Vertex duplication and removal dynamics on large graphs

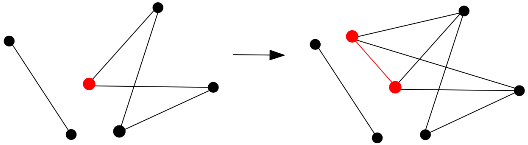
Valentin Féray
joint work with Kelvin Rivera-Lopez, arXiv:2512.20338

CNRS, Institut Élie Cartan de Lorraine (IECL)

17th edition of the French Romanian Colloquium on Applied Mathematics

Nancy, June 23rd, 2026

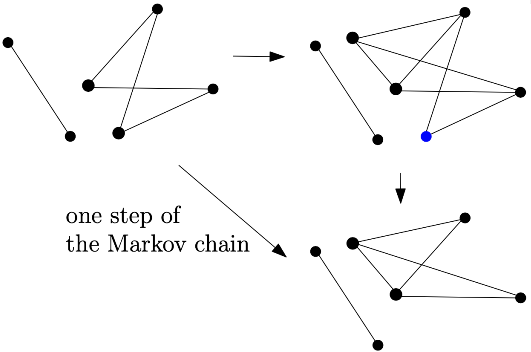
A simple Markov chain on graphs



Upstep : duplicate a uniform random element/vertex.

Then the duplicated vertex and its copy are connected by an edge with probability p.

A simple Markov chain on graphs



Downstep: delete a uniform random element/vertex.

In this talk: behaviour on large graphs, stationary distribution and mixing time.

Motivations

- Large literature on dynamics on large graphs to model evolving networks (social networks, the internet, etc.)

Motivations

- Large literature on dynamics on large graphs to model evolving networks (social networks, the internet, etc.)
- This precise model is inspired by works on up-down chains on partitions in *integrable probability* (Petrov 2009-2013, Borodin–Olshanski 2009, Olshanski 2010, ...).

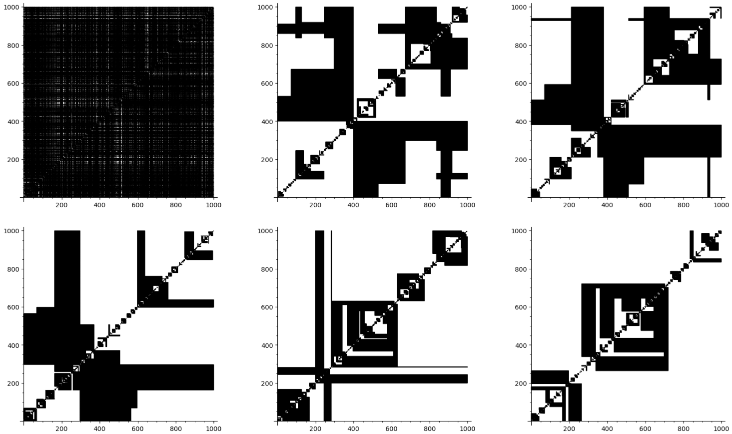
Motivations

- Large literature on dynamics on large graphs to model evolving networks (social networks, the internet, etc.)
- This precise model is inspired by works on up-down chains on partitions in *integrable probability* (Petrov 2009-2013, Borodin–Olshanski 2009, Olshanski 2010, . . .).
- Unlike many models in the literature (e.g., Garbe, Hladký, Šileikis and Skerman, 2022), which have deterministic large graph limits, our model exhibits completely different behaviour.

Motivations

- Large literature on dynamics on large graphs to model evolving networks (social networks, the internet, etc.)
- This precise model is inspired by works on up-down chains on partitions in *integrable probability* (Petrov 2009-2013, Borodin–Olshanski 2009, Olshanski 2010, . . .).
- Unlike many models in the literature (e.g., Garbe, Hladký, Šileikis and Skerman, 2022), which have deterministic large graph limits, our model exhibits completely different behaviour.
- It would be interesting to study more general models which don't alternate strictly between vertex duplication and deletion, mixes edge/vertex local moves, . . .

Simulation



Simulation of our up-down chain with $p = 1/2$ and $n = 1000$.
For $m \in \{0, 1, 2, 3, 4, 5\} \cdot 50000$, we plot the adjacency matrix of the graph obtained after m steps, black/white dots representing 0/1 entries.

Graphons: an overview

- Introduced in 2006 by Borgs, Chayes, Lovász, Sós, and Vesztergombi, and an active topic in graph theory ever since ;

Graphons: an overview

- Introduced in 2006 by Borgs, Chayes, Lovász, Sós, and Vesztergombi, and an active topic in graph theory ever since ;
- Provides a notion of limits for large (dense) graphs ;

Graphons: an overview

- Introduced in 2006 by Borgs, Chayes, Lovász, Sós, and Vesztergombi, and an active topic in graph theory ever since ;
- Provides a notion of limits for large (dense) graphs ;
- Also gives some non-homogeneous models of (random) graphs (generalization of the stochastic block model), used in modelization (population dynamics, game theory) ; usually the next step after the mean field model. . .

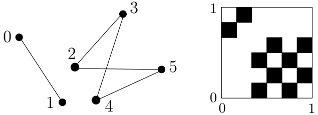
Graphons: an overview

- Introduced in 2006 by Borgs, Chayes, Lovász, Sós, and Vesztergombi, and an active topic in graph theory ever since ;
- Provides a notion of limits for large (dense) graphs ;
- Also gives some non-homogeneous models of (random) graphs (generalization of the stochastic block model), used in modelization (population dynamics, game theory) ; usually the next step after the mean field model. . .
- Applications in large deviation theory and extremal combinatorics (not in this talk).

Crash course on graphons (1/2)

Graph function: with a graph G on vertex set $\{0,\dots,n-1\}$, we associate its *rescaled adjacency matrix* (sometimes called pixel picture) $W_G : [0,1]^2 \to [0,1]$

$$W_G(x,y) = \begin{cases} 1 & \text{if } \{\lfloor nx \rfloor, \lfloor ny \rfloor\} \in E_G; \\ 0 & \text{otherwise.} \end{cases}$$

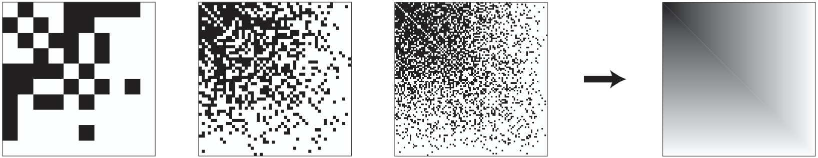


Crash course on graphons (1/2)

Graph function: with a graph G on vertex set $\{0,\dots,n-1\}$, we associate its *rescaled adjacency matrix* (sometimes called pixel picture) $W_G : [0,1]^2 \to [0,1]$

$$W_G(x,y) = \begin{cases} 1 & \text{if } \{\lfloor nx \rfloor, \lfloor ny \rfloor\} \in E_G; \\ 0 & \text{otherwise.} \end{cases}$$

Idea: a sequence of graphs G_n converges if W_{G_n} converges...



©Glasscock, What is...a graphon?, Notices AMS, 2015

Crash course on graphons (1/2)

Graph function: with a graph G on vertex set $\{0,\dots,n-1\}$, we associate its *rescaled adjacency matrix* (sometimes called pixel picture) $W_G : [0,1]^2 \to [0,1]$

$$W_G(x,y) = \begin{cases} 1 & \text{if } \{\lfloor nx \rfloor, \lfloor ny \rfloor\} \in E_G; \\ 0 & \text{otherwise.} \end{cases}$$

Idea: a sequence of graphs G_n converges if W_{G_n} converges... Yes, but

- it's better to use an ad hoc norm

$$\|W\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x,y) dxdy \right|.$$

- The function depends on the labelling of the vertices. We need to identify functions W and W' whenever

$$\exists \varphi \text{ Lebesgue-preserving} : W(\varphi(x), \varphi(y)) = W'(x,y).$$

Crash course on graphons (2/2)

A (*a priori*) different approach:

- To describe a large graph, we may look (asymptotically) at its number of edges ;

Crash course on graphons (2/2)

A (*a priori*) different approach:

- To describe a large graph, we may look (asymptotically) at its number of edges ;
- Not quite enough:

$$\left(\begin{array}{c} \text{a clique of size } n/2 \\ \text{and } n/2 \text{ isolated vertices} \end{array} \right) \not\approx \left(\begin{array}{c} \text{a graph with } n \text{ vertices} \\ \text{and } n^2/8 \text{ randomly chosen edges} \end{array} \right)$$

$\rightarrow$ look also at the number of triangles?

Crash course on graphons (2/2)

A (*a priori*) different approach:

- To describe a large graph, we may look (asymptotically) at its number of edges ;
- Not quite enough:

$$\begin{pmatrix} \text{a clique of size } n/2 \\ \text{and } n/2 \text{ isolated vertices} \end{pmatrix} \not\approx \begin{pmatrix} \text{a graph with } n \text{ vertices} \\ \text{and } n^2/8 \text{ randomly chosen edges} \end{pmatrix}$$

$\rightarrow$ look also at the number of triangles?

Theorem (Borgs, Chayes, Lovász, Sós, Vesztergombi, 2008)

The following are equivalent

- W_{G_n} *converges to* W*;*
- *All subgraph proportions in* G_n *converge to explicit functionals of* W*,*

$$\frac{\#\text{ edges in } G_n}{\binom{|V(G_n)|}{2}} \to \int_{[0,1]^2} W(x,y)dxdy, \quad \frac{\#\text{triangles in } G_n}{\binom{|V(G_n)|}{3}} \to \int_{[0,1]^2} W(x,y)W(y,z)W(z,x)dxdydz, \ \dots$$

Scaling limit

Theorem (F., Rivera-Lopez, 2025)

For $n \geq 1$, let $(X_n(m))_{m \geq 1}$ be the above defined Markov chain on graphs of size n, from an initial (possibly random) graph $G_{n,0}$. Assume that $G_{n,0}$ converges to some (possibly random) graphon W.

Scaling limit

Theorem (F., Rivera-Lopez, 2025)

For $n \geq 1$, let $(X_n(m))_{m\geq 1}$ be the above defined Markov chain on graphs of size n, from an initial (possibly random) graph $G_{n,0}$. Assume that $G_{n,0}$ converges to some (possibly random) graphon W.

Then there exists a continuous Feller diffusion $F = F_W$ in the space $\mathscr{G}$ of graphons with initial state W such that

$$\big(X_n(\lfloor n^2 t \rfloor)\big)_{t\geq 0} \Longrightarrow \big(F(t)\big)_{t\geq 0},$$

in distribution in the Skorokhod space $D([0,+\infty),\mathscr{G})$.

About the limiting process

We can describe the generator A of F on subgraph densities d_H:

$$A\, d_H = k(k-1)\Big(-d_H + \sum_{H' : |V_{H'}| = |V_H| - 1} p^{\uparrow}(H', H)\, d_{H'} \Big). \tag{1}$$

Here,

- $d_H(G)$ is the density of copies of H in G, e.g.
$$d_{\bullet\!-\!\bullet}(G) = \frac{\#\text{ edges in } G_n}{\binom{|V(G_n)|}{2}}, \quad d_{\triangle}(G) = \frac{\#\text{triangles in } G_n}{\binom{|V(G_n)|}{3}}, \dots$$
- $p^{\uparrow}(H', H)$ is the probability to obtain H by applying a random duplication to H'.

(Recall the definition of the generator of a Feller process (X_t):

$$A f(x) := \lim_{t \to \infty} \frac{\mathbb{E}(f(X_t) | X_0 = x) - f(x)}{t}.)$$

About the limiting process

We can describe the generator A of F on subgraph densities d_H:

$$A\,d_H = k(k-1)\Big(-d_H + \sum_{H':|V_{H'}|=|V_H|-1} p^{\uparrow}(H',H)\,d_{H'}\Big). \qquad (1)$$

☹ The generator formula does not give a direct description of F (e.g. giving W as a function of some standard stochastic process; we can only approximate it by pure-jump processes on $\mathscr{G}$).

About the limiting process

We can describe the generator A of F on subgraph densities d_H:

$$A\, d_H = k(k-1)\Big(-d_H + \sum_{H' : |V_{H'}| = |V_H| - 1} p^{\uparrow}(H', H)\, d_{H'} \Big). \tag{1}$$

☹ The generator formula does not give a direct description of F (e.g. giving W as a function of some standard stochastic process; we can only approximate it by pure-jump processes on $\mathscr{G}$).

From (1), we do not even know if the evolution is deterministic or random.

About the limiting process

We can describe the generator A of F on subgraph densities d_H:

$$A\, d_H = k(k-1)\Big(-d_H + \sum_{H' : |V_{H'}| = |V_H| - 1} p^{\uparrow}(H', H)\, d_{H'} \Big). \tag{1}$$

☹ The generator formula does not give a direct description of F (e.g. giving W as a function of some standard stochastic process; we can only approximate it by pure-jump processes on $\mathscr{G}$).

From (1), we do not even know if the evolution is deterministic or random.

☺ Nevertheless, we can show that F is ergodic (i.e. it has a unique stationary distribution and converges to it) and describe the stationary distribution (which is random, implying that the evolution must be random!)

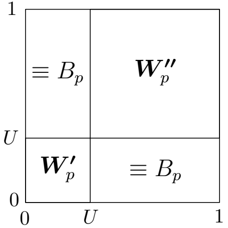
Stationary distribution

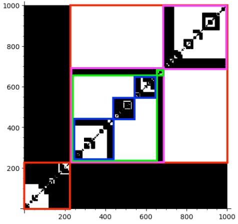
Proposition (F., Rivera-Lopez, 2025)

The limiting process F is ergodic and its stationary distribution is the law of the unique random graphon $\boldsymbol{W}_p$ which satisfies

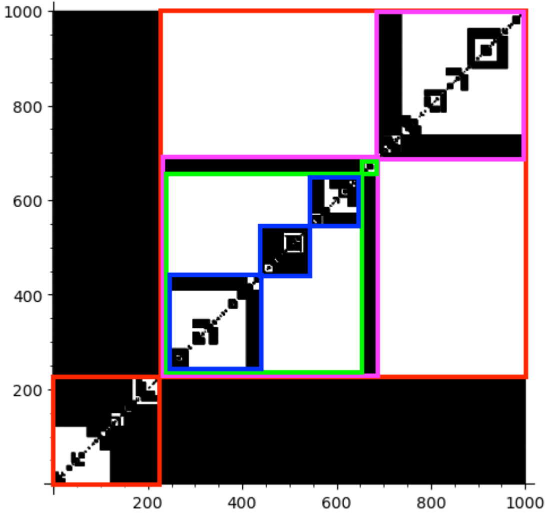
$$\boldsymbol{W}_p(x,y) \overset{law}{=} \begin{cases} \boldsymbol{W}'_p\left(\frac{x}{U}, \frac{y}{U}\right) & \text{if } x,y \in [0,U]; \\ \boldsymbol{W}''_p\left(\frac{x-U}{1-U}, \frac{y-U}{1-U}\right) & \text{if } x,y \in [U,1]; \\ B_p & \text{otherwise,} \end{cases}$$

where $\boldsymbol{W}'_p$ and $\boldsymbol{W}''_p$ are independent copies of $\boldsymbol{W}_p$, and $B_p \sim$ Bernoulli(p).

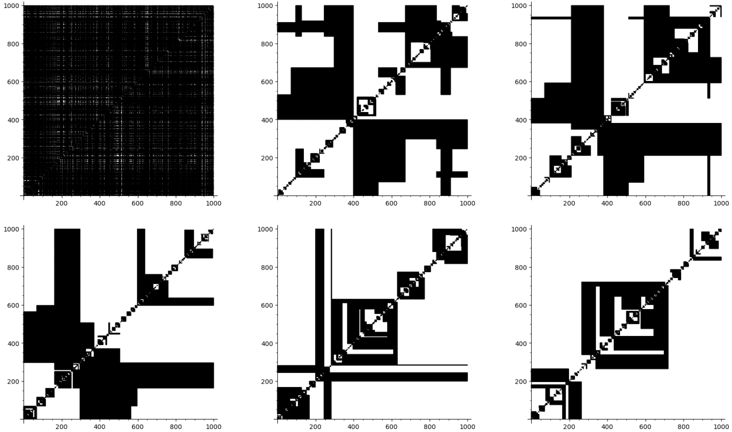




Stationary distribution



Back to simulation



Simulation of our up down-chain with $p = 1/2$, $n = 100$.
For $m \in \{0,1,2,3,4,5\} \cdot 50000$, we plot the adjacency matrix of the graph obtained after m steps, black/white dots representing 0/1 respectively.

Transition

Some proof elements

(Later: results on speed of convergence to the stationary distribution.)

Key identity: the commutation relation

- Let $p_n^{\uparrow} \in \mathscr{M}(\mathscr{G}_n \times \mathscr{G}_{n+1})$ be the up-step transition matrix, i.e. $p_n^{\uparrow}(H, H')$ is the probability to find H' when duplicating a uniform random point in H.
- Let $p_{n+1}^{\downarrow} \in \mathscr{M}(\mathscr{G}_{n+1} \times \mathscr{G}_n)$ be the down-step transition matrix, i.e. $p_{n+1}^{\downarrow}(H', H)$ is the probability of finding H when deleting a uniform random point in H'.

Then the transition matrix of our chain on $\mathscr{G}_n$ is $p_n^{\uparrow} p_{n+1}^{\downarrow}$.

Key identity: the commutation relation

- Let $p_n^{\uparrow} \in \mathscr{M}(\mathscr{G}_n \times \mathscr{G}_{n+1})$ be the up-step transition matrix, i.e. $p_n^{\uparrow}(H,H')$ is the probability to find H' when duplicating a uniform random point in H.
- Let $p_{n+1}^{\downarrow} \in \mathscr{M}(\mathscr{G}_{n+1} \times \mathscr{G}_n)$ be the down-step transition matrix, i.e. $p_{n+1}^{\downarrow}(H',H)$ is the probability of finding H when deleting a uniform random point in H'.

Then the transition matrix of our chain on $\mathscr{G}_n$ is $p_n^{\uparrow} p_{n+1}^{\downarrow}$.

Proposition

For any $n \geq 2$, we have

$$p_n^{\uparrow} p_{n+1}^{\downarrow} = \frac{n-1}{n+1}\, p_n^{\downarrow} p_{n-1}^{\uparrow} + \frac{2}{n+1}\, \mathrm{Id}_{\mathscr{G}_n}, \tag{C}$$

Key identity: the commutation relation

Proposition

For any $n \geq 2$, we have

$$p_n^{\uparrow} p_{n+1}^{\downarrow} = \frac{n-1}{n+1}\, p_n^{\downarrow} p_{n-1}^{\uparrow} + \frac{2}{n+1}\, \mathrm{Id}_{\mathscr{G}_n}, \tag{C}$$

Intuitive explanation: The term $\frac{2}{n+1}\,\mathrm{Id}_{\mathscr{G}_n}$ corresponds to the case where the deleted vertex is the duplicated one or its copy: in this case, the graph remains the same.

Otherwise, deleting first a vertex and then duplicating another vertex would yield the same result, hence the term $\frac{n-1}{n+1}\, p_n^{\downarrow} p_{n-1}^{\uparrow}$.

Key identity: the commutation relation

Proposition

For any $n \geq 2$, we have

$$p_n^{\uparrow} p_{n+1}^{\downarrow} = \frac{n-1}{n+1}\, p_n^{\downarrow} p_{n-1}^{\uparrow} + \frac{2}{n+1}\, \mathrm{Id}_{\mathscr{G}_n}, \tag{C}$$

Intuitive explanation: The term $\frac{2}{n+1}\,\mathrm{Id}_{\mathscr{G}_n}$ corresponds to the case where the deleted vertex is the duplicated one or its copy: in this case, the graph remains the same.
Otherwise, deleting first a vertex and then duplicating another vertex would yield the same result, hence the term $\frac{n-1}{n+1}\, p_n^{\downarrow} p_{n-1}^{\uparrow}$.

In fact, our results hold more generally for up-down chains satisfying such a commutation relation.

Computing the spectrum

Reminder from linear algebra: if A, B are rectangular $m \times p$ and $p \times m$ matrices respectively, with $m > p$, then, we have

$$\mathrm{Sp}(AB) = \mathrm{Sp}(BA) \uplus \{\underbrace{0, \dots, 0}_{m-p \text{ times}}\}. \tag{Sp}$$

Sp(M): multiset of eigenvalues of M.

Computing the spectrum

Reminder from linear algebra: if A, B are rectangular $m \times p$ and $p \times m$ matrices respectively, with $m > p$, then, we have

$$\mathrm{Sp}(AB) = \mathrm{Sp}(BA) \uplus \{\underbrace{0, \dots, 0}_{m-p \text{ times}}\}. \tag{Sp}$$

$\rightarrow$ we can compute the spectrum of $p_n^{\uparrow} p_{n+1}^{\downarrow}$ by induction:

$$\begin{aligned}\mathrm{Sp}(p_n^{\uparrow} p_{n+1}^{\downarrow}) &\overset{(C)}{=} \frac{n-1}{n+1} \otimes \mathrm{Sp}(p_n^{\downarrow} p_{n-1}^{\uparrow}) \oplus \frac{2}{n+1} \\ &\overset{(Sp)}{=} \frac{n-1}{n+1} \otimes \left(\mathrm{Sp}(p_{n-1}^{\uparrow} p_n^{\downarrow}) \uplus \{ \underbrace{0, \dots, 0}_{|\mathscr{G}_n| - |\mathscr{G}_{n-1}| \text{ times}} \} \right) \oplus \frac{2}{n+1}.\end{aligned}$$

Notation: $c \otimes A = \{c \cdot a, a \in A\}$ and $A \oplus c = \{a + c, a \in A\}$.

Computing the spectrum

Reminder from linear algebra: if A, B are rectangular $m \times p$ and $p \times m$ matrices respectively, with $m > p$, then, we have

$$\mathrm{Sp}(AB) = \mathrm{Sp}(BA) \uplus \{\underbrace{0,\dots,0}_{m-p \text{ times}}\}. \tag{Sp}$$

→ we can compute the spectrum of $p_n^{\uparrow} p_{n+1}^{\downarrow}$ by induction:

$$\begin{aligned}\mathrm{Sp}(p_n^{\uparrow} p_{n+1}^{\downarrow}) &\overset{(C)}{=} \frac{n-1}{n+1} \otimes \mathrm{Sp}(p_n^{\downarrow} p_{n-1}^{\uparrow}) \oplus \frac{2}{n+1} \\ &\overset{(Sp)}{=} \frac{n-1}{n+1} \otimes \left(\mathrm{Sp}(p_{n-1}^{\uparrow} p_n^{\downarrow}) \uplus \{ \underbrace{0,\dots,0}_{|\mathcal{G}_n| - |\mathcal{G}_{n-1}| \text{ times}} \} \right) \oplus \frac{2}{n+1}.\end{aligned}$$

Corollary (Fulman, 2009, FKR, 2025)

The transition matrix $p_n = p_n^{\uparrow} p_{n+1}^{\downarrow}$ of the up-down chain has eigenvalues $1 - \frac{i(i-1)}{n(n+1)}$, with multiplicities $|\mathcal{G}_i| - |\mathcal{G}_{i-1}|$.
(for $1 \le i \le n$, with the convention $|\mathcal{G}_0| = 0$.)

Additional proof elements. . .

What remains:

- find the eigenvectors, and relate them to the subgraph densities;

Additional proof elements. . .

What remains:

- find the eigenvectors, and relate them to the subgraph densities;
- show that the generators of the discrete chains converge to the one of the limiting process F;

Additional proof elements...

What remains:

- find the eigenvectors, and relate them to the subgraph densities;
- show that the generators of the discrete chains converge to the one of the limiting process F;
- Using (C) again, we show that the stationary measure of the chain on $\mathscr{G}_n$ is
$$\delta_\bullet p_1^\uparrow p_2^\uparrow \dots p_{n-1}^\uparrow,$$
i.e. the law of the graph obtained by applying random duplication to the single vertex graph.

Additional proof elements...

What remains:

- find the eigenvectors, and relate them to the subgraph densities;
- show that the generators of the discrete chains converge to the one of the limiting process F;
- Using (C) again, we show that the stationary measure of the chain on $\mathscr{G}_n$ is
$$\delta_\bullet p_1^\uparrow p_2^\uparrow \dots p_{n-1}^\uparrow,$$
i.e. the law of the graph obtained by applying random duplication to the single vertex graph.
$\Rightarrow$ the large n limit of the stationary measure must have the claimed self-similarity property.

Separation distance

Standard question for Markov chains: how quickly does it converge to the stationary distribution?

Separation distance

Standard question for Markov chains: how quickly does it converge to the stationary distribution?

We use the separation distance (Aldous–Diaconis, 1987)

$$\Delta_n(m) = \max_{\substack{x,y \in \mathscr{G}_n \\ M_n(y) \neq 0}} \left(1 - \frac{\mathbb{P}_x\big(X_n(m) = y\big)}{M_n(y)} \right),$$

where M_n is the stationary distribution of the chain on $\mathscr{G}_n$.

Separation distance

Standard question for Markov chains: how quickly does it converge to the stationary distribution?

We use the separation distance (Aldous–Diaconis, 1987)

$$\Delta_n(m) = \max_{\substack{x,y \in \mathscr{G}_n \\ M_n(y) \neq 0}} \left(1 - \frac{\mathbb{P}_x\big(X_n(m) = y\big)}{M_n(y)} \right),$$

where M_n is the stationary distribution of the chain on $\mathscr{G}_n$.

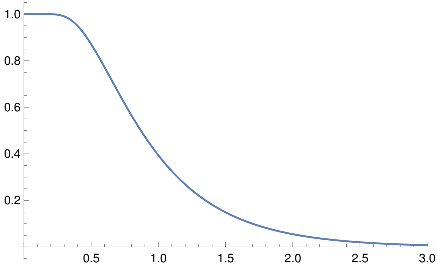
We can also define a continuous analog $\Delta_F(t)$.

Asymptotics of the separation distance

Theorem (F., Rivera-Lopez, 2025)

For any $t > 0$, we have

$$\lim_{n\to\infty} \Delta_n(\lfloor n^2 t \rfloor) = \Delta_F(t) = \sum_{j=1}^{+\infty} (-1)^{j-1}(2j+1)e^{-tj(j+1)}.$$



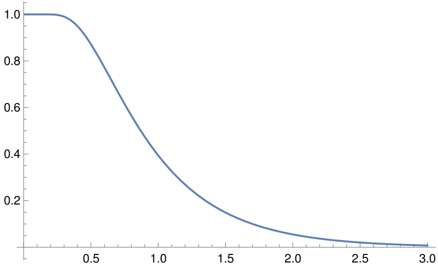
Plot of the function $\Delta_F(t)$

Asymptotics of the separation distance

Theorem (F., Rivera-Lopez, 2025)

For any $t > 0$, we have

$$\lim_{n\to\infty} \Delta_n(\lfloor n^2 t \rfloor) = \Delta_F(t) = \sum_{j=1}^{+\infty} (-1)^{j-1}(2j+1)e^{-tj(j+1)}.$$



Plot of the function $\Delta_F(t)$

Asymptotics:

- as $t \to +\infty$, $\Delta_F(t) \sim 3e^{-2t}$;
- as $t \to 0$, not so clear a priori...

Small t asymptotics: Jacobi identity and modular form

Remarkably, using an identity of Jacobi

$$\sum_{j=0}^{+\infty}(-1)^j(2j+1)q^{\binom{j+1}{2}} = \left(\prod_{i=1}^{+\infty}(1-q^i)\right)^3,$$

we can rewrite $\Delta_F(t)$ as

$$\Delta_F(t) = 1 - \prod_{j=1}^{\infty}(1-e^{-2jt})^3.$$

Small t asymptotics: Jacobi identity and modular form

Remarkably, using an identity of Jacobi

$$\sum_{j=0}^{+\infty}(-1)^j(2j+1)q^{\binom{j+1}{2}} = \left(\prod_{i=1}^{+\infty}(1-q^i)\right)^3,$$

we can rewrite $\Delta_F(t)$ as

$$\Delta_F(t) = 1 - \prod_{j=1}^{\infty}(1-e^{-2jt})^3.$$

But, defining $\eta(\tau) := q^{1/24}\prod_{i=1}^{+\infty}(1-q^i)$ (with $q = e^{2\pi i\tau}$), the function η is a modular form and satisfies $\eta(-1/\tau) = \sqrt{-i\tau}\eta(\tau)$.

Small t asymptotics: Jacobi identity and modular form

Remarkably, using an identity of Jacobi

$$\sum_{j=0}^{+\infty}(-1)^j(2j+1)q^{\binom{j+1}{2}} = \left(\prod_{i=1}^{+\infty}(1-q^i)\right)^3,$$

we can rewrite $\Delta_F(t)$ as

$$\Delta_F(t) = 1 - \prod_{j=1}^{\infty}(1-e^{-2jt})^3.$$

But, defining $\eta(\tau) := q^{1/24}\prod_{i=1}^{+\infty}(1-q^i)$ (with $q = e^{2\pi i\tau}$), the function η is a modular form and satisfies $\eta(-1/\tau) = \sqrt{-i\tau}\,\eta(\tau)$.

Hence

$$1-\Delta_F(t) = \exp\left(-\frac{\pi^2}{4t}+\frac{t}{4}\right)\left(\frac{\pi}{t}\right)^{3/2}\left(1-\Delta_F\left(\frac{\pi^2}{t}\right)\right)$$

and for small t, we have $\Delta_F(t) = 1-\exp\left(-\frac{\pi^2}{4t}\right)\left(\frac{\pi}{t}\right)^{3/2}\left(1+O(e^{-2\pi^2/t})\right)$.

Thank you for your attention