



HAL
open science

Optimal dimensionality reduction of Markov chains using graph transformation

Deepti Kannan, Daniel Sharpe, T D Swinburne, David Wales

► **To cite this version:**

Deepti Kannan, Daniel Sharpe, T D Swinburne, David Wales. Optimal dimensionality reduction of Markov chains using graph transformation. *Journal of Chemical Physics*, 2020, 153 (24), pp.244108. 10.1063/5.0025174 . hal-03090008

HAL Id: hal-03090008

<https://hal.science/hal-03090008>

Submitted on 4 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal dimensionality reduction of Markov chains using graph transformation

Deepti Kannan,^{1, a)} Daniel J. Sharpe,^{1, a)} Thomas D. Swinburne,^{2, b)} and David J. Wales^{1, c)}

¹⁾*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

²⁾*Aix-Marseille Université, CNRS, CINaM UMR 7325, Campus de Luminy, 13288 Marseille, France*

(Dated: 20 July 2020)

Markov chains can accurately model the state-to-state dynamics of a wide range of complex systems, but the underlying transition matrix is ill-conditioned when the dynamics feature a separation of timescales. Graph transformation (GT) provides a numerically stable method to compute exact mean first passage times (MFPTs) between states, which are the usual dynamical observables, in continuous-time Markov chains (CTMCs). Here, we generalize the GT algorithm to discrete-time Markov chains (DTMCs), which are commonly estimated from simulation data, for example in the Markov State Model approach. We then consider the dimensionality reduction of CTMCs and DTMCs, which aids model interpretation and facilitates more expensive computations, including sampling of pathways. We perform a detailed numerical analysis of existing methods to compute the optimal reduced CTMC, given a partitioning of the network into metastable communities (macrostates) of nodes (microstates). We show that approaches based on linear algebra encounter numerical problems that arise from the requisite metastability. We propose an alternative approach, using GT to compute the matrix of intermicrostate MFPTs in the original Markov chain, from which a matrix of weighted intermacrostate MFPTs can be obtained. We also propose an approximation to the weighted-MFPT matrix in the strongly metastable limit. Inversion of the weighted-MFPT matrix, which is better conditioned than the matrices that must be inverted in alternative dimensionality reduction schemes, then yields the optimal reduced Markov chain. The superior numerical stability of the GT approach therefore enables us to realize optimal Markovian coarse-graining of systems with rare event dynamics.

Keywords: graph transformation, Markov chain, Markov process, Markov State Model, dimensionality reduction, lumping, stochastic dynamics, master equation, rare events, metastability, kinetic Monte Carlo

I. INTRODUCTION

Many stochastic processes can be represented by discrete- or continuous-time Markov chains, which generate memoryless dynamics for transitions between nodes of a network.^{1–9} The usual dynamical observable in such models is the mean first passage time (MFPT) from a set of initial nodes \mathcal{B} to a set of absorbing nodes \mathcal{A} .^{10–22} Standard linear algebra methods to compute MFPTs encounter numerical issues for metastable Markov chains because the separation of slow and fast timescales in the system dynamics leads to severe ill-conditioning.^{23–28} Since rare events are ubiquitous in realistic models of stochastic dynamical processes,^{29–43} more numerically stable algorithms are often required for the analysis of Markov chain dynamics. Graph transformation^{44–49} (GT) provides an exact method to compute the $\mathcal{A} \leftarrow \mathcal{B}$ MFPT that retains numerical precision even for strongly metastable Markov chains.⁴⁸ The method does not assume that the Markov chain satisfies the detailed balance condition,⁴ nor are there any constraints on the initial occupation probability distribution within \mathcal{B} .

Graph transformation was originally introduced to compute MFPTs in continuous-time Markov Chains (CTMCs),^{45,47} which are parameterized by transition rates.³ Here we show that the GT algorithm can be adapted to treat discrete-time Markov chains (DTMCs), which are parameterized by transition probabilities for a fixed time step.² This generalization is important because many methods have been developed to construct DTMCs from sparse simulation data.^{50,51} For example, master equation approaches based on explicit dynamics^{52–56} are designed to construct a DTMC from first passage time data in a continuous state space. Estimation of a CTMC from trajectory data is also possible,⁵⁷ but is significantly more difficult.^{58–60} Since the analysis of DTMCs featuring rare events encounters numerical issues due to finite precision,^{24–28} as for continuous-time counterparts, a discrete-time formulation of the GT algorithm offers an attractive method for computing $\mathcal{A} \leftarrow \mathcal{B}$ MFPTs in the metastable regime.

The remainder of this contribution is concerned with the dimensionality reduction, or coarse-graining, of Markov chains using exact MFPTs. Reducing the dimensionality of a Markov chain, while accurately preserving global dynamical properties of the original model, is an active area of research.^{1,61–74} Coarse-graining facilitates computational analyses that are prohibitively expensive for the original model, such as sampling of the $\mathcal{A} \leftarrow \mathcal{B}$ transition path ensemble by kinetic Monte Carlo (kMC)

^{a)}These two authors contributed equally

^{b)}Electronic mail: swinburne@cinam.univ-mrs.fr

^{c)}Electronic mail: dw34@cam.ac.uk

methods.^{75–81} Moreover, the reduced Markov chain only preserves the slowest dynamical processes and so is less ill-conditioned than the original representation. The coarse-grained model may therefore be amenable to analyses based on linear algebra, such as computation of the fundamental matrix,^{10,82–86} from which many insightful dynamical properties can be derived.⁵ The reduced model is also easier to interpret, since it only describes dynamics between key metastable macrostates (i.e. communities of nodes).^{87,88}

It may seem intuitive that the MFPTs between communities of nodes could be used to directly estimate a coarse-grained DTMC or CTMC, by expressions similar to those employed in milestoning.^{89–91} However, Ref. 92 showed that, for the reduced Markov chain to preserve the stationary distribution of the original Markov chain, the intercommunity MFPTs must satisfy a constraint defined by the Kemeny constant,^{1,92–99} which we show arises as a direct consequence of the detailed balance condition. As a result, evaluation of the matrix of appropriately weighted intercommunity MFPTs, introduced in Ref. 92, requires computation of the intermicrostate MFPTs for transitions between all pairs of nodes in the original Markov chain. Inversion of this “weighted-MFPT” matrix then yields the “optimal” reduced CTMC for a given partitioning of the original Markov chain,¹⁰⁰ as defined in Sec. III A.⁹²

We demonstrate that the GT algorithm provides a numerically stable route to obtain the matrix of weighted-MFPTs, and hence the optimal intercommunity transition rates for a given community structure. A careful numerical comparison between our approach and alternative formulations⁹², including the Hummer-Szabo relation,¹⁰⁰ shows that GT greatly extends the regime in which optimal Markovian coarse-graining of metastable systems is feasible. Moreover, our analysis confirms that the optimal reduced CTMC represents the dynamics of the original system more accurately than the coarse-grained Markov chain obtained from the simpler local equilibrium approximation.^{100,101} We also suggest more scalable approaches to dimensionality reduction based on the GT algorithm, including an approximation to the optimal reduced CTMC that is valid in the strongly metastable limit.

II. MEAN FIRST PASSAGE TIMES AND THE GRAPH TRANSFORMATION ALGORITHM

We consider the $\mathcal{A} \leftarrow \mathcal{B}$ mean first passage time (MFPT) $\mathcal{T}_{\mathcal{A}\mathcal{B}}$ for the transition to the absorbing macrostate (community of nodes) \mathcal{A} from the initial macrostate \mathcal{B} , associated with a specified initial probability distribution $p_b(0)$, $b \in \mathcal{B}$. The $\mathcal{A} \leftarrow \mathcal{B}$ MFPT is defined as the average first hitting time^{4,102} for trajectories initialised in \mathcal{B} to reach the boundary nodes of the target set \mathcal{A} . The discrete state space, comprising V microstates (nodes), is partitioned into the endpoint

macrostates, \mathcal{A} and \mathcal{B} , and the set of intervening nodes, denoted $\mathcal{I} = (\mathcal{A} \cup \mathcal{B})^c$.

A. Graph transformation for continuous-time Markov chains

A CTMC is parameterized by a transition rate matrix \mathbf{K} , where the elements K_{ij} are the $i \leftarrow j$ intermicrostate transition rates, and the diagonal terms ensure that the sum of elements in any column of the matrix is zero, $K_{jj} = -\sum_{\gamma \neq j} K_{\gamma j}$. Each node j has a mean waiting time $\tau_j = 1/\sum_{\gamma \neq j} K_{\gamma j}$, from which the branching probability matrix¹⁰³ \mathbf{P} emerges as $P_{ij} = K_{ij}\tau_j$.¹⁰⁴ The graph transformation (GT) algorithm is a method to iteratively remove nodes from a CTMC while preserving the MFPT for the transition to the absorbing macrostate \mathcal{A} .^{44–49} In the elimination of node x by the GT algorithm, the waiting times and branching probabilities of the remaining nodes are updated according to⁴⁷

$$\tau_j \rightarrow \tau'_j = \tau_j + \frac{P_{xj}\tau_x}{1 - P_{xx}}, \quad (1)$$

$$P_{ij} \rightarrow P'_{ij} = P_{ij} + \frac{P_{ix}P_{xj}}{1 - P_{xx}}, \quad (2)$$

respectively. The renormalized $i \leftarrow j$ transition probabilities account for pathways consisting of a transition to node x from j , followed by an arbitrary number of self-loop (i.e. $x \leftarrow x$) transitions in x , before a final transition to node i . The transformation also introduces self-loop transitions such that the probability flow is conserved, $\sum_{\gamma} P'_{\gamma j} = 1$, for all remaining nodes j . Similarly, the waiting times (Eq. 1) are renormalized to account for the expected time spent in the eliminated node x , weighted by the probability of transitioning to x from j . The GT algorithm can also be extended to eliminate multiple nodes simultaneously.^{49,105}

To compute the MFPT between two sets of endpoint nodes \mathcal{A} and \mathcal{B} , the GT algorithm is first used to iteratively eliminate all nodes of the set \mathcal{I} . The renormalized network comprising only the nodes of the set $\mathcal{A} \cup \mathcal{B}$ is then stored. For each node $b \in \mathcal{B}$, the nodes of the set $\mathcal{B} \setminus b$ are eliminated from this stored network, leaving only the nodes in $\mathcal{A} \cup b$. The $\mathcal{A} \leftarrow b$ MFPT is then the expected escape time from b , which reads

$$\mathcal{T}_{\mathcal{A}b} = \frac{\tau'_b}{1 - P'_{bb}}. \quad (3)$$

The overall $\mathcal{A} \leftarrow \mathcal{B}$ MFPT is given by a weighted average of MFPTs for transitions to the absorbing macrostate from nodes of the initial macrostate,

$$\mathcal{T}_{\mathcal{A}\mathcal{B}} = \frac{1}{\sum_{b \in \mathcal{B}} p_b(0)} \sum_{b \in \mathcal{B}} \mathcal{T}_{\mathcal{A}b} p_b(0). \quad (4)$$

The GT procedure to compute the $\mathcal{A} \leftarrow \mathcal{B}$ MFPT is illustrated in Fig. 1. The $\mathcal{B} \leftarrow \mathcal{A}$ MFPT is computed using

expressions analogous to Eqs. 3 and 4 for the reverse direction. Since the renormalized network with state space $\mathcal{A} \cup \mathcal{B}$ is stored for the computation of $\mathcal{T}_{\mathcal{A}\mathcal{B}}$, $\mathcal{T}_{\mathcal{B}\mathcal{A}}$ can be obtained with little additional computational effort.

The efficiency and numerical stability of GT is essentially unaffected by the metastability of the Markov chain.⁴⁸ The GT algorithm retains numerical precision because the $(1 - P_{xx})$ factors that appear in the renormalization equations (Eqs. 1 and 2) can be computed indirectly via $\sum_{\gamma \neq x} P_{\gamma x} = 1 - P_{xx}$ when $P_{xx} \rightarrow 1$.⁴⁷ This improvement in numerical stability corresponds to the ‘‘GTH advantage’’¹⁰⁶ in the Grassmann-Taksar-Heyman (GTH) algorithm for computation of the stationary probability distribution.^{11,107}

B. Graph transformation for discrete-time Markov chains

The GT procedure described in Sec. II A can be applied to a discrete-time Markov chain (DTMC) by replacing the CTMC branching probabilities $P_{ij} = K_{ij}\tau_j$ and mean waiting times τ_j with the discrete-time transition probabilities $\mathbf{T}_{ij}(\Delta)$ and uniform lag times $\Delta_j \equiv \Delta \forall j$. The discrete- and continuous-time formulations are related via⁸

$$\mathbf{T}(\Delta) = \exp(\mathbf{K}\Delta) \quad \text{and} \quad \mathbf{K} = \lim_{\Delta \rightarrow 0} (\mathbf{T}(\Delta) - \mathbb{I})/\Delta, \quad (5)$$

where \mathbb{I} is the identity matrix. Since the columns of \mathbf{K} sum to zero, the columns of $\mathbf{T}(\Delta)$ sum to unity.⁴ In a CTMC, the time for the $i \leftarrow j$ transition is exponentially distributed with mean τ_j , because outgoing transitions from node j are competing Poisson processes.¹⁰⁴ In a DTMC, the time between transitions is fixed by the lag time Δ . Hence, the transition matrix $\mathbf{T}(\Delta)$ propagates the occupation probability distribution $\mathbf{p}_n \equiv \mathbf{p}(t_n)$ via the Chapman-Kolmogorov equation $\mathbf{p}_{n+1} = \mathbf{T}(\Delta)\mathbf{p}_n$ at discrete time intervals $t_n = n\Delta$.⁴ We show below that the elimination of nodes in a DTMC using renormalization (i.e. discrete-time analogs of Eqs. 1 and 2) preserves MFPTs to an absorbing macrostate.

For a CTMC, the MFPT between two endpoint macrostates can be written as a sum of contributions from individual paths weighted by the corresponding path probabilities.⁴⁵ Consider a particular discrete path ξ from the initial to the absorbing macrostate, written as a sequence of n nodes, $\xi = \{i_n \leftarrow i_{n-1} \leftarrow \dots \leftarrow i_1\}$, with starting node $i_1 \equiv b \in \mathcal{B}$ and target node $i_n \equiv a \in \mathcal{A}$. The path probability \mathcal{W}_ξ is simply a product of branching probabilities for all transitions ($i \leftarrow j$) along the path ξ , weighted by the initial occupation probability for the starting node b , $\mathcal{W}_\xi = p_b(0) \prod_{(i,j) \in \xi} P_{ij}$.^{45,105} The MFPT associated with the path ξ is a sum of mean waiting times for microstates visited along the path prior to hitting the absorbing node,⁴⁷ $\mathcal{T}_\xi = \sum_{i=1}^{n-1} \tau_i$. Therefore, the overall $\mathcal{A} \leftarrow b$ MFPT is

$$\mathcal{T}_{Ab} = \sum_{a \in \mathcal{A}} \sum_{\xi \in \mathcal{A} \leftarrow b} \mathcal{W}_\xi \mathcal{T}_\xi. \quad (6)$$

It can be shown that renormalization of the waiting times and transition probabilities by GT (Eqs. 1 and 2) preserves the pathwise sum over product nodes (Eq. 6).^{45,47} To apply the GT algorithm to a DTMC, we consider an analogous formulation of the MFPT as a weighted sum of contributions from individual discrete-time $\mathcal{A} \leftarrow \mathcal{B}$ paths.

The probability of a discrete-time path is a product of discrete-time transition probabilities $\mathbf{T}_{ij}(\Delta)$ weighted by the occupation probability of the initial node, directly analogous to the continuous-time case. Hence, the renormalization of the discrete-time transition probabilities by an equation of the same form as Eq. 2 will correctly preserve the individual path probabilities for each member of the ensemble of $\mathcal{A} \leftarrow \mathcal{B}$ paths, following the proof for the continuous-time case in Ref. 47. Alternative proofs that renormalization of transition probabilities via Eq. 2 is valid in both continuous and discrete time appear in other contexts, for example, in the theory of stochastic complementation (see e.g. Ref. 108).

The renormalization of the waiting times associated with nodes in a DTMC requires more careful consideration. In a CTMC, the renormalized mean waiting time τ'_j (Eq. 1) accounts for the average time to leave node j (possibly returning to j) via an arbitrary number of self-loop transitions in the eliminated node x . In a DTMC, the probability $\mathbf{T}_{ij}(\Delta)$ of the $i \leftarrow j$ transition is associated with a *fixed* waiting time Δ . For the weighted sum of contributions to the $\mathcal{A} \leftarrow \mathcal{B}$ MFPT from individual pathways (Eq. 6) to be conserved, the renormalized lag times in a DTMC must also account for the average number of steps $\langle N_j \rangle$ required to leave a node j (possibly returning to j) via an arbitrary number of self-loop transitions in the eliminated node x . This effect will increment the lag time on average by $\Delta'_j = \langle N_j \rangle \Delta$, meaning the initially uniform lag time $\Delta_j = \Delta \forall j$ becomes node-dependent as nodes are eliminated via GT. We therefore introduce the V -dimensional vector of waiting times, Δ , associated with the generalized transition probability matrix $\mathbf{T}'(\Delta)$ for the renormalized DTMC.

Following this reasoning, the renormalized DTMC is formally equivalent to an artificial CTMC where the mean waiting time for the j -th node is Δ_j and the $i \leftarrow j$ transition probabilities are given by $\mathbf{T}'_{ij}(\Delta)$. The renormalization equations in the GT algorithm applied to a DTMC are therefore (by analogy with Eqs. 1 and 2)

$$\Delta_j \rightarrow \Delta'_j = \Delta_j + \frac{\mathbf{T}_{xj}(\Delta)\Delta_x}{1 - \mathbf{T}_{xx}(\Delta)}, \quad (7)$$

$$\mathbf{T}_{ij}(\Delta) \rightarrow \mathbf{T}'_{ij}(\Delta) = \mathbf{T}_{ij}(\Delta) + \frac{\mathbf{T}_{ix}(\Delta)\mathbf{T}_{xj}(\Delta)}{1 - \mathbf{T}_{xx}(\Delta)}. \quad (8)$$

Because the lag time is node-dependent, the elapsed time for a trajectory of n steps, $\sum_{i=1}^{n-1} \Delta_i$, is path-dependent, unlike a standard DTMC, where the elapsed time is always $n\Delta$. That is, the discrete time interval at which the occupation probability distribution for a renormalized DTMC is propagated represents an *average* over

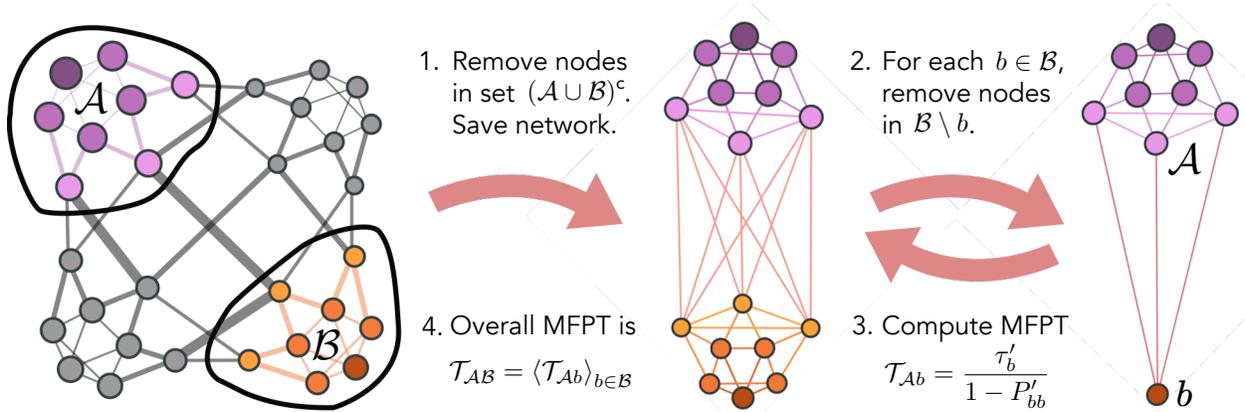


FIG. 1. The graph transformation (GT) method to compute mean first passage times (MFPTs) from a set of initial nodes \mathcal{B} to a set of absorbing nodes \mathcal{A} , illustrated for a model 32-node network. Here, τ'_b and P'_{bb} are the renormalized waiting times and self-loop transition probabilities (Eqs. 1 and 2, respectively) from node b in the graph-transformed network comprising nodes of the set $(\mathcal{A} \cup b)$. The overall $\mathcal{A} \leftarrow \mathcal{B}$ MFPT is obtained as a weighted average of $\mathcal{A} \leftarrow b \in \mathcal{B}$ MFPTs (Eq. 4). The MFPT for the reverse ($\mathcal{B} \leftarrow \mathcal{A}$) direction can be obtained by repeating steps 2-4, removing nodes $a \in \mathcal{A}$.

multiple possible pathways. Hence, the probability distribution evolves through a modified Chapman-Kolmogorov equation $\mathbf{p}_{n+1} = \mathbf{T}(\Delta)\mathbf{p}_n$, where the time evolution is defined in expectation, $\langle t_{n+1} \rangle = (\Delta^\top \mathbf{p}_n) + \langle t_n \rangle$, as opposed to $t_{n+1} = \Delta + t_n$ for a uniform lag time Δ . Therefore $\mathcal{A} \leftarrow \mathcal{B}$ trajectories can still be sampled on a renormalized DTMC by incrementing the simulation clock using the renormalized, node-dependent lag time Δ'_j for a $i \leftarrow j$ transition (Eq. 7), and noting that the trajectory time is an *expectation* over visits to eliminated nodes.

A formal proof that Eqs. 7 and 8 preserve the $\mathcal{A} \leftarrow \mathcal{B}$ MFPT for a DTMC is presented in Appendix C. Algorithms for the computation of MFPTs that are very similar to the GT algorithm have previously been described in the context of DTMCs.^{12,27,82,83} For instance, Hunter proposed an extension of the GTH algorithm^{11,107} for the computation of the stationary distribution vector, which is valid for both discrete- and continuous-time stochastic matrices,¹⁰⁶ to compute the matrix of all pairwise intermicrostate MFPTs.^{13,15} However, to the best of our knowledge, the GT algorithm of Ref. 47 has not been applied to DTMCs, and the interpretation of the renormalized waiting times in discrete-time (Eq. 7), based on the proof presented in Appendix C, has not been discussed.

The continuous-time branching probability matrix \mathbf{P} is, in general, more sparse than a corresponding discrete-time transition probability matrix $\mathbf{T}(\Delta)$ (Eq. 5), since \mathbf{P} is derived from the transition probabilities per unit time for an infinitesimally small time step.^{4,104} In addition, the discrete-time transition probability matrix may contain self-loop transitions even before the application of the GT algorithm. For these reasons, a GT computation will be less efficient for a DTMC than for a corresponding CTMC. Therefore, if the stochastic model to be analyzed is a CTMC, then there is no advantage to transforming the system to an equivalent DTMC via Eq. 5. How-

ever, the reverse transformation to obtain a CTMC from a given DTMC is often not feasible.^{59,60} Thus the application of the GT algorithm to DTMCs is a valuable approach to circumvent numerical issues when analyzing a discrete-time model featuring rare events.

In a standard DTMC with uniform lag time Δ , the $\mathcal{A} \leftarrow \mathcal{B}$ MFPT is simply the average number of steps along $\mathcal{A} \leftarrow \mathcal{B}$ paths, \mathcal{L}_{AB} , multiplied by Δ , i.e. $\mathcal{T}_{AB} = \Delta \mathcal{L}_{AB}$. \mathcal{L}_{AB} can thus be obtained for DTMCs or CTMCs using the GT algorithm (Eqs. 1-3) by setting the initial waiting times for all nodes to unity. Although \mathcal{L}_{AB} is not a dynamical observable, it is nonetheless of theoretical interest.⁵ For instance, the number of steps along stochastic paths has been extensively used to characterize metastable macrostates in the theory of dynamical phase transitions,⁴³ where \mathcal{L}_{AB} is referred to as the $\mathcal{A} \leftarrow \mathcal{B}$ mean *dynamical activity*.¹⁰⁹ \mathcal{L}_{AB} can be compared to the numbers of steps along $\mathcal{A} \leftarrow \mathcal{B}$ paths determined by k shortest paths algorithms with appropriate edge weights,¹¹⁰ highlighting the paths that exhibit numerous revisits ('flickering') within metastable macrostates. Knowledge of \mathcal{L}_{AB} is helpful in choosing a kinetic Monte Carlo (kMC) method for the simulation of $\mathcal{A} \leftarrow \mathcal{B}$ trajectories, since standard kMC methods⁷⁵ become unfeasible when trajectories become too long. In contrast, more advanced kMC algorithms are agnostic to the number of steps along paths.⁷⁶⁻⁸¹

C. Linear algebra methods for computing MFPTs in Markov chains

A straightforward linear algebra approach to compute $\mathcal{A} \leftarrow \mathcal{B}$ MFPTs for a DTMC or CTMC is based on the theory of absorbing Markov chains. Consider the $V' \times V'$ transition rate matrix \mathbf{K}' comprising only the

V' microstates of the nonabsorbing set $\mathcal{S} \equiv \mathcal{A}^c$. Transitions to absorbing nodes in this modified CTMC remain accounted for in the waiting times of nodes in \mathcal{S} with direct connections to \mathcal{A} ; thus, the sums of elements in the corresponding columns of \mathbf{K}' will be less than zero. Hence, this absorbing Markov chain “leaks” probability, and MFPTs for transitions to \mathcal{A} are related to the probability flux out of the nonabsorbing region. Unlike \mathbf{K} , \mathbf{K}' does not have a zero eigenvalue and is therefore invertible, giving^{29,105}

$$\mathcal{T}_{\mathcal{A}\mathcal{B}} = -\mathbf{1}_{V'}^\top [\mathbf{K}']^{-1} \mathbf{p}'(0), \quad (9)$$

where $\mathbf{1}_{V'}$ is a V' -dimensional column vector with all elements equal to unity, and $\mathbf{p}'(0)$ is the normalized initial occupation probability distribution constrained to \mathcal{B} , i.e. $p'_b(0) = p_b(0) / [\sum_{b' \in \mathcal{B}} p_{b'}(0)]$. The above expression can also be evaluated by solving the linear problem $\mathbf{K}'\mathbf{x} = \mathbf{p}'(0)$ from which $\mathcal{T}_{\mathcal{A}\mathcal{B}} = -\mathbf{1}_{V'}^\top \mathbf{x}$, or through eigen-decomposition of \mathbf{K}' to give^{57,92}

$$\mathcal{T}_{\mathcal{A}\mathcal{B}} = -\sum_k \frac{1}{\gamma_k} \mathbf{1}_{V'}^\top (\boldsymbol{\psi}'_k \otimes \boldsymbol{\phi}'_k) \mathbf{p}'(0), \quad (10)$$

where $\boldsymbol{\psi}'_k$ and $\boldsymbol{\phi}'_k$ are the k -th right and left (orthonormal) eigenvectors of \mathbf{K}' , respectively, and $\gamma'_k < 0$ is the associated eigenvalue. A DTMC shares the same set of right and left eigenvectors as an equivalent CTMC,⁵⁷ and the k -th eigenvalue is related to the corresponding eigenvalue of an equivalent CTMC via $\lambda_k = \exp(\gamma_k \Delta)$, cf. Eq. 5.⁵² Hence, Eq. 10 also provides a means to compute $\mathcal{A} \leftarrow \mathcal{B}$ MFPTs for DTMCs via construction of an absorbing transition probability matrix $\mathbf{T}'_{ij}(\Delta)$. In addition to the matrix inversion (Eq. 9) and eigendecomposition solutions (Eq. 10), the formulation of the absorbing transition rate matrix \mathbf{K}' allows MFPTs to be computed from the elements of the *absorbing fundamental matrix*,^{1,5} which is discussed in Appendix B.

An additional linear algebra method to compute the MFPT between two microstates $i \leftarrow j$ is provided by the expression (Appendix E)

$$\mathcal{T}_{ij} = \frac{1}{\pi_i} (Z_{ij}^{\text{CT}} - Z_{ii}^{\text{CT}}), \quad (11)$$

where

$$\mathbf{Z}^{\text{CT}} = (\boldsymbol{\pi} \mathbf{1}_V^\top + \mathbf{K})^{-1} - \boldsymbol{\pi} \mathbf{1}_V^\top, \quad (12)$$

is the *ergodic fundamental matrix*^{10,82–86} of an irreducible CTMC, which exists when the Markov chain has a defined stationary distribution $\boldsymbol{\pi}$. In Appendix D, we show that there is some freedom in the choice of \mathbf{Z}^{CT} .^{10,15,111,112} In Ref. 92, Eqs. 11 and 12 were related to a difference of correlation functions¹⁰⁰ of the Markov chain. Eqs. 11 and 12 can be used to compute a particular $\mathcal{A} \leftarrow \mathcal{B}$ MFPT by constructing a modified Markov chain in which the nodes of the set \mathcal{A} are represented by a single supernode,¹¹³ so that the elements of the fundamental matrix yield the MFPTs $\mathcal{T}_{\mathcal{A}j} \forall j \notin \mathcal{A}$, and $\mathcal{T}_{\mathcal{A}\mathcal{B}}$ can then be obtained via Eq. 4.

In the discrete-time case, the most common definition for the fundamental matrix is given by^{1,84}

$$\mathbf{Z}^{\text{DT}} = (\boldsymbol{\pi} \mathbf{1}_V^\top + \mathbf{I} - \mathbf{T}(\Delta))^{-1}, \quad (13)$$

in which case it can be shown that the $i \leftarrow j$ MFPT is¹

$$\mathcal{T}_{ij} = \Delta(\delta_{ij} + Z_{ii}^{\text{DT}} - Z_{ij}^{\text{DT}})/\pi_i. \quad (14)$$

The factor of the lag time Δ in Eq. 14 is required for consistency with the continuous-time result in Eq. 11. The discrete-time formulation (Eq. 14) yields $\mathcal{T}_{ii} = \Delta/\pi_i$, in accordance with Kac’s Lemma,¹¹⁴ whereas the diagonal elements of the MFPT matrix are strictly zero in continuous time (Eq. 11).⁹²

Computation of MFPTs via the elements of the fundamental matrix (Eq. 11) yields the $V \times V$ matrix \mathcal{T} of MFPTs for transitions between *all* pairs of microstates of the Markov chain in a single matrix inversion operation. This matrix can be used to estimate the optimal coarse-grained CTMC for a given partitioning of the nodes (Sec. III). The matrix \mathcal{T} also provides a means to convert from a discrete- to a continuous-time formulation via⁹²

$$\mathbf{K} = \mathcal{T}^{-1}(-\mathbf{1}_V \mathbf{1}_V^\top + \text{diag}(\boldsymbol{\pi})^{-1}), \quad (15)$$

where $\text{diag}(\boldsymbol{\pi})$ is the diagonal matrix with elements corresponding to the stationary (equilibrium occupation) probabilities of the V microstates, $[\text{diag}(\boldsymbol{\pi})]_{ij} = \pi_i \delta_{ij}$, and the diagonal elements \mathcal{T}_{ii} of the MFPT matrix are taken to be zero. In general, the conditions required for a DTMC to be “embeddable”, i.e. have an underlying CTMC, are unknown, and are not always satisfied.^{59,60} In practice, a CTMC can be estimated from a DTMC by solving a constrained optimization problem for the “closest” valid rate matrix satisfying Eq. 5, but the optimization is non-convex and scales poorly.⁵⁸ Evaluation of Eq. 15 provides an alternative approach to perform the conversion, but requires inversion of a V -dimensional square matrix, which is not scalable and is numerically unstable for metastable Markov chains (Sec. II E).^{23,48} Hence, it is often not possible to convert a DTMC to a corresponding CTMC.

D. Derivation of the Kemeny constant from the detailed balance condition

In addition to the MFPT, another important dynamical observable is the Kemeny constant,^{1,92–99} ζ , defined as

$$\zeta = \sum_i \pi_i \mathcal{T}_{ij} \quad \forall j. \quad (16)$$

The Kemeny constant is interpreted as the average mixing time of an irreducible Markov chain.^{115,116} Eq. 16 implies that the weighted sum of MFPTs to all possible target microstates i is the same for all initial microstates

j .⁵ The Kemeny constant can be expressed in terms of the trace of the fundamental matrix^{10,82–86} (Eqs. 12 and 13) as⁹⁸

$$\zeta = -\text{Tr}(\mathbf{Z}^{\text{CT}}) = \Delta \text{Tr}(\mathbf{Z}^{\text{DT}}), \quad (17)$$

for continuous and discrete time, respectively. In Appendix E, we derive the Kemeny constant, ζ , and the MFPT matrix, \mathcal{T} , from the fundamental matrix in continuous time, \mathbf{Z}^{CT} , and show that ζ represents an average mixing time.

It is insightful to show that the Kemeny constant arises straightforwardly from the detailed balance condition $\mathbf{K} \text{diag}(\boldsymbol{\pi}) = \text{diag}(\boldsymbol{\pi}) \mathbf{K}^\top$. Multiplying both sides of this expression by $\mathbf{1}_V^\top \mathcal{T}^{-1} \mathcal{T}$ and substituting for $\mathcal{T} \mathbf{K} \text{diag}(\boldsymbol{\pi})$ using Eq. 15 gives

$$\mathbf{1}_V^\top \mathcal{T}^{-1} (\mathbf{I} - \mathbf{1}_V \boldsymbol{\pi}^\top) = \mathbf{1}_V^\top \text{diag}(\boldsymbol{\pi}) \mathbf{K}^\top = \boldsymbol{\pi}^\top \mathbf{K}^\top = 0, \quad (18)$$

which follows from the fact that $\boldsymbol{\pi}$ is the unique right eigenvector of \mathbf{K} associated with the zero eigenvalue. Hence, $\mathbf{1}_V^\top \mathcal{T}^{-1} = \mathbf{1}_V^\top \mathcal{T}^{-1} \mathbf{1}_V \boldsymbol{\pi}^\top$. Postmultiplying both sides of this constraint by \mathcal{T} and rearranging yields

$$\boldsymbol{\pi}^\top \mathcal{T} = \mathbf{1}_V^\top / (\mathbf{1}_V^\top \mathcal{T}^{-1} \mathbf{1}_V) \equiv \zeta \mathbf{1}_V^\top. \quad (19)$$

Eq. 19 further implies that $\boldsymbol{\pi}^\top \mathcal{T} \mathbf{p}(0) = \zeta$, meaning that the mixing time is independent of the initial occupation probability distribution $\mathbf{p}(0)$. Although this remarkable fact follows from the detailed balance condition, the alternative derivation of the Kemeny constant given in Appendix E demonstrates that Eq. 19 holds even in the absence of detailed balance, provided that the global balance equation, $\mathbf{K} \boldsymbol{\pi} = \mathbf{0}$, is satisfied.

In Sec. III B, we will use the fact that there exists a constraint of the same form as Eq. 19 on the Kemeny constant of the optimal reduced Markov chain for a given partitioning of the network, which was first noted in Ref. 92.

E. Computational complexity and numerical stability of graph transformation and linear algebra methods

In general, linear algebra methods to compute MFPTs have superior time complexity compared to the graph transformation (GT) algorithm. However, GT retains numerical precision even in strongly metastable Markov chains owing to the GTH advantage,¹⁰⁶ whereas linear algebra methods suffer severe conditioning issues.

Computation of the matrix \mathcal{T} of all pairwise intermicrostate MFPTs via evaluation of the fundamental matrix (Eqs. 11 and 12) requires only a single matrix inversion operation, and hence has time complexity $\mathcal{O}(V^3)$.²³ The methods to compute a particular $\mathcal{A} \leftarrow \mathcal{B}$ MFPT based on inversion, linear solution, and eigendecomposition operations of an absorbing Markov chain (Eqs. 9 and 10) also have a formal complexity of $\mathcal{O}(V^3)$, but sparse linear algebra methods¹¹⁹ can provide a significant

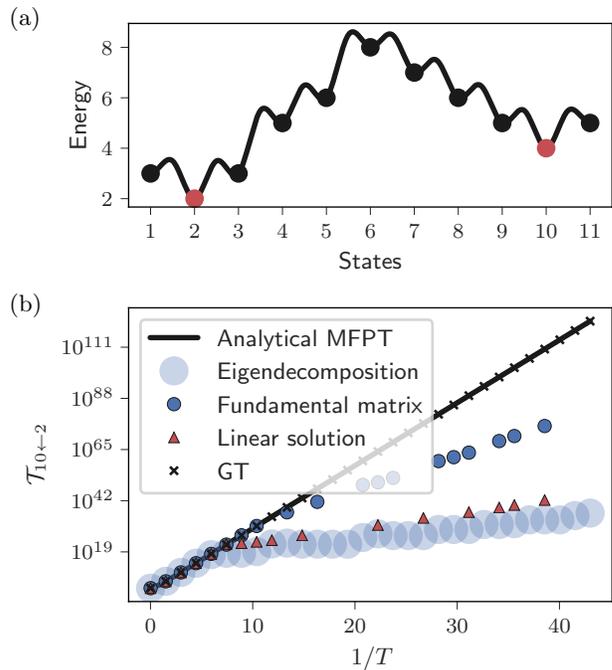


FIG. 2. (a) One-dimensional potential energy landscape for the 11-microstate unbranched chain. (b) Mean first passage time for the transition between microstates $2 \leftarrow 10$ computed using graph transformation (GT), inversion to obtain the fundamental matrix, eigendecomposition, and linear solution, compared to the exact analytical result from the recursive formula due to Weiss.^{117,118}

speedup when solving the linear problem $\mathbf{K}' \mathbf{x} = \mathbf{p}'(0)$ in certain cases. These methods can be used to compute the matrix of all pairwise intermicrostate MFPTs \mathcal{T} by considering V separate absorbing Markov chains, with each target node i specified as the absorbing node in turn. A single iteration of this procedure therefore yields the individual MFPTs \mathcal{T}_{ij} from all source nodes $j \neq i$. Computation of \mathcal{T} using the absorbing Markov chain formulation therefore has a formal time complexity of $\mathcal{O}(V^4)$.

Despite the $\mathcal{O}(V^3)$ scaling, eigendecomposition can in principle be achieved efficiently even for high-dimensional systems^{120,121} using Krylov subspace methods,¹²² such as the Lanczos algorithm,¹²³ applied to the symmetrized^{29,57} transition probability (or rate) matrix. However, this approach does not mitigate the severe ill-conditioning issues for systems exhibiting rare event dynamics.^{23,48} In such systems, the second dominant eigenvalues of DTMCs and CTMCs tend to unity and zero,¹²⁴ respectively, with increasing separation of timescales in the dynamics.^{24–28} Moreover, for metastable systems, there may be several eigenvalues associated with slow dynamical eigenmodes that are very similar in magnitude,^{41,125} which also causes difficulty for the Lanczos algorithm.^{126,127} For metastable

Markov chains, small perturbations in the transition probabilities (or rates) can lead to profound changes in global dynamical quantities, such as the stationary distribution^{26,128–131} and, by extension, MFPTs.^{132–134} This effect complicates the use of Krylov subspace methods for eigendecomposition of sparse matrices, since the iterative updates of the solution vector may fail to converge,²⁵ and the choice of initial approximation to the solution vector is crucial to achieving convergence.¹³⁵ The sensitivity of global dynamical properties to perturbations in the transition matrix elements also affects the numerical stability of matrix inversion methods and dense methods for eigendecomposition,¹³⁶ owing to the propagation of roundoff error in the floating point operations.

To illustrate the superior numerical stability of GT, we consider a simple 11-microstate unbranched CTMC, where the nodes and edges correspond to the local minima and transition states of a one-dimensional potential energy landscape (Fig. 2).¹⁰¹ The unbranched chain provides a useful benchmark for testing the numerical stability of methods to compute MFPTs because the exact intermicrostate MFPTs can be obtained from the recursive formula due to Weiss (see Supplementary Material Sec. II).^{117,118} Since the harmonic transition state theory¹³⁷ rate constants have an Arrhenius form, the effective temperature T governs the metastability of the Markov chain. The GT algorithm yields precise agreement with the exact recursive formula across the entire temperature range considered, whereas computations based on inversion to obtain the fundamental matrix⁹² (Eqs. 11 and 12), linear solution (Eq. 9), and eigendecomposition^{23,29} (Eq. 10) break down in the metastable regime. The failure of linear algebra methods can be attributed solely to the fact that the rate matrix is ill-conditioned, because the dimensionality of the model system is not prohibitive for any of the methods employed. DTMCs exhibiting rare event dynamics are affected by the same ill-conditioning and numerical problems as their continuous-time counterparts.^{24–28} Hence, we anticipate that the GT algorithm will be equally useful for computing MFPTs in DTMCs (Sec. IIB).

The numerical results presented in Fig. 2 were obtained using machine double precision.¹⁵ Employing higher-precision floating point data types only marginally improves the stability of linear algebra methods, while significantly increasing the required computational time and memory. Thus, it is preferable to employ the GT method, which benefits from the stability conferred by the GTH advantage,¹⁰⁶ in the metastable regime.

The time complexity of a single GT computation is roughly $\mathcal{O}(V^4)$ for sparse random networks,⁴⁵ and each calculation yields both the forward and reverse MFPTs for the chosen endpoint sets of nodes. Matrix generalisations of GT^{49,105,108} enable simultaneous elimination of blocks of B nodes, with a reduced time complexity of $\mathcal{O}(VB^2)$. The computation of \mathcal{T} using GT therefore requires $V(V-1)/2$ separate calculations, resulting in

an overall time complexity of approximately $\mathcal{O}(V^6)$ when eliminating nodes individually, or $\mathcal{O}(V^3B^3)$ when removing nodes in blocks of B nodes. The efficiency of a GT-based approach to computing \mathcal{T} can be improved using the formulation of GT employed in kinetic path sampling (kPS),^{79,80} in which transitions from eliminated to non-eliminated nodes are preserved. This algorithm is analogous to a LU decomposition of a matrix.⁷⁹ It is then not necessary to redo the complete GT computation for each pair of nodes, and instead individual GT iterations can be undone, resulting in approximately $\mathcal{O}(V^5)$ overall time complexity with careful bookkeeping of transition probabilities and waiting times. Although the computation of \mathcal{T} using GT scales less favourably, it remains feasible for sparse networks comprising a few thousand microstates. We also note that the time complexity of a GT computation is strongly dependent on the sparsity of the network and on the heterogeneity of the node degree distribution,⁴⁵ and time complexity better than $\mathcal{O}(V^4)$ may be observed for some networks.⁷⁹

III. DIMENSIONALITY REDUCTION OF MARKOV CHAINS USING MEAN FIRST PASSAGE TIMES

We now consider the problem of constructing a coarse-grained CTMC that accurately preserves the global dynamical properties of the original model, given a partitioning of a discrete- or continuous-time Markov chain into $N < V$ communities $\mathcal{C} \equiv \{I, J, \dots\}$.^{1,61–74} To ensure that a reduced Markovian model is a valid approximation to the original system, the community structure \mathcal{C} must appropriately characterize the metastable macrostates. Determination of \mathcal{C} is particularly challenging in the metastable regime, where spectral methods^{71–73,138} are numerically unstable (Sec. IIE), and is beyond the scope of the present work. We have previously shown that multi-level regularized Markov clustering⁸¹ and recursive regrouping based on a transition rate threshold²⁹ provide viable approaches to community structure detection¹³⁹ in metastable Markov chains.

The reduced CTMC should satisfy the detailed balance condition and preserve the stationary distribution of the original Markov chain, so that the equilibrium occupation probability of the J -th community is $\Pi_J = \sum_{j \in J} \pi_j$. We employ capital letter indices to denote microstates of the reduced Markov chain, which correspond to macrostates (groups of microstates) of the original Markov chain. The stationary distribution $\boldsymbol{\pi}$ of the original model is required to compute a reduced CTMC. To avoid the propagation of numerical error, $\boldsymbol{\pi}$ should be computed via a numerically robust method, such as the GTH algorithm,^{11,107} uncoupling-coupling via stochastic complementation,¹⁰⁸ or iterative aggregation-disaggregation^{140,141} (IAD) methods, all of which are related to the GT algorithm (Sec. IIA).

A. Definition of the optimal reduced CTMC

Dimensionality reduction simplifies the dynamics of the original system, and therefore will induce error. A “perfect” coarse-grained trajectory can in principle be generated from a trajectory on the original model, by replacing the instantaneous node occupation (or indicator¹⁰⁰) function $n_j(t) \in [0, 1]$ at time t , for nodes $j \in J$, with the corresponding coarse-grained node occupation function $n_J(t) \in [0, 1]$.¹⁴² The error incurred in a dimensionality reduction procedure can then be gauged by comparing the time-dependent macrostate occupation correlation functions¹⁴³ produced from the ensemble of these “perfect” coarse-grained trajectories with those produced by the reduced system.

Hummer and Szabo¹⁰⁰ proposed a dimensionality reduction strategy that minimizes the error in the occupation correlation functions under the constraint that the dynamics of the coarse-grained model are also Markovian. This constraint requires that the time integrals of the macrostate correlation functions agree in the coarse-grained and original models. The resulting expression, which we rederive in the Supplementary Materials, is the Hummer-Szabo (HS) relation^{100,144}

$$\mathbf{K}_C = \mathbf{\Pi} \mathbf{\Pi}_N^\top - \text{diag}(\mathbf{\Pi})(\mathbf{A}^\top (\boldsymbol{\pi} \mathbf{1}_V^\top - \mathbf{K})^{-1} \text{diag}(\boldsymbol{\pi}) \mathbf{A})^{-1}. \quad (20)$$

Here, the elements $[\mathbf{K}_C]_{IJ}$ are the optimal $I \leftarrow J$ intercommunity transition rates, $\text{diag}(\mathbf{\Pi})$ is the diagonal matrix with elements $[\text{diag}(\mathbf{\Pi})]_{IJ} = \Pi_J \delta_{IJ}$, and \mathbf{A} is the $V \times N$ aggregation matrix with non-zero elements $A_{iJ} = 1$ if $i \in J$. Kells *et al.*⁹² derived an alternative expression for \mathbf{K}_C based on the matrix of all pairwise intermicrostate MFPTs \mathcal{T} (Sec. II C), which is therefore applicable to the dimensionality reduction of both DTMCs and CTMCs, given by

$$\mathbf{K}_C = \mathbf{\Pi} \mathbf{\Pi}_N^\top - \text{diag}(\mathbf{\Pi})[\mathbf{\Pi} \mathbf{\Pi}^\top + \mathbf{A}^\top \text{diag}(\boldsymbol{\pi}) \mathcal{T} (\boldsymbol{\pi} \mathbf{\Pi}^\top - \text{diag}(\boldsymbol{\pi}) \mathbf{A})]^{-1}. \quad (21)$$

We refer to Eq. 21 as the KKRA relation, and to the matrix \mathbf{K}_C as the optimal reduced CTMC.

B. Optimal reduced CTMC from graph transformation

A third expression for the optimal reduced CTMC has recently been derived, which takes an analogous form to the relation between intermicrostate transition rates and MFPTs (Eq. 15),⁹²

$$\mathbf{K}_C = \tilde{\mathcal{T}}_C^{-1} (-\mathbf{1}_N \mathbf{1}_N^\top + \text{diag}(\mathbf{\Pi})^{-1}), \quad (22)$$

where $\tilde{\mathcal{T}}_C$ is a $N \times N$ matrix of *appropriately weighted* intercommunity MFPTs. In Ref. 92, it was shown that for Eq. 22 to hold, $\tilde{\mathcal{T}}_C$ must satisfy a constraint analogous to Eq. 19,

$$\mathbf{\Pi}^\top \tilde{\mathcal{T}}_C = \zeta_C \mathbf{1}_N^\top, \quad (23)$$

where ζ_C is the Kemeny constant of the coarse-grained Markov chain,

$$\zeta_C = (\mathbf{1}_N^\top \tilde{\mathcal{T}}_C^{-1} \mathbf{1}_N)^{-1} = \sum_I \Pi_I [\tilde{\mathcal{T}}_C]_{IJ} \quad \forall J, \quad (24)$$

analogous to Eq. 16. Eq. 24 therefore enforces that the sum of weighted-MFPTs to all possible target communities is the same for all initial communities J .

Ref. 92 showed there exist an infinite number of matrices $\tilde{\mathcal{T}}_C$ that satisfy the Kemeny constant constraint (Eq. 23) and therefore preserve detailed balance, with the form

$$\tilde{\mathcal{T}}_C = \boldsymbol{\theta} - \mathbf{w} \mathbf{1}_N^\top, \quad (25)$$

where

$$\theta_{IJ} = \frac{1}{\Pi_I \Pi_J} \sum_{i \in I} \sum_{j \in J} \pi_i \mathcal{T}_{ij} \pi_j, \quad (26)$$

and \mathbf{w} is an arbitrary column vector. All choices of $\tilde{\mathcal{T}}_C$ consistent with Eqs. 25 and 26 yield the same rate matrix in Eq. 22, which coincides with the optimal reduced CTMC given by the HS (Eq. 20) and KKRA (Eq. 21) expressions.⁹² Note that, since there are an infinite number of valid weighted-MFPT matrices $\tilde{\mathcal{T}}_C$, the Kemeny constants of the original and reduced Markov chains are not necessarily equal, although this condition can be enforced through a specific choice for \mathbf{w} in Eq. 25.⁹²

The “weighted-MFPT”, $[\tilde{\mathcal{T}}_C]_{IJ}$, takes into account dynamics *within* the target community I , and is therefore distinct from the true MFPT $[\mathcal{T}_C]_{IJ} = \Pi_J^{-1} \sum_{j \in J} \mathcal{T}_{Ij} \pi_j$, which measures the mean time at which $I \leftarrow J$ trajectories first hit a boundary node of community I . Therefore, $\tilde{\mathcal{T}}_C$, but not \mathcal{T}_C , yields a coarse-grained Markov chain that correctly preserves the stationary distribution of the original Markov chain and satisfies the detailed balance condition.⁹²

Since the GT algorithm (Eqs. 1 and 2) only preserves the MFPT to a *set* of absorbing nodes, and not the *individual* MFPTs to *particular* absorbing nodes,⁴⁷ the weighted-MFPT cannot be computed via a single GT calculation. That is, the sum over product nodes is preserved in Eq. 6, but the contributions to the sum from sets of paths terminating at particular absorbing nodes are not individually preserved. However, GT can be used to indirectly evaluate the weighted-MFPTs by computing the matrix of all pairwise intermicrostate MFPTs \mathcal{T} appearing in Eq. 26, via $V(V-1)/2$ separate GT calculations, as discussed in Sec. II E. Computation of the optimal reduced CTMC via inversion of the weighted-MFPT matrix (Eq. 22), exploiting the GT algorithm, therefore has approximate time complexity $\mathcal{O}(V^6)$, or $\mathcal{O}(V^5)$ using the LU decomposition formulation of GT,⁷⁹ since obtaining \mathcal{T} via GT is the computational bottleneck.

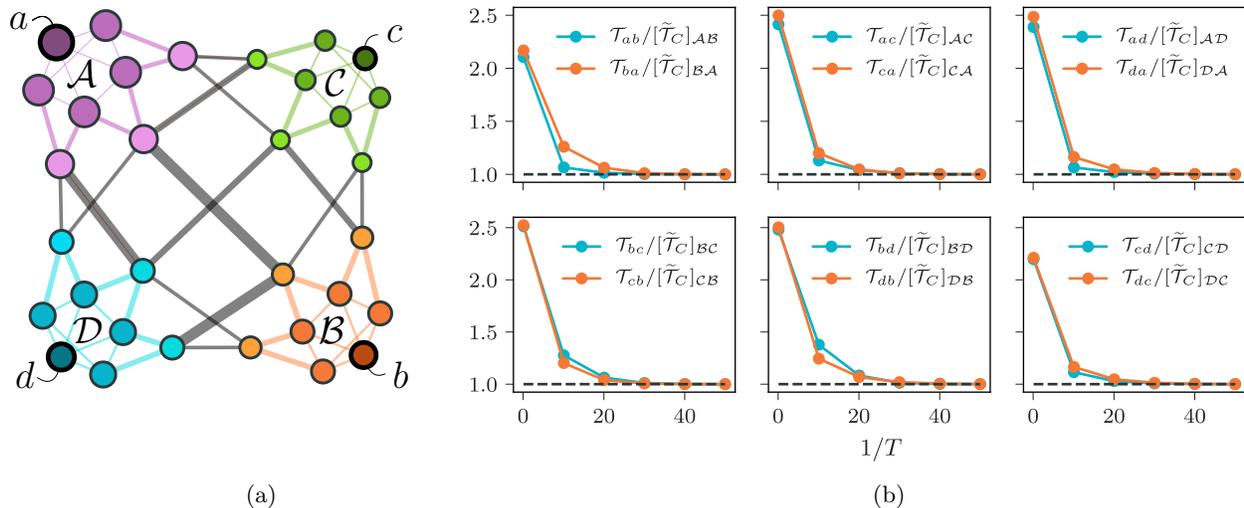


FIG. 3. (a) Model 32-microstate network with four macrostates (communities of microstates), visualized using Gephi.¹⁴⁵ Nodes are colored according to the communities to which they belong. Darker, larger nodes are associated with larger stationary probabilities, and thicker edges indicate slower transition rates. All edges are bidirectional and the network is asymmetric. The lower-case letters label the “attractor” nodes⁸⁷ of each community, defined as the nodes with the largest stationary probability in each community. (b) Ratio of intermicrostate MFPT ($\mathcal{T}_{i^*j^*}$) to intercommunity weighted-MFPT ($[\tilde{\mathcal{T}}_C]_{IJ}$), where $i^* \in I$ and $j^* \in J$ are the attractor nodes of communities I and J . In the strongly metastable regime ($1/T > 20$), $\mathcal{T}_{i^*j^*}$ accurately approximates $[\tilde{\mathcal{T}}_C]_{IJ}$. Since $[\tilde{\mathcal{T}}_C]_{II} = 0 \forall I$, only $N(N-1)/2$ GT calculations, one per pair of communities, are required to compute the approximation to the matrix weighted-MFPT $\tilde{\mathcal{T}}_C$, as opposed to the $V(V-1)/2$ separate GT calculations required to compute $\tilde{\mathcal{T}}_C$ exactly.

C. Approximations to the optimal reduced CTMC

An approximation to the weighted-MFPT expression for the optimal reduced CTMC (Eqs. 22, 25, and 26) can be derived by noting that, in the limit of strongly metastable communities, a single node of a community can be arbitrarily chosen to represent the entire macrostate. To obtain the approximation, we begin by choosing the weighted-MFPT matrix such that the diagonal elements are equal to zero,

$$[\tilde{\mathcal{T}}_C]_{IJ} = \frac{1}{\prod_I \prod_J} \sum_{i \in I} \sum_{j \in J} \pi_i \mathcal{T}_{ij} \pi_j - \frac{1}{\prod_I \prod_I} \sum_{i \in I} \sum_{i' \in I} \pi_{i'} \mathcal{T}_{i'i} \pi_i, \quad (27)$$

as for the matrix of true pairwise intermacrostate MFPTs, \mathcal{T}_C . This choice of $\tilde{\mathcal{T}}_C$ leads to the valid weighted-MFPT matrix that is most readily interpretable, and the choice of \mathbf{w} in Eq. 25 does not affect the computational performance. Hence, we use the definition in Eq. 27 for $\tilde{\mathcal{T}}_C$ throughout the present work.

For a Markov chain that is ideally metastable with respect to the community structure \mathcal{C} , the intermicrostate MFPTs between nodes within a community are vanishingly small compared to the timescale to cross a boundary between communities. That is, $\mathcal{T}_{i'i} \ll \mathcal{T}_{ij}$ and $\mathcal{T}_{j'j} \ll \mathcal{T}_{ij}$ for $i, i' \in I$ and $j, j' \in J \neq I$. It follows

that the $i \leftarrow j$ MFPT will be approximately independent of the particular initial microstate $i \in I$ or target microstate $j \in J$. We can thus approximate each off-diagonal element $[\tilde{\mathcal{T}}_C]_{IJ}$, $I \neq J$, using a single intermicrostate MFPT between an arbitrary pair of nodes $i^* \in I$ and $j^* \in J$, so that

$$[\tilde{\mathcal{T}}_C]_{IJ} \simeq \mathcal{T}_{i^*j^*}. \quad (28)$$

Using this strategy, only one intermicrostate MFPT needs to be computed for each pair of communities, and therefore the total number of GT computations required to obtain the approximate weighted-MFPT matrix is $N(N-1)/2$. Hence, the overall time complexity to obtain the approximation to the optimal reduced CTMC using GT is roughly $\mathcal{O}(V^4 N^2)$.

The approximation in Eq. 28 is similar in spirit to the local equilibrium approximation (LEA),¹⁰⁰ which also invokes the assumption that the dynamics within a community are equilibrated prior to escape.¹⁰¹ In the LEA, the $I \leftarrow J$ intercommunity transition rate is a sum of intermicrostate transition rates across the I - J community boundary, weighted by the relative equilibrium occupation probabilities of the boundary nodes in the source community,

$$[\mathbf{K}_{\text{LEA}}]_{IJ} = \prod_J^{-1} \sum_{i \in I} \sum_{j \in J} k_{ij} \pi_j. \quad (29)$$

The discrete-time analog of Eq. 29, $[\mathbf{T}_{\text{LEA}}(\Delta)]_{IJ}$, is instead a weighted sum of intermicrostate transition proba-

bilities $T_{i \in I, j \in J}(\Delta)$ across the I - J community boundary. The LEA in discrete time is more accurate for longer lag times, since there is then more time for the dynamics to relax to a local equilibrium distribution within the source macrostate. However, the choice of a longer lag time sacrifices the time resolution at which the dynamics are modeled. The continuous time formulation of the LEA corresponds to the limit $\Delta \rightarrow 0$,¹⁰⁰ and is therefore expected to perform poorly if the communities are not strongly metastable. The accuracy of the LEA is sensitive to the precise definition of the community structure \mathcal{C} , because $[\mathbf{K}_{\text{LEA}}]_{IJ}$ depends explicitly on the choice of boundary nodes for communities I and J .¹⁴⁶ In the Supplementary Materials, we describe how the optimal reduced CTMC and the LEA arise from alternative constraints on the correlation functions of the coarse-grained model.

We analyze the accuracy of Eq. 28 for a model 32-microstate network shown in Fig. 3a, for which we compute the optimal reduced CTMC via various routes in Sec. III D. As for the unbranched chain shown in Fig. 2a, the temperature T determines the metastability of this system. We choose the “attractor” nodes⁸⁷ to represent the communities. Although the approximation (Eq. 28) assumes that all $i \in I \leftarrow j \in J \neq I$ MFPTs are approximately equal, it is sensible to choose the representative nodes of the communities to be those that have the largest stationary probabilities and which therefore have a dominant effect on the intercommunity dynamics. While Eq. 28 is a poor approximation at higher effective temperatures, the ratio $\mathcal{T}_{i^*j^*}/[\tilde{\mathcal{T}}_{\mathcal{C}}]_{IJ}$ approaches unity at low temperatures (Fig. 3b). Thus, the approximation leads to weighted-MFPTs that are essentially identical to the values for the optimal reduced CTMC in the strongly metastable regime, and is significantly more accurate than the LEA (see Fig. 4).

D. Numerical comparison of coarse-graining approaches

Although the HS (Eq. 20), KKRA (Eq. 21), and weighted-MFPT (Eq. 22) expressions lead to precisely the same (optimal) coarse-grained rate matrix $\mathbf{K}_{\mathcal{C}}$, in practice, there are important considerations that may lead to one formulation being preferred. Since the MFPT matrix \mathcal{T} is defined for both DTMCs and CTMCs, the KKRA and weighted-MFPT expressions can be applied to discrete-time models, without requiring conversion to a CTMC via Eq. 15.^{23,48} Since \mathcal{T} can be computed via GT, these expressions also invite the possibility of obtaining the optimal reduced CTMC by a method that is more numerically stable than approaches based on linear algebra (Sec. II C).

We analyze the numerical stability of these alternative formulations to obtain the optimal reduced CTMC for the 32-microstate Markov chain illustrated in Fig. 3a. The network is partitioned into four communities, each comprising eight nodes. The communities have inter-

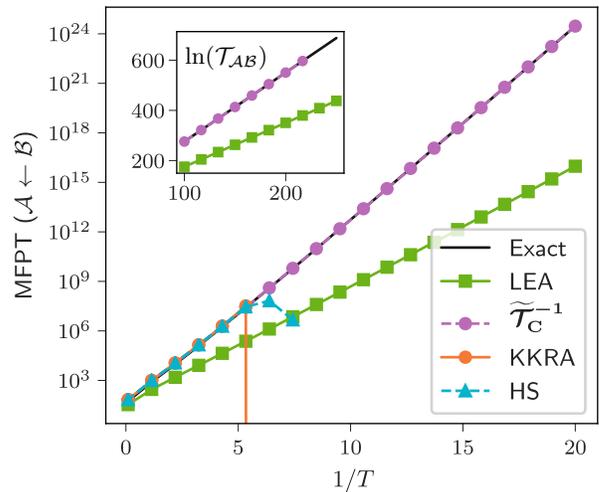


FIG. 4. MFPTs \mathcal{T}_{AB} for the transition from community \mathcal{B} to \mathcal{A} in the 32-microstate model Markov chain illustrated in Fig. 3a, as a function of inverse temperature. The MFPTs for the optimal reduced 4-node CTMC estimated by the HS (Eq. 20), KKRA (Eq. 21), and weighted-MFPT (i.e. via inversion of $\tilde{\mathcal{T}}_{\mathcal{C}}$, cf. Eq. 22) methods are compared to the exact values for the original network, obtained using GT. Evaluation of the optimal intercommunity transition rates via computation of the weighted-MFPT matrix $\tilde{\mathcal{T}}_{\mathcal{C}}$, using GT to compute the matrix of all pairwise intermicrostate MFPTs \mathcal{T} , is more numerically stable than alternative methods to obtain the optimal reduced CTMC. Indeed, this method is numerically stable even in the strongly metastable regime, whereas evaluation of the HS and KKRA expressions fails at the intermediate inverse temperature of $1/T \approx 6$. The $\mathcal{A} \leftarrow \mathcal{B}$ MFPTs for the optimal reduced CTMC are significantly more accurate than for the CTMC given by the local equilibrium approximation (LEA) (Eq. 29).

nal structure; there is a well-defined “attractor” node⁸⁷ for each community, which is connected to neighbouring nodes of the same community by fast transition rates. There is a net flow of probability to the attractor node from all other nodes of the same community when the network is not at equilibrium. Hence, the attractor node has the largest relative stationary probability within a community and acts as a dynamical hub.⁸⁸ Only three nodes are at the boundary of each community, and these nodes are connected to neighboring nodes of the same community via slower transitions, so that the boundaries between communities are not clearly defined. These nonideal properties are designed to mimic the topology of typical Markovian networks representing realistic dynamical processes.^{88,110,147,148}

In Fig. 4, we use the GT algorithm to compute the exact $\mathcal{A} \leftarrow \mathcal{B}$ MFPT as a function of temperature T for the original Markov chain, and compare these values to the corresponding MFPTs obtained from coarse-grained CTMCs corresponding to the optimal and LEA rate ma-

trices. Evaluation of the intermacrostate transition rates given by the LEA (Eq. 29) is trivial, but the approximation systematically underestimates the true $\mathcal{A} \leftarrow \mathcal{B}$ MFPT. Since the LEA only takes into account transitions between boundary nodes, the contribution to the MFPT from the slower intracommunity transitions is neglected. The $\mathcal{A} \leftarrow \mathcal{B}$ MFPT associated with the optimal reduced CTMC is very close to the value for the original Markov chain. Hence, for this system, which lacks a clear separation of timescales between intracommunity and intercommunity transitions, it is necessary to employ more advanced methods for Markovian coarse-graining than the LEA.

Preservation of the intermacrostate MFPTs does not guarantee that the detailed dynamics are accurately represented in a coarse-grained model.¹⁴⁹ To show that the time-dependent macrostate occupation probability distribution vectors $p_I(t) \forall I \in \mathcal{C}$ associated with the original and optimal reduced CTMCs are consistent, we compare trajectories simulated on the original network with propagation of an initial distribution using the optimal reduced rate matrix, $p_I(t) = \sum_{J \in \mathcal{C}} [e^{\mathbf{K}_C t}]_{IJ} p_J(0)$. The results of this Chapman-Kolmogorov test^{52,150} are illustrated in Fig. 5. The test was performed at an effective temperature of $T = 0.15$, where evaluation of \mathbf{K}_C by the HS and KKRA relations becomes numerically unstable (Fig. 4). Trajectories on the original Markov chain were obtained using kinetic path sampling^{79,80} to avoid the problem of flickering within metastable communities.⁸¹ We find that the optimal reduced CTMC provides a close approximation to the simulated distributions $p_I(t)$ for all communities $I \in \mathcal{C}$ of the original network.

Computation of the optimal reduced rate matrix \mathbf{K}_C by inversion of the matrix of weighted-MFPTs $\tilde{\mathcal{T}}_C$, where the required matrix of intermicrostate MFPTs \mathcal{T} is obtained using GT, retains numerical precision even in the strongly metastable regime. In contrast, the matrix inversion operations in the evaluation of the HS and KKRA expressions fail at the intermediate effective temperature of $1/T \approx 5$. We investigate the origin of these numerical issues in Fig. 6, which shows the condition number¹⁵¹ of the matrices that are inverted in the various expressions for the optimal reduced CTMC. From this figure, it is clear that the weighted-MFPT matrix is much less ill-conditioned than the matrices inverted in the evaluation of the HS (Eq. 20) and KKRA (Eq. 21) expressions. Note that the condition number of a matrix is essentially independent of its dimensionality.^{136,151} Hence, use of the KKRA expression, which requires inversion of a $N \times N$ matrix, is not necessarily preferable to evaluation of the HS relation, which requires inversion of a $V \times V$ matrix. However, N is often orders of magnitude smaller than V ,⁸⁷ so it may be feasible to invert the weighted-MFPT matrix, or evaluate the KKRA expression (Eq. 21), using higher precision calculations than for the evaluation of the HS relation (Eq. 22).

It is unsurprising that the weighted-MFPT matrix $\tilde{\mathcal{T}}_C$ is not severely ill-conditioned, since the intercommunity

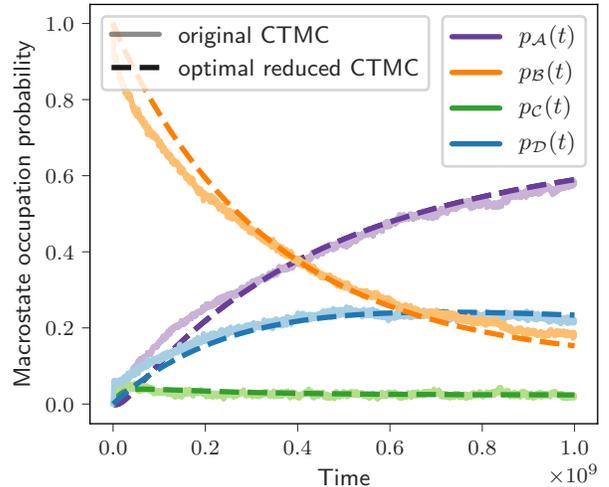


FIG. 5. Time-dependent macrostate occupation probability distributions for each of the communities of the 32-microstate network shown in Fig. 3a, with the initial distribution localized within community \mathcal{B} , at an effective temperature of $T = 0.15$. The distributions derived from simulated trajectories, obtained using kinetic path sampling,^{79–81} on the original Markov chain are compared with the result of propagating the initial distribution according to the optimal coarse-grained rate matrix \mathbf{K}_C , i.e. $p_I(t) = e^{\mathbf{K}_C t} p_I(0)$.

MFPTs should be relatively consistent with one another, corresponding to the observation timescale at which the chosen communities appear metastable. By design, appropriate community detection¹³⁹ algorithms ought to yield macrostates that satisfy this criterion.^{87,101} That is, ideal metastable systems, by definition, belong to the regime in which eigendecomposition or inversion of the original (V -dimensional) Markov chain is numerically unstable, but there exists a community structure \mathcal{C} yielding a N -dimensional matrix $\tilde{\mathcal{T}}_C$ that does not encompass the great disparity of timescales present in the original Markov chain. The condition number of the weighted-MFPT matrix is system-dependent, and will be strongly influenced by the asymmetry of the network and the extent to which the chosen partitioning \mathcal{C} appropriately characterizes the metastable macrostates.

In Fig. 4, the matrix of all pairwise intermicrostate MFPTs \mathcal{T} used in the KKRA and weighted-MFPT expressions was obtained using GT. Fig. S1 of the Supplementary Material shows an analogous figure in which \mathcal{T} is obtained via linear algebra methods (Sec. IIC). Owing to the numerical instability in computing \mathcal{T} by matrix inversion or eigendecomposition operations, the evaluation of the optimal reduced CTMC via Eq. 22 breaks down at $1/T \approx 15$. In contrast, using the GT algorithm to compute \mathcal{T} enables us to access the optimal reduced CTMC in the strongly metastable regime (in this case, $15 < 1/T < 215$).

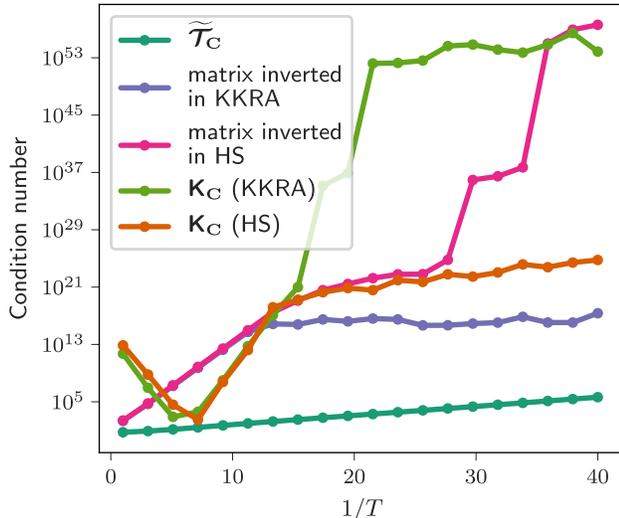


FIG. 6. Condition numbers of the matrices that are inverted in the evaluation of the optimal reduced CTMC \mathbf{K}_C via various expressions, as a function of inverse temperature. Condition numbers are estimated by the GESDD singular value decomposition routine in LAPACK.¹⁵² The weighted-MFPT matrix ($\tilde{\mathcal{T}}_C$, cf. Eqs. 22, 25 and 26) is significantly better-conditioned than the matrix inverted in the KKRA expression (Eq. 21), namely $(\mathbf{\Pi}\mathbf{\Pi}^\top + \mathbf{A}^\top \text{diag}(\boldsymbol{\pi})\mathcal{T}(\boldsymbol{\pi}\mathbf{\Pi}^\top - \text{diag}(\boldsymbol{\pi})\mathbf{A}))$, even though both matrices have the same dimension ($N \times N$). The $V \times V$ matrix inverted in the HS relation (Eq. 20), namely $(\boldsymbol{\pi}\mathbf{1}_V^\top - \mathbf{K})$, is similarly ill-conditioned. Thus, the HS and KKRA expressions to yield \mathbf{K}_C are expected to lose numerical precision at relatively high temperatures compared to the expression based on the weighted-MFPT matrix (Eq. 22).

IV. CONCLUSIONS

We have demonstrated that the graph transformation^{44–49} (GT) algorithm (Eqs. 1 and 2) for computation of the $\mathcal{A} \leftarrow \mathcal{B}$ MFPT in continuous-time Markov chains (CTMCs) can be readily extended to discrete-time Markov chains (DTMCs) (Eqs. 7 and 8). This extension expands the utility of the GT algorithm, since it is typically not possible to convert a DTMC to a corresponding CTMC (Eq. 15).^{59,60} Furthermore, DTMCs are popular models for stochastic processes^{2,7,53–56} because they can often be constructed more easily from trajectory data than an equivalent CTMC representation.^{50,51} Hence, methods for constructing Markovian network models from continuous-state simulation data, such as the Markov State Model (MSM) framework,^{52–56} typically yield a DTMC. DTMCs and CTMCs are similarly ill-conditioned in the metastable regime,^{24–28} where linear algebra methods (Sec. II C) encounter numerical problems arising from finite precision,^{23,48} but the GT algorithm remains numerically stable (Sec. II E). GT therefore provides an attractive approach for computing MFPTs in DTMCs

featuring a separation of characteristic timescales.

Obtaining a reduced representation of a Markov chain aids model interpretation and facilitates more expensive numerical analyses.^{1,61–74} We have shown how the GT algorithm can be utilized to compute the matrix of all pairwise intermicrostate MFPTs, from which the matrix of appropriately weighted intercommunity MFPTs is derived (Eq. 25),⁹² and hence the optimal reduced CTMC via a matrix inversion operation (Eq. 22). We find that this route to computing the optimal reduced CTMC for a given partitioning of the original Markov chain has superior numerical stability to alternative approaches based on linear algebra (Sec. III D). Our approach greatly increases the practicality of optimal Markovian coarse-graining,^{92,100,144} where the existence of metastable macrostates is a necessary condition to yield a reduced Markov chain that accurately represents the dynamics of the original network. We have also suggested an approximation to the optimal reduced CTMC that is valid in the strongly metastable regime (Eq. 28), which can be computed more efficiently than the exact optimal reduced CTMC using the GT algorithm. This approximation is more accurate than the simple local equilibrium approximation (LEA) (Eq. 29).^{100,101} Moreover, our approximation is much less sensitive to the precise definition of the community structure than the LEA,¹⁴⁶ since it is based on a particular choice of internal microstates representing the communities, and does not involve the boundary nodes.

In future work, we will pursue alternative, more scalable, approaches to estimating coarse-grained Markov chains based on the GT algorithm. Possible strategies include using trajectory data obtained from kinetic path sampling (kPS),^{79,80} which extends the GT algorithm with an iterative reverse randomisation procedure to sample the full first passage time distribution.⁸¹ A coarse-grained DTMC or CTMC could then be estimated¹⁴⁶ from maximum-likelihood^{50,57,58} or Gibbs sampling,^{50,51} or from the MFPTs and fluxes between milestones.^{89–91} Alternatively, the GT algorithm could be repurposed to iteratively eliminate microstates that do not contribute significantly to intercommunity dynamics, yielding a renormalized Markov chain that closely approximates the original.¹⁵³ Since GT does *not* preserve the MFPTs to *individual* absorbing nodes,⁴⁷ this approach would require a careful analysis of the error incurred in the dynamical properties of the resulting renormalized network.

The numerical stability of the GT algorithm^{48,106} confers a significant advantage to these dimensionality reduction strategies, and to the method based on exact MFPTs that we have described. Exploiting the GT method in approaches to Markovian coarse-graining therefore provides valuable tools to aid the interpretation and analysis of Markov chains exhibiting rare event dynamics.

V. SUPPLEMENTARY MATERIAL

Further mathematical detail pertaining to the results of the present work can be found in the Supplementary Materials; namely, a formal definition of occupancy number correlation functions, a derivation of the LEA and of the Hummer-Szabo relation for the optimal reduced CTMC (following Refs. 92 and 100), and a statement of the exact recursive formula for the MFPTs in an unbranched CTMC. A figure analogous to Fig. 4, in which linear algebra methods (rather than GT) are used to obtain the matrix of all pairwise intermicrostate MFPTs, is also included in the Supplementary Materials.

We provide various software to perform GT computations for arbitrary discrete- and continuous-time Markov chains, including the dimensionality reduction methodology described in Sec. III, all available online under the GNU General Public License. PyGT, a Python package for graph transformation (GT) computations, including the removal of blocks of nodes, is available at <https://pygt.readthedocs.io/en/latest/>. A Fortran implementation of the iterative GT algorithm is part of the PATHSAMPLE program, available at <http://www-wales.ch.cam.ac.uk/PATHSAMPLE/>. A C++ implementation of the LU decomposition formulation of the GT algorithm is included in the DISCOTRESS (DIScrete State COntinuous Time Rare Event Simulation Suite) software, available at <https://github.com/danieljsharpe/DISCOTRESS>.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

VI. ACKNOWLEDGEMENT

DK gratefully recognises support from the Marshall Scholarship. DJS gratefully acknowledges the Cambridge Commonwealth, European and International Trust for a PhD scholarship. TDS recognises support from the Agence Nationale de Recherche, via the MEMOPAS project ANR-19-CE46-0006-1. DJW gratefully acknowledges support from the EPSRC.

DK and DJS contributed equally to this work.

Appendix A: Glossary of symbols and abbreviations

γ_k	k -th dominant eigenvalue of CTMC
Δ	lag time of DTMC
ζ	Kemeny constant
ζ_C	Kemeny constant of reduced Markov chain
θ	$N \times N$ matrix related to $\tilde{\mathcal{T}}_C$ (Eq. 25)
λ_k	k -th dominant eigenvalue of DTMC
ξ	discrete path
Π	stationary distribution of reduced Markov chain
π	stationary distribution of original Markov chain
τ_j	mean waiting time for the j -th node

ϕ_k	k -th dominant left eigenvector of Markov chain
ψ_k	k -th dominant right eigenvector of Markov chain
\mathbf{A}	$V \times N$ aggregation matrix
\mathcal{A}	set of absorbing nodes
\mathcal{B}	set of initial nodes
CTMC	continuous-time Markov chain
\mathcal{C}	set of communities $\{I, J, \dots\}$
DTMC	discrete-time Markov chain
\mathbf{G}_S	absorbing fundamental matrix in continuous time
GT	graph transformation algorithm
GTH	Grassmann-Taksar-Heyman algorithm
HS	Hummer-Szabo relation (Eq. 20)
\mathbf{I}	identity matrix
\mathcal{I}	set of intervening nodes $\mathcal{I} \equiv (\mathcal{A} \cup \mathcal{B})^c$
i, j	node (microstate) indices
i^*, j^*	attractor node indices
I, J	community (macrostate) indices
\mathbf{K}	transition rate matrix of CTMC
\mathbf{K}'	absorbing transition rate matrix of CTMC
$\mathbf{K}^\#$	group inverse of rate matrix
\mathbf{K}_C	rate matrix for optimal reduced CTMC for given \mathcal{C}
\mathbf{K}_{LEA}	rate matrix for reduced CTMC from the LEA (Eq. 29) for given \mathcal{C}
KKRA	Kells-Koskin-Rosta-Annibale relation (Eq. 21)
LEA	local equilibrium approximation
MFPT	mean first passage time
N	number of communities (macrostates)
\mathbf{N}_S	absorbing fundamental matrix in discrete time for nodes $S \equiv \mathcal{A}^c$
\mathbf{P}	branching probability matrix of CTMC
$\mathbf{p}(0)$	initial occupation probability distribution
\mathbf{p}_n	probability distribution of DTMC after n steps
\mathcal{S}	set of source nodes $S \equiv \mathcal{A}^c$
T	effective temperature
\mathbf{T}_C	optimal reduced DTMC for given \mathcal{C}
$\mathbf{T}(\Delta)$	transition probability matrix of DTMC
\mathcal{T}	$V \times V$ matrix of pairwise intermicrostate MFPTs
\mathcal{T}_{AB}	$\mathcal{A} \leftarrow \mathcal{B}$ MFPT
$\tilde{\mathcal{T}}_C$	$N \times N$ matrix of intercommunity weighted-MFPTs (Eqs. 25 and 26)
\mathcal{W}_ξ	probability of discrete path ξ
V	number of nodes in the original Markov chain
\mathbf{Z}^{CT}	ergodic fundamental matrix in continuous time
\mathbf{Z}^{DT}	ergodic fundamental matrix in discrete time
$\mathbf{1}_V$	V -dimensional column vector with elements equal to unity

Appendix B: Derivation of MFPTs from the absorbing fundamental matrix

In Sec. II C, it was stated that the absorbing rate matrix \mathbf{K}' is associated with an absorbing fundamental matrix,^{1,5} which provides another linear algebra approach to compute MFPTs. In this appendix, we prove this statement in both discrete- and continuous-time.

Consider the MFPT to a target set of nodes \mathcal{A} from the set of source nodes $\mathcal{S} \equiv \mathcal{A}^c$. Using the first-step relation⁶ for a DTMC with uniform lag time Δ and $s' \leftarrow s$ transi-

tion probabilities $T_{s's}$, the $\mathcal{A} \leftarrow s \in \mathcal{S}$ MFPT is

$$\mathcal{T}_{\mathcal{A}s} = \Delta + \sum_{s' \in \mathcal{S}} \mathcal{T}_{\mathcal{A}s'} T_{s's}. \quad (\text{B1})$$

Rearranging, we obtain

$$\Delta = \sum_{s' \in \mathcal{S}} \mathcal{T}_{\mathcal{A}s'} (\delta_{s's} - T_{s's}), \quad (\text{B2})$$

and so the MFPT from a source node s to the absorbing macrostate \mathcal{A} is $[\mathbf{\Delta}_S^\top (\mathbf{I}_S - \mathbf{T}_{SS})^{-1}]_s$, where \mathbf{I}_S is the identity matrix of dimension $|\mathcal{S}|$, the elements of \mathbf{T}_{SS} are $T_{s's} \forall s, s' \in \mathcal{S}$, $\mathbf{\Delta}_S = \Delta \mathbf{1}_S$, and $\mathbf{1}_S$ is a $|\mathcal{S}|$ -dimensional column vector with elements equal to unity. To average over an initial probability distribution within the source macrostate, $\mathbf{p}_S(0)$, we simply write

$$\mathcal{T}_{\mathcal{A}S} = \mathbf{\Delta}_S^\top (\mathbf{I}_S - \mathbf{T}_{SS})^{-1} \mathbf{p}_S(0) \equiv \mathbf{\Delta}_S^\top \mathbf{N}_S \mathbf{p}_S(0), \quad (\text{B3})$$

where \mathbf{N}_S is the fundamental matrix of the absorbing DTMC.^{1,5} The elements $[\mathbf{N}_S]_{s's}$ are interpreted as the mean number of visits to node s' along escape trajectories initialized in node s and terminating at the absorbing boundary of \mathcal{A} . Setting $\Delta = 1$ in Eq. B3, we see that the mean number of steps to the \mathcal{A} region starting from node s is $[\mathbf{1}_S^\top \mathbf{N}_S]_s$.

The $\mathcal{A} \leftarrow \mathcal{S}$ MFPT for a CTMC parameterized by the vector of node-dependent mean waiting times $\boldsymbol{\tau}_S$ and branching probability matrix \mathbf{P}_{SS} for source nodes is

$$\mathcal{T}_{\mathcal{A}S} = \boldsymbol{\tau}_S^\top (\mathbf{I}_S - \mathbf{P}_{SS})^{-1} \mathbf{p}_S(0), \quad (\text{B4})$$

analogous to Eq. B3, where $\mathbf{G}_S \equiv (\mathbf{I}_S - \mathbf{P}_{SS})^{-1}$ is the fundamental matrix of the absorbing CTMC.¹⁰⁵

The formula derived above for $\mathcal{T}_{\mathcal{A}S}$ holds for any distribution of waiting times within \mathcal{S} . If all the waiting times are zero except for $\tau_{s'}$, then the MFPT to \mathcal{A} starting from s is given by $\tau_{s'} [\mathbf{G}_S]_{s's}$, and we can therefore associate this quantity with the time spent in node s' for pathways starting in node s prior to absorption.^{103,154} Hence, the formula derived in Eq. B4 for $\mathcal{T}_{\mathcal{A}S}$ can be interpreted as a weighted average of the *residence times*^{1,5} of all nodes in \mathcal{S} prior to absorption. That is, the MFPT to reach \mathcal{A} starting from node $s \in \mathcal{S}$ is $\sum_{s' \in \mathcal{S}} \tau_{s'} [\mathbf{G}_S]_{s's}$, which defines the residence time in node s .^{103,154}

It is also insightful to derive the time to absorption from the master equation for the occupation probabilities of the nodes in \mathcal{S} ,

$$\dot{\mathbf{p}}_S(t) = \mathbf{K}_{SS} \mathbf{p}_S(t), \quad (\text{B5})$$

where the elements of \mathbf{K}_{SS} can be written as

$$\begin{aligned} K_{s's} &= k_{s's} - \delta_{s's} \sum_m k_{ms} \\ &= (P_{s's} - \delta_{s's}) / \tau_s = -[\mathbf{G}_S^{-1}]_{s's} / \tau_s, \end{aligned} \quad (\text{B6})$$

which is equivalent to $\mathbf{K}_{SS}^{-1} = -\text{diag}(\boldsymbol{\tau}_S) \mathbf{G}_S$. The formal solution to the master equation is $\mathbf{p}_S(t) =$

$\exp(\mathbf{K}_{SS} t) \mathbf{p}_S(0)$. Hence, the expected time spent in node s' before absorption for trajectories initialized from node s is the integral

$$\int_0^\infty \exp(K_{s's} t) dt = -[\mathbf{K}^{-1}]_{s's} = \tau_{s'} [\mathbf{G}_S]_{s's}, \quad (\text{B7})$$

as we obtained by considering $\mathcal{T}_{\mathcal{A}S} = (\boldsymbol{\tau}_S)^\top \mathbf{G}_S \mathbf{p}_S(0)$ above.

Appendix C: Conservation of MFPTs in the discrete-time formulation of the GT algorithm

The mapping in Sec. IIB defines a DTMC with node-dependent lag times, which has the same pathwise averages as a CTMC with corresponding waiting times and branching probabilities. Hence, in analogy to the argument for the continuous-time case, the GT procedure applied to a DTMC conserves the MFPT averaged over a set of target nodes.^{45,47} This result can be shown by treating \mathcal{A} as an absorbing boundary and analyzing the master equation of the absorbing Markov chain,¹⁰⁵ or by breaking down the sum over $\mathcal{A} \leftarrow \mathcal{S}$ pathways into segments that travel through different regions of the network.^{45,47,105} In this appendix, we follow the latter approach to analyze pathwise sums in discrete time.

We first note that the elements of the fundamental matrix $\mathbf{N}_S = (\mathbf{I}_S - \mathbf{T}_{SS})^{-1}$ for the source region \mathcal{S} , $[\mathbf{N}_S]_{s's}$, correspond to the sum of probabilities for all paths within \mathcal{S} that start at node s and end at node s' . This result can be proved by writing the $s' \leftarrow s$ path probability as a product of transition probabilities, and summing over paths of arbitrary length by exponentiating the transition matrix: $\mathbf{N}_S = \sum_{n=0}^\infty \mathbf{T}_{SS}^n = (\mathbf{I}_S - \mathbf{T}_{SS})^{-1}$.¹⁰⁵ To obtain the probability for paths that start in \mathcal{S} and end with absorption in \mathcal{A} , we need to combine the \mathbf{N}_S matrix with the branching probability from \mathcal{S} to \mathcal{A} . Hence, the overall probability for all paths that begin at node $s \in \mathcal{S}$ and end at a particular absorbing node $a \in \mathcal{A}$ is $[\mathbf{T}_{\mathcal{A}S} \mathbf{N}_S]_{as}$. The sum over target nodes, $[\mathbf{1}_{\mathcal{A}}^\top \mathbf{T}_{\mathcal{A}S} \mathbf{N}_S]_s$, is unity for all $s \in \mathcal{S}$, since all probability flow ends in \mathcal{A} ; the corresponding result for a CTMC is $[\mathbf{1}_{\mathcal{A}}^\top \mathbf{P}_{\mathcal{A}S} \mathbf{G}_S]_s = 1 \forall s \in \mathcal{S}$.

The $\mathcal{A} \leftarrow \mathcal{S}$ MFPT, $\mathcal{T}_{\mathcal{A}S}$, is the accumulated sum of lag times for each node visited along $\mathcal{A} \leftarrow \mathcal{S}$ paths, where each path is weighted by the pathway probability. We can calculate this sum by defining a modified transition matrix $\tilde{\mathbf{T}}_{SS}$ with elements $\tilde{T}_{s's} \exp(\zeta \Delta_s)$, where Δ_s is the lag time for node s , and the lag times for nodes are not necessarily uniform. Differentiating the product $\mathbf{1}_{\mathcal{A}}^\top \tilde{\mathbf{T}}_{\mathcal{A}S} \tilde{\mathbf{N}}_S$ with respect to ζ and then setting $\zeta = 0$ produces the corresponding sum of waiting times for each path, multiplied by the path probability.^{45,47}

The required derivatives are¹⁰⁵

$$\frac{\partial}{\partial \zeta} \tilde{\mathbf{T}}_{\mathcal{A}\mathcal{S}} \Big|_{\zeta=0} = \mathbf{T}_{\mathcal{A}\mathcal{S}} \text{diag}(\boldsymbol{\Delta}_{\mathcal{S}}), \quad (\text{C1})$$

$$\frac{\partial}{\partial \zeta} \tilde{\mathbf{N}}_{\mathcal{S}} \Big|_{\zeta=0} = \mathbf{N}_{\mathcal{S}} \mathbf{T}_{\mathcal{S}\mathcal{S}} \text{diag}(\boldsymbol{\Delta}_{\mathcal{S}}) \mathbf{N}_{\mathcal{S}}, \quad (\text{C2})$$

so that

$$\begin{aligned} \frac{\partial}{\partial \zeta} \mathbf{1}_{\mathcal{A}}^{\top} \tilde{\mathbf{T}}_{\mathcal{A}\mathcal{S}} \tilde{\mathbf{N}}_{\mathcal{S}} \Big|_{\zeta=0} &= \mathbf{1}_{\mathcal{A}}^{\top} \mathbf{T}_{\mathcal{A}\mathcal{S}} \text{diag}(\boldsymbol{\Delta}_{\mathcal{S}}) \mathbf{N}_{\mathcal{S}} \\ &\quad + \mathbf{1}_{\mathcal{A}}^{\top} \mathbf{T}_{\mathcal{A}\mathcal{S}} \mathbf{N}_{\mathcal{S}} \mathbf{T}_{\mathcal{S}\mathcal{S}} \text{diag}(\boldsymbol{\Delta}_{\mathcal{S}}) \mathbf{N}_{\mathcal{S}} \\ &= (\mathbf{1}_{\mathcal{A}}^{\top} \mathbf{T}_{\mathcal{A}\mathcal{S}} + \mathbf{1}_{\mathcal{S}}^{\top} \mathbf{T}_{\mathcal{S}\mathcal{S}}) \text{diag}(\boldsymbol{\Delta}_{\mathcal{S}}) \mathbf{N}_{\mathcal{S}} \\ &= \mathbf{1}_{\mathcal{S}}^{\top} \text{diag}(\boldsymbol{\Delta}_{\mathcal{S}}) \mathbf{N}_{\mathcal{S}} = \boldsymbol{\Delta}_{\mathcal{S}}^{\top} \mathbf{N}_{\mathcal{S}}, \quad (\text{C3}) \end{aligned}$$

where we have used $\mathbf{1}_{\mathcal{A}}^{\top} \mathbf{T}_{\mathcal{A}\mathcal{S}} \mathbf{N}_{\mathcal{S}} = \mathbf{1}_{\mathcal{S}}^{\top}$. Therefore $\mathcal{T}_{\mathcal{A}\mathcal{S}} = \boldsymbol{\Delta}_{\mathcal{S}}^{\top} \mathbf{N}_{\mathcal{S}} \mathbf{p}_{\mathcal{S}}(0)$, as in Eq. B3. The equivalent result for a CTMC is $\mathcal{T}_{\mathcal{A}\mathcal{S}} = \boldsymbol{\tau}_{\mathcal{S}}^{\top} \mathbf{G}_{\mathcal{S}} \mathbf{p}_{\mathcal{S}}(0)$.

Now we divide the source region $\mathcal{S} \equiv \mathcal{I} \cup \mathcal{B}$ into the initial macrostate \mathcal{B} and the set of nodes to be eliminated \mathcal{I} , and derive the $\mathcal{A} \leftarrow b$ pathway probabilities for all nodes $b \in \mathcal{B}$. Pathways initialized in the macrostate \mathcal{B} either proceed directly to \mathcal{A} or via a detour of arbitrary length through the region \mathcal{I} ; the total transition probability to \mathcal{A} can thus be written as $\mathbf{T}_{\mathcal{A}\mathcal{B}} + \mathbf{T}_{\mathcal{A}\mathcal{I}} \mathbf{N}_{\mathcal{I}} \mathbf{T}_{\mathcal{I}\mathcal{B}} \equiv \tilde{\mathbf{T}}_{\mathcal{A}\mathcal{B}}^{\mathcal{I}}$, where the factor $\mathbf{N}_{\mathcal{I}}$ is the sum of pathway probabilities for all paths in \mathcal{I} starting and finishing in nodes $i, i' \in \mathcal{I}$. The complete sum of $\mathcal{A} \leftarrow \mathcal{B}$ path probabilities must also take into account non-reactive paths that start and finish in \mathcal{B} without leaving $\mathcal{B} \cup \mathcal{I}$, i.e. without reaching \mathcal{A} . This factor is $\mathbf{N}_{\mathcal{B}}^{\mathcal{I}} = (\mathbf{I}_{\mathcal{B}} - \mathbf{T}_{\mathcal{B}\mathcal{B}}^{\mathcal{I}})^{-1}$, where $\mathbf{T}_{\mathcal{B}\mathcal{B}}^{\mathcal{I}} \equiv \mathbf{T}_{\mathcal{B}\mathcal{B}} + \mathbf{T}_{\mathcal{B}\mathcal{I}} \mathbf{N}_{\mathcal{I}} \mathbf{T}_{\mathcal{I}\mathcal{B}}$. Hence, $[\mathbf{1}_{\mathcal{A}}^{\top} \tilde{\mathbf{T}}_{\mathcal{A}\mathcal{B}}^{\mathcal{I}} \mathbf{N}_{\mathcal{B}}^{\mathcal{I}}]_b$ is the sum of pathway probabilities over all possible $a \in \mathcal{A} \leftarrow b$ paths, i.e. $[\mathbf{1}_{\mathcal{A}}^{\top} \tilde{\mathbf{T}}_{\mathcal{A}\mathcal{B}}^{\mathcal{I}} \mathbf{N}_{\mathcal{B}}^{\mathcal{I}}]_b = \sum_{a \in \mathcal{A}} \sum_{\xi \in a \leftarrow b} \mathcal{W}_{\xi}$, which is unity for all $b \in \mathcal{B}$.¹⁰⁵

The MFPT $\mathcal{T}_{\mathcal{A}\mathcal{B}}$ can again be calculated by differentiating the product $\mathbf{1}_{\mathcal{A}}^{\top} \tilde{\mathbf{T}}_{\mathcal{A}\mathcal{B}}^{\mathcal{I}} \tilde{\mathbf{N}}_{\mathcal{B}}^{\mathcal{I}}$ with respect to ζ and then setting $\zeta = 0$. The result is¹⁰⁵

$$\begin{aligned} \mathcal{T}_{\mathcal{A}\mathcal{B}} &= \left(\boldsymbol{\Delta}_{\mathcal{B}}^{\top} + \boldsymbol{\Delta}_{\mathcal{I}}^{\top} \mathbf{N}_{\mathcal{I}} \mathbf{T}_{\mathcal{I}\mathcal{B}} \right) \\ &\quad \left[\mathbf{I}_{\mathcal{B}} - (\mathbf{T}_{\mathcal{B}\mathcal{B}} + \mathbf{T}_{\mathcal{B}\mathcal{I}} \mathbf{N}_{\mathcal{I}} \mathbf{T}_{\mathcal{I}\mathcal{B}}) \right]^{-1} \mathbf{p}_{\mathcal{B}}(0) \\ &= (\boldsymbol{\Delta}_{\mathcal{B}}^{\mathcal{I}})^{\top} (\mathbf{I}_{\mathcal{B}} - \mathbf{T}_{\mathcal{B}\mathcal{B}}^{\mathcal{I}})^{-1} \mathbf{p}_{\mathcal{B}}(0) \\ &\equiv (\boldsymbol{\Delta}_{\mathcal{B}}^{\mathcal{I}})^{\top} \mathbf{N}_{\mathcal{B}}^{\mathcal{I}} \mathbf{p}_{\mathcal{B}}(0), \quad (\text{C4}) \end{aligned}$$

with $\boldsymbol{\Delta}_{\mathcal{B}}^{\mathcal{I}}$ the renormalized lag time vector in the graph-transformed absorbing DTMC where all nodes in \mathcal{I} have been eliminated (Eq. 7), and $\mathbf{T}_{\mathcal{B}\mathcal{B}}^{\mathcal{I}}$ is the corresponding GT-renormalized transition probability matrix (Eq. 8). Hence, we can apply the GT procedure to remove all the nodes in the \mathcal{I} region and recover the same MFPT $\mathcal{T}_{\mathcal{A}\mathcal{B}}$ if the initial occupation probability distribution $\mathbf{p}_{\mathcal{S}}(0)$ is localized in \mathcal{B} .

Appendix D: Fundamental matrices in continuous-time

The choice of ergodic fundamental matrix \mathbf{Z}^{CT} in continuous-time^{111,112} used in Eq. 12 is such that the Ke-

meny constant is obtained directly as the negative trace, $\zeta = -\text{Tr}(\mathbf{Z}^{\text{CT}})$ (Eq. 17). In fact, there is some freedom in the choice of fundamental matrix, from which all key dynamical properties of the Markov chain can be computed, including the matrix of pairwise intermicrostate MFPTs, \mathcal{T} (Eq. 11), and the Kemeny constant.⁹³ Eq. 12 is a particular case of the group inverse¹⁰ $\mathbf{K}^{\#}$ of the transition rate matrix \mathbf{K} , which in general has the form

$$\begin{aligned} \mathbf{K}^{\#} &= - \sum_{k>1} \frac{1}{|\gamma_k|} \boldsymbol{\psi}_k \otimes \boldsymbol{\phi}_k \\ &= (\alpha \boldsymbol{\pi} \mathbf{1}_V^{\top} + \mathbf{K})^{-1} - \frac{1}{\alpha} \boldsymbol{\pi} \mathbf{1}_V^{\top}, \quad (\text{D1}) \end{aligned}$$

for any real $\alpha \neq 0$. Eq. D1 follows from eigendecomposition of the rate matrix,⁵⁷ subsequent addition of a term $\alpha(\boldsymbol{\psi}^{(1)} \otimes \boldsymbol{\phi}^{(1)})$, shifting the zero eigenvalue of \mathbf{K} (namely γ_1 , with corresponding left eigenvector $\boldsymbol{\phi}^{(1)} = \mathbf{1}_V^{\top}$ and right eigenvector $\boldsymbol{\psi}^{(1)} = \boldsymbol{\pi}$) to α , and inversion of the resulting matrix. As for the rate matrix \mathbf{K} , all other eigenvalues of the group inverse $\mathbf{K}^{\#}$ are negative, and the eigenvectors satisfy $\boldsymbol{\phi}_l \boldsymbol{\psi}_k = \delta_{lk}$.

The group inverse of the rate matrix (Eq. D1) satisfies several relations:

$$\begin{aligned} \mathbf{K} \mathbf{K}^{\#} \mathbf{K} &= \left(\sum_{k=1}^V \gamma_k \boldsymbol{\psi}_k \otimes \boldsymbol{\phi}_k \right) \left(\sum_{l=2}^V \boldsymbol{\psi}_l \otimes \boldsymbol{\phi}_l \right) \\ &= \sum_{k=1}^V \gamma_k \boldsymbol{\psi}_k \otimes \boldsymbol{\phi}_k = \mathbf{K}, \quad (\text{D2a}) \end{aligned}$$

$$\begin{aligned} \mathbf{K}^{\#} \mathbf{K} \mathbf{K}^{\#} &= \left(\sum_{k=2}^V \boldsymbol{\psi}_k \otimes \boldsymbol{\phi}_k \right) \left(\sum_{l=2}^V \frac{1}{\gamma_l} \boldsymbol{\psi}_l \otimes \boldsymbol{\phi}_l \right) \\ &= \sum_{l=2}^V \frac{1}{\gamma_l} \boldsymbol{\psi}_l \otimes \boldsymbol{\phi}_l = \mathbf{K}^{\#}, \quad (\text{D2b}) \end{aligned}$$

$$\begin{aligned} \mathbf{K}^{\#} \mathbf{K} &= \left(\sum_{l=2}^V \frac{1}{\gamma_l} \boldsymbol{\psi}_l \otimes \boldsymbol{\phi}_l \right) \left(\sum_{k=1}^V \gamma_k \boldsymbol{\psi}_k \otimes \boldsymbol{\phi}_k \right) \\ &= \sum_{l=2}^V \boldsymbol{\psi}_l \otimes \boldsymbol{\phi}_l = \mathbf{K} \mathbf{K}^{\#}. \quad (\text{D2c}) \end{aligned}$$

To appreciate the utility of the group inverse $\mathbf{K}^{\#}$, consider the action of $\mathbf{K}^{\#} \mathbf{K}$ or $\mathbf{K} \mathbf{K}^{\#}$ on an arbitrary vector \mathbf{x} expanded in the eigenvector basis as $\mathbf{x} = \sum_k \boldsymbol{\psi}_k (\boldsymbol{\phi}_k \mathbf{x})$:

$$\begin{aligned} \mathbf{K}^{\#} \mathbf{K} \mathbf{x} &= \left(\sum_{l=2}^V \boldsymbol{\psi}_l \otimes \boldsymbol{\phi}_l \right) \left(\sum_k \boldsymbol{\psi}_k (\boldsymbol{\phi}_k \mathbf{x}) \right) \\ &= \sum_{l=2}^V \boldsymbol{\psi}_l (\boldsymbol{\phi}_l \mathbf{x}) = \mathbf{x} - \boldsymbol{\psi}^{(1)} (\boldsymbol{\phi}^{(1)} \mathbf{x}) \\ &= \mathbf{x} - \boldsymbol{\pi} (\mathbf{1}_V^{\top} \mathbf{x}). \quad (\text{D3}) \end{aligned}$$

Hence, $\mathbf{K}^{\#} \mathbf{K} \mathbf{x} = (\mathbf{I}_V - \boldsymbol{\pi} \mathbf{1}_V^{\top}) \mathbf{x}$ or $\mathbf{x} = \mathbf{K}^{\#} \mathbf{K} \mathbf{x} + \boldsymbol{\pi} \mathbf{1}_V^{\top} \mathbf{x}$, which we use in Appendix E to derive expressions for MFPTs and the Kemeny constant in terms of the elements of the group inverse.

Any *generalized inverse* of \mathbf{K} , defined as satisfying the condition Eq. D2a, is a *fundamental* matrix of the Markov chain in the sense that key dynamical quantities, including \mathcal{T} , can be written straightforwardly in terms of the matrix (and the stationary distribution),¹⁵⁵ as we show is true for the group inverse $\mathbf{K}^\#$ (Eq. D1) in Appendix E. In discrete-time, the same logic applies to generalized inverses of the Markovian kernel $\mathbf{I} - \mathbf{T}(\Delta)$.^{156–159}

Appendix E: MFPTs and the Kemeny constant from fundamental matrices

To derive the Kemeny constant from the group inverse, we begin by constructing an absorbing Markov chain for a single target state p , such that all transitions out of p are set to zero, i.e. $K_{ip} = 0 \quad \forall i$. The dynamics in the set $\bar{p} = \{j, j \neq p\}$ comprising the $V - 1$ states aside from p are preserved prior to the absorption event. We consider an initial distribution where the probability density is localized at a single node, q , such that $\mathbf{p}_{\bar{p}}(0) = \hat{\mathbf{e}}_q$ where $[\hat{\mathbf{e}}_q]_k = \delta_{qk}$. Then the $p \leftarrow q$ MFPT is given by $\mathcal{T}_{pq} = \mathbf{1}_{\bar{p}}^\top \mathbf{x}_{\bar{p}}$, where $\mathbf{x}_{\bar{p}}$ is the solution vector of dimension $|V - 1|$ to the linear system $\mathbf{K}_{\bar{p}\bar{p}} \mathbf{x}_{\bar{p}} = -\hat{\mathbf{e}}_q$ and $\mathbf{K}_{\bar{p}\bar{p}} \equiv \mathbf{K}^*$ is the absorbing rate matrix describing transitions in \bar{p} prior to absorption (Sec. II C).

We can also write $\mathcal{T}_{pq} = \mathbf{1}_V^\top \mathbf{x}$, where \mathbf{x} is now a V -dimensional column vector that solves the modified linear system involving the original rate matrix \mathbf{K} ,

$$\mathbf{K}\mathbf{x} = \begin{bmatrix} \mathbf{K}_{\bar{p}\bar{p}} & \mathbf{K}_{\bar{p}p} \\ \mathbf{K}_{p\bar{p}} & \mathbf{K}_{pp} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\bar{p}} \\ x_p \end{bmatrix} = -\hat{\mathbf{e}}_q + \mu \hat{\mathbf{e}}_p, \quad (\text{E1})$$

where μ is to be determined and $\hat{\mathbf{e}}_q$ and $\hat{\mathbf{e}}_p$ are of dimension V . Here, we have partitioned the rate matrix \mathbf{K} into the non-absorbing region \bar{p} and the absorbing region $\{p\}$. The top row of this matrix equation reads $\mathbf{K}_{\bar{p}\bar{p}} \mathbf{x}_{\bar{p}} + \mathbf{K}_{\bar{p}p} x_p = -\hat{\mathbf{e}}_q$. Since $\mathbf{K}_{\bar{p}\bar{p}} \mathbf{x}_{\bar{p}} = -\hat{\mathbf{e}}_q$, this result implies that $x_p = 0$ and $\mu = \mathbf{K}_{p\bar{p}} \mathbf{x}_{\bar{p}}$, a scalar. Furthermore, the columns of \mathbf{K} sum to zero, and hence we can also use the fact that $\mathbf{K}_{p\bar{p}} = -\mathbf{1}_{V-1}^\top \mathbf{K}_{\bar{p}\bar{p}}$ to obtain $\mu = -\mathbf{1}_{V-1}^\top \mathbf{K}_{\bar{p}\bar{p}} \mathbf{x}_{\bar{p}} = \mathbf{1}_{V-1}^\top \mathbf{K}_{\bar{p}\bar{p}} \mathbf{K}_{\bar{p}\bar{p}}^{-1} \hat{\mathbf{e}}_q = 1$. Thus, we have

$$\mathbf{K}\mathbf{x} = \hat{\mathbf{e}}_p - \hat{\mathbf{e}}_q, \quad \text{with} \quad \hat{\mathbf{e}}_p \cdot \mathbf{x} = 0, \quad \mathcal{T}_{pq} = \mathbf{1}_V^\top \mathbf{x}. \quad (\text{E2})$$

Now we can use the identity $\mathbf{x} = \mathbf{K}^\# \mathbf{K}\mathbf{x} + \boldsymbol{\pi} \mathbf{1}_V^\top \mathbf{x}$ derived above and substitute for $\mathbf{K}\mathbf{x}$ and $\mathbf{1}_V^\top \mathbf{x}$ to obtain

$$\mathbf{x} = \mathbf{K}^\# [\hat{\mathbf{e}}_p - \hat{\mathbf{e}}_q] + \mathcal{T}_{pq} \boldsymbol{\pi}. \quad (\text{E3})$$

Since the p component of \mathbf{x} vanishes, we have $[\mathbf{K}^\#]_{pp} - [\mathbf{K}^\#]_{pq} + \mathcal{T}_{pq} \pi_p = 0$, or

$$\mathcal{T}_{pq} = \frac{[\mathbf{K}^\#]_{pq} - [\mathbf{K}^\#]_{pp}}{\pi_p}, \quad (\text{E4})$$

which is Eq. 11. Hence, we have proved that the matrix of pairwise intermicrostate MFPTs can be computed from any group inverse (Eq. D1).^{10,15}

To derive the Kemeny constant, we observe that summing over p and converting to a matrix formulation gives

$$\boldsymbol{\pi}^\top \mathcal{T} = \mathbf{1}_V^\top \mathbf{K}^\# - \text{Tr}(\mathbf{K}^\#) \mathbf{1}_V^\top = -\text{Tr}(\mathbf{K}^\#) \mathbf{1}_V^\top, \quad (\text{E5})$$

where we have noted from the definition of the group inverse (Eq. D1) that $\mathbf{1}_V^\top \mathbf{K}^\#$ vanishes by orthogonality, since $\mathbf{1}_V^\top$ is the left eigenvector $\phi^{(1)}$. This formulation highlights a key property of the row vector $\boldsymbol{\pi}^\top \mathcal{T}$, namely that its average value over any normalised probability distribution is simply $-\text{Tr}(\mathbf{K}^\#)$, independent of the components. Hence, for arbitrary $\mathbf{p}(0)$, we have $\boldsymbol{\pi}^\top \mathcal{T} \mathbf{p}(0) = -\text{Tr}(\mathbf{K}^\#) \equiv \zeta$, which defines the Kemeny constant, ζ . That is, *any* initial occupation probability distribution $\mathbf{p}(0)$ will decay to the stationary distribution on an average timescale that is the Kemeny constant, ζ , which therefore represents an average mixing time.^{115,160}

- 1 J. G. Kemeny and J. L. Snell, *Finite Markov Chains* (Van Nostrand, New Jersey, USA, 1960).
- 2 D. T. Gillespie, *Markov Processes: An Introduction for Physical Scientists* (Academic Press, New York, USA, 1992).
- 3 N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, Netherlands, 1992).
- 4 J. R. Norris, *Markov Chains* (Cambridge University Press, New York, USA, 1997).
- 5 C. M. Grinstead and J. L. Snell, *Introduction to Probability* (American Mathematical Society, Providence, Rhode Island, 1997).
- 6 H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, 3rd ed. (Academic Press, London, UK, 1998).
- 7 L. J. S. Allen, *An Introduction to Stochastic Processes with Applications to Biology* (Prentice Hall, Upper Saddle River, New Jersey, 2003).
- 8 L. J. S. Allen, "An introduction to stochastic epidemic models," in *Mathematical Epidemiology*, edited by F. Brauer, P. van den Driessche, and J. Wu (Springer-Verlag, Berlin, 2008) pp. 81–130.
- 9 J. Goutsias and G. Jenkinson, *Phys. Rep.* **529**, 199–264 (2013).
- 10 C. D. Meyer Jr, *SIAM Rev.* **17**, 443–464 (1975).
- 11 W. K. Grassmann, M. I. Taksar, and D. P. Heyman, *Oper. Res.* **33**, 1107–1116 (1985).
- 12 J. Kohlas, *Zeit. Oper. Res.* **30**, 197–207 (1986).
- 13 J. J. Hunter, *Spec. Matrices* **4**, 151–175 (2016).
- 14 J. J. Hunter, *Linear Algebra Appl.* **511**, 176–202 (2016).
- 15 J. J. Hunter, *Linear Algebra Appl.* **549**, 100–122 (2018).
- 16 T. Dayar and N. Akar, *SIAM J. Matrix Anal. Appl.* **27**, 396–412 (2005).
- 17 Z. Zhang, A. Julaiti, B. Hou, H. Zhang, and G. Chen, *Eur. Phys. J. B* **84**, 691–697 (2011).
- 18 Z. Zhang, Y. Sheng, Z. Hu, and G. Chen, *Chaos* **22**, 043129 (2012).
- 19 Z. Zhang, T. Shan, and G. Chen, *Phys. Rev. E* **87**, 012112 (2013).
- 20 N. F. Polizzi, M. J. Therien, and D. N. Beratan, *Isr. J. Chem.* **56**, 816–824 (2016).
- 21 M. Torchala, P. Chelminiak, M. Kurzynski, and P. A. Bates, *BMC Systems Biology* **7**, 130 (2013).
- 22 D. J. Bicout and A. Szabo, *J. Chem. Phys.* **106**, 10292–10298 (1997).
- 23 T. J. Frankcombe and S. C. Smith, *Theor. Chem. Acc.* **124**, 303–317 (2009).
- 24 D. P. Heyman and A. Reeves, *ORSA J. Comp.* **1**, 52–60 (1989).
- 25 B. Philippe, Y. Saad, and W. J. Stewart, *Oper. Res.* **40**, 1156–1179 (1992).
- 26 C. D. Meyer, *SIAM J. Matrix Anal. Appl.* **15**, 715–728 (1994).
- 27 D. P. Heyman and D. P. O’Leary, *SIAM J. Matrix Anal. Appl.* **19**, 534–540 (1998).

- 28 J. L. Barlow, *SIAM J. Matrix Anal. Appl.* **22**, 230–241 (2000).
- 29 T. D. Swinburne, D. Kannan, D. J. Sharpe, and D. J. Wales, *J. Chem. Phys.* **153**, 134115 (2020).
- 30 D. J. Aldous and M. Brown, “Inequalities for rare events in time-reversible Markov chains I,” in *IMS Lecture Notes in Statistics, Vol. 22: Stochastic Inequalities*, edited by M. Shaked and Y. L. Tong (Institute of Mathematical Statistics, Ohio, USA, 1992) pp. 1–16.
- 31 P. Heidelberger, *ACM Trans. Model. Comput. Simul.* **5**, 43–85 (1995).
- 32 P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic, *Oper. Res.* **47**, 585–600 (1999).
- 33 A. Bovier, M. Eckhoff, V. Gayraud, and M. Klein, *J. Phys. A.: Math. Gen.* **33**, L447–L451 (2000).
- 34 A. Bovier, M. Eckhoff, V. Gayraud, and M. Klein, *Commun. Math. Phys.* **228**, 219–255 (2002).
- 35 S. Juneja and P. Shahabuddin, *Manage. Sci.* **47**, 547–562 (2001).
- 36 J. Beltrán and C. Landim, *J. Stat. Phys.* **140**, 1065–1114 (2010).
- 37 E. Vanden-Eijnden and J. Weare, *Commun. Pure Appl. Math* **65**, 1770–1803 (2012).
- 38 O. Benois, C. Landim, and M. Mourragui, *J. Stat. Phys.* **153**, 967–990 (2013).
- 39 C. Hartmann, R. Banisch, M. Sarich, T. Badoewski, and C. Schütte, *Entropy* **16**, 350–376 (2014).
- 40 M. Sarich, R. Banisch, C. Hartmann, and C. Schütte, *Entropy* **16**, 258–286 (2014).
- 41 M. K. Cameron, *J. Chem. Phys.* **141**, 184113 (2014).
- 42 T. Gan and M. Cameron, *J. Nonlinear Sci.* **27**, 927–972 (2017).
- 43 F. Bouchet, J. Rolland, and J. Wouters, *Chaos* **29**, 080402 (2019).
- 44 S. A. Trygubenko and D. J. Wales, *Mol. Phys.* **104**, 1497–1507 (2006).
- 45 S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.* **124**, 234110 (2006).
- 46 D. J. Wales, *Int. Rev. Phys. Chem.* **25**, 237–282 (2006).
- 47 D. J. Wales, *J. Chem. Phys.* **130**, 204111 (2009).
- 48 J. D. Stevenson and D. J. Wales, *J. Chem. Phys.* **141**, 041104 (2014).
- 49 R. S. MacKay and J. D. Robinson, *Phil. Trans. Roy. Soc. A* **376**, 20170232 (2018).
- 50 B. Trendelkamp-Schroer and F. Noé, *J. Chem. Phys.* **138**, 164113 (2013).
- 51 B. Trendelkamp-Schroer, H. Wu, F. Paul, and F. Noé, *J. Chem. Phys.* **143**, 174101 (2015).
- 52 J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, *J. Chem. Phys.* **134**, 174105 (2011).
- 53 G. R. Bowman, V. S. Pande, and F. Noé, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, 1st ed. (Springer, Netherlands, 2014).
- 54 V. S. Pande, K. Beauchamp, and G. R. Bowman, *Methods* **52**, 99–105 (2010).
- 55 B. E. Husic and V. S. Pande, *J. Am. Chem. Soc.* **140**, 2386–2896 (2018).
- 56 A. Mardt, L. Pasquali, H. Wu, and F. Noé, *Nat. Commun.* **9**, 5 (2018).
- 57 N.-V. Buchete and G. Hummer, *J. Phys. Chem. B* **112**, 6057–6069 (2008).
- 58 R. T. McGibbon and V. S. Pande, *J. Chem. Phys.* **143**, 034109 (2015).
- 59 D. T. Crommelin and E. Vanden-Eijnden, *J. Comput. Phys.* **217**, 782–805 (2006).
- 60 P. Metzner, E. Dittmer, T. Jahnke, and C. Schütte, *J. Comput. Phys.* **227**, 353–375 (2007).
- 61 D. R. Barr and M. U. Thomas, *Oper. Res.* **25**, 1028–1031 (1977).
- 62 F. Ball and G. F. Yeo, *J. Appl. Probab.* **30**, 518–528 (1993).
- 63 P. Buchholz, *J. Appl. Probab.* **31**, 59–75 (1994).
- 64 T. Dayar and W. J. Stewart, *SIAM J. Matrix Anal. Appl.* **18**, 482–498 (1997).
- 65 S. Derisavi, H. Hermanns, and W. H. Sanders, *Inf. Proc. Lett.* **87**, 309–315 (2003).
- 66 W. E, T. Li, and E. Vanden-Eijnden, *Proc. Natl. Acad. Sci. USA* **105**, 7907–7912 (2008).
- 67 V. B. Yap, *J. Appl. Probab.* **46**, 497–506 (2009).
- 68 A. M. S. Barreto and M. D. Fragoso, *IFAC Proceedings Volumes* **44**, 4206–4211 (2011).
- 69 M. N. Katehakis and L. C. Smit, *Probab. Eng. Inf. Sci.* **26**, 483–508 (2012).
- 70 J. A. Ward and M. López-García, *Appl. Netw. Sci.* **4**, 108 (2019).
- 71 P. Deuffhard, W. Huisinga, A. Fischer, and C. Schütte, *Linear Algebra Appl.* **315**, 39–59 (2000).
- 72 P. Deuffhard and M. Weber, *Linear Algebra Appl.* **398**, 161–184 (2005).
- 73 S. Kube and M. Weber, *J. Chem. Phys.* **126**, 024103 (2007).
- 74 W. Wang, T. Liang, F. K. Sheong, X. Fan, and X. Huang, *J. Chem. Phys.* **149**, 072337 (2018).
- 75 A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, *J. Comput. Phys.* **17**, 10–18 (1975).
- 76 M. A. Novotny, *Phys. Rev. Lett.* **74**, 1–5 (1995).
- 77 R. J. Allen, C. Valerani, and P. R. ten Wolde, *J. Phys.: Condens. Matter* **21**, 463102 (2009).
- 78 J. T. Berryman and T. Schilling, *J. Chem. Phys.* **133**, 244101 (2010).
- 79 M. Athènes and V. V. Bulatov, *Phys. Rev. Lett.* **113**, 230601 (2014).
- 80 M. Athènes, S. Kaur, G. Adjanor, T. Vanacker, and T. Jourdan, *Phys. Rev. Materials* **3**, 103802 (2019).
- 81 D. J. Sharpe and D. J. Wales, *J. Chem. Phys.* **153**, 024121 (2020).
- 82 D. P. Heyman, *SIAM J. Matrix Anal. Appl.* **16**, 954–963 (1995).
- 83 T. J. Sheskin, *Int. J. Math. Educ. Sci. Technol.* **26**, 729–735 (1995).
- 84 D. P. Heyman and D. P. O’Leary, “What is fundamental for Markov chains: First passage times, fundamental matrices, and group generalized inverses,” in *Computations with Markov Chains*, edited by W. J. Stewart (Springer, Boston, MA, 1995) pp. 151–161.
- 85 I. Sonin, *Adv. Math.* **145**, 159–188 (1999).
- 86 I. Sonin and J. Thornton, *SIAM J. Matrix Anal. Appl.* **23**, 209–224 (2001).
- 87 D. Gfeller, P. De Los Rios, A. Cafisch, and F. Rao, *Proc. Natl. Acad. Sci. USA* **104**, 1817–1822 (2007).
- 88 G. R. Bowman and V. S. Pande, *Proc. Natl. Acad. Sci. USA* **107**, 10890–10895 (2010).
- 89 E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber, *J. Chem. Phys.* **129**, 174102 (2008).
- 90 R. Elber, *Q. Rev. Biophys.* **50**, e8 (2017).
- 91 A. M. Berezhkovskii and A. Szabo, *J. Chem. Phys.* **150**, 054106 (2019).
- 92 A. Kells, V. Koskin, E. Rosta, and A. Annibale, *J. Chem. Phys.* **152**, 104108 (2020).
- 93 J. G. Kemeny, *Linear Algebra Appl.* **38**, 193–206 (1981).
- 94 M. Levene and G. Loizou, *Amer. Math. Monthly* **109**, 741–745 (2002).
- 95 P. Doyle, “The Kemeny constant of a Markov chain,” (2009), (<https://arxiv.org/abs/0909.2636>).
- 96 J. J. Hunter, *Commun. Stat. - Theory Methods* **43**, 1309–1321 (2014).
- 97 D. Bini, J. J. Hunter, G. Latouche, B. Meini, and P. Taylor, *J. Appl. Probab.* **55**, 1025–1036 (2018).
- 98 J. Pitman and W. Tang, *Bernoulli* **24**, 1942–1972 (2018).
- 99 J. Berkhout and B. F. Heidergott, *Oper. Res.* **67**, 892–904 (2019).
- 100 G. Hummer and A. Szabo, *J. Phys. Chem. B* **119**, 9029–9037 (2015).
- 101 D. J. Wales and P. Salamon, *Proc. Natl. Acad. Sci. USA* **111**, 617–622 (2014).
- 102 P. Metzner, C. Schütte, and E. Vanden-Eijnden, *Multiscale Model. Simul.* **7**, 1192–1219 (2009).
- 103 J. G. Kemeny and J. L. Snell, *Theory Prob. Its Appl.* **6**, 101–105

- (1961).
- ¹⁰⁴S. A. Serebrinsky, Phys. Rev. E **83**, 037701 (2011).
- ¹⁰⁵T. D. Swinburne and D. J. Wales, J. Chem. Theory Comput. **16**, 2661–2679 (2020).
- ¹⁰⁶W. Grassmann and D. A. Stanford, “Matrix analytic methods,” in *Computational Probability*, edited by W. Grassmann (Springer, New York, 2000) pp. 153–203.
- ¹⁰⁷T. J. Sheskin, Oper. Res. **33**, 228–235 (1985).
- ¹⁰⁸C. D. Meyer Jr., SIAM Rev. **31**, 240–272 (1989).
- ¹⁰⁹R. L. Jack and P. Sollich, Prog. Theor. Phys. Supp. **184**, 304–317 (2010).
- ¹¹⁰D. J. Sharpe and D. J. Wales, J. Chem. Phys. **151**, 124101 (2019).
- ¹¹¹D. D. Yao, J. Appl. Probab. **22**, 939–945 (1985).
- ¹¹²P. Coolen-Schrijner and E. A. van Doorn, Probab. Eng. Inf. Sci. **16**, 351–366 (2002).
- ¹¹³J.-H. Prinz, M. Held, J. C. Smith, and F. Noé, Multiscale Model. Simul. **9**, 545–567 (2011).
- ¹¹⁴M. Kac, Amer. Math. Soc. **53**, 1002–1010 (1947).
- ¹¹⁵J. J. Hunter, Linear Algebra Appl. **417**, 108–123 (2006).
- ¹¹⁶S. Kirkland, Linear Algebra Appl. **433**, 1988–1996 (2010).
- ¹¹⁷G. H. Weiss, Adv. Chem. Phys. **13**, 1–18 (1967).
- ¹¹⁸I. Procaccia, S. Mukamel, and J. Ross, J. Chem. Phys. **68**, 3244–3253 (1978).
- ¹¹⁹T. A. Davis, ACM Trans. Math. Software **30**, 196–199 (2004).
- ¹²⁰R. G. Grimes, J. G. Lewis, and H. D. Simon, SIAM J. Matrix Anal. Appl. **15**, 228–272 (1994).
- ¹²¹D. C. Sorensen, “Implicitly restarted Arnoldi/Lanczos methods for large scale eigenvalue calculations,” in *Parallel Numerical Algorithms*, edited by D. Keyes, A. Sameh, and V. Venkatakrishnan (Springer, Dordrecht, 1997) pp. 119–165.
- ¹²²Y. Saad, Linear Algebra Appl. **34**, 269–295 (1980).
- ¹²³C. C. Paige, Linear Algebra Appl. **34**, 235–258 (1980).
- ¹²⁴C. R. MacCluer, SIAM Rev. **42**, 487–498 (2000).
- ¹²⁵M. K. Cameron and E. Vanden-Eijnden, J. Stat. Phys. **156**, 427–454 (2014).
- ¹²⁶H. D. Simon, Linear Algebra Appl. **61**, 101–131 (1984).
- ¹²⁷G. Meurant, *The Lanczos and Conjugate Gradient Algorithms: From Theory to Finite Precision Computations* (SIAM, Philadelphia, USA, 2006).
- ¹²⁸R. E. Funderlic and C. D. Meyer, Linear Algebra Appl. **76**, 1–17 (1986).
- ¹²⁹G. H. Golub and C. D. Meyer, SIAM J. Alg. Discr. Meth. **7**, 273–281 (1986).
- ¹³⁰G. D. Zhang, SIAM J. Matrix Anal. Appl. **14**, 1112–1123 (1993).
- ¹³¹I. C. F. Ipsen and C. D. Meyer, SIAM J. Matrix Anal. Appl. **15**, 1061–1074 (1994).
- ¹³²G. E. Cho and C. D. Meyer, Linear Algebra Appl. **316**, 21–28 (2000).
- ¹³³J. J. Hunter, “Perturbed Markov chains,” in *Contributions to Probability and Statistics: Applications and Challenges*, edited by P. Brown, S. Liu, and D. Sharma (World Scientific, Singapore, 2006) pp. 99–112.
- ¹³⁴J. J. Hunter, Linear Algebra Appl. **410**, 217–243 (2005).
- ¹³⁵Y. Saad, “Preconditioned Krylov subspace methods for the numerical solution of Markov chains,” in *Computations with Markov Chains*, edited by W. J. Stewart (Springer, Boston, MA, 1995) pp. 49–64.
- ¹³⁶Y. Saad, *Numerical Methods for Large Eigenvalue Problems* (SIAM, Philadelphia, USA, 2011).
- ¹³⁷B. Peters, *Reaction Rate Theory and Rare Events* (Elsevier, Oxford, UK, 2017).
- ¹³⁸M. K. Cameron and T. Gan, Mol. Simul. **42**, 1410–1428 (2016).
- ¹³⁹S. Fortunato, Phys. Rep. **486**, 75–174 (2010).
- ¹⁴⁰M. Haviv, SIAM J. Numer. Anal. **24**, 952–966 (1987).
- ¹⁴¹T. Dayar and W. J. Stewart, SIAM J. Sci. Comput. **17**, 287–303 (1996).
- ¹⁴²R. Zwanzig, J. Stat. Phys. **30**, 255–262 (1983).
- ¹⁴³A. Berezhkovski, G. Hummer, and A. Szabo, J. Chem. Phys. **130**, 205102 (2009).
- ¹⁴⁴A. Kells, Z. E. Mihálka, A. Annibale, and E. Rosta, J. Chem. Phys. **150**, 134107 (2019).
- ¹⁴⁵M. Bastian, S. Heymann, and M. Jacomy, “Gephi: An open source software for exploring and manipulating networks,” International AAAI Conference on Weblogs and Social Media (2009).
- ¹⁴⁶D. J. Sharpe and D. J. Wales, “Dimensionality reduction of markovian networks using efficient dynamical simulations,” (2020), (in preparation).
- ¹⁴⁷J. K. Weber and V. S. Pande, J. Chem. Theory Comput. **7**, 3405–3411 (2011).
- ¹⁴⁸J. K. Weber and V. S. Pande, J. Chem. Phys. **142**, 215105 (2015).
- ¹⁴⁹A. M. A. West, R. Elber, and D. Shalloway, J. Chem. Phys. **126**, 145104 (2007).
- ¹⁵⁰F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, Proc. Natl. Acad. Sci. USA **106**, 19011–19016 (2009).
- ¹⁵¹J. R. Bunch, Linear Algebra Appl. **88/89**, 49–66 (1987).
- ¹⁵²E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users’ Guide*, 3rd ed. (SIAM, Philadelphia, PA, 1999).
- ¹⁵³D. Kannan, D. J. Sharpe, T. D. Swinburne, and D. J. Wales, “Dimensionality reduction of complex networks using partial graph transformation,” (2020), (unpublished).
- ¹⁵⁴H. Caswell, Numer. Linear Algebra Appl. **18**, 901–917 (2011).
- ¹⁵⁵J. J. Hunter, Res. Lett. Inf. Math. Sci. **1**, 25–36 (2000).
- ¹⁵⁶J. J. Hunter, Linear Algebra Appl. **45**, 157–198 (1982).
- ¹⁵⁷J. J. Hunter, Linear Algebra Appl. **102**, 121–142 (1988).
- ¹⁵⁸J. J. Hunter, Linear Algebra Appl. **127**, 71–84 (1990).
- ¹⁵⁹J. J. Hunter, Linear Algebra Appl. **447**, 38–55 (2014).
- ¹⁶⁰J. J. Hunter, Linear Algebra Appl. **429**, 1135–1162 (2008).