in [0, t_0)$ , then  $\Phi^t W$  is covered by the ball around  $\Phi^\tau W$  for the largest  $\tau$  with  $\tau \leq t$ . To see this, recall that by (36),  $\Phi^\tau W$  and  $\Phi^t W$  at most  $\binom{k}{2}(t - \tau) \leq \varepsilon$  apart in the  $L^\infty$ -norm (and hence also in  $d_\square$ ). □

The strongest possible (positive) answer to Question 6.18 (and also to Question 6.17) would be a trajectory whose closure covers the entire graphon space. However, there is no such trajectory.

**Proposition 6.20.** *There is no flip process and a graphon  $W$  such that  $\overline{\{\Phi^t W : t \geq 0\}} = \mathcal{W}_0$ .*

**Proof.** Suppose that such a flip process exists, say with a rule  $\mathcal{R}$  of order  $k$ , and let  $W$  be a graphon whose closure of the trajectory covers  $\mathcal{W}_0$ .

Graphon  $W$  cannot have a set of twins of positive measure. Indeed, if it had, Proposition 6.5 and Lemma 3.8 would tell us that we can approximate only graphons with that set of twins, making it impossible to approximate graphons without a set of twins of positive measure. In particular,

$$(77) \quad W \text{ is not a graphon representation of any } k\text{-vertex graph.}$$

The following claim is the core of the proof.

**Claim.** *For every  $G \in \mathcal{H}_k$  we have  $\mathcal{R}_{G,G} = 1$ .*

**Proof of Claim.** Suppose that this is not the case, that is, there exists a graph  $H \in \mathcal{H}_k \setminus \{G\}$  such that  $p := \mathcal{R}_{G,H} > 0$ . Fix a pair  $ij \in E(G) \ominus E(H)$ . Let us assume that (\*)  $ij \notin E(G)$  and  $ij \in E(H)$ . The argument in the case (\*\*)  $ij \in E(G)$  and  $ij \notin E(H)$  is symmetric and we will not include it. However, to clarify the adjustments needed and also for its own sake, let us give an overview of the proof of (\*). We shall consider a graphon representation  $W_G$  of  $G$  with respect to a fixed partition  $(\Omega_v)_{v \in [k]}$ . Our focus will be on the square  $\Omega_i \times \Omega_j$  where  $W_G$  is constant 0. The core argument (expressed in (79)) is that if the trajectory  $U = \Phi^t W$  gets somewhat close to  $W_G$  then necessarily  $\int_{\Omega_i \times \Omega_j} \mathfrak{W} U$  is positive (with an explicit positive lower bound) thus repelling from  $W_G$  on  $\Omega_i \times \Omega_j$ . Then inapproximability of  $W_G$  follows. Obviously, if we have (\*\*) instead of (\*) then the counterpart of (79) is a negative upper bound on  $\int_{\Omega_i \times \Omega_j} \mathfrak{W} U$  showing that an ambition of having a trajectory approaching  $W_G$  must fail due to  $(W_G)|_{\Omega_i \times \Omega_j} = 1$ .

Let  $W_G$  be the graphon representation of  $G$  with respect to an arbitrary fixed partition  $(\Omega_v)_{v \in [k]}$ . By (77) we have that  $\Phi^0 W \neq W_G$ . Also, as  $W_G$  is  $\{0, 1\}$ -valued, we cannot have  $\Phi^t W = W_G$  for any  $t > 0$  by Proposition 6.7. In particular this means that there are times  $t'_1 < t_1 < t'_2 < t_2 < \dots$  such that

$$(78) \quad d_{\square}(W_G, \Phi^{t'_\ell} W) > \frac{1}{4}, \quad d_{\square}(W_G, \Phi^{t_\ell} W) < \frac{1}{\ell}.$$

To see this, note that there are two  $\{0, 1\}$ -valued graphons with distance more than  $1/2$  between them (say, representation of complete and empty graphs); hence at least one of them,  $F$ , is such that  $d_{\square}(W_F, W_G) > 1/4$ ; since  $W_F$  also does not lie on the trajectory, the trajectory has to alternate between small neighbourhoods of  $W_G$  and  $W_F$ .

Let  $C_{\square} = C_{\square}(k, 0)$  be given by Lemma 4.10. Set  $\delta := pk^{-k}/(2C_{\square})$ . We claim that for every graphon  $U$  with  $d_{\square}(U, W_G) \leq \delta$  we have

$$(79) \quad \int_{\Omega_i \times \Omega_j} \mathfrak{W} U \geq \frac{pk^{-k}}{2}.$$

To this end, first we compute  $\mathfrak{W} W_G$ , focusing on the value on the square  $\Omega_i \times \Omega_j$  (on which  $\mathfrak{W} W_G$  is constant 0). So, let us fix  $x \in \Omega_i, y \in \Omega_j$  and look at the summands in the representation (26). Since  $W_G(x, y) = 0$ , each term in (26) is nonnegative. Keeping the term for  $ab = ij$ , we get (recall that  $\mathbf{U}_v$ s are independent  $\pi$ -random variables)

$$\begin{aligned} \mathfrak{W} W_G(x, y) &\geq \mathbb{P}(\mathbb{G}(k, W_G) = G \mid \mathbf{U}_i = x, \mathbf{U}_j = y) \mathcal{R}_{G,H} \\ &\geq \mathbb{P}(\mathbf{U}_v \in \Omega_v, \forall v \in [k]) \mathcal{R}_{G,H} = k^{-k} p. \end{aligned}$$

By Lemma 4.10,  $\int_{\Omega_i \times \Omega_j} \mathfrak{W} U \geq \int_{\Omega_i \times \Omega_j} W_G - C_{\square} d_{\square}(U, W_G) \geq pk^{-k}/2$ . This proves (79).

Let us show how (79) leads to a contradiction, thus proving the Claim. Letting  $\tau_\ell$  be the last time (see (78)) smaller than  $t_\ell$  such that  $d_{\square}(W_G, \Phi^{\tau} W) \geq \delta$ , we have that

$$d_{\square}(\Phi^{\tau_\ell} W, \Phi^{t_\ell} W) \geq \delta - \frac{1}{\ell},$$

and hence (36) (taking into account that  $\|\cdot\|_{\square} \leq \|\cdot\|_{\infty}$ ) implies that  $t_{\ell} - \tau_{\ell} \geq (\delta - 1/\ell)/(k)_2$ . Starting with the integral form for the trajectory (see Remark 4.9),  $\Phi^{t_{\ell}} W = \Phi^{\tau_{\ell}} W + \int_{\tau_{\ell}}^{t_{\ell}} \mathfrak{V} \Phi^t W dt$  and making use of the fact that for each  $t \in [\tau_{\ell}, t_{\ell}]$  we can apply (79) with  $U = \Phi^t W$ , we get

$$(80) \quad \begin{aligned} \int_{\Omega_i \times \Omega_j} \Phi^{t_{\ell}} W &= \int_{\Omega_i \times \Omega_j} (\Phi^{\tau_{\ell}} W) + \int_{\Omega_i \times \Omega_j} \left( \int_{\tau_{\ell}}^{t_{\ell}} \mathfrak{V} \Phi^t W dt \right) \\ &\geq 0 + \frac{\delta - \frac{1}{\ell}}{(k)_2} \cdot \frac{pk^{-k}}{2}. \end{aligned}$$

To justify the last inequality, note that by definition of the integral (see Appendix C) we can approximate (in  $L^{\infty}$ -norm and thus in the cut norm)  $\int_{\tau_{\ell}}^{t_{\ell}} \mathfrak{V} \Phi^t W dt$  by a linear combination of kernels taking values in  $\{\mathfrak{V} \Phi^t W : t \in [\tau_{\ell}, t_{\ell}]\}$  and applying (79) to each of them.

By choosing  $\ell$  large enough we get that  $\int_{\Omega_i \times \Omega_j} \Phi^{t_{\ell}} W > \frac{1}{\ell}$ . On the other hand,  $ij \notin E(G)$  implies  $\int_{\Omega_i \times \Omega_j} W_G = 0$ . We conclude that

$$d_{\square}(\Phi^{t_{\ell}} W, W_G) \geq \int_{\Omega_i \times \Omega_j} (\Phi^{t_{\ell}} W - W_G) > 1/\ell,$$

contradicting (78). □

The Claim indeed proves the main statement, since by applying it to all graphs  $G \in \mathcal{H}_k$  we deduce that the flip process in question is trivial. This is clearly a contradiction. □

**Remark 6.21.** A strengthening of Proposition 6.20 can be proved for the factor space  $\widetilde{\mathcal{W}}_0$  (recall Remark 3.1). Namely, there is no trajectory whose closure in the cut distance covers  $\widetilde{\mathcal{W}}_0$ . This follows by making minor changes to the proof of Proposition 6.20: in (78) we allow the graphon representation  $W_G$  to depend on  $\ell$ .

**Remark 6.22.** In this section, we looked at the closure of the *forward* orbit  $\overline{\{\Phi^t W : t \geq 0\}}$ . As an alternative, we could look at the closure of the portion of the “whole” orbit  $\{\Phi^t W : t \in I_W\}$ . Obviously, the qualitative difference between these two images can be only in the case  $\text{age}(W) = \infty$ . Question 6.18 is also open in this alternative setting as is the question about non-emptiness of the set

$$\omega_{-}(W) := \{U \in \mathcal{W}_0 : \|\Phi^{t_n} W - U\|_{\square} \rightarrow 0 \text{ for some sequence } t_n \rightarrow -\infty\}.$$

Nevertheless, a counterpart to Proposition 6.19 holds if we assume, in addition, that  $\Phi^t W$  converges as  $t \rightarrow -\infty$ . Also, Proposition 6.20 holds even in this stronger setting. Indeed, the repulsion argument used to show the inapproximability of  $W_G$  holds uniformly even at negative times.

### 6.12. Speed of convergence

For now let us focus on convergent flip processes. For each such corresponding rule  $\mathcal{R}$ , we can define for each  $\delta > 0$  *lower convergence time*

$$\tau_{\mathcal{R}}^{-}(\delta) = \sup\{\inf\{T : d_{\square}(\Phi^T W, \text{dest}_{\mathcal{R}}(W)) < \delta\} : W \in \mathcal{W}_0\},$$

and the *upper convergence time*,

$$\tau_{\mathcal{R}}^{+}(\delta) = \sup\{\sup\{T : d_{\square}(\Phi^T W, \text{dest}_{\mathcal{R}}(W)) \geq \delta\} : W \in \mathcal{W}_0\}.$$

**Remark 6.23.** The definition of the cut norm as

$$\|W\|_{\square} = \sup_{A, B \subseteq \Omega} \left| \int_{A \times B} W d\pi^2 \right|$$

is not the only natural one, as sometimes it makes more sense to work only with

$$\|W\|_{\square} = \sup_{A \subseteq \Omega} \left| \int_{A \times A} W d\pi^2 \right| \quad \text{or} \quad \|W\|_{\square} = \sup_{A, B \subseteq \Omega: A \cap B = \emptyset} \left| \int_{A \times B} W d\pi^2 \right|.$$

As is well known (see, e.g., [8]), these alternative definitions are off from the original one by a factor of at most 4. Thus, when pinning down the speed of convergence precisely, it matters which of these definitions is used.

**Remark 6.24.** In view of Conjecture 6.4, we might also define the speed using the  $L^1$ -distance.

Below, we show that  $\tau_{\mathcal{R}}^-(\delta)$  is finite for flip processes with a single destination.

**Proposition 6.25.** *If  $\mathcal{R}$  is a convergent flip processes with a single destination, then for every  $\delta \in (0, 1]$  we have  $\tau_{\mathcal{R}}^-(\delta) < \infty$ .*

**Proof.** Let  $U$  be the unique destination (and thus the unique fixed point). Hence by Proposition 6.10 graphon  $U$  is constant.

Suppose for contradiction that the statement does not hold. Then for each  $\ell \in \mathbb{N}$  we have a graphon  $W_\ell$  such that  $\inf\{T : d_{\square}(\Phi^T W_\ell, U) < \delta\} > \ell$ . By the Lovász–Szegeedy compactness theorem, there exists a graphon  $Z$  which is an accumulation point of  $W_1, W_2, \dots$  in the cut distance. Without loss of generality, let us assume that  $Z$  is actually a cut distance limit of  $W_1, W_2, \dots$ . Since the trajectory of  $Z$  converges to  $U$ , there exists a time  $T$  such that  $d_{\square}(\Phi^T Z, U) < \delta/2$ . Combining this with Theorem 4.16, for any  $W$  sufficiently close to  $Z$  in the cut norm, we have  $d_{\square}(\Phi^T W, U) < \delta/2$ . Since  $U$  is a constant graphon, by Section 6.3, the same conclusion holds for any  $W$  sufficiently close to  $Z$  in the cut distance. Taking  $W = W_\ell$  for sufficiently large  $\ell$  gives a contradiction.  $\square$

The same question is open for the quantity  $\tau_{\mathcal{R}}^+(\delta)$ .

Sometimes the convergence time is infinite for trivial reasons. For example if, given a rule  $\mathcal{R}$ , there is a graphon  $W$  with  $\text{age}(W) = \infty$  such that  $\lim_{t \rightarrow -\infty} \Phi^t W$  exists and  $\mathfrak{A} W \neq 0$ , then one can make the convergence time arbitrarily long by choosing the initial graphon as  $\Phi^{-t} W$  with  $t$  large enough.

**Question 6.26.** Suppose that  $\mathcal{R}$  is a convergent flip processes in which every graphon is either a fixed point or has finite age. Is it then true that for each  $\delta \in (0, 1]$  we have  $\tau_{\mathcal{R}}^-(\delta) < \infty$ ? Is the same conclusion true if, in addition,  $\mathcal{R}$  has only finitely many destinations?

We can go back to our motivating examples. Indeed, (1) tells us that for the Erdős–Rényi flip process (whose only destination is the constant-1 graphon) we have  $\tau^+(\delta) = \ln(-\delta)/2$ . The triangle removal flip process (generalised in Section 2.3) is much more interesting. In particular, in [2] a uniform bound on convergence time is derived using the Szemerédi’s regularity lemma, thus giving extremely poor bounds.

It can be shown that for any flip process  $\mathcal{R}$  with a single destination we have a lower bound  $\tau^+(\delta) = \Omega(\ln(-\delta))$ . Let us discuss the possibility of any general upper bounds. Note that there is no ‘slowest process’. Indeed, we can slow down any rule  $\mathcal{R}$  by any factor  $C > 1$ . That is, we define a new rule  $\mathcal{R}'$  which, for any drawn graph  $H$ , replaces it with  $H$  with probability  $1 - 1/C$  and replaces it with a graph sampled according to  $\mathcal{R}$  with probability  $1/C$ . Still, this slowing down does not change the order of magnitude of the function  $\tau_{\mathcal{R}}^-(\delta)$  as  $\delta \rightarrow 0$ . Hence, we could ask what is the slowest order of magnitude of convergence.

### 6.12.1. Variants

Many variants of these questions exist. For example, one might consider local versions of convergence times, such as follows. Suppose that we have a convergent flip process with a stable destination  $U$  of radius  $\gamma^*$ . Then define

$$\tau_{\mathcal{R}}^-(\delta, \gamma) = \sup\{\inf\{T : d_{\square}(\Phi^T W, U) < \delta\} : W \in \mathcal{W}_0, d_{\square}(W, U) < \gamma\},$$

for  $\gamma^* > \gamma > \delta > 0$ .

### 6.13. Uniqueness of the rule

In an upcoming preprint, Hng [14] provides a complete characterisation of equivalence classes of flip process rules (of the same order) with the same trajectories. As an application, Hng demonstrates that the symmetric deterministic rules<sup>8</sup> and the rules of order 2 are precisely the rules which are unique in their equivalence classes. The class of symmetric deterministic rules includes several examples from Section 2, namely the complementing rules, the component completion rules, the extremist rules, and the clique removal rules.

<sup>8</sup>A rule  $\mathcal{R}$  of order  $k$  is *symmetric* if we have for each permutation  $\psi$  on  $[k]$  (which naturally acts on  $\mathcal{H}_k$ ) and each  $H, J \in \mathcal{H}_k$  that  $\mathcal{R}_{H,J} = \mathcal{R}_{\psi(H),\psi(J)}$ . A rule  $\mathcal{R}$  is *deterministic* if for each  $H \in \mathcal{H}_k$  we have that  $\mathcal{R}_{H,J} = 0$  for all except one  $J \in \mathcal{H}_k$ .

6.14. Large deviations

Our Theorem 5.1 provides a description of the typical evolution for flip processes. It would be interesting to work out a large deviation theory for flip processes as well. For flip processes of order 2 such a large deviation theory was worked out in [9], but in a variant with updates along a Poisson point process (cf. Section 5.2).<sup>9</sup>

**Appendix A: Probability**

We recall the Hoeffding–Azuma martingale concentration inequality including necessary preliminaries. To simplify some definitions we assume that the sample space  $\Omega$  of the underlying probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  is finite. Consider a sequence of sigma-algebras  $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_t$ , such that  $\mathcal{F}_i \subseteq \mathcal{F}$ , which we call a *filtration*. We say that a sequence of random variables  $Y_1, \dots, Y_t$  is a *martingale difference sequence* if  $Y_i$  is  $\mathcal{F}_i$ -measurable and  $\mathbb{E}(Y_i | \mathcal{F}_{i-1}) = 0$  for  $i = 1, \dots, t$ .

The following lemma is a special case of Theorem 3.10 in [17].

**Lemma A.1.** *Let  $c$  be a constant. If  $Y_1, \dots, Y_t$  is a martingale difference sequence with  $|Y_i| \leq c$  for each  $i$ , then for any  $x \geq 0$ ,*

$$\mathbb{P}(|Y_1 + \dots + Y_t| \geq x) \leq \exp\left(-\frac{x^2}{2tc^2}\right).$$

**Appendix B: Graphons**

Here we include some more facts about graphons and the proofs for Section 3.2.

The next two facts give alternative ways of bounding the cut norm. Variants appear in the literature, but we provide proofs for completeness.

**Fact B.1 (Inequality (7.1) in [8]).** For each  $W \in \mathcal{W}$ , we have  $\|W\|_{\square} \leq 2 \sup_{A \subseteq \Omega} |\int_{A \times A} W \, d\pi^2|$ .

**Proof.** To bound  $\|W\|_{\square}$ , let  $A$  and  $B$  be arbitrary as in (6). Using symmetry of  $W$ ,

$$\begin{aligned} \left| \int_{A \times B} W \right| &= \frac{1}{2} \left| \int_{A \times B} W + \int_{B \times A} W \right| = \frac{1}{2} \left| \int_{(A \cup B)^2} W + \int_{(A \cap B)^2} W - \int_{(A \setminus B)^2} W - \int_{(B \setminus A)^2} W \right| \\ &\leq \frac{1}{2} \cdot 4 \sup_{C \subseteq \Omega} \left| \int_{C \times C} W \right|. \end{aligned} \quad \square$$

**Fact B.2.** Suppose that  $\mathcal{A}$  is a partition of  $\Omega$  into finitely many sets. For each  $W \in \mathcal{W}$  which is constant on each cell of  $\mathcal{A} \times \mathcal{A}$  we have

$$(81) \quad \|W\|_{\square} \leq \sup_{S, T \subseteq \mathcal{A}} \left| \int_{S \times \cup T} W \, d\pi^2 \right|.$$

**Proof.** To bound  $\|W\|_{\square}$ , let  $A$  and  $B$  be arbitrary as in (6). We shall prove that for an arbitrary  $D \in \mathcal{A}$  we have

$$(82) \quad \left| \int_{A \times B} W \right| \leq \max \left\{ \left| \int_{(A \cup D) \times B} W \right|, \left| \int_{(A \setminus D) \times B} W \right| \right\}.$$

A similar ‘rounding’ statement could be obtained for  $B$  as well. Once this is done, (81) follows by rounding each of  $A$  and  $B$  on each cell of  $\mathcal{A}$ .

Let  $\alpha := \pi(A \cap D) / \pi(D)$ . Since  $x \mapsto \int_{y \in B} W(x, y)$  is constant on  $D$ , we have

$$\int_{A \times B} W = \alpha \cdot \int_{(A \cup D) \times B} W + (1 - \alpha) \cdot \int_{(A \setminus D) \times B} W.$$

Since  $\alpha \in [0, 1]$ , inequality (82) follows from the fact that  $x \mapsto |x|$  is a convex function. □

<sup>9</sup>As we explain in Section 5.2 there is essentially no difference between updates at integer times and along a Poisson point process for the typical behaviour. However, there usually is a difference between these two regimes from the point of large deviations.

Here, we give a self-contained proof of Lemma 3.2. This proof was suggested to us by Jan Grebík and Guus Regts.

**Proof of Lemma 3.2.** The proof shall use the concept of weak\* convergence of graphons (which we treat as elements of  $L^\infty(\Omega^2)$  with a predual space  $L^1(\Omega^2)$ ), which we quickly recall. We refer the reader to [13] for details. A sequence of graphons  $W_1, W_2, \dots$  converges weak\* to a graphon  $W$  if for every  $S, T \subset \Omega$  and every  $\varepsilon > 0$  there exists a number  $N_{S,T,\varepsilon} \in \mathbb{N}$  such that

$$(83) \quad \left| \int_{S \times T} W_n - \int_{S \times T} W \right| < \varepsilon \quad \text{for every } n \geq N_{S,T,\varepsilon}.$$

So, the cut norm topology can be regarded as a certain uniformisation of the weak\* topology, and indeed convergence in the former topology implies convergence in the latter. It follows from the Banach–Alaoglu theorem that the closed unit ball in  $L^\infty(\Omega^2)$  is compact with respect to the weak\* topology. Since the set of graphons  $\mathcal{W}_0$  is contained in this unit ball and is closed with respect to the  $L^\infty$  norm (and therefore closed in the weak\* topology),  $\mathcal{W}_0$  is compact in the weak\* topology, too.

We are now ready to prove the lemma. Suppose that  $W_1, W_2, \dots$  is a Cauchy sequence in  $d_\square$ . That is, for every  $\varepsilon > 0$  there exists an  $M_\varepsilon \in \mathbb{N}$  such that

$$(84) \quad \sup_{S,T \subset \Omega} \left| \int_{S \times T} W_n - \int_{S \times T} W_m \right| < \varepsilon \quad \text{for every } n, m \geq M_\varepsilon.$$

From compactness of  $\mathcal{W}_0$  and (84) it easily follows that  $(W_n)$  converges weak\* to some graphon  $W$ . We claim that  $W$  is the sought cut norm limit. Suppose that  $\varepsilon > 0$  is arbitrary. We then have that for  $n \geq M_\varepsilon$

$$\begin{aligned} d_\square(W_n, W) &= \sup_{S,T \subset \Omega} \left| \int_{S \times T} W_n - \int_{S \times T} W \right| \\ &\leq \sup_{S,T \subset \Omega} \left\{ \left| \int_{S \times T} W_n - \int_{S \times T} W_{\max\{M_\varepsilon, N_{S,T,\varepsilon}\}} \right| \right. \\ &\quad \left. + \left| \int_{S \times T} W_{\max\{M_\varepsilon, N_{S,T,\varepsilon}\}} - \int_{S \times T} W \right| \right\}, \end{aligned}$$

where the terms  $N_{S,T,\varepsilon}$  are defined as in (83). From (83) and (84) it follows that  $d_\square(W_n, W) \leq \varepsilon + \varepsilon$ , which implies  $\lim_{n \rightarrow \infty} d_\square(W_n, W) = 0$ , as was needed.  $\square$

Next, we turn to proving Lemma 3.5. We will make use of the following fact.

**Fact B.3.** Given a real number  $M > 0$ , any numbers  $a_1, \dots, a_\ell, b_1, \dots, b_\ell$  of modulus at most  $M$  satisfy the inequality

$$\left| \prod_i a_i - \prod_i b_i \right| \leq M^{\ell-1} \sum_i |a_i - b_i|.$$

**Proof.** To apply induction on  $\ell$ , write  $A = \prod_{i \leq \ell-1} a_i$ ,  $B = \prod_{i \leq \ell-1} b_i$ , so that the left-hand-side is  $|Aa_\ell - Bb_\ell|$ . We have  $Aa_\ell - Bb_\ell = A(a_\ell - b_\ell) + (A - B)b_\ell$ . Hence, by the triangle inequality and the induction hypothesis

$$\begin{aligned} |Aa_\ell - Bb_\ell| &\leq |A||a_\ell - b_\ell| + |A - B||b_\ell| \\ &\leq M^{\ell-1}|a_\ell - b_\ell| + |A - B|M \\ &\leq M^{\ell-1} \sum_{i=1}^\ell |a_i - b_i|, \end{aligned}$$

as was needed.  $\square$

**Proof of Lemma 3.5.** This follows by fixing  $(x, y)$ , writing  $(T_F^{a,b}U)(x, y)$  and  $(T_F^{a,b}W)(x, y)$  as integrals (see (10)) with the same variables of integration, applying the inequality  $|\int f| \leq \int |f|$  and then applying Fact B.3 to the integrand. (Note that  $W(x_i, x_j)$  and  $1 - W(x_i, x_j)$  have modulus at most  $1 + \varepsilon$  because of  $W \in \mathcal{W}_0^\varepsilon$ .)  $\square$

We want to move on to prove Lemma 3.7. For this we need to cite a counting lemma from [16, Lemma 10.24].



**Lemma B.4 (Counting lemma for vectors of graphons).** *Given a simple graph  $F = (V, E)$  and two tuples  $\mathbf{U} = (U_e : e \in E)$  and  $\mathbf{W} = (W_e : e \in E)$  of graphons, we have*

$$|t(F, \mathbf{U}) - t(F, \mathbf{W})| \leq \sum_{e \in E} \|U_e - W_e\|_{\square}.$$

**Proof of Lemma 3.7.** Every  $W \in \mathcal{W}_0^{\varepsilon}$  can be written as  $W = -\varepsilon + (1 + 2\varepsilon)\hat{W} = (1 + \varepsilon)\hat{W} - \varepsilon(1 - \hat{W})$  for some graphon  $\hat{W} \in \mathcal{W}_0$ . We have, always assuming that the arguments of  $W_{ij}$  and  $\hat{W}_{ij}$  are  $(x_i, x_j)$ ,

$$\begin{aligned} t(F, \mathbf{W}) &= \int \prod_{e \in E} W_e \, d\pi^V \\ &= \int \prod_{e \in E} ((1 + \varepsilon)\hat{W}_e - \varepsilon(1 - \hat{W}_e)) \, d\pi^V \\ &= \sum_{S \subseteq E} (1 + \varepsilon)^{|S|} (-\varepsilon)^{|E|-|S|} \int \prod_{e \in S} \hat{W}_e \prod_{e \in E \setminus S} (1 - \hat{W}_e) \, d\pi^V \\ &= \sum_{S \subseteq E} (1 + \varepsilon)^{|S|} (-\varepsilon)^{|E|-|S|} t(F, \hat{\mathbf{W}}_S), \end{aligned}$$

where we defined  $\hat{\mathbf{W}}_S = (\hat{W}_{S,e} : e \in E)$  by setting  $\hat{W}_{S,e} = \hat{W}_e$  if  $e \in S$  and  $\hat{W}_{S,e} = 1 - \hat{W}_e$  otherwise.

Analogously defining  $\hat{\mathbf{U}}_S$  and expressing  $t(F, \mathbf{U})$ , it follows that

$$\begin{aligned} |t(F, \mathbf{U}) - t(F, \mathbf{W})| &\leq \sum_{S \subseteq E} (1 + \varepsilon)^{|S|} \varepsilon^{|E|-|S|} |t(F, \hat{\mathbf{U}}_S) - t(F, \hat{\mathbf{W}}_S)| \\ &\stackrel{\text{Lemma B.4}}{\leq} \sum_{S \subseteq E} (1 + \varepsilon)^{|S|} \varepsilon^{|E|-|S|} \sum_{e \in E} \|\hat{U}_{S,e} - \hat{W}_{S,e}\|_{\square} \\ &\stackrel{\hat{U}_{S,e} - \hat{W}_{S,e} = (U_e - W_e)/(1 + 2\varepsilon)}{=} \sum_{S \subseteq E} (1 + \varepsilon)^{|S|} \varepsilon^{|E|-|S|} \cdot \frac{\sum_{e \in E} \|U_e - W_e\|_{\square}}{1 + 2\varepsilon} \\ &\stackrel{\text{binomial theorem}}{=} (1 + 2\varepsilon)^{|E|-1} \sum_{e \in E} \|U_e - W_e\|_{\square}. \quad \square \end{aligned}$$

Lastly, we add the proof of Lemma 3.8.

**Proof of Lemma 3.8.** Fix a measurable  $T \subset \Omega$  and let  $f_n(x) = \int_T U_n(x, \cdot) \, d\pi$  and  $f(x) = \int_T U(x, \cdot) \, d\pi$ . Given a set of positive measure  $B$  and  $g \in L^{\infty}(\Omega)$ , let us write  $\mathbb{E}_B g = \frac{1}{\pi(B)} \int_B g \, d\pi$ . Since  $A$  is a twin-set for  $U_n$ , for every  $B \subset A$  of positive measure we have

$$\mathbb{E}_B f_n = \frac{1}{\pi(B)} \int_{B \times T} U_n = \frac{1}{\pi(A)} \int_{A \times T} U_n = \mathbb{E}_A f_n.$$

Since  $U_n \rightarrow U$  in the cut norm, this implies that

$$(85) \quad \mathbb{E}_B f = \frac{1}{\pi(B)} \int_{B \times T} U = \frac{1}{\pi(A)} \int_{A \times T} U = \mathbb{E}_A f.$$

We claim that  $f$  is constant a.e. on  $A$ . Let  $\varepsilon > 0$  be arbitrary. Let  $B = \{x \in A : f(x) \geq \varepsilon + \mathbb{E}_A f\}$ . If  $\pi(B) > 0$ , then  $\mathbb{E}_B f \geq \varepsilon + \mathbb{E}_A f$ , contradicting (85). Since  $\varepsilon$  was arbitrary,  $f \leq \mathbb{E}_A f$  a.e., and by a similar argument  $f \geq \mathbb{E}_A f$ .

Recalling that the sigma-algebra of  $\Omega$  is separable, let  $\mathcal{C}$  be a countable set of measurable sets which approximate any measurable  $S \subset \Omega$ . Writing  $f_T(x) = \int_T U(x, \cdot) \, d\pi$ , by what we showed above, for every  $T \in \mathcal{C}$  we have  $f_T = c_T := \mathbb{E}_A f_T$  a.e. Hence the set  $A' := \{x \in A : \forall T \in \mathcal{C} \, f_T(x) = c_T\}$  is a conull subset of  $A$ .

Fix  $x, y \in A'$  and  $\varepsilon > 0$  and let  $B_{\varepsilon} = \{z \in \Omega : U(x, z) - U(y, z) \geq \varepsilon\}$ . For any  $T \in \mathcal{C}$  we have

$$\varepsilon \pi(B_{\varepsilon}) \leq \int_{B_{\varepsilon}} (U(x, \cdot) - U(y, \cdot)) \, d\pi \leq \pi(B_{\varepsilon} \cap T) \cdot 2\|U\|_{\infty} + \int_T (U(x, \cdot) - U(y, \cdot)) \, d\pi.$$

Since  $\int_T (U(x, \cdot) - U(y, \cdot)) = f_n(x) - f_n(y) = 0$  and  $\pi(B_\varepsilon \ominus T)$  can be made arbitrarily small, we get that  $\pi(B_\varepsilon) = 0$ , that is,  $U(x, \cdot) \leq U(y, \cdot) + \varepsilon$  a.e. Since  $x, y, \varepsilon$  were arbitrary, this implies that  $A$  is a twin-set for  $U$ .  $\square$

### Appendix C: Calculus of normed-space-valued functions

Let  $\mathcal{E}$  be a Banach space with norm  $\|\cdot\|$ .

**Definition C.1.** Suppose that a function  $f : I \rightarrow \mathcal{E}$  is defined on some subset  $I \subset \mathbb{R}$ . If there is an element  $z \in \mathcal{E}$  such that

$$\lim_{I \ni \varepsilon \rightarrow 0} \left\| \frac{f(t + \varepsilon) - f(t)}{\varepsilon} - z \right\| = 0,$$

then we say that  $f$  is differentiable at  $t$  and call  $\frac{df}{dt}(t) = f'(t) = z$  the derivative of  $f$  at  $t$ .

For any  $a, b \in \mathbb{R}$ , let  $\mathcal{S}([a, b], \mathcal{E})$  be the set of step functions  $f : [a, b] \rightarrow \mathcal{E}$ , that is, functions such that  $f$  is fixed on each set of some finite partition of  $[a, b]$  into intervals  $(P_1, \dots, P_n)$ . For a step function  $f$ , the integral  $\int_a^b f$  is an element of  $\mathcal{E}$  defined by

$$\int_a^b f = \sum_i |P_i| f_i, \quad \text{where } f(t) = f_i \text{ on } P_i.$$

Considering the space of bounded functions from  $[a, b]$  to  $\mathcal{E}$  with uniform norm  $\|f\|_\infty := \sup_{t \in [a, b]} \|f(t)\|$ , it is easy to show that any continuous function  $f : [a, b] \rightarrow \mathcal{E}$  can be approximated by step functions (see [1, p. 54]), that is, there is a sequence  $f_n \in \mathcal{S}([a, b], \mathcal{E})$  such that  $\lim_{n \rightarrow \infty} \|f - f_n\|_\infty = 0$ . Moreover, for any such approximating sequence the elements  $\int_a^b f_n$  converge to the same element of  $\mathcal{E}$  which we denote by  $\int_a^b f$  and call the Cauchy–Bochner integral of  $f$  over  $[a, b]$ . This integral satisfies the following properties

$$(86) \quad \left\| \int_a^b f \right\| \leq \int_a^b \|f\| \leq |b - a| \|f\|_\infty,$$

$$(87) \quad \int_a^b f = \int_a^c f + \int_c^b f \quad \text{with a convention that } \int_a^b f := - \int_b^a f.$$

For proofs below, whenever we approximate a continuous function  $f$  by a step function  $f^{(n)}$ ,  $n \in \mathbb{N}$ , we let  $[a, b] = P_1 \cup \dots \cup P_n$  be a partition into intervals of equal lengths and assume that  $f^{(n)}$  on  $P_i$  takes some (constant) value which  $f$  takes at some point in  $P_i$ . We may occasionally choose this value not arbitrarily; say if  $f(t)$  has some property for almost every  $t$ , then we may want  $f^{(n)}$  to inherit this property and to achieve this we will choose the value in  $P_i$  avoiding the exceptional measure-zero set for  $t$ .

Since  $f$  is continuous on a compact interval, it is also uniformly continuous. Hence the sequence  $(f^{(n)})_n$  converges to  $f$  in the uniform norm implying that  $(\int f^{(n)})_n$  converges to  $\int f$ .

We now state the fundamental theorem of calculus for normed space-valued functions.

**Theorem C.2 (Theorem 2.3.11 in [1]).** *If  $f : [a, b] \rightarrow \mathcal{E}$  is continuous, differentiable on  $(a, b)$  and if  $f'$  extends to a continuous function on  $[a, b]$ , then*

$$f(b) - f(a) = \int_a^b f'.$$

Further, given  $f, g \in L^\infty(\Omega^2)$  we write  $f \geq g$  meaning  $f \geq g$  a.e.

**Proposition C.3.** *Given a continuous function  $f : [a, b] \rightarrow L^\infty(\Omega^2)$  such that for every  $t$  we have  $f(t) \geq 0$ , we have that*

$$\int_a^b f(t) dt \geq 0.$$

**Proof.** Let  $f_n$  be the explicit step functions defined above so that

$$I_n := \int_a^b f_n(t) dt \xrightarrow{\infty} \int_a^b f(t) dt =: I.$$

Since  $f_n$  takes values in the range of  $f$ , for every  $t$  we have that  $f_n(t) \geq 0$  and therefore  $I_n \geq 0$ . Assume for contradiction that  $I \leq -\varepsilon$  on a set  $A$  of measure  $\varepsilon > 0$ . Then

$$-\varepsilon^2 \geq \int_A I = \int_A I_n + \int_A (I_n - I) \geq 0 - \|I_n - I\|_1 \geq -\|I_n - I\|_\infty \rightarrow 0,$$

giving a contradiction. □

The following fact says that for  $L^\infty$ -valued integrals one can calculate the integral pointwise.

**Proposition C.4.** *Let  $\mathcal{E}$  be the normed space  $L^\infty(\Omega^2)$ . Consider a continuous function  $[a, b] \ni t \mapsto f_t \in \mathcal{E}$ . One can modify, for every  $t \in [a, b]$ ,  $f_t$  on a set of measure zero (that possibly depends on  $t$ ) so that for every  $(x, y) \in \Omega^2$  function  $t \mapsto f_t(x, y)$  is continuous (and thus integrable) and the function  $J \in \mathcal{E}$  defined by  $J(x, y) = \int_a^b [f_t(x, y)] dt$  represents the same element of  $\mathcal{E}$  as  $I := \int_a^b f_t dt$ .*

*Furthermore, if we have continuous functions  $t \mapsto f_t$  and  $t \mapsto g_t$  such that  $f_t \geq g_t$  for every  $t \in [a, b]$ , then each of them can be modified to satisfy the conclusion above and also  $f_t(x, y) \geq g_t(x, y)$  for every  $(x, y) \in \Omega^2$  and  $t \in [a, b]$ .*

**Proof.** Since  $[a, b]$  is compact,  $f$  is uniformly continuous, and so there is a function  $\omega$  such that  $\omega(x) \rightarrow 0, x \rightarrow 0$ , for which

$$(88) \quad \|f_{t_1} - f_{t_2}\|_\infty \leq \omega(|t_1 - t_2|) \quad \text{for every } t_1, t_2 \in [a, b].$$

Let us choose  $D_1 \subset \Omega^2$  of full measure such that

$$(89) \quad \sup_{(x,y) \in D_1} |f_{t_1}(x, y) - f_{t_2}(x, y)| \leq \omega(|t_1 - t_2|) \quad \text{for every rational } t_1, t_2 \in [a, b].$$

For  $(x, y) \in D_1$ , the function  $t \mapsto f_t(x, y)$  is continuous on  $[a, b] \cap \mathbb{Q}$ , so we can extend it to a continuous function  $t \mapsto \tilde{f}_t(x, y)$  defined on the whole interval  $[a, b]$ . For  $(x, y) \in \Omega^2 \setminus D_1$  and  $t \in [a, b]$ , we define  $\tilde{f}_t(x, y) = 0$ , say. We claim that  $\|f_t - \tilde{f}_t\|_\infty = 0$  for every  $t \in [a, b]$ . By the triangle inequality, for any rational  $\tau \in [a, b]$ ,

$$\begin{aligned} \|f_t - \tilde{f}_t\|_\infty &\leq \|f_t - f_\tau\|_\infty + \|f_\tau - \tilde{f}_\tau\|_\infty + \|\tilde{f}_\tau - \tilde{f}_t\|_\infty \\ &\leq |t - \tau| + 0 + |\tau - t| = 2|t - \tau|. \end{aligned}$$

By sending  $\tau$  to  $t$  the claim follows, implying that for every  $t \in [a, b]$ ,  $f_t = \tilde{f}_t$  almost everywhere.

We have  $\int f_t dt = \int \tilde{f}_t dt$ , since  $f_t$  and  $\tilde{f}_t$  are the same element of  $\mathcal{E}$ . We declare  $\tilde{f}_t$  to be the claimed modification of  $f_t$ , for every  $t$ . For convenience let us further drop the tilde and write  $f_t$  instead of  $\tilde{f}_t$ . By construction, we have that

$$(90) \quad \forall x, y \in \Omega \quad \forall t_1, t_2 \in [a, b] \quad |f_{t_1}(x, y) - f_{t_2}(x, y)| \leq \omega(|t_1 - t_2|).$$

In particular,  $t \mapsto f_t(x, y)$  is continuous, so  $J(x, y) := \int_a^b f_t(x, y) dt$  is well defined for every  $(x, y) \in \Omega^2$  (on  $\Omega^2 \setminus D_1$  we have  $J(x, y) = 0$ ). It remains to show that  $\|I - J\|_\infty = 0$ .

Let  $I$  be some bounded measurable function  $I : \Omega^2 \rightarrow \mathbb{R}$  that represents  $\int_a^b f_t dt$ . Let  $f_t^{(n)}$  be a  $n$ -step function that approximates  $f_t$  (recalling the convention we made above), let  $I_n(x, y) := (\int_a^b f_t^{(n)} dt)(x, y) = \int f_t^{(n)}(x, y) dt$ , where the equality follows from the definition of the integral for the step function  $f_t^{(n)}$ .

By (86) and (88) we get

$$\begin{aligned}
 \|I_n - I\|_\infty &\leq \sum_{i \in [n]} \left\| \int_{P_i} (f_i^{(n)} - f_i) dt \right\|_\infty \\
 (91) \qquad &\leq \sum_{i \in [n]} |P_i| \sup_{t \in P_i} \|f_i^{(n)} - f_i\|_\infty \\
 &\leq \sum_{i \in [n]} |P_i| \omega(|P_i|) = (b - a) \omega\left(\frac{b - a}{n}\right) =: d_n.
 \end{aligned}$$

We define a set

$$D_2 := \{(x, y) \in \Omega^2 : \forall n \in \mathbb{N} \ |I_n(x, y) - I(x, y)| \leq d_n\},$$

which by (91) has full measure.

Using (90), we have, for every  $x, y \in \Omega$ ,

$$\begin{aligned}
 |I_n(x, y) - J(x, y)| &= \left| \int_a^b [f_t^{(n)}(x, y) - f_t(x, y)] dt \right| \\
 &\leq \sum_{i \in [n]} |P_i| \sup_{t \in P_i} |f_t^{(n)}(x, y) - f_t(x, y)| \\
 &\leq (b - a) \omega\left(\frac{b - a}{n}\right) = d_n.
 \end{aligned}$$

We get that for  $(x, y) \in D_2$

$$\begin{aligned}
 |I(x, y) - J(x, y)| &\leq |I(x, y) - I_n(x, y)| + |I_n(x, y) - J(x, y)| \\
 &\leq 2d_n.
 \end{aligned}$$

Since  $d_n \rightarrow 0$  as  $n \rightarrow \infty$ , we obtain  $I = J$  on  $D_2$  and thus  $I = J$  as elements of  $\mathcal{E}$ .

To prove the “furthermore” part, we modify the set  $D_1$  we chose at the beginning of the proof (still maintaining that it has full measure) so that (89) is satisfied for both  $f$  and  $g$ , while, in addition, for every rational  $t$  we have  $f_t(x, y) \geq g_t(x, y)$  for  $(x, y) \in D_1$ . Then we notice that the continuous extensions of  $t \mapsto f_t(x, y)$  and  $t \mapsto g_t(x, y)$  satisfy  $\tilde{f}_t(x, y) \geq \tilde{g}_t(x, y)$  for  $(x, y) \in D_1$  and every  $t$ . Since we defined  $\tilde{f}_t = \tilde{g}_t = 0$  outside of  $D_1$ , the inequality everywhere follows.  $\square$

**Acknowledgments**

We thank Hanka Řada, Joel Spencer and Lutz Warnke for their comments. We also thank an anonymous referee for their feedback.

**Funding**

FG: Parts of the work were done while affiliated with the Institute of Mathematics of the Czech Academy of Sciences. Research supported by GAČR project 18-01472Y and RVO: 67985840. Furthermore, this work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 648509). This publication reflects only its authors’ view; the European Research Council Executive Agency is not responsible for any use that may be made of the information it contains. This author was also supported by the MUNI Award in Science and Humanities of the Grant Agency of Masaryk University.

JH: Research supported by Czech Science Foundation Project 21-21762X.

MŠ: Research supported by Czech Science Foundation Project 20-27757Y, with Institutional Support RVO:67985807.

FS: Research supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP), the project AI4Research at Uppsala University, and a Simons-Berkeley Research Fellowship.

## References

- [1] R. Abraham, J. E. Marsden and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*, 2nd edition. *Applied Mathematical Sciences* **75**. Springer-Verlag, New York, 1988. MR0960687 <https://doi.org/10.1007/978-1-4612-1029-0>
- [2] P. Araújo, J. Hladký, E. K. Hng and M. Šileikis. Prominent examples of flip processes. Available at [arXiv:2206.03884](https://arxiv.org/abs/2206.03884).
- [3] S. Athreya, F. den Hollander and A. Röllin. Graphon-valued stochastic processes from population genetics. *Ann. Appl. Probab.* **31** (2021) 1724–1745. MR4312844 <https://doi.org/10.1214/20-aap1631>
- [4] T. Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.* **5** (2008) 80–145. MR2426176 <https://doi.org/10.1214/08-PS124>
- [5] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science* **286** (1999) 509–512. MR2091634 <https://doi.org/10.1126/science.286.5439.509>
- [6] T. Bohman, A. Frieze and E. Lubetzky. Random triangle removal. *Adv. Math.* **280** (2015) 379–438. MR3350225 <https://doi.org/10.1016/j.aim.2015.04.015>
- [7] T. Bohman and P. Keevash. The early evolution of the  $H$ -free process. *Invent. Math.* **181** (2010) 291–336. MR2657427 <https://doi.org/10.1007/s00222-010-0247-x>
- [8] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztegombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** (2008) 1801–1851. MR2455626 <https://doi.org/10.1016/j.aim.2008.07.008>
- [9] P. Braunsteins, F. den Hollander and M. Mandjes. A sample-path large deviation principle for dynamic Erdős–Rényi random graphs. *Ann. Appl. Probab.* **33** (2023) 3278–3320. <https://doi.org/10.1214/22-AAP1892>
- [10] F. R. K. Chung, R. L. Graham and R. M. Wilson. Quasi-random graphs. *Combinatorica* **9** (1989) 345–362. MR1054011 <https://doi.org/10.1007/BF02125347>
- [11] H. Crane. Dynamic random networks and their graph limits. *Ann. Appl. Probab.* **26** (2016) 691–721. MR3476622 <https://doi.org/10.1214/15-AAP1098>
- [12] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)* **28** (2008) 33–61. MR2463439
- [13] M. Doležal, J. Grebík, J. Hladký, I. Rocha and V. Rozhoň. Relating the cut distance and the weak\* topology for graphons. *J. Combin. Theory Ser. B* **147** (2021) 252–298. MR4192828 <https://doi.org/10.1016/j.jctb.2020.04.003>
- [14] E. K. Hng. Characterisation of flip process rules with the same trajectories. Available at [arXiv:2305.19925](https://arxiv.org/abs/2305.19925).
- [15] D. Keliger. Markov processes on quasi-random graphs. Available at [arXiv:2105.13446](https://arxiv.org/abs/2105.13446).
- [16] L. Lovász. *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. American Mathematical Society, Providence, RI, 2012. MR3012035 <https://doi.org/10.1090/coll/060>
- [17] C. McDiarmid. Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics* 195–248. *Algorithms Combin.* **16**. Springer, Berlin, 1998. MR1678578 [https://doi.org/10.1007/978-3-662-12788-9\\_6](https://doi.org/10.1007/978-3-662-12788-9_6)
- [18] L. Perko. *Differential Equations and Dynamical Systems*, 3rd edition. *Texts in Applied Mathematics* **7**. Springer-Verlag, New York, 2001. MR1801796 <https://doi.org/10.1007/978-1-4613-0003-8>
- [19] G. Regts and A. Schrijver. Compact orbit spaces in Hilbert spaces and limits of edge-colouring models. *European J. Combin.* **52** (2016) 389–395. MR3425986 <https://doi.org/10.1016/j.ejc.2015.07.013>
- [20] O. Riordan and L. Warnke. Achlioptas process phase transitions are continuous. *Ann. Appl. Probab.* **22** (2012) 1450–1464. MR2985166 <https://doi.org/10.1214/11-AAP798>
- [21] M. Royden and P. M. Fitzpatrick. *Real Analysis*. Prentice Hall, New York, NY, 2010.
- [22] E. Szemerédi. Regular partitions of graphs. In *Problèmes Combinatoires et Théorie des Graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)* 399–401. *Colloq. Internat. CNRS* **260**. CNRS, Paris, 1978. MR0540024
- [23] T. Tao. *Nonlinear Dispersive Equations: Local and Global Analysis*. *CBMS Regional Conference Series in Mathematics* **106**. American Mathematical Society, Providence, RI, 2006. Published for the Conference Board of the Mathematical Sciences, Washington, DC. MR2233925 <https://doi.org/10.1090/cbms/106>
- [24] G. Teschl. *Ordinary Differential Equations and Dynamical Systems*. *Graduate Studies in Mathematics* **140**. American Mathematical Society, Providence, RI, 2012. MR2961944 <https://doi.org/10.1090/gsm/140>
- [25] A. Thomason. Pseudorandom graphs. In *Random Graphs '85 (Poznań, 1985)* 307–331. *North-Holland Math. Stud.* **144**. North-Holland, Amsterdam, 1987. MR0930498
- [26] N. C. Wormald. Differential equations for random processes and random graphs. *Ann. Appl. Probab.* **5** (1995) 1217–1235. MR1384372