Finding Koopman Invariant Subspaces via Personalized PageRank

Hyukpyo Hong[*]
Department of Mathematics
University of Wisconsin–Madison
Madison, WI 53706, USA
hhong78@wisc.edu

Matthew J. Colbrook
DAMTP, University of Cambridge
Cambridge, CB3 0WA, UK
m.colbrook@damtp.cam.ac.uk

Qin Li [†]
Department of Mathematics
University of Wisconsin–Madison
Madison, WI 53706, USA
qinli@math.wisc.edu

Hanbaek Lyu[†]
Department of Mathematics
University of Wisconsin–Madison
Madison, WI 53706, USA
hlyu@math.wisc.edu

Abstract

Selecting a finite dictionary of observables whose span is Koopman-invariant is a central challenge in data-driven Koopman operator approximation. We address this problem by exploiting zero-block structure in Extended Dynamic Mode Decomposition (EDMD) matrices. We show that any sub-dictionary whose span is Koopman-invariant induces an exact zero block in the EDMD matrix, even for finite data. We then show that such blocks can be detected by applying PageRank to a row-normalized EDMD matrix constructed from a large initial dictionary. The theory extends to approximately invariant subspaces and yields stronger guarantees for personalized PageRank (PPR) when the seed observables lie inside the target block and reach all observables in that block. Combining EDMD concentration bounds with PageRank perturbation theory gives end-to-end detection guarantees with $O(1/\sqrt{M})$ finite-sample scaling and explicit constants. More generally, without assuming an invariant subspace exists, high PPR mass on a sub-dictionary controls discounted multi-step leakage from the seed observables. Numerical experiments on the Duffing oscillator, Van der Pol oscillator, Lorenz system, and a three-well Ramachandran potential suggest that the method identifies compact, interpretable dictionaries with accurate predictions.

1 Introduction

Predicting long-term behavior of a nonlinear dynamical system $\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k)$ is fundamentally difficult: trajectories can exhibit sensitive dependence on initial conditions, complex bifurcation structures, and chaotic attractors. The Koopman operator [1, 2] offers an alternative perspective by shifting attention from the state $\mathbf{x} \in \mathcal{M} \subset \mathbb{R}^d$ to *observables*, i.e., scalar-valued functions $f : \mathcal{M} \to \mathbb{C}$. The operator $\mathcal{K}$ acts by $\mathcal{K}f = f \circ \mathbf{F}$, transforming the nonlinear state-space dynamics into a *linear* evolution in a function space. The price paid for linearity is infinite-dimensionality, but finite-dimensional projections of $\mathcal{K}$ still capture physically meaningful dynamical features. Its growing adoption sparked by [3, 4] is reflected in the coinage "Koopmanism" [5] and thousands of articles over the last decade (see [6, 7, 8, 9, 10] for comprehensive reviews). Some popular

[*]Present address: Korea Institute for Advanced Study, Seoul 02455, Republic of Korea. Email: hhong78@kias.re.kr.
[†]Corresponding authors.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

applications span fluid dynamics [6], system identification [11], neuroscience [12], and molecular dynamics [13].

Extended Dynamic Mode Decomposition (EDMD) [14] approximates $\mathcal{K}$ by projecting onto a finite *dictionary* of observables $\mathcal{D}_{\tilde{N}} = \{\psi_1, \ldots, \psi_{\tilde{N}}\}$, yielding an $\tilde{N} \times \tilde{N}$ matrix $K_{\tilde{N},M}$ estimated from M data pairs $(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})_{i=1,\ldots,M}$ via least squares where $\mathbf{y}^{(i)} = \mathbf{F}(\mathbf{x}^{(i)})$. While several convergence results are now well understood [15, 16, 17, 18], these guarantees depend on the choice of dictionary. In addition, the choice of dictionary strongly affects performance of the approximation [9, 19]. This raises a central practical question: *which observables should the dictionary contain?*

The dictionary learning problem. A natural strategy is to start with a large initial dictionary, for example, all polynomials up to a certain degree, or a dictionary of radial basis functions (RBFs), and hope that they capture relevant dynamics. However, using a large dictionary introduces several difficulties. First, when $\tilde{N}$ is large relative to M, the EDMD matrix can be poorly conditioned, leading to unstable predictions. Second, a large dictionary obscures interpretability: the practitioner cannot tell which observables are relevant and which are harmful. Third, and most fundamentally, EDMD accuracy is not monotone in dictionary size due to spectral pollution [9, 18].

These difficulties motivate the question: *given a large initial dictionary $\mathcal{D}_{\tilde{N}}$, how can we select a small sub-dictionary $\mathcal{D}_N \subset \mathcal{D}_{\tilde{N}}$ that captures essential dynamics?* The ideal sub-dictionary would span a *Koopman invariant subspace* (i.e., a subspace closed under $\mathcal{K}$) so that the EDMD matrix on this subspace exactly represents the dynamics without leakage to the complementary observables.

Our approach. We propose an algorithm that selects a sub-dictionary by exploiting a *zero block structure* of an EDMD matrix. This zero block can be efficiently detected by applying the PageRank (PR) algorithm to the row-normalized EDMD matrix, which reinterprets the matrix as a Markov chain on observables and identifies those forming a "closed community." We show that *personalized PR* (PPR), seeded at observables whose prediction matters most, achieves stronger detection guarantees by eliminating the bias-variance tradeoff inherent in standard PR's uniform teleportation. Because the initial dictionary consists of observables with known functional forms (e.g., polynomials, RBFs), the selected sub-dictionary yields an explicit, interpretable lifted dynamical system.

Contributions.

1. **Zero block structure and invariant subspaces.** We show that any sub-dictionary spanning a Koopman invariant subspace produces an exact zero block in the EDMD matrix, even with finitely many data (Theorem 2). When a spanned subspace is not exactly invariant, the norm of the off-diagonal block serves as a computable surrogate for prediction error (Propositions 1, 8).
2. **PR/PPR-based detection.** To detect zero blocks, we propose Algorithm 1 that reinterprets the row-normalized EDMD matrix as a Markov chain on observables and ranks them via PR or PPR. For an *exact* zero block, PPR succeeds at *any* damping factor $\alpha \in (0, 1)$, whereas standard PR requires a mixing condition on the internal block (Theorem 3). For an *approximate* zero block, PPR still succeeds with a range of a damping factor broader than standard PR.
3. **Finite-sample guarantees.** Combining the perturbation theory with the EDMD matrix concentration bounds yields end-to-end detection guarantees at rate $O(1/\sqrt{M})$ with explicit constants (Theorem 5). The resulting sample-complexity bounds for both PR and PPR show that PPR requires fewer samples than PR (Corollary 5.1). These guarantees hold under i.i.d. sampling and under exponentially mixing trajectory sampling of Markovian or deterministic dynamics (Section 3.4).
4. **Unconditional leakage bounds.** Even without assuming an invariant subspace exists, we bound discounted multi-step leakage from the seed observables by the PPR mass outside the selected sub-dictionary (Proposition 10). This gives a principled justification for PPR-based selection in terms of dynamical leakage, without any block-structure assumption.

1.1 Related work

Koopman operator theory and dictionary learning. Dictionary choices range from classical bases (polynomials or RBFs [14]), kernelized EDMD [14, 20], time-delay embeddings [21], to

neural-network-learned dictionaries [22, 23, 24, 25, 26, 27]. Time-delay dictionaries also connect to the Arnoldi interpretation of DMD [28]. Conradie et al. [29] used a principal angle decomposition for compressing an initial dictionary, selecting modes whose Koopman image remains approximately within the subspace. They also derived error bounds for selected modes, provided a selection criterion, and used multistep bounds to train a dictionary. Haseli and Cortés [30, 31] proposed a subspace decomposition for iterative dictionary pruning, but provide no sample complexity guarantees. In contrast, our method selects an explicit subset of the original dictionary rather than a compressed linear combination of modes, and it provides perturbation and finite-sample guarantees for the resulting selection rule. Our PPR-based sub-dictionary selection criterion has a principled justification through Markov chain theory.

A complementary line of work approximates the Koopman (or transfer) operator directly by a *stochastic matrix*. Classical Ulam-type discretizations of the Perron–Frobenius operator partition the state space and estimate transition probabilities between cells, yielding a row-stochastic matrix whose stationary and spectral properties approximate those of the transfer operator [32, 33, 34]. Transition matrices in our work can be viewed as an EDMD-level analogue of this construction: instead of discretizing the state space, we row-normalize (the transpose of) the learned EDMD matrix to obtain a Markov chain *on observables* whose PPR structure encodes sub-dictionary selection.

PageRank, block structure detection, and spectral methods. The PR algorithm [35] computes stationary distributions of regularized Markov chains [36], with well-studied perturbation theory [37, 38]. Our perturbation analysis builds on the theory of nearly completely decomposable Markov chains [39, 40]. PPR, introduced by Haveliwala [41], was connected to local graph partitioning by Andersen, Chung, and Lang [42]. Block structure detection relates to spectral clustering [43], community detection [44], and Perron cluster analysis (PCCA/PCCA+) [45, 46]; Lai et al. [47] studied PR for block detection in stochastic matrices. Our approach differs in starting from EDMD matrices (requiring row normalization and its perturbation analysis) and targeting Koopman invariant subspaces rather than communities or metastable states. Classical invariant subspace algorithms [48, 49] do not exploit the specific block structure that Theorem 2 guarantees.

2 Algorithm

Given a large initial dictionary $\mathcal{D}_{\tilde{N}} = \{\psi_1, \dots, \psi_{\tilde{N}}\}$, our goal is to select a sub-dictionary $\mathcal{D}_N \subset \mathcal{D}_{\tilde{N}}$ of N observables that captures essential dynamics. We view this sub-dictionary selection problem as a *ranking problem*. Observables receive high PPR scores when they are strongly coupled, through the projected Koopman dynamics, to the seed observables of interest. Our algorithm is essentially a PPR algorithm run on a directed network constructed from the EDMD matrix. Before we state our algorithm, we briefly review EDMD and PPR.

Koopman operator and EDMD. For deterministic dynamics, consider a discrete-time dynamical system $\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k)$ with $\mathbf{F} : \mathcal{M} \to \mathcal{M}$ and $\mathcal{M} \subset \mathbb{R}^d$. Let $\mathcal{F} = L^2(\mathcal{M}, \mu)$ where μ is the data distribution from which data $\mathbf{x}^{(i)}$ is sampled. The induced *Koopman operator* $\mathcal{K} : \mathcal{F} \to \mathcal{F}$ is defined by $\mathcal{K}f = f \circ \mathbf{F}$ for $f \in \mathcal{F}$. Throughout, we assume $\mathcal{K}$ is a bounded operator (see Appendix A for details). When the dynamics is stochastic, i.e. $\mathbf{x}_{k+1} \sim p(\mathbf{x}_k, \cdot)$ for a Markov kernel p, we use the *stochastic Koopman operator* $\mathcal{K}f(\mathbf{x}) = \mathbb{E}[f(\mathbf{x}_{k+1}) \mid \mathbf{x}_k = \mathbf{x}]$, which is a contraction on $L^2(\mathcal{M}, \mu)$ whenever μ is invariant, so boundedness holds automatically. Given a *dictionary* of linearly independent observables $\mathcal{D}_N = \{\psi_1, \dots, \psi_N\} \subset \mathcal{F}$, let $\mathcal{F}_N = \operatorname{span}\{\psi_1, \dots, \psi_N\}$.

With M data pairs $(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})$, where $\mathbf{y}^{(i)} = \mathbf{F}(\mathbf{x}^{(i)})$ for deterministic dynamics and $\mathbf{y}^{(i)} \sim p(\mathbf{x}^{(i)}, \cdot)$ for stochastic dynamics, let

$$\mathbf{\Psi}(\mathbf{x}) = [\psi_1(\mathbf{x}), \dots, \psi_N(\mathbf{x})] \in \mathbb{C}^{1\times N}, \quad \mathbf{\Psi}(\mathbf{X}) = \begin{bmatrix} \mathbf{\Psi}(\mathbf{x}^{(1)})^\intercal & \cdots & \mathbf{\Psi}(\mathbf{x}^{(M)})^\intercal \end{bmatrix}^\intercal \in \mathbb{C}^{M\times N}, \tag{1}$$

and similarly $\mathbf{\Psi}(\mathbf{Y}) \in \mathbb{C}^{M\times N}$. The *EDMD matrix* $K_{N,M}$ is a minimizer of the least-squares problem

$$K_{N,M} \;=\; \mathbf{\Psi}(\mathbf{X})^+ \mathbf{\Psi}(\mathbf{Y}) \;\in\; \operatorname*{arg\,min}_{\tilde{K}\in\mathbb{C}^{N\times N}} \left\| \mathbf{\Psi}(\mathbf{Y}) - \mathbf{\Psi}(\mathbf{X})\tilde{K} \right\|_F^2, \tag{2}$$

where $^+$ represents the Moore-Penrose pseudo inverse. Note that if $\mathcal{F}_N$ is $\mathcal{K}$-invariant, the Frobenius error vanishes and $\mathbf{\Psi}(\mathbf{X})\, K_{N,M} = \mathbf{\Psi}(\mathbf{Y})$, so the EDMD matrix *exactly* captures the Koopman dynamics on $\mathcal{F}_N$.

When the matrix $\mathbf{\Psi}(\mathbf{X})$ has full column rank, $K_{N,M}$ admits the closed-form expression given by the normal equations (see Appendix A.3 for details). To ensure this, we impose the following assumption throughout.

Assumption 1. For a set of N observables $\{\psi_1, \dots, \psi_N\}$ and a set of M data points $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}\}$, the matrix $\mathbf{\Psi}(\mathbf{X})$ in (1) has full column rank.

This is a mild condition asserting that the linear independence of the observables carries over to that of their pointwise evaluations at the data points (see Remark 1 for its measure-theoretic form).

Sampling. We allow independent data and trajectory data. For a Markov kernel p with invariant measure μ, write $d(k) := \sup_{\mathbf{x}\in\mathcal{M}} \|p^k(\mathbf{x}, \cdot) - \mu\|_{\mathrm{TV}}$.

Assumption 2. The M data pairs are generated in one of the following ways.

(S1) *Independent pairs*: the pairs $(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})$ are i.i.d., with $\mathbf{x}^{(i)} \sim \mu$ and either $\mathbf{y}^{(i)} = \mathbf{F}(\mathbf{x}^{(i)})$ or $\mathbf{y}^{(i)} \sim p(\mathbf{x}^{(i)}, \cdot)$.

(S2) *Deterministic trajectory*: $\mathbf{x}^{(i)} = \mathbf{F}^{\,i-1}(\mathbf{x}^{(1)})$ and $\mathbf{y}^{(i)} = \mathbf{x}^{(i+1)}$, where $\mathbf{F}$ and each ψ_i are continuous, μ is a compactly supported *physical measure* of $\mathbf{F}$, the orbit of $\mathbf{x}^{(1)}$ is bounded, and $\mathbf{x}^{(1)}$ lies in the basin $B(\mu)$ (almost surely, if $\mathbf{x}^{(1)}$ is random). Here $B(\mu)$ is the set of $\mathbf{x}$ with $\frac{1}{n}\sum_{k=0}^{n-1} g(\mathbf{F}^k\mathbf{x}) \to \int g\, d\mu$ for every bounded continuous g, and μ is physical if $B(\mu)$ has positive Lebesgue measure.

(S3) *Markov trajectory*: $\mathbf{x}^{(i)} = \mathbf{x}_i$ and $\mathbf{y}^{(i)} = \mathbf{x}_{i+1}$ for a Markov chain $(\mathbf{x}_k)_{k\geq 0}$ with kernel p, invariant measure μ and arbitrary initial distribution, which is uniformly ergodic: $d(k) \to 0$ as $k \to \infty$.

(S3) is an exponential mixing condition. (S2) gives convergence of time averages but no rate (see Section 3.4 for finite-sample results). Neither needs a stationary start. Duffing and Van der Pol use (S1), Ramachandran uses (S3), and Lorenz uses (S2), as the Lorenz attractor carries a physical measure [50]; Lorenz is thus covered asymptotically but not by the finite-sample theory (Remark 3).

Personalized PageRank. Let $Q \in \mathbb{R}^{n\times n}$ be row stochastic, $\mathbf{s} \in \mathbb{R}^n$ a preference vector ($\mathbf{s} \geq 0$, $\mathbf{s}^\intercal\mathbf{1} = 1$), and $\alpha \in [0, 1)$ a damping factor. The *PPR vector* $\pi_{\mathbf{s}}$ is the stationary distribution of the transition kernel $G_{\mathbf{s}} = \alpha Q + (1-\alpha)\mathbf{1}\mathbf{s}^\intercal$, which describes a walker who follows Q with probability α and teleports to a node drawn from $\mathbf{s}$ with probability $1-\alpha$. From the stationary equation $\pi_{\mathbf{s}}^\intercal = \alpha\pi_{\mathbf{s}}^\intercal Q + (1-\alpha)\mathbf{s}^\intercal$ (see Appendix A for the full derivation):

$$\pi_{\mathbf{s}}^\intercal = (1-\alpha)\,\mathbf{s}^\intercal(I - \alpha Q)^{-1}. \tag{3}$$

Since the resolvent $(I-\alpha Q)^{-1}$ is independent of $\mathbf{s}$, the map $\mathbf{s} \mapsto \pi_{\mathbf{s}}$ is *linear*. For a *seed set* $\mathcal{S} \subset \{1, \dots, n\}$ with $\mathbf{s} = \frac{1}{|\mathcal{S}|}\sum_{v\in\mathcal{S}} \mathbf{e}_v$, the *multi-seed PPR vector* is $\pi_{\mathcal{S}} = \frac{1}{|\mathcal{S}|}\sum_{v\in\mathcal{S}} \pi_v$, where π_v is the *single-seed* PPR vector with $\mathcal{S} = \{v\}$. *Standard PR* [35, 36] is a case where $\mathcal{S} = \{1, \dots, n\}$. Thus, the *PR vector* $\pi^\intercal$ is given by $\pi^\intercal = \frac{1}{n}\sum_{v=1}^n \pi_v^\intercal$.

Algorithm: EDMD+PPR. The key idea in our observable ranking algorithm is to interpret the EDMD matrix as a Markov chain on observables to discover zero block structure, indicating Koopman invariant subspaces. Define $P := \mathcal{R}(|K_{\tilde{N},M}^\intercal|)$ where $\mathcal{R}$ is the row-normalization, i.e., the (i,j)-entry of $\mathcal{R}([a_{ij}])$ is $a_{ij}/\sum_{k=1}^n a_{ik}$. The transition probability from observable i to j, $P[i,j]$, is proportional to $|K_{\tilde{N},M}[j,i]|$, the coefficient of ψ_j in the expansion of $\mathcal{K}\psi_i$ — i.e., the coupling strength from ψ_i to ψ_j under the projected dynamics. We then apply PPR with a prescribed *seed set* $\mathcal{S}$ of observables to rank all $\tilde{N}$ observables (Algorithm 1).

If a row of $|K_{\tilde{N},M}^\intercal|$ has zero sum, the row cannot be normalized; such rows (and corresponding columns) are discarded in Step 2. Step 4 *recomputes* the EDMD matrix with the selected N observables, yielding the optimal predictor within $\mathcal{F}_N$, rather than extracting a submatrix of $K_{\tilde{N},M}$.

The seed set may contain observables whose predictions matter most for specific application (e.g., coordinates for state prediction, slow mode observables), or could be taken to be the whole set of $\tilde{N}$ observables if no prior information is available. See Remark 4 for details. The PPR scores on the

Algorithm 1 Sub-dictionary selection via Personalized PageRank on the EDMD matrix

Input: Dictionary $\mathcal{D}_{\tilde{N}} = \{\psi_1, \ldots, \psi_{\tilde{N}}\}$, target size N, seed set $\mathcal{S} \subset \{1, \ldots, \tilde{N}\}$ with $|\mathcal{S}| \leq N$, data $(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})_{i=1}^{M}$, and damping α.
Step 1: Compute the EDMD matrix $K_{\tilde{N},M} = \mathbf{\Psi}_{\tilde{N}}(\mathbf{X})^{+}\, \mathbf{\Psi}_{\tilde{N}}(\mathbf{Y})$.
Step 2: Row-normalize $|K_{\tilde{N},M}^{\intercal}|$ to obtain the row stochastic matrix P.
Step 3: Compute the multi-seed PPR vector $\pi_{\mathcal{S},M} = \frac{1}{|\mathcal{S}|}\sum_{v\in\mathcal{S}} \pi_{v,M}$ of P at damping α; select the set of indices, S_N, with the top-N PPR scores and the corresponding sub-dictionary $\mathcal{D}_N$.
Step 4: Recompute EDMD on the selected sub-dictionary: $K_{N,M} = \mathbf{\Psi}_N(\mathbf{X})^{+}\, \mathbf{\Psi}_N(\mathbf{Y})$.
Output: $\mathcal{D}_N$ and $K_{N,M}$.

seed observables are usually high due to the teleportation, so they are likely to be included in the selected sub-dictionary $\mathcal{D}_N$ consisting of the top N observables with the highest PPR scores.

For a fixed seed preference and damping factor, the selected sub-dictionary $\mathcal{D}_N$ maximizes the sum of the selected PPR scores among all sub-dictionaries of size N. This can be interpreted in terms of multi-step prediction. Let $r_{\max,M} = \|K_{\tilde{N},M}^{\intercal}\|_\infty$ and assume $\alpha < r_{\max,M}$ so that $\gamma := \alpha / r_{\max,M} \in (0,1)$ (automatic when $r_{\max,M} \geq 1$). For any preference vector $\mathbf{s}$ supported on the selected indices S_N, we denote the *γ-discounted multi-step leakage* by $\Lambda_{\mathbf{s}}^{\gamma}(S_N)$. This quantity measures the total γ-discounted weight that escapes S_N under repeated application of $|K_{\tilde{N},M}^{\intercal}|$. We show this satisfies $\Lambda_{\mathbf{s}}^{\gamma}(S_N) \leq \frac{1-\gamma}{1-\alpha}(1 - \pi_{\mathbf{s},M}(S_N))$, where $\pi_{\mathbf{s},M}$ is the PPR vector found in Step 3 (see Proposition 10 in Appendix E.2 for the precise statement). Thus, selecting observables with the top-N PPR scores minimizes this upper bound on discounted multi-step leakage from the seed set. The bound passes through row normalization of $|K_{\tilde{N},M}^{\intercal}|$, which discards row-magnitude information (see Appendix E.2 for its effect on algorithm outputs).

3 Theoretical results

We first present error bounds for prediction accuracy (Section 3.1), develop detection theory for exact and approximate invariant subspaces (Section 3.2), and give finite-sample extensions (Section 3.4). A worked three-state example is in Appendix C, finite-sample details in Appendix F.2, key lemmas in Section 3.3, and proofs in Appendix E. Note that Table 2 summarizes the notation.

3.1 EDMD prediction error

We derive error bounds that reveal the central role of off-diagonal blocks in the EDMD matrix, motivating the entire framework. Let $\mathcal{D} = \{\psi_1, \psi_2, \ldots\}$ be a countable basis for the Hilbert space $\mathcal{F} = L^2(\mathcal{M}, \mu)$. Writing $\mathbf{\Psi} = [\psi_1, \psi_2, \ldots]$ as an infinite row vector of observables and K^{true} for the (infinite) matrix whose j-th column collects the expansion coefficients of $\mathcal{K}\psi_j$ in $\mathcal{D}$, the image of each basis function under the Koopman operator is

$$\mathcal{K}\psi_i \;=\; \mathbf{\Psi}\, K^{\text{true}}\, \mathbf{e}_i \;=\; [\mathbf{\Psi}_N \quad \mathbf{\Psi}_{N^c}] \begin{bmatrix} K^{11,\text{true}} & K^{12,\text{true}} \\ K^{21,\text{true}} & K^{22,\text{true}} \end{bmatrix} \mathbf{e}_i, \tag{4}$$

where $\mathbf{\Psi}_N = [\psi_1, \ldots, \psi_N]$ and $\mathbf{\Psi}_{N^c} = [\psi_{N+1}, \psi_{N+2}, \ldots]$, and the block partition is by the first N rows/columns and their complements. For $i \leq N$, the selector $\mathbf{e}_i$ lies in the left block, so that $\mathcal{K}\psi_i = \mathbf{\Psi}_N K^{11,\text{true}}\mathbf{e}_i + \mathbf{\Psi}_{N^c} K^{21,\text{true}}\mathbf{e}_i$; the second summand is the "leakage" of $\mathcal{K}\psi_i$ outside $\mathcal{F}_N$, carried by column i of the bottom-left block $K^{21,\text{true}}$.

The EDMD approximation of $\mathcal{K}\psi_i$ with a finite dictionary $\mathcal{D}_N$ and finitely many data points $\{\mathbf{x}^{(i)}\}_{i=1}^{M}$ is given by $\mathcal{K}_{N,M}\psi_i := \sum_{j=1}^{N} K_{N,M}[j,i]\psi_j$ where $K_{N,M}$ is the EDMD matrix in (2). We want to reduce the L^2-error of this approximation where the dictionary $\mathcal{D}_N$ is a subset of a larger initial dictionary $\mathcal{D}_{\tilde{N}}$. Thus, we use the EDMD matrix with the initial dictionary $\mathcal{D}_{\tilde{N}}$ to provide an upper bound of the L^2-error. Specifically, we partition the EDMD matrix with $\mathcal{D}_{\tilde{N}}$ as

$$K_{\tilde{N},M} = \begin{bmatrix} K_{\tilde{N},M}^{11} & K_{\tilde{N},M}^{12} \\ K_{\tilde{N},M}^{21} & K_{\tilde{N},M}^{22} \end{bmatrix}, \tag{5}$$

where the rows and columns are divided into two groups: from 1 to N and from $N+1$ to $\tilde{N}$. Then we prove the following proposition.

Proposition 1. *Let $\mathcal{D}_{\tilde{N}} = \{\psi_1, \ldots, \psi_{\tilde{N}}\}$ be an initial dictionary and $\mathcal{D}_N = \{\psi_1, \ldots, \psi_N\}$ be a sub-dictionary of $\mathcal{D}_{\tilde{N}}$. Suppose $\mathcal{D}_{\tilde{N}}$ is an orthonormal set and satisfies Assumption 1 with the data $\{\mathbf{x}^{(i)}\}_{i=1}^M$ generated under Assumption 2. For each $\varepsilon > 0$, almost surely there exists $M_0 > 0$ such that for all $M \geq M_0$,*

$$\sum_{i=1}^{N} \|\mathcal{K}\psi_i - \mathcal{K}_{N,M}\,\psi_i\|_{L^2}^2 \leq 2\left(\|K_{\tilde{N},M}^{21}\|_F^2 + \|K^{\text{true}}[(\tilde{N}+1):\infty,\, 1:\tilde{N}]\|_F^2\right) + \varepsilon. \tag{6}$$

Importantly, Proposition 1 applies to *any* sub-dictionary of size N formed from $\mathcal{D}_{\tilde{N}}$: relabeling the observables by any permutation of $\{1, \ldots, \tilde{N}\}$ and taking the first N elements under the permutation yields a valid instance of the proposition, with the off-diagonal block $K_{\tilde{N},M}^{21}$ recomputed accordingly.

On the right-hand side, the first term $\|K_{\tilde{N},M}^{21}\|_F^2$ is *directly computable from data*; the second reflects truncation beyond $\tilde{N}$ and depends on the initial dictionary $\mathcal{D}_{\tilde{N}}$, not the sub-dictionary $\mathcal{D}_N$. The practical implication is therefore: *to find the best sub-dictionary of size N, one should search for the permutation $\{\psi_1, \ldots, \psi_{\tilde{N}}\}$ that minimizes the Frobenius norm of the resulting off-diagonal block, $\|K_{\tilde{N},M}^{21}\|_F^2$.* The ideal case is when this block is exactly zero for some permutation. Exact zero blocks correspond precisely to Koopman invariant subspaces within the span of the initial dictionary.

We now show that EDMD "discovers" these invariant subspaces automatically. Note that the initial dictionary $\mathcal{D}_{\tilde{N}}$ in Theorem 2 is not necessarily orthonormal, unlike Proposition 1.

Theorem 2. *For deterministic dynamics $\mathbf{y}^{(i)} = \mathbf{F}(\mathbf{x}^{(i)})$, let $\mathcal{D}_{\tilde{N}} = \{\psi_1, \ldots, \psi_{\tilde{N}}\}$ be an initial dictionary (not necessarily spanning an invariant subspace) and $\{\mathbf{x}^{(i)}\}_{i=1}^M$ be a set of data points, which satisfy Assumption 1. If the sub-dictionary $\mathcal{D}_N = \{\psi_1, \ldots, \psi_N\} \subset \mathcal{D}_{\tilde{N}}$ spans a Koopman invariant subspace, then*

(i) *The top-left $N \times N$ submatrix of $K_{\tilde{N},M}$ is identical to $K_{N,M}$.*

(ii) *The bottom-left $(\tilde{N} - N) \times N$ submatrix of $K_{\tilde{N},M}$ (i.e., $K_{\tilde{N},M}^{21}$) is identically zero.*

This theorem suggests that if there is any sub-dictionary that spans a Koopman invariant subspace such a zero block structure emerges up to permutation. Similar results when $M = \infty$ or $\mathcal{D}_N$ is chosen as the initial dictionary were noted in literature [14, 34, 15, 51]. We emphasize that the result holds for $M < \infty$ as well even with a larger initial dictionary, making ours more computationally feasible. See Appendix B for illustration with a toy model.

The proof of Theorem 2 is an algebraic identity that uses the data only through Assumption 1 and the exact one-step pairs $\mathbf{y}^{(i)} = \mathbf{F}(\mathbf{x}^{(i)})$; independence of $\{\mathbf{x}^{(i)}\}$ is not used. Thus, the exact zero block holds verbatim for data that is non-independent, such as correlated trajectory data in regime (S2).

These results recast the problem of finding an invariant subspace as identifying a zero block structure in the EDMD matrix. However, exhaustive search over all $\binom{\tilde{N}}{N}$ sub-dictionaries to obtain the clear zero block submatrix at the bottom-left corner is combinatorially prohibitive. Our framework (Algorithm 1) circumvents this obstacle by considering the Markov chain induced by the EDMD matrix, while also handling the practically relevant case of approximate invariance.

3.2 Finding Koopman invariant subspaces via personalized PageRank

We now develop a finer theory for the case when the EDMD matrix admits (approximate) zero block structure. Let $P = \mathcal{R}(|K_{\tilde{N},M}^{\intercal}|)$ be the row normalization of $|K_{\tilde{N},M}^{\intercal}|$. Let $K_{\tilde{N},M}^{(0)}$ denote $K_{\tilde{N},M}$ with the bottom-left block zeroed, and define the row-stochastic matrix $P^{(0)} := \mathcal{R}(|K_{\tilde{N},M}^{(0)|\intercal}|)$. Equivalently, $P^{(0)}$ is obtained from P by zeroing its top-right block P^{12} and *renormalizing the rows of the top block*, so that $P^{11,(0)} = \mathcal{R}(P^{11})$, $P^{21,(0)} = P^{21}$, $P^{22,(0)} = P^{22}$, with the top-right block identically zero, i.e., $P^{12,(0)} = O$. All matrices are partitioned conformally with (5). The coupling parameter is $\eta := 1 - \|P^{22}\|_\infty > 0$.

Algorithm 1 ranks observables by the PR or PPR vector $\pi_{\mathbf{s}}$ of P (with $\mathbf{s} = \mathbf{1}/\tilde{N}$ for PR, $\mathbf{s} = |\mathcal{S}|^{-1}\sum_{v\in\mathcal{S}} \mathbf{e}_v$ for PPR), so it correctly identifies $\mathcal{D}_N$ iff the corresponding gap on P is positive:

$$\Delta_{\text{PR}} := \min_{i\leq N} \pi(i) - \max_{j>N} \pi(j) > 0, \qquad \Delta_{\text{PPR}} := \min_{i\leq N} \pi_{\mathcal{S}}(i) - \max_{j>N} \pi_{\mathcal{S}}(j) > 0. \tag{7}$$

Theorem 3 provides sufficient conditions for detection.

Theorem 3 (Koopman invariant subspace detection)**.** *Under Assumption 1 and $\alpha \in (0,1)$:*

(i) Standard PR *($\mathcal{S} = \{1, \dots, \tilde{N}\}$). Let $(\mathbf{q}^\alpha)^\intercal := \frac{1-\alpha}{N} \mathbf{1}_N^\intercal (I - \alpha P^{11,(0)})^{-1}$ denote the standard PR vector of $P^{11,(0)}$ and let $q^\alpha_{\min} := \min_{i \le N} q^\alpha_i$. Then detection is successful (i.e., $\Delta_{\text{PR}} > 0$) provided*

$$\|P^{12}\|_\infty \;<\; \frac{(1-\alpha)N}{4\alpha\tilde{N}} \left[q^\alpha_{\min} - \frac{(\tilde{N}-N)(1-\alpha)}{N(1-\alpha+\alpha\eta)} \right]. \tag{8}$$

(ii) Multi-seed PPR. *Let $\mathcal{S} \subset \{1, \dots, N\}$ satisfy the reachability condition: every $i \le N$ is reachable from some $v \in \mathcal{S}$ in $P^{11,(0)}$. Then detection is successful (i.e., $\Delta_{\text{PPR}} > 0$) provided*

$$\|P^{12}\|_\infty \;<\; \frac{(1-\alpha)^2}{4\alpha|\mathcal{S}|} \min_{i \le N} \sum_{v \in \mathcal{S}} \left[(I - \alpha P^{11,(0)})^{-1} \right]_{vi}. \tag{9}$$

In particular, when $\mathcal{F}_N$ is $\mathcal{K}$-invariant, $P^{12} = O$ by Theorem 2, so (9) holds: $\pi_{\mathcal{S}} = \pi^{(0)}_{\mathcal{S}}$, $\pi_{\mathcal{S}}(j) = 0$ for all $j > N$, and the recomputed $K_{N,M}$ coincides with the population-level Koopman matrix $K_N := \lim_{M \to \infty} K_{N,M}$.

The proof of the above result is in Appendix E.3. A brief sketch of the proof is provided in Sec. 3.3. To demonstrate the implications of Theorem 3, in Figure 1 we plot the detection window $\alpha^*(\epsilon) := \sup\{\alpha : \text{Theorem 3(i) or (ii) holds}\}$ as a function of leakage $\epsilon := \|P^{12}\|_\infty$ for a simple 3-dimensional example. A wider window means detection succeeds for a broader range of the damping factor α. PPR (red) dominates PR (blue) in both panels (see Appendix C for details). In Panel (a) (well-mixed internal block), PPR tolerates leakage up to $\epsilon = 1/4 = 0.25$, whereas PR fails beyond $\epsilon = 1/8 = 0.125$. The gap is more dramatic in Panel (b), where the internal block has a *starved node* (a node receiving no inflow from within the block): PR's window collapses beyond $\epsilon = 1/24 \approx 0.042$, while PPR remains effective up to $\epsilon = 1/4 = 0.25$. The underlying reason is structural: PR requires the internal block $P^{11,(0)}$ to be well-mixed, a condition that fails whenever a starved node is present, whereas PPR requires only that the seed set $\mathcal{S}$ can reach every node in the block, which is a strictly weaker condition that holds independently of α.

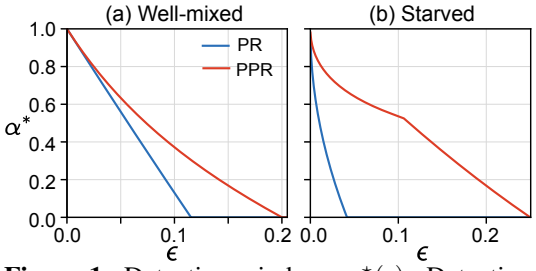

Figure 1: Detection windows $\alpha^\star(\epsilon)$. Detection holds below each curve.

3.3 Sketch of proof for Theorem 3

Full proofs of our main results are in Appendix E.3. Here we give a brief sketch of the proof of Theorem 3. Observe that, due to the block lower-triangular structure, it may be easier to detect on the proxy matrix $P^{(0)}$. Accordingly, define the auxiliary gaps

$$\Delta^{(0)}_{\text{PR}} := \min_{i \le N} \pi^{(0)}(i) - \max_{j > N} \pi^{(0)}(j), \qquad \Delta^{(0)}_{\text{PPR}} := \min_{i \le N} \pi^{(0)}_{\mathcal{S}}(i) - \max_{j > N} \pi^{(0)}_{\mathcal{S}}(j), \tag{10}$$

where $\pi^{(0)}, \pi^{(0)}_{\mathcal{S}}$ are the PR/PPR vectors of $P^{(0)}$. If the off-diagonal block P^{12} is small, then P and $P^{(0)}$ are close so they should have similar PR/PPR vectors. More precisely, we show (see Lemma 15 in Appendix E.3) the PR/PPR vectors of P and $P^{(0)}$ differ in ℓ^1 by at most $\frac{2\alpha}{1-\alpha}\|P^{12}\|_\infty$.

Therefore, detection on P succeeds if detection on $P^{(0)}$ holds by a large enough margin compared to the off-diagonal block P^{12}: For method $\in \{\text{PR}, \text{PPR}\}$,

$$\Delta_{\text{method}} > 0 \quad \Longleftarrow \quad \|P^{12}\|_\infty \;<\; \frac{1-\alpha}{4\alpha} \Delta^{(0)}_{\text{method}}. \tag{11}$$

Now, due to the block-triangular structure of $P^{(0)}$, we can establish the following closed-form expressions for the auxiliary gaps $\Delta^{(0)}_{\text{PR}}$ and $\Delta^{(0)}_{\text{PPR}}$ in terms of the diagonal blocks of the resolvent $(I - \alpha P^{(0)})^{-1}$, namely $R_{11} := (I - \alpha P^{11,(0)})^{-1}$ and $R_{22} := (I - \alpha P^{22,(0)})^{-1}$:

Lemma 4 (Closed-form auxiliary gaps)**.** *For $\alpha \in (0,1)$, both R_{11} and R_{22} are entrywise nonnegative with finite entries, and*

$$\Delta_{\text{PR}}^{(0)} \;=\; \frac{1-\alpha}{\tilde{N}}\left[\min_{i\le N}\sum_{v\le N} s_v^{\text{eff}}\,[R_{11}]_{vi} \;-\; \max_{j>N}\sum_{w>N}[R_{22}]_{wj}\right],\; \Delta_{\text{PPR}}^{(0)} \;=\; \min_{i\le N}\frac{1-\alpha}{|\mathcal{S}|}\sum_{v\in\mathcal{S}}[R_{11}]_{vi}, \tag{12}$$

where $s_v^{\text{eff}} := 1+\alpha\sum_{w>N}[R_{22}P^{21,(0)}]_{wv} \ge 1$ for $v \le N$. Moreover, $\pi_{\mathcal{S}}^{(0)}(j) = 0$ for every $j > N$, so the max-leakage term in (10) *vanishes identically. Here, row and column indices of R_{22} start at $N+1$, consistent with the full $\tilde{N}\times\tilde{N}$ indexing.*

Obtaining explicit lower bounds on the auxiliary gaps $\Delta_{\text{method}}^{(0)}$ from the above expressions and combining with (11) then yields the sufficient condition for detection on P in Theorem 3. The two expressions enter Theorem 3 in structurally different ways.

PPR: direct use in (ii). The closed form (12) depends only on R_{11} evaluated at the seed set $\mathcal{S}$ and is directly usable as a detection threshold. Positivity $\Delta_{\text{PPR}}^{(0)} > 0$ reduces to reachability—every $i \le N$ must be reachable from some $v \in \mathcal{S}$ in $P^{11,(0)}$—because $[R_{11}]_{vi} > 0$ iff v reaches i in $P^{11,(0)}$.

PR: relaxation to a usable bound for (i). The closed form (12) couples R_{11} to bottom block quantities through s^{eff} and the column sums of R_{22}, obstructing a clean condition on $P^{11,(0)}$ alone. Dropping $s_v^{\text{eff}} \ge 1$ to 1 and bounding $\|R_{22}\|_\infty \le 1/(1-\alpha+\alpha\eta)$ (from $\|P^{22}\|_\infty \le 1-\eta$) yields the *mixing relaxation*

$$\Delta_{\text{PR}}^{(0)} \;\ge\; \frac{N}{\tilde{N}}\,q_{\min}^{\alpha} \;-\; \frac{(\tilde{N}-N)(1-\alpha)}{\tilde{N}(1-\alpha+\alpha\eta)}, \tag{13}$$

where $(\mathbf{q}^\alpha)^\intercal := \frac{1-\alpha}{N}\mathbf{1}_N^\intercal R_{11}$ and $q_{\min}^\alpha := \min_{i\le N} q_i^\alpha$. Positivity of the right-hand side requires the *mixing condition* $q_{\min}^\alpha > (\tilde{N}-N)(1-\alpha)/\big(N(1-\alpha+\alpha\eta)\big)$, which is stricter than reachability.

3.4 Finite-sample guarantees

To establish finite-sample error bounds and detection guarantees, we additionally assume the dictionary and its Koopman images are uniformly bounded:

Assumption 3. There exists a positive constant $D > 0$ such that $\sup_{\mathbf{x}\in\mathcal{M}}|\psi_i(\mathbf{x})| \le D$ and $\sup_{\mathbf{x}\in\mathcal{M}}|\mathcal{K}\psi_i(\mathbf{x})| \le D$ for all $i \in \mathbb{N}$.

Under Assumption 3, matrix Bernstein inequalities yield $\|K_{\tilde{N},M}^\intercal - K_{\tilde{N}}^\intercal\|_\infty \le \varepsilon_M = O(1/\sqrt{M})$ for M sufficiently large (Proposition 17 in Appendix F). This concentration bound is the *only* place where the sampling regime enters the finite-sample analysis. We therefore state the end-to-end result in terms of ε_M, and record below what ε_M is in each regime. Under (S3), the same bound holds with M replaced by M/τ_{mix} for a mixing constant τ_{mix} (Proposition 18). Let $\lambda_{\min}$ denote the smallest eigenvalue (in modulus) of the Gram matrix $G_{\tilde{N}} := [\langle\psi_i,\psi_j\rangle_{L^2(\mu)}]_{i,j=1}^{\tilde{N}}$. Throughout this subsection, let $P_{\text{pop}} = \mathcal{R}(|K_{\tilde{N}}^\intercal|)$ be the row-normalized population-level EDMD matrix with top-right block P_{pop}^{12}, and let $P_{\text{pop}}^{(0)} := \mathcal{R}(|K_{\tilde{N}}^{(0)}|^\intercal)$ be the row-stochastic matrix obtained from P_{pop} by zeroing P_{pop}^{12} and renormalizing the rows of the top block. Let $K_{\tilde{N}}^{(0)}$ denote $K_{\tilde{N}}$ with its bottom-left block zeroed, and set $r_{\min}^{(0)} := \min_i\sum_j|(K_{\tilde{N}}^{(0)})^\intercal[i,j]|$ and $r_{\max} := \|K_{\tilde{N}}^\intercal\|_\infty$. Define $\varepsilon_0 := \|(K_{\tilde{N}}^{(0)} - K_{\tilde{N}})^\intercal\|_\infty = \max_{i\le N}\sum_{j>N}|K_{\tilde{N}}^\intercal[i,j]|$ as the population operator-level leakage, which satisfies $\|P_{\text{pop}}^{12}\|_\infty \le \varepsilon_0/r_{\min}^{(0)}$. Let $\pi_{\text{pop}}^{(0)}$ denote the PR (or PPR) vector of the row-stochastic matrix $P_{\text{pop}}^{(0)}$. Combining with the perturbation theory:

Theorem 5 (End-to-end detection)**.** *Let $\mathcal{D}_{\tilde{N}}$ satisfy Assumptions 1 and 3 with the data satisfying Assumption 2, and let $\rho \in (0,1/2)$. Under (S1), let $\varepsilon_M = C_{\text{EDMD}}\sqrt{(2/M)\log(2\tilde{N}/\rho)}$ and $M \ge \frac{32\,\tilde{N}^2D^4}{\lambda_{\min}^2}\log\frac{2\tilde{N}}{\rho}$ (see Appendix F for the definition of C_{EDMD}); under (S3), let $\varepsilon_M = \varepsilon_M^{\text{mix}}$ and M be as in Proposition 18. Let π_M be the PR (or PPR) vector of $P = \mathcal{R}(|K_{\tilde{N},M}^\intercal|)$. Then with probability $\ge 1-2\rho$, we have $\|\pi_M - \pi_{\text{pop}}^{(0)}\|_1 \;\le\; \frac{2\alpha(\varepsilon_0+\varepsilon_M)}{(1-\alpha)\,r_{\min}^{(0)}}$.*

Theorem 5 also holds for deterministic trajectories under a variant (S2′) of (S2), introduced in Appendix F.1, which adds an exponential concentration inequality satisfied by systems modelled by Young towers. This result, combined with Theorem 3, provides the sample-complexity bounds for a correct detection.

Corollary 5.1 (Sample complexity for PPR detection with $\alpha = 1/2$)**.** *Let $\Delta^{(0)}_{\mathrm{PPR}}$ be the PPR auxiliary gap* (10) *of Section 3.3 computed from $P^{(0)}_{\mathrm{pop}}$. Under (S1) of Assumption 2, if $\Delta^{(0)}_{\mathrm{PPR}} > 0$ and $\|P^{12}_{\mathrm{pop}}\|_\infty < r^{(0)}_{\min}\Delta^{(0)}_{\mathrm{PPR}}/(4r_{\max})$, detection by Algorithm 1 (i.e. $\Delta_{\mathrm{PPR}} > 0$) holds with probability $\geq 1 - 2\rho$ for*

$$M \;\geq\; \max\left\{ \frac{32\,\tilde{N}^2 D^4}{\lambda^2_{\min}} \log\frac{2\tilde{N}}{\rho},\;\; \frac{32\log(2\tilde{N}/\rho)\,C^2_{EDMD}}{(r^{(0)}_{\min})^2\big(\Delta^{(0)}_{\mathrm{PPR}} - 4r_{\max}\|P^{12}_{\mathrm{pop}}\|_\infty/r^{(0)}_{\min}\big)^2} \right\}. \tag{14}$$

The bound is of the form $M \gtrsim C^2_{\mathrm{EDMD}}\log(\tilde{N}/\rho)/\big(\Delta^{(0)}_{\mathrm{PPR}} - c\,(r_{\max}/r^{(0)}_{\min})\|P^{12}_{\mathrm{pop}}\|_\infty\big)^2$: the leakage shrinks the effective gap from $\Delta^{(0)}_{\mathrm{PPR}}$ to $\Delta^{(0)}_{\mathrm{PPR}} - c\,(r_{\max}/r^{(0)}_{\min})\|P^{12}_{\mathrm{pop}}\|_\infty$ and inflates the sample requirement quadratically. The $\alpha = 1/2$ specialization eliminates any α-dependent prefactor: because PPR leaks no mass outside $\mathcal{D}_N$, $\Delta^{(0)}_{\mathrm{PPR}}$ stays positive without requiring α near 1. The proof, the analogous PR sample complexity bound (Corollary 18.1), and the related finite-sample leakage bound (Corollary 18.2) are deferred to Appendix F.2.

4 Numerical experiments

We now test Algorithm 1 on four systems of increasing complexity. For error plots, we use 20 random seeds and depict the mean±1SD by shaded regions. Full experimental details (dictionaries, parameters, precise definitions of errors, additional figures) are in Appendix D, and all relevant codes can be found at `https://github.com/HyukpyoHong/Koopman-PageRank`.

Duffing and Van der Pol oscillators. We first apply the method to the Duffing oscillator ($\dot{x} = y$, $\dot{y} = -\delta y + \gamma x - \beta x^3$) and the Van der Pol oscillator, using 90 bivariate Laguerre polynomials with $M = 2000$ samples. In both systems, the PPR-ordered EDMD matrix reveals a clear approximate zero block structure (Figs. 5–7 in Appendix D). The PPR-selected sub-dictionary achieves $\sim 10^{-12}$ one-step error with only 5 observables, outperforming the others in multi-step state predictions as well (Fig. 2). For the tunable symmetric subspace decomposition (TSSD) [31], we varied the accuracy parameter ϵ from 0 to 1 with 0.01 intervals, but only three sizes $N = 3, 45, 65$ are detected. For the Van der Pol oscillator, the advantage of PPR over standard PR is more pronounced (Fig. 6 in Appendix D).

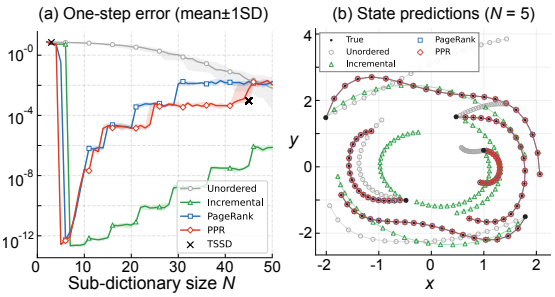

Figure 2: Duffing oscillator. (a) One-step error vs. sub-dictionary size. (b) 20-step state predictions with 5 observables.

Three-well Ramachandran potential. We test on a molecular dynamics benchmark: overdamped Langevin diffusion on a three-well Ramachandran potential [52] (see Appendix D for details). The dictionary consists of $\tilde{N} = 196$ observables (Fourier modes, cross terms, Gaussian RBFs, and 4 state coordinates $\sin\phi, \cos\phi, \sin\psi, \cos\psi$), with $M = 10^5$ and 20 random seeds. We compare PPR (seeded at the 4 state coordinates, since predicting the angular state is the primary objective), standard PR, time-lagged independent component analysis (TICA) [53], PCCA+ [45], and random ordering. The data satisfy (S3) of Assumption 2 (Remark 3).

Table 1: Ratio of the error for each ordering to the error for a random ordering, computed within each trajectory and reported as mean ± standard deviation over 20 random seeds. Values below 1 indicate improvement over random ordering.

Method	$N=5$	$N=10$	$N=20$
PPR	**0.29 ± 0.05**	**0.38 ± 0.07**	**0.52 ± 0.07**
PR	1.02 ± 0.14	1.07 ± 0.17	0.97 ± 0.15
PCCA+	0.67 ± 0.09	0.78 ± 0.13	0.94 ± 0.14
TICA	1.01 ± 0.16	1.07 ± 0.18	1.13 ± 0.20

For one-step predictions, PPR outperforms all other methods at small sub-dictionary sizes (Table 1): at $N = 5$, PPR reduces one-step error by 71% over Random, compared to 33% for PCCA+, while TICA performs comparably to Random. The advantage persists throughout the plotted range of N. Fig. 3 shows that PPR achieves the lowest one-step and multi-step prediction errors at all horizons, confirming good internal dynamics propagation.

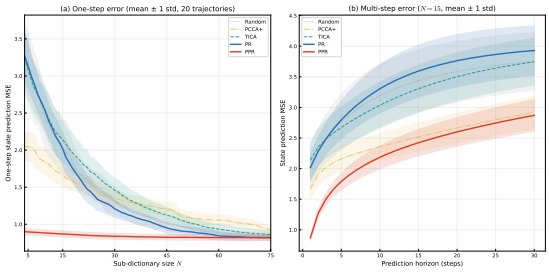

Figure 3: Three-well Ramachandran potential ($\tilde{N} = 196$). (a) One-step error vs. sub-dictionary size. (b) Multi-step prediction errors at $N = 15$.

The composition of selected observables is physically revealing: PPR places the 4 state coordinates at ranks 1–4, followed by basin-boundary RBFs (Fig. 9 in Appendix D). The four state coordinates capture three macrostates. On the other hand, the basin-boundary RBFs seem to effectively capture inter-basin transitions, yielding accurate predictions with smaller sub-dictionaries.

Lorenz system. We apply Algorithm 1 to the Lorenz system in the chaotic regime ($\sigma = 10, \rho = 28, \beta = 8/3$), using a time-delay embedding with $\tilde{N} = 300$ observables and $M = 2 \times 10^5$ snapshot pairs. The data satisfy (S2) of Assumption 2, so our asymptotic results cover this experiment (Lemma 7), but our finite-sample theorem does not (Remark 3). Rather than prediction error, we examine whether a PPR-selected sub-dictionary recovers dynamically relevant pseudo-eigenfunctions. Fig. 4 shows the real parts of the pseudo-eigenfunctions at frequency ≈ 6 rad/s: PPR captures the characteristic oscillatory structure wrapping around unstable periodic orbits [9, 18] at $N = 30$–60, while Incremental ordering requires $N \geq 120$. This demonstrates that PPR identifies a compressed, informative sub-dictionary for Koopman spectral analysis.

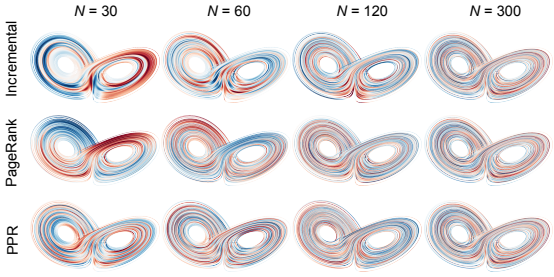

Figure 4: Pseudo-eigenfunctions (≈ 6 rad/s) on the Lorenz attractor for varying sub-dictionary sizes N. PPR recovers the oscillatory structure at $N = 30$–60, while Incremental requires $N \geq 120$.

5 Limitations

PPR can be sensitive to the damping factor α, although held-out prediction error in Ramachandran remains stable across the tested values (see Table 3 in Appendix D.5). Also, PPR performance depends on the choice of seed observables. When seed observables are not obvious, one fallback is to perform PR first and pick top observables from PR as seed observables for PPR (see Table 4 in Appendix D.5). In our framework, row normalization discards row-magnitude information and thus can change observable rankings (Appendix E.2). Least-squares EDMD fitting is sensitive to outliers, and robustness to contaminated data is not addressed here. For stochastic systems, our analysis concerns conditional-mean prediction; pathwise prediction variance remains open, and stochastic ResDMD [54] is a natural tool for it.

6 Conclusions

We developed a framework connecting zero block structures in EDMD matrices to Koopman invariant subspaces, and proposed Algorithm 1, which applies PPR to a row-normalized EDMD matrix to select sub-dictionaries. The framework provides finite-sample detection guarantees. More broadly, discounted multi-step Koopman leakage is bounded by the PPR mass outside the selected set (Proposition 10), providing a leakage-based justification for PPR selection without assuming invariant subspace existence.

Experiments on the Duffing and Van der Pol oscillators, the Lorenz system, and the three-well Ramachandran potential confirm that PPR identifies compact, interpretable dictionaries, reducing one-step prediction error by up to $\sim 70\%$ over random selection at a small sub-dictionary size N (see Table 1). Future directions include constructing invariant structures beyond subsets of a prescribed dictionary and adaptive selection of the damping factor α.

Acknowledgments and Disclosure of Funding

Q.L. acknowledges support from NSF-DMS-2308440, ONR-GRANT14408071 and Vilas Associate Award. H.L. is partially supported by NSF CAREER grant DMS-2541878 and NSF DMS-2206296. H.H. was supported in part by a KIAS Individual Grant (AP111101) via the Center for Artificial Intelligence and Natural Sciences at Korea Institute for Advanced Study.

The authors confirm that there is no conflict of interest related to the manuscript.

References

[1] Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

[2] B. O. Koopman and J. v. Neumann. Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences*, 18(3):255–263, 1932.

[3] Igor Mezić and Andrzej Banaszuk. Comparison of systems with complex behavior. *Physica D: Nonlinear Phenomena*, 197(1):101–133, 2004.

[4] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1):309–325, Aug 2005.

[5] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 12 2012.

[6] Igor Mezić. Analysis of fluid flows via spectral properties of the koopman operator. *Annual Review of Fluid Mechanics*, 45(Volume 45, 2013):357–378, 2013.

[7] Peter J. Schmid. Dynamic mode decomposition and its variants. *Annual Review of Fluid Mechanics*, 54:225–254, 2022.

[8] Steven L. Brunton, Marko Budišić, Eurika Kaiser, and J. Nathan Kutz. Modern koopman theory for dynamical systems. *SIAM Review*, 64(2):229–340, 2022.

[9] Matthew J. Colbrook. Chapter 4 - the multiverse of dynamic mode decomposition algorithms. In Siddhartha Mishra and Alex Townsend, editors, *Numerical Analysis Meets Machine Learning*, volume 25 of *Handbook of Numerical Analysis*, pages 127–230. Elsevier, 2024.

[10] Matthew Colbrook, Zlatko Drmač, and Andrew Horning. An introductory guide to koopman learning. In Daniel Alpay, Fabrizio Colombo, and Irene Sabadini, editors, *Operator Theory*, pages 1–49. Springer Basel, Basel, 2026.

[11] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

[12] Bingni W. Brunton, Lise A. Johnson, Jeffrey G. Ojemann, and J. Nathan Kutz. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1–15, 2016.

[13] Hao Wu, Feliks Nüske, Fabian Paul, Stefan Klus, Péter Koltai, and Frank Noé. Variational koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations. *The Journal of chemical physics*, 146(15), 2017.

[14] Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.

[15] Milan Korda and Igor Mezić. On convergence of extended dynamic mode decomposition to the koopman operator. *Journal of Nonlinear Science*, 28(2):687–710, 2018.

[16] Vladimir Kostic, Karim Lounici, Pietro Novelli, and Massimiliano Pontil. Sharp spectral rates for koopman operator learning. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 32328–32339. Curran Associates, Inc., 2023.

[17] Feliks Nüske, Sebastian Peitz, Friedrich Philipp, Manuel Schaller, and Karl Worthmann. Finite-data error bounds for koopman-based prediction and control. *Journal of Nonlinear Science*, 33(1):14, 2023.

[18] Matthew J Colbrook and Alex Townsend. Rigorous data-driven computation of spectral properties of koopman operators for dynamical systems. *Communications on Pure and Applied Mathematics*, 77(1):221–283, 2024.

[19] Samuel E Otto and Clarence W Rowley. Koopman operators for estimation and control of dynamical systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):59–87, 2021.

[20] Yoshinobu Kawahara. Dynamic mode decomposition with reproducing kernels for koopman spectral analysis. *Advances in neural information processing systems*, 29, 2016.

[21] Hassan Arbabi and Igor Mezic. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator. *SIAM Journal on Applied Dynamical Systems*, 16(4):2096–2126, 2017.

[22] Daniel J Alford-Lago, Christopher W Curtis, Alexander T Ihler, and Opal Issan. Deep learning enhanced dynamic mode decomposition. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(3), 2022.

[23] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning koopman invariant subspaces for dynamic mode decomposition. *Advances in neural information processing systems*, 30, 2017.

[24] Qianxiao Li, Felix Dietrich, Erik M Bollt, and Ioannis G Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10), 2017.

[25] Christoph Wehmeyer and Frank Noé. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of chemical physics*, 148(24), 2018.

[26] Yuhuang Meng, Jianguo Huang, and Yue Qiu. Koopman operator learning using invertible neural networks. *Journal of Computational Physics*, 501:112795, 2024.

[27] Enoch Yeung, Soumya Kundu, and Nathan Hodas. Learning deep neural network representations for koopman operators of nonlinear dynamical systems. In *2019 American Control Conference (ACC)*, pages 4832–4839. IEEE, 2019.

[28] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1(2):391–421, 2014.

[29] Gustav Conradie, Nicolas Boullé, Jean-Christophe Loiseau, Steven L Brunton, and Matthew J Colbrook. Trustworthy koopman operator learning: Invariance diagnostics and error bounds. *arXiv preprint arXiv:2603.15091*, 2026.

[30] Masih Haseli and Jorge Cortés. Learning koopman eigenfunctions and invariant subspaces from data: Symmetric subspace decomposition. *IEEE Transactions on Automatic Control*, 67(7):3442–3457, 2021.

[31] Masih Haseli and Jorge Cortés. Generalizing dynamic mode decomposition: Balancing accuracy and expressiveness in koopman approximations. *Automatica*, 153:111001, 2023.

[32] Stanislaw M Ulam. *Problems in modern mathematics*. Courier Corporation, 2004.

[33] Tien-Yien Li. Finite approximation for the frobenius-perron operator. a solution to ulam's conjecture. *Journal of Approximation Theory*, 17(2):177–186, 1976.

[34] Stefan Klus, Péter Koltai, and Christof Schütte. On the numerical approximation of the perron-frobenius and koopman operator. *Journal of Computational Dynamics*, 3(1):51–79, 2016.

[35] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.

[36] Amy N Langville and Carl D Meyer. Deeper inside pagerank. *Internet Mathematics*, 1(3):335–380, 2004.

[37] Ilse C. F. Ipsen and Rebecca S. Wills. Mathematical properties and analysis of Google's PageRank. *Boletín de la Sociedad Española de Matemática Aplicada*, 34:191–196, 2006.

[38] Carl D. Meyer. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Review*, 31(2):240–272, 1989.

[39] Herbert A. Simon and Albert Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29(2):111–138, 1961.

[40] Pierre-Jacques Courtois. *Decomposability: Queueing and Computer System Applications*. Academic Press, New York, 1977.

[41] Taher H. Haveliwala. Topic-sensitive PageRank: A context-sensitive ranking algorithm for web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):784–796, 2003.

[42] Reid Andersen, Fan Chung, and Kevin Lang. Local graph partitioning using PageRank vectors. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 475–486. IEEE, 2006.

[43] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[44] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3–5):75–174, 2010.

[45] Peter Deuflhard and Marcus Weber. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications*, 398:161–184, 2005.

[46] Susanna Röblitz and Marcus Weber. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Advances in Data Analysis and Classification*, 7(2):147–179, 2013.

[47] Darong Lai, Hongtao Lu, and Christine Nardini. Finding communities in directed networks by pagerank random walk induced network embedding. *Physica A: Statistical Mechanics and its Applications*, 389(12):2443–2454, 2010.

[48] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 4th edition, 2013.

[49] G. W. Stewart. *Matrix Algorithms, Volume II: Eigensystems*. SIAM, 2001.

[50] Vitor Araújo, Maria José Pacifico, Enrique R. Pujals, and Marcelo Viana. Singular-hyperbolic attractors are chaotic. *Transactions of the American Mathematical Society*, 361(5):2431–2485, 2009.

[51] Igor Mezić. Koopman operator, geometry, and learning of dynamical systems. *Not. Am. Math. Soc*, 68(7):1087–1105, 2021.

[52] G N Ramachandran, C Ramakrishnan, and V Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, 1963.

[53] Guillermo Pérez-Hernández, Fabian Paul, Toni Giorgino, Gianni De Fabritiis, and Frank Noé. Identification of slow molecular order parameters for Markov model construction. *The Journal of Chemical Physics*, 139(1):015102, 2013.

[54] Matthew J Colbrook, Qin Li, Ryan V Raut, and Alex Townsend. Beyond expectations: residual dynamic mode decomposition and variance for stochastic dynamical systems. *Nonlinear Dynamics*, 112(3):2037–2061, 2024.

[55] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010.

[56] Clarence W Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S Henningson. Spectral analysis of nonlinear flows. *Journal of fluid mechanics*, 641:115–127, 2009.

[57] J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[58] I Kevrekidis, Clarence W Rowley, and M Williams. A kernel-based method for data-driven koopman spectral analysis. *Journal of Computational Dynamics*, 2(2):247–265, 2016.

[59] Christophe Zhang and Enrique Zuazua. A quantitative analysis of koopman operator methods for system identification and predictions. *Comptes Rendus. Mécanique*, 351(S1):1–31, 2023.

[60] Joel A Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1-2):1–230, 2015.

[61] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12:389–434, 2012.

[62] Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PLOS ONE*, 11(2):1–19, 02 2016.

[63] Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2 edition, 2009.

[64] Liam Llamazares-Elias, Samir Llamazares-Elias, Jonas Latz, and Stefan Klus. Data-driven approximation of koopman operators and generators: Convergence rates and error bounds, 2024.

[65] Jean-René Chazottes and Sébastien Gouëzel. Optimal concentration inequalities for dynamical systems. *Communications in Mathematical Physics*, 316(3):843–889, 2012.

[66] Lai-Sang Young. Statistical properties of dynamical systems with some hyperbolicity. *Annals of Mathematics*, 147(3):585–650, 1998.

[67] Lai-Sang Young. What are SRB measures, and which dynamical systems have them? *Journal of Statistical Physics*, 108(5):733–754, 2002.

[68] Daniel Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability*, 20:1–32, 2015. Corrected version: arXiv:1212.2015v5 (2018).

[69] Warwick Tucker. A rigorous ODE solver and Smale's 14th problem. *Foundations of Computational Mathematics*, 2(1):53–117, 2002.

[70] Vitor Araújo and Ian Melbourne. Exponential decay of correlations for nonuniformly hyperbolic flows with a $c^{1+\alpha}$ stable foliation, including the classical Lorenz attractor. *Annales Henri Poincaré*, 17(11):2975–3004, 2016.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The four main contributions (zero block structure, PR/PPR detection, finite-sample guarantees, unconditional leakage bounds) are clearly stated in the abstract and the Contributions paragraph of the introduction, and are each supported by corresponding theorems and numerical experiments.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5 discusses the dependence on seed observables, the loss of row-magnitude information under row normalization, and the sensitivity of least-squares EDMD fitting to outliers.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.

- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All theorems, lemmas and propositions are numbered and cross-referenced. Full proofs are provided in Appendices E and F, with a proof sketch in Section 3.3.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Appendix D gives the dictionaries, sample sizes, parameters, sampling procedures, and evaluation protocols. Prediction-error statistics use 20 random seeds; the Lorenz spectral comparison uses a single trajectory as described in Appendix D.6. Code is linked at `https://github.com/HyukpyoHong/Koopman-PageRank`.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code used in this work is available at `https://github.com/HyukpyoHong/Koopman-PageRank`. Experimental settings and evaluation procedures are described in Appendix D.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Full experimental details are provided in Appendix D.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: One-step prediction error plots (Figs. 2a,6a, and 3) report mean $\pm$1SD over 20 independent random seeds. The source of variability (random data sampling and random permutation for the Unordered baseline) is clearly stated in Appendix D. The improvement ratio table (Table 1) also reports mean $\pm$1SD over 20 experiments with different random seeds. Our work does not include hypothesis testing.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix D.7 reports the CPU, RAM, software implementation, and estimated total runtime.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research adheres to the NeurIPS Code of Ethics. It involves only synthetic datasets.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This paper presents foundational mathematical methodology for Koopman operator approximation and dictionary selection in dynamical systems. It does not introduce a deployable system, dataset, or application with a direct path to negative societal impact. No foreseeable negative societal consequences are associated with the work.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release pretrained models, scraped datasets, or any assets that carry a risk of misuse. The only outputs are mathematical algorithms and numerical experiment code for well-studied benchmark dynamical systems.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All methods compared against (TSSD, TICA, PCCA+) are cited with their original references. Experimental data are generated synthetically. The code package includes an MIT license and lists its software dependencies in `requirements.txt`.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The paper introduces Algorithm 1 as a new computational asset. The algorithm is fully described in the main text, with sufficient detail for reimplementation. All relevant code is available via the repository: `https://github.com/HyukpyoHong/Koopman-PageRank`. No separate dataset is released, as all data are synthetically generated from the described dynamical systems.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: No human subjects are involved in any part of this research. IRB approval is not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs.

    Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.

A Background

This appendix provides a more comprehensive overview on the backgrounds on related literature summarized in Section 2.

A.1 Notation

For convenience, we collect the key notation used throughout the paper in Table 2.

Table 2: Summary of key notation.

Symbol	Meaning
$\mathcal{D}_N$	Dictionary of size N, i.e., $\mathcal{D}_N = \{\psi_1, \dots, \psi_N\}$
$\mathcal{F}_N$	Subspace spanned by $\mathcal{D}_N$
$K_{N,M}$	EDMD matrix (with M data points, dictionary of size N)
K_N	Population-level Koopman matrix (i.e., $K_N = \lim_{M\to\infty} K_{N,M}$)
$K[i,j]$	(i,j)-entry of matrix K; $K[i_1 : i_2,\ j_1 : j_2]$ denotes a submatrix
$K^{(0)}_{\tilde{N}}$	Matrix $K_{\tilde{N}}$ with bottom-left block zeroed out
$K^{(0)}_{\tilde{N},M}$	Matrix $K_{\tilde{N},M}$ with bottom-left block zeroed out
$\lambda_{\min}$	Smallest eigenvalue (in modulus) of the population Gram matrix $G_{\tilde{N}}$
$r^{(0)}_{\min}$	Minimum absolute row sum of $(K^{(0)}_{\tilde{N}})^\intercal$
$r_{\max}$	Maximum absolute row sum of $K^\intercal_{\tilde{N}}$
ε_M	High-probability bound on the EDMD matrix estimation error
$\mathcal{R}(A)$	Row stochastic matrix obtained by row normalizing matrix A
P	Row-normalized $\lvert K^\intercal_{\tilde{N},M}\rvert$, i.e., $P = \mathcal{R}(\lvert K^\intercal_{\tilde{N},M}\rvert)$, interpreted as a Markov chain
$P^{(0)}$	Row-normalized $\lvert K^{(0)}_{\tilde{N},M}\rvert^\intercal$, i.e., $P^{(0)} = \mathcal{R}(\lvert K^{(0)}_{\tilde{N},M}\rvert^\intercal)$
π	Generic notation for PR/PPR vectors
$\pi_v, \pi_{\mathcal{S}}, \pi_{\mathbf{s}}$	PPR vectors with a single seed v, a seed set $\mathcal{S}$, and a preference vector $\mathbf{s}$, respectively.
$\pi_M, \pi_{\mathbf{s},M}$	PR/PPR vectors associated EDMD matrices from M data samples
$\pi^{(0)}, \pi^{(0)}_{\mathcal{S}}$	PR/PPR vectors of $P^{(0)}$
$\Delta_{\text{PR}}, \Delta_{\text{PPR}}$	PR/PPR detection gaps on P
$\Delta^{(0)}_{\text{PR}}, \Delta^{(0)}_{\text{PPR}}$	Auxiliary PR/PPR detection gaps on $P^{(0)}$
α	Damping factor $\in (0,1)$ for PR and PPR algorithms
η	Coupling parameter $= 1 - \lVert P^{22}\rVert_\infty$.

A.2 The Koopman operator

The Koopman operator framework, introduced by Koopman [1] and extended by Koopman and von Neumann [2], lifts a nonlinear dynamical system on the state space $\mathcal{M}$ to a linear (but infinite-dimensional) operator on a function space of *observables*. This linearization is the foundation of modern data-driven analysis of nonlinear dynamics: dynamic mode decomposition (DMD) [55, 56, 57] and its extensions [14, 34, 58] approximate the Koopman operator from trajectory data, and its spectral theory underpins mode decomposition, long-time forecasting, coherent-structure discovery, and reduced-order modeling across fluid mechanics, molecular dynamics, power systems, and control [6, 19, 51, 54].

We consider a discrete-time dynamical system $\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k)$, $k = 0, 1, 2, \dots$ with $\mathbf{F} : \mathcal{M} \to \mathcal{M}$ where $\mathbf{x} \in \mathcal{M}$ denotes the state of the system and $\mathcal{M} \subset \mathbb{R}^d$ is the state space. Let $\mathcal{F}$ be the Hilbert space $L^2(\mathcal{M}, \mu)$ where μ is the distribution on $\mathcal{M}$ from which data points are sampled. Throughout, we take μ to be a (positive) Borel probability measure on $\mathcal{M}$ and assume $\mathbf{F}$ is μ-*nonsingular*, i.e. $\mathbf{F}_{\#}\mu \ll \mu$; this ensures that the pointwise composition $f \circ \mathbf{F}$ is well-defined on μ-equivalence classes and hence on $\mathcal{F}$. The induced *Koopman operator* $\mathcal{K} : \mathcal{F} \to \mathcal{F}$, $\mathcal{K}f = f \circ \mathbf{F}$, is bounded precisely when the Radon–Nikodym derivative $d(\mathbf{F}_{\#}\mu)/d\mu \in L^\infty(\mu)$, in which case $\lVert\mathcal{K}\rVert^2_{\mathcal{F}\to\mathcal{F}} =$

$\|d(\mathbf{F}_{\#}\mu)/d\mu\|_{L^\infty}$; it is an isometry when μ is $\mathbf{F}$-invariant. We assume this nonsingularity condition throughout. Given linearly independent functions $\psi_i \in \mathcal{F}$, $i = 1, \ldots, N$, define a finite-dimensional subspace $\mathcal{F}_N = \operatorname{span}\{\psi_1, \ldots, \psi_N\} \subset \mathcal{F}$ and denote the Koopman operator restricted on this subspace by $\mathcal{K}_{|\mathcal{F}_N} : \mathcal{F}_N \to \mathcal{F}$.

Linearity of $\mathcal{K}$. The defining feature of the Koopman formalism is that $\mathcal{K}$ is a *linear* operator, irrespective of whether the underlying dynamics $\mathbf{F}$ is linear or nonlinear. Linearity is immediate from the definition: for $f, g \in \mathcal{F}$ and $a, b \in \mathbb{C}$,

$$\mathcal{K}(af + bg)(\mathbf{x}) = (af + bg)\big(\mathbf{F}(\mathbf{x})\big) = a\, f(\mathbf{F}(\mathbf{x})) + b\, g(\mathbf{F}(\mathbf{x})) = a\,(\mathcal{K}f)(\mathbf{x}) + b\,(\mathcal{K}g)(\mathbf{x}) \quad (15)$$

for every $\mathbf{x} \in \mathcal{M}$, so $\mathcal{K}(af + bg) = a\mathcal{K}f + b\mathcal{K}g$. This identity holds pointwise on $\mathcal{M}$ and hence in $\mathcal{F} = L^2(\mathcal{M}, \mu)$.

Linearity comes at the cost of dimension. Linearity on observables of $\mathcal{K}$ is a *global and exact* property, but it is purchased by leaving the finite-dimensional state space $\mathcal{M} \subset \mathbb{R}^d$ in exchange for the typically infinite-dimensional function space $\mathcal{F}$. This is qualitatively different from the Jacobian linearization $D\mathbf{F}(\mathbf{x}^\star)$, which is a *local* linear approximation valid only in a neighborhood of a fixed point: $\mathcal{K}$ represents the full nonlinear dynamics everywhere on $\mathcal{M}$, losslessly. The central modeling question raised by this trade-off, and the one that motivates our paper, is whether a finite-dimensional subspace $\mathcal{F}_N \subset \mathcal{F}$ can be chosen so that $\mathcal{K}|_{\mathcal{F}_N}$ is a faithful finite-matrix restriction of $\mathcal{K}$. For example, a subspace spanned by Koopman eigenfunctions is invariant under the action given by $\mathcal{K}$.

Spectral consequences. Because $\mathcal{K}$ is linear, the full machinery of operator theory applies. It has a spectrum $\sigma(\mathcal{K}) \subset \mathbb{C}$, and eigenpairs (λ_j, φ_j) with $\mathcal{K}\varphi_j = \lambda_j\varphi_j$ are the *Koopman eigenvalues* and *eigenfunctions*. Any observable $f \in \mathcal{F}$ admitting an eigenfunction expansion $f = \sum_j c_j\varphi_j$ evolves diagonally under iteration:

$$(\mathcal{K}^k f)(\mathbf{x}) = f(\mathbf{F}^k(\mathbf{x})) = \sum_j c_j\, \lambda_j^k\, \varphi_j(\mathbf{x}), \quad (16)$$

so nonlinear trajectories decompose into a superposition of geometrically scaling/rotating *Koopman modes* [6, 56]. When $\mathbf{F}$ preserves the measure μ, $\mathcal{K}$ is an isometry on $L^2(\mathcal{M}, \mu)$ with $\sigma(\mathcal{K}) \subset \{z \in \mathbb{C} : |z| \le 1\}$, and the unimodular part of the spectrum encodes the ergodic and quasi-periodic structure of the dynamics. These consequences, which are inaccessible from nonlinear state-space analysis alone, are what make the Koopman framework powerful for mode decomposition, reduced-order modeling, and long-horizon forecasting of nonlinear systems [51, 19, 54].

Finite-dimensional restriction. Linearity also underlies the matrix representation used throughout this paper. If $\mathcal{F}_N = \operatorname{span}\{\psi_1, \ldots, \psi_N\}$ is $\mathcal{K}$-invariant, then for each j there exist unique coefficients $K_{ij} \in \mathbb{C}$ with $\mathcal{K}\psi_j = \sum_{i=1}^N K_{ij}\psi_i$, and the matrix $K = (K_{ij}) \in \mathbb{C}^{N\times N}$ represents $\mathcal{K}|_{\mathcal{F}_N}$ in the basis $\{\psi_i\}$. When $\mathcal{F}_N$ is only *approximately* invariant, the corresponding finite matrix captures only the Galerkin projection of $\mathcal{K}$ onto $\mathcal{F}_N$, and the deviation of $\mathcal{K}\psi_j$ from $\mathcal{F}_N$ is the source of the residual structure analyzed in the sequel.

A.3 Extended dynamic mode decomposition

Convention. Throughout this appendix and all proofs in Appendix E, we use the same row-vector convention as in the main text: $\mathbf{\Psi}(\mathbf{x}) \in \mathbb{C}^{1\times N}$, $\mathbf{\Psi}(\mathbf{X}) \in \mathbb{C}^{M\times N}$, and $K_{N,M} = \mathbf{\Psi}(\mathbf{X})^+\,\mathbf{\Psi}(\mathbf{Y})$ as in (2). To follow the convention of transition matrices in Markov chain theory, we define the row-stochastic matrix $P = \mathcal{R}(|K_{\tilde{N},M}^{\mathsf{T}}|)$ by row-normalizing the *transpose* of $|K_{\tilde{N},M}|$.

Extended dynamic mode decomposition (EDMD), introduced by Williams, Kevrekidis, and Rowley [14] as a generalization of the original DMD algorithm [55, 56], approximates the Koopman operator by Galerkin projection onto the span of a user-chosen dictionary of observables. Convergence to the L^2-projected Koopman operator as $M \to \infty$ was established by Korda and Mezić [15]; kernel and Hilbert-space extensions appear in [58, 34], and recent work has produced rigorous finite-sample rates and spectral guarantees [59, 17, 18, 16]. The quality of the approximation hinges on whether the dictionary spans an (approximately) Koopman invariant subspace, which is the motivation for the sub-dictionary selection problem considered in this paper.

Let $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(M)}\} \subset \mathcal{M}$ be i.i.d. samples from μ and $\{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}\}$ their one-step evolutions, i.e., $\mathbf{y}^{(i)} = \mathbf{F}(\mathbf{x}^{(i)})$. A *dictionary* of linearly independent observables is $\mathcal{D}_N = \{\psi_1, \ldots, \psi_N\} \subset \mathcal{F}$. With $\mathbf{\Psi}(\mathbf{x}) = [\psi_1(\mathbf{x}), \ldots, \psi_N(\mathbf{x})] \in \mathbb{C}^{1\times N}$, define

$$\mathbf{\Psi}(\mathbf{X}) = \begin{bmatrix} \mathbf{\Psi}(\mathbf{x}^{(1)}) \\ \vdots \\ \mathbf{\Psi}(\mathbf{x}^{(M)}) \end{bmatrix} \in \mathbb{C}^{M\times N}, \quad \mathbf{\Psi}(\mathbf{Y}) = \begin{bmatrix} \mathbf{\Psi}(\mathbf{y}^{(1)}) \\ \vdots \\ \mathbf{\Psi}(\mathbf{y}^{(M)}) \end{bmatrix} \in \mathbb{C}^{M\times N}. \tag{17}$$

The *EDMD matrix* with the dictionary $\mathcal{D}_N$ and the data set $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(M)}\}$ is $K_{N,M} = \mathbf{\Psi}(\mathbf{X})^+ \mathbf{\Psi}(\mathbf{Y})$. If $\mathbf{\Psi}(\mathbf{X})$ has full column rank, then $K_{N,M} = \left(\mathbf{\Psi}(\mathbf{X})^* \mathbf{\Psi}(\mathbf{X})\right)^{-1} \mathbf{\Psi}(\mathbf{X})^* \mathbf{\Psi}(\mathbf{Y})$. Here we make the following remark on Assumption 1 that states $\mathbf{\Psi}(\mathbf{X})$ has full column rank.

Remark 1. Assumption 1 is ensured *almost surely* with respect to the distribution μ by the following condition of measure-theoretic form: $\mu(\{\mathbf{x} \in \mathcal{M} : c_1\psi_1(\mathbf{x}) + \ldots + c_N\psi_N(\mathbf{x}) = 0\}) = 0$ for any $(c_1, \ldots, c_N) \in \mathbb{R}^N \setminus \{\mathbf{0}\}$, which follows Assumption 1 in [15]. This holds, for instance, for real-analytic basis functions when μ is not supported on a proper real-analytic subvariety of $\mathcal{M}$; more generally, it suffices that the span of the $\{\psi_i\}$ contains no function that vanishes on a μ-positive set. We state the condition in this measure-theoretic form so that it is meaningful even when $\mathcal{M}$ has no interior (e.g., an attractor or fractal invariant set): what matters is that μ charges enough of $\mathcal{M}$ to distinguish every nonzero linear combination of dictionary functions, not that $\mathcal{M}$ has full topological support.

When $\mathcal{F}_N$ is invariant under $\mathcal{K}$, the Frobenius error vanishes, giving

$$\mathbf{\Psi}(\mathbf{X})\, K_{N,M} = \mathbf{\Psi}(\mathbf{Y}). \tag{18}$$

That is, if a dictionary spans a Koopman invariant subspace, the finite-dimensional EDMD matrix exactly captures the Koopman dynamics on that subspace.

Galerkin / L^2-projection interpretation. In the infinite-data limit $M \to \infty$, the Gram matrices converge (by the strong law of large numbers) to their population counterparts $G_{ij} = \langle \psi_i, \psi_j \rangle_\mu$ and $A_{ij} = \langle \mathcal{K}\psi_j, \psi_i \rangle_\mu$, and $K_{N,M}$ converges to $K_N = G^{-1}A$, the matrix of the *Galerkin projection* $\mathcal{P}_{\mathcal{F}_N}\mathcal{K}|_{\mathcal{F}_N}$ in the basis $\{\psi_i\}$ [15]. Here $\mathcal{P}_{\mathcal{F}_N}$ is the $L^2(\mathcal{M}, \mu)$-orthogonal projection onto $\mathcal{F}_N$. Thus EDMD is not an arbitrary fit: it recovers the best $\mathcal{F}_N$-representation of $\mathcal{K}$ in the L^2 sense. When $\mathcal{F}_N$ is exactly $\mathcal{K}$-invariant, $\mathcal{P}_{\mathcal{F}_N}\mathcal{K}|_{\mathcal{F}_N} = \mathcal{K}|_{\mathcal{F}_N}$ and EDMD is exact, so the residual $\mathbf{\Psi}(\mathbf{Y}) - \mathbf{\Psi}(\mathbf{X})\, K_{N,M}$ captures the failure of invariance.

Relation to DMD. The original DMD algorithm [55, 56, 57] corresponds to EDMD with the choice of identity dictionary $\mathbf{\Psi}(\mathbf{x}) = \mathbf{x}$: the EDMD matrix reduces to the best linear fit $Y \approx AX$ between snapshot matrices, exact only when $\mathbf{F}$ is itself linear. EDMD generalizes this by letting $\mathbf{\Psi}$ contain arbitrary nonlinear observables, such as monomials, radial basis functions, Hermite polynomials, or data-driven features, so that the method can capture nonlinear dynamics that admit a finite-dimensional linear representation on the lifted space. Kernel DMD [58] sidesteps the explicit choice of dictionary by operating in a reproducing kernel Hilbert space; Klus et al. unifies EDMD, kernel EDMD, and variational approaches to Koopman and Perron–Frobenius approximation [34].

Finite-sample and spectral guarantees. Quantitative finite-sample error bounds for EDMD have been a recent focus. Zhang and Zuazua [59], Nüske et al. [17], and Colbrook et al. [18] give $1/\sqrt{M}$-type rates with explicit constants depending on the dictionary Gram condition number and sub-Gaussian tail parameters of $\mathbf{\Psi}(\mathbf{x})$. Kostic et al. [16] obtain sharper operator-norm bounds in RKHS settings via concentration of the empirical covariance; Colbrook et al. address the difficult problem of *spurious spectra*—eigenvalues of $K_{N,M}$ that do not correspond to any spectral object of $\mathcal{K}$—through pseudospectral and residual-DMD techniques [18, 54]. The concentration-of-measure tradition built on matrix Bernstein inequalities forms the technical backbone of these results [60, 61].

Dictionary design and learning. The practical performance of EDMD is dominated by dictionary choice. Hand-crafted dictionaries include monomials up to a certain degree, radial basis functions [14], and Hermite polynomials; data-driven approaches learn the dictionary jointly with $K_{N,M}$ via neural networks [24, 27, 23, 25] or regression-with-sparsity schemes [11, 62]. When samples come from a

single trajectory rather than i.i.d. draws from μ, ergodic variants of EDMD are required [21, 20]. In all these settings, the underlying question is the same: does the chosen dictionary span (or nearly span) a $\mathcal{K}$-invariant subspace [30, 31]? Answering this question a posteriori, from the structure of $K_{N,M}$ itself, is the problem addressed by the PPR-based detection method developed in this paper.

A.4 Personalized PageRank

PageRank (PR) was introduced by Page and Brin [35] to rank web pages by modeling a random surfer on the hyperlink graph who occasionally teleports to a uniformly random page. The algorithm's mathematical properties, such as convergence, sensitivity to the damping parameter, and robustness of the stationary distribution, were further developed in [36, 37]. Personalized PageRank (PPR), due to Haveliwala [41], replaces the uniform teleportation distribution by a user-specified *preference vector*, localizing the ranking around a chosen seed set. Andersen, Chung, and Lang [42] established Cheeger-type guarantees for local clustering via sweep cuts on PPR vectors, and PPR has since become a canonical tool in graph learning, community detection, and semi-supervised learning on networks. In this paper we exploit PPR in a novel setting: we interpret the EDMD matrix itself as a weighted graph on observables and use PPR mass from a seed set of "target" observables to rank sub-dictionaries by the strength of their Koopman coupling.

Let $Q \in \mathbb{R}^{n\times n}$ be a row stochastic matrix, $\mathbf{s} \in \mathbb{R}^n$ a *preference vector* satisfying $\mathbf{s}^\intercal \mathbf{1} = 1$ and $s_i \geq 0$, and $\alpha \in [0, 1)$ a *damping factor*. Define the row stochastic *transition kernel*

$$G_{\mathbf{s}} := \alpha Q + (1-\alpha)\,\mathbf{1}\mathbf{s}^\intercal. \tag{19}$$

This describes a random walker who, at each step, follows Q with probability α and teleports to a node drawn from $\mathbf{s}$ with probability $1-\alpha$. Since every entry satisfies $G_{\mathbf{s}}[i,j] \geq (1-\alpha)s_j$, the chain is primitive whenever $\mathbf{s}$ has full support, guaranteeing a unique stationary distribution by the Perron–Frobenius theorem.

The *PPR vector* $\pi_{\mathbf{s}}$ is the stationary distribution of $G_{\mathbf{s}}$. From $\pi_{\mathbf{s}}^\intercal = \pi_{\mathbf{s}}^\intercal G_{\mathbf{s}}$ and $\pi_{\mathbf{s}}^\intercal \mathbf{1} = 1$:

$$\pi_{\mathbf{s}}^\intercal = \alpha\,\pi_{\mathbf{s}}^\intercal Q + (1-\alpha)\,\mathbf{s}^\intercal. \tag{20}$$

Rearranging, $\pi_{\mathbf{s}}^\intercal(I-\alpha Q) = (1-\alpha)\,\mathbf{s}^\intercal$. Since Q is row stochastic, all eigenvalues of Q have modulus at most 1, and $\alpha < 1$ ensures that $I-\alpha Q$ is nonsingular. Therefore $\pi_{\mathbf{s}}^\intercal = (1-\alpha)\,\mathbf{s}^\intercal(I-\alpha Q)^{-1}$. Expanding the resolvent as a Neumann series $(I-\alpha Q)^{-1} = \sum_{k=0}^{\infty}\alpha^k Q^k$ gives $\pi_{\mathbf{s}}^\intercal = (1-\alpha)\sum_{k=0}^{\infty}\alpha^k\,\mathbf{s}^\intercal Q^k$, which converges absolutely since $\alpha < 1$ and $\|Q^k\|_\infty = 1$. The entry $\pi_{\mathbf{s}}(j)$ equals the expected discounted fraction of time at node j; the factor α^k geometrically discounts k-step transitions. A key structural property is *linearity in the preference vector*: since the resolvent $(I-\alpha Q)^{-1}$ is independent of $\mathbf{s}$, the map $\mathbf{s} \mapsto \pi_{\mathbf{s}}$ is linear. All special cases are instances of this linearity.

Multi-seed PPR. For a seed set $\mathcal{S} \subset \{1,\dots,n\}$, setting $\mathbf{s} = \frac{1}{|\mathcal{S}|}\sum_{v\in\mathcal{S}}\mathbf{e}_v$ gives the multi-seed PPR vector $\pi_{\mathcal{S}} = \frac{1}{|\mathcal{S}|}\sum_{v\in\mathcal{S}}\pi_v$ by linearity. The corresponding transition kernel is $G_{\mathbf{s}} = \alpha Q + \frac{1-\alpha}{|\mathcal{S}|}\sum_{v\in\mathcal{S}}\mathbf{1}\mathbf{e}_v^\intercal$: the walker teleports uniformly to a node in $\mathcal{S}$.

Single-seed PPR. The special case $\mathcal{S} = \{v\}$ (equivalently, $\mathbf{s} = \mathbf{e}_v$) gives

$$\pi_v(j) = (1-\alpha)\sum_{k=0}^{\infty}\alpha^k\,Q^k[v,j], \tag{21}$$

the expected discounted occupation of node j for a walker starting at v and teleporting back to v at rate $1-\alpha$. PPR was introduced by Haveliwala [41] for topic-sensitive web search; Andersen, Chung, and Lang [42] proved that sweep cuts on single-seed PPR vectors achieve Cheeger-type guarantees for detecting low-conductance clusters.

Standard PageRank. Setting $\mathcal{S} = \{1,\dots,n\}$ recovers the standard PR vector. The transition kernel is the *Google matrix* $G = \alpha Q + \frac{1-\alpha}{n}\mathbf{1}\mathbf{1}^\intercal$ [35]. Since every entry of G is at least $(1-\alpha)/n > 0$, G is primitive and π is the unique positive left eigenvector with $\pi^\intercal\mathbf{1} = 1$.

For the limiting case $\alpha = 1$, the kernel reduces to $G_{\mathbf{s}} = Q$, and every stationary distribution concentrates on the closed communicating classes of Q, assigning zero probability to transient states.

B Toy example for illustrating Theorem 2

Here we illustrate zero block structure in an EDMD matrix guaranteed by Theorem 2 using the following toy model with $\tilde{N} = 9$ observables in the initial dictionary.

Example 6. Consider the following discrete-time dynamical system with two variables x_1 and x_2:

$$\begin{aligned} x_1(k+1) &= x_1(k) - \Delta t \cdot a x_1(k), \\ x_2(k+1) &= x_2(k) - \Delta t \cdot b(x_2(k) - x_1(k)^2), \end{aligned} \tag{22}$$

where $a > 0$ and $b > 0$, which is a discrete-time approximation of the continuous-time system:

$$\begin{aligned} \dot{x}_1 &= -a x_1, \\ \dot{x}_2 &= -b(x_2 - x_1^2). \end{aligned}$$

This system (22) admits the invariant subspace spanned by $\mathcal{D}_3 \coloneqq \{x_1, x_2, x_1^2\}$ because their images under the action by the Koopman operator are contained in the same subspace. In other words, their one-step evolutions at time $k+1$ can be expressed as linear combinations of their values at time k as follows:

$$\begin{bmatrix} (\mathcal{K}\psi_1)(\mathbf{x}(k)) \\ (\mathcal{K}\psi_2)(\mathbf{x}(k)) \\ (\mathcal{K}\psi_3)(\mathbf{x}(k)) \end{bmatrix}^\intercal = \begin{bmatrix} \psi_1(\mathbf{x}(k+1)) \\ \psi_2(\mathbf{x}(k+1)) \\ \psi_3(\mathbf{x}(k+1)) \end{bmatrix}^\intercal = \begin{bmatrix} \psi_1(\mathbf{x}(k)) \\ \psi_2(\mathbf{x}(k)) \\ \psi_3(\mathbf{x}(k)) \end{bmatrix}^\intercal \begin{bmatrix} 1 - \Delta t \cdot a & 0 & 0 \\ 0 & 1 - \Delta t \cdot b & 0 \\ 0 & \Delta t \cdot b & (1 - \Delta t \cdot a)^2 \end{bmatrix}, \tag{23}$$

where $\psi_1(\mathbf{x}) = x_1, \psi_2(\mathbf{x}) = x_2$, and $\psi_3(\mathbf{x}) = x_1^2$. This 3×3 matrix form of the Koopman operator is identical to the EDMD matrix obtained with the dictionary $\mathcal{D}_3$ as EDMD completely captures the behavior of the Koopman operator on the invariant subspace [31]. Indeed, the EDMD method with $M = 100$ i.i.d. samples from the uniform distribution on the square region $[-2, 2]^2$ provides the following matrix:

$$\begin{bmatrix} 0.92 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0.2 & 0.846 \end{bmatrix} \tag{24}$$

which coincides with the matrix form in (23) where the parameter values are $\Delta t = 0.2$, $a = 0.4$, and $b = 1.0$.

According to Theorem 2, EDMD still captures the invariant subspace with a larger dictionary. Here, we perform EDMD with an extended dictionary of polynomials of degree less than 4, $\mathcal{D}_9 \coloneqq \{x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3\}$. The EDMD matrix is given by

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \\ x_1^3 \\ x_1^2x_2 \\ x_1x_2^2 \\ x_2^3 \end{array} \begin{array}{c} \begin{array}{ccccccccc} x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 & x_1^3 & x_1^2x_2 & x_1x_2^2 & x_2^3 \end{array} \\ \begin{bmatrix} 0.920 & 0.000 & 0.000 & 0.000 & 0.002 & 0.000 & 0.009 & -0.104 & -0.170 \\ 0.000 & 0.800 & 0.000 & 0.000 & 0.003 & 0.000 & 0.013 & -0.069 & -0.113 \\ 0.000 & 0.200 & 0.846 & 0.000 & 0.118 & 0.000 & 0.499 & 0.052 & 0.417 \\ 0.000 & 0.000 & 0.000 & 0.736 & 0.003 & 0.000 & 0.012 & 0.722 & 0.018 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.627 & 0.000 & -0.057 & -0.050 & 0.379 \\ 0.000 & 0.000 & 0.000 & 0.184 & -0.003 & 0.779 & -0.015 & 0.147 & 0.089 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.324 & 0.000 & 0.696 & 0.018 & 0.286 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.002 & 0.000 & 0.008 & 0.616 & -0.043 \\ 0.000 & 0.000 & 0.000 & 0.000 & -0.003 & 0.000 & -0.014 & -0.007 & 0.559 \end{bmatrix} \end{array}. \tag{25}$$

It is clear that the top-left corner 3×3 submatrix is identical to the EDMD matrix in (24) obtained with the dictionary of the three observables, $\mathcal{D}_3 = \{x_1, x_2, x_1^2\}$, which spans the Koopman invariant subspace. Also, the bottom-left corner 6×3 matrix is a zero matrix, $O_{6\times 3}$, as stated in Theorem 2.

There are more zero entries in the matrix (25) besides the zero block at the bottom-left corner. They are not just numerical artifacts; they are consequences of other invariant subspaces. We can identify other subsets that span other Koopman invariant subspaces by permuting the rows and columns of the EDMD matrix. For instance, an ordered dictionary $\{x_1, x_1x_2, x_1^3, x_2, x_1^2, x_2^2, x_1^2x_2, x_1x_2^2, x_2^3\}$ (yet

with the same elements as $\mathcal{D}_9$) provides the following EDMD matrix:

$$\begin{array}{c} \\ x_1 \\ x_1x_2 \\ x_1^3 \\ x_2 \\ x_1^2 \\ x_2^2 \\ x_1^2x_2 \\ x_1x_2^2 \\ x_2^3 \end{array} \begin{array}{c} \begin{array}{ccccccccc} x_1 & x_1x_2 & x_1^3 & x_2 & x_1^2 & x_2^2 & x_1^2x_2 & x_1x_2^2 & x_2^3 \end{array} \\ \begin{bmatrix} 0.920 & 0.000 & 0.000 & 0.000 & 0.000 & 0.002 & 0.009 & -0.104 & -0.170 \\ 0.000 & 0.736 & 0.000 & 0.000 & 0.000 & 0.003 & 0.012 & 0.722 & 0.018 \\ 0.000 & 0.184 & 0.779 & 0.000 & 0.000 & -0.003 & -0.015 & 0.147 & 0.089 \\ 0.000 & 0.000 & 0.000 & 0.800 & 0.000 & 0.003 & 0.013 & -0.069 & -0.113 \\ 0.000 & 0.000 & 0.000 & 0.200 & 0.846 & 0.118 & 0.499 & 0.052 & 0.417 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.627 & -0.057 & -0.050 & 0.379 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.324 & 0.696 & 0.018 & 0.286 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.002 & 0.008 & 0.616 & -0.043 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.003 & -0.014 & -0.007 & 0.559 \end{bmatrix} \end{array}. \tag{26}$$

Now zero block structures are more noticeable. The 6×3 submatrix at the bottom-left corner is a zero matrix. Also, the 4×5 submatrix at the bottom-left corner is a zero matrix. These observations respectively indicate the facts that two dictionaries $\{x_1, x_1x_2, x_1^3\}$ and $\{x_1, x_1x_2, x_1^3, x_2, x_1^2\}$ span Koopman invariant subspaces. This allows the identification of *all* Koopman invariant subspaces spanned by subsets of the dictionary $\mathcal{D}_9$.

C Three-state example: exact detection windows

To concretely illustrate the mixing-versus-reachability dichotomy of Theorem 3, we carry out the full computation of the auxiliary gaps $\Delta_{\text{PR}}^{(0)}$, $\Delta_{\text{PPR}}^{(0)}$ and the detection windows $\{\alpha \in (0,1) :$ Theorem 3(i) or (ii) holds$\}$ on two minimal examples with $\tilde{N} = 3$, $N = 2$. In each example, we specify a row-stochastic matrix $P = \mathcal{R}(|K^\intercal|)$ as if we start from $K = P^\intercal$, so $r_{\max} = 1$. The matrix $P^{(0)} = \mathcal{R}(|K^{(0)}|^\intercal)$ is the row-normalization of P after zeroing P^{12} (equivalently, the rows of the top block are renormalized to make $P^{(0)}$ row-stochastic), with $P^{11,(0)} = \mathcal{R}(P^{11})$, $P^{22,(0)} = P^{22}$, $P^{21,(0)} = P^{21}$. Both share the bottom-block row $P^{22,(0)} = 1 - 2\beta$ and $P^{21,(0)} = (\beta, \beta)$ with $\beta \in (0, 1/2]$.

The leakage parameter is $\epsilon := \|P^{12}\|_\infty \in (0,1)$. In both, $r_{\min}^{(0)} = 1 - \epsilon$. Because the leakage is removed before renormalization, $P^{11,(0)} = \mathcal{R}(P^{11})$ is row-stochastic and the auxiliary gaps $\Delta_{\text{PR}}^{(0)}, \Delta_{\text{PPR}}^{(0)}$ become ϵ-independent: ϵ enters detection only through $\|P^{12}\|_\infty = \epsilon$. The two examples differ only in the internal block $P^{11,(0)}$: Example A has a well-mixed internal block, while Example B has a "starved" node with zero column in $P^{11,(0)}$.

Throughout this section we write $\Delta_{\text{PR}}^{(0)}, \Delta_{\text{PPR}}^{(0)}$ for the auxiliary gaps defined in (10).

C.1 Example A: well-mixed internal block

Setup. Take

$$P_A \;=\; \begin{pmatrix} 0 & 1-\epsilon & \epsilon \\ 1-\epsilon & 0 & \epsilon \\ \beta & \beta & 1-2\beta \end{pmatrix}. \tag{27}$$

Each row of P_A sums to 1, so $\|P^{12}\|_\infty = \epsilon$. The top-left block $P_A^{11} = (1-\epsilon)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ has each row summing to $1 - \epsilon$, so after zeroing P^{12} and renormalizing the rows of the top block we obtain the row-stochastic

$$P_A^{(0)} \;=\; \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ \beta & \beta & 1-2\beta \end{pmatrix}, \qquad P_A^{11,(0)} \;=\; Q := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{28}$$

The internal block $P_A^{11,(0)}$ is the row-stochastic swap matrix, irreducible with stationary distribution $\mathbf{p} = (1/2, 1/2)$.

Resolvent. A geometric series gives

$$R_{11} \;\coloneqq\; (I-\alpha Q)^{-1} \;=\; \frac{1}{1-\alpha^2}\begin{pmatrix}1 & \alpha\\ \alpha & 1\end{pmatrix}. \tag{29}$$

Standard PR. With $(\mathbf{q}^\alpha)^\intercal = \frac{1-\alpha}{2}\mathbf{1}_2^\intercal R_{11}$,

$$q_1^\alpha \;=\; q_2^\alpha \;=\; \frac{(1-\alpha)(1+\alpha)}{2(1-\alpha^2)} \;=\; \frac{1}{2}.$$

So $q_{\min}^\alpha = 1/2$, independent of both α and ϵ. With $\eta = 2\beta$, the mixing condition $q_{\min}^\alpha > (\tilde{N}-N)(1-\alpha)/(N(1-\alpha+\alpha\eta))$ reduces to

$$\tfrac{1}{2} > \tfrac{1-\alpha}{2(1-\alpha+2\alpha\beta)} \iff \alpha\beta > 0, \tag{30}$$

which holds trivially for any $\alpha, \beta > 0$. Renormalization eliminates the original sub-stochastic-form mixing constraint $\epsilon < 2\beta$.

Exact PR gap. Since $P_A^{(0)}$ is block lower triangular with the right-top block zero, the resolvent $R \coloneqq (I-\alpha P_A^{(0)})^{-1}$ is block lower triangular with diagonal blocks R_{11} from (29) and $R_{22} = (1-\alpha(1-2\beta))^{-1}$, and lower-left block $R_{21} = \alpha R_{22} P^{21,(0)} R_{11}$. Writing $\pi^{(0),\intercal} = \frac{1-\alpha}{3}\mathbf{1}^\intercal R$ and computing column sums,

$$\pi_1^{(0)} \;=\; \pi_2^{(0)} \;=\; \frac{1-\alpha+3\alpha\beta}{3(1-\alpha+2\alpha\beta)}, \qquad \pi_3^{(0)} \;=\; \frac{1-\alpha}{3(1-\alpha+2\alpha\beta)},$$

from which

$$\Delta_{\mathrm{PR}}^{(0)}(\alpha) \;=\; \pi_2^{(0)} - \pi_3^{(0)} \;=\; \frac{\alpha\beta}{1-\alpha+2\alpha\beta}. \tag{31}$$

$\Delta_{\mathrm{PR}}^{(0)}(\alpha) > 0$ for all $\alpha \in (0,1)$ and $\beta > 0$, monotone increasing in α from 0 to $\beta/(1+2\beta)$.

PPR with seed $\mathcal{S} = \{1\}$. Reachability holds ($\{1\} \to \{1,2\}$ via the 2-cycle). From (29), $e_1^\intercal R_{11} = (1,\alpha)/(1-\alpha^2)$, so

$$\bar{\pi}_1^{(0)} \;=\; \frac{1}{1+\alpha}, \qquad \bar{\pi}_2^{(0)} \;=\; \frac{\alpha}{1+\alpha}, \qquad \bar{\pi}_3^{(0)} \;=\; 0.$$

The minimum of the PPR masses corresponding to the top block is $\bar{\pi}_2^{(0)}$, giving

$$\Delta_{\mathrm{PPR}}^{(0)}(\alpha) \;=\; \frac{\alpha}{1+\alpha}. \tag{32}$$

Both $\Delta_{\mathrm{PR}}^{(0)}$ and $\Delta_{\mathrm{PPR}}^{(0)}$ are ϵ-independent under renormalization: ϵ enters detection only through $\|P^{12}\|_\infty = \epsilon$, not through the auxiliary gaps on $P^{(0)}$.

Detection windows. The sufficient condition from Theorem 3—detection on P succeeds whenever $\|P^{12}\|_\infty < \frac{1-\alpha}{4\alpha}\Delta_{\mathbf{s}}^{(0)}$—specializes here with $\|P^{12}\|_\infty = \epsilon$ to $\epsilon < \frac{1-\alpha}{4\alpha}\Delta_{\mathbf{s}}^{(0)}(\alpha)$. With $\beta = 1/2$ for concreteness, $\Delta_{\mathrm{PR}}^{(0)}(\alpha) = \alpha/2$ and $\Delta_{\mathrm{PPR}}^{(0)}(\alpha) = \alpha/(1+\alpha)$. The PR condition becomes $\epsilon < (1-\alpha)/8$, the PPR condition $\epsilon < (1-\alpha)/(4(1+\alpha))$. Solving for α gives detection windows $(0, \alpha^\star)$ with

$$\alpha_{\mathrm{PR}}^\star(\epsilon) \;=\; 1-8\epsilon, \tag{33}$$

$$\alpha_{\mathrm{PPR}}^\star(\epsilon) \;=\; \frac{1-4\epsilon}{1+4\epsilon}. \tag{34}$$

Window closure ($\alpha^\star = 0$):

$$\epsilon_{\mathrm{PR}}^{\max} \;=\; 1/8 \;=\; 0.125, \qquad \epsilon_{\mathrm{PPR}}^{\max} \;=\; 1/4 \;=\; 0.250. \tag{35}$$

The two curves (33) and (34) are plotted in Panel (a) of Figure 1: the blue curve is $\alpha_{\mathrm{PR}}^\star(\epsilon)$ and the red curve is $\alpha_{\mathrm{PPR}}^\star(\epsilon)$.

C.2 Example B: starved non-mixing internal block

Setup. Take

$$P_B \;=\; \begin{pmatrix} 1-\epsilon & 0 & \epsilon \\ 1 & 0 & 0 \\ \beta & \beta & 1-2\beta \end{pmatrix}. \tag{36}$$

The top-left block $P_B^{11} = \begin{pmatrix} 1-\epsilon & 0 \\ 1 & 0 \end{pmatrix}$ has row 1 summing to $1-\epsilon$ and row 2 summing to 1. After zeroing P^{12} and renormalizing the rows of the top block, we obtain the row-stochastic

$$P_B^{(0)} \;=\; \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \beta & \beta & 1-2\beta \end{pmatrix}, \qquad P_B^{11,(0)} \;=\; \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}. \tag{37}$$

Column 2 of $P_B^{11,(0)}$ is zero: node 2 receives no inflow within top-left block—a *starved node* in the sense of the discussion after Theorem 3. Row 1 is a self-loop, and row 2 jumps deterministically to node 1. The starved-node property is preserved under renormalization.

Resolvent. Direct inversion gives

$$R_{11} \;\coloneqq\; (I - \alpha P_B^{11,(0)})^{-1} \;=\; \begin{pmatrix} 1/(1-\alpha) & 0 \\ \alpha/(1-\alpha) & 1 \end{pmatrix}. \tag{38}$$

Standard PR. Column sums of (38) yield

$$q_1^\alpha \;=\; \frac{1+\alpha}{2}, \qquad q_2^\alpha \;=\; \frac{1-\alpha}{2}.$$

So $q_{\min}^\alpha = q_2^\alpha = (1-\alpha)/2$, the pure-teleportation value, again *independent of* ϵ. The mixing condition becomes $1-\alpha+2\alpha\beta > 1 \Leftrightarrow \beta > 1/2$, α- and ϵ-independent: no choice of α or reduction of leakage can rescue mixing once $P_B^{11,(0)}$ has a starved node.

Exact PR gap. Using (38) and the block lower triangular structure,

$$\pi_1^{(0)} = \frac{(1+\alpha)(1-\alpha+3\alpha\beta)}{3(1-\alpha+2\alpha\beta)}, \quad \pi_2^{(0)} = \frac{(1-\alpha)(1-\alpha+3\alpha\beta)}{3(1-\alpha+2\alpha\beta)}, \quad \pi_3^{(0)} = \frac{1-\alpha}{3(1-\alpha+2\alpha\beta)}.$$

Since $\pi_1^{(0)}/\pi_2^{(0)} = (1+\alpha)/(1-\alpha) \geq 1$, the minimum of the first two PR masses is at the starved node 2, and the gap is

$$\Delta_{\text{PR}}^{(0)}(\alpha) \;=\; \pi_2^{(0)} - \pi_3^{(0)} \;=\; \frac{\alpha(1-\alpha)(3\beta-1)}{3(1-\alpha+2\alpha\beta)}. \tag{39}$$

Positive only if $\beta > 1/3$. Interior max in α, vanishing at both endpoints with the $(1-\alpha)$ prefactor.

PPR with seed $\mathcal{S} = \{2\}$. Seed $\{1\}$ fails reachability (row 1 of $P_B^{11,(0)}$ is a self-loop), but seed $\{2\}$ reaches $\{1,2\}$. From (38), $e_2^\intercal R_{11} = (\alpha/(1-\alpha), 1)$, giving

$$\bar{\pi}_1^{(0)} \;=\; \alpha, \qquad \bar{\pi}_2^{(0)} \;=\; 1-\alpha, \qquad \bar{\pi}_3^{(0)} \;=\; 0.$$

The minimum of $\bar{\pi}_1^{(0)}$ and $\bar{\pi}_2^{(0)}$ crosses over at $\alpha_c \coloneqq 1/2$: $\bar{\pi}_1^{(0)} \lesseqgtr \bar{\pi}_2^{(0)} \Leftrightarrow \alpha \lesseqgtr 1/2$. So

$$\Delta_{\text{PPR}}^{(0)}(\alpha) \;=\; \begin{cases} \alpha & \alpha \leq 1/2, \\ 1-\alpha & \alpha \geq 1/2, \end{cases} \tag{40}$$

continuous at α_c with value $1/2$ and ϵ-*independent*. Note that $\Delta_{\text{PPR}}^{(0)} > 0$ on all of $(0,1)$ — reachability delivers a positive gap even though PR's mixing condition is uniformly violated.

Detection windows. For $\beta = 1/2$, (39) simplifies to $\Delta_{\text{PR}}^{(0)}(\alpha) = \alpha(1-\alpha)/6$, so the PR detection condition $\epsilon < (1-\alpha)\Delta_{\text{PR}}^{(0)}/(4\alpha) = (1-\alpha)^2/24$ inverts to

$$\alpha_{\text{PR}}^{\star}(\epsilon) \;=\; \max\bigl(0,\; 1-\sqrt{24\epsilon}\bigr), \qquad \epsilon_{\text{PR}}^{\max} = 1/24 \approx 0.0417. \tag{41}$$

For PPR, the detection condition $\epsilon < (1-\alpha)\Delta_{\text{PPR}}^{(0)}(\alpha)/(4\alpha)$ on each branch is solved separately. On the lower branch ($\alpha \le 1/2$, $\Delta_{\text{PPR}}^{(0)} = \alpha$), it becomes $\epsilon < (1-\alpha)/4$, i.e. $\alpha < 1-4\epsilon$. On the upper branch ($\alpha \ge 1/2$, $\Delta_{\text{PPR}}^{(0)} = 1-\alpha$), it becomes $\epsilon < (1-\alpha)^2/(4\alpha)$, which solves to $\alpha < (1+2\epsilon) - 2\sqrt{\epsilon(1+\epsilon)}$. The two branches meet at $\alpha = 1/2$, corresponding to $\epsilon = 1/8$, so

$$\alpha_{\text{PPR}}^{\star}(\epsilon) \;=\; \begin{cases} (1+2\epsilon) - 2\sqrt{\epsilon(1+\epsilon)} & \epsilon \le 1/8, \\ 1-4\epsilon & \epsilon \ge 1/8, \end{cases} \tag{42}$$

with window closure $\epsilon_{\text{PPR}}^{\max} = 1/4$, attained as $\alpha \to 0$ on the lower branch. The two curves are plotted in Panel (b) of Figure 1, with the branch change at $\epsilon = 1/8$ visible as a slight kink.

Takeaway. PR tolerates roughly $2\times$ less leakage than PPR in the well-mixed Example A ($1/8$ vs $1/4$) and $6\times$ less in the starved Example B ($1/24$ vs $1/4$). In Example B, PR's detection window collapses by $\alpha \to 0$ at any $\epsilon > 1/24$, while PPR stays open up to $\epsilon = 1/4$; this is the mixing-versus-reachability separation in its starkest form.

C.3 Remarks on the two examples

(a) *Role of α.* In both examples and both methods, the detection window is an interval $(0, \alpha^{\star})$ opening at $\alpha = 0$; no interior-α tuning is required. The $1/\alpha$ factor in the perturbation amplifier $2\alpha/((1-\alpha)r_{\min}^{(0)})$ dominates any interior maximum of the baseline gap, so small α always yields the widest tolerance to leakage. This is consistent with the remark in Section 3.2 that PPR "does not require α near 1."

(b) *What the starved example rules out.* Example B shows that PR's mixing condition is not merely a sufficient abstraction but a genuine obstruction: once $P^{11,(0)}$ has a starved node, $q_{\min}^{\alpha}$ collapses to $(1-\alpha)/N$ for every α, and the mixing condition becomes a condition on β alone, independent of both α and ϵ. In contrast, PPR's reachability condition is structural and passes trivially when any seed v has $P_{vi}^{11,(0)} \neq 0$ cumulatively reaching all $i \in \{1, \ldots, N\}$.

(c) *Size of the PPR advantage.* In the well-mixed example, PPR's maximum tolerable leakage is $\epsilon_{\text{PPR}}^{\max}/\epsilon_{\text{PR}}^{\max} = (1/4)/(1/8) = 2\times$ larger than PR's. In the starved example, the ratio is $(1/4)/(1/24) = 6\times$. The reachability-over-mixing advantage compounds the α-decoupling advantage from zero leakage when internal dynamics are non-mixing.

D Experimental details

This appendix provides full experimental setup, dictionary construction, and additional figures for the experiments in Section 4. All relevant codes can be found at `https://github.com/HyukpyoHong/Koopman-PageRank`. For the Duffing and Van der Pol oscillators, we used bivariate Laguerre polynomials, which are not bounded on $\mathbb{R}^2$, while our finite-sample error bound (Proposition 17) assumes boundedness of observables (Assumption 3). Nevertheless, the data distribution μ is compactly supported on $[-2, 2]^2$, so any fixed polynomial is bounded on $\text{supp}(\mu)$; the finite-sample bounds of Theorem 5 apply with $D = \sup_{\mathbf{x} \in [-2,2]^2} |\psi_i(\mathbf{x})|$. Also, we always include the first two Laguerre polynomials, $1-x$ and $1-y$, in sub-dictionaries for the two oscillator examples in order to reconstruct the state variables, x and y, and measure state prediction errors. For the three-well Ramachandran potential and Lorenz system, we chose the set of state coordinates as the set of seed nodes for PPR.

D.1 Quantifying one-step/multi-step state prediction errors

For the two oscillator examples, the one-step state prediction error (Figs. 2a and 6a) is computed as follows. The state variables x and y are recovered from the first two Laguerre basis functions

$\psi_1(\mathbf{x}) = 1 - x$ and $\psi_2(\mathbf{x}) = 1 - y$, which are always included in every sub-dictionary. Given a sub-dictionary $\mathcal{D}_N$ of size N, we fit the EDMD matrix $K_{N,M}$ and measure the one-step prediction error by

$$\mathcal{E}(N) = \left(\sum_{j=1}^{2} \|\mathcal{K}\psi_j - \mathcal{K}_{N,M}\psi_j\|^2_{L^2(\mu_{\text{test}})} \right)^{1/2}, \tag{43}$$

where $\mathcal{K}\psi_j$ is approximated by evaluating ψ_j at the true next state $\mathbf{F}(\mathbf{x})$ on the test data, and $\mathcal{K}_{N,M}\psi_j$ is the prediction obtained by applying $\mathcal{K}_{N,M}$ to the observable vector at the current state. Note that $\mathcal{K}_{N,M}$ is the operator naturally associated with the EDMD matrix $K_{N,M}$, i.e., $\mathcal{K}_{N,M}\psi_i \coloneqq \sum_{j=1}^{N} K_{N,M}[j,i]\psi_j$. Here, $\|\cdot\|_{L^2(\mu_{\text{test}})}$ denotes the empirical L^2 norm over $M_{\text{test}} = 2000$ i.i.d. test points drawn from the same distribution as the training data. Since the state coordinates are affine functions of ψ_1 and ψ_2, this error directly measures how accurately the sub-dictionary EDMD matrix propagates the physical state one step forward.

For the three-well Ramachandran potential, the system evolves on the torus $[-\pi, \pi]^2$, so the state variables ϕ and ψ are not globally recoverable from a single linear observable. Instead, we embed the state via the four circular coordinates Coord $\coloneqq \{\sin\phi, \cos\phi, \sin\psi, \cos\psi\}$, which are always included in every sub-dictionary and serve as the seed set for PPR. The one-step state prediction error is then

$$\mathcal{E}(N) = \left(\sum_{f \in \text{Coord}} \|\mathcal{K}f - \mathcal{K}_{N,M}f\|^2_{L^2(\mu_{\text{test}})} \right)^{1/2}. \tag{44}$$

Since $(\sin\phi, \cos\phi, \sin\psi, \cos\psi)$ uniquely determines the state on the torus, this error faithfully measures one-step prediction accuracy for the physical state (Fig. 3a). The same error quantification extends naturally to multi-step prediction (Fig. 3b): starting from a test point, we apply $\mathcal{K}$ and $\mathcal{K}_{N,M}$ repeatedly to propagate the full observable vector forward, reading off the four circular coordinates at each step and comparing against the true trajectory.

D.2 Compared methods

For the two oscillator examples, we compare four dictionary ordering methods alongside the tunable symmetric subspace decomposition (TSSD) baseline. **Unordered** uses a uniformly random permutation of all $\tilde{N}$ observables as a baseline, so that the first N selected observables carry no structural information, except that the state variables are always included in the sub-dictionary. **Incremental** selects the first N observables in lexicographic order of total degree, i.e., the natural ordering of the Laguerre polynomial basis from lowest to highest degree. **PageRank (PR)** and **Personalized PageRank (PPR)** apply Algorithm 1 to the full EDMD matrix $K_{\tilde{N},M}$: the $\tilde{N}$ observables are ranked by their PR or PPR scores, and the top-N are selected to form the sub-dictionary $\mathcal{D}_N$. For PPR, the seed set is taken to be the state-coordinate observables.

TSSD [30, 31] iteratively prunes the dictionary by retaining only directions that are approximately invariant under both the forward and backward EDMD operators; the tolerance parameter $\epsilon \in [0, 1]$ controls how strictly invariance is enforced, with $\epsilon = 0$ recovering exact symmetric subspace decomposition (SSD). We sweep ϵ over a uniform grid of $[0, 1]$ with spacing 0.01 and report each detected sub-dictionary size as a scatter point in the error plots. Despite the fine grid, TSSD identifies only a small number of distinct sizes: three for the Duffing oscillator ($N = 3, 45, 65$; Fig. 2a) and seven for the Van der Pol oscillator ($N = 3, 24, 25, 26, 27, 28, 51$; Fig. 6a), reflecting the coarse granularity of the invariance structure detected by the algorithm.

D.3 Duffing oscillator

The Duffing oscillator is

$$\dot{x} = y, \quad \dot{y} = -\delta y + \gamma x - \beta x^3, \tag{45}$$

with $\delta = 0.3$, $\gamma = 1$, $\beta = 1$, discretized at $\Delta t = 0.1$. We use the dictionary consisting of 90 bivariate Laguerre polynomials of order up to 12 with $M = 2000$ data points, which are i.i.d. samples from the two-dimensional uniform distribution on $[-2, 2]^2$. Fig. 5 shows the EDMD matrix before and after PR-based reordering, with entries rescaled via $a_{ij} \mapsto -1 + 2/(1 + \exp(-a_{ij}))$. For the

one-step prediction error (Fig. 2a), we use 20 random seeds and depict mean±1SD for each curve. For the multi-step state prediction (Fig. 2b), we use one representative among the 20 random seeds. Nevertheless, the overall tendency remains the same regardless of the random seed value, as evidenced by the small deviation in the one-step prediction error plot in Panel (a).

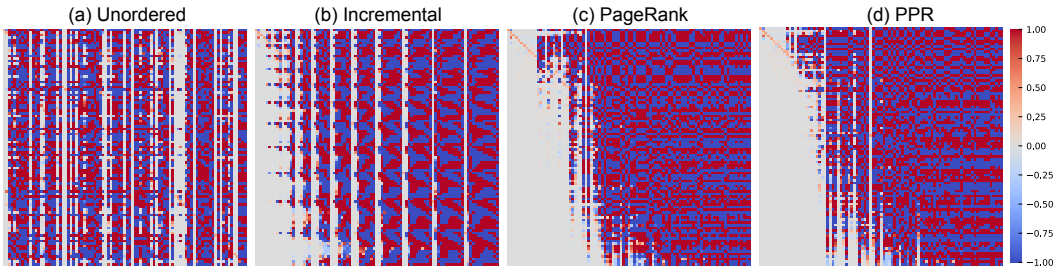

Figure 5: EDMD matrices for the Duffing oscillator with various permutations. (a) Randomly permuted. (b) Lexicographic order. (c) PR-ordered. (d) PPR-ordered.

D.4 Van der Pol oscillator

For the Van der Pol oscillator $\dot{x} = y$, $\dot{y} = \mu(1 - x^2)y - x$ with $\mu = 1.1$, discretized at $\Delta t = 0.1$, we use the same 90 Laguerre polynomials and $M = 2000$ data points, which are i.i.d. samples from the two dimensional uniform distribution on $[-2, 2]^2$. Figs. 6 and 7 show the EDMD prediction errors and heatmaps for the state (x, y). As done in the Duffing oscillator example above, we use 20 random seeds and depict mean±1SD for each curve, for the one-step prediction error (Fig. 6a). For the multi-step state prediction (Fig. 6b), we use one representative among the 20 random seeds. Nevertheless, the overall tendency remains the same regardless of the random seed value, as evidenced by the small deviation in the one-step prediction error plot in Panel (a).

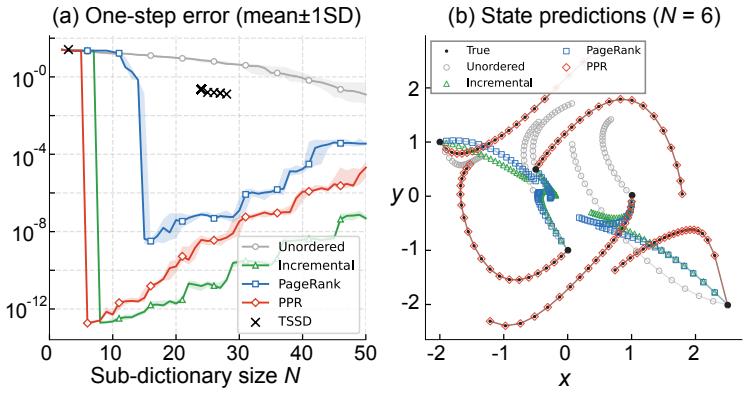

Figure 6: Van der Pol oscillator. (a) One-step errors. The curves represent mean±1SD over 20 random seeds. The tunable symmetric subspace decomposition (TSSD) detects only 7 sizes: 3, 24, 25, 26, 27, 28 and 51. (b) 20-step state predictions with 6 observables.

D.5 Three-well Ramachandran potential

Potential and dynamics. The potential on the torus $[-\pi, \pi]^2$ is

$$V(\phi, \psi) = \sum_{k=1}^{3} A_k \exp\left(-\frac{d(\phi, \phi_k^0)^2 + d(\psi, \psi_k^0)^2}{2\sigma_k^2}\right) + 0.3\left(\cos 2\phi + \cos 2\psi\right), \tag{46}$$

with geodesic distance $d(\cdot, \cdot)$ on the circle, well centers $(\phi_k^0, \psi_k^0) = (-1.0, -1.0)$, $(-1.0, 1.2)$, $(1.1, -0.3)$, depths $A_k = -6, -6, -4$, widths $\sigma_k = 0.55, 0.55, 0.65$. The dynamics follow the overdamped Langevin equation $d\phi = -\partial_\phi V\, dt + \sqrt{2\beta^{-1}}\, dW_\phi$, $d\psi = -\partial_\psi V\, dt + \sqrt{2\beta^{-1}}\, dW_\psi$ at $\beta = 1.0$, discretized with $\Delta t = 0.005$, producing $M = 100{,}000$

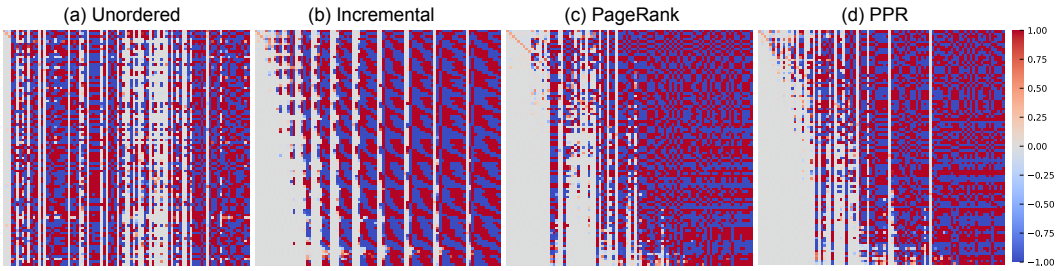

Figure 7: EDMD matrices for the Van der Pol oscillator with various permutations. (a) Randomly permuted. (b) Lexicographic order. (c) PR-ordered. (d) PPR-ordered.

frames. In other words, with 20 different random seeds, we generated 20 stochastic trajectories of the length 100,000.

Dictionary ($\tilde{N} = 196$). (i) state coordinates $\sin\phi, \cos\phi, \sin\psi, \cos\psi$ (4 terms, placed first); (ii) Fourier: $\sin(k\phi)$, $\cos(k\phi)$, $\sin(l\psi)$, $\cos(l\psi)$ for $k, l = 2, \ldots, 8$ (28 terms); (iii) cross terms: $\sin(k\phi)\cos(l\psi)$, etc., for $k, l = 1, \ldots, 4$ (64 terms); (iv) Gaussian RBFs on a 10×10 grid with $\sigma_{\text{RBF}} = 0.45$ (100 terms).

Comparison methods. (a) PR ($\alpha = 0.85$); (b) PPR ($\alpha = 0.85$, multi-seed at the 4 state coordinates); (c) TICA at lag 1, ranking by maximum absolute loading on top-10 slow components; (d) PCCA+ with 50 microstates (k-means), lag 5, 3 macrostates, ranking by Fisher discriminant; (e) Random. All methods rank the same $\tilde{N} = 196$ dictionary; EDMD is recomputed on the selected N observables via ridge-regularized least squares. All data-dependent preprocessing, ranking, and fitting use the first 80% of each trajectory; prediction errors are evaluated on the remaining 20%.

For the 196-observable dictionary, we assess damping sensitivity over ten values of α from 0.05 to 0.95 at $N \in \{5, 10, 20, 40\}$, using the same train-only fitting and state-coordinate seeds as in the main experiment. Table 3 shows that held-out prediction error is stable across the tested damping values, although the selected dictionaries can change. All four state-coordinate observables remain selected throughout.

Table 3: Damping sensitivity of PPR on the 196-observable Ramachandran dictionary over 20 trajectories. The MSE range is the minimum and maximum, over the tested α values, of the trajectory-averaged held-out one-step MSE. Mean CV averages the trajectory-wise coefficients of variation across α. The final column gives the minimum, over α, of the trajectory-averaged Jaccard index relative to the selection at $\alpha = 0.85$, where $J(A, B) = |A \cap B|/|A \cup B|$.

N	Mean MSE range	Mean CV (%)	Min. mean Jaccard
5	0.8960–0.8969	0.092	0.750
10	0.8797–0.8839	0.340	0.518
20	0.8559–0.8600	0.550	0.538
40	0.8265–0.8297	0.229	0.647

We also assess seed dependence using the same train-only protocol at $\alpha = 0.85$ and $N \in \{5, 10, 20, 40\}$. We compare the four state-coordinate seeds with a PR-to-PPR fallback that uses the four highest-ranked observables from standard PR as seeds. Table 4 reports held-out one-step MSE over 20 trajectories. State-coordinate seeds give lower errors in every trajectory at each size, illustrating that globally central seeds need not be aligned with state prediction.

Using the same 196-observable dictionary, training/test split, state-coordinate seeds, and ridge readout, we compare row-normalized PPR with the globally normalized variant at $\alpha = 0.85$ over 20 trajectories. For $N = 5, 10, 20, 40$, the mean held-out one-step MSEs are $(0.8966, 0.8826, 0.8559, 0.8285)$ for row-normalized PPR and $(0.8961, 0.8830, 0.8650, 0.8321)$ for the globally normalized variant. The variant does not consistently reduce prediction error.

Table 4: Seed dependence for the 196-observable Ramachandran dictionary. Held-out one-step MSE is reported as mean ± one standard deviation over 20 trajectories. Both policies use four seeds; state-coordinate observables are not forced into the selected dictionaries.

N	State-coordinate seeds	PR-to-PPR fallback
5	0.897 ± 0.041	3.208 ± 0.300
10	0.883 ± 0.040	2.595 ± 0.338
20	0.856 ± 0.042	1.640 ± 0.206
40	0.829 ± 0.038	1.005 ± 0.090

Additional figures. Fig. 8 shows the EDMD matrix under different orderings: both PR and PPR reveal approximate zero block structure invisible under random, TICA, PCCA+ ordering. These heatmaps show log absolute values for visibility. Fig. 9a shows that PPR selects all 4 state coordinates at ranks 1–4 followed by basin-boundary RBFs (white dots). Such basin-boundary RBFs seem to effectively capture inter-basin transitions, yielding accurate predictions with smaller sub-dictionaries. Fig. 9c,d shows the score distributions: PPR exhibits a sharp two-tier structure (state coordinates receiving ≈17% of total mass, with a factor-of-6.7 drop to the next tier). Compared with the randomly permuted sub-dictionaries, PPR yielded the largest improvement (Fig. 9e).

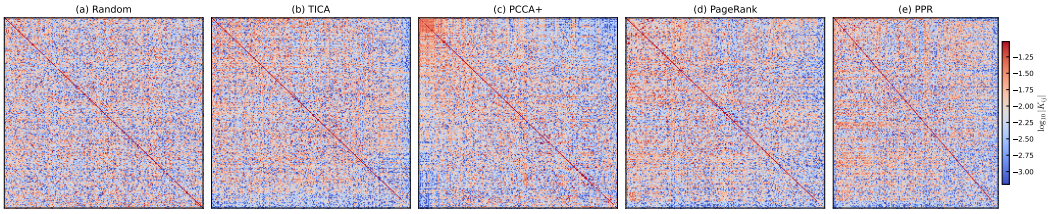

Figure 8: EDMD matrices for the Ramachandran potential with various permutations. (a) Random. (b) TICA. (c) PCCA+ (d) PR. (e) PPR.

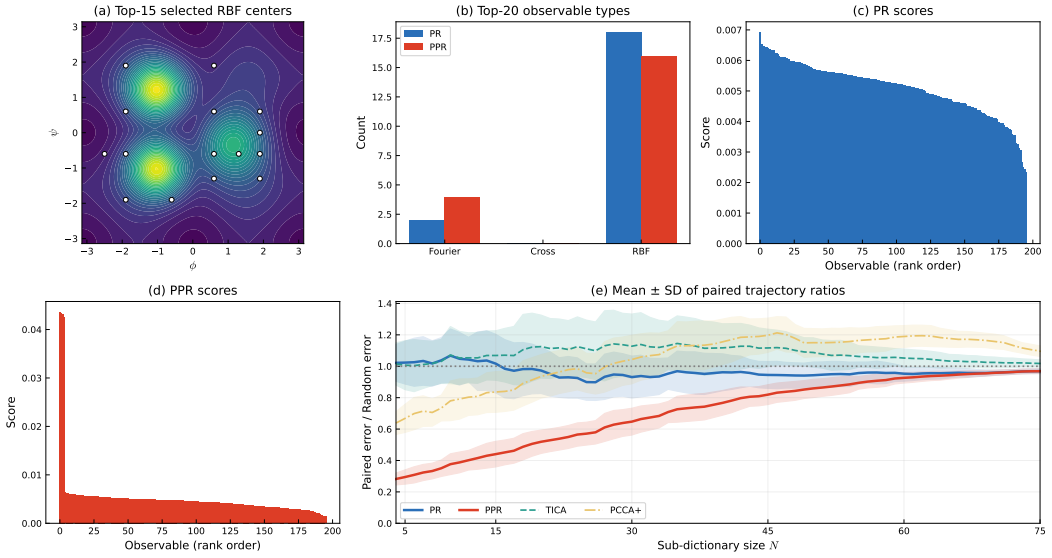

Figure 9: Three-well Ramachandran potential with one representative experiment for illustration. (a) Potential energy landscape. White dots represent the centers of top 15 RBFs detected by PPR. (b) Types of top 20 observables chosen by PageRank and PPR. (c) Standard PR scores. (d) PPR scores: sharp two-tier structure. (e) Improvement ratio vs. N.

D.6 Lorenz system

We generate 3×10^5 data points from a single initial condition at $\Delta t = 0.001$, discard the first 10^5 as burn-in, and use the remaining $M = 2 \times 10^5$ data points for EDMD. The dictionary is a time-delay embedding of the three-dimensional state (x, y, z) with $N_{\text{delay}} = 100$ delay levels and stride $h = 100$ fine steps, yielding $\tilde{N} = 300$ observables and $M = 2 \times 10^5$ snapshot pairs. These delayed coordinates form a Krylov-type dictionary under repeated Koopman action, connecting the construction to the Arnoldi interpretation of DMD [28]. The observable matrix is

$$\mathbf{\Psi}(\mathbf{X}) = \begin{pmatrix} \mathbf{x}(0) & \mathbf{x}(h\Delta t) & \cdots & \mathbf{x}((N_{\text{delay}}-1)h\Delta t) \\ \mathbf{x}(\Delta t) & \mathbf{x}(\Delta t + h\Delta t) & \cdots & \mathbf{x}(\Delta t + (N_{\text{delay}}-1)h\Delta t) \\ \vdots & \vdots & & \vdots \\ \mathbf{x}((M-1)\Delta t) & \cdots & \cdots & \mathbf{x}((M-1)\Delta t + (N_{\text{delay}}-1)h\Delta t) \end{pmatrix} \in \mathbb{R}^{M \times \tilde{N}}.$$

For the Incremental, PR, and PPR orderings, we reconstruct the EDMD with $N_r \in \{30, 60, 120, 300\}$ and identify the pseudo-eigenfunction at frequency $\log(\lambda)/(i\Delta t) \approx 6$ rad/s, consistent with prior work [9, 18].

We note that a direct one-step prediction error comparison analogous to Figs. 2 and 6 is not presented for the Lorenz system. This is because the time-delay embedding dictionary directly contains the state coordinates as its first-level observables, so the recomputed EDMD matrix on any sub-dictionary that includes the first delay level already achieves accurate one-step state prediction. The eigenfunction comparison of Fig. 4 therefore provides a more discriminating quality measure for this dictionary class, probing whether the selected sub-dictionary recovers the correct spectral geometry of the Koopman operator, which is a strictly harder task than one-step prediction accuracy alone.

D.7 Computational resources

All experiments were conducted on a standard laptop computer equipped with an Apple M5 CPU and 32 GB of RAM. No GPU acceleration was used. All code is implemented in Python, using `numpy` for linear algebra and `numba` for JIT-compiled simulation of the Ramachandran potential; the overdamped Langevin time-stepping loop is compiled to native machine code via LLVM at the first call, with a short warmup run performed before production to amortize the compilation cost. The total compute time for all experiments reported in the paper is estimated at approximately 30–40 minutes on the hardware described above.

E Proofs

This appendix contains all proofs. We organize them following the logical structure of the paper: error bounds and zero block structure (Appendix E.1), the coupling bound and unconditional leakage guarantee (Appendix E.2), the PR and PPR lemmas on block-structured matrices (Appendix E.3), and finite-sample extensions (Appendix F).

E.1 Error bounds and zero block structure

Goal. This subsection proves the three results that justify reading sub-dictionary quality off the off-diagonal block of the EDMD matrix: Proposition 8 (intermediate prediction-error bound in the infinite-dimensional Koopman matrix), Proposition 1 (computable surrogate for the infinite tail), and Theorem 2 (the ideal case of a Koopman invariant sub-dictionary).

Strategy. We proceed in two stages. Proposition 8 decomposes the EDMD prediction error into an approximation term (projection error within $\mathcal{F}_N$) and a statistical term (finite-sample EDMD error); the orthonormal basis structure and Parseval's identity turn the approximation term into $\|K^{21,\text{true}}\|_F^2$. Proposition 1 then replaces this infinite-dimensional quantity with the computable finite block $\|K^{21}_{\tilde{N},M}\|_F^2$ plus a truncation error, so the bound becomes evaluable from data. Finally, Theorem 2 shows the ideal case: when the sub-dictionary spans a Koopman invariant subspace, the off-diagonal block vanishes identically; the proof exploits the Schur complement structure of the augmented Gram matrix.

We first record the convergence of the EDMD matrix under Assumption 2; this is the only probabilistic input to the proofs of Propositions 8 and 1.

Lemma 7 (Almost-sure EDMD convergence)**.** *Let $\mathcal{D}_N$ be linearly independent in $L^2(\mathcal{M},\mu)$ and let the data satisfy Assumption 2. Then almost surely Assumption 1 holds for all sufficiently large M and $K_{N,M} \to K_N = G_N^{-1} A_N$, where $A_N[i,j] = \langle \mathcal{K}\psi_j, \psi_i\rangle_\mu$ and $\mathcal{K}$ is the deterministic or stochastic Koopman operator as appropriate. Consequently $\|\mathcal{K}_N\psi_i - \mathcal{K}_{N,M}\psi_i\|_{L^2} \to 0$ almost surely for each $i \le N$.*

Proof. Write $\widehat{G}_M = \frac{1}{M}\mathbf{\Psi}(\mathbf{X})^*\mathbf{\Psi}(\mathbf{X})$ and $\widehat{A}_M = \frac{1}{M}\mathbf{\Psi}(\mathbf{X})^*\mathbf{\Psi}(\mathbf{Y})$, so that $K_{N,M} = \widehat{G}_M^{-1}\widehat{A}_M$ whenever $\widehat{G}_M$ is invertible. It suffices to prove $\widehat{G}_M \to G_N$ and $\widehat{A}_M \to A_N$ entrywise almost surely: invertibility of G_N then forces $\widehat{G}_M$ to be invertible for all large M, which is Assumption 1, and the conclusion follows by continuity of matrix inversion together with the equivalence of norms on the finite-dimensional space $\mathcal{F}_N$. In regime (S1) this is the strong law of large numbers applied to $\psi_i^*\psi_j$ and $\psi_i^*(\psi_j \circ \mathbf{F})$, both of which lie in $L^1(\mu)$. In regime (S2), the entries are averages of $\psi_i^*\psi_j$ and $\psi_i^*(\psi_j \circ \mathbf{F})$ along the orbit of $\mathbf{x}^{(1)}$. These functions are continuous, and on the compact set formed by the closure of the orbit and $\operatorname{supp}\mu$ they agree with bounded continuous functions, so $\mathbf{x}^{(1)} \in B(\mu)$ gives convergence to $\langle\psi_i,\psi_j\rangle_\mu$ and $\langle\mathcal{K}\psi_j,\psi_i\rangle_\mu$. This replaces Birkhoff's theorem in the argument of [15, §4.1], which gives convergence only for μ-a.e. $\mathbf{x}^{(1)}$. In regime (S3), the pair process $\mathbf{z}_k = (\mathbf{x}_k, \mathbf{x}_{k+1})$ is uniformly ergodic (given $\mathbf{z}_0$, the law of $\mathbf{z}_k$ is within total variation $d(k-1)$ of its stationary law), hence positive Harris recurrent, and the strong law of large numbers for positive Harris chains [63, Thm. 17.0.1], which holds for every initial distribution, applied to $h_{ij}(\mathbf{x},\mathbf{y}) = \psi_i(\mathbf{x})^*\psi_j(\mathbf{y})$ gives

$$\widehat{A}_M[i,j] \to \mathbb{E}_{\mathbf{x}_0\sim\mu,\mathbf{x}_1|\mathbf{x}_0\sim P(\mathbf{x}_0,\cdot)}[\psi_i(\mathbf{x}_0)^*\psi_j(\mathbf{x}_1)] \tag{47}$$

$$= \mathbb{E}_{\mathbf{x}_0\sim\mu}\left[\mathbb{E}_{\mathbf{x}_1|\mathbf{x}_0\sim P(\mathbf{x}_0,\cdot)}\left[\psi_i(\mathbf{x}_0)^*\psi_j(\mathbf{x}_1)\,\middle|\,\mathbf{x}_0\right]\right] \tag{48}$$

$$= \mathbb{E}_{\mathbf{x}_0\sim\mu}\left[\psi_i(\mathbf{x}_0)^*\mathbb{E}_{\mathbf{x}_1|\mathbf{x}_0\sim P(\mathbf{x}_0,\cdot)}\left[\psi_j(\mathbf{x}_1)\,\middle|\,\mathbf{x}_0\right]\right] \tag{49}$$

$$= \mathbb{E}_\mu[\psi_i(\mathbf{x}_0)^*(\mathcal{K}\psi_j)(\mathbf{x}_0)] = A_N[i,j], \tag{50}$$

where the first equality uses the iterated expectation and the second equality uses the definition of the stochastic Koopman operator, while $\widehat{G}_M \to G_N$ as before; for i.i.d. pairs the strong law replaces Birkhoff in the same computation. □

The one-step evolution of ψ_i decomposes as in (4). The prediction error is first bounded in terms of the infinite-dimensional Koopman matrix:

Proposition 8. *Under the same assumptions as in Proposition 1, for each $\varepsilon > 0$, almost surely there exists $M_0 > 0$ such that for all $M \ge M_0$, $\sum_{i=1}^N \|\mathcal{K}\psi_i - \mathcal{K}_{N,M}\psi_i\|_{L^2}^2 \le 2\|K^{21,\text{true}}\|_F^2 + \varepsilon$.*

Since $K^{21,\text{true}}$ is not computable from data, Proposition 1 reformulates the bound in terms of the finite EDMD matrix: the first term $\|K^{21}_{\tilde{N},M}\|_F^2$ is then computable from data, and the second is a truncation error decreasing in $\tilde{N}$.

Proof of Proposition 8. We first derive the error term for a single ψ_i and then the one for the summation over $i = 1,\dots,N$. For any $i \in \{1,\dots,N\}$, we have

$$\|\mathcal{K}\psi_i - \mathcal{K}_{N,M}\,\psi_i\|_{L^2}^2 \le 2\left(\|\mathcal{K}\psi_i - \mathcal{K}_N\,\psi_i\|_{L^2}^2 + \|\mathcal{K}_N\psi_i - \mathcal{K}_{N,M}\,\psi_i\|_{L^2}^2\right), \tag{51}$$

where $\mathcal{K}_N$ denotes $\mathcal{P}_{\mathcal{F}_N}\mathcal{K}_{|\mathcal{F}_N}$ and $\mathcal{P}_{\mathcal{F}_N} : \mathcal{F} \to \mathcal{F}_N$ is the projection operator onto $\mathcal{F}_N$. Note that this projection is well defined as $\mathcal{F}_N$ is a finite-dimensional subspace of $\mathcal{F} = L^2(\mathcal{M},\mu)$ and thus a closed subspace. Under Assumption 1, Lemma 7 (which reduces to Theorem 2 of [15] in regime (S1)) implies that the following holds:

$$\lim_{M\to\infty}\|\mathcal{K}_N\psi_i - \mathcal{K}_{N,M}\,\psi_i\|_{L^2} = 0.$$

Therefore, there exists $M > 0$ such that

$$\|\mathcal{K}_N\psi_i - \mathcal{K}_{N,M}\,\psi_i\|_{L^2} < \sqrt{\frac{\varepsilon}{2N}},$$

and thus

$$\|\mathcal{K}_N\psi_i - \mathcal{K}_{N,M}\,\psi_i\|_{L^2}^2 < \frac{\varepsilon}{2N}. \tag{52}$$

From (4), the term, $\|\mathcal{K}\psi_i - \mathcal{K}_N\,\psi_i\|_{L^2}^2$, in (51) can be expressed as follows:

$$\begin{aligned}\|\mathcal{K}\psi_i - \mathcal{K}_N\,\psi_i\|_{L^2}^2 &= \|\mathbf{\Psi}\,K^{\text{true}}\mathbf{e}_i - \mathbf{\Psi}_N\,K^{11,\text{true}}\mathbf{e}_i\|_{L^2}^2 \\ &= \|\mathbf{\Psi}_{N^c}\,K^{21,\text{true}}\mathbf{e}_i\|_{L^2}^2,\end{aligned}$$

which is identical to $\|K^{21,\text{true}}\mathbf{e}_i\|_{\ell^2}^2$ by Parseval's identity for the orthonormal set.

Combining this with (51) and (52), we have the following inequality for each $i \in \{1,\dots,N\}$:

$$\begin{aligned}\|\mathcal{K}\psi_i - \mathcal{K}_{N,M}\,\psi_i\|_{L^2}^2 &\le 2\left(\|\mathcal{K}\psi_i - \mathcal{K}_N\,\psi_i\|_{L^2}^2 + \|\mathcal{K}_N\psi_i - \mathcal{K}_{N,M}\,\psi_i\|_{L^2}^2\right) \\ &\le 2\left(\|K^{21,\text{true}}\mathbf{e}_i\|_{\ell^2}^2 + \frac{\varepsilon}{2N}\right).\end{aligned}$$

Taking the sum over $i = 1,\dots,N$, we obtain:

$$\sum_{i=1}^{N}\|\mathcal{K}\psi_i - \mathcal{K}_{N,M}\,\psi_i\|_{L^2}^2 \le 2\|K^{21,\text{true}}\|_F^2 + \varepsilon,$$

which completes the proof. □

The next step splits $K^{21,\text{true}}$ (which has infinitely many rows) into a finite computable part $K^{21}_{\tilde{N},M}$ and a tail that decreases with $\tilde{N}$, then uses EDMD convergence to relate the finite part to data.

Proof of Proposition 1. Given Proposition 8, it is sufficient to show

$$\|K^{21,\text{true}}\|_F^2 \le \|K^{21}_{\tilde{N},M}\|_F^2 + \|K^{\text{true}}[(\tilde{N}+1):\infty,\,1:\tilde{N}]\|_F^2 + \frac{\varepsilon}{2}$$

for a large $M > 0$. By definition,

$$\|K^{21,\text{true}}\|_F^2 = \|K^{\text{true}}[(N+1):\tilde{N},\,1:N]\|_F^2 + \|K^{\text{true}}[(\tilde{N}+1):\infty,\,1:N]\|_F^2,$$

and

$$\|K^{\text{true}}[(\tilde{N}+1):\infty,\,1:N]\|_F^2 \le \|K^{\text{true}}[(\tilde{N}+1):\infty,\,1:\tilde{N}]\|_F^2$$

since the left-hand side sums over columns $j = 1,\dots,N$ while the right-hand side sums over the larger set of columns $j = 1,\dots,\tilde{N}$. Moreover,

$$\|K^{21}_{\tilde{N},M}\|_F^2 \to \|K^{\text{true}}[(N+1):\tilde{N},\,1:N]\|_F^2 \quad \text{as} \quad M \to \infty$$

since $\mathcal{K}_{\tilde{N},M}$ converges to the projection of $\mathcal{K}$ onto $\mathcal{F}_{\tilde{N}}$ (Lemma 7; Theorem 2 in [15] for regime (S1)) and the matrix $K^{21}_{\tilde{N},M}$ contains only finitely many terms. Therefore, there exists $M > 0$ such that

$$\|K^{\text{true}}[(N+1):\tilde{N},\,1:N]\|_F^2 \le \|K^{21}_{\tilde{N},M}\|_F^2 + \frac{\varepsilon}{2},$$

which completes the proof. □

Theorem 2 is the exact-case companion: when $\mathcal{F}_N$ is $\mathcal{K}$-invariant, the off-diagonal block vanishes identically. The proof uses the block structure of the augmented evaluation matrices $\mathbf{\Psi}_{\tilde{N}}(\mathbf{X}), \mathbf{\Psi}_{\tilde{N}}(\mathbf{Y}) \in \mathbb{C}^{M\times\tilde{N}}$, combined with the Schur complement formula for block matrix inversion. The crucial step is substituting the invariance relation $\mathbf{\Psi}_N(\mathbf{X})\,K_{N,M} = \mathbf{\Psi}_N(\mathbf{Y})$ (which holds exactly when $\mathcal{F}_N$ is $\mathcal{K}$-invariant) into the block product; the off-diagonal block $K^{21}_{\tilde{N},M}$ then collapses to zero through an algebraic cancellation involving the Schur complement $W = A - BD^{-1}C$.

Proof of Theorem 2. Let $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(M)}\}$ be the data points in the theorem, with $\mathbf{y}^{(i)} = \mathbf{F}(\mathbf{x}^{(i)})$. With the initial dictionary $\mathcal{D}_{\tilde{N}} = \{\psi_1, \ldots, \psi_{\tilde{N}}\}$ and the sub-dictionary $\mathcal{D}_N = \{\psi_1, \ldots, \psi_N\}$, the EDMD matrices are

$$K_{\tilde{N},M} = \mathbf{\Psi}_{\tilde{N}}(\mathbf{X})^+ \mathbf{\Psi}_{\tilde{N}}(\mathbf{Y}) \;=\; \big(\mathbf{\Psi}_{\tilde{N}}(\mathbf{X})^* \mathbf{\Psi}_{\tilde{N}}(\mathbf{X})\big)^{-1} \mathbf{\Psi}_{\tilde{N}}(\mathbf{X})^* \mathbf{\Psi}_{\tilde{N}}(\mathbf{Y}), \tag{53}$$

$$K_{N,M} = \mathbf{\Psi}_{N}(\mathbf{X})^+ \mathbf{\Psi}_{N}(\mathbf{Y}) \;=\; \big(\mathbf{\Psi}_{N}(\mathbf{X})^* \mathbf{\Psi}_{N}(\mathbf{X})\big)^{-1} \mathbf{\Psi}_{N}(\mathbf{X})^* \mathbf{\Psi}_{N}(\mathbf{Y}). \tag{54}$$

Partition the columns of the augmented evaluation matrices into the first N (i.e. $\mathcal{D}_N$) and the remaining $\tilde{N} - N$ (i.e. $\mathcal{D}_{N_2} := \mathcal{D}_{\tilde{N}} \setminus \mathcal{D}_N$):

$$\mathbf{\Psi}_{\tilde{N}}(\mathbf{X}) = \begin{bmatrix} \mathbf{\Psi}_N(\mathbf{X}) & \mathbf{\Psi}_{N_2}(\mathbf{X}) \end{bmatrix} \quad \text{and} \quad \mathbf{\Psi}_{\tilde{N}}(\mathbf{Y}) = \begin{bmatrix} \mathbf{\Psi}_N(\mathbf{Y}) & \mathbf{\Psi}_{N_2}(\mathbf{Y}) \end{bmatrix},$$

with $\mathbf{\Psi}_N(\mathbf{X}), \mathbf{\Psi}_N(\mathbf{Y}) \in \mathbb{C}^{M \times N}$ and $\mathbf{\Psi}_{N_2}(\mathbf{X}), \mathbf{\Psi}_{N_2}(\mathbf{Y}) \in \mathbb{C}^{M \times (\tilde{N}-N)}$. With this partition, (53) reads

$$K_{\tilde{N},M} = \begin{bmatrix} \mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{X}) & \mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_{N_2}(\mathbf{X}) \\ \mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{X}) & \mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_{N_2}(\mathbf{X}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{Y}) & \mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_{N_2}(\mathbf{Y}) \\ \mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{Y}) & \mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_{N_2}(\mathbf{Y}) \end{bmatrix}. \tag{55}$$

To further simplify (55), we use the block matrix inversion formula

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} W^{-1} & -W^{-1}BD^{-1} \\ -D^{-1}CW^{-1} & D^{-1} + D^{-1}CW^{-1}BD^{-1} \end{bmatrix}, \tag{56}$$

valid whenever D is invertible and $W := A - BD^{-1}C$ (i.e., the Schur complement of D) is invertible. Applying (56) with $A = \mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{X})$, $B = \mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_{N_2}(\mathbf{X})$, $C = \mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{X})$, $D = \mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_{N_2}(\mathbf{X})$, we rewrite (55) as

$$K_{\tilde{N},M} = \begin{bmatrix} W^{-1} & -W^{-1}BD^{-1} \\ -D^{-1}CW^{-1} & D^{-1} + D^{-1}CW^{-1}BD^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{Y}) & \mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_{N_2}(\mathbf{Y}) \\ \mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{Y}) & \mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_{N_2}(\mathbf{Y}) \end{bmatrix}. \tag{57}$$

By Assumption 1, the Gram matrix $\mathbf{\Psi}_{\tilde{N}}(\mathbf{X})^* \mathbf{\Psi}_{\tilde{N}}(\mathbf{X})$ and its block D are invertible, so the Schur complement W is also invertible; hence (57) is valid.

Part (i). From (57), the top-left $N \times N$ block of $K_{\tilde{N},M}$ equals

$$\begin{aligned} & W^{-1}\mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{Y}) - W^{-1}BD^{-1}\mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{Y}) \\ &\quad = W^{-1}\big\{\mathbf{\Psi}_N(\mathbf{X})^* - BD^{-1}\mathbf{\Psi}_{N_2}(\mathbf{X})^*\big\}\mathbf{\Psi}_N(\mathbf{Y}) && (58) \\ &\quad = W^{-1}\big\{\mathbf{\Psi}_N(\mathbf{X})^* - BD^{-1}\mathbf{\Psi}_{N_2}(\mathbf{X})^*\big\}\mathbf{\Psi}_N(\mathbf{X})\, K_{N,M} && (59) \\ &\quad = W^{-1}\big\{\mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{X}) - BD^{-1}\mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{X})\big\}K_{N,M} \\ &\quad = W^{-1}\big\{A - BD^{-1}C\big\}K_{N,M} \;=\; W^{-1} \cdot W \cdot K_{N,M} \;=\; K_{N,M}. \end{aligned}$$

From (58) to (59) we used the invariance of $\mathcal{F}_N$, namely $\mathbf{\Psi}_N(\mathbf{Y}) = \mathbf{\Psi}_N(\mathbf{X})\, K_{N,M}$, as in (18).

Part (ii). Again from (57), the bottom-left $(\tilde{N} - N) \times N$ block of $K_{\tilde{N},M}$ equals

$$\begin{aligned} & -D^{-1}CW^{-1}\mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{Y}) + \big(D^{-1} + D^{-1}CW^{-1}BD^{-1}\big)\mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{Y}) \\ &\quad = -D^{-1}CW^{-1}\mathbf{\Psi}_N(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{X})\, K_{N,M} + D^{-1}\mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{X})\, K_{N,M} \\ &\qquad\qquad + D^{-1}CW^{-1}BD^{-1}\mathbf{\Psi}_{N_2}(\mathbf{X})^* \mathbf{\Psi}_N(\mathbf{X})\, K_{N,M} \\ &\quad = -D^{-1}CW^{-1}A\, K_{N,M} + D^{-1}C\, K_{N,M} + D^{-1}CW^{-1}BD^{-1}C\, K_{N,M} \\ &\quad = D^{-1}C\, W^{-1}\big(-A + W + BD^{-1}C\big)\, K_{N,M} \\ &\quad = D^{-1}C\, W^{-1} \cdot O_{N \times N} \cdot K_{N,M} \;=\; O_{(\tilde{N}-N)\times N}. \end{aligned}$$

This completes the proof. □

Remark 2 (Remark for the proof)**.** It is noteworthy that the data set to obtain the EDMD matrix $K_{\tilde{N},M}$ does not need to be the same as the one for $K_{N,M}$. Since $\mathcal{F}_N$ is invariant, $K_{N,M}$ is identical to the matrix form (i.e., K_N) of the projection of the restricted Koopman operator, $\mathcal{K}_N$, under Assumption 1. Therefore, $K_{N,M} = K_N$ has no data dependency, and this proof can be simply reproduced with different data sets for the two EDMD matrices, $K_{\tilde{N},M}$ and $K_{N,M}$.

E.2 Coupling bound and unconditional leakage

Goal. This subsection proves the unconditional leakage guarantee (Proposition 10), which bounds multi-step Koopman leakage by the PPR mass outside the selected set.

Strategy. The bridge between multi-step Koopman dynamics and the PPR Neumann series is a single entrywise domination result (Lemma 9) showing that the entries of $K_{\tilde{N}}^k$ are controlled by those of the row-normalized matrix P^k up to a factor $r_{\max}^k$. Given this control, the discounting rates align via the change of variables $\gamma = \alpha/r_{\max}$, which converts the Koopman-side geometric sum into the PPR Neumann series and reveals $(1-\gamma)/(1-\alpha) \cdot (1 - \pi_{\mathbf{s}}(S_N))$ as the leakage bound.

Lemma 9 (Entrywise coupling to a row-normalized matrix)**.** *Let A be a square matrix with absolute row sums $r_i := \sum_l |A[i,l]| > 0$, row normalization $\hat{A}[i,j] := |A[i,j]|/r_i$, and $r_{\max} := \max_i r_i$. Let $|A|$ denote the entrywise modulus of A. Then for all v, j and $k \geq 0$:*

$$|A^k[v,j]| \leq |A|^k[v,j] \leq r_{\max}^k \, \hat{A}^k[v,j]. \tag{60}$$

Proof. The first inequality is the entrywise triangle inequality applied to the k-fold product:

$$|A^k[v,j]| = \Big| \sum_{l_1,\ldots,l_{k-1}} A[v,l_1] \cdots A[l_{k-1},j] \Big| \leq \sum_{l_1,\ldots,l_{k-1}} |A[v,l_1]| \cdots |A[l_{k-1},j]| = |A|^k[v,j].$$

We prove the second inequality by induction on k. The base case $k = 0$ is straightforward. For $k = 1$: $|A|[v,j] = r_v \, \hat{A}[v,j] \leq r_{\max} \, \hat{A}[v,j]$, using $|A[v,j]| = r_v \hat{A}[v,j]$ by definition of $\hat{A}$. For the inductive step $k \to k+1$:

$$\begin{aligned} |A|^{k+1}[v,j] &= \sum_l |A|^k[v,l] \, |A[l,j]| \\ &\leq \sum_l r_{\max}^k \, \hat{A}^k[v,l] \cdot r_l \, \hat{A}[l,j] \leq r_{\max}^{k+1} \sum_l \hat{A}^k[v,l] \, \hat{A}[l,j] = r_{\max}^{k+1} \, \hat{A}^{k+1}[v,j], \end{aligned}$$

where the inequality uses the inductive hypothesis $|A|^k[v,l] \leq r_{\max}^k \hat{A}^k[v,l]$, the identity $|A[l,j]| = r_l \hat{A}[l,j]$, and $r_l \leq r_{\max}$. □

Proposition 10 (Multi-step leakage and PPR)**.** *Let $P = \mathcal{R}(|K_{\tilde{N},M}^{\intercal}|)$, $r_{\max,M} = \|K_{\tilde{N},M}^{\intercal}\|_\infty$, and $S_N \subset \{1,\ldots,\tilde{N}\}$ with $|S_N| = N$. Assume $\alpha < r_{\max,M}$ and set $\gamma := \alpha/r_{\max,M} \in (0,1)$ (when $r_{\max,M} \geq 1$ this holds for every $\alpha \in (0,1)$; when $r_{\max,M} < 1$, the damping factor must satisfy $\alpha < r_{\max,M}$). For any preference vector $\mathbf{s}$ supported on S_N, the γ-*discounted multi-step leakage *defined by*

$$\Lambda_{\mathbf{s}}^{\gamma}(S_N) \coloneqq (1-\gamma) \sum_{k=1}^{\infty} \gamma^k \sum_{j \notin S_N} [\mathbf{s}^{\intercal} |K_{\tilde{N},M}^{\intercal}|^k]_j \tag{61}$$

satisfies

$$\Lambda_{\mathbf{s}}^{\gamma}(S_N) \leq \frac{1-\gamma}{1-\alpha} \left(1 - \pi_{\mathbf{s},M}(S_N)\right),$$

where $\pi_{\mathbf{s},M}$ is the PPR vector of P with preference $\mathbf{s}$ and damping α.

Proof. Write $P_v^{(k)} = \sum_{j \notin S_N} P^k[v,j]$. For $v \in S_N$, we have $P_v^{(0)} = 0$. Apply Lemma 9 with $A = K_{\tilde{N},M}^{\intercal}$; then $\hat{A} = \mathcal{R}(|K_{\tilde{N},M}^{\intercal}|) = P$, and equation (60) gives $|K_{\tilde{N},M}^{\intercal}|^k[v,j] \leq r_{\max,M}^k \, P^k[v,j]$, so

$$\sum_{j \notin S_N} |K_{\tilde{N},M}^{\intercal}|^k[v,j] \leq r_{\max,M}^k P_v^{(k)}.$$

Expanding the definition of $\Lambda_{\mathbf{s}}^{\gamma}(S_N) = (1-\gamma) \sum_{k=1}^{\infty} \gamma^k \sum_{j \notin S_N} [\mathbf{s}^{\intercal} |K_{\tilde{N},M}^{\intercal}|^k]_j$ and using $[\mathbf{s}^{\intercal} |K_{\tilde{N},M}^{\intercal}|^k]_j = \sum_v s_v |K_{\tilde{N},M}^{\intercal}|^k[v,j]$,

$$\Lambda_{\mathbf{s}}^{\gamma}(S_N) \leq \sum_v s_v \cdot (1-\gamma) \sum_{k=1}^{\infty} \gamma^k r_{\max,M}^k \, P_v^{(k)}.$$

Setting $\gamma = \alpha / r_{\max,M}$ gives $\gamma^k r_{\max,M}^k = \alpha^k$, so

$$\Lambda_{\mathbf{s}}^{\gamma}(S_N) \le \sum_v s_v \cdot (1-\gamma) \sum_{k=1}^{\infty} \alpha^k P_v^{(k)} = \sum_v s_v \cdot \frac{1-\gamma}{1-\alpha}(1 - \pi_{v,M}(S_N)),$$

where the last equality uses the single-seed PPR Neumann series (21): $(1-\alpha)\sum_{k=1}^{\infty} \alpha^k P_v^{(k)} = 1 - \pi_{v,M}(S_N)$. By linearity of $\mathbf{s} \mapsto \pi_{\mathbf{s},M}$, $\sum_v s_v \, \pi_{v,M}(S_N) = \pi_{\mathbf{s},M}(S_N)$, giving

$$\Lambda_{\mathbf{s}}^{\gamma}(S_N) \le \frac{1-\gamma}{1-\alpha} \left(1 - \pi_{\mathbf{s},M}(S_N)\right). \qquad \square$$

Effect of row normalization. Consider the nonnegative weight matrix

$$W = \begin{pmatrix} 0.90 & 0.05 & 0.05 \\ 0.60 & 0.40 & 0 \\ 0.03 & 0 & 0.07 \end{pmatrix}, \qquad \mathbf{s} = \mathbf{e}_1, \qquad \alpha = 0.85.$$

The third row has sum 0.1, but row normalization raises its self-loop from 0.07 to 0.7, above the second row's 0.4. The resulting PPR scores are $\pi_2 \approx 0.0551$ and $\pi_3 \approx 0.0897$. Global normalization uses $A = W/r_{\max} = W$ and the score $\mathbf{q} = (1-\alpha)(I - \alpha A^{\intercal})^{-1}\mathbf{s}$, giving $q_2 \approx 0.0481$ and $q_3 \approx 0.0337$. These scores select $\{1,3\}$ and $\{1,2\}$, respectively, at $N = 2$. Thus, row normalization can change the selected observables by discarding row-magnitude information. The globally normalized scores need not sum to one.

E.3 Proof of Theorem 3

Goal. This subsection proves Theorem 3. The main-text sketch (§3.3) reduces detection on P to (a) a perturbation bound between P and $P^{(0)}$ and (b) explicit lower bounds on the auxiliary gaps $\Delta_{\text{PR}}^{(0)}$, $\Delta_{\text{PPR}}^{(0)}$ via the closed-form expressions of Lemma 4 (only this lemma is stated in the main text).

Strategy. We first prove Lemma 4 (§E.3.1) by computing $(I - \alpha P^{(0)})^{-1}$ via the block-lower-triangular structure of $P^{(0)}$. We then derive the explicit lower bounds: on $\Delta_{\text{PR}}^{(0)}$ (Lemma 11, by dropping $s^{\text{eff}} \ge 1$ to 1 and bounding $\|R_{22}\|_{\infty}$ via the coupling parameter η), and on $\Delta_{\text{PPR}}^{(0)}$ (Lemma 12, by reading off the closed form). Next, the shared perturbation bound (Lemma 15) follows from a generic resolvent perturbation lemma (Lemma 14). Finally, an abstract detection criterion (Lemma 16) composes these ingredients to yield Theorem 3. The row-normalization stability lemma (Lemma 13) is also stated here but is used only in the finite-sample proof of Theorem 5.

E.3.1 Closed-form detection gaps

We begin with the closed-form detection gaps of Lemma 4. The proof proceeds in three steps: a block-lower-triangular form of the resolvent $(I - \alpha P^{(0)})^{-1}$ (Step 1), from which the standard PR (Step 2) and multi-seed PPR (Step 3) closed forms follow by direct computation.

Proof of Lemma 4. Step 1: block form of the resolvent. Since $P^{(0)}$ has zero right-top block, an induction on k shows that $(P^{(0)})^k$ inherits the same zero block for every $k \ge 0$, with top-left block $(P^{11,(0)})^k$ and bottom-right block $(P^{22,(0)})^k$: if $(P^{(0)})^k = \begin{bmatrix} (P^{11,(0)})^k & O \\ C_k & (P^{22,(0)})^k \end{bmatrix}$, then

$$(P^{(0)})^{k+1} = P^{(0)}(P^{(0)})^k = \begin{bmatrix} (P^{11,(0)})^{k+1} & O \\ P^{21,(0)}(P^{11,(0)})^k + P^{22,(0)} C_k & (P^{22,(0)})^{k+1} \end{bmatrix},$$

with $C_0 = O$. Summing the Neumann series $(I - \alpha P^{(0)})^{-1} = \sum_{k \ge 0} \alpha^k (P^{(0)})^k$ yields

$$(I - \alpha P^{(0)})^{-1} = \begin{bmatrix} R_{11} & O \\ \alpha R_{22} P^{21,(0)} R_{11} & R_{22} \end{bmatrix}, \tag{62}$$

where $R_{11} := (I - \alpha P^{11,(0)})^{-1}$ and $R_{22} := (I - \alpha P^{22,(0)})^{-1}$. The lower-left block follows either from the block-inverse identity or from $\sum_{k \ge 0} \alpha^k C_k$ via the recursion above. Since $P^{11,(0)}$, $P^{22,(0)}$

are entrywise nonnegative with $\|P^{11,(0)}\|_\infty = 1$ (row-stochastic) and $\|P^{22,(0)}\|_\infty \le 1$ and $\alpha < 1$, each Neumann series converges absolutely; hence R_{11}, R_{22} are entrywise nonnegative with finite entries.

Step 2: standard PR closed form. Writing $\pi^{(0),\intercal} = \frac{1-\alpha}{\tilde{N}}\mathbf{1}^\intercal(I - \alpha P^{(0)})^{-1}$ and splitting $\mathbf{1}^\intercal = [\mathbf{1}_N^\intercal \;\; \mathbf{1}_{\tilde{N}-N}^\intercal]$, equation (62) gives, for $i \le N$,

$$\pi_i^{(0)} \;=\; \frac{1-\alpha}{\tilde{N}}\big[\mathbf{1}_N^\intercal R_{11} + \alpha\,\mathbf{1}_{\tilde{N}-N}^\intercal R_{22}P^{21,(0)}R_{11}\big]_i \;=\; \frac{1-\alpha}{\tilde{N}}\sum_{v\le N} s_v^{\text{eff}}\,[R_{11}]_{vi}, \tag{63}$$

where $s_v^{\text{eff}} := 1 + \alpha\sum_{w>N}[R_{22}P^{21,(0)}]_{wv}$. Since $R_{22}, P^{21,(0)} \ge 0$ entrywise, $s_v^{\text{eff}} \ge 1$ for every $v \le N$. For $j > N$, the top-right block of the resolvent vanishes, so only the $\mathbf{1}_{\tilde{N}-N}^\intercal R_{22}$ contribution survives:

$$\pi_j^{(0)} \;=\; \frac{1-\alpha}{\tilde{N}}\sum_{w>N}[R_{22}]_{wj}. \tag{64}$$

Taking $\min_{i\le N}$ of (63) and $\max_{j>N}$ of (64) and subtracting yields the PR closed form in (12).

Step 3: multi-seed PPR closed form and vanishing leakage. For $v \le N$, define the single seed PPR vector $(\pi_v^{(0)})^\intercal := (1-\alpha)\,\mathbf{e}_v^\intercal(I - \alpha P^{(0)})^{-1}$. Because the top-right block of the resolvent (62) vanishes, $\pi_v^{(0)}(j) = 0$ for all $j > N$, and $\pi_v^{(0)}(i) = (1-\alpha)[R_{11}]_{vi}$ for $i \le N$. Averaging over $v \in \mathcal{S} \subset \{1,\dots,N\}$,

$$\pi_{\mathcal{S}}^{(0)}(j) = 0 \quad (j > N), \qquad \pi_{\mathcal{S}}^{(0)}(i) = \frac{1-\alpha}{|\mathcal{S}|}\sum_{v\in\mathcal{S}}[R_{11}]_{vi} \quad (i \le N). \tag{65}$$

Hence the max-leakage term in (10) vanishes identically, and $\Delta_{\text{PPR}}^{(0)} = \min_{i\le N}\pi_{\mathcal{S}}^{(0)}(i)$ takes the form displayed in (12).

□

The next two lemmas specialize the closed-form expressions (63)–(65) of Lemma 4 into forms used by the detection theorem: the relaxed PR baseline that replaces $s_v^{\text{eff}} \ge 1$ by 1 and bounds $\|R_{22}\|_\infty$ via the coupling parameter η, and the multi-seed PPR baseline that reads off the vanishing-leakage form directly.

Lemma 11 (Standard PR baseline)**.** *Let $\alpha \in (0,1)$ be a damping factor. Let $(\mathbf{q}^\alpha)^\intercal := \frac{1-\alpha}{N}\mathbf{1}_N^\intercal(I - \alpha P^{11,(0)})^{-1}$ and $q_{\min}^\alpha := \min_{i\le N} q_i^\alpha$. Then*

$$\max_{j>N}\pi_j^{(0)} \;\le\; \frac{(1-\alpha)(\tilde{N}-N)}{\tilde{N}(1-\alpha+\alpha\eta)}, \tag{66}$$

$$\min_{i\le N}\pi_i^{(0)} \;\ge\; \frac{N}{\tilde{N}}\,q_{\min}^\alpha, \tag{67}$$

and consequently

$$\Delta_{\text{PR}}^{(0)} \;\ge\; \frac{N}{\tilde{N}}\,q_{\min}^\alpha \;-\; \frac{(1-\alpha)(\tilde{N}-N)}{\tilde{N}(1-\alpha+\alpha\eta)}, \tag{68}$$

positive whenever $q_{\min}^\alpha > (\tilde{N}-N)(1-\alpha)/\big(N(1-\alpha+\alpha\eta)\big)$.

Proof. From (63) and $s_v^{\text{eff}} \ge 1$,

$$\pi_i^{(0)} \;\ge\; \frac{1-\alpha}{\tilde{N}}\big[\mathbf{1}_N^\intercal R_{11}\big]_i \;=\; \frac{N}{\tilde{N}}\,q_i^\alpha,$$

using $(\mathbf{q}^\alpha)^\intercal = \frac{1-\alpha}{N}\mathbf{1}_N^\intercal R_{11}$; taking $\min_{i\le N}$ gives (67). Since $R_{11} = \sum_{k\ge0}\alpha^k(P^{11,(0)})^k$ contains the identity term, $[\mathbf{1}_N^\intercal R_{11}]_i \ge 1$, so $q_{\min}^\alpha \ge (1-\alpha)/N$.

For the bottom-block upper bound, from (64) and $R_{22} = (I - \alpha P^{22,(0)})^{-1} = \sum_{k\ge0}\alpha^k(P^{22,(0)})^k \ge 0$ entrywise,

$$\pi_j^{(0)} = \frac{1-\alpha}{\tilde{N}}\sum_{w>N}[R_{22}]_{wj} \ge 0,$$

where row and column indices of R_{22} start at $N+1$, consistent with the full $\tilde{N} \times \tilde{N}$ indexing. Specifically, R_{22} has the size of $(\tilde{N}-N) \times (\tilde{N}-N)$ so its row and column indices are $N+1, N+2, \ldots, \tilde{N}$. Summing over $j > N$ and using $\sum_j [R_{22}]_{wj} \le \frac{1}{1-\alpha\|P^{22,(0)}\|_\infty} \le \frac{1}{1-\alpha(1-\eta)}$ (since $\|P^{22,(0)}\|_\infty \le 1-\eta$ and $R_{22} = \sum_{k\ge 0} \alpha^k (P^{22,(0)})^k$):

$$\|\pi_2^{(0)}\|_1 = \sum_{j>N} \pi_j^{(0)} \le \frac{(1-\alpha)(\tilde{N}-N)}{\tilde{N}(1-\alpha+\alpha\eta)},$$

which implies (66) since $\pi_j^{(0)} \ge 0$. Subtracting (66) from (67) gives (68).

As $\alpha \to 1$, (66) gives $\max_{j>N} \pi_j^{(0)} \to 0$, and from (67) we get $\Delta_{\text{PR}}^{(0)} \ge (N/\tilde{N}) q_{\min}^{\alpha} \to (N/\tilde{N}) \lim_{\alpha\to 1} q_{\min}^{\alpha}$. Note that $q_{\min}^{\alpha} \ge (1-\alpha)/N \to 0$, so this does not immediately give a positive lower bound; the bound is non-vacuous only when $q_{\min}^{\alpha}$ decays slower than $1-\alpha$, which occurs when $P^{11,(0)}$ is "well-mixed" (Theorem 3(i)). □

Lemma 12 (PPR baseline)**.** *For $v \le N$, define the PPR vector $(\pi_v^{(0)})^\intercal := (1-\alpha)\,\mathbf{e}_v^\intercal (I - \alpha P^{(0)})^{-1}$ with a single seed v, and for any seed set $\mathcal{S} \subset \{1,\ldots,N\}$ define $\pi_{\mathcal{S}}^{(0)} := \frac{1}{|\mathcal{S}|}\sum_{v\in\mathcal{S}} \pi_v^{(0)}$. Then $\pi_{\mathcal{S}}^{(0)}(j) = 0$ for all $j > N$, and*

$$\Delta_{\text{PPR}}^{(0)} \;=\; \min_{i\le N} \pi_{\mathcal{S}}^{(0)}(i) \;=\; \min_{i\le N} \frac{1-\alpha}{|\mathcal{S}|} \sum_{v\in\mathcal{S}} \big[(I - \alpha P^{11,(0)})^{-1}\big]_{vi}. \tag{69}$$

Proof. This is Step 3 of the proof of Lemma 4: formulas (65) read off directly, and substituting into $\Delta_{\text{PPR}}^{(0)} = \min_{i\le N} \pi_{\mathcal{S}}^{(0)}(i) - \max_{j>N} \pi_{\mathcal{S}}^{(0)}(j)$ gives (69). □

E.3.2 Perturbation bounds

The shared perturbation bound (Lemma 15) follows directly from a resolvent perturbation bound (Lemma 14 below): since $P^{(0)} = \mathcal{R}(|K^{(0)}|^\intercal)$ is obtained from P by zeroing P^{12} *and renormalizing the rows of the top block*, a direct calculation (in the proof) gives $\|P - P^{(0)}\|_\infty = 2\|P^{12}\|_\infty$, and Lemma 14 then gives the bound with sensitivity constant $2\alpha/(1-\alpha)$. The row-normalization stability lemma (Lemma 13 below) is used in the finite-sample proof of Theorem 5 to bound $\|\mathcal{R}(|K^\intercal_{\tilde{N},M}|) - \mathcal{R}(|K^\intercal_{\tilde{N}}|)\|_\infty$. Both lemmas are independent of the preference vector, so they cover PR and PPR uniformly.

We begin with row-normalization stability. The row normalization $\mathcal{R}(A)$ divides each entry by its absolute row sum, introducing a denominator that couples all entries in a row; perturbations affect both the numerator (individual entries) and the denominator (absolute row sum), leading to the factor of 2.

Lemma 13 (Stability of row normalization)**.** *Let $A, B \in \mathbb{R}^{n\times n}$ have all nonzero rows, and let $r_{\min}^A := \min_i \sum_j |A[i,j]|$. Then $\|\mathcal{R}(A) - \mathcal{R}(B)\|_\infty \le \frac{2}{r_{\min}^A}\|A - B\|_\infty$.*

Proof. Write $r_i^A = \sum_l |A[i,l]|$ and $r_i^B = \sum_l |B[i,l]|$ for the row sums. For each row i:

$$\begin{aligned}
\mathcal{R}(A)[i,j] - \mathcal{R}(B)[i,j] &= \frac{|A[i,j]|}{r_i^A} - \frac{|B[i,j]|}{r_i^B} \\
&= \frac{|A[i,j]| - |B[i,j]|}{r_i^A} + |B[i,j]|\left(\frac{1}{r_i^A} - \frac{1}{r_i^B}\right) \\
&= \frac{|A[i,j]| - |B[i,j]|}{r_i^A} + \frac{|B[i,j]|(r_i^B - r_i^A)}{r_i^A \cdot r_i^B}.
\end{aligned}$$

Summing moduli over j:

$$\begin{aligned}
\sum_j |\mathcal{R}(A)[i,j] - \mathcal{R}(B)[i,j]| &\le \frac{1}{r_i^A}\sum_j \big||A[i,j]| - |B[i,j]|\big| + \frac{|r_i^A - r_i^B|}{r_i^A \cdot r_i^B}\sum_j |B[i,j]| \\
&\le \frac{\delta_i}{r_i^A} + \frac{\delta_i}{r_i^A},
\end{aligned}$$

where $\delta_i = \sum_j |A[i,j] - B[i,j]|$ is the i-th row sum of $|A - B|$. In the second inequality, we used $\big||A[i,j]| - |B[i,j]|\big| \le |A[i,j] - B[i,j]|$ (reverse triangle inequality), the bound $|r_i^A - r_i^B| \le \delta_i$ (same argument applied to row sums), and $\sum_j |B[i,j]| = r_i^B$. Taking the maximum over i:

$$\|\mathcal{R}(A) - \mathcal{R}(B)\|_\infty \le \max_i \frac{2\delta_i}{r_i^A} \le \frac{2}{r_{\min}^A} \max_i \delta_i = \frac{2}{r_{\min}^A} \|A - B\|_\infty. \qquad \square$$

Next, the resolvent perturbation bound. The identity $(I-\alpha S')^{-1} - (I-\alpha S)^{-1} = (I-\alpha S)^{-1}\alpha(S' - S)(I - \alpha S')^{-1}$ shows that the PPR difference factors through the matrix difference $S' - S$, with the resolvent norms contributing the $\alpha/(1-\alpha)$ amplification factor.

Lemma 14 (Resolvent perturbation bound)**.** *Let S, S' be nonnegative matrices with $\|S\|_\infty \le 1$, $\|S'\|_\infty \le 1$, and $\alpha \in (0,1)$. For any $\mathbf{s} \ge 0$ with $\|\mathbf{s}\|_1 = 1$, define*

$$\mathbf{q}^\intercal := (1-\alpha)\mathbf{s}^\intercal (I - \alpha S)^{-1} \quad \textit{and} \quad (\mathbf{q}')^\intercal := (1-\alpha)\mathbf{s}^\intercal (I - \alpha S')^{-1}.$$

Then $\|\mathbf{q} - \mathbf{q}'\|_1 \le \frac{\alpha}{1-\alpha}\|S - S'\|_\infty$.

Proof. Since $\|S\|_\infty \le 1$ and $\|S'\|_\infty \le 1$ with $\alpha < 1$, the Neumann series $(I - \alpha S)^{-1} = \sum_{k \ge 0} \alpha^k S^k$ and $(I - \alpha S')^{-1} = \sum_{k \ge 0} \alpha^k (S')^k$ both converge absolutely, so $\mathbf{q}$ and $\mathbf{q}'$ are well defined. Using the resolvent identity $(I-\alpha S')^{-1} - (I-\alpha S)^{-1} = (I-\alpha S)^{-1} \cdot \alpha(S' - S) \cdot (I - \alpha S')^{-1}$, we obtain

$$(\mathbf{q}')^\intercal - \mathbf{q}^\intercal = \alpha \cdot \mathbf{q}^\intercal \cdot (S' - S) \cdot (I - \alpha S')^{-1}. \tag{70}$$

For any row vector $\mathbf{v}^\intercal \in \mathbb{R}^{1 \times n}$ and matrix $A \in \mathbb{R}^{n \times n}$, the ℓ^1-norm satisfies

$$\|\mathbf{v}^\intercal A\|_1 = \sum_j \left| \sum_i v_i A_{ij} \right| \le \sum_j \sum_i |v_i||A_{ij}| = \sum_i |v_i| \sum_j |A_{ij}| \le \|\mathbf{v}\|_1 \|A\|_\infty. \tag{71}$$

Since $\|S'\|_\infty \le 1$, we have $\|(S')^k\|_\infty \le 1$ for all $k \ge 0$, so $\|(I - \alpha S')^{-1}\|_\infty \le \sum_{k=0}^\infty \alpha^k = \frac{1}{1-\alpha}$. Similarly, $\|(I - \alpha S)^{-1}\|_\infty \le \frac{1}{1-\alpha}$, which gives

$$\|\mathbf{q}\|_1 = (1-\alpha)\|\mathbf{s}^\intercal (I - \alpha S)^{-1}\|_1 \le (1-\alpha)\|\mathbf{s}\|_1 \cdot \frac{1}{1-\alpha} = 1.$$

Applying (71) twice to (70):

$$\begin{aligned} \|\mathbf{q}' - \mathbf{q}\|_1 &= \|\alpha \cdot \mathbf{q}^\intercal (S' - S)(I - \alpha S')^{-1}\|_1 \\ &\le \alpha \cdot \|\mathbf{q}^\intercal (S' - S)\|_1 \cdot \|(I - \alpha S')^{-1}\|_\infty \\ &\le \alpha \cdot \underbrace{\|\mathbf{q}\|_1}_{\le 1} \cdot \|S' - S\|_\infty \cdot \frac{1}{1-\alpha} \;\le\; \frac{\alpha}{1-\alpha}\|S' - S\|_\infty. \end{aligned} \qquad \square$$

As a corollary, specializing Lemma 14 to $S = P^{(0)}$ and $S' = P$ yields the perturbation bound used in the end-to-end detection theorem (Theorem 5).

Lemma 15 (Perturbation bound for PR/PPR vectors on P and $P^{(0)}$)**.** *Let $P = \mathcal{R}(|K_{\tilde{N},M}^\intercal|)$ be row stochastic and $P^{(0)} := \mathcal{R}(|K_{\tilde{N},M}^{(0)|\intercal})$ be the row-stochastic matrix obtained from P by zeroing its top-right block P^{12} and renormalizing the rows in the top block. For any $\mathbf{s} \ge 0$ with $\|\mathbf{s}\|_1 = 1$, define the PR/PPR vectors $\pi_{\mathbf{s},M}^\intercal := (1-\alpha)\mathbf{s}^\intercal (I - \alpha P)^{-1}$ and $(\pi_{\mathbf{s},M}^{(0)})^\intercal := (1-\alpha)\mathbf{s}^\intercal (I - \alpha P^{(0)})^{-1}$. Then $\|P - P^{(0)}\|_\infty = 2\|P^{12}\|_\infty$, and*

$$\|\pi_{\mathbf{s},M} - \pi_{\mathbf{s},M}^{(0)}\|_1 \;\le\; \frac{2\alpha}{1-\alpha}\|P^{12}\|_\infty. \tag{72}$$

Proof. P and $P^{(0)}$ agree on rows $i > N$. For $i \le N$, write $r_i = \sum_l |K_{\tilde{N},M}^\intercal[i,l]|$, $a_{ij} = |K_{\tilde{N},M}^\intercal[i,j]|$ for $j \le N$, and $b_{ij} = |K_{\tilde{N},M}^\intercal[i,j]|$ for $j > N$. Then $P[i,j] = a_{ij}/r_i$ and $P[i,j] = b_{ij}/r_i$ on the

two blocks, while $P^{(0)}[i,j] = a_{ij}/(r_i - \sum_k b_{ik})$ for $j \le N$ and 0 for $j > N$ where $\sum_k$ ranges over $k \in \{N+1, \dots, \tilde{N}\}$. A direct calculation gives

$$\sum_{j \le N} |P[i,j] - P^{(0)}[i,j]| = \sum_{j \le N} a_{ij} \frac{\sum_k b_{ik}}{r_i(r_i - \sum_k b_{ik})} = \frac{\sum_k b_{ik}}{r_i}$$

and

$$\sum_{j > N} |P[i,j] - P^{(0)}[i,j]| = \frac{\sum_k b_{ik}}{r_i},$$

so the row-i ℓ^1 difference is $2\sum_k b_{ik}/r_i = 2\|P^{12}_{i,:}\|_1$. Maximizing over $i \le N$ gives $\|P - P^{(0)}\|_\infty = 2\|P^{12}\|_\infty$. Both P and $P^{(0)}$ are row stochastic so $\|\cdot\|_\infty \le 1$; Lemma 14 applies with $S = P$ and $S' = P^{(0)}$, giving $\|\pi_{\mathbf{s},M} - \pi^{(0)}_{\mathbf{s},M}\|_1 \le \frac{\alpha}{1-\alpha}\|P - P^{(0)}\|_\infty = \frac{2\alpha}{1-\alpha}\|P^{12}\|_\infty$. □

E.4 Proof of detection theorem

Now using the perturbation bound in Lemma 15, we obtain the following abstract detection criterion.

Lemma 16 (Abstract detection criterion)**.** *Let $\pi_{\mathbf{s},M}$ and $\pi^{(0)}_{\mathbf{s},M}$ be as in Lemma 15, with gap $\Delta_{\mathbf{s}} = \min_{i \le N} \pi^{(0)}_{\mathbf{s},M}(i) - \max_{j > N} \pi^{(0)}_{\mathbf{s},M}(j)$. If $\Delta_{\mathbf{s}} > 0$ and*

$$\|P^{12}\|_\infty \;<\; \frac{1-\alpha}{4\alpha} \cdot \Delta_{\mathbf{s}}, \tag{73}$$

then $\min_{i \le N} \pi_{\mathbf{s},M}(i) > \max_{j > N} \pi_{\mathbf{s},M}(j)$, so Algorithm 1 (using $\pi_{\mathbf{s},M}$) correctly identifies $\{1, \dots, N\}$ as the top-N coordinates. In particular, this holds for standard PR ($\mathbf{s} = \mathbf{1}/\tilde{N}$, $\Delta_{\mathbf{s}} = \Delta^{(0)}_{\text{PR}}$) and multi-seed PPR ($\mathbf{s} = |\mathcal{S}|^{-1}\sum_{v \in \mathcal{S}} \mathbf{e}_v$ with $\mathcal{S} \subset \{1, \dots, N\}$, $\Delta_{\mathbf{s}} = \Delta^{(0)}_{\text{PPR}}$).

Proof. Algorithm 1 correctly identifies $\{1, \dots, N\}$ as the top-N coordinates when $\min_{i \le N} \pi_{\mathbf{s},M}(i) > \max_{j > N} \pi_{\mathbf{s},M}(j)$. Since

$$\pi_{\mathbf{s},M}(i) \ge \pi^{(0)}_{\mathbf{s},M}(i) - \|\pi_{\mathbf{s},M} - \pi^{(0)}_{\mathbf{s},M}\|_\infty, \qquad \pi_{\mathbf{s},M}(j) \le \pi^{(0)}_{\mathbf{s},M}(j) + \|\pi_{\mathbf{s},M} - \pi^{(0)}_{\mathbf{s},M}\|_\infty,$$

it is sufficient to have $\|\pi_{\mathbf{s},M} - \pi^{(0)}_{\mathbf{s},M}\|_\infty < \Delta_{\mathbf{s}}/2$, and a fortiori $\|\pi_{\mathbf{s},M} - \pi^{(0)}_{\mathbf{s},M}\|_1 < \Delta_{\mathbf{s}}/2$. By Lemma 15,

$$\|\pi_{\mathbf{s},M} - \pi^{(0)}_{\mathbf{s},M}\|_1 \le \frac{2\alpha}{1-\alpha}\|P^{12}\|_\infty,$$

so the sufficient condition $\|\pi_{\mathbf{s},M} - \pi^{(0)}_{\mathbf{s},M}\|_1 < \Delta_{\mathbf{s}}/2$ is implied by (73). Standard PR and multi-seed PPR follow by linearity of $\mathbf{s} \mapsto \pi_{\mathbf{s},M}$. □

With Lemma 16 in hand, the detection theorem follows by substituting the closed-form gap lower bounds from the preceding subsection.

Proof of Theorem 3. Following the main-text sketch (§3.3), Lemma 16 reduces detection on P to a sufficient condition on $\|P^{12}\|_\infty$ in terms of the auxiliary gap on $P^{(0)}$: for either method,

$$\Delta_{\text{method}} > 0 \;\Longleftarrow\; \|P^{12}\|_\infty \;<\; \frac{1-\alpha}{4\alpha}\,\Delta^{(0)}_{\text{method}}. \tag{74}$$

It remains to substitute explicit lower bounds on $\Delta^{(0)}_{\text{method}}$ to derive the conditions stated in Theorem 3.

Part (i): Standard PR. Under the mixing condition—equivalent to positivity of the bracket in (8)—Lemma 11 gives the explicit lower bound

$$\Delta^{(0)}_{\text{PR}} \;\ge\; \frac{N}{\tilde{N}}\left[q^{\alpha}_{\min} - \frac{(\tilde{N}-N)(1-\alpha)}{N(1-\alpha+\alpha\eta)}\right] \;>\; 0,$$

so substituting into (74) yields (8).

Part (ii): Multi-seed PPR. Under the reachability hypothesis, Lemma 12 gives the closed-form lower bound

$$\Delta_{\text{PPR}}^{(0)} \;=\; \min_{i\le N} \frac{1-\alpha}{|\mathcal{S}|} \sum_{v\in\mathcal{S}} \big[(I-\alpha P^{11,(0)})^{-1}\big]_{vi} \;>\; 0,$$

so substituting into (74) yields (9).

Invariant subspace case. When $\mathcal{F}_N$ is $\mathcal{K}$-invariant, Theorem 2 gives the bottom-left block $K^{21}_{\tilde{N},M} = O$, hence $P^{12} = O$, so (9) holds trivially. By Lemma 15, $\pi_{\mathcal{S},M} = \pi^{(0)}_{\mathcal{S},M}$, and by Lemma 12, $\pi^{(0)}_{\mathcal{S},M}(j) = 0$ for all $j > N$; hence $\pi_{\mathcal{S},M}(j) = 0$ for all $j > N$. □

F Finite-sample extensions

Goal. This subsection states and proves the EDMD concentration bound (Proposition 17) and proves the end-to-end PR/PPR perturbation bound (Theorem 5). The PR/PPR sample complexity (Corollary 5.1) and the finite-sample leakage bound (Corollary 18.2) are proved in §F.2 below, combining Proposition 17 with the row-normalization and PPR perturbation lemmas of Appendix E.3.

Strategy. A single triangle inequality organizes the argument:

$$\|\pi_M - \pi^{(0)}_{\text{pop}}\|_1 \;\le\; \underbrace{\|\pi_M - \pi_{\text{pop}}\|_1}_{\text{finite-sample error}} + \underbrace{\|\pi_{\text{pop}} - \pi^{(0)}_{\text{pop}}\|_1}_{\text{population approximation error}},$$

where π_{pop} is the PR (or PPR) vector of the row-normalized population-level matrix $P_{\text{pop}} = \mathcal{R}(|K^{\intercal}_{\tilde{N}}|)$. The population term is controlled by $\varepsilon_0 = \|K^{\intercal}_{\tilde{N}} - (K^{(0)}_{\tilde{N}})^{\intercal}\|_\infty$ through the shared perturbation bound (Lemma 15). The finite-sample term is controlled by $\varepsilon_M = \|K^{\intercal}_{\tilde{N},M} - K^{\intercal}_{\tilde{N}}\|_\infty$, which we obtain from matrix Bernstein (Proposition 17) and then plug into the same two-step composition (row-normalization stability + resolvent perturbation) used in the population case. The sample complexity and leakage corollaries then follow by inverting $\varepsilon_M = O(1/\sqrt{M})$ against the relevant gap.

We begin by stating the EDMD concentration bound, which supplies ε_M via matrix Bernstein applied to the Gram and cross-covariance matrices. The boundedness assumption (Assumption 3) activates the concentration machinery.

Relation to existing finite-sample EDMD bounds. Concentration-type finite-sample guarantees for EDMD have been developed in several complementary settings. Korda and Mezić [15] established almost-sure convergence of $K_{N,M}$ to the L^2-projected Koopman operator as $M \to \infty$, without an explicit rate. Quantitative rates appeared first for the Gram/cross-covariance blocks under sub-Gaussian and boundedness hypotheses: Llamazares-Elias et al. [64], Nüske et al. [17], Kostic et al. [16] give operator-norm bounds of order $1/\sqrt{M}$ with explicit constants. The proposition below specializes such arguments to the infinity-norm quantity required by the row-normalization and PPR perturbation machinery of this paper.

Proposition 17 (EDMD concentration)**.** *Let $\mathcal{D}_{\tilde{N}}$ be a linearly independent set satisfying Assumption 3 with constant $D > 0$. Let $\lambda_{\min}$ be the minimum of the moduli of the eigenvalues of the Gram matrix $G_{\tilde{N}} := [\langle \psi_i, \psi_j\rangle_{L^2(\mu)}]_{i,j=1}^{\tilde{N}} \in \mathbb{C}^{\tilde{N}\times\tilde{N}}$. Let $\rho \in (0, 1/2)$. Then for $M \ge \frac{32\,\tilde{N}^2 D^4}{\lambda_{\min}^2} \log\frac{2\tilde{N}}{\rho}$, with probability $\ge 1 - 2\rho$:*

$$\|K^{\intercal}_{\tilde{N},M} - K^{\intercal}_{\tilde{N}}\|_\infty \le \varepsilon_M \tag{75}$$

where

$$\varepsilon_M = C_{EDMD}\sqrt{2\log(2\tilde{N}/\rho)/M} \quad \textit{and} \quad C_{EDMD} = \frac{4\tilde{N}^{3/2}D^2}{\lambda_{\min}}\left(1 + \frac{\|G_{\tilde{N}}\|_2}{\lambda_{\min}}\right). \tag{76}$$

Proof. Under Assumption 3, Proposition 5.6 of [64] gives, with probability $\ge 1 - 2\rho$,

$$\|K_{\tilde{N},M} - K_{\tilde{N}}\|_2 \le \frac{4\tilde{N}D^2}{\lambda_{\min}}\left(1 + \frac{\|G_{\tilde{N}}\|_2}{\lambda_{\min}}\right)\sqrt{\frac{2\log(2\tilde{N}/\rho)}{M}}.$$

Since $\|B^{\intercal}\|_\infty \le \sqrt{\tilde{N}}\|B\|_2$ for any $\tilde{N}\times\tilde{N}$ matrix B, this yields (75) with $C_{\text{EDMD}} = 4\tilde{N}^{3/2}D^2(1 + \|G_{\tilde{N}}\|_2/\lambda_{\min})/\lambda_{\min}$. □

F.1 Concentration under dependent sampling

Proposition 17 is the only ingredient of the finite-sample analysis that uses independence. It also covers (S1) with stochastic dynamics, since the pairs are then still i.i.d. with entries bounded by D^2. We record here its analogues in regimes (S2) and (S3). Because Theorem 5 depends on the data only through ε_M, its perturbation bound extends to these regimes once ε_M is supplied. The same substitution in the argument leading to Corollary 5.1 yields the corresponding sample-complexity bounds. For trajectory data, exponential mixing replaces independence. Under (S3) of Assumption 2 it is part of the assumption. Under (S2), time averages converge but without a rate, so for deterministic trajectories we use the variant (S2′) below, which assumes an exponential concentration inequality. The concentration inequalities we need follow from (S2′) and (S3), as we now record.

Young towers. We use the definition of [65, §3 and §7.1], which follows [66]. A *uniform Young tower* is a space Δ with a map $T : \Delta \to \Delta$ such that: (i) Δ is partitioned into sets $\Delta_{a,\ell}$, $a \in \mathbb{N}$, $0 \le \ell < \phi(a)$, for an integer-valued return time ϕ, and T maps $\Delta_{a,\ell}$ bijectively onto $\Delta_{a,\ell+1}$ if $\ell < \phi(a) - 1$, and $\Delta_{a,\phi(a)-1}$ onto the base $\Delta_0 := \bigcup_a \Delta_{a,0}$; (ii) the distance is $d_\Delta(x, y) = \beta^{s(x,y)}$ for some $\beta < 1$, where the separation time $s(x, y)$ is the first n such that $T^n x$ and $T^n y$ lie in different elements of the partition; (iii) there is a T-invariant probability measure $\hat{\mu}$ whose inverse Jacobian g satisfies $|g(x)/g(y) - 1| \le C\, d_\Delta(Tx, Ty)$ for x, y in the same element of the partition; (iv) $\gcd\{\phi(a) : a \in \mathbb{N}\} = 1$. The tower has *exponential tails* if $\int_{\Delta_0} e^{c_0 \phi}\, d\hat{\mu} < \infty$ for some $c_0 > 0$. A map $\mathbf{F}$ on a bounded set $\mathcal{M}$ with invariant measure μ is *modelled by a uniform Young tower with exponential tails* if there are such a tower (for invertible $\mathbf{F}$, its natural extension, with the bilateral separation time) and a projection $\pi : \Delta \to \mathcal{M}$ with $\mathbf{F} \circ \pi = \pi \circ T$, $\mu = \hat{\mu} \circ \pi^{-1}$, and $\|\pi x - \pi y\| \le d_\Delta(x, y)$ for x, y in the same element of the partition. The measure μ is then the SRB (Sinai–Ruelle–Bowen) measure of $\mathbf{F}$, that is, an invariant measure whose conditional measures on unstable manifolds are absolutely continuous with respect to the volume on these manifolds [67]. Examples include Axiom A attractors, Hénon attractors for Benedicks–Carleson parameters, and some billiards with convex scatterers [65, §7.1].

(S2′) *Exponentially concentrating deterministic trajectory*: $\mathbf{x}^{(i)} = \mathbf{F}^{\,i-1}(\mathbf{x}^{(1)})$ and $\mathbf{y}^{(i)} = \mathbf{x}^{(i+1)}$, where μ is $\mathbf{F}$-invariant, $\mathbf{x}^{(1)} \sim \nu$ with $\kappa := \|d\nu/d\mu\|_\infty < \infty$, each ψ_i is Lipschitz with constant at most L_ψ, and the process $(\mathbf{x}, \mathbf{F}\mathbf{x}, \mathbf{F}^2\mathbf{x}, \dots)$ with $\mathbf{x} \sim \mu$ satisfies the *exponential concentration inequality* [65, Eq. (1.1)]: there is a constant $C_{\mathbf{F}} < \infty$ such that for every $n \ge 1$ and every separately Lipschitz $K : \mathcal{M}^n \to \mathbb{R}$, with $\mathrm{Lip}_j(K) := \sup |K(\dots, \mathbf{x}_j, \dots) - K(\dots, \mathbf{x}'_j, \dots)|/\|\mathbf{x}_j - \mathbf{x}'_j\|$ (the supremum over all arguments, with only the j-th one changed),

$$\mathbb{E}_{\mathbf{x}\sim\mu}\Big[e^{K(\mathbf{x},\mathbf{F}\mathbf{x},\dots,\mathbf{F}^{n-1}\mathbf{x})-\mathbb{E}_\mu K}\Big] \;\le\; \exp\Big(C_{\mathbf{F}} \sum_{j=1}^{n} \mathrm{Lip}_j(K)^2\Big). \tag{77}$$

By [65, Thm. 7.1], (77) holds whenever $(\mathcal{M}, \mathbf{F}, \mu)$ is modelled by a uniform Young tower with exponential tails. Theorem 5 also holds under (S2′), with $\varepsilon_M = \varepsilon_M^{\mathrm{mix}}$ and M as in Proposition 18; its proof below covers this case.

Concentration from (S2′) and (S3). Under (S2′), by Markov's inequality applied to $e^{\pm\lambda(K-\mathbb{E}_\mu K)}$, $\mu(|K - \mathbb{E}_\mu K| > t) \le 2\exp(-t^2/(4C_{\mathbf{F}} \sum_j \mathrm{Lip}_j(K)^2))$, and the same bound holds with $\mathbf{x}^{(1)} \sim \nu$ after multiplying by κ. Under (S3), let $t_{\mathrm{mix}} := \min\{k \ge 1 : d(k) \le 1/4\}$. McDiarmid's inequality for Markov chains [68, Cor. 2.11], which holds for any initial distribution, gives for every $K : \mathcal{M}^n \to \mathbb{R}$ with $|K(\dots, \mathbf{x}_j, \dots) - K(\dots, \mathbf{x}'_j, \dots)| \le c_j$ and all $t > 0$

$$\mathbb{P}\big(|K(\mathbf{x}_0, \dots, \mathbf{x}_{n-1}) - \mathbb{E}K(\mathbf{x}_0, \dots, \mathbf{x}_{n-1})| \ge t\big) \;\le\; 2\exp\Big(-\frac{2t^2}{9\, t_{\mathrm{mix}} \|c\|_2^2}\Big), \tag{78}$$

where we used $\tau_{\min} \le 9\, t_{\mathrm{mix}}$ for the mixing parameter $\tau_{\min}$ of [68] (take $\varepsilon = 1/2$ there; laws from two initial points are within $2d(k)$). Moreover, $|\mathbb{E}f(\mathbf{x}_k, \mathbf{x}_{k+1}) - \mathbb{E}_{\mathbf{x}\sim\mu,\, \mathbf{y}\sim p(\mathbf{x},\cdot)} f(\mathbf{x}, \mathbf{y})| \le 2\|f\|_\infty d(k)$ for bounded f, and $\sum_{k\ge 0} d(k) \le 2t_{\mathrm{mix}}$ since $d(\ell\, t_{\mathrm{mix}}) \le 2^{-\ell}$. We define the *mixing constant*

$$\tau_{\mathrm{mix}} := C_{\mathbf{F}}\, L_\psi^2/D^2 \;\text{ under (S2′)}, \qquad \tau_{\mathrm{mix}} := 9\, t_{\mathrm{mix}} \;\text{ and }\; \kappa := 1 \;\text{ under (S3)}. \tag{79}$$

Proposition 18 (EDMD concentration under mixing)**.** *Let $\mathcal{D}_{\tilde{N}}$ satisfy Assumption 3 with constant D, let $\lambda_{\min}$ and C_{EDMD} be as in Proposition 17, and let $\rho \in (0, 1/2)$. Under (S2′), or (S3) of*

Assumption 2, with τ_{mix} and κ as in (79)*, there are absolute constants c_1, c_2 such that, for $M \geq c_1 \tau_{\text{mix}} \tilde{N}^2 D^4 \lambda_{\min}^{-2} \log(2\kappa\tilde{N}/\rho)$, with probability $\geq 1 - 2\rho$,*

$$\|K^{\mathsf{T}}_{\tilde{N},M} - K^{\mathsf{T}}_{\tilde{N}}\|_\infty \;\leq\; \varepsilon_M^{\text{mix}} \coloneqq c_2\, C_{\text{EDMD}} \sqrt{\frac{2\tau_{\text{mix}}}{M} \log \frac{2\kappa\tilde{N}}{\rho}}.$$

This is the bound of Proposition 17 with the effective sample size M/τ_{mix} in place of M.

Proof. Each entry of the empirical Gram and cross-Gram matrices $\widehat{G}_M, \widehat{A}_M$ is an average of bounded observables of the pair process $\mathbf{z}_k = (\mathbf{x}_k, \mathbf{x}_{k+1})$, with absolute value at most D^2. Under (S3), each state enters at most two summands, so each entry average has bounded differences $c_j \leq 4D^2/M$ and $\|c\|_2^2 \leq 32D^4/M$. Applying (78) to the real and imaginary parts gives deviations from the expectation larger than $t/2$ with probability at most $4\exp[-Mt^2/(64\,\tau_{\text{mix}} D^4)]$, and by the bias bound above the expectation is within $4D^2 t_{\text{mix}}/M \leq t/2$ of the population value under the sample-size condition (after enlarging c_1). Under (S2′), each entry average is a separately Lipschitz functional of $M+1$ successive states, with coordinate Lipschitz constants at most $2DL_\psi/M$. Their squared sum is at most $8D^2L_\psi^2/M$, so (77) gives entrywise tails bounded by $4\kappa\exp[-Mt^2/(32\,C_{\mathbf{F}} D^2 L_\psi^2)] = 4\kappa\exp[-Mt^2/(32\,\tau_{\text{mix}} D^4)]$. Taking a union bound over the $2\tilde{N}^2$ entries, applying $\|B\|_2 \leq \tilde{N}\max_{i,j}|B_{ij}|$, and using the Gram-inversion perturbation bound yields the stated estimate after converting to $\|\cdot^{\mathsf{T}}\|_\infty$ at a cost of $\sqrt{\tilde{N}}$. The sample-size condition ensures that $\widehat{G}_M$ is invertible. □

Remark 3 (Status of the experiments)**.** The Ramachandran system of Appendix D.5 is an overdamped Langevin diffusion on the compact torus $[-\pi, \pi]^2$ with smooth, everywhere-positive transition density, hence a uniformly ergodic Markov chain. The simulated Euler–Maruyama chain is uniformly ergodic as well, since its transition density is bounded below on the torus, so (S3) holds from any initial point. Applying Theorem 5 to this experiment additionally requires compatibility of the dictionary and regularized EDMD estimator with the stated assumptions. Alternatively, under their stated assumptions, the ergodic-sampling bounds of Nüske et al. [17, Thm. 14] for unregularized EDMD can supply ε_M after converting their Frobenius-norm bound to $\|\cdot^{\mathsf{T}}\|_\infty$, at the cost of polynomial rather than exponential tails. For the Lorenz system of Appendix D.6 the Koopman operator of the measure-preserving flow is an $L^2(\mu)$-isometry, so Markov-chain spectral-gap arguments do not apply. Theorem 2 is an exact finite-sample statement under its stated assumptions. The Lorenz data come from a single initial condition after a burn-in. The classical Lorenz attractor exists [69] and carries an SRB measure that is physical [50], so (S2) holds for Lebesgue-a.e. initial condition in its basin of attraction and Lemma 7 applies. Theorems 2 and 3 also apply: they concern the computed matrix and use no sampling assumption. Our finite-sample theorem would require (S2′) for the time-Δt map of the Lorenz flow. We do not know whether (77) holds for this map, for instance through a Young tower with exponential tails, although the classical Lorenz attractor is robustly exponentially mixing [70], so we claim no finite-sample guarantee for the Lorenz experiment.

Plugging the EDMD bound into the two-step (row-normalization + resolvent) composition yields the end-to-end perturbation bound.

Proof of Theorem 5. By the triangle inequality:

$$\|\pi_M - \pi_{\text{pop}}^{(0)}\|_1 \leq \|\pi_M - \pi_{\text{pop}}\|_1 + \|\pi_{\text{pop}} - \pi_{\text{pop}}^{(0)}\|_1.$$

Population approximation error. By Lemma 15 applied to P_{pop},

$$\|\pi_{\text{pop}} - \pi_{\text{pop}}^{(0)}\|_1 \leq \frac{2\alpha}{1-\alpha}\|P_{\text{pop}}^{12}\|_\infty \leq \frac{2\alpha\,\varepsilon_0}{(1-\alpha)\,r_{\min}^{(0)}}, \tag{80}$$

using $\|P_{\text{pop}}^{12}\|_\infty \leq \varepsilon_0/r_{\min}^{(0)}$. *Finite-sample error.* By Proposition 17 under (S1), or Proposition 18 under (S2′) or (S3), on an event of probability at least $1-2\rho$: $\|K^{\mathsf{T}}_{\tilde{N},M} - K^{\mathsf{T}}_{\tilde{N}}\|_\infty \leq \varepsilon_M$. To apply Lemma 13 with tight constants, we use $K^{\mathsf{T}}_{\tilde{N}}$ as the reference matrix and verify that its minimum

absolute row sum (equivalently, the minimum absolute column sum of $|K_{\tilde{N}}|$) is at least $r_{\min}^{(0)}$. For columns $i \le N$, the block structure of $K_{\tilde{N}}^{(0)}$ gives

$$\sum_j |K_{\tilde{N}}^{\intercal}[i,j]| = \sum_j |K_{\tilde{N}}[j,i]| \ge \sum_j |K_{\tilde{N}}^{(0)}[j,i]| = \sum_j |(K_{\tilde{N}}^{(0)})^{\intercal}[i,j]| \ge r_{\min}^{(0)},$$

since $K_{\tilde{N}}$ and $K_{\tilde{N}}^{(0)}$ agree outside the bottom-left block and any additional entries only increase the absolute column sum. For columns $i > N$, $K_{\tilde{N}}$ and $K_{\tilde{N}}^{(0)}$ agree on column i, so $\sum_j |K_{\tilde{N}}^{\intercal}[i,j]| = \sum_j |(K_{\tilde{N}}^{(0)})^{\intercal}[i,j]| \ge r_{\min}^{(0)}$. Therefore the minimum absolute row sum of $|K_{\tilde{N}}^{\intercal}|$ is at least $r_{\min}^{(0)}$. Applying Lemma 13 with $A = K_{\tilde{N}}^{\intercal}$ and $B = K_{\tilde{N},M}^{\intercal}$ (so $r_{\min}^{A} \ge r_{\min}^{(0)}$):

$$\|\mathcal{R}(|K_{\tilde{N},M}^{\intercal}|) - \mathcal{R}(|K_{\tilde{N}}^{\intercal}|)\|_\infty \le \frac{2}{r_{\min}^{(0)}}\,\varepsilon_M.$$

Since $P = \mathcal{R}(|K_{\tilde{N},M}^{\intercal}|)$ and $P_{\text{pop}} = \mathcal{R}(|K_{\tilde{N}}^{\intercal}|)$ are both row stochastic (hence $\|\cdot\|_\infty = 1$), Lemma 14 applies with $S = P$ and $S' = P_{\text{pop}}$:

$$\|\pi_M - \pi_{\text{pop}}\|_1 \;\le\; \frac{\alpha}{1-\alpha}\|P - P_{\text{pop}}\|_\infty \;\le\; \frac{2\alpha\,\varepsilon_M}{(1-\alpha)\,r_{\min}^{(0)}}. \tag{81}$$

Combining (80) and (81):

$$\|\pi_M - \pi_{\text{pop}}^{(0)}\|_1 \le \frac{2\alpha\,\varepsilon_0}{(1-\alpha)\,r_{\min}^{(0)}} + \frac{2\alpha\,\varepsilon_M}{(1-\alpha)\,r_{\min}^{(0)}} = \frac{2\alpha(\varepsilon_0+\varepsilon_M)}{(1-\alpha)\,r_{\min}^{(0)}},$$

which is the desired bound

$$\|\pi_M - \pi_{\text{pop}}^{(0)}\|_1 \;\le\; \frac{2\alpha(\varepsilon_0+\varepsilon_M)}{(1-\alpha)\,r_{\min}^{(0)}}. \tag{82}$$

□

The sample-complexity corollary (Corollary 5.1) follows by requiring the right-hand side of (82) to be smaller than half the PPR detection gap and inverting $\varepsilon_M = O(1/\sqrt{M})$; its proof, the analogous PR bound (Corollary 18.1), and the finite-sample leakage bound (Corollary 18.2) are given in §F.2 below.

F.2 Sample complexity and finite-sample leakage

Building on Theorem 5, we now prove the three finite-sample corollaries: the PPR sample complexity (Corollary 5.1, stated in Section 3.4), the PR sample complexity (Corollary 18.1), and the finite-sample leakage guarantee (Corollary 18.2, the population analogue of Proposition 10). For both sample-complexity corollaries, requiring the total perturbation $\|\pi_M - \pi_{\text{pop}}^{(0)}\|_1 < \Delta^{(0)}/2$ (where $\Delta^{(0)}$ is the relevant auxiliary gap) preserves the top-N ordering; the general bound (85) is derived once in the PPR proof and then specialized to PR. For Corollary 18.2, ε_M is combined with the population leakage bound via the triangle inequality applied to the PPR mass on S_N.

Proof of Corollary 5.1. Let $\pi_{\text{pop}}^{(0)}$ be the PPR vector of $P_{\text{pop}}^{(0)} = \mathcal{R}(|K_{\tilde{N}}^{(0)}|^{\intercal})$ and π_M be the PPR vector of $P = \mathcal{R}(|K_{\tilde{N},M}^{\intercal}|)$.

Step 1: gap-based identification criterion. On the event $\{\|\pi_M - \pi_{\text{pop}}^{(0)}\|_1 < \Delta_{\text{PPR}}^{(0)}/2\}$, for any $i \le N$ and $j > N$,

$$(\pi_M)_i \;\ge\; (\pi_{\text{pop}}^{(0)})_i - \tfrac{\Delta_{\text{PPR}}^{(0)}}{2}, \qquad (\pi_M)_j \;\le\; (\pi_{\text{pop}}^{(0)})_j + \tfrac{\Delta_{\text{PPR}}^{(0)}}{2},$$

since $|(\pi_M)_k - (\pi_{\text{pop}}^{(0)})_k| \le \|\pi_M - \pi_{\text{pop}}^{(0)}\|_1 < \Delta_{\text{PPR}}^{(0)}/2$ for every coordinate k. By Lemma 12, $(\pi_{\text{pop}}^{(0)})_i \ge \Delta_{\text{PPR}}^{(0)}$ for all $i \le N$ and $(\pi_{\text{pop}}^{(0)})_j = 0$ for all $j > N$, so $(\pi_M)_i - (\pi_M)_j > 0$ for every such pair, i.e. $\Delta_{\text{PPR}} > 0$ (detection success).

Step 2: bound $\|\pi_M - \pi^{(0)}_{\mathrm{pop}}\|_1$ in terms of $\|P^{12}_{\mathrm{pop}}\|_\infty$ and ε_M. Provided $M \geq 32\,\tilde{N}^2 D^4 \log(2\tilde{N}/\rho)/\lambda^2_{\min}$, Theorem 5 applies and gives, with probability $\geq 1-2\rho$:

$$\|\pi_M - \pi^{(0)}_{\mathrm{pop}}\|_1 \;\leq\; \frac{2\alpha}{1-\alpha}\,\|P^{12}_{\mathrm{pop}}\|_\infty + \frac{2\alpha\,\varepsilon_M}{(1-\alpha)\,r^{(0)}_{\min}}. \tag{83}$$

Since $r_{\max} \geq r^{(0)}_{\min}$, the first term satisfies $\frac{2\alpha}{1-\alpha}\|P^{12}_{\mathrm{pop}}\|_\infty \leq \frac{2\alpha\,r_{\max}}{(1-\alpha)\,r^{(0)}_{\min}}\|P^{12}_{\mathrm{pop}}\|_\infty$, so we may combine:

$$\|\pi_M - \pi^{(0)}_{\mathrm{pop}}\|_1 \;\leq\; \frac{2\alpha}{(1-\alpha)r^{(0)}_{\min}}\big(r_{\max}\|P^{12}_{\mathrm{pop}}\|_\infty + \varepsilon_M\big). \tag{84}$$

Step 3: invert against the gap. Requiring (84) $< \Delta^{(0)}_{\mathrm{PPR}}/2$, substituting $\varepsilon_M = C_{\mathrm{EDMD}}\sqrt{2\log(2\tilde{N}/\rho)/M}$ from (76), squaring, and solving for M yields

$$M \;\geq\; \frac{32\,\alpha^2\log(2\tilde{N}/\rho)\,C^2_{\mathrm{EDMD}}}{(1-\alpha)^2(r^{(0)}_{\min})^2\,\big(\Delta^{(0)}_{\mathrm{PPR}} - 4\alpha r_{\max}\|P^{12}_{\mathrm{pop}}\|_\infty/((1-\alpha)r^{(0)}_{\min})\big)^2}. \tag{85}$$

Setting $\alpha = 1/2$ so that $\alpha^2/(1-\alpha)^2 = 1$ and $4\alpha/((1-\alpha)r^{(0)}_{\min}) = 4/r^{(0)}_{\min}$ gives the second argument of the max in (14); combining with the Step 2 requirement $M \geq 32\,\tilde{N}^2 D^4\log(2\tilde{N}/\rho)/\lambda^2_{\min}$ yields the full bound. □

Corollary 18.1 (Sample complexity for PR detection)**.** *Let $\Delta^{(0)}_{\mathrm{PR}}$ be the PR auxiliary gap of Section 3.3 computed from $P^{(0)}_{\mathrm{pop}}$. If $\Delta^{(0)}_{\mathrm{PR}} > 0$ and $\|P^{12}_{\mathrm{pop}}\|_\infty < (1-\alpha)r^{(0)}_{\min}\Delta^{(0)}_{\mathrm{PR}}/(4\alpha r_{\max})$, detection ($\Delta_{\mathrm{PR}} > 0$) holds with probability $\geq 1-2\rho$ for*

$$M \;\geq\; \max\left\{\frac{32\,\tilde{N}^2 D^4}{\lambda^2_{\min}}\log\frac{2\tilde{N}}{\rho},\;\; \frac{32\,\alpha^2\log(2\tilde{N}/\rho)\,C^2_{EDMD}}{(1-\alpha)^2(r^{(0)}_{\min})^2\big(\Delta^{(0)}_{\mathrm{PR}} - 4\alpha r_{\max}\|P^{12}_{\mathrm{pop}}\|_\infty/((1-\alpha)r^{(0)}_{\min})\big)^2}\right\}. \tag{86}$$

Compared with Corollary 5.1, this bound carries an additional $\alpha^2/(1-\alpha)^2$ prefactor. Since $\Delta^{(0)}_{\mathrm{PR}} > 0$ requires the mixing condition of Theorem 3(i), which may force α close to 1, it is not always possible to set $\alpha = 1/2$. Therefore, the prefactor $\alpha^2/(1-\alpha)^2$ can be significantly larger than 1.

Proof. Specialize (85) (derived in the proof above) to $\pi^{(0)}_{\mathrm{pop}}$ the standard PR vector and $\Delta = \Delta^{(0)}_{\mathrm{PR}}$, giving (86). □

Corollary 18.2 (Finite-sample leakage)**.** *Let $\Lambda^{\gamma,\mathrm{pop}}_{\mathbf{s}}(S_N)$ denote the leakage defined with $K_{\tilde{N}} := \lim_{M\to\infty} K_{\tilde{N},M}$ as in (61) in place of $K_{\tilde{N},M}$. For any $\mathbf{s}$ supported on S_N and any $M \geq \frac{32\,\tilde{N}^2D^4}{\lambda^2_{\min}}\log\frac{2\tilde{N}}{\rho}$, with probability $\geq 1-2\rho$:*

$$\Lambda^{\gamma,\mathrm{pop}}_{\mathbf{s}}(S_N) := (1-\gamma)\sum_{k=1}^{\infty}\gamma^k\sum_{j\notin S_N}[\mathbf{s}^\intercal|K^\intercal_{\tilde{N}}|^k]_j \leq \frac{1-\gamma}{1-\alpha}\left(1 - \pi_{\mathbf{s},M}(S_N) + \frac{2\alpha\varepsilon_M}{(1-\alpha)r^{(0)}_{\min}}\right). \tag{87}$$

Proof. Let $P_{\mathrm{pop}} = \mathcal{R}(|K^\intercal_{\tilde{N}}|)$ and $P = \mathcal{R}(|K^\intercal_{\tilde{N},M}|)$ denote the population and empirical row-normalized matrices, and let $\pi_{\mathbf{s}}$, $\pi_{\mathbf{s},M}$ denote their respective PPR vectors with preference $\mathbf{s}$ and damping α.

Step 1: population leakage bound. The proof of Proposition 10 applies verbatim with $K_{\tilde{N},M}$ replaced by $K_{\tilde{N}}$ and $\pi_{\mathbf{s},M}$ replaced by $\pi_{\mathbf{s}}$, yielding

$$\Lambda^{\gamma,\mathrm{pop}}_{\mathbf{s}}(S_N) \;\leq\; \frac{1-\gamma}{1-\alpha}\big(1-\pi_{\mathbf{s}}(S_N)\big), \tag{88}$$

where $\pi_{\mathbf{s}}$ is the PPR vector of P_{pop}.

Step 2: bound $\pi_{\mathbf{s}}(S_N) - \pi_{\mathbf{s},M}(S_N)$. By Proposition 17, $\|K^{\mathsf{T}}_{\tilde{N},M} - K^{\mathsf{T}}_{\tilde{N}}\|_\infty \le \varepsilon_M$ with probability $\ge 1-2\rho$. Lemma 13 with $A = K^{\mathsf{T}}_{\tilde{N}}$ and $B = K^{\mathsf{T}}_{\tilde{N},M}$ gives $\|P - P_{\text{pop}}\|_\infty \le (2/r^{(0)}_{\min})\varepsilon_M$. Lemma 14 then yields

$$\|\pi_{\mathbf{s}} - \pi_{\mathbf{s},M}\|_1 \;\le\; \frac{2\alpha\,\varepsilon_M}{(1-\alpha)\,r^{(0)}_{\min}}, \tag{89}$$

and since $|\pi_{\mathbf{s}}(S_N) - \pi_{\mathbf{s},M}(S_N)| \le \|\pi_{\mathbf{s}} - \pi_{\mathbf{s},M}\|_1$,

$$\pi_{\mathbf{s}}(S_N) \;\ge\; \pi_{\mathbf{s},M}(S_N) - \frac{2\alpha\,\varepsilon_M}{(1-\alpha)\,r^{(0)}_{\min}}. \tag{90}$$

Step 3: combine. Substituting (90) into (88):

$$\Lambda^{\gamma,\text{pop}}_{\mathbf{s}}(S_N) \;\le\; \frac{1-\gamma}{1-\alpha}\Big(1 - \pi_{\mathbf{s},M}(S_N) + \frac{2\alpha\,\varepsilon_M}{(1-\alpha)\,r^{(0)}_{\min}}\Big),$$

which is (87). □

Remark 4 (Practical seed selection)**.** The seed set $\mathcal{S}$ should be chosen to reflect the practitioner's prediction objective. Natural choices include state coordinates (when state prediction is the goal), slow observables identified by a preliminary spectral analysis, or any observables whose future evolution is of primary interest. When no domain knowledge is available, one can use standard PR ($\mathcal{S} = \{1, \dots, \tilde{N}\}$) or run standard PR first to identify candidate seeds, then refine with PPR. The Andersen–Chung–Lang theory [42] provides complementary guarantees via sweep cuts.