Markov Random Geometric Graph (MRGG): A Growth Model for Temporal Dynamic Networks

Yohann De Castro
Institut Camille Jordan, École Centrale de Lyon, Lyon, France
yohann.de-castro@ec-lyon.fr
&
Quentin Duchemin*
LAMA, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France.
quentin.duchemin@univ-eiffel.fr

June 11, 2020

Abstract

We introduce Markov Random Geometric Graphs (MRGGs), a growth model for temporal dynamic networks. It is based on a Markovian latent space dynamic: consecutive latent points are sampled on the Euclidean Sphere using an unknown Markov kernel; and two nodes are connected with a probability depending on a unknown function of their latent geodesic distance.

More precisely, at each stamp-time k we add a latent point X_k sampled by jumping from the previous one X_{k-1} in a direction chosen uniformly Y_k and with a length r_k drawn from an unknown distribution called the *latitude function*. The connection probabilities between each pair of nodes are equal to the *envelope function* of the distance between these two latent points. We provide theoretical guarantees for the non-parametric estimation of the latitude and the envelope functions.

We propose an efficient algorithm that achieves those non-parametric estimation tasks based on an ad-hoc Hierarchical Agglomerative Clustering approach, and we deploy this analysis on a real data-set given by exchange of messages on a social network.

1 Introduction

In Random Geometric Graph (RGG), nodes are sampled independently in latent space $\mathbb{R}^d$. Two nodes are connected if their distance is smaller than a threshold. A thorough probabilistic study of RGGs can be found in [26]. RGGs have been widely studied recently due to their ability to provide a powerful modeling tool for networks with spatial structure. We can mention applications in bioinformatics [17] or analysis of social media [18]. One main feature is to uncover hidden representation of nodes using latent space and to model interactions by relative positions between latent points.

Furthermore, nodes interactions may evolve with time. In some applications, this evolution is given by the arrival of new nodes as in online collection growth [24], online social network growth [4, 20], or outbreak modeling [32] for instance. The network is growing as more nodes are entering. Other time evolution modelings have been studied, we refer to [28] for a review.

A natural extension of RGG consists in accounting this time evolution. In [12], the expected length of connectivity and dis-connectivity periods of the Dynamic Random Geometric Graph is studied: each node choose at random an angle in $[0, 2\pi)$ and make a constant step size move in that direction. In [30], a random walk model for RGG on the hypercube is studied where at each time step a vertex

*This work was supported by grants from Région Ile-de-France.

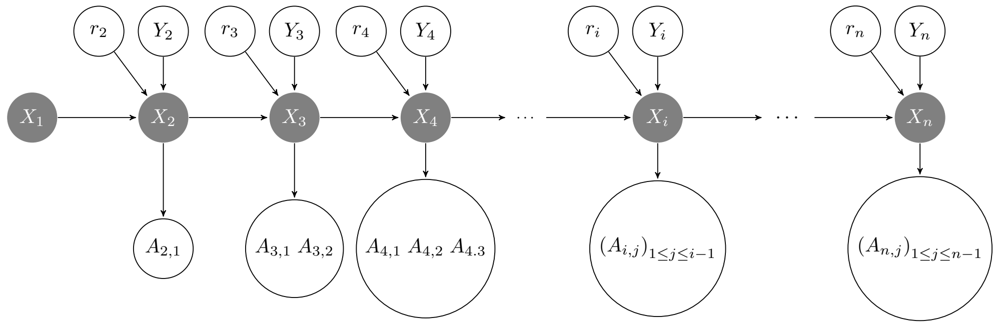


Figure 1: Graphical model of MRGG model: Markovian dynamics on Euclidean sphere where we jump from X_k onto X_{k+1}. The Y_k encodes direction of jump while r_k encodes its distance, see (2).

is either appended or deleted from the graph. Their model falls into the class of Geometric Markovian Random Graphs that are generally defined in [9].
As far as we know, there is no extension of RGG to growth model for temporal dynamic networks. For the first time, in this paper, we consider a Markovian dynamic on the latent space where the new latent point is drawn with respect to the latest latent point and some Markov kernel to be estimated.

Estimation of graphon in RGGs: the Euclidean sphere case Random graphs with latent space can be defined using a *graphon*, see [25]. A graphon is a kernel function that defines edge distribution. In [31], Tang and al. prove that spectral method can recover the matrix formed by graphon evaluated at latent points up to an orthogonal transformation, assuming that graphon is a positive definite kernel (PSD). Going further, algorithms have been designed to estimate graphons, as in [21] which provide sharp rates for the Stochastic Block Model (SBM). Recently, the paper [8] provides an non-parametric algorithm to estimate RGGs on Euclidean spheres, without PSD assumption.
We present here RGG on Euclidean sphere. Given n points $X_1, X_2, \dots, X_n$ on the Euclidean sphere $\mathbb{S}^{d-1}$, we set an edge between nodes i and j (where $i, j \in [n]$, $i \neq j$) with independent probability $p(\langle X_i, X_j \rangle)$. The unknown function $p : [-1, 1] \to [0, 1]$ is called the *envelope function*. This RGG is a graphon model with a symmetric kernel W given by $W(x, y) = p(\langle x, y \rangle)$. Once the latent points are given, independently draw the random undirected adjacency matrix A by

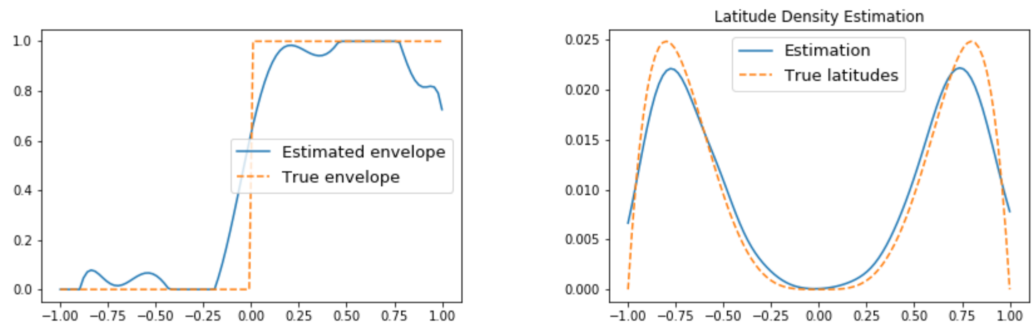
$$A_{i,j} \sim \mathcal{B}(p(\langle X_i, X_j \rangle)) \,, \quad i < j$$

with Bernoulli r.v. drawn independently (set zero on the diagonal and complete by symmetry), and set

$$T_n := \frac{1}{n} \left(p(\langle X_i, X_j \rangle) \right)_{i,j \in [n]} \quad \text{and} \quad \hat{T}_n := \frac{1}{n} A, \tag{1}$$

We do not observe the latent point and we have to estimate the envelope p from A only. A standard strategy is to remark that $\hat{T}_n$ is a random perturbation of T_n and to dig into T_n to uncover p.
One important feature of this model is that the interactions between nodes is depicted by a simple object: the envelope function p. The envelope summarises how individuals connect each others given their latent positions. Standard examples [7] are given by $p_\tau(t) = \mathbb{1}_{\{t \geq \tau\}}$ where one connects two points as soon as their geodesic distance is below some threshold. The non-parametric estimation of p is given by [8] where the authors assume that latent points X_i are independently and uniformly distributed on the sphere, which will not be the case in the present paper.

A new growth model: the latent Markovian dynamic Consider RGGs where latent points are sampled with Markovian jumps, the Graphical Model under consideration can be found in Figure 1.



(a) True and estimated envelope functions. (b) True and estimated latitude functions.

Figure 2: Non-parametric estimation of envelope and latitude functions using algorithms of Sections 2 and 3. We built a graph of 1500 nodes sampled on the sphere $\mathbb{S}^2$ and using envelope and latitude (dot orange curves) defined in Section 5 by (10). The estimated envelope is thresholded to get a function in $[0, 1]$ and the estimated latitude function is normalized with integral 1 (plain blue lines).

Namely, we sample n points $X_1, X_2, \ldots, X_n$ on the Euclidean sphere $\mathbb{S}^{d-1}$ using a Markovian dynamic. We start by sampling randomly X_1 on $\mathbb{S}^{d-1}$. Then, for any $i \in \{2, \ldots, n\}$, we sample

- a unit vector $Y_i \in \mathbb{S}^{d-1}$ uniformly, orthogonal to X_{i-1}.
- a real $r_i \in [-1, 1]$ encoding the distance between X_{i-1} and X_i, see (3). r_i is sampled from a distribution $f_{\mathcal{L}} : [-1, 1] \to [0, 1]$, called the *latitude function*.

then X_i is defined by

$$X_i = r_i \times X_{i-1} + \sqrt{1 - r_i^2} \times Y_i \,. \tag{2}$$

This dynamic can be pictured as follows. Consider that X_{i-1} is the north pole, then chose uniformly a direction (i.e., a longitude) and, in a independent manner, randomly move along the latitudes (the longitude being fixed by the previous step). The geodesic distance γ_i drawn on the latitudes satisfies

$$\gamma_i = \arccos(r_i) \,, \tag{3}$$

where random variable $r_i = \langle X_i, X_{i-1} \rangle$ has density $f_{\mathcal{L}}(r_i)$. The resulting model will be referred to as the Markov Random Geometric Graph (MRGG) and is described with Figure 1.

Temporal Dynamic Networks: MRGG estimation strategy Seldom growth models exist for temporal dynamic network modeling, see [28] for a review. In our model, we add one node at a time making a Markovian jump from the previous latent position. It results in

$$\text{the observation of } (A_{i,j})_{1 \leq j \leq i-1} \text{ at time } T = i \,,$$

as pictured in Figure 1. Namely, we observe how a new node connects to the previous ones. For such dynamic, we aim at estimating the model, namely envelope p and respectively latitude $f_{\mathcal{L}}$. These functions capture in a simple function on $\Omega = [-1, 1]$ the range of interaction of nodes (represented by p) and respectively the dynamic of the jumps in latent space (represented by $f_{\mathcal{L}}$), where, in abscissa Ω, values $r = \langle X_i, X_j \rangle$ near 1 corresponds to close point $X_i \simeq X_j$ while values close to -1 corresponds to antipodal points $X_i \simeq -X_j$. These functions may be non-parametric.
From snapshots of the graph at different time steps, can we recover envelope and latitude functions? This paper proves that it is possible under mild conditions on the Markovian dynamic of the latent points and our approach is summed up with Figure 3.

Fundamental result		Guarantee for the recovery of:		Algorithm
Spectral convergence of $\hat{T}_n$ under Markovian dynamic, see Section 2.1	$\Rightarrow$	(a) envelope p, see (5)	$\leftrightarrow$	**SCCHEi**
		(b) latent distances r_i, see (9)	$\leftrightarrow$	**HEiC** [2]

Figure 3: Presentation of our method to recover the envelope and the latitude functions.

Define $\lambda(T_n) := (\lambda_1, \dots, \lambda_n)$ and resp. $\lambda(\hat{T}_n) := (\hat{\lambda}_1, \dots, \hat{\lambda}_n)$ the spectrum of T_n and resp. $\hat{T}_n$, see (1). Building clusters from $\lambda(\hat{T}_n)$, Algorithm 1 (SCCHEi) estimates the spectrum of envelope p while Algorithm 2 [2] (HEiC, see Section E in Appendix) extracts d eigenvectors of $\hat{T}_n$ to uncover the Gram matrix of the latent positions. Both can then be used to estimate the unknown functions of our model (see Figure 2).

Previous works Non-parametric estimation of RGGs on Euclidean sphere has been investigated in [8] with i.i.d. latent points. Estimation of latent point relative distances with HEiC Algorithm has been introduced in [2] under i.i.d. latent points assumption. Phase transitions on the detection of geometry in RGGs (against Erdös Rényi alternatives) has been investigated in [7].
For the first time, we introduce latitude function and non-parametric estimations of envelope and latitude using new results on kernel matrices concentration with dependent variables (see Appendix).

Outline Sections 2 and 3 present the estimation method with new theoretical results under Markovian dynamic. These news results are random matrices operator norm control and resp. U-statistics control under Markovian dynamic, shown in the Appendix at Section H and resp. Section G. The envelope *adaptive* estimate is built from a size constrained clustering (Algorithm 1) tuned by slope heuristic (6), and the latitude function estimate (see Section 3.1) is derived from estimates of latent distances r_i. Section 5 investigates synthetic and real data experiments. Our method can handle random graphs with logarithmic growth node degree (i.e., new comer at time $T = n$ connects to $\mathcal{O}(\log n)$ previous nodes), referred to as *relatively sparse* model, see Section 4.

Notations Consider a dimension $d \geq 3$. Denote by $\|\cdot\|_2$ (resp. $\langle \cdot, \cdot \rangle$) the Euclidean norm (resp. inner product) on $\mathbb{R}^d$. Consider the d-dimensional sphere $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d \;:\; \|x\|_2 = 1\}$ and denote by σ the uniform distribution on $\mathbb{S}^{d-1}$. For two real valued sequences $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$, denote $u_n \underset{n\to\infty}{=} \mathcal{O}(v_n)$ if there exist $k_1 > 0$ and $n_0 \in \mathbb{N}$ such that $\forall n > n_0$, $|u_n| \leq k_1 |v_n|$.
Given two sequences x, y of reals–completing finite sequences by zeros–such that $\sum_i x_i^2 + y_i^2 < \infty$, we define the ℓ_2 rearrangement distance $\delta_2(x, y)$ as

$$\delta_2^2(x, y) := \inf_{\pi \in \mathfrak{S}} \sum_i (x_i - y_{\pi(i)})^2 \,,$$

where $\mathfrak{S}$ is the set of permutations with finite support. This distance is useful to compare two spectra.

2 Nonparametric estimation of the envelope function

One can associate with $W(x, y) = p(\langle x, y \rangle)$ the integral operator $\mathbb{T}_W : L^2(\mathbb{S}^{d-1}) \to L^2(\mathbb{S}^{d-1})$:

$$\forall g \in L^2(\mathbb{S}^{d-1}),\ \forall x \in \mathbb{S}^{d-1}, \quad (\mathbb{T}_W g)(x) = \int_{\mathbb{S}^{d-1}} g(y) p(\langle x, y \rangle) \mathrm{d}\sigma(y),$$

where $\mathrm{d}\sigma$ is the Lebesgue measure on $\mathbb{S}^{d-1}$. The operator $\mathbb{T}_W$ is Hilbert-Schmidt and it has a countable number of bounded eigenvalues λ_k^* with zero as only accumulation point. The eigenfunctions of $\mathbb{T}_W$ have the remarkable property that they do not depend on p (see [10] Lemma 1.2.3): they are given by the real Spherical Harmonics. We denote $\mathcal{H}_l$ the space of real Spherical Harmonics of degree l with dimension d_l and with orthonormal basis $(Y_{l,j})_{j\in[d_l]}$ where

$$d_l := \dim(\mathcal{H}_l) = \begin{cases} 1 & \text{if } l = 0 \\ d & \text{if } l = 1 \\ \binom{l+d-1}{l} - \binom{l+d-3}{l-2} & \text{otherwise.} \end{cases}$$

We define also for all $R \in \mathbb{N}$, $\tilde{R} := \sum_{l=0}^{R} d_l$. We end up with the following spectral decomposition

$$p(\langle x, y\rangle) = \sum_{l\geq 0} p_l^* \sum_{1\leq j\leq d_l} Y_{l,j}(x)Y_{l,j}(y) = \sum_{k\geq 0} p_k^* c_k G_k^\beta(\langle x, y\rangle)\,, \tag{4}$$

where $\lambda(\mathbb{T}_W) = \{p_0^*, p_1^*, \dots, p_1^*, \dots, p_l^*, \dots, p_l^*, \dots\}$ meaning that each eigenvalue p_l^* has multiplicity d_l; and G_k^β is the Gegenbauer polynomial of degree k with parameter $\beta := \frac{d-2}{2}$ and $c_k := \frac{2k+d-2}{d-2}$ (see Appendix C). Since p is bounded, one has $p \in L^2((-1,1), w_\beta)$ where the weight function w_β is defined by $w_\beta(t) := (1-t^2)^{\beta-\frac{1}{2}}$, and it can be decomposed as $p \equiv \sum_{k\geq} p_k^* c_k G_k^\beta$ and the Gegenbauer polynomials G_k^β are an orthogonal basis of $w_\beta(t) := (1-t^2)^{\beta-\frac{1}{2}}$.

Weighted Sobolev space The space $Z^s_{w_\beta}((-1,1))$ with smoothness $s > 0$ is defined as the set of functions $g = \sum_{k\geq 0} g_k^* c_k G_k^\beta \in L^2((-1,1), w_\beta)$ such that

$$\|g\|^*_{Z^s_{w_\beta}((-1,1))} := \left[\sum_{l=0}^{\infty} d_l |g_l^*|^2 \left(1 + (l(l+2\beta))^s\right)\right]^{1/2} < \infty.$$

2.1 Integral operator spectrum estimation with dependent variables

One key result is a new control of U-statistics with latent Markov variables (see Section G) and it makes use of a Talagrand's concentration inequality for Markov chains (see [1]). This article follows the hypothesis made on the Markov chain $(X_i)_{i\geq 1}$ by [1]. Namely, we referred to as "*mild conditions*" the assumption that latitude function $f_\mathcal{L}$ is bounded away from 0 with $\|f_\mathcal{L}\|_\infty < \infty$; and the first and the second regeneration times associated with the *split* chain to have sub-exponential tail. Those assumptions are fully described in section G.
Theorem 1 is a theoretical guarantee for a random matrix approximation of spectrum of integral operator with **dependent** latent variables. Theorem 3 in Appendix H gives explicitly the constants hidden in the big O below which depend on the spectral gap of the Markov chain $(X_i)_{i\geq 1}$.

Theorem 1 *Assume mild conditions on the Markov chain $(X_i)_{i\geq 1}$ and assume the envelope p has regularity $s > 0$. Then, it holds*

$$\mathbb{E}\left[\delta_2^2(\lambda(\mathbb{T}_W), \lambda(T_n)) \vee \delta_2^2(\lambda(\mathbb{T}_W), \lambda^{R_{opt}}(\hat{T}_n))\right] = \mathcal{O}\left(\left[\frac{n}{\log^2(n)}\right]^{-\frac{2s}{2s+d-1}}\right),$$

with $\lambda^{R_{opt}}(\hat{T}_n) = (\hat{\lambda}_1, \dots, \hat{\lambda}_{\tilde{R}_{opt}}, 0, 0, \dots)$ and $R_{opt} = \lfloor (n/\log^2(n))^{\frac{1}{2s+d-1}} \rfloor$ where $\hat{\lambda}_1, \dots, \hat{\lambda}_n$ are the eigenvalues of $\hat{T}_n$ sorted in decreasing order of magnitude.

Remark In Theorem 1 and Theorem 2, note that we recover, up to a log factor, the *minimax rate of non-parametric estimation* of s-regular functions on a space of (Riemannian) dimension $d-1$. Even with i.i.d. latent variables, it is still an open question to know if this rate is the minimax rate of non-parametric estimation of RGGs.
Eq.(4) shows that one could use an approximation of $(p_k^*)_{k\geq 1}$ to estimate the envelope p and Theorem 1 states we can recover $(p_k^*)_{k\geq 1}$ up to a permutation. The problem of finding such a permutation is NP-hard and we introduce in the next section an efficient algorithm to fix this issue.

2.2 Size Constrained Clustering Algorithm

Note the spectrum of $\mathbb{T}_W$ is given by $(p_l^*)_{l\geq 0}$ where p_l^* has multiplicity d_l. In order to recover envelope p, we build clusters from eigenvalues of $\hat{T}_n$ while respecting the dimension d_l of each eigen-space of $\mathbb{T}_W$. In [8], an algorithm is proposed testing all permutations of $\{0,\dots,R\}$ for a given maximal resolution R. To bypass the high computational cost of such approach, we propose an efficient method based on the tree built from *Hierarchical Agglomerative Clustering* (HAC).
In the following, for any $\nu_1,\dots,\nu_n \in \mathbb{R}$, we denote by $\mathrm{HAC}(\{\nu_1,,\dots,\nu_n\},d_c)$ the tree built by a HAC on the real values $\nu_1,\dots,\nu_n$ using the complete linkage function d_c defined by $\forall A,B\subset\mathbb{R}$, $d_c(A,B)=\max_{a\in A}\max_{b\in B}\|a-b\|_2$. Algorithm 1 describes our approach.

Algorithm 1 Size Constrained Clustering for Harmonic Eigenvalues (SCCHEi).

Data: Resolution R, matrix $\hat{T}_n=\frac{1}{n}A$, dimensions $(d_k)_{k=0}^R$.

```
1:  Let λ̂_1, ..., λ̂_n be the eigenvalues of T̂_n sorted in decreasing order of magnitude.
2:  Set P := {λ̂_1, ..., λ̂_R̃} and dims = [d_0, d_1, ..., d_R].
3:  while All eigenvalues in P are not clustered do
4:     tree ← HAC(nonclustered eigenvalues in P, d_c)
5:     for d ∈ dims do
6:        Search for a cluster of size d in tree as close as possible to the root.
7:        if such a cluster C_d exists then Update(dims, tree, C_d, d).
8:     for d ∈ dims do
9:        Search for the group C in tree with a size larger than d and as close as possible to d.
10:       if such a group exists then Update(dims, tree, C, d) else Go to step 3.
```

Return: $\mathcal{C}_{d_0},\dots,\mathcal{C}_{d_R},\{\hat{\lambda}_{\tilde{R}+1},\dots,\hat{\lambda}_n\}$

Update($dims, tree, \mathcal{C}, d$).

Save the subset $\mathcal{C}_d$ consisting of the d eigenvalues in $\mathcal{C}$ with the largest absolute values.
Delete from *tree* all occurrences to eigenvalues in $\mathcal{C}_d$ and delete d from *dims*.

2.3 Adaptation: Slope heuristic as model selection of Resolution

A data-driven choice of model size R can be done by *slope heuristic*, see [3] for a nice review. One main idea of slope heuristic is to penalize the empirical risk by $\kappa\,\mathrm{pen}(\tilde{R})$ and to calibrate $\kappa>0$. If the sequence $(\mathrm{pen}(\tilde{R}))_{\tilde{R}}$ is equivalent to the sequence of variances of the population risk of empirical risk minimizer (ERM) as model size $\tilde{R}$ grows, then, penalizing the empirical risk (as done in (6)), one may ultimately uncover an empirical version of the U-shaped curve of the population risk. Hence, minimizing it, one builds a model size $\hat{R}$ that balances between bias (*under-fitting* regime) and variance (*over-fitting* regime). First, note that empirical risk is given by the intra-class variance below.

Definition 1 *For any output $(\mathcal{C}_{d_0},\dots,\mathcal{C}_{d_R},\Lambda)$ of the Algorithm SCCHEi, the thresholded intra-class variance is defined by*

$$\mathcal{I}_R := \frac{1}{n}\left[\sum_{k=0}^{R}\sum_{\lambda\in\mathcal{C}_{d_k}}\left(\lambda-\frac{1}{d_k}\sum_{\lambda'\in\mathcal{C}_{d_k}}\lambda'\right)^2+\sum_{\lambda\in\Lambda}\lambda^2\right],$$

and the estimations $(\hat{p}_k)_{k\geq 0}$ of the eigenvalues $(p_k^*)_{k\geq 0}$ is given by

$$\forall k\in\mathbb{N},\quad \hat{p}_k=\begin{cases}\frac{1}{d_k}\sum_{\lambda\in\mathcal{C}_{d_k}}\lambda & \text{if } k\in\{0,\dots,\hat{R}\}\\ 0 & \text{otherwise.}\end{cases} \tag{5}$$

Second, as underlined in the proof of Theorem 1 (see Theorem 3 in the Appendix), the estimator's variance of our estimator scales linearly in $\tilde{R}$.
Hence, we apply Algorithm SCCHEi for R varying from 0 to $R_{\max} := \max\{R \geq 0 \ : \ \tilde{R} \leq n\}$ to compute the *thresholded intra-class variance* $\mathcal{I}_R$ (see Definition 1) and given some $\kappa > 0$, we select

$$R(\kappa) \in \underset{R \in \{0,\dots,R_{max}\}}{\arg\min} \left\{ \mathcal{I}_R + \kappa \frac{\tilde{R}}{n} \right\}. \tag{6}$$

The hyper-parameter κ controlling the bias-variance trade-off is set to $2\kappa_0$ where κ_0 is the value of $\kappa > 0$ leading to the "*largest jump*" of the function $\kappa \mapsto R(\kappa)$. Once $\hat{R} := R(2\kappa_0)$ has been computed, we approximate the envelope function p using (5) (see (14) in Appendix for the closed form). In Appendix D, we describe this slope heuristic on real data and our results can be reproduced using the notebook *Experiments*[1].

3 Nonparametric estimation of the latitude function

3.1 Our approach to estimate the latitude function in a nutshell

In Theorem 2 (see below), we show that we are able to estimate consistently the pairwise distances encoded by the Gram matrix G^* where

$$G^* := \frac{1}{n} \left(\langle X_i, X_j \rangle \right)_{i,j \in [n]}.$$

Taking the diagonal just above the main diagonal (referred to as *superdiagonal*) of $\hat{G}$, we get estimates of the i.i.d. random variables $(\langle X_i, X_{i-1} \rangle)_{2 \leq i \leq n} = (r_i)_{2 \leq i \leq n}$ sampled from $f_{\mathcal{L}}$. Using $(\hat{r}_i)_{2 \leq i \leq n}$ the superdiagonal of $n\hat{G}$, we can build a kernel density estimator of the latitude function $f_{\mathcal{L}}$. In the following, we describe the algorithm used to build our estimator $\hat{G}$ with theoretical guarantees.

3.2 Spectral gap condition and Gram matrix estimation

The Gegenbauer polynomial of degree one is defined by $G_1^{\beta}(t) = 2\beta t, \ \forall t \in [-1, 1]$. As a consequence, the Gram matrix G^* is related to the Gegenbauer polynomial of degree one by

$$G^* = \frac{1}{2\beta n} \left(G_1^{\beta}(\langle X_i, X_j \rangle) \right)_{i,j \in [n]} = \frac{1}{nd} \sum_{k=1}^{d} Y_{1,k}(X_i) Y_{1,k}(X_j), \tag{7}$$

using the *addition theorem* (see [10, Lem.1.2.3 and Thm.1.2.6]). Denoting $V^* \in \mathbb{R}^{n \times d}$ the matrix with columns $v_k^* := \frac{1}{\sqrt{n}} \left(Y_{1,k}(X_1), \dots, Y_{1,k}(X_n) \right)$ for $k \in [d]$, (7) becomes

$$G^* := \frac{1}{d} V^* (V^*)^{\top}.$$

We will prove that for n large enough there exists a matrix $\hat{V} \in \mathbb{R}^{n \times d}$ where each column is an eigenvector of $\hat{T}_n$, such that $\hat{G} := \frac{1}{d} \hat{V} \hat{V}^{\top}$ approximates G^* well, in the sense that the norm $\|G^* - \hat{G}\|_F$ converges to 0. To choose the d eigenvectors of the matrix $\hat{T}_n$ that we will use to build the matrix $\hat{V}$, we need the following spectral gap condition

$$\Delta^* := \min_{k \in \mathbb{N}, \ k \neq 1} |p_1^* - p_k^*| > 0. \tag{8}$$

This condition will allow us to apply Davis-Kahan type inequalities.

[1] `https://github.com/quentin-duchemin/Markovian-random-geometric-graph`

Now, thanks to Theorem 1, we know that the spectrum of the matrix $\hat{T}_n$ converges towards the spectrum of the integral operator $\mathbb{T}_W$. Then, based on (7), one can naturally think that extracting the d eigenvectors of the matrix $\hat{T}_n$ related with the eigenvalues that converge towards p_1^*, we can approximate the Gram matrix G^* of the latent positions. Theorem 2 proves that the latter intuition is true with high probability under the spectral gap condition (8). The algorithm HEiC [2] (See Section E for a presentation) aims at identifying the above mentioned d eigenvectors of the matrix $\hat{T}_n$ to build our estimate of the Gram matrix G^*.

Theorem 2 *Assume mild conditions on the Markov chain $(X_i)_{i\geq 1}$, assume $\Delta^* > 0$, and assume that graphon W has regularity $s > 0$. We denote $\hat{V} \in \mathbb{R}^{n\times d}$ the d eigenvectors of the matrix $\hat{T}_n$ associated with the eigenvalues returned by the algorithm HEiC and we define $\hat{G} := \frac{1}{d}\hat{V}\hat{V}^\top$. Then for n large enough and for some constant $D > 0$, it holds with probability at least $1 - 5/n^2$,*

$$\|G^* - \hat{G}\|_F \leq D \left(\frac{n}{\log^2(n)} \right)^{\frac{-s}{2s+d-1}} . \tag{9}$$

4 Relatively Sparse Regime

As highlighted in [2], one can use the algorithm HEiC to estimate the underlying dimension d by running the algorithm with different dimensions and keeping the one that leads to the larger gap (as defined in Algorithm 2). Although this paper deals with the so-called *dense* regime (i.e. when the expected number of neighbors of each node scales linearly with n), our results may be generalized to the *relatively sparse* model connecting nodes i and j with probability $W(X_i, X_j) = \zeta_n p(\langle X_i, X_j\rangle)$ where $\zeta_n \in (0, 1]$ satisfies $\liminf_n \zeta_n n/\log n \geq Z$ for some universal constant $Z > 0$.

In the relatively sparse model, one can show following the proof of Theorem 1 that the resolution should be chosen as $\hat{R} = \left(\frac{n\zeta_n}{1+\zeta_n \log^2 n}\right)^{\frac{1}{2s+d-1}}$. Specifying that $\lambda^* = (p_0^*, p_1^*, \ldots, p_1^*, p_2^*, \ldots)$ and $\hat{T}_n = A/n$, Theorem 1 becomes for a graphon with regularity $s > 0$

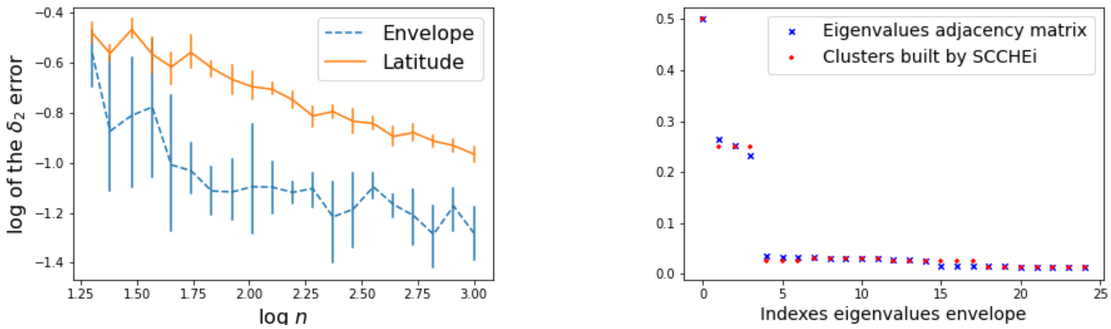
$$\mathbb{E}\left[\delta_2^2\left(\lambda^*, \frac{\lambda(\hat{T}_n)}{\zeta_n}\right)\right] = \mathcal{O}\left(\left(\frac{n\zeta_n}{1+\zeta_n \log^2 n}\right)^{\frac{-2s}{2s+d-1}}\right).$$

5 Experiments

Experiments with simulated data We tested our method using $d = 3$ and considering the following functions

$$p : x \mapsto \mathbb{1}_{x\geq 0}, \quad \text{and} \quad f_{\mathcal{L}} : x \mapsto \begin{cases} \frac{1}{2}g(1-x; 2, 2) & \text{if } x \geq 0 \\ \frac{1}{2}g(1+x; 2, 2) & \text{otherwise} \end{cases}, \tag{10}$$

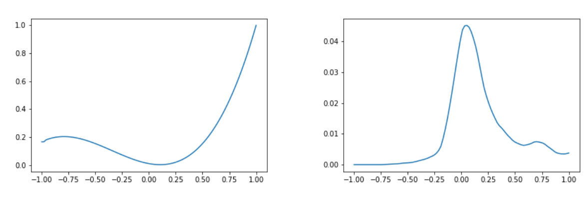
where $g(\cdot; 2, 2)$ is the density of the beta distribution with parameters $(2, 2)$. Figure 4.(a) presents the δ_2-error between the spectra of the true envelope (resp. latitude) function and the estimated one when the size of the graph is increasing. In Figure 4.(b), we propose a visualization of the clustering performed by SCCHEi with $n = 1000$ and $R = 4$. Blue crosses represent the $\tilde{R}$ eigenvalues of $\hat{T}_n$ with the largest magnitude, which are used to form clusters corresponding to the five-first spherical harmonic spaces. The red circles are the estimated eigenvalues $(\hat{p}_k)_{0\leq k\leq 4}$ (plotted with multiplicity) defined from the clustering given by our algorithm SCCHEi (see (5)). Those results show that SCCHEi achieves a relevant clustering of the eigenvalues of $\hat{T}_n$ which allows us to recover the envelope function (see Figure 2.(a)).



(a) δ_2 error on envelope and latitude functions. (b) Clustering adjacency eigenvalues ($R = 4$).

Figure 4: Non-parametric estimation of envelope and latitude using algorithms described in Sections 2 and 3. In (a), bars represent standard deviation of δ_2 errors between true and estimated functions.

Experiments with real data We test our algorithm on real data provided by [29] and collected from a social network involving students at University of California, Irvine. The dataset includes dated messages between users. We consider a self-avoiding random walk starting from the student that first sent a message and where we always move to the addressee of the older message sent by the current user. If at some point all the contacts of the current user have been already seen, we restart the random walk following the same procedure with the remaining users. A portion of this walk between two restarts is called an *excursion*. We keep only the excursions of length at least 20, which leads to a graph with 302 nodes where two nodes are connected if the corresponding users exchanged at least one message. Working with a dimension $d = 3$, we apply our algorithm on this network. Figure 5.(a) presents the estimated envelope function which proves that students who share messages have most of the time latent representations which are highly correlated. Figure 5.(b) shows the estimated latitude function. The peak at 0 corresponds to the transitions in the self-avoiding random walk where we restart the process by choosing a new user. The right tail of the distribution is slowly decreasing and remains strictly positive, corresponding to transitions of the random walk where we stay in the same group of contacts (i.e. without restarting the random walk). The higher the average number of contacts per user, the smaller the peak at 0 of the latitude function and the larger the right tail of the distribution. Those results can be reproduced using the notebook *Experiments*[1] and Appendix D gives the details of the slope heuristic used to choose the resolution $\hat{R}$. In Appendix F, we propose another application of our algorithm with another dataset.



(a) Envelope function (b) Latitude function

Figure 5: We run Algorithm SCCHEi with $d = 3$ and $\hat{R} = 3$. Figure (a) shows the estimated function p which was rescaled in order to have 1 as maximum value on $[0, 1]$. For the latitude function, we removed 10% of the $(\hat{r}_i)_i$ with the largest magnitude and we rescaled those remaining so that the maximum magnitude is 1. Normalizing to get a density on [-1,1], we obtain the latitude function presented in Figure 5 (b).

References

[1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to markov chains. *Electronic Journal of Probability*, 13, 10 2007. doi: 10.1214/EJP.v13-521.

[2] E. Araya and Y. De Castro. Latent distance estimation for random geometric graphs. In *Advances in Neural Information Processing Systems*, pages 8721–8731, 2019.

[3] S. Arlot. Minimal penalties and the slope heuristics: a survey. *Journal de la Société Française de Statistique*, 160(3):1–106, 2019.

[4] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 44–54, 2006.

[5] A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 07 2016. doi: 10.1214/15-AOP1025. URL `https://doi.org/10.1214/15-AOP1025`.

[6] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer New York, 1996. ISBN 9780387948461. URL `https://books.google.fr/books?id=F4hRy1F1M6QC`.

[7] S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, Jan 2016. ISSN 1042-9832. doi: 10.1002/rsa.20633. URL `http://dx.doi.org/10.1002/rsa.20633`.

[8] Y. D. Castro, C. Lacour, and T. M. P. Ngoc. Adaptive estimation of nonparametric geometric graphs. *Mathematical Statistics and Learning*, 2020.

[9] A. E. Clementi, F. Pasquale, A. Monti, and R. Silvestri. Information spreading in stationary markovian evolving graphs. *2009 IEEE International Symposium on Parallel & Distributed Processing*, May 2009. doi: 10.1109/ipdps.2009.5160986. URL `http://dx.doi.org/10.1109/IPDPS.2009.5160986`.

[10] Y. Dai, F.and Xu. *Approximation theory and harmonic analysis on spheres and balls*. Springer, 2013.

[11] C. Dorea and A. Pereira. A note on a variation of doeblin's condition for uniform ergodicity of markov chains. *Acta Mathematica Hungarica*, 110:287–292, 01 2006. doi: 10.1007/s10474-006-0023-y.

[12] J. Díaz, D. Mitsche, and X. Pérez-Giménez. On the connectivity of dynamic random geometric graphs, 01 2008.

[13] S. A. M. . J. L. E. Cho. *Friendship and Mobility: User Movement in Location-Based Social Networks*. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2011. URL `https://snap.stanford.edu/data/loc-gowalla.html`.

[14] J. Fan, B. Jiang, and Q. Sun. Hoeffding's lemma for markov chains and its applications to statistical learning, 2018.

[15] P. Gilles. *The volume of convex bodies and Banach space geometry*. Cambridge University Press, 1989.

[16] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*, volume 40. Cambridge University Press, 2016.

[17] D. J. Higham, M. Rašajski, and N. Pržulj. Fitting a geometric graph to a protein–protein interaction network. *Bioinformatics*, 24(8):1093–1099, 03 2008. ISSN 1367-4803. doi: 10.1093/bioinformatics/btn079.

[18] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002. doi: 10.1198/016214502388618906. URL https://doi.org/10.1198/016214502388618906.

[19] B. Jiang, Q. Sun, and J. Fan. Bernstein's inequality for general markov chains, 2018.

[20] E. M. Jin, M. Girvan, and M. E. Newman. Structure of growing social networks. *Physical review E*, 64(4):046132, 2001.

[21] O. Klopp, A. B. Tsybakov, and N. Verzelen. Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1):316–354, Feb 2017. ISSN 0090-5364. doi: 10.1214/16-aos1454. URL http://dx.doi.org/10.1214/16-AOS1454.

[22] V. Koltchinskii and E. Giné. Random matrix approximation of spectra of integral operators. *Bernoulli*, 6, 02 2000. doi: 10.2307/3318636.

[23] I. Kontoyiannis and S. P. Meyn. Geometric ergodicity and the spectral gap of non-reversible markov chains. *Probability Theory and Related Fields*, 154(1-2):327–339, Jun 2011. ISSN 1432-2064. doi: 10.1007/s00440-011-0373-4. URL http://dx.doi.org/10.1007/s00440-011-0373-4.

[24] C. Lo, J. Cheng, and J. Leskovec. Understanding online collection growth over time: A case study of pinterest. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, page 545–554, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee. ISBN 9781450349147. doi: 10.1145/3041021.3054189. URL https://doi.org/10.1145/3041021.3054189.

[25] L. Lovász. *Large Networks and Graph Limits*. American Mathematical Society, 01 2012.

[26] M. Penrose. Random geometric graphs, 01 2003.

[27] G. O. Roberts and J. S. Rosenthal. General state space markov chains and mcmc algorithms. *Probability Surveys*, 1(0):20–71, 2004. ISSN 1549-5787. doi: 10.1214/154957804100000024. URL http://dx.doi.org/10.1214/154957804100000024.

[28] G. Rossetti and R. Cazabet. Community discovery in dynamic networks: a survey. *ACM Computing Surveys (CSUR)*, 51(2):1–37, 2018.

[29] R. A. Rossi and N. K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015. URL http://networkrepository.com.

[30] R. S. . G. Staples. Dynamic geometric graph processes : Adjacency operator approach, 2009.

[31] M. Tang, D. L. Sussman, and C. E. Priebe. Universally consistent vertex classification for latent positions graphs. *The Annals of Statistics*, 41(3):1406–1430, Jun 2013. ISSN 0090-5364. doi: 10.1214/13-aos1112. URL http://dx.doi.org/10.1214/13-AOS1112.

[32] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16):5962–5966, 2012.

Appendix

Guidelines for the Appendix

Sections A to C: Basic definitions and Complements

In Section A we recall basic definitions on Markov chains which are required for Section B where we describe some properties verified by the Markov chain $(X_i)_{i\geq 1}$. Section C provides complementary results on the Harmonic Analysis on $\mathbb{S}^{d-1}$ which will be useful for our proofs.

Sections D to F: Algorithms and Experiments

Section D describes precisely the slope heuristic used to perform the adaptive selection of the model dimension $\tilde{R}$ on real data. Section E provides a complete description of the HEiC alogorithm used to extract d-eigenvectors of the adjacency matrix that will be used to estimate the Gram matrix of the latent positions. Section F is another application on real data of our algorithms where no significant dependence structure is revealed by the estimation the latitude function even if the RGG model still provides interesting information through the estimation of the envelope function.

Sections G to I: Proofs of theoretical results

Thereafter, we dig into the most theoretical part of the Appendix. In Section G, we provide a full description of the assumptions we made on the Markov chain $(X_i)_{i\geq 1}$ and which we have been referring to so far by *mild conditions*. Section G is also dedicated to the proof of a concentration result for a particular U-statistic of the Markov chain $(X_i)_{i\geq 1}$ that is an essential element of the proof of Theorem 1 which is provided in Section H. Finally, the proof of Theorem 2 can be found in Section I.

A Definitions for general Markov chains

We consider a state space E and a sigma-algebra Σ on E which is a standard Borel space. We denote by $(X_i)_{i\geq 1}$ a time homogeneous Markov chain on the state space (E, Σ) with transition kernel P.

A.1 Ergodic and reversible Markov chains

Definition 2 *[27, section 3.2] (ϕ-irreducible Markov chains)*
The Markov chain $(X_i)_{i\geq 1}$ is said ϕ-irreducible if there exists a non-zero σ-finite measure ϕ on E such that for all $A \subset E$ with $\phi(A) > 0$, and for all $x \in E$, there exists a positive integer $n = n(x, A)$ such that $P^n(x, A) > 0$ (where $P^n(x, \cdot)$ denotes the distribution of X_{n+1} conditioned on $X_1 = x$).

Definition 3 *[27, section 3.2] (Aperiodic Markov chains)*
The Markov chain $(X_i)_{i\geq 1}$ with invariant distribution π is aperiodic if there do not exist $m \geq 2$ and disjoint subsets $A_1, \dots, A_m \subset E$ with $P(x, A_{i+1}) = 1$ for all $x \in A_i$ ($1 \leq i \leq m-1$), and $P(x, A_1) = 1$ for all $x \in A_m$, such that $\pi(A_1) > 0$ (and hence $\pi(A_i) > 0$ for all i).

Definition 4 *[27, section 3.4] (Geometric ergodicity)*
The Markov chain $(X_i)_{i\geq 1}$ is said geometrically ergodic if there exists an invariant distribution π and functions $\rho : E \to (0, 1)$ and $C : E \to [1, \infty)$ such that

$$\|P^n(x, \cdot) - \pi\|_{TV} \leq C(x)\rho(x)^n, \qquad \forall n \geq 0,\ \pi\text{−a.e } x \in E,$$

where $\|\mu\|_{TV} := \sup_{A\in\Sigma} |\mu(A)|$.

Definition 5 *[27, section 3.3] (Uniform ergodicity)*
The Markov chain $(X_i)_{i\geq 1}$ is said uniformly ergodic if there exists an invariant distribution π and constants $0 < \rho < 1$ and $L > 0$ such that

$$\|P^n(x, \cdot) - \pi\|_{TV} \leq L\rho^n, \qquad \forall n \geq 0,\ \pi\text{−a.e } x \in E,$$

where $\|\mu\|_{TV} := \sup_{A\in\Sigma} |\mu(A)|$.

Remark: A Markov chain geometrically or uniformly ergodic admits a unique invariant distribution.

Definition 6 *A Markov chain is said reversible if there exists a distribution π satisfying*

$$\pi(dx)P(x,dy) = \pi(dy)P(y,dx).$$

A.2 Spectral gap

This section is largely inspired from [14]. Let us consider that the Markov chain $(X_i)i \geq 1$ admits a unique invariant distribution π on $\mathbb{S}^{d-1}$.
For any real-valued, Σ-measurable function $h : E \to \mathbb{R}$, we define $\pi(h) := \int h(x)\pi(dx)$. The set

$$\mathcal{L}_2(E,\Sigma,\pi) := \{h : \pi(h^2) < \infty\}$$

is a Hilbert space endowed with the inner product

$$\langle h_1, h_2\rangle_\pi = \int h_1(x)h_2(x)\pi(dx),\ \forall h_1, h_2 \in \mathcal{L}^2(E,\Sigma,\pi).$$

The map

$$\|\cdot\|_\pi : h \in \mathcal{L}_2(E,\Sigma,\pi) \mapsto \|h\|_\pi = \sqrt{\langle h,h\rangle_\pi},$$

is a norm on $\mathcal{L}_2(E,\Sigma,\pi)$. $\|\cdot\|_\pi$ naturally allows to define the norm of a linear operator T on $\mathcal{L}_2(E,\Sigma,\pi)$ as

$$N_\pi(T) = \sup\{\|Th\|_\pi : \|h\|_\pi = 1\}.$$

To each transition probability kernel $P(x,B)$ with $x \in E$ and $B \in \Sigma$ invariant with respect to π, we can associate a bounded linear operator $h \mapsto \int h(y)P(\cdot,dy)$ on $\mathcal{L}_2(E,\Sigma,\pi)$. Denoting this operator P, we get

$$Ph(x) = \int h(y)P(x,dy),\ \forall x \in E,\ \forall h \in \mathcal{L}_2(E,\Sigma,\pi).$$

Let $\mathcal{L}_2^0(\pi) := \{h \in \mathcal{L}_2(E,\Sigma,\pi)\ :\ \pi(h) = 0\}$. We define the absolute spectral gap of a Markov operator.

Definition 7 *(Spectral gap) A Markov operator P reversible admits a spectral gap $1-\lambda$ if*

$$\lambda := \sup\left\{\frac{\|Ph\|_\pi}{\|h\|_\pi}\ :\ h \in \mathcal{L}_2^0(\pi),\ h \neq 0\right\} < 1.$$

The next result provides a connection between spectral gap and geometric ergodicity for reversible Markov chains.

Proposition 1 *[23, Prop 1.2]*
A reversible, ϕ-irreducible and aperiodic Markov chain is geometrically ergodic if and only if P admits a spectral gap.

B Properties of the Markov chain

Let P be the Markov operator of the Markov chain $(X_i)_{i\geq 1}$. By abuse of notation, we will also denote $P(x,\cdot)$ the density of the measure $P(x,dz)$ with respect to $d\sigma$, the uniform measure on $\mathbb{S}^{d-1}$. For any $x,z\in\mathbb{S}^{d-1}$, we denote $R_x^z\in\mathbb{R}^{d\times d}$ a rotation matrix sending x to z (i.e. $R_x^z x=z$) and keeping $\mathrm{Span}(x,z)^{\perp}$ fixed. In the following, we denote $e_d:=(0,0,\dots,0,1)$.

B.1 Invariant distribution and reversibility for the Markov chain

Reversibility of the Markov chain $(X_i)_{i\geq 1}$.

Lemma 1 *For all $x,z\in\mathbb{S}^{d-1}$, $P(x,z)=P(z,x)=P(e_d,R_z^{e_d}x)$.*

Proof of Lemma 1.
Using our model described in section 2, we get $X_2=rX_1+\sqrt{1-r^2}Y$ where conditionally on X_1, Y is uniformly sampled on $\mathcal{S}(X_1):=\{q\in\mathbb{S}^{d-1}\ :\ \langle q,X_1\rangle=0\}$, and where r has density $f_{\mathcal{L}}$ on $[-1,1]$. Let us consider a gaussian vector $W\sim\mathcal{N}(0,I_d)$. Using the Cochran's theorem and Lemma 2, we know that conditionally on X_1, the random variable $\frac{W-\langle W,X_1\rangle X_1}{\|W-\langle W,X_1\rangle X_1\|_2}$ is distributed uniformly on $\mathcal{S}(X_1)$.

Lemma 2 *Let $W\sim\mathcal{N}(0,I_d)$. Then, $\frac{W}{\|W\|_2}$ is distributed uniformly on the sphere $\mathbb{S}^{d-1}$.*

Denoting $\overset{\mathcal{L}}{=}$ the equality in distribution sense, we have

$$X_2\overset{\mathcal{L}}{=}rX_1+\sqrt{1-r^2}\frac{W-\langle W,X_1\rangle X_1}{\|W-\langle W,X_1\rangle X_1\|_2}=rX_1+\sqrt{1-r^2}\frac{R_{X_2}^{X_1}W'-\langle R_{X_2}^{X_1}W',X_1\rangle X_1}{\|R_{X_2}^{X_1}W'-\langle R_{X_2}^{X_1}W',X_1\rangle X_1\|_2},$$

where $W':=R_{X_1}^{X_2}W$. Note that $W'\in\mathbb{R}^d$ is also a standard centered gaussian vector because this distribution is invariant by rotation. Since $\langle R_{X_2}^{X_1}W',X_1\rangle=\langle W',X_2\rangle$ and $\|R_{X_2}^{X_1}q\|_2=\|q\|_2,\ \forall q\in\mathbb{S}^{d-1}$, we deduce that

$$X_2-rX_1\overset{\mathcal{L}}{=}R_{X_2}^{X_1}\left[\sqrt{1-r^2}\frac{W'-\langle W',X_2\rangle X_2}{\|W'-\langle W',X_2\rangle X_2\|_2}\right]. \tag{11}$$

$R_{X_1}^{X_2}$ is the rotation that sends X_1 to X_2 keeping the other dimensions fixed. Let us denote $a_1:=X_1$, $a_2:=\frac{X_2-rX_1}{\|X_2-rX_1\|_2}$ and complete the linearly independent family (a_1,a_2) in an orthonormal basis of $\mathbb{R}^d$ given by $a:=(a_1,a_2,\dots,a_d)$. Then, the matrix of $R_{X_1}^{X_2}$ in the basis a is

$$\begin{bmatrix} r & -\sqrt{1-r^2} & 0_{d-2}^{\top}\\ \sqrt{1-r^2} & r & 0_{d-2}^{\top}\\ 0_{d-2} & 0_{d-2} & I_{d-2}\end{bmatrix}.$$

We deduce that

$$\begin{aligned}\left(R_{X_2}^{X_1}\right)^{-1}(X_2-rX_1)&=R_{X_1}^{X_2}(X_2-rX_1)\\&=\|X_2-rX_1\|_2R_{X_1}^{X_2}\left(\frac{X_2-rX_1}{\|X_2-rX_1\|_2}\right)\\&=\|X_2-rX_1\|_2R_{X_1}^{X_2}a_2\\&=\|X_2-rX_1\|_2\left[-\sqrt{1-r^2}a_1+ra_2\right]\\&=-\sqrt{1-r^2}\|X_2-rX_1\|_2X_1+rX_2-r^2X_1\\&=-(1-r^2)X_1+rX_2-r^2X_1\\&=-X_1+rX_2.\end{aligned}$$

Going back to (11), we deduce that

$$X_1 \overset{\mathcal{L}}{=} rX_2 + \sqrt{1-r^2}\frac{\tilde{W} - \langle \tilde{W}, X_2\rangle X_2}{\|\tilde{W} - \langle \tilde{W}, X_2\rangle X_2\|_2}, \tag{12}$$

where $\tilde{W} = -W'$ is also a standard centered gaussian vector in $\mathbb{R}^d$. Thus, we proved the first equality of Lemma 1. Based on (12) we have,

$$\begin{aligned} R^{e_d}_{X_2} X_1 &\overset{\mathcal{L}}{=} rR^{e_d}_{X_2} X_2 + \sqrt{1-r^2}\frac{R^{e_d}_{X_2}\tilde{W} - \langle \tilde{W}, X_2\rangle R^{e_d}_{X_2} X_2}{\|\tilde{W} - \langle \tilde{W}, X_2\rangle X_2\|_2} \\ &= re_d + \sqrt{1-r^2}\frac{R^{e_d}_{X_2}\tilde{W} - \langle R^{e_d}_{X_2}\tilde{W}, e_d\rangle e_d}{\|R^{e_d}_{X_2}\tilde{W} - \langle R^{e_d}_{X_2}\tilde{W}, e_d\rangle e_d\|_2}, \end{aligned}$$

which proves that $P(e_d, R^{e_d}_{x_2} x_1) = P(x_2, x_1)$ for any $x_1, x_1 \in \mathbb{S}^{d-1}$ because $R^{e_d}_{X_2}\tilde{W}$ is again a standard centered gaussian vector in $\mathbb{R}^d$.
■

Invariant distribution of the Markov chain.

Proposition 2 *The uniform distribution on the sphere $\mathbb{S}^{d-1}$ is an invariant distribution of the Marokv chain $(X_i)_{i\geq 1}$.*

Proof of Proposition 2.
Let us consider $z \in \mathbb{S}^{d-1}$. We denote $d\sigma \equiv d\sigma_d$ the Lebesgue measure on $\mathbb{S}^{d-1}$ and $d\sigma_{d-1}$ the Lebesgue measure on $\mathbb{S}^{d-2}$. Using [10, section 1.1], it holds $b_d := \int_{\mathbb{S}^{d-1}} d\sigma = \frac{2\pi^{d/2}}{\Gamma(d/2)}$. We have

$$\begin{aligned} & \int_{x\in\mathbb{S}^{d-1}} P(x,z)\frac{d\sigma(x)}{b_d} \\ = & \int_{x\in\mathbb{S}^{d-1}} P(e_d, R^{e_d}_z x)\frac{d\sigma(x)}{b_d} \quad \text{(Using Lemma 1)} \\ = & \int_{x\in\mathbb{S}^{d-1}} P(e_d, x)\frac{d\sigma(x)}{b_d} \quad \text{(Using the change of variable } x \mapsto R^{e_d}_z x) \\ = & \int_0^\pi \int_{\mathbb{S}^{d-2}} P\left(e_d, \begin{bmatrix} \xi \sin\theta \\ \cos\theta \end{bmatrix}\right) (\sin\theta)^{d-2} d\theta \frac{d\sigma_{d-1}(\xi)}{b_d} \quad \text{(Using [10, Eq.(1.5.4) Section 1.5])} \\ = & \int_{-1}^1 \int_{\mathbb{S}^{d-2}} P\left(e_d, \begin{bmatrix} \xi\sqrt{1-r^2} \\ r \end{bmatrix}\right) (1-r^2)^{\frac{d-3}{2}} dr \frac{d\sigma_{d-1}(\xi)}{b_d} \\ = & \int_{-1}^1 f_{\mathcal{L}}(r) dr \times \int_{\mathbb{S}^{d-2}} \frac{1}{b_{d-1}} \frac{d\sigma_{d-1}}{b_d} = \frac{1}{b_d}, \end{aligned}$$

which proves that the uniform distribution on the sphere is an invariant distribution of the Markov chain.
■

B.2 Ergodicity of the Markov chain

Lemma 3 *We consider that $f_{\mathcal{L}}$ is bounded away from 0. Then, the Markov chain $(X_i)_{i\geq 1}$ is ϕ-irreducible and aperiodic.*

Proof of Lemma 3.

Considering for ϕ the uniform distribution on $\mathbb{S}^{d-1}$, we get that for any $x \in \mathbb{S}^{d-1}$ and any $A \subset \mathbb{S}^{d-1}$ with $\phi(A) > 0$,

$$
\begin{aligned}
P(x, A) &= \int_{z \in A} P(x, z) \frac{d\sigma(z)}{b_d} \\
&= \int_{z \in A} P(e_d, R_x^{e_d} z) \frac{d\sigma(z)}{b_d} \quad \text{(Using Lemma 1)} \\
&= \int_{z \in R_x^{e_d} A} P(e_d, z) \frac{d\sigma(z)}{b_d} \\
&\text{(Using the change of variable } z \mapsto R_x^{e_d} z \text{ with } R_x^{e_d} A = \{R_x^{e_d} a : a \in A\}) \\
&= \int_{r \in [-1,1]} \int_{\xi \in \mathbb{S}^{d-2}} f_{\mathcal{L}}(r) 1_{(\xi^\top, r)^\top \in R_x^{e_d} A} dr \frac{d\sigma_{d-1}(\xi)}{b_{d-1} b_d} > 0,
\end{aligned}
$$

since ϕ is invariant by rotation, $\phi(A) > 0$ and $f_{\mathcal{L}}$ is bounded away from 0.
Let us prove now that the chain is aperiodic. Suppose that A_1 and A_2 are disjoint subsets of E both of positive σ measure, with $P(x, A_2) = 1$ for all $x \in A_1$. Consider any $x \in A_1$, then since A_1 must have positive Lebesgue measure, we have $P(x, A_1) > 0$ which concludes the proof.
■

Lemma 4 *We consider that $f_{\mathcal{L}}$ is bounded away from 0. Then the Markov chain $(X_i)_{i \geq 1}$ uniformly ergodic.*

Proof of Lemma 4. The condition on $f_{\mathcal{L}}$ implies that for all $x \in \mathbb{S}^{d-1}$, the density $P(x, \cdot)$ with respect to the Lebesgue measure on $\mathbb{S}^{d-1}$ is strictly positive everywhere. This implies that the Markov chain $(X_i)_{i \geq 1}$ is uniformly ergodic (see [11, section 1]).
■
Remark: From Lemmas 1 and 3 and Proposition 2, we know that the Markov chain $(X_i)_{i \geq 1}$ is ϕ-irreducible, aperiodic and reversible with unique invariant distribution the uniform distribution on the sphere. Thanks to Proposition 1, we deduce that the Markov chain has a spectral gap if and only if the chain is geometrically ergodic. With Proposition 4, we know that the Markov chain $(X_i)_{i \geq 1}$ is uniformly ergodic and thus is geometrically ergodic. Hence, $(X_i)_{i \geq 1}$ has a spectral gap. In the following subsection, we show that this spectral gap is equal to 1.

B.3 Computation of the spectral gap of the Markov chain

Keeping notations of Appendix A, let us consider $h \in \mathcal{L}_2^0(\sigma)$ such that $\|h\|_\sigma = 1$. Then

$$
\begin{aligned}
\|Ph\|_\sigma^2 &= \int_{x \in \mathbb{S}^{d-1}} \left(\int_{y \in \mathbb{S}^{d-1}} P(x, dy) h(y) \right)^2 d\sigma(x) \\
&= \int_{x \in \mathbb{S}^{d-1}} \left(\int_{y \in \mathbb{S}^{d-1}} P(x, y) h(y) d\sigma(y) \right)^2 d\sigma(x) \\
&= \int_{x \in \mathbb{S}^{d-1}} \left(\int_{y \in \mathbb{S}^{d-1}} P(e_d, R_y^{e_d} x) h(y) d\sigma(y) \right)^2 d\sigma(x) \quad \text{(Using Lemma 1)} \\
&= \int_{x \in \mathbb{S}^{d-1}} \left(\int_{y \in \mathbb{S}^{d-1}} P(e_d, x) h(y) d\sigma(y) \right)^2 d\sigma(x) \quad \text{(Using the rotational invariance of } \sigma) \\
&= \int_{x \in \mathbb{S}^{d-1}} P(e_d, x)^2 \left(\int_{y \in \mathbb{S}^{d-1}} h(y) d\sigma(y) \right)^2 d\sigma(x) \\
&= 0,
\end{aligned}
$$

where the last equality comes from $h \in \mathcal{L}_2^0(\sigma)$. Hence, the Markov chain $(X_i)_{i \geq 1}$ has 1 for spectral gap.

C Complement on Harmonic Analysis on the sphere

This section completes the brief introduction to Harmonic Analysis on the sphere $\mathbb{S}^{d-1}$ provided in Section 2. We will need in our proof the following result which states that fixing one variable and integrating with respect to the other one with the uniform measure on $\mathbb{S}^{d-1}$ gives $\|W - W_R\|_2^2$.

Lemma 5 *For any $x \in \mathbb{S}^{d-1}$,*

$$\mathbb{E}_{X\sim\pi}[(W - W_R)^2(x, X)] = \|W - W_R\|_2^2,$$

where π is the uniform measure on the $\mathbb{S}^{d-1}$.

Proof of Lemma 5.

$$\begin{aligned}
\mathbb{E}_{X\sim\pi}[(W - W_R)^2(x, X)] &= \int_y (W - W_R)^2(x, y)\pi(dy) \\
&= \int_y \left(\sum_{r>R} p_r^* \sum_{l=1}^{d_r} Y_{r,l}(x)Y_{r,l}(y) \right)^2 \pi(dy) \\
&= \int_y \sum_{r_1,r_2>R} p_{r_1}^* p_{r_2}^* \sum_{l_1=1}^{d_r} \sum_{l_2=1}^{d_r} Y_{r_1,l_1}(x)Y_{r_1,l_1}(y)Y_{r_2,l_2}(x)Y_{r_2,l_2}(y)\pi(dy) \\
&= \sum_{r_1,r_2>R} p_{r_1}^* p_{r_2}^* \sum_{l_1=1}^{d_r} \sum_{l_2=1}^{d_r} Y_{r_1,l_1}(x)Y_{r_2,l_2}(x) \int_y Y_{r_1,l_1}(y)Y_{r_2,l_2}(y)\pi(dy).
\end{aligned}$$

Since $\int_y Y_{r,l}(y)Y_{r',l'}\pi(dy)$ is 1 if $r = r'$ and $l = l'$ and 0 otherwise, we have that

$$\begin{aligned}
\mathbb{E}_{X\sim\pi}[(W - W_R)^2(x, X)] &= \sum_{r>R} (p_r^*)^2 \sum_{l=1}^{d_r} Y_{r,l}(x)^2 \\
&= \sum_{r>R} (p_r^*)^2 d_r \quad \text{(Using [10, Eq.(1.2.9)])} \\
&= \|W - W_R\|_2^2.
\end{aligned}$$

■

Let us consider $\beta := \frac{d-2}{2}$ and the weight function $w_\beta(t) := (1 - t^2)^{\beta - \frac{1}{2}}$. As highlighted in section 2, any envelope function $p \in L^2([-1, 1], w_\beta)$ can be decomposed as $p \equiv \sum_{k=0}^{R} p_k^* c_k G_k^\beta$ where G_l^β is the Gegenbauer polynomial of degree l with parameter β and where $c_k := \frac{2k+d-2}{d-2}$. The Gegenbauer polynomials are orthonormal polynomials on $[-1, 1]$ associated with the weight function w_β.

The eigenvalues $(p_k^*)_{k\geq 0}$ of the envelope function can be computed numerically through the formula

$$\forall l \geq 0, \quad p_l^* = \left(\frac{c_l b_d}{d_l} \right) \int_{-1}^{1} p(t) G_l^\beta(t) w_\beta(t) dt, \tag{13}$$

where $b_d := \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{d}{2}-\frac{1}{2})}$ with Γ the Gamma function. Hence, it is possible to recover the envelope function p thanks to the identity

$$p = \sum_{l\geq 0} \sqrt{d_l} p_l^* \frac{G_l^\beta}{\|G_l^\beta\|_{L^2([-1,1],w_\beta)}} = \sum_{l\geq 0} p_l^* c_l G_l^\beta. \tag{14}$$

For $R \geq 0$, let us define $\tilde{R} = \sum_{k=0}^{R} d_k$ which corresponds to the dimension of the space of Spherical Harmonics with degree at most R. We introduce the truncated graphon W_R which is obtained from W by keeping only the $\tilde{R}$ first eigenvalues, that is

$$\forall x, y \in \mathbb{S}^{d-1}, \quad W_R(x, y) := \sum_{k=0}^{R} p_k^* \sum_{l=1}^{d_k} Y_{k,l}(x) Y_{k,l}(y).$$

Similarly, we denote for all $t \in [0, 1]$, $p_R(t) = \sum_{k=0}^{R} p_k^* c_k G_k^{\beta}(t)$.

D Slope heuristic on real data

We propose a detailed analysis of the slope heuristic described in section 2.2 on the real data presented in section 5. We recall that $R(\kappa)$ represents the optimal value of R to minimize the bias-variance decomposition defined by (6) for a given hyperparameter κ. Figure 6 shows the evolution of $\tilde{R}(\kappa)$ with respect to κ which is sampled on a logscale. $\tilde{R}(\kappa)$ is the dimension of the space of Spherical Harmonics with degree at most $R(\kappa)$. Our slope heuristic consists in choosing the value κ_0 leading to the larger jump of the function $\kappa \mapsto \tilde{R}(\kappa)$. In our case, Figure 6 shows that $\kappa_0 = 10^{-3.5}$. As described in Section 2.2, the resolution level $\hat{R}$ selected to cluster the eigenvalues of the matrix $\hat{T}_n$ is given by $R(2\kappa_0)$. In our case, we choose $\hat{R} = 3$.

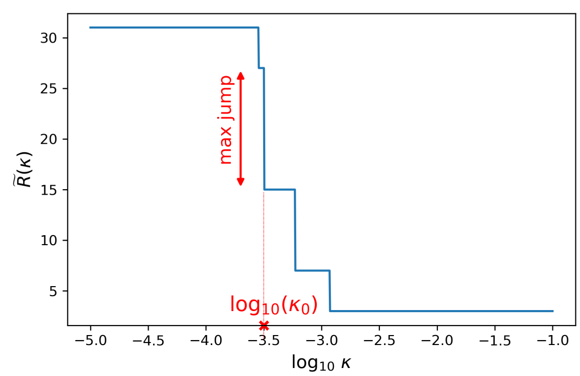


Figure 6: We sample the parameter κ on a logscale between 10^{-5} and 10^{-1} and we compute the corresponding $R(\kappa)$ defined in (6). We plot the values of $\tilde{R}(\kappa)$ with respect to κ. The larger jump allows us to define κ_0.

E Reminder on Harmonic EigenCluster(HEiC)

Before presenting the algorithm HEiC, let us define for a given set of indices $i_1, \ldots, i_d \in [n]$

$$\mathrm{Gap}_1(\hat{T}_n; i_1, \ldots, i_d) := \min_{i \notin \{i_1, \ldots, i_d\}} \max_{j \in \{i_1, \ldots, i_d\}} |\hat{\lambda}_i - \hat{\lambda}_j|.$$

Algorithm 2 Harmonic EigenCluster(HEiC) algorithm.

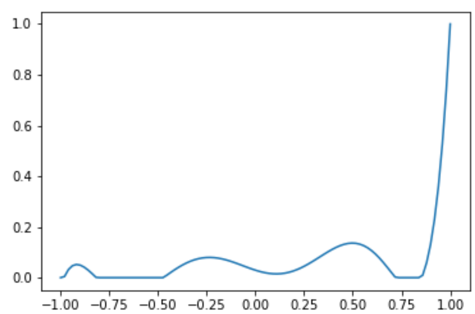
Data: Adjacency matrix A. Dimension d.

$(\hat{\lambda}_1^{sort}, \ldots, \hat{\lambda}_n^{sort}) \leftarrow$ eigenvalues of $\hat{T}_n$ sorted decreasing order.
$\Lambda_1 \leftarrow \{\hat{\lambda}_1^{sort}, \ldots, \hat{\lambda}_d^{sort}\}$.
Initialize $i = 2$ and gap$= \mathrm{Gap}_1(\hat{T}_n; 1, 2, \ldots, d)$.
while $i \leq n - d + 1$ **do**
  if $\mathrm{Gap}_1(\hat{T}_n; i, i+1, \ldots, i+d-1) >$ gap **then**
    $\Lambda_1 \leftarrow \{\hat{\lambda}_i^{sort}, \ldots, \hat{\lambda}_{i+d-1}^{sort}\}$
  $i = i + 1$

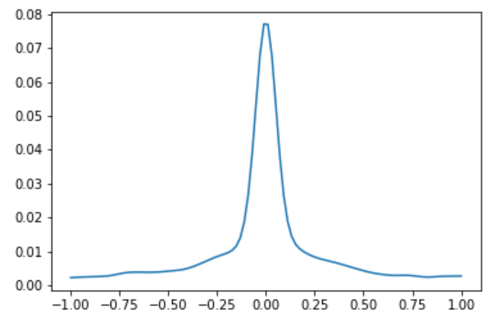
Return: Λ_1, gap.

F Results on the Gowalla dataset

We propose another application of our algorithm on a friendship dataset provided in [13] and collected from Gowalla, a popular location-based social network where users share their locations by checking-in. Our graph contains $6,442,890$ nodes, corresponding to all the check-ins between February 2009 and October 2010, ordered by timestamps. Two nodes in our graph are connected if the corresponding users are friends on Gowalla. Results are provided with Figure 7 working with a dimension $d = 3$. Figure 7.(a) presents the estimated envelope function while (b) shows the estimated latitude function. Figure 7.(b) looks like a centered normal distribution with a small variance. Therefore, we cannot conclude that there is a dependence between the latent representations of two consecutive nodes in the graph.



(a) Envelope function.



(b) Latitude function.

Figure 7: Presentation of the results working with the dataset from [13]. On a graph with $6,442,890$ nodes, we run the algorithm SCCHEi with $d = 3$ and $\hat{R} = 2$ to get a clustering of the largest $\tilde{R}_{opt}$ eigenvalues of $\hat{T}_n$. Figure (a) shows the estimated function p which was first rescaled (in order to have 1 as maximum value on $[0, 1]$) and then thresholded (to have positive values). For the latitude function, we removed the tail of the distribution by considering the 95% fraction of the $(\hat{r}_i)_i$ with the smallest magnitude and rescaling them so that the maximum magnitude is 1. The estimated latitude function is presented with Figure 7 (b).

G Concentration inequality for U-statistics with Markov chains

In this section, we prove a concentration inequality for a U-statistic of the Markov chain $(X_i)_{i\geq 1}$ which is a key result to prove Theorem 1. In the first subsection, we describe precisely the assumptions made on the Markovian dynamic, namely Assumption A and Assumption B. Those conditions allow us to fall into the framework of [1, Section 3] and to use a Talagrand's concentration for Markov chains.

G.1 Assumptions and notations for the Markov chain

Assumption A The latitude function $f_{\mathcal{L}}$ is bounded away from 0 and $\|f_{\mathcal{L}}\|_\infty < \infty$.
Assumption A guarantees that there exist $\delta_m, \delta_M > 0$ such that

$$\forall x \in \mathbb{S}^{d-1}, \forall A \in \mathcal{B}(\mathbb{S}^{d-1}), \quad \delta_m \nu(A) \leq P(x, A) \leq \delta_M \nu(A),$$

for some probability measure ν (e.g. the uniform measure on the sphere π). Assumption A also implies that there exists a *small set*, namely there exists a set $C \in B(S^{d-1})$ such that

$$\forall x \in \mathbb{S}^{d-1},\ \exists m, \quad P^m(x, C) > 0.$$

One can take $m = 1$ and $C = \mathbb{S}^{d-1}$. The latter has the important consequence that the Markov chain $(X_i)_{i\geq 1}$ is **uniformly ergodic** (see [11, section 1]), with associated constants $L > 0$ and $0 < \rho < 1$ (see Definition 5).

We introduce some additional notations which will be useful for the proof of the next subsection in particular to apply Talagrand concentration result from [1]. We extend the Markov chain $(X_i)_{i\geq 1}$ to a new (so called *split*) chain $(\tilde{X}_n, R_n) \in \mathbb{S}^{d-1} \times \{0,1\}$, satisfying the following properties.

- $(\tilde{X}_n)_n$ is again a Markov chain with transition kernel P and initial distribution π We recall that π is the uniform distribution on the $\mathbb{S}^{d-1}$.
- if we define $T_1 = \inf\{n > 0 : R_n = 1\}$,
$$T_{i+1} = \inf\{n > 0 : R_{T_1+\dots+T_i+n} = 1\},$$
then $T_1, T_2, \dots$ are well defined, independent moreover $T_2, T_3, \dots$ are i.i.d.
- if we define $S_i = T_1 + \dots + T_i$, then the "blocks"
$$Y_0 = (X_1, \dots, X_{T_1}),$$
$$Y_i = (X_{S_i+1}, \dots, X_{S_{i+1}}), i > 0,$$
form a one-dependent sequence (i.e. for all i, $\sigma((Y_j)_{j<i})$ and $\sigma((Y_j)_{j>i})$ are independent). Moreover, the sequence $Y_1, Y_2, \dots$ is stationary and the variables $Y_0, Y_1, \dots$ are independent. In consequence, for $f : \mathbb{S}^{d-1} \to \mathbb{R}$, the variables
$$Z_i = Z_i(f) = \sum_{j=S_i+1}^{S_{i+1}} f(X_j), \quad i \geq 1,$$
constitute an i.i.d. sequence.
- the distribution of T_1 depends only on π, P, C, δ_m, ν, whereas the law of T_2 only on P, C, δ_m and ν.

To derive concentration of measure inequality, we assume the exponential integrability of the regeneration time described with Assumption B.

Assumption B $\|T_1\|_{\psi_1} < \infty$ and $\|T_2\|_{\psi_1} < \infty$, where $\|\cdot\|_{\psi_1}$ is the 1-Orlicz norm introduced in Definition 8.

Definition 8 *For $\alpha > 0$, define the function $\psi_\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ with the formula $\psi_\alpha(x) = \exp(x\alpha) - 1$. Then for a random variable X, the α-Orlicz norm is given by*
$$\|X\|_{\psi_\alpha} = \inf\left\{\lambda > 0 : \mathbb{E}[\psi_\alpha(|X|/\lambda)] \leq 1\right\}.$$

G.2 Concentration inequality of U-statistic for Markov chain

One key result to prove Theorem 1 is the concentration of the following U-statistic
$$U_{stat}(n) = \sum_{1\leq i<j\leq n} \left[(W - W_R)^2(X_i, X_j) - \|W - W_R\|_2^2\right].$$

Lemma 6 *Let us consider $\gamma \in (0,1)$ satisfying $\log(e/\gamma) \leq n/\log^2 n$. Then it holds with probability at least $1-\gamma$,*
$$U_{stat}(n) \leq M \frac{\|p - p_R\|_\infty^2 \log n}{n} \left(\log(e/\gamma) \vee \log n\right),$$
where $M > 0$ only depends on constants related to the Markov chain $(X_i)_{i\geq 1}$, namely $\delta_m, \delta_M, \|T_1 - T_1\|$, $\mathbb{E}T_2, \|T_1\|_{\psi_1}, \|T_2\|_{\psi_1}, L$ and ρ.

Note that $\|W - W_R\|_2^2$ corresponds to the expectation of the kernel $(W - W_R)^2(\cdot,\cdot)$ under the uniform distribution on $\mathbb{S}^{d-1}$ which is known to be the unique invariant distribution π of the Markov chain $(X_i)_{i\geq 1}$ (see Appendix B). More precisely, for any $x \in \mathbb{S}^{d-1}$, it holds

$$\|W - W_R\|_2^2 = \mathbb{E}_{X\sim\pi}[(W - W_R)^2(x, X)] = \mathbb{E}_{(X,X')\sim\pi\otimes\pi}[(W - W_R)^2(X, X')],$$

see Lemma 5 for a proof. Our proof is inspired from [16, Section 3.4.3] where a Bernstein-type inequality is shown for U-statistics of order 2 in the independence setting. Their proof relies on the *canonical* property of the kernel functions which endowed the U-statistic with a martingale structure. We want to use a similar argument and we decompose $U_{stat}(n)$ to recover the martingale property for each term (except for the last one). Considering for any $l \geq 1$ the σ-algebra $G_l = \sigma(X_1, \ldots, X_l)$, the notation $\mathbb{E}_l$ refers to the conditional expectation with respect to G_l. Then we decompose $U_{stat}(n)$ as follows,

$$\begin{aligned} U_{stat}(n) = &\sum_{k=1}^{t_n}\sum_{i<j}\left(\mathbb{E}_{j-k+1}[(W - W_R)^2(X_i, X_j)] - \mathbb{E}_{j-k}[(W - W_R)^2(X_i, X_j)]\right) \\ &+ \sum_{i<j}\left(\mathbb{E}_{j-t_n}[(W - W_R)^2(X_i, X_j)] - \|W - W_R\|_2^2\right), \end{aligned} \tag{15}$$

where t_n is an integer that scales logarithmically with n and that will be specified latter. By convention, we assume here that for all $k < 1$, $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot]$. Hence the first term that we will consider is given by

$$U_n = \sum_{1\leq i<j\leq n} h(X_i, X_{j-1}, X_j),$$

where for all $x, y, z \in \mathbb{S}^{d-1}$,

$$h(x, y, z) = (W - W_R)^2(x, z) - \int_w (W - W_R)^2(x, w)P(y, dw).$$

We provide a detailed proof of a concentration result for U_n by taking advantage of its martingale structure following the work of [16, Section 3.4.3]. Reasoning by induction, we show that the $t_n - 1$ following terms involved in the decomposition (15) of $U_{stat}(n)$ can be handled using a similar approach. Since the last term of the decomposition (15) has not a martingale property, another argument is required. We deal with the last term exploiting the uniform ergodicity of the Markov chain $(X_i)_{i\geq 1}$ which is guaranteed by Assumption A (see the previous Subsection or [27, Theorem 8]).

G.3 Concentration of the first term of the decomposition of the U-statistic

Martingale structure of the U-statistic Defining $Y_j = \sum_{i=1}^{j-1} h(X_i, X_{j-1}, X_j)$, U_n can be written as $U_n = \sum_{j=2}^{n} Y_j$. Since

$$\mathbb{E}_{j-1}[Y_j] = \mathbb{E}[Y_j \mid X_1, \ldots, X_{j-1}] = 0,$$

we know that $(U_k)_{k\geq 2}$ is a martingale relative to the σ-algebras G_l, $l \geq 2$. This martingale can be extended to $n = 0$ and $n = 1$ by taking $U_0 = U_1 = 0$, $G_0 = \{\emptyset, \mathbb{S}^{d-1}\}$, $G_1 = \sigma(X_1)$. We will use the martingale structure of $(U_n)_n$ through the following Lemma.

Lemma 7 *(see [16, Lemma 3.4.6])*
Let $(U_m, G_m), m \in \mathbb{N}$, be a martingale with respect to a filtration G_m such that $U_0 = U_1 = 0$. For each $m \geq 1$ and $k \geq 2$, define the angle brackets $A_m^k = A_m^k(U)$ of the martingale U by

$$A_m^k = \sum_{i=1}^{m} \mathbb{E}_{i-1}[(U_i - U_{i-1})^k]$$

(and note $A_1^k = 0$ for all k). Suppose that for $\lambda > 0$ and all $i \geq 1$, $\mathbb{E}[e^{\lambda|U_i - U_{i-1}|}] < \infty$. Then

$$\left(\epsilon_m := e^{\lambda U_m - \sum_{k\geq 2} \lambda^k A_m^k / k!}, G_m\right), m \in \mathbb{N},$$

is a supermartingale. In particular, $\mathbb{E}[\epsilon_m] \leq \mathbb{E}[\epsilon_1] = 1$, so that, if $A_m^k \leq w_m^k$ for constants $w_m^k \geq 0$; then

$$\mathbb{E}[e^{\lambda U_m}] \leq e^{\sum_{k\geq 2} \lambda^k w_m^k / k!}.$$

For all $k \geq 2$ and $n \geq 1$, we have :

$$A_n^k = \sum_{j=2}^{n} \mathbb{E}_{j-1}\left[\sum_{i=1}^{j-1} h(X_i, X_{j-1}, X_j)\right]^k \leq V_n^k := \sum_{j=2}^{n} \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1} h(X_i, X_{j-1}, X_j)\right|^k$$
$$\leq \delta_M \sum_{j=2}^{n} \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1} h(X_i, X_{j-1}, X_j')\right|^k,$$

where the random variables $(X_j')_j$ are i.i.d. with distribution ν.

Lemma 8 *(see [16, Ex.1 Section 3.4]) Let Z_i be independent random variables with respective probability laws P_i. Let $k > 1$, and consider the space $\mathbb{H} = \{(f_1(Z_1), \ldots, f_N(Z_N)) \ : \ f_i \in L^k(P_i)\}$. Then the duality of L^p spaces and the independence of the variables Z_i imply that*

$$\left(\sum_{i=1}^{N} \mathbb{E}\left[|f_i(Z_i)|^k\right]\right)^{1/k} = \sup_{\sum_{i=1}^{N} \mathbb{E}|\xi_i(Z_i)|^{k/(k-1)}=1} \sum_{i=1}^{N} \mathbb{E}\left[f_i(Z_i)\xi_i(Z_i)\right],$$

where the sup runs over $\xi_i \in L^{k/(k-1)}(P_i)$.

Then by the duality result of Lemma 8,

$$(V_n^k)^{1/k} \leq \delta_M \sup_{\xi\in\mathcal{L}_k} \sum_{j=2}^{n}\sum_{i=1}^{j-1} \mathbb{E}_{j-1}\left[h(X_i, X_{j-1}, X_j')\xi_j(X_j')\right]$$

$$\textit{where } \mathcal{L}_k = \left\{\xi = (\xi_2, \ldots, \xi_n) \textit{ s.t. } \forall 2 \leq j \leq n,\ \xi_j \in L^{k/(k-1)}(\nu) \textit{ with } \sum_{j=2}^{n} \mathbb{E}|\xi_j(X_j')|^{k/(k-1)} = 1\right\}.$$

$$= \delta_M \sup_{\xi\in\mathcal{L}_k} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathbb{E}_{j-1}\left[h(X_i, X_{j-1}, X_j')\xi_j(X_j')\right]$$

If we define the random vectors $\mathbf{X}_i$ for $i = 1, \ldots, n-1$ on $\mathbb{R}^n$ by

$$\mathbf{X}_i = (0, \ldots, 0, h(X_i, X_i, x_{i+1}), \ldots, h(X_i, X_{n-1}, x_n)),$$

and for $\xi = (\xi_2, \ldots, \xi_n) \in \prod_{i=2}^{n} L^{k/(k-1)}(\nu)$, the function $f_\xi(h_2, \ldots, h_n) = \sum_{j=2}^{n} \int h_j(x)\xi_j(x)d\nu(x)$, then setting $\mathcal{F} = \{f_\xi : \sum_{j=2}^{n} \mathbb{E}|\xi_j(X_j')|^{k/(k-1)} = 1\}$, we have

$$\left(V_n^k\right)^{1/k} \leq \delta_M \sup_{f\in\mathcal{F}} \left|\sum_{i=1}^{n-1} f_\xi(\mathbf{X_i})\right|.$$

Using a similar approach, one can prove that

$$\delta_m \sup_{f\in\mathcal{F}} \left|\sum_{i=1}^{n-1} f_\xi(\mathbf{X_i})\right| \leq \left(V_n^k\right)^{1/k}.$$

By the separability of the L^p spaces of finite measures, $\mathcal{F}$ can be replaced by a countable subset $\mathcal{F}_0$. To upper-bound the tail probabilities of U_n, we will bound the variable V_n^k on sets of large probability using Talagrand's inequality. Then we will use Lemma 7 on these sets by means of optional stopping.

Application of Talagrand's inequality for Markov chains Let us denote

$$Z = \sup_{f\in\mathcal{F}} \left| \sum_{i=1}^{n-1} f_\xi(\mathbf{X_i}) \right|$$

and let us define the asymptotic weak variance

$$\sigma_k^2 = \sup_{f\in\mathcal{F}} \mathrm{Var} Z_1(f)/\mathbb{E}T_2.$$

We also consider $\tau = \|T_1\|_{\psi_1} \vee \|T_2\|_{\psi_1}$ and $b_k = \sup_{f\in\mathcal{F}} \|f\|_\infty$.
Now, we define for all $i \in [n]$, $V_i = (0, \dots, 0, X_{n-i+1}, \dots, X_n) \in \mathbb{R}^n$. Then, thanks to the reversibility of the Markov chain $(X_i)_{i\geq 1}$, V_i is a Markov chain with transition kernel directly given by the transition kernel of $(X_i)_{i\geq 1}$: P. Since $\mathbf{X}_i$ is $\sigma(V_{n-i+1})$ measurable, we can apply Theorem 7 from [1], and we get that for any $t \geq 0$,

$$\mathbb{P}\left(Z \geq K\mathbb{E}[Z] + t\right) \leq K \exp\left(-\frac{1}{K}\min\left(\frac{t^2}{n\sigma_k^2}, \frac{t}{\tau^3(\mathbb{E}T_2)^{-1} b_k \log n}\right)\right).$$

We deduce that for any $t \geq 1$,

$$\mathbb{P}\left(\left(V_n^k\right)^{1/k} \geq \frac{\delta_M}{\delta_m} K\mathbb{E}[\left(V_n^k\right)^{1/k}] + \delta_M t\right) \leq K \exp\left(-\frac{1}{K}\min\left(\frac{t^2}{n\sigma_k^2}, \frac{t}{\tau^3(\mathbb{E}T_2)^{-1} b_k \log n}\right)\right),$$

which implies that for any $x \geq 0$,

$$\mathbb{P}\left(\left(V_n^k\right)^{1/k} \geq \frac{\delta_M}{\delta_m} K\mathbb{E}[\left(V_n^k\right)^{1/k}] + \delta_M\sqrt{n\sigma_k^2 x} + \delta_M x\tau^3(\mathbb{E}T_2)^{-1} b_k \log n\right) \leq K\exp\left(-\frac{x}{K}\right).$$

K is a universal constant that we will assume equal to 1 to simplify notations. Using the change of variable $x = ku$ with $u \geq 0$ in the previous inequality leads to

$$\mathbb{P}\left(\bigcup_{k=2}^{\infty} \left(V_n^k\right)^{1/k} \geq \frac{\delta_M}{\delta_m} \mathbb{E}[\left(V_n^k\right)^{1/k}] + \delta_M\sqrt{n\sigma_k^2 ku} + \delta_M ku\tau^3(\mathbb{E}T_2)^{-1} b_k \log n\right) \leq 1.62e^{-u},$$

because

$$1 \wedge \sum_{k=2}^{\infty} \exp(-ku) \leq 1 \wedge \frac{1}{e^u(e^u-1)} = \left(e^u \wedge \frac{1}{e^u - 1}\right) e^{-u} \leq \frac{1+\sqrt{5}}{2} e^{-u} \leq 1.62e^{-u}.$$

Bounding b_k. Using Hölder's inequality we have,

$$\begin{aligned}
b_k &= \sup_{\sum_{j=2}^n \mathbb{E}|\xi_j(X_j')|^{k/(k-1)}=1} \max_i \sup_{x,y} \left| \sum_{j=i+1}^{n} \mathbb{E}\left[h(x,y,X_j')\xi_j(X_j')\right] \right| \\
&\leq \sup_{\sum_{j=2}^n \mathbb{E}|\xi_j(X_j')|^{k/(k-1)}=1} \max_i \sup_{x,y} \sum_{j=i+1}^{n} \left(\mathbb{E}\left|h(x,y,X_j')\right|^k\right)^{1/k} \left(\mathbb{E}\left|\xi_j(X_j')\right|^{k/(k-1)}\right)^{(k-1)/k} \\
&\leq \sup_{\sum_{j=2}^n \mathbb{E}|\xi_j(X_j')|^{k/(k-1)}=1} \max_i \sup_{x,y} \left(\sum_{j=i+1}^{n} \mathbb{E}\left|h(x,y,X_j')\right|^k\right)^{1/k} \left(\sum_{j=i+1}^{n} \mathbb{E}\left|\xi_j(X_j')\right|^{k/(k-1)}\right)^{(k-1)/k} \\
&\leq \max_i \sup_{x,y} \left(\sum_{j=i+1}^{n} \mathbb{E}\left|h(x,y,X_j')\right|^k\right)^{1/k} \\
&\leq (B^2 A^{k-2})^{1/k},
\end{aligned}$$

where

$$A := \|h\|_\infty \leq 2\|p - p_R\|_\infty^2$$

and

$$B^2 := nb^2 \text{ where } b^2 = \max\left\{\left\|\mathbb{E}_{X'\sim\nu}\left[h^2(\cdot,\cdot,X')\right]\right\|_\infty, \left\|\mathbb{E}_{X''\sim\pi}\left[h^2(X'',\cdot,\cdot)\right]\right\|_\infty\right\}.$$

Bounding the variance.

$$\begin{aligned}
\mathbb{E}[T_2]\sigma_k^2 &= \sup_{f\in\mathcal{F}} \mathrm{Var}\left(\sum_{i=T_1+1}^{T_2} f(\mathbf{X}_i)\right)\\
&= \sup_{\xi\in\mathcal{L}_k} \mathrm{Var}\left[\sum_{i=T_1+1}^{T_2}\sum_{j=i+1}^{n} \mathbb{E}_{j-1}[h(X_i, X_{j-1}, X_j')\xi_j(X_j')]\right]\\
&\leq 2\left(B^2A^{k-2}\right)^{2/k}\mathbb{E}[(T_2-T_1)^2],
\end{aligned}$$

where the last inequality comes from the following (where we use twice Holder's inequality),

$$\begin{aligned}
& \sup_{\xi\in\mathcal{L}_k}\mathbb{E}\left[\left(\sum_{i=T_1+1}^{T_2}\sum_{j=i+1}^{n}\mathbb{E}_{j-1}[h(X_i, X_{j-1}, X_j')\xi_j(X_j')]\right)^2\right]\\
\leq\ & \sup_{\xi\in\mathcal{L}_k}\mathbb{E}\left[\left(\sum_{i=T_1+1}^{T_2}\sum_{j=i+1}^{n}\mathbb{E}_{j-1}[|h(X_i, X_{j-1}, X_j')|^k]^{1/k}\mathbb{E}[|\xi_j(X_j')|^{k/(k-1)}]^{(k-1)/k}\right)^2\right]\\
\leq\ & \sup_{\xi\in\mathcal{L}_k}\mathbb{E}\left[\left(\sum_{i=T_1+1}^{T_2}\left(\sum_{j=i+1}^{n}\mathbb{E}_{j-1}[|h(X_i, X_{j-1}, X_j')|^k]\right)^{1/k}\left(\sum_{j=i+1}^{n}\mathbb{E}[|\xi_j(X_j')|^{k/(k-1)}]\right)^{(k-1)/k}\right)^2\right]\\
\leq\ & \mathbb{E}\left[\left(\sum_{i=T_1+1}^{T_2}\left(\sum_{j=i+1}^{n}\mathbb{E}_{j-1}[|h(X_i, X_{j-1}, X_j')|^k]\right)^{1/k}\right)^2\right]\\
\leq\ & \left(B^2A^{k-2}\right)^{2/k}\mathbb{E}[(T_2-T_1)^2].
\end{aligned}$$

Now notice that for all $\theta_1, \theta_2 \geq 0$ and $0 < \epsilon \leq 1$ by convexity,

$$\left(\frac{\theta_1+\theta_2}{1+\epsilon}\right)^k = \left(\frac{\theta_1}{1+\epsilon} + \frac{\epsilon\epsilon^{-1}\theta_2}{1+\epsilon}\right)^k \leq \frac{1}{1+\epsilon}\theta_1^k + \frac{\epsilon}{1+\epsilon}\epsilon^{-k}\theta_2^k,$$

so that

$$(\theta_1+\theta_2)^k \leq (1+\epsilon)^{k-1}\theta_1^k + \epsilon^{-(k-1)}(1+\epsilon)^{k-1}\theta_2^k = (1+\epsilon)^{k-1}\theta_1^k + (1+\epsilon^{-1})^{k-1}\theta_2^k.$$

By symmetry, this inequality holds for all $\epsilon \geq 0$, that is, for all $\theta_1, \theta_2, \epsilon \geq 0$,

$$(\theta_1+\theta_2)^k \leq (1+\epsilon)^{k-1}\theta_1^k + (1+\epsilon^{-1})^{k-1}\theta_2^k.$$

Using this inequality twice and the bounds obtained on b_k and σ_k^2 gives for $u > 0$,

$$\left[\frac{\delta_M}{\delta_m}\left(\mathbb{E}V_n^k\right)^{1/k} + \delta_M\sqrt{n\sigma_k^2ku} + \delta_M ku\tau^3(\mathbb{E}T_2)^{-1}b_k\log n\right]^k$$

$$
\begin{aligned}
&\le \left[\frac{\delta_M}{\delta_m}\left(\mathbb{E}V_n^k\right)^{1/k} + \sqrt{2}\delta_M(B^2A^{k-2})^{1/k}\|T_2-T_1\|_2(\mathbb{E}T_2)^{-1/2}\sqrt{nku}\right. \\
&\qquad \left. + \delta_M ku\tau^3(\mathbb{E}T_2)^{-1}(B^2A^{k-2})^{1/k}\log n\right]^k \\
&\le (1+\epsilon)^{k-1}\left(\frac{\delta_M}{\delta_m}\right)^k \mathbb{E}V_n^k + (1+\epsilon^{-1})^{k-1}\left[\delta_M ku\tau^3(\mathbb{E}T_2)^{-1}(B^2A^{k-2})^{1/k}\log n\right. \\
&\qquad \left. + \sqrt{2}\delta_M(B^2A^{k-2})^{1/k}\|T_2-T_1\|_2(\mathbb{E}T_2)^{-1/2}\sqrt{nku}\right]^k \\
&\le (1+\epsilon)^{k-1}\left(\frac{\delta_M}{\delta_m}\right)^k \mathbb{E}V_n^k + (1+\epsilon^{-1})^{2k-2}(\delta_M ku)^k\tau^{3k}(\mathbb{E}T_2)^{-k}B^2A^{k-2}\log^k n \\
&\qquad + (1+\epsilon)^{k-1}(1+\epsilon^{-1})^{k-1}\delta_M^k B^2A^{k-2}\|T_2-T_1\|_2^k(\mathbb{E}T_2)^{-k/2}(2nku)^{k/2}.
\end{aligned}
$$

So, setting

$$
\begin{aligned}
w_n^k := (1+\epsilon)^{k-1}\left(\frac{\delta_M}{\delta_m}\right)^k \mathbb{E}V_n^k + (1+\epsilon^{-1})^{2k-2}(\delta_M ku)^k\tau^{3k}(\mathbb{E}T_2)^{-k}B^2A^{k-2}\log^k n \\
+ (2+\epsilon+\epsilon^{-1})^{k-1}\delta_M^k B^2A^{k-2}\|T_2-T_1\|_2^k(\mathbb{E}T_2)^{-k/2}(2nku)^{k/2},
\end{aligned}
$$

we have

$$
\mathbb{P}\left(V_n^k \le w_n^k \quad \forall k\ge 2\right) \ge 1 - 1.62e^{-u}, \tag{16}
$$

where the dependence in u of w_n^k is leaved implicit.

Upper-bounding U_n using the martingale structure Let

$$
T+1 := \inf\{l\in\mathbb{N} : V_l^k \ge w_l^k \text{ for some } k\ge 2\}.
$$

Then, the event $\{T\le l\}$ depends only on $X_1,\dots,X_l$ for all $l\ge 1$. Hence, T is a stopping time for the filtration $(\mathcal{G}_l)_l$ where $\mathcal{G}_l = \sigma((X_i)_{i\in[l]})$ and we deduce that $U_l^T := U_{l\wedge T}$ for $l=0,\dots,n$ is a martingale with respect to $(\mathcal{G}_l)_l$ with $U_0^T = U_0 = 0$ and $U_1^T = U_1 = 0$. We remark that $U_j^T - U_{j-1}^T = U_j - U_{j-1}$ if $T\ge j$ and zero otherwise, and that $\{T\ge j\}$ is $\mathcal{G}_{j-1}$ measurable. Then, the angle brackets of this martingale admit the following bound:

$$
\begin{aligned}
A_n^k(U^T) &= \sum_{j=2}^n \mathbb{E}_{j-1}[(U_j^T - U_{j-1}^T)^k] \\
&\le V_n^k(U^T) \\
&= \sum_{j=2}^n \mathbb{E}_{j-1}|U_j - U_{j-1}|^k \mathbb{1}_{T\ge j} \\
&= \sum_{j=2}^n \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1} h(X_i, X_{j-1}, X_j)\right|^k \mathbb{1}_{T\ge j} \\
&= \sum_{j=2}^{n-1} V_j^k \mathbb{1}_{T=j} + V_n^k\mathbb{1}_{T\ge n} \\
&\le w_n^k\left(\sum_{j=2}^{n-1}\mathbb{1}_{T=j} + \mathbb{1}_{T\ge n}\right)
\end{aligned}
$$

$$\leq w_n^k,$$

since, by definition of T, $V_j^k \leq w_n^k$ for all k on $\{T \geq j\}$. Hence, Lemma 7 applied to the martingale U_n^T implies

$$\mathbb{E}e^{\lambda U_n^T} \leq \exp\left(\sum_{k\geq 2} \frac{\lambda^k}{k!} w_n^k\right).$$

Also, since V_n^k is nondecreasing in n for each k, inequality (16) implies that

$$\mathbb{P}(T < n) \leq \mathbb{P}\left(V_n^k \geq w_n^k \quad \text{for some } k \geq 2\right) \leq 1.62e^{-u}.$$

Thus we deduce that for all $s \geq 0$,

$$\mathbb{P}(U_n \geq s) \leq \mathbb{P}(U_n^T \geq s, T \geq n) + \mathbb{P}(T < n) \leq e^{-\lambda s} \exp\left(\sum_{k\geq 2} \frac{\lambda^k}{k!} w_n^k\right) + 1.62e^{-u}. \tag{17}$$

The final step of the proof consists in simplifying $\exp\left(\sum_{k\geq 2} \frac{\lambda^k}{k!} w_n^k\right)$.

$$\begin{aligned}
\sum_{k\geq 2} \frac{\lambda^k}{k!} w_n^k &= \sum_{k\geq 2} \frac{\lambda^k}{k!} (1+\epsilon)^{k-1} \left(\frac{\delta_M}{\delta_m}\right)^k \mathbb{E}V_n^k \\
&\quad + \sum_{k\geq 2} \frac{\lambda^k}{k!} (2+\epsilon+\epsilon^{-1})^{k-1} \delta_M^k B^2 A^{k-2} \|T_2 - T_1\|_2^k (\mathbb{E}T_2)^{-k/2} (2nku)^{k/2} \\
&\quad + \sum_{k\geq 2} \frac{\lambda^k}{k!} (1+\epsilon^{-1})^{2k-2} (\delta_M ku)^k \tau^{3k} (\mathbb{E}T_2)^{-k} B^2 A^{k-2} \log^k n \\
&:= a_1 + a_2 + a_3.
\end{aligned}$$

Bounding a_3. Using the inequality $k! \geq (k/e)^k$, we have noting $\delta(\epsilon) := e(1+\epsilon^{-1})^2 (\mathbb{E}T_2)^{-1} \tau^3 \delta_M$,

$$\begin{aligned}
a_3 &\leq \sum_{k\geq 2} \lambda^k (1+\epsilon^{-1})^{2k-2} (\delta_M eu)^k \tau^{3k} (\mathbb{E}T_2)^{-k} B^2 A^{k-2} \log^k n \\
&= \sum_{k\geq 2} \lambda^k (1+\epsilon^{-1})^{-2} \delta(\epsilon)^k u^k B^2 A^{k-2} \log^k n \\
&= \lambda^2 (1+\epsilon^{-1})^{-2} \delta(\epsilon)^2 u^2 B^2 \log^2 n \sum_{k\geq 0} \lambda^k \delta(\epsilon)^k u^k A^k \log^k n \\
&= \frac{\left(B(1+\epsilon^{-1})^{-1} u\delta(\epsilon) \log n\right)^2 \lambda^2}{1 - u\lambda A\delta(\epsilon) \log n},
\end{aligned}$$

for $\lambda < (uA\delta(\epsilon)\log n)^{-1}$.

Bounding a_2. We use the inequality $k! \geq k^{k/2}$ because $(k/e)^k > k^{k/2}$ for $k \geq e^2$ and for k smaller, the inequality follows by direct verification. Defining $\eta(\epsilon) := \sqrt{2}(2+\epsilon+\epsilon^{-1})\delta_M \|T_2 - T_1\| (\mathbb{E}T_2)^{-1/2}$, we have

$$\begin{aligned}
a_2 &\leq \sum_{k\geq 2} \lambda^k (2+\epsilon+\epsilon^{-1})^{k-1} \delta_M^k B^2 A^{k-2} \|T_2 - T_1\|_2^k (\mathbb{E}T_2)^{-k/2} (2nu)^{k/2} \\
&= \sum_{k\geq 2} \lambda^k (2+\epsilon+\epsilon^{-1})^{-1} \eta(\epsilon)^k B^2 A^{k-2} (nu)^{k/2} \\
&= (2+\epsilon+\epsilon^{-1})^{-1} \eta(\epsilon)^2 B^2 (nu) \lambda^2 \sum_{k\geq 0} \lambda^k \eta(\epsilon)^k A^k (nu)^{k/2} \\
&= \frac{(2+\epsilon+\epsilon^{-1})^{-1} \left(\lambda \eta(\epsilon) B\right)^2 nu}{1 - \lambda A \eta(\epsilon)\sqrt{nu}},
\end{aligned}$$

for $\lambda < (A\eta(\epsilon)\sqrt{nu})^{-1}$.

Bounding a_1. Let us recall that

$$\mathbb{E}[V_n^k] = \mathbb{E}\sum_{j=2}^n \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1} h(X_i, X_{j-1}, X_j)\right|^k = \mathbb{E}\sum_{j=2}^n \mathbb{E}\left[\left|\sum_{i=1}^{j-1} h(X_i, X_{j-1}, X_j)\right|^k \,\middle|\, X_{j-1}, X_j\right].$$

Thus denoting $\kappa = \frac{\delta_M}{\delta_m}$,

$$a_1 = \frac{1}{1+\epsilon}\mathbb{E}\sum_{j=2}^n \left(\mathbb{E}_{|X_{j-1},X_j}\left[e^{\lambda(1+\epsilon)\kappa|C_j|}\right] - \lambda(1+\epsilon)\kappa\mathbb{E}_{|X_{j-1},X_j}\left[|C_j|\right] - 1\right),$$

where $C_j = \sum_{i=1}^{j-1} h(X_i, X_{j-1}, X_j)$ and where the notation $\mathbb{E}_{|X_{j-1},X_j}$ refers to the expectation conditionally to the σ-algebra $\sigma(X_{j-1}, X_j)$.
Now we use a symmetrization trick: since $e^x - x - 1 \geq 0$ for all x and since $e^{a|x|} + e^{-a|x|} = e^{ax} + e^{-ax}$, adding $\mathbb{E}_{|X_{j-1},X_j}[\exp(-\lambda(1+\epsilon)\kappa|C_j|)] + \lambda(1+\epsilon)\kappa\mathbb{E}_{|X_{j-1},X_j}[|C_j|] - 1$ to a_1 gives

$$a_1 \leq \frac{1}{1+\epsilon}\mathbb{E}\sum_{j=2}^n \left(\mathbb{E}_{|X_{j-1},X_j}[e^{\lambda(1+\epsilon)\kappa C_j}] - 1 + \mathbb{E}_{|X_{j-1},X_j}[e^{-\lambda(1+\epsilon)\kappa C_j}] - 1\right). \tag{18}$$

Let us consider some $j \in \{2, \ldots, n\}$. Conditionally on $\sigma(X_{j-1}, X_j)$, C_j is a sum of bounded functions (by A) depending on the Markov chain [2]. We denote

$$v_j(X_{j-1}, X_j) = \sum_{i=1}^{j-1} \mathbb{E}_{X_i\sim\pi}[h^2(X_i, X_{j-1}, X_j)|X_{j-1}, X_j] \leq B^2$$

and $V = \sum_{j=2}^n \mathbb{E}v_j^k(X_{j-1}, X_j) \leq C^2 B^{2(k-1)}$ (with $C^2 = nB^2$).
Remark that

$$\begin{aligned}
& \mathbb{E}_{X_i\sim\pi}[h(X_i, X_{j-1}, X_j)|X_{j-1}, X_j] \\
= \;& \mathbb{E}_{X_i\sim\pi}\left[(W - W_R)^2(X_i, X_j) - \mathbb{E}_{j-1}[(W - W_R)^2(X_i, X_j)] \mid X_{j-1}, X_j\right] \\
= \;& \int_{x_i} (W - W_R)^2(x_i, X_j)\pi(dx_i) \\
& - \int_{x_i}\int_{x_j} (W - W_R)^2(x_i, x_j)\pi(dx_i)P(X_{j-1}, dx_j) \\
= \;& \|W - W_R\|_2^2 - \|W - W_R\|_2^2 \\
= \;& 0,
\end{aligned}$$

[2]Strictly speaking, we consider the time reversal Markov chain starting at X_{j-1} which has the same kernel due to the reversibility of the chain.

where we used that

$$\forall x \in \mathbb{S}^{d-1}, \quad \mathbb{E}_{X\sim\pi}[(W-W_R)^2(X,x)] = E_{X\sim\pi}[(W-W_R)^2(x,X)] = \|W-W_R\|_2^2,$$

as stated by Lemma 5. We will use a Bernstein inequality for Markov chain (see [19]). Note that the time reversal Markov chain with kernel P starting at X_{j-1}, that we consider here, admits the uniform distribution on $\mathbb{S}^{d-1}$ has invariant measure. Moreover, its spectral gap is equal to 1 (see Appendix B). Finally, notice from Taylor expansion that $(1-p/3)(e^p-p-1) \leq p^2/2$ for all $p \geq 0$. We can now apply [19, Eq.(4.5) Thm 1.1] with $t = \lambda(1+\epsilon)\kappa$ and $c = A$ (using their notations). We get that for $\lambda < [(1+\epsilon)\kappa(A/3+B/\sqrt{2})]^{-1}$,

$$\begin{aligned}
&\mathbb{E}_{|X_{j-1},X_j}[e^{\lambda(1+\epsilon)\kappa|C_j|}]\\
\leq\ & \mathbb{E}_{|X_{j-1},X_j}\left[\exp\left(\frac{\lambda^2(1+\epsilon)^2\kappa^2 v_j(X_{j-1},X_j)}{2-2A\lambda(1+\epsilon)\kappa/3}\right)\right]\\
\leq\ & \mathbb{E}_{|X_{j-1},X_j}\left[\exp\left(\frac{\lambda^2(1+\epsilon)^2\kappa^2 v_j(X_{j-1},X_j)}{2-2A\lambda(1+\epsilon)\kappa/3}\right)\right].
\end{aligned}$$

Considering $\lambda < [(1+\epsilon)\kappa(A/3+B/\sqrt{2})]^{-1}$, $\epsilon < 1$ and using (18), this leads to

$$\begin{aligned}
a_1 &\leq \frac{2}{1+\epsilon}\sum_{j=2}^{n}\mathbb{E}\left[\exp\left(\frac{\lambda^2(1+\epsilon)^2\kappa^2 v_j(X_{j-1},X_j)}{2-2A\lambda(1+\epsilon)\kappa/3}\right)-1\right]\\
&\leq \frac{2}{1+\epsilon}\sum_{k=1}^{\infty}\frac{\lambda^{2k}(1+\epsilon)^{2k}\kappa^{2k}V}{(2-2A\lambda(1+\epsilon)\kappa/3)^k}\\
&\leq \frac{2}{1+\epsilon}\sum_{k=1}^{\infty}\frac{\lambda^{2k}(1+\epsilon)^{2k}\kappa^{2k}C^2B^{2(k-1)}}{(2-2A\lambda(1+\epsilon)\kappa/3)^k}\\
&= \frac{(1+\epsilon)C^2\lambda^2\kappa^2}{1-A\lambda(1+\epsilon)\kappa/3-\lambda^2(1+\epsilon)^2\kappa^2B^2/2}\\
&\leq \frac{(1+\epsilon)C^2\lambda^2\kappa^2}{1-\lambda(1+\epsilon)\kappa(A/3+B/\sqrt{2})}.
\end{aligned}$$

Putting altogether we obtain

$$\exp\left(\sum_{k\geq 2}\frac{\lambda^k}{k!}w_n^k\right) \leq \exp\left(\frac{\lambda^2W^2}{1-\lambda c}\right),$$

where

$$W = (1+\epsilon)^{1/2}C\kappa + (1+\epsilon+\epsilon^{-1})^{-1/2}\eta(\epsilon)B\sqrt{nu} + B(1+\epsilon^{-1})^{-1}u\delta(\epsilon)\log n,$$

and

$$c = \max\left[(1+\epsilon)\kappa(A/3+B/\sqrt{2}),\, A\eta(\epsilon)\sqrt{nu},\, A\delta(\epsilon)u\log n\right].$$

Using this estimate in (17) and taking $s = 2W\sqrt{u}+cu$ and $\lambda = \sqrt{u}/(W+c\sqrt{u})$ in this inequality yields

$$\mathbb{P}\left(U_n \geq 2W\sqrt{u}+cu\right) \leq 2.62e^{-u}.$$

By taking $\epsilon = 1/2$, we deduce that for any $u \geq 0$, it holds with probability at least $1-e^{1-u}$

$$\begin{aligned}
&\sum_{i<j}h_{i,j}(X_i,X_j)\\
\leq\ & \sqrt{6}C\kappa\sqrt{u} + 6\delta_M\|T_2-T_1\|(\mathbb{E}T_2)^{-1/2}B\sqrt{n}u + 6e\delta_M(\mathbb{E}T_2)^{-1}\tau^3Bu^{3/2}\log n\\
&+\frac{\kappa}{2}(A+3B)u + 9A\delta_M\|T_2-T_1\|(\mathbb{E}T_2)^{-1/2}\sqrt{n}u^{3/2} + 12A(\mathbb{E}T_2)^{-1}\tau^3e\delta_Mu^2\log n,
\end{aligned}$$

Using

$$B \leq \sqrt{n}A, \quad C = \sqrt{n}B \leq nA \quad and \quad A \leq 2\|p - p_R\|_\infty^2,$$

it holds with probability at least $1 - e^{1-u}$,

$$\sum_{i<j} h_{i,j}(X_i, X_j) \leq A\left[M_0 n\sqrt{u} + M_1 nu + M_2\sqrt{n}\log(n)u^{3/2} + M_3\log(n)u^2\right]$$

with

$$M_0 = \sqrt{6}\kappa, \qquad M_1 = 6\delta_M\|T_2 - T_1\|(\mathbb{E}T_2)^{-1/2} + 3\kappa,$$
$$M_2 = 9\delta_M\|T_2 - T_1\|(\mathbb{E}T_2)^{-1/2} + 6e\delta_M(\mathbb{E}T_2)^{-1}\tau^3 e, \text{ and } M_3 = 12(\mathbb{E}T_2)^{-1}\tau^3\delta_M.$$

Let us consider some $\gamma \in (0,1)$. Then, denoting $M = 8\max(M_0, M_1, M_2, M_3)$, we have with probability at least $1-\gamma$,

$$\begin{aligned}
&\frac{2}{n(n-1)}\sum_{i<j} h_{i,j}(X_i, X_j) \\
\leq\quad & \frac{M}{2}\frac{\|p - p_R\|_\infty^2}{n(n-1)}\left[n\log\left(\frac{e}{\gamma}\right) \vee \sqrt{n}\log n\log\left(\frac{e}{\gamma}\right)^{3/2} \vee \log(n)\log\left(\frac{e}{\gamma}\right)^2\right]. \qquad (19)
\end{aligned}$$

In particular, for any $\gamma \in (0,1)$ satisfying $\log(e/\gamma) \leq n/\log^2 n$ it holds with probability at least $1-\gamma$,

$$\frac{2}{n(n-1)}\sum_{i<j} h_{i,j}(X_i, X_j) \leq M\frac{\|p - p_R\|_\infty^2}{n}\log\left(\frac{e}{\gamma}\right). \qquad (20)$$

G.3.1 Reasoning by descending induction with a logarithmic depth

As previously explained, we apply a proof similar the one of the previous subsection on the t_n first terms of the decomposition (15). Let us give the key elements to justify such approach by considering the second term of the decomposition (15), namely

$$\begin{aligned}
&\sum_{i<j}\left(\mathbb{E}_{j-1}\left[(W - W_R)^2(X_i, X_j)\right] - \mathbb{E}_{j-2}\left[(W - W_R)^2(X_i, X_j)\right]\right) \\
=\quad & \sum_{i=1}^{n-2}\sum_{j=i+2}^{n} h^{(1)}(X_i, X_{j-2}, X_{j-1}) + \sum_{i=1}^{n-1} g(X_i) \\
=\quad & \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1} h^{(1)}(X_i, X_{j-1}, X_j) + \sum_{i=1}^{n-1} g(X_i), \qquad (21)
\end{aligned}$$

where

$$h^{(1)}(x,y,z) = \int_w (W - W_R)^2(x,w)P(z,dw) - \int_w (W - W_R)^2(x,w)P^2(y,dw)$$

and

$$g(x) = \int_w (W - W_R)^2(x,w)P(x,dw) - \|W - W_R\|_2^2.$$

We can upper-bound directly $\sum_{i=1}^{n-1} g(X_i)$ by $2n\|p - p_R\|_\infty^2$ and we aim at proving a concentration result for the term

$$U_{n-1}^{(1)} := \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1} h^{(1)}(X_i, X_{j-1}, X_j) = \sum_{j=2}^{n-1}\sum_{i=1}^{j-1} h^{(1)}(X_i, X_{j-1}, X_j),$$

using an approach similar to the one of the previous subsection. One can use exactly the same sketch of proof.

- Martingale structure
  Using the notation $Y_j^{(1)} = \sum_{i=1}^{j-1} h^{(1)}(X_i, X_{j-1}, X_j)$, we have $U_{n-1}^{(1)} = \sum_{j=2}^{n-1} Y_j^{(1)}$ which shows that $(U_n^{(1)})_n$ is a martingale with respect to the σ-algebras G_l. Indeed, we have $\mathbb{E}_{j-1}[Y_j^{(1)}] = 0$.

- Talagrand's inequality
  One can use the same duality trick to show that the V_n^k can be controlled using a sum of functions of the Markov chain $(V_i)_{i\geq 1}$ (as defined in the previous section).

- Bounding $\exp(w_n^k \lambda^k / k!)$
  The terms a_2 and a_3 can be bounded in a similar way. For the term a_1, we only need to show that $h^{(1)}$ satisfies $\mathbb{E}_{X_i\sim\pi}|h^{(1)}(X_i, y, z)] = 0,\ \forall y, z \in \mathbb{S}^{d-1}$ in order to apply as previously a Bernstein's type inequality.

$$\begin{aligned}
&\mathbb{E}_{X_i\sim\pi}|h^{(1)}(X_i, y, z)] \\
&= \int_{x_i} \pi(dx_i) \left(\int_w (W - W_R)^2(x_i, w) P(z, dw) - \int_w (W - W_R)^2(x_i, w) P^2(y, dw) \right) \\
&= \int_w \left(P(z, dw) \int_{x_i} \pi(dx_i)(W - W_R)^2(x_i, w) - P^2(y, dw) \int_{x_i} \pi(dx_i)(W - W_R)^2(x_i, w) \right) \\
&= \|W - W_R\|_2^2 - \|W - W_R\|_2^2 \quad \text{(Using Lemma 5)} \\
&= 0.
\end{aligned}$$

- Conclusion of the proof
  The key remark is that the constants A and B are respectively bounded by $2\|p - p_R\|_\infty^2$ and $2n\|p - p_R\|_\infty^2$ as previously. This allows us to get a concentration inequality strictly similar to the one of the previous subsection, namely for any $\gamma \in (0, 1)$ satisfying $\log(e/\gamma) \leq n/\log^2 n$, it holds with probability at least $1 - \gamma$,

$$\frac{2}{n(n-1)} h^{(1)}(X_i, X_{j-1}, X_j) \leq M \frac{\|p - p_R\|_\infty^2}{n} \log\left(\frac{e}{\gamma}\right).$$

  Going back to (21), we get that for any $\gamma \in (0, 1)$ satisfying $\log(e/\gamma) \leq n/\log^2 n$ it holds with probability at least $1 - \gamma$,

$$\begin{aligned}
&\sum_{i<j} \left(\mathbb{E}_{j-1}\left[(W - W_R)^2(X_i, X_j)\right] - \mathbb{E}_{j-2}\left[(W - W_R)^2(X_i, X_j)\right] \right) \\
\leq\ & M \frac{\|p - p_R\|_\infty^2}{n} \log\left(\frac{e}{\gamma}\right) + 8 \frac{\|p - p_R\|_\infty^2}{n} \qquad (22)
\end{aligned}$$

One can do the same analysis for the t_n first terms in the decomposition (15). Hence for any $\gamma \in (0, 1)$ satisfying $\log(e/\gamma) \leq n/\log^2 n$, it holds with probability at least $1 - \gamma$,

$$\begin{aligned}
&\sum_{k=1}^{t_n} \sum_{i<j} \left(\mathbb{E}_{j-k+1}[(W - W_R)^2(X_i, X_j)] - \mathbb{E}_{j-k}[(W - W_R)^2(X_i, X_j)] \right) \\
\leq\ & M t_n \frac{\|p - p_R\|_\infty^2}{n} \log\left(\frac{e}{\gamma}\right) + 8(t_n - 1) \frac{\|p - p_R\|_\infty^2}{n}. \qquad (23)
\end{aligned}$$

G.3.2 Bounding the remaining statistic with uniform ergodicity

In the previous steps of the proof, we decompose U_n in $t_n + 1$ terms (see (15)). The martingale structure of the first t_n terms of this decomposition allowed us to derive a concentration inequality for each of them. It remains to control the last term of this decomposition, namely

$$\sum_{i<j} \left(\mathbb{E}_{j-t_n}\left[(W-W_R)^2(X_i,X_j)\right] - \|W-W_R\|_2^2\right),$$

where $t_n = \lfloor q \log n \rfloor$ with $0 < q < (\log(1/\rho))^{-1}$. Using our convention which states that for all $k < 1$, $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot]$, we need to control

$$\begin{aligned}
& \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\mathbb{E}_{j-t_n}\left[(W-W_R)^2(X_i,X_j)\right] - \|W-W_R\|_2^2\right) \\
= & \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}_{j-t_n}\left[(W-W_R)^2(X_i,X_j) - (W-W_R)^2(X_i,X_j')\right],
\end{aligned}$$

where $(X_j')_j$ are i.i.d random variables with distribution π (the uniform distribution on $\mathbb{S}^{d-1}$), and independent of $(X_i)_{i\geq 1}$ (see Lemma 5). We deduce that

$$\begin{aligned}
& \left|\sum_{i<j} \left(\mathbb{E}_{j-t_n}\left[(W-W_R)^2(X_i,X_j)\right] - \|W-W_R\|_2^2\right)\right| \\
\leq & \sum_{i=1}^{n-1} \sum_{j=(i+1)\vee(t_n+1)}^{n} \left|\mathbb{E}_{j-t_n}\left[(W-W_R)^2(X_i,X_j) - (W-W_R)^2(X_i,X_j')\right]\right| \\
& \quad + \sum_{i=1}^{n-1} \sum_{j=(i+1)}^{n\wedge t_n} \left|\mathbb{E}\left[(W-W_R)^2(X_i,X_j)\right] - \|W-W_R\|_2^2\right| \\
\leq & \ (1) + (2) + (3) + (4) + 2nt_n\|p-p_R\|_\infty^2,
\end{aligned}$$

with, denoting $H_{ij} = \mathbb{E}_{j-t_n}\left[(W-W_R)^2(X_i,X_j) - (W-W_R)^2(X_i,X_j')\right]$,

$$(1) := \sum_{i=1}^{t_n} \sum_{j=(t_n+1)}^{2t_n} |H_{ij}| \leq 2q^2 \log^2(n) \|W-W_R\|_\infty^2,$$

$$(2) := \sum_{i=1}^{t_n} \sum_{j=(2t_n+1)}^{n} |H_{ij}|,$$

$$(3) := \sum_{i=t_n+1}^{n-1} \sum_{j=i+1}^{n\wedge(t_n+i)} |H_{ij}| \leq 2qn\log(n)\|W-W_R\|_\infty^2,$$

$$(4) := \sum_{i=t_n+1}^{n-1} \sum_{j=t_n+i+1}^{n} |H_{ij}|.$$

Let us upper-bound (2) and (4) to conclude the proof. First note that for $i \leq j - t_n$, it holds

$$E_{j-t_n}\left[(W-W_R)^2(X_i,X_j)\right] = \int_{z\in\mathbb{S}^{d-1}} (W-W_R)^2(X_i,z) P^{t_n}(X_{j-t_n}, dz).$$

We start by upper-bounding (2),

$$
\begin{aligned}
(2) &= \sum_{i=1}^{t_n} \sum_{j=2t_n+1}^{n} \left| \mathbb{E}_{j-t_n}\left[(W-W_R)^2(X_i,X_j) - (W-W_R)^2(X_i,X_j') \right] \right| \\
&\leq \sum_{i=1}^{t_n} \sup_{x_i \in \mathbb{S}^{d-1}} \sum_{j=2t_n+1}^{n} \left| \mathbb{E}_{j-t_n}\left[(W-W_R)^2(x_i,X_j) - (W-W_R)^2(x_i,X_j') \right] \right| \\
&\leq \sum_{i=1}^{t_n} \sup_{x_i \in \mathbb{S}^{d-1}} \sum_{j=2t_n+1}^{n} \left| \int_z (W-W_R)^2(x_i,z) \left(P^{t_n}(X_{j-t_n},dz) - \pi(dz) \right) \right| \\
&\leq \|W-W_R\|_\infty^2 \sum_{i=1}^{t_n} \sum_{j=(2t_n+1)}^{n} \int_z \left| P^{t_n}(X_{j-t_n},dz) - \pi(dz) \right| \\
&\leq \|W-W_R\|_\infty^2 \sum_{i=1}^{t_n} \sum_{j=2t_n+1}^{n} L\rho^{t_n} \\
&\leq \|W-W_R\|_\infty^2 n t_n L \rho^{t_n}.
\end{aligned}
$$

where $L > 0$ and $0 < \rho < 1$ are constants related to the uniform ergodicity of the Markov chain (see Definition 5). With analogous computations, we upper-bound the term (4)

$$
\begin{aligned}
(4) &= \sum_{i=t_n+1}^{n-1} \sum_{j=t_n+i+1}^{n} \left| \mathbb{E}_{j-t_n}\left[(W-W_R)^2(X_i,X_j) - (W-W_R)^2(X_i,X_j') \right] \right| \\
&\leq \sum_{i=t_n+1}^{n-1} \sup_{x_i \in \mathbb{S}^{d-1}} \sum_{j=t_n+i+1}^{n} \left| \mathbb{E}_{j-t_n}\left[(W-W_R)^2(x_i,X_j) - (W-W_R)^2(x_i,X_j') \right] \right| \\
&\leq \sum_{i=t_n+1}^{n-1} \sup_{x_i \in \mathbb{S}^{d-1}} \sum_{j=t_n+i+1}^{n} \left| \int_z (W-W_R)^2(x_i,z) \left(P^{t_n}(X_{j-t_n},dz) - \pi(dz) \right) \right| \\
&\leq \|W-W_R\|_\infty^2 \sum_{i=t_n+1}^{n-1} \sum_{j=t_n+i+1}^{n} \int_z \left| P^{t_n}(X_{j-t_n},dz) - \pi(dz) \right| \\
&\leq \|W-W_R\|_\infty^2 \sum_{i=t_n+1}^{n-1} \sum_{j=t_n+i+1}^{n} L\rho^{t_n} \\
&\leq \|W-W_R\|_\infty^2 n^2 L \rho^{t_n}.
\end{aligned}
$$

We deduce that

$$
\begin{aligned}
&\frac{2}{n(n-1)} \sum_{i<j} \left(\mathbb{E}_{j-t_n}\left[(W-W_R)^2(X_i,X_j) \right] - \|W-W_R\|_2^2 \right) \\
\leq\ & 10(1+q+q^2)\|W-W_R\|_\infty^2 \frac{n\log^2 n + L n^2 \rho^{q\log(n)}}{n(n-1)} \\
=\ & \|W-W_R\|_\infty^2 \mathcal{O}(n^{-1}\log^2 n), \qquad (24)
\end{aligned}
$$

because $\rho^{q\log(n)} = \mathcal{O}(n^{-1})$. Indeed,

$$
n\rho^{q\log(n)} = n n^{q\log(\rho)} = n^{1+q\log(\rho)},
$$

with $1 + q\log(\rho) < 0$ because we choose q such that $0 < q < (\log(1/\rho))^{-1}$.
We proved that the last term of the decomposition (15) can be bounded by $M'\|W-W_R\|_\infty^2 n^{-1}\log^2(n)$ where M' is a constant that depends only on L and ρ. Coupling this result with the concentration result (23) concludes the proof of Lemma 6.

H Proof of Theorem 1

The proof of Theorem 1 mainly lies in the following result which is proved in Appendix H.1. Coupling the convergence of the spectrum of the matrix of probability T_n with a concentration result on the spectral norm of random matrices with independent entries (see [5]), we show the convergence in metric δ_2 of the spectrum of $\hat{T}_n$ towards the spectrum of the integral operator $\mathbb{T}_W$.

Theorem 3 *Let us consider $\gamma \in (0,1)$ satisfying $\log(e/\gamma) \leq (n/\log^2 n) \wedge (n/(13\tilde{R}))$. Then it holds with probability at least $1-\gamma$,*

$$
\begin{aligned}
& \delta_2\left(\lambda(\mathbb{T}_W), \lambda(T_n)\right) \\
\leq\quad & 2\|p - p_R\|_2 + 8\sqrt{\frac{\tilde{R}}{n}\ln(e/\gamma)} + M\|p - p_R\|_\infty \sqrt{\frac{\log n}{n}} \left(\log(e/\gamma) \vee \log n\right)^{1/2},
\end{aligned}
$$

where $M > 0$ only depends on constants related to the Markov chain $(X_i)_{i\geq 1}$ (see Lemma 6).

First part of the proof for Theorem 1 We start by establishing the convergence rate for $\delta_2\left(\lambda(\mathbb{T}_W), \lambda(T_n)\right)$. We keep notations of Theorem 3. Let us consider $\gamma \in (0,1)$ satisfying $(n/\log^2 n) \wedge (n/(13\tilde{R}))$, and assume that $p \in Z^s_{w_\beta}((-1,1))$ with $s > 0$.
Let us define the event

$$
\begin{aligned}
\Omega(\gamma) := \Bigg\{ & \delta_2\left(\lambda(\mathbb{T}_W), \lambda(T_n)\right) \leq 2\|p - p_R\|_2 + 8\sqrt{\frac{\tilde{R}}{n}\ln(e/\gamma)} \\
& + M\|p - p_R\|_\infty \sqrt{\frac{\log n}{n}} \left(\log(e/\gamma) \vee \log n\right)^{1/2} \Bigg\}.
\end{aligned}
$$

Using Theorem 3, it holds $\mathbb{P}\left(\Omega(\gamma)\right) \geq 1-\gamma$. Remarking further that

$$
\delta_2\left(\lambda(\mathbb{T}_W), \lambda(T_n)\right) \leq \delta_2\left(\lambda(\mathbb{T}_W), 0\right) + \delta_2\left(0, \lambda(T_n)\right) \leq \|p\|_2 + \sqrt{n} \leq \sqrt{2} + \sqrt{n},
$$

we have

$$
\begin{aligned}
& \mathbb{E}[\delta_2^2(\lambda(\mathbb{T}_W), \lambda(T_n))] \\
=\quad & \mathbb{E}[\delta_2^2(\lambda(\mathbb{T}_W), \lambda(T_n))\mathbb{1}_{\Omega(\gamma)}] + (1+\sqrt{2})^2 n \mathbb{P}(\Omega(\gamma)^c) \\
\leq\quad & c\|p - p_R\|_2^2 + c\frac{\tilde{R}}{n}\log(e/\gamma) + c\|p - p_R\|_\infty^2 \frac{\log n}{n}\left(\log(e/\gamma) \vee \log n\right) \\
& + (1+\sqrt{2})^2 n\gamma,
\end{aligned}
$$

where $c > 0$ is a constant that does not depend on R, d or n.
Since

$$
\|p - p_R\|_2^2 = \sum_{k>R} (p_k^*)^2 d_k \frac{(1+k(k+2\beta))^s}{(1+k(k+2\beta))^s} \leq C(p,s,d) R^{-2s}, \tag{25}
$$

and since

$$
\tilde{R} = O(R^{d-1}), \tag{26}
$$

we have choosing $\gamma = 1/n^2$

$$
\mathbb{E}[\delta_2^2(\lambda(\mathbb{T}_W), \lambda(T_n))] \leq D' \left[R^{-2s} + R^{d-1}\frac{\log(n)}{n} + \|p - p_R\|_\infty^2 \frac{\log^2(n)}{n} \right], \tag{27}
$$

where $D' > 0$ is a constant independent of n and R.

Let us show that choosing $R = \lfloor \left(n/\log^2(n)\right)^{\frac{1}{2s+d-1}} \rfloor$ concludes the proof.
Since $\|G_k^\beta\|_\infty = G_k^\beta(1) = d_k/c_k$, we get that

$$\|p_R\|_\infty \le \sum_{k=0}^{R} |p_k^*| c_k G_k^\beta(1) = \sum_{k=0}^{R} |p_k^*| d_k \le \sqrt{\tilde{R}} \|p_R\|_2,$$

and using (33), we deduce that

$$\|p - p_R\|_\infty \le \|p\|_\infty + \|p_R\|_\infty \le 1 + \sqrt{2\tilde{R}}. \tag{28}$$

Hence, (27) becomes

$$\mathbb{E}[\delta_2^2(\lambda(\mathbb{T}_W), \lambda(T_n))] \le D'' \left[R^{-2s} + R^{d-1} \frac{\log(n)}{n} + \tilde{R} \frac{\log^2(n)}{n} \right],$$

where D'' is a constant that does not depend on n or R.
Choosing $R = \lfloor \left(n/\log^2(n)\right)^{\frac{1}{2s+d-1}} \rfloor$ and using (26) we get

$$\begin{aligned} & \mathbb{E}[\delta_2^2(\lambda(\mathbb{T}_W), \lambda(T_n))] \\ \le \; & D'' \left[\left(\frac{n}{\log^2(n)} \right)^{\frac{-2s}{2s+d-1}} + 2 \left(\frac{n}{\log^2(n)} \right)^{\frac{d-1}{2s+d-1}} \frac{\log^2(n)}{n} \right] \\ \le \; & 3D'' \left(\frac{n}{\log^2(n)} \right)^{\frac{-2s}{2s+d-1}}. \end{aligned}$$

Second part of the proof for Theorem 1 Let us recall that in the statement of Theorem 1, $\lambda^{R_{opt}}(\hat{T}_n)$ is the sequence of the $\tilde{R}_{opt}$ first eigenvalues (sorted in decreasing absolute values) of the matrix $\hat{T}_n$ where R_{opt} is the value of the parameter R leading to the optimal bias-variance trade off, namely

$$\lambda^{R_{opt}}(\hat{T}_n) = (\hat{\lambda}_1, \dots, \hat{\lambda}_{\tilde{R}_{opt}}, 0, 0, \dots).$$

From the computations of the first part of the proof, we know that $R_{opt} = \lfloor \left(n/\log^2(n)\right)^{\frac{1}{2s+d-1}} \rfloor$. That corresponds to the situation where we choose optimally R and it is in practice possible to approximate this best model dimension using e.g. the slope heuristic. Therefore, $\delta_2\left(\lambda(\mathbb{T}_W), \lambda^{R_{opt}}(\hat{T}_n)\right)$ is the quantity of interest since it represents the distance between the eigenvalues used to built our estimates $(\hat{p}_k)_k$ and the true spectrum of the envelope function p. Since $\tilde{R} = \mathcal{O}\left(R^{d-1}\right)$ for all integer $R \ge 0$, we have $\tilde{R}_{opt} = \mathcal{O}\left(\left(n/\log^2(n)\right)^{\frac{d-1}{2s+d-1}}\right)$. We deduce that for n large enough $2\tilde{R}_{opt} \le n$ and using [8, Proposition 15] we obtain

$$\begin{aligned} & \delta_2\left(\lambda^{R_{opt}}(\hat{T}_n), \lambda(\mathbb{T}_{W_{R_{opt}}})\right) \\ \le \; & \delta_2\left(\lambda(T_n), \lambda(\mathbb{T}_{W_{R_{opt}}})\right) + \sqrt{2\tilde{R}_{opt}} \|\hat{T}_n - T_n\| \\ \le \; & \delta_2\left(\lambda(T_n), \lambda(\mathbb{T}_W)\right) + \delta_2\left(\lambda(\mathbb{T}_W), \lambda(\mathbb{T}_{W_{R_{opt}}})\right) + \sqrt{2\tilde{R}_{opt}} \|\hat{T}_n - T_n\|, \end{aligned} \tag{29}$$

where $\lambda(\mathbb{T}_{W_{R_{opt}}}) = (\lambda_1^*, \dots, \lambda_{\tilde{R}_{opt}}^*, 0, 0, \dots)$. Let us consider $\gamma \in (0,1)$. Using Theorem 3, we know that with probability at least $1-\gamma$ it holds for n large enough

$$\begin{aligned} & \delta_2\left(\lambda(T_n), \lambda(\mathbb{T}_W)\right) \\ \le \; & 2\|p - p_{R_{opt}}\|_2 + 8\sqrt{\frac{\tilde{R}_{opt}}{n} \ln(e/\gamma)} + M \|p - p_{R_{opt}}\|_\infty \sqrt{\frac{\log n}{n}} \left(\log(e/\gamma) \vee \log n\right)^{1/2}. \end{aligned}$$

Using (25), (28) and the fact that $\tilde{R} = \mathcal{O}(R^{d-1})$, it holds with probability at least $1 - 1/n^2$,

$$\begin{aligned}\delta_2^2\left(\lambda(T_n), \lambda(\mathbb{T}_W)\right) &\leq c\left[R_{opt}^{-2s} + R_{opt}^{d-1}\frac{\log n}{n} + MR_{opt}^{d-1}\frac{\log^2 n}{n}\right]\\ &\leq (M')^2(n/\log^2 n)^{\frac{-2s}{2s+d-1}},\end{aligned}$$

where $c > 0$ is a numerical constant and $M' > 0$ depends on constants related to the Markov chain $(X_i)_{i\geq 1}$ (see Theorem 3 for details). Moreover,

$$\delta_2^2\left(\lambda(\mathbb{T}_W), \lambda(\mathbb{T}_{W_{R_{opt}}})\right) = \|p - p_{R_{opt}}\|_2^2 \leq C(p,s,d)R_{opt}^{-2s} = \mathcal{O}\left((n/\log^2 n)^{\frac{-2s}{2s+d-1}}\right), \tag{30}$$

where we used (25). Finally, using the concentration of spectral norm for random matrices with independent entries from [5], there exists a universal constant $C_0 > 0$ such that it holds with probability at least $1 - 1/n^2$,

$$\|T_n - \hat{T}_n\| \leq \frac{3}{\sqrt{2n}} + C_0\frac{\sqrt{\log(n^3)}}{n}.$$

Using again $\tilde{R} = \mathcal{O}(R^{d-1})$, this implies that for n large enough, it holds with probability at least $1 - 1/n^2$,

$$\sqrt{2\tilde{R}_{opt}}\|T_n - \hat{T}_n\| \leq D(n/\log^2 n)^{\frac{-s}{2s+d-1}},$$

where $D > 0$ is a numerical constant.
From (29), we deduce that $\mathbb{P}(\Omega) \geq 1 - 2/n^2$ where the event Ω is defined by

$$\Omega = \left\{\delta_2^2\left(\lambda^{R_{opt}}(\hat{T}_n), \lambda(\mathbb{T}_{W_{R_{opt}}})\right) \leq \left(C(p,s,d)^{1/2} + D + M'\right)^2 (n/\log^2 n)^{\frac{-2s}{2s+d-1}}\right\}.$$

Remarking finally,

$$\delta_2\left(\lambda^{R_{opt}}(\hat{T}_n), \lambda(\mathbb{T}_{W_{R_{opt}}})\right) \leq \delta_2\left(\lambda(\mathbb{T}_{W_{R_{opt}}}), 0\right) + \delta_2\left(0, \lambda(\hat{T}_n)\right) \leq \|p\|_2 + \sqrt{n} \leq \sqrt{2} + \sqrt{n},$$

we obtain

$$\begin{aligned}
& \mathbb{E}\left[\delta_2^2\left(\lambda^{R_{opt}}(\hat{T}_n), \lambda(\mathbb{T}_{W_{R_{opt}}})\right)\right] && (31)\\
\leq\ & \mathbb{E}\left[\delta_2^2\left(\lambda^{R_{opt}}(\hat{T}_n), \lambda(\mathbb{T}_{W_{R_{opt}}})\right) \mid \Omega\right] + \mathbb{P}(\Omega^c)(\sqrt{2} + \sqrt{n})^2 \\
\leq\ & \left(C(p,s,d)^{1/2} + D + M'\right)^2 (n/\log^2 n)^{\frac{-2s}{2s+d-1}} + 2\frac{(\sqrt{2}+\sqrt{n})^2}{n^2} \\
=\ & \mathcal{O}\left((n/\log^2 n)^{\frac{-2s}{2s+d-1}}\right). && (32)
\end{aligned}$$

Using the triangle inequality, (30) and (32) leads to

$$\begin{aligned}\mathbb{E}\left[\delta_2^2\left(\lambda^{R_{opt}}(\hat{T}_n), \lambda(\mathbb{T}_W)\right)\right] &\leq 3\mathbb{E}\left[\delta_2^2\left(\lambda^{R_{opt}}(\hat{T}_n), \lambda(T_{W_{R_{opt}}})\right)\right] + 3\delta_2^2\left(\lambda(T_{W_{R_{opt}}}), \lambda(T_W)\right)\\ &= \mathcal{O}\left((n/\log^2 n)^{\frac{-2s}{2s+d-1}}\right),\end{aligned}$$

which concludes the proof of Theorem 1.
■

H.1 Proof of Theorem 3

We follow the same sketch of proof as in [8].
Let $R \geq 1$ and define,

$$\begin{aligned}
\Phi_{k,l} &= \frac{1}{\sqrt{n}} \left[Y_{k,l}(X_1), \dots, Y_{k,l}(X_n) \right] \in \mathbb{R}^n, \\
E_{R,n} &= \left(\langle \Phi_{k,l}, \Phi_{k',l'} \rangle - \delta_{(k,l),(k',l')} \right)_{(k,k') \in [R],\ l \in \{1,\dots,d_k\},\ l' \in \{1\dots,d_{k'}\}} \in \mathbb{R}^{\tilde{R} \times \tilde{R}}, \\
X_{R,n} &= [\Phi_{0,1}, \Phi_{1,1}, \Phi_{1,2}, \dots, \Phi_{R,d_R}] \in \mathbb{R}^{n \times \tilde{R}}, \\
A_{R,n} &= \left(X_{R,n}^\top X_{R,n} \right)^{1/2} \text{ with } A_{R,n}^2 = Id_{\tilde{R}} + E_{R,n}, \\
K_R &= Diag(\lambda_1(\mathbb{T}_W), \dots, \lambda_{\tilde{R}}(\mathbb{T}_W)), \\
T_{R,n} &= \sum_{k=0}^{R} p_k^* \sum_{l=1}^{d_k} \Phi_{k,l} (\Phi_{k,l})^\top = X_{R,n} K_R X_{R,n}^\top \in \mathbb{R}^{n \times n} \\
\tilde{T}_{R,n} &= ((1 - \delta_{i,j}) T_{R,n})_{i,j \in [n]} \in \mathbb{R}^{n \times n}, \\
T_{R,n}^* &= A_{R,n} K_R A_{R,n}^\top \in \mathbb{R}^{\tilde{R} \times \tilde{R}}, \\
W_R(x,y) &= \sum_{k=0}^{R} p_k^* \sum_{l=1}^{d_k} Y_{k,l}(x) Y_{k,l}(y).
\end{aligned}$$

It holds

$$\delta_2(\lambda(\mathbb{T}_W), \lambda(\mathbb{T}_{W_R})) = \left(\sum_{k > R} d_k (p_k^*)^2 \right)^{1/2}.$$

We point out the equality between spectra of the operator $\mathbb{T}_{W_R}$ and the matrix K_R. Using the SVD decomposition of $X_{R,n}$, one can also easily prove that $\lambda(T_{R,n}) = \lambda(T_{R,n}^*)$. We deduce that

$$\delta_2\left(\lambda(\mathbb{T}_{W_R}), \lambda(T_{R,n})\right) = \delta_2\left(\lambda(K_R), \lambda(T_{R,n}^*)\right) \leq \|T_{R,n}^* - K_R\|_F = \|A_{R,n} K_R A_{R,n} - K_R\|_F,$$

with the Hoffman-Wielandt inequality. Using equation (4.8) at ([22] p.127) gives

$$\delta_2\left(\lambda(\mathbb{T}_{W_R}), \lambda(T_{R,n})\right) \leq \sqrt{2} \|K_R\|_F \|E_{R,n}\| = \sqrt{2} \|W_R\|_2 \|E_{R,n}\|.$$

Using again the Hoffman-Wielandt inequality we get

$$\delta_2(\lambda(T_{R,n}), \lambda(\tilde{T}_{R,n})) \leq \|\tilde{T}_{R,n} - T_{R,n}\|_F = \left[\frac{1}{n^2} \sum_{i=1}^{n} W_R(X_i, X_i)^2 \right]^{1/2},$$

and

$$\delta_2\left(\lambda(\tilde{T}_{R,n}), \lambda(T_n)\right) \leq \|\tilde{T}_{R,n} - T_n\|_F = \left[\frac{1}{n^2} \sum_{i \neq j} (W - W_R)^2 (X_i, X_j) \right]^{1/2}.$$

Now, we invoke Lemmas 9, 10 and 6 to conclude the proof. Proofs of those lemmas are provided in Appendix H.2, H.3 and G respectively.

Lemma 9 *Let us consider $\gamma > 0$ and assume that $13\tilde{R}\ln(e/\gamma) \leq n$. Then it holds with probability at least $1 - \gamma$*

$$\|E_{R,n}\| \leq 4 \sqrt{\frac{\tilde{R}}{n} \ln(2/\gamma)}.$$

Lemma 10 *Let $R \geq 1$. We have*

$$\frac{1}{n^2}\sum_{i=1}^{n} W_R(X_i,X_i)^2 = \frac{1}{n}\left(\sum_{k=0}^{R} p_k^* d_k\right)^2 .$$

For any $\gamma \in (0,1)$ with $\log(e/\gamma) \leq (n/\log^2 n) \wedge (n/(13\tilde{R}))$, it holds with probability at least $1-\gamma$,

$$\begin{aligned}
&\delta_2\left(\lambda(\mathbb{T}_W),\lambda(T_n)\right)\\
\leq\ & \delta_2\left(\lambda(\mathbb{T}_W),\lambda(\mathbb{T}_{W_R})\right) + \delta_2\left(\lambda(\mathbb{T}_{W_R}),\lambda(T_{R,n})\right) + \delta_2\left(\lambda(T_{R,n}),\lambda(\tilde{T}_{R,n})\right)\\
&+ \delta_2\left(\lambda(\tilde{T}_{R,n}),\lambda(T_n)\right)\\
\leq\ & 4\sqrt{\frac{\tilde{R}}{n}\ln(2/\gamma)} + \sqrt{2}\left(\sum_{k=0}^{R} d_k (p_k^*)^2\right)^{1/2} + \frac{1}{\sqrt{n}}\left|\sum_{k=0}^{R} p_k^* d_k\right| + 2\|p-p_R\|_2\\
&+ M\|p-p_R\|_\infty \sqrt{\frac{\log n}{n}}\left(\log(e/\gamma)\vee \log n\right)^{1/2},
\end{aligned}$$

where $M>0$ depends only on constants related to the Markov chain $(X_i)_{i\geq 1}$.
Now remark that

$$\left|\sum_{k=0}^{R} p_k^* d_k\right| \leq \left(\sum_{k=0}^{R} d_k\right)^{1/2}\left(\sum_{k=0}^{R} d_k (p_k^*)^2\right)^{1/2} = \sqrt{\tilde{R}}\|p_R\|_2,$$

and that

$$\|p_R\|_2^2 \leq \|p\|_2^2 \leq 2, \tag{33}$$

because p_R is the orthogonal projection of p, and $|p|\leq 1$.
We deduce that

$$\begin{aligned}
&\delta_2\left(\lambda(\mathbb{T}_W),\lambda(T_n)\right)\\
\leq\ & 2\|p-p_R\|_2 + 4\sqrt{\frac{\tilde{R}}{n}\ln(2/\gamma)} + \sqrt{\frac{2\tilde{R}}{n}}\\
&+ M\|p-p_R\|_\infty \sqrt{\frac{\log n}{n}}\left(\log(e/\gamma)\vee \log n\right)^{1/2}\\
\leq\ & 2\|p-p_R\|_2 + 8\sqrt{\frac{\tilde{R}}{n}\ln(e/\gamma)}\\
&+ M\|p-p_R\|_\infty \sqrt{\frac{\log n}{n}}\left(\log(e/\gamma)\vee \log n\right)^{1/2}.
\end{aligned} \tag{34}$$

■

H.2 Proof of Lemma 9

Observe that $nE_{R,n} = \sum_{i=1}^n \left(Z_i Z_i^\top - Id_{\tilde{R}}\right)$ where for all $i\in[n]$, $Z_i \in \mathbb{R}^{\tilde{R}}$ is defined by

$$Z_i := Z(X_i) := \left(Y_{0,1}(X_i), Y_{1,1}(X_i), Y_{1,2}(X_i),\dots,Y_{1,d_1}(X_i),\dots,Y_{R,1}(X_i),\dots,Y_{R,d_R}(X_i)\right).$$

By definition of the spectral norm for a hermitian matrix,

$$\|\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^\top - Id_{\tilde{R}}\| = \max_{x,\ \|x\|=1}\left|x^\top\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^\top\right)x - 1\right|.$$

We use a covering set argument based on the following Lemma.

Lemma 11 *(see [15])*
Let us consider an integer $D \geq 2$. For any $\epsilon_0 > 0$, there exists a set $Q \subset \mathbb{S}^{D-1}$ of cardinality at most $(1+2/\epsilon_0)^D$ such that

$$\forall \alpha \in \mathbb{S}^{D-1}, \quad \exists q \in Q, \quad \|\alpha - q\|_2 \leq \epsilon_0.$$

We consider Q the set given by Lemma 11 with $D = d$ and $\epsilon_0 \in (0, 1/2)$. Let us define $x_0 \in \mathbb{S}^{d-1}$ such that $|x_0^\top E_{R,n} x_0| = \|E_{R,n}\|$ and $q_0 \in Q$ such that $\|x_0 - q_0\|_2 \leq \epsilon_0$. Then,

$$\begin{aligned}
|x_0^\top E_{R,n} x_0| - |q_0^\top E_{R,n} q_0| &\leq |x_0^\top E_{R,n} x_0 - q_0^\top E_{R,n} q_0| \text{ (by triangle inequality)}\\
&= |x_0^\top E_{R,n}(x_0 - q_0) - (q_0 - x_0)^\top E_{R,n} q_0|\\
&\leq \|x_0\|_2 \|E_{R,n}\| \|x_0 - q_0\|_2 + \|q_0 - x_0\|_2 \|E_{R,n}\| \|q_0\|_2\\
&\leq 2\epsilon_0 \|E_{R,n}\|.
\end{aligned}$$

which leads to

$$|x_0^\top E_{R,n} x_0| = \|E_{R,n}\| \leq |q_0^\top E_{R,n} q_0| + 2\epsilon_0 \|E_{R,n}\|.$$

Hence,

$$\|E_{R,n}\| \leq \frac{1}{1 - 2\epsilon_0} \max_{q \in Q} |q^\top E_{R,n} q|.$$

We introduce for any $q \in Q$ the function

$$F_q : x = (x_1, \dots, x_n) \mapsto \frac{1}{n} \sum_{i=1}^n q^\top \left(Z_i Z_i^\top - 1 \right) q := \frac{1}{n} \sum_{i=1}^n f_q(x_i),$$

where $f_q(x) = q^\top \left(Z(x) Z(x)^\top - 1 \right) q$.
Let us consider $t > 0$. We want to apply Bernstein's inequality for Markov chains from [19, Theorem 1.1]. We remark that $\mathbb{E}_\pi[f_q(X)] = 0$ and that $\|f_q\|_\infty \leq \tilde{R} - 1$. For all $m \in [\tilde{R}]$, we denote $\phi_m = Y_{r,l}$ with $r \in \{0, \dots, R\}$ and $l \in [d_r]$ such that $m = l + \sum_{i=0}^r d_i - 1$. Then, for any $x \in \mathbb{S}^{d-1}$, and for all $k, l \in [\tilde{R}]$, $\left((Z(x)^\top Z(x))^2\right)_{k,l} = \sum_{m=1}^{\tilde{R}} \phi_l(x) \phi_m(x)^2 \phi_k(x) = \tilde{R} \phi_l(x) \phi_k(x) = \tilde{R} \left(Z(x) Z(x)^\top \right)_{k,l}$ where we used [10, Eq.(1.2.9)]. We deduce that

$$\begin{aligned}
\mathbb{E}_\pi[f_q(X)^2] &= \mathbb{E}_\pi[q^\top Z(X) Z(X)^\top q q^\top Z(x) Z(x)^\top q] - 2\mathbb{E}_\pi[q^\top Z(X) Z(X)^\top q] + 1\\
&= \mathbb{E}_\pi[q^\top \underbrace{(Z(X) Z(X)^\top)^2}_{=\tilde{R} \cdot Z(X) Z(X)^\top} q] - 2q^\top \underbrace{\mathbb{E}_\pi[Z(X) Z(X)^\top]}_{=\mathrm{Id}} q + 1\\
&= \tilde{R} \cdot q^\top \mathbb{E}_\pi[Z(X) Z(X)^\top] q - 1\\
&= \tilde{R} - 1.
\end{aligned}$$

Using that the Markov chain $(X_i)_{i \geq 1}$ has a spectral gap equals to 1 (see Appendix B.3), we get from [19, Eq. (1.6)] that

$$\mathbb{P}\left(|F_q(X)| \geq t\right) = \mathbb{P}\left(|q^\top E_{R,n} q| \geq t\right) \leq 2 \exp\left(\frac{-nt^2}{4(\tilde{R} - 1) + 10(\tilde{R} - 1)t} \right),$$

which leads to

$$\mathbb{P}\left(\max_{q \in Q} |q^\top E_{R,n} q| \geq t \right) \leq \mathbb{P}\left(\bigcup_{q \in Q} |q^\top E_{R,n} q| \geq t \right) \leq 2 \exp\left(\frac{-nt^2 / (\tilde{R} - 1)}{4 + 10t} \right) (1 + 2/\epsilon_0)^{\tilde{R}}.$$

Choosing $\epsilon_0 = 2\left(\exp\left(\frac{nt^2/2}{(\tilde{R}-1)\tilde{R}(4+10t)} \right) - 1 \right)^{-1}$ in order to satisfy $(1 + 2/\epsilon_0)^{\tilde{R}} = \exp(nt^2 (\tilde{R} - 1)^{-1} (4 + 10t)^{-1}/2)$, we get

$$\mathbb{P}\left(\max_{q \in Q} |q^\top E_{R,n} q| \geq t \right) \leq 2 \exp\left(\frac{-nt^2}{(\tilde{R} - 1)(8 + 20t)} \right).$$

We deduce that if $\frac{25}{2}\ln(2/\alpha)\tilde{R} \le n$, it holds with probability at least $1-\alpha$,

$$\max_{q\in Q} |q^\top E_{R,n} q| \le 16\sqrt{\frac{\tilde{R}}{n}\ln(2/\alpha)}.$$

Assuming that $200\ln(7)\tilde{R}^3\ln(2/\alpha) \le n^3$ in order to have $1/(1-2\epsilon_0) \le 1/4$, it holds with probability at least $1-\alpha$

$$\|E_{R,n}\| \le \frac{1}{1-2\epsilon_0}\max_{q\in Q}|q^\top E_{R,n}q| \le 4\sqrt{\frac{\tilde{R}}{n}\ln(2/\alpha)}.$$

■

H.3 Proof of Lemma 10

Reminding that for all $x \in \mathbb{S}^{d-1}$ and for all $k \ge 0$, $\sum_{l=1}^{d_k} Y_{k,l}(x)^2 = d_k$ (see Corollary 1.2.7 from [10]), we get

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^{n} W_R(X_i,X_i)^2 &= \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{k=0}^{R} p_k^* \sum_{l=1}^{d_k} Y_{k,l}(X_i)^2\right)^2 \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{k=0}^{R} p_k^* d_k\right)^2 \\ &= \frac{1}{n}\left(\sum_{k=0}^{R} p_k^* d_k\right)^2.\end{aligned}$$

■

I Proof of Theorem 2

Proposition 3 is the counterpart of Proposition 1 in [2] in our dependent framework. This result is the cornerstone of Theorem 2 and is proved in Appendix I.1.

Proposition 3 *We assume that $\Delta^* > 0$. Let us consider $\gamma > 0$ and define the event*

$$\mathcal{E} := \left\{ \delta_2(\lambda(T_n), \lambda(\mathbb{T}_W)) \vee \frac{2^{\frac{9}{2}}\sqrt{d}}{\Delta^*}\|T_n - \hat{T}_n\| \le \frac{\Delta^*}{4} \right\}.$$

Then for n large enough,

$$\mathbb{P}(\mathcal{E}) \ge 1-\gamma/2.$$

Moreover, on the event $\mathcal{E}$, there exists one and only one set Λ_1, consisting of d eigenvalues of $\hat{T}_n$, whose diameter is smaller that $\Delta^*/2$ and whose distance to the rest of the spectrum of $\hat{T}_n$ is at least $\Delta^*/2$. Furthermore, on the event $\mathcal{E}$, the algorithm HEiC returns the matrix $\hat{G} = \frac{1}{d}\hat{V}\hat{V}^\top$, where $\hat{V}$ has by columns the eigenvectors corresponding to the eigenvalues in Λ_1.

In the following, we work on the event $\mathcal{E}$. Let us consider $\gamma \in (0,1)$.
We choose $R = (n/\log^2 n)^{\frac{1}{2s+d-1}}$. Reminding that W_R is the rank R approximation of W, the Gram matrix associated with the kernel W_R is

$$T_{R,n} = \sum_{k=0}^{R} p_k^* \sum_{l=1}^{d_k} \Phi_{k,l}(\Phi_{k,l})^\top = X_{R,n}K_R X_{R,n}^\top \in \mathbb{R}^{n\times n}$$

where

$$\Phi_{k,l} = \frac{1}{\sqrt{n}} [Y_{k,l}(X_1), \dots, Y_{k,l}(X_n)] \in \mathbb{R}^n,$$
$$X_{R,n} = [\Phi_{0,1}, \Phi_{1,1}, \Phi_{1,2}, \dots, \Phi_{R,d_R}] \in \mathbb{R}^{n\times\tilde{R}} \text{ and}$$
$$K_R = Diag(\lambda_1(\mathbb{T}_W), \dots, \lambda_{\tilde{R}}(\mathbb{T}_W)).$$

Let us denote now $\tilde{V}$ (resp. $\tilde{V}_R$) the orthonormal matrix formed by the eigenvectors of the matrix T_n (resp. $T_{R,n}$). We have the following eigenvalue decompositions

$$T_n = \tilde{V}\Lambda\tilde{V}^\top \text{ and } T_{R,n} = \tilde{V}_R\Lambda_R\tilde{V}_R^\top,$$

where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ are the eigenvalues of the matrix T_n and where
$\Lambda_R = (p_0^*, p_1^*, \dots, p_1^*, \dots, p_R^*, \dots, p_R^*, 0, \dots, 0) \in \mathbb{R}^n$ where each p_k^* has multiplicity d_k. Then, we note by $V \in \mathbb{R}^{n\times d}$ (resp. V_R) the matrix formed by the columns $2, \dots, d$ of the matrix $\tilde{V}$ (resp. $\tilde{V}_R$). The matrix $V^* \in \mathbb{R}^{n\times d}$ is the orthonormal matrix with $i-$th column $\frac{1}{\sqrt{n}}(Y_{1,1}(X_i), \dots, Y_{1,d}(X_i))$. The matrices G^*, G, G_R and G^*_{proj} are defined as follows

$$G^* := \frac{1}{c_1} V^*(V^*)^\top$$
$$G := \frac{1}{c_1} VV^\top$$
$$G_R := \frac{1}{c_1} V_R V_R^\top$$
$$G^*_{proj} := V^*((V^*)^\top V^*)^{-1}(V^*)^\top.$$

G^*_{proj} is the projection matrix for the columns span of the matrix V^*. Using the triangle inequality we have

$$\|G^* - G\|_F \le \|G^* - G^*_{proj}\|_F + \|G^*_{proj} - G_R\|_F + \|G_R - G\|_F.$$

Step 1: Bounding $\|G - G_R\|_F$. Since the columns of the matrices V and V_R correspond respectively to the eigenvectors of the matrices T_n and $T_{R,n}$, applying the Davis Kahan sinus Theta Theorem (see Theorem 4) gives that there exists $O \in \mathbb{R}^{d\times d}$ such that

$$\|VO - V_R\|_F \le \frac{2^{3/2}\|T_n - T_{R,n}\|_F}{\Delta},$$

where $\Delta := \min_{k\in\{0,2,3,\dots,R\}} |p_1^* - p_k^*| \ge \Delta^* = \min_{k\in\mathbb{N},\, k\neq 1} |p_1^* - p_k^*|$. Using Lemma 12, we get that

$$\|G - G_R\|_F = \frac{1}{d}\|VO(VO)^\top - V_R V_R^\top\|_F \le 2\|VO - V_R\|_F.$$

Hence, using the proof of Theorem 1, we get that with probability at least $1 - 1/n^2$,

$$\|G - G_R\|_F \le 2\|VO - V_R\|_F \le \frac{C}{\Delta^*}\left(\frac{n}{\log^2 n}\right)^{-\frac{s}{2s+d-1}},$$

where $C > 0$ is a constant.

Step 2: Bounding $\|G^* - G^*_{proj}\|_F$. To bound $\|G^* - G^*_{proj}\|_F$, we apply first Lemma 13 with $B = V^*$. This leads to

$$\|G^* - G^*_{proj}\|_F \le \|\mathrm{Id}_d - (V^*)^\top V^*\|_F \le \sqrt{d}\|\mathrm{Id}_d - (V^*)^\top V^*\|.$$

Using a proof rigorously analogous to the proof of Lemma 9, it holds with probability at least $1-\gamma$ and for n large enough,

$$\|\mathrm{Id}_d - (V^*)^\top V^*\| \leq 4\sqrt{\frac{d\log(e/\gamma)}{n}}.$$

We get by choosing $\gamma = 1/n^2$ that it holds with probability at least $1-1/n^2$,

$$\|\mathrm{Id}_d - (V^*)^\top V^*\| \leq C'\sqrt{\frac{d\log(n)}{n}},$$

where $C' > 0$ is a universal constant.

Step 3: Bounding $\|G^*_{proj} - G_R\|_F$. We proceed exactly like in [2] but we provide here the proof for completeness. Since G^*_{proj} and G_R are projectors we have, using for example [6, p.202],

$$\|G^*_{proj} - G_R\|_F = 2\|G^*_{proj} G_R^{\perp}\|_F. \tag{35}$$

We use Theorem 5 with $E = G^*_{proj}$, $F = G_R^{\perp}$, $B = T_{R,n}$ and $A = T_{R,n} + H$ where

$$H = \tilde{X}_{R,n} K_R \tilde{X}_{R,n}^\top - X_{R,n} K_R X_{R,n},$$

where the columns of the matrix $\tilde{X}_{R,n}$ are obtained using a Gram-Schmidt orthonormalization process on the columns of $X_{R,n}$. Hence there exists a matrix L such that $\tilde{X}_{R,n} = X_{R,n}(L^{-1})^\top$. This matrix L is such that a Cholesky decomposition of $X_{R,n}^\top X_{R,n}$ reads as $LL^\top$.
A and B are symmetric matrices thus we can apply Thoerem 5. On the event $\mathcal{E}$, we can take $S_1 = (\lambda_1 - \frac{\Delta^*}{8}, \lambda_1 + \frac{\Delta^*}{8})$ and $S_2 = \mathbb{R}\backslash(\lambda_1 - \frac{7\Delta^*}{8}, \lambda_1 + \frac{7\Delta^*}{8})$. By Theorem 5 we get

$$\|G^*_{proj} G_R^{\perp}\|_F \leq \frac{\|A-B\|_F}{\Delta^*} = \frac{\|H\|_F}{\Delta^*}. \tag{36}$$

We only need to bound $\|H\|_F$.

$$\begin{aligned} \|H\|_F &\leq \|L^{-\top} K_R L^{-1} - K_R\|_F \|X_{R,n}^\top X_{R,n}\| \\ &\leq \|K_R\|_F \|L^{-1}L^{-\top} - \mathrm{Id}_{\tilde{R}}\| \|X_{R,n}^\top X_{R,n}\|, \end{aligned} \tag{37}$$

where the last inequality comes from Lemma 14. From the previous remarks on the matrix L, we directly get

$$\|L^{-1}L^{-\top} - \mathrm{Id}_{\tilde{R}}\| = \| \left(X_{R,n}^\top X_{R,n}\right)^{-1} - \mathrm{Id}_{\tilde{R}}\|.$$

Using the notations of the proof of Theorem 3 which is provided in Appendix H.1, we get

$$\|L^{-1}L^{-\top} - \mathrm{Id}_{\tilde{R}}\| \|X_{R,n}^\top X_{R,n}\| = \|X_{R,n}^\top X_{R,n} - \mathrm{Id}_{\tilde{R}}\| = \|E_{R,n}\|.$$

Noticing further that $\|K_R\|_F^2 \leq \sum_{k\geq 0} (p_k^*)^2 d_k = \|p\|_2^2 \leq 2$ (because $|p| \leq 1$), (37) becomes

$$\|H\|_F \leq \sqrt{2}\|E_{R,n}\|. \tag{38}$$

Using Lemma 9, it holds with probability at least $1-\gamma$ and for n large enough,

$$\|E_{R,n}\| \leq 4\sqrt{\frac{\tilde{R}}{n}\ln(2/\gamma)}. \tag{39}$$

Since $\tilde{R} = \mathcal{O}\left(R^{d-1}\right)$ and $R = \mathcal{O}\left(\left(n/\log^2 n\right)^{\frac{1}{2s+d-1}}\right)$, we obtain using (35), (36), (38) and (39) that with probability at least $1-1/n^2$ it holds

$$\|G^*_{proj} - G_R\|_F = 2\|G^*_{proj} G_R^{\perp}\|_F \leq \frac{C_d}{\Delta^*}\left(\frac{n}{\log^2(n)}\right)^{\frac{-s}{2s+d-1}},$$

where $C_d > 0$ is a constant that may depend on d and on constants related to the Markov chain $(X_i)_{i\geq 1}$.

Conclusion. We proved that on the event $\mathcal{E}$, it holds with probability at least $1 - 3/n^2$,

$$\|G^* - G\|_F \leq D_1 \left(\frac{n}{\log^2(n)} \right)^{\frac{-s}{2s+d-1}},$$

where $D_1 > 0$ is a constant that depends on Δ^*, d and on constants related to the Markov chain $(X_i)_{i\geq 1}$. Moreover, Eq. (43) from the proof of Proposition 3 gives that on the event $\mathcal{E}$, we have

$$\|G - \hat{G}\|_F = \frac{1}{d}\|VV^\top - \hat{V}\hat{V}^\top\|_F \leq \frac{2^{\frac{9}{2}}\sqrt{d}\|T_n - \hat{T}_n\|}{3\Delta^*}.$$

Using the concentration result from [5] on spectral norm of centered random matrix with independent entries we get that there exists some constant $D_2 > 0$ such that with probability at least $1 - 1/n^2$ it holds

$$\|G - \hat{G}\|_F \leq D_2 \frac{\sqrt{\log n}}{n}.$$

Using again Proposition 3, we know that for n large enough, $\mathbb{P}(\mathcal{E}) \geq 1 - 1/n^2$. We conclude that for n large enough, it holds with probability at least $1 - 5/n^2$,

$$\|G^* - \hat{G}\|_F \leq D_3 \left(\frac{n}{\log^2(n)} \right)^{\frac{-s}{2s+d-1}},$$

for some constant $D_3 > 0$ that depends on Δ^*, d and on constants related to the Markov chain $(X_i)_{i\geq 1}$ (see Theorem 3 for details).
■

I.1 Proof of Proposition 3

First part of the proof Let us consider $\gamma > 0$.
Using the concentration of spectral norm for random matrices with independent entries from [5], there exists a universal constant C_0 such that

$$\mathbb{P}\left(\|T_n - \hat{T}_n\| \leq \frac{3\sqrt{2D_0}}{n} + C_0 \frac{\sqrt{\log n/\gamma}}{n} \right) \leq \gamma,$$

where denoting $Y = T_n - \hat{T}_n$, we define $D_0 := \max_{1\leq i\leq n} \sum_{j=1}^{n} Y_{i,j}\left(1 - Y_{i,j}\right)$. We deduce that for n large enough, it holds with probability at least $1 - \gamma/4$,

$$\|T_n - \hat{T}_n\| \leq \frac{(\Delta^*)^2}{2^{\frac{13}{2}}\sqrt{d}}. \tag{40}$$

Using now Theorem 1, it holds with probability at least $1 - \gamma/4$ for n large enough

$$\delta_2\left(\lambda(T_n), \lambda(\mathbb{T}_W)\right) \leq C \left(\frac{\log^2 n}{n} \right)^{\frac{s}{2s+d-1}} \leq \frac{\Delta^*}{8}. \tag{41}$$

Putting together (40) and (41), we deduce that for n large enough,

$$\mathbb{P}\left(\mathcal{E}\right) \geq 1 - \gamma/2.$$

Second part of the proof In the following, we work on the event $\mathcal{E}$. Since $\Delta^* > 0$ by assumption, we get that $p_1^* = \lambda_1^* = \cdots = \lambda_d^*$ is the only eigenvalue of $\mathbb{T}_W$ with multiplicity d. Indeed, all eigenvalue p_k^* with $k > d$ has multiplicity $d_k > d$ and p_0^* has multiplicity 1. Moreover, from (41), we have that there exists a unique set of d eigenvalues of T_n, denoted $\lambda_{i_1}, \lambda_{i_2}, \dots, \lambda_{i_d}$, such that they are at a distance least $3\Delta^*/4$ away from the other eigenvalues, i.e.

$$\Delta := \min_{\nu_1 \in \lambda(T_n)\backslash\{\lambda_{i_1},\lambda_{i_2},\dots,\lambda_{i_d}\}} \max_{\nu_2 \in \{\lambda_{i_1},\lambda_{i_2},\dots,\lambda_{i_d}\}} |\nu_1 - \nu_2| \geq \frac{3\Delta^*}{4}. \tag{42}$$

Let us form the matrix $V \in \mathbb{R}^{n\times d}$ where the k-th column is the eigenvector of T_n associated with the eigenvalue λ_{i_k}. We denote further $G := VV^\top/d$. Let $\hat{V} \in \mathbb{R}^{n\times d}$ be the matrix with columns corresponding to the eigenvectors associated to eigenvalues $\hat{\lambda}_{i_1}, \hat{\lambda}_{i_2}, \dots, \hat{\lambda}_{i_d}$ of $\hat{T}_n$ and $\hat{G} := \hat{V}\hat{V}^\top/d$. Using Theorem 4 there exists some orthonormal matrix $O \in \mathbb{R}^{d\times d}$ such that

$$\|VO - \hat{V}\|_F \leq \frac{2^{\frac{3}{2}} \min\{\sqrt{d}\|T_n - \hat{T}_n\|, \|T_n - \hat{T}_n\|_F\}}{\Delta}.$$

Denoting $\lambda_{i_1}^{sort} \geq \lambda_{i_2}^{sort} \geq \cdots \geq \lambda_{i_d}^{sort}$ (resp. $\hat{\lambda}_{i_1}^{sort} \geq \hat{\lambda}_{i_2}^{sort} \geq \cdots \geq \hat{\lambda}_{i_d}^{sort}$) the sorted version of the eigenvalues $\lambda_{i_1}, \lambda_{i_2}, \dots, \lambda_{i_d}$ (resp. $\hat{\lambda}_{i_1}, \hat{\lambda}_{i_2}, \dots, \hat{\lambda}_{i_d}$), we have

$$\begin{aligned}
&\left[\sum_{k=1}^{d} \left(\lambda_{i_k}^{sort} - \hat{\lambda}_{i_k}^{sort}\right)^2\right]^{1/2} \\
\leq\ & \|VV^\top - \hat{V}\hat{V}^\top\|_F \quad \text{(Hoffman-Wielandt inequality [6, Thm VI.4.1])} \\
\leq\ & 2\|VO - \hat{V}\|_F \quad \text{(Using Lemma 12)} \\
\leq\ & \frac{2^{\frac{5}{2}} \min\{\sqrt{d}\|T_n - \hat{T}_n\|, \|T_n - \hat{T}_n\|_F\}}{\Delta} \\
\leq\ & \frac{2^{\frac{9}{2}} \min\{\sqrt{d}\|T_n - \hat{T}_n\|, \|T_n - \hat{T}_n\|_F\}}{3\Delta^*} \quad \text{(Using (42))} \\
\leq\ & \Delta^*/8. \quad \text{(Using (40))}
\end{aligned} \tag{43}$$

Using the triangle inequality, we get that

$$\hat{\Delta} := \min_{\nu_1 \in \lambda(\hat{T}_n)\backslash\{\hat{\lambda}_{i_1},\hat{\lambda}_{i_2},\dots,\hat{\lambda}_{i_d}\}} \max_{\nu_2 \in \{\hat{\lambda}_{i_1},\hat{\lambda}_{i_2},\dots,\hat{\lambda}_{i_d}\}} |\nu_1 - \nu_2| \geq \frac{\Delta^*}{2}. \tag{44}$$

We proved that on the event $\mathcal{E}$, the eigenvalues in $\Lambda_1 := \{\hat{\lambda}_{i_1}, \dots, \hat{\lambda}_{i_d}\}$ are at distance at least $\Delta^*/2$ from the other eigenvalues of $\hat{T}_n$ (see (44)) and are at distance at most $\Delta^*/8$ of the eigenvalues $\lambda_{i_1}, \dots, \lambda_{i_d}$ of T_n. We could have done this analysis for different eigenvalues. Let us consider some $k \geq 0$. Eq. (41) shows that on the event $\mathcal{E}$, there exists a set of d_k eigenvalues of T_n which concentrate around p_k^* and such that it has diameter at most $\Delta^*/4$. Weyl's inequality (see [6, p.63]) proves that there exist d_k eigenvalues of $\hat{T}_n$ that are at distance at most $\Delta^*/4$ from p_k^*. If we consider now a subset $L \neq \Lambda_1$ of d eigenvalues of $\hat{T}_n$, then the previous analysis shows that there exists some eigenvalue $\hat{\lambda}$ of $\hat{T}_n$ which is not in L and that is at distance at most $\Delta^*/4$ from one eigenvalue in L. Using (42), we deduce that Algorithm (HEiC) returns $\hat{G} = \hat{V}\hat{V}^\top/d$ where the columns of $\hat{V}$ correspond to the eigenvectors of $\hat{T}_n$ associated to the eigenvalues in Λ_1.

■

I.2 Useful results

Lemma 12 *Let A, B be two matrices in $\mathbb{R}^{n\times d}$ then*

$$\|AA^\top - BB^\top\|_F \leq (\|A\| + \|B\|)\|A - B\|_F.$$

If $A^\top A = B^\top B = \mathrm{Id}$ *then*

$$\|AA^\top - BB^\top\|_F \leq 2\|A - B\|_F.$$

Proof of Lemma 12.

$$\begin{aligned}
\|AA^\top - BB^\top\|_F &= \|(A-B)A^\top + B(A^\top - B^\top)\|_F \\
&= \|A(A-B)^\top\|_F + \|(B-A)B^\top\|_F \\
&\leq \|(A \otimes Id_n)vec(A-B)\|_2 + \|(Id_d \otimes B)vec(A-B)^\top\|_2 \\
&\leq (\|A \otimes Id_n\| + \|Id_d \otimes B\|)\,\|A-B\|_F \\
&= (\|A\| + \|B\|)\|A-B\|_F,
\end{aligned}$$

where $vec(\cdot)$ represent the vectorization of a matrix that its transformation into a column vector and $\otimes$ is the notation for the Kronecker product between two matrices.

■

Theorem 4 *(Davis-Kahan Theorem) Let* Σ *and* $\hat{\Sigma}$ *be two symmetric* $\mathbb{R}^{n\times n}$ *matrices with eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ *and* $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_n$ *respectively. For* $1 \leq r \leq s \leq n$ *fixed, we assume that* $\min\{\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s-1}\} > 0$ *where* $\lambda_0 := \infty$ *and* $\lambda_{n+1} = -\infty$*. Let* $d = s - r + 1$ *and* V *and* $\hat{V}$ *two matrices in* $\mathbb{R}^{n\times d}$ *with columns* $(v_r, v_{r+1}, \dots, v_s)$ *and* $(\hat{v}_r, \hat{v}_{r+1}, \dots, \hat{v}_s)$ *respectively, such that* $\Sigma v_j = \lambda_j v_j$ *and* $\hat{\Sigma}\hat{v}_j = \lambda_j \hat{v}_j$*. Then there exists an orthogonal matrix* $\hat{O}$ *in* $\mathbb{R}^{d\times d}$ *such that*

$$\|\hat{V}\hat{O} - V\|_F \leq \frac{2^{3/2}\min\{\sqrt{d}\|\Sigma - \hat{\Sigma}\|, \|\Sigma - \hat{\Sigma}\|_F\}}{\min\{\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}\}}.$$

Lemma 13 *Let* B *be a* $n \times d$ *matrix with full column rank. Then we have*

$$\|BB^\top - B(B^\top B)^{-1}B^\top\|_F = \|\mathrm{Id}_d - B^\top B\|_F.$$

Proof of Lemma 13. Using the cyclic property of the trace, we have

$$\begin{aligned}
\|BB^\top - B(B^\top B)^{-1}B^\top\|_F^2 &= \|B\left(Id_d - (B^\top B)^{-1}\right)B^\top\|_F^2 \\
&= \mathrm{Tr}\left(B\left(Id_d - (B^\top B)^{-1}\right)B^\top B\left(Id_d - (B^\top B)^{-1}\right)B^\top\right) \\
&= \mathrm{Tr}\left(B^\top B\left(Id_d - (B^\top B)^{-1}\right)B^\top B\left(Id_d - (B^\top B)^{-1}\right)\right) \\
&= \mathrm{Tr}\left(\left(B^\top B - Id_d\right)\left(B^\top B - Id_d\right)\right) \\
&= \|\mathrm{Id}_d - B^\top B\|_F^2.
\end{aligned}$$

■

Theorem 5 *(see [6, Thm VII.3.4]) Let* A *and* B *be two normal operators and* S_1 *and* S_2 *two sets separated by a strip of size* δ*. Let* E *be the orthogonal projection matrix of the eigenspaces of* A *with eigenvalues inside* S_1 *and* F *be the orthogonal projection matrix of the eigenspaces of* B *with eigenvalues inside* S_2*. Then*

$$\|EF\|_F \leq \frac{1}{\delta}\|E(A-B)F\|_F \leq \frac{1}{\delta}\|A-B\|_F.$$

Lemma 14 *(Ostrowski's inequality) Let* $A \in \mathbb{R}^{n\times n}$ *be a Hermitian matrix and* $S \in \mathbb{R}^{d\times n}$ *be a general matrix then*

$$\|SAS^\top - A\|_F \leq \|A\|_F \times \|S^\top S - \mathrm{Id}_n\|.$$