# Graph transformation and shortest paths algorithms for finite Markov chains 

Daniel J. Sharpe ${ }^{1}$ and David J. Wales ${ }^{1,}$ a)<br>Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

(Dated: 3 November 2020)
The graph transformation (GT) algorithm robustly computes the mean first passage time to an absorbing state in a finite Markov chain. Here we present a concise overview of the iterative and block formulations of the GT procedure and generalize the GT formalism to the case of any path property that is a sum of contributions from individual transitions. In particular, we examine the path action, which directly relates to the path probability, and analyze the first passage path ensemble for a model Markov chain that is metastable and therefore numerically challenging. We compare the mean first passage path action, obtained using GT, with the full path action probability distribution simulated efficiently using kinetic path sampling, and with values for the highest-probability paths determined by the recursive enumeration algorithm (REA). In Markov chains representing realistic dynamical processes, the probability distributions of first passage path properties are typically fat-tailed and therefore difficult to converge by sampling, which motivates the use of exact and numerically stable approaches to compute the expectation. We find that the kinetic relevance of the set of highest-probability paths depends strongly on the metastability of the Markov chain, and so the properties of the dominant first passage paths may be unrepresentative of the global dynamics. Use of a global measure for edge costs in the REA, based on net productive fluxes, allows the total reactive flux to be decomposed into a finite set of contributions from simple flux-paths. By considering transition flux-paths, a detailed quantitative analysis of the relative importance of competing dynamical processes is possible even in the metastable regime.
Keywords: graph transformation, Markov chain, master equation, stochastic dynamics, random walks, rare events, metastability, state reduction, shortest paths

## I. INTRODUCTION

Diverse stochastic phenomena are conveniently represented by finite Markov chains; ${ }^{1}$ probabilistic network models for which the future dynamics depend only on the currently occupied state and not on the prior history of the trajectory. ${ }^{2}$ Discrete-time Markov chains ${ }^{3}$ (DTMCs) are commonly estimated from trajectory data on a continuous potential energy landscape in the Markov State Model (MSM) framework. ${ }^{4-8}$ In a complementary approach, continuous-time Markov chains ${ }^{1}$ (CTMCs) can be mapped from a potential energy landscape by geometry optimization ${ }^{9}$ of local stationary points in the discrete path sampling (DPS) framework. ${ }^{10-13}$ CTMCs with a countably-infinite state space ${ }^{14,15}$ are widely used to represent the number of each species in population dynamics ${ }^{16-18}$ processes such as chemical and biochemical reaction cycles, ${ }^{19-24}$ and can be transformed to finite Markov chains with negligible error by truncating the state space. ${ }^{25,26}$

In previous work we have considered a discrete-state Markov reward process ${ }^{27}$ on a finite state space $\mathcal{S}$, where individual $i \leftarrow j$ transitions in the Markov chain are associated with a reward $R_{i j}$ that depends only on the identity of the currently occupied node $j$ and not the next node $i$ (i.e. $R_{i j} \equiv R_{j} \forall i$ ). The graph transformation

[^0](GT) algorithm ${ }^{28-32}$ can be used to compute the average reward along first passage ${ }^{33,34}$ trajectories from an initial set of nodes $\mathcal{B}$ to an absorbing set of nodes $\mathcal{A}$ in this case. An important example of a path property of this kind is the path time. The mean time elapsed along an $\mathcal{A} \leftarrow \mathcal{B}$ first passage path ${ }^{33,34}$ in a CTMC $^{2}$ is a sum of mean waiting times $\tau_{j}$ for transitions from nodes $j$ in the path. ${ }^{30}$ For a DTMC, ${ }^{3}$ the fixed lag time associated with transitions is uniform for all nodes, $\tau_{j} \equiv \tau \forall j .{ }^{35}$ The $\mathcal{A} \leftarrow \mathcal{B}$ mean first passage time (MFPT), ${ }^{36} \mathcal{T}_{\mathcal{A B}}$, is a sum of path times, taken over all possible first passage paths, weighted by the associated path probabilities. ${ }^{35}$

The GT algorithm is numerically stable, and therefore valuable in many practical applications. ${ }^{31}$ Markov chains representing realistic dynamical processes are frequently observed to encompass a separation of characteristic timescales, and the corresponding transition probability or rate matrix is therefore ill-conditioned. ${ }^{37-48}$ This feature arises in Markovian networks constructed using the MSM and DPS frameworks because of the exponential sensitivity of estimated transition probabilities or rates to the structure of the underlying energy landscape. ${ }^{49,50}$ Metastability also emerges in reaction networks where the rate constants for alternative competing reactions are disparate. ${ }^{19-24}$ Markov chains that harbour metastable communities of nodes pose numerical challenges, since dynamical simulations become unfeasibly inefficient ${ }^{51,52}$ and conventional linear algebra methods lead to a severe propagation of numerical error. The GT algorithm pro-
vides a powerful alternative approach to compute MFPTs in high-dimensional and ill-conditioned Markov chains. ${ }^{31}$

In the present contribution we generalize the GT algorithm ${ }^{28-32}$ to the case of rewards $R_{i j}$ that are different for transitions to alternative destination nodes $i$ from the currently occupied node $j$. Relevant examples of such rewards include the path action, ${ }^{53}$ which directly relates to the path probability, and the entropy flow, ${ }^{54}$ which quantifies the reversibility of a trajectory. ${ }^{55,56} \mathrm{Al}$ though they are not dynamical observables, the average path action and entropy flow have rigorous interpretations, and the probability distributions for these path properties yield important insight into the characteristics of a Markov chain. For instance, the expectation of the path action is the Shannon entropy ${ }^{57}$ associated with the ensemble of first passage paths. ${ }^{58-62}$ A similar quantity is employed in the maximum caliber and maximum entropy frameworks as the objective function in a variational principle to estimate Markovian transition probabilities or rates for a discrete set of states, given constraints on the stationary distribution and additional global dynamical information. ${ }^{63,64}$ The entropy flow is a central quantity in stochastic thermodynamics, ${ }^{65}$ since the average entropy production is governed by an integral fluctuation theorem. ${ }^{66}$ Previous analytical results considering paths in Markov chains weighted by arbitrary rewards are limited and do not lend themselves to the design of computational procedures that have the desirable scalability and stability of our generalized GT algorithm. ${ }^{67}$

Following derivations of the expected rewards for paths on renormalized Markov chains (Sec. IIB), and of the iterative and block formulations for the generalized GT algorithm (Sec. II C), we compute the mean first passage path action for a model metastable Markov chain (Sec. III). We compare the expectation for the path action with the full probability distribution simulated efficiently using kinetic path sampling, ${ }^{68,69}$ and with the values for the highest-probability paths determined by the recursive enumeration algorithm (REA) (Sec. II D). ${ }^{70}$ We demonstrate that the probability distributions of first passage path properties are typically fat-tailed, and that the fraction of the total probability flux to the absorbing state accounted for by the dominant first passage paths depends strongly on the metastability of the Markov chain. Hence, it is often challenging to obtain an accurate numerical estimate for the expectation of a first passage path property by sampling trajectories, and it may be unfeasible to converge the pathwise sum for the expectation using shortest paths algorithms. We propose an alternative shortest paths analysis to provide quantitative information on the relative importance of alternative $\mathcal{A} \leftarrow \mathcal{B}$ processes, using edge costs in the REA that are based on net reactive fluxes. ${ }^{71}$ This formulation allows the total $\mathcal{A} \leftarrow \mathcal{B}$ reactive flux to be decomposed into a sum of contributions from a finite set of simple fluxpaths (Sec. IIE). We find that the total reactive flux becomes increasingly localized among a small subset of
transition flux-paths with increasing metastability.

## II. THEORY

## A. Mathematical definitions

We consider arbitrary discrete- ${ }^{3}$ and continuous-time ${ }^{72}$ finite Markov chains. A DTMC is parameterized by $i \leftarrow j$ transition probabilities $T_{i j}(\tau)$ for a fixed time step $\tau$. A CTMC is parameterized by $i \leftarrow j \neq i$ delete infinitesimal transition rates $K_{i j} .{ }^{1}$ Equivalently, a CTMC can be specified by a branching probability matrix ${ }^{73} \mathbf{P}$ with off-diagonal elements $P_{i j}=K_{i j} / \sum_{\gamma \neq j} K_{\gamma j}$ and diagonal elements $P_{j j}=0$, and a vector of mean waiting times for transitions from nodes $j$, with elements $\tau_{j}=1 / \sum_{\gamma \neq j} K_{\gamma j}$. In the present work, we denote the stochastic matrix of a Markov chain $(\mathbf{T}(\tau)$ for a DTMC and $\mathbf{P}$ for a CTMC) by $\mathbf{T}$ for generality. We denote the state space of the Markov chain (i.e. the complete set of nodes) as $\mathcal{S}$, and consider two disjoint sets of endpoint nodes $\mathcal{A}$ and $\mathcal{B}$, where $\mathcal{A} \cup \mathcal{B} \subseteq \mathcal{S}$, which are the target and initial states, respectively.

Let the $i \leftarrow j$ transition be associated with a reward $R_{i j}$, which does not modify the dynamics but instead is used to assign a weight $\mathcal{R}[\xi]$ to paths $\xi$. The total reward along a particular $\mathcal{A} \leftarrow \mathcal{B}$ first passage path $\xi \equiv\left\{a \in \mathcal{A} \leftarrow i_{n} \leftarrow i_{n-1} \leftarrow \cdots \leftarrow i_{1} \leftarrow b \in \mathcal{B}\right\}$, where $i_{1}, \ldots, i_{n} \notin \mathcal{A}$, is a sum of contributions from individual transitions along $\xi, \mathcal{R}[\xi]=\sum_{(i \leftarrow j) \in \xi} R_{i j}$. An important example of a path property of this type is the path action, ${ }^{57}-\ln \mathcal{W}[\xi]=-\sum_{(i \leftarrow j) \in \xi} \ln T_{i j}$. Here, $\mathcal{W}[\xi]$ denotes the product of transition probabilities along the path $\xi$, i.e. the path weight. ${ }^{30}$ The path probability $\mathcal{P}[\xi]$ is equal to this probability weighted by the probability $p_{b}(0)$ of starting at the initial node $b \in \mathcal{B}$ of the path $\xi, \mathcal{P}[\xi]=p_{b}(0) \mathcal{W}[\xi] .{ }^{74}$ Another tangible example of a reward is the entropy flow. ${ }^{54}$ In discrete time, the path entropy flow is ${ }^{75} \mathcal{S}[\xi]=\sum_{(i \leftarrow j) \in \xi} \ln \left(T_{j i} / T_{i j}\right)$ (in units of the Boltzmann constant), and in continuous time ${ }^{76}$ $\mathcal{S}[\xi]=\sum_{(i \leftarrow j) \in \xi} \ln \left(K_{j i} / K_{i j}\right)$. The numerical results presented in Sec. III are concerned with the path action.

In addition to rewards $\mathcal{R}[\xi]$ along individual trajectories $\xi$, we are interested in the ensemble average reward $\mathcal{R}_{\mathcal{A B}}$, considering all trajectories that start in the state $\mathcal{B}$ and are absorbed upon hitting the state $\mathcal{A}$, including revisits to $\mathcal{B} .{ }^{30}$ We refer to this set of trajectories as the first passage path ensemble ${ }^{58-62}$ (FPPE) and $\mathcal{R}_{\mathcal{A B}}$ as the mean first passage reward (MFPR).

## B. Expected rewards for individual paths on censored Markov chains

Our generalized GT algorithm to calculate the MFPR (Sec. II C) utilizes the concept of a censored Markov chain. ${ }^{16,77-82}$ We begin by considering the effect of renor-
malization to eliminate a single node ${ }^{83} n$ on the rewards associated with paths on the resulting censored network. For pairs of nodes $i$ and $j$ for which there exist $i \leftarrow n$ and $n \leftarrow j$ transitions, it is possible to define renormalized transition probabilities $T_{i j}^{\prime}$ that account for the average contribution of $i \leftarrow j$ transitions proceeding via the eliminated node $n$. Specifically, the total probability of the $i \leftarrow j$ transition in the renormalized Markov chain is ${ }^{30}$

$$
\begin{equation*}
T_{i j}^{\prime}=T_{i j}+\frac{T_{i n} T_{n j}}{1-T_{n n}} \tag{1}
\end{equation*}
$$

Here, the first contribution corresponds to direct $i \leftarrow j$ transitions on the original network, and the second contribution corresponds to indirect ('round-trip') transitions, $i \leftarrow n \leftarrow \ldots \leftarrow n \leftarrow j$, where the eliminated node $n$ is visited an arbitrary number of times. ${ }^{68}$ The updated transition probabilities of Eq. 1 naturally yield a new stochastic matrix and correctly preserves the normalization. ${ }^{30}$

We wish to derive the renormalized reward $R_{i j}^{\prime}$ associated with the $i \leftarrow j$ transition in the censored Markov chain for which the $n$-th node is eliminated. We must account for the fact that the expected reward associated with $i \leftarrow j$ transitions proceeding indirectly, via the eliminated (censored) node $n$, is different from the reward for the direct $i \leftarrow j$ transition, which does not involve the censored node. The conditional probability that a $i \leftarrow j$ transition is direct is given by $T_{i j} / T_{i j}^{\prime}(c f$. Eq. 1), and the reward for the direct transition is simply $R_{i j}$. The contribution to the renormalized $i \leftarrow j$ reward arising from indirect transitions via node $n$ is more complicated. On average, a trajectory at node $n$ will transition from $n$ a total of $\left(1-T_{n n}\right)^{-1}$ times before leaving $n$, including the final transition to escape from $n .{ }^{3}$ Thus the expected reward for an indirect $i \leftarrow j$ transition is

$$
\begin{equation*}
\left\langle R_{i n}^{\text {indir }}\right\rangle=R_{i n}+R_{n j}+R_{n n}\left(\frac{1}{1-T_{n n}}-1\right) \tag{2}
\end{equation*}
$$

The average reward associated with the $i \leftarrow j$ transition for the renormalized (censored) network is the sum of direct and average indirect rewards weighted by the conditional probabilities of direct and indirect $i \leftarrow j$ transitions, respectively,

$$
\begin{equation*}
R_{i j}^{\prime}=\frac{1}{T_{i j}^{\prime}}\left(T_{i j} R_{i j}+\frac{T_{i n} T_{n j}}{1-T_{n n}}\left\langle R_{i n}^{\mathrm{indir}}\right\rangle\right) \tag{3}
\end{equation*}
$$

Here $T_{i j}^{\prime}$ is given by Eq. 1 and $\left\langle R_{i n}^{\text {indir }}\right\rangle$ by Eq. 2. Eq. 3 conserves the average rewards $\mathcal{R}[\xi]$ for all individual paths $\xi$, with arbitrary initial and final nodes, on a censored Markov chain. ${ }^{77}$ The rewards associated with trajectories on the censored Markov chain are strictly an expectation with respect to the contributions of path segments that visit censored nodes. ${ }^{77}$ Reducing the dimensionality of Markov chains by renormalization provides a strategy to facilitate the sampling of trajectories, ${ }^{84}$ and Eq. 3 allows
for the probability distributions of path rewards on the transformed network to be estimated within this framework.

Our result in Eq. 3 can also be exploited to compute the overall $\mathcal{A} \leftarrow \mathcal{B}$ MFPR. Following elimination of all nodes of the set $(\mathcal{A} \cup b)^{c}$ using renormalization of the transition probabilities and rewards (Eqs. 1 and 3, respectively), where $b \in \mathcal{B}$ is a single node of the initial state, the average reward associated with the ensemble of $\mathcal{A} \leftarrow b$ trajectories is

$$
\begin{equation*}
\mathcal{R}_{\mathcal{A} b}=\left(\frac{1}{1-T_{b b}^{\prime}}-1\right) R_{b b}^{\prime}+\sum_{a \in \mathcal{A}} \frac{T_{a b}^{\prime} R_{a b}^{\prime}}{1-T_{b b}^{\prime}} \tag{4}
\end{equation*}
$$

Here, we have again used the result that the expected number of transitions from node $b$ before hitting a different node is $\left(1-T_{b b}^{\prime}\right)^{-1}$, all of which except the final transition are $b \leftarrow b$ self-loop transitions, and the probability that the node $a \in \mathcal{A}$ is hit upon leaving $b$ is $T_{a b}^{\prime} /\left(1-T_{b b}^{\prime}\right)$. The average reward for paths of the $\mathcal{A} \leftarrow \mathcal{B}$ FPPE, ${ }^{58-62}$ $\mathcal{R}_{\mathcal{A B}}$, is simply a weighted average of rewards $\mathcal{R}_{\mathcal{A} b}$ (Eq. 4) with respect to the initial occupation probability distribution $p_{b}(0)$ for nodes $b \in \mathcal{B} .{ }^{31}$

## C. Mean first passage reward computed using a generalized graph transformation procedure

Let the set of transient (nonabsorbing) nodes of the Markov chain be denoted $\mathcal{Q}$, and the complete set of nodes as $\mathcal{S} \equiv \mathcal{Q} \cup \mathcal{A}$. The results of Sec. II B demonstrate that the $\mathcal{A} \leftarrow \mathcal{B}$ MFPR can be computed by iteratively renormalizing the elements of a reward matrix $\mathbf{R}$ that is initially of dimensions $|\mathcal{Q}| \times|\mathcal{Q}|$, and from which the $n$ th row and column are removed when eliminating node $n$. In fact, it is only necessary to consider an initial $|\mathcal{Q}|-$ dimensional vector of mean rewards for transitions from the transient nodes in order to compute the MFPR to the absorbing state, as we now show.

For a general Markov chain, the sum of path probabilities to the absorbing state $\mathcal{A}$ from a transient node $q \in \mathcal{Q}$ is given by the component $\left[\mathbf{1}_{\mathcal{A}}^{\top} \mathbf{T}_{\mathcal{A Q}} \mathbf{N}_{\mathcal{Q Q}}\right]_{q}$, and is unity for all $q .^{85}$ Here, $\mathbf{N}_{\mathcal{Q Q}}=\left(\mathbb{I}_{\mathcal{Q Q}}-\mathbf{T}_{\mathcal{Q Q}}\right)^{-1}$ is the fundamental matrix ${ }^{3}$ associated with the absorbing Markov chain parameterized by the substochastic matrix $\mathbf{T}_{\mathcal{Q} \mathcal{Q}}$, for transitions between nodes of the set $\mathcal{Q}$, and $\mathbb{I}_{\mathcal{Q Q}}$ denotes the $|\mathcal{Q}|$-dimensional identity matrix.

To produce a general formula for the MFPR from the set of transient nodes to the absorbing state, $\mathcal{R}_{\mathcal{A Q}}$, we introduce the reweighted $i \leftarrow j$ transition probabilities $\widehat{T}_{i j}=T_{i j} \exp \left(\zeta R_{i j}\right)$. In component form we have $\partial \widehat{T}_{i j} /\left.\partial \zeta\right|_{\zeta=0}=T_{i j} R_{i j}$, and so

$$
\begin{equation*}
\left.\frac{\partial \widehat{\mathbf{T}}_{\mathcal{A Q}}}{\partial \zeta}\right|_{\zeta=0}=\mathbf{T}_{\mathcal{A Q}} \circ \mathbf{R}_{\mathcal{A Q}} \tag{5}
\end{equation*}
$$

where $\circ$ denotes the elementwise (Hadamard) product,
which we write as $\mathbf{C}_{\mathcal{A Q}}$. Here, $\mathbf{T}_{\mathcal{A Q}}$ is the substochastic matrix for transitions from transient to absorbing nodes, with dimensions $|\mathcal{A}| \times|\mathcal{Q}|$, and $\mathbf{R}_{\mathcal{A Q}}$ is the corresponding matrix of associated rewards for the transitions.

The $\mathcal{A} \leftarrow \mathcal{Q}$ MFPR can be computed from the fundamental matrix $\mathbf{N}_{\mathcal{Q} \mathcal{Q}}$ of the absorbing Markov chain as

$$
\begin{equation*}
\mathcal{R}_{\mathcal{A Q}}=\left.\frac{\partial}{\partial \zeta} \mathbf{1}_{\mathcal{A}}^{\top} \widehat{\mathbf{T}}_{\mathcal{A Q}} \widehat{\mathbf{N}}_{\mathcal{Q Q}}\right|_{\zeta=0} \mathbf{p}_{\mathcal{Q}}(0) \tag{6}
\end{equation*}
$$

where $\mathbf{p}_{\mathcal{Q}}(0)$ is the initial occupation probability distribution within $\mathcal{Q}$. We require the derivative

$$
\begin{align*}
\left.\frac{\partial \widehat{\mathbf{N}}_{\mathcal{Q Q}}}{\partial \zeta}\right|_{\zeta=0} & =\left.\mathbf{N}_{\mathcal{Q Q}} \frac{\partial \widehat{\mathbf{T}}_{\mathcal{A Q}}}{\partial \zeta}\right|_{\zeta=0} \mathbf{N}_{\mathcal{Q Q}} \\
& =\mathbf{N}_{\mathcal{Q} \mathcal{Q}} \mathbf{C}_{\mathcal{Q} \mathcal{Q}} \mathbf{N}_{\mathcal{Q Q}} \tag{7}
\end{align*}
$$

which gives

$$
\begin{align*}
\mathcal{R}_{\mathcal{A Q}} & =\mathbf{1}_{\mathcal{A}}^{\top}\left(\mathbf{C}_{\mathcal{A Q}}+\mathbf{T}_{\mathcal{A Q}} \mathbf{N}_{\mathcal{Q} \mathcal{Q}} \mathbf{C}_{\mathcal{Q Q}}\right) \mathbf{N}_{\mathcal{Q Q}} \mathbf{p}_{\mathcal{Q}}(0) \\
& =\left(\mathbf{1}_{\mathcal{A}}^{\top} \mathbf{C}_{\mathcal{A Q}}+\mathbf{1}_{\mathcal{Q}}^{\top} \mathbf{C}_{\mathcal{Q Q}}\right) \mathbf{N}_{\mathcal{Q Q}} \mathbf{p}_{\mathcal{Q}}(0) \\
& =\mathbf{r}_{\mathcal{Q}}^{\top} \mathbf{N}_{\mathcal{Q Q}} \mathbf{p}_{\mathcal{Q}}(0) \tag{8}
\end{align*}
$$

where the $q$-th component of the column vector $\mathbf{r}_{\mathcal{Q}}$ is the average reward for transitions from node $q$ :

$$
\begin{equation*}
\left[\mathbf{r}_{\mathcal{Q}}\right]_{q}=\sum_{\gamma} T_{\gamma q} R_{\gamma q} \tag{9}
\end{equation*}
$$

For comparison, the corresponding formula for the $\mathrm{MFPT}^{85}$ is $\mathcal{T}_{\mathcal{A Q}}=\boldsymbol{\tau}^{\top} \mathbf{N}_{\mathcal{Q} \mathcal{Q}} \mathbf{p}_{\mathcal{Q}}(0)$, where $\boldsymbol{\tau}$ is the vector of mean waiting times (for a CTMC) or lag times (for a DTMC) for transitions from the nodes. Eq. 8 also demonstrates that $\left[\mathbf{N}_{\mathcal{Q}}\right]_{i j}$ is the expected number of times the $i$-th node is visited prior to absorption for first passage paths initialized from the $j$-th node. ${ }^{3}$ Using Eqs. 8 and 9, the $\mathcal{A} \leftarrow q$ MFPRs for all transient nodes $q \in \mathcal{Q}$ can be computed simultaneously by inversion of a matrix with dimensions $|\mathcal{Q}| \times|\mathcal{Q}|$.

For metastable Markov chains, the matrix inversion operation to compute $\mathbf{N}_{\mathcal{Q} \mathcal{Q}}$ is numerically unstable. We can instead iteratively eliminate blocks of one or more nodes to compute $\mathcal{A} \leftarrow \mathcal{B}$ MFPRs by renormalization of an average reward vector (Eq. 9) and a transition probability matrix. delete Consider that The set of initial nodes $\mathcal{B}$ forms a subset of the transient state, with the set of other (intervening) nodes denoted $\mathcal{I}$, i.e. $\mathcal{Q} \equiv \mathcal{B} \cup \mathcal{I}$. After eliminating nodes of the state $\mathcal{I}$, so that the Markov chain comprises only nodes of the set $\mathcal{B} \cup \mathcal{A}$, the corresponding path probabilities can be written as $\mathbf{1}_{\mathcal{A}}^{\top} \mathbf{T}_{\mathcal{A B}}^{\mathcal{I}} \mathbf{N}_{\mathcal{B} \mathcal{B}}^{\mathcal{I}}$, where ${ }^{32,77,85}$

$$
\begin{align*}
& \mathbf{T}_{\mathcal{A B}}^{\mathcal{I}}=\mathbf{T}_{\mathcal{A B}}+\mathbf{T}_{\mathcal{A I}} \mathbf{N}_{\mathcal{I I}} \mathbf{T}_{\mathcal{I B}},  \tag{10a}\\
& \mathbf{N}_{\mathcal{B B}}^{\mathcal{I}}=\left(\mathbb{I}_{\mathcal{B B}}-\mathbf{T}_{\mathcal{B B}}^{\mathcal{I}}\right)^{-1} . \tag{10b}
\end{align*}
$$

Here, we have used the superscript $\mathcal{I}$ to indicate that nodes of the set $\mathcal{I}$ have been eliminated by renormaliza-
tion. Introducing the reweighted transition probabilities $\widehat{T}_{i j}$ and following a derivation analogous to that for Eq. 8 yields the following expression for the $\mathcal{A} \leftarrow \mathcal{B}$ MFPR:

$$
\begin{equation*}
\mathcal{R}_{\mathcal{A B}}=\left(\mathbf{r}_{\mathcal{B}}^{\top}+\mathbf{r}_{\mathcal{I}}^{\top} \mathbf{N}_{\mathcal{I I}} \mathbf{T}_{\mathcal{I B}}\right) \mathbf{N}_{\mathcal{B} \mathcal{B}}^{\mathcal{I}} \mathbf{p}_{\mathcal{B}}(0) \tag{11}
\end{equation*}
$$

Since the initial occupation probability distribution is localized within $\mathcal{B}$, the MFPR will be conserved if we iteratively eliminate blocks of nodes $\mathcal{N} \subseteq \mathcal{I}$, renormalizing the probabilities for transitions from nodes in the set $\mathcal{Q}^{\prime} \equiv \mathcal{Q} \backslash \mathcal{N}$ according to the usual GT formula ( $c f$. Eq. 10a),

$$
\begin{equation*}
\mathbf{T}_{\mathcal{S} \mathcal{Q}^{\prime}}^{\mathcal{N}}=\mathbf{T}_{\mathcal{S} \mathcal{Q}^{\prime}}+\mathbf{T}_{\mathcal{S N}} \mathbf{N}_{\mathcal{N N}} \mathbf{T}_{\mathcal{N} \mathcal{Q}^{\prime}} \tag{12}
\end{equation*}
$$

and updating the average rewards according to

$$
\begin{equation*}
\mathbf{r}_{\mathcal{Q}^{\prime}}^{\mathcal{N}}=\mathbf{r}_{\mathcal{Q}^{\prime}}^{\top}+\mathbf{r}_{\mathcal{N}}^{\top} \mathbf{N}_{\mathcal{N} \mathcal{N}} \mathbf{T}_{\mathcal{N} \mathcal{Q}^{\prime}} \tag{13}
\end{equation*}
$$

Eq. 12 is the block analogue of Eq. 1. That is, Eq. 12 yields the same renormalized stochastic matrix as the repeated application of Eq. 1 to iteratively eliminate the nodes of the set $\mathcal{N}$ in any order. The generalized GT procedure to compute the $\mathcal{A} \leftarrow \mathcal{B}$ MFPR based on Eqs. 12 and 13 is both numerically stable and efficient if nodes in blocks $\mathcal{N}$ to be eliminated simultaneously belong to the same metastable community. ${ }^{32,85}$ The communities can be determined a priori by an appropriate clustering algorithm. ${ }^{52,86}$

Eq. 13 is analogous to the result for the renormalized waiting times that preserve the MFPT, ${ }^{30,85}$ with the mean rewards for transitions from transient nodes $q \in \mathcal{Q}$ in place of the mean waiting (or lag) times. Eliminating a single node $n \in \mathcal{I}$ by renormalization, Eq. 12 reduces to Eq. 1, and Eq. 13 reduces to

$$
\begin{equation*}
\left[\mathbf{r}_{\mathcal{Q}^{\prime}}\right]_{q}=\left[\mathbf{r}_{\mathcal{Q}^{\prime}}\right]_{q}+\frac{\left[\mathbf{r}_{\mathcal{Q}^{\prime}}\right]_{n} T_{n q}}{1-T_{n n}} \tag{14}
\end{equation*}
$$

Exploiting the relation $1-T_{n n}=\sum_{\gamma \neq n} T_{\gamma n}$ when $T_{n n} \rightarrow$ 1 avoids the propagation of significant roundoff error in the finite precision arithmetic, and this algorithm is numerically stable. ${ }^{37-47}$ The time complexity of the iterative procedure depends on the average degree of nodes and on the heterogeneity of the degree distribution, ${ }^{29,68}$ and varies between $\mathcal{O}\left(|\mathcal{Q}|^{3}\right)$ and $\mathcal{O}\left(|\mathcal{Q}|^{4}\right)$.

## D. Recursive enumeration algorithm

Formally, the expected $\mathcal{A} \leftarrow \mathcal{B}$ reward is a sum of contributions $\mathcal{R}[\xi]$ from all paths $\xi$ of the first passage path ensemble ${ }^{35,57,62}$

$$
\begin{equation*}
\mathcal{R}_{\mathcal{A B}}=\sum_{\xi \in\{\mathcal{A} \leftarrow \mathcal{B}\}} p_{b}(0) \mathcal{W}[\xi] \mathcal{R}[\xi] \tag{15}
\end{equation*}
$$

The weighted sum in Eq. 15 has an infinite number of terms for Markov chains featuring loops, but the con-
tributions to the sum from paths related by additional traversals of a particular loop converge. delete to zero. It is a finite sum with an infinite number of terms, and the loop traversals converge, but not to zero. The highestprobability first passage paths, ${ }^{87}$ and their contribution to the $\mathcal{A} \leftarrow \mathcal{B}$ reward sum in Eq. 15, can be determined by a $k$ shortest paths algorithm ${ }^{88-90}$ where the $i \leftarrow j$ edge cost is $-\ln T_{i j}$, i.e. the contribution of the transition to the path action. ${ }^{74,91}$ For DTMCs, self-loop transitions for nodes can be eliminated by renormalization using $T_{i j} \Rightarrow T_{i j} /\left(1-T_{j j}\right) \forall i \neq j{ }^{30}$ while preserving the MFPR using $R_{i j} \Rightarrow R_{i j}+R_{j j}\left[\left(1-T_{j j}\right)^{-1}-1\right]$ (see Sec. IIB).

For metastable Markov chains, successive shortest paths tend to differ by small modifications, such as a single additional loop traversal or a path that differs by a few alternative nodes. ${ }^{74}$ We therefore choose to employ the recursive enumeration algorithm (REA) of Jiménez and Marzal, ${ }^{70}$ which is particularly efficient in cases where the set of $k$ shortest paths share most of their nodes in common, and consist of a small fraction of the total number of nodes in the network. ${ }^{70}$ The REA has worst-case time complexity $\mathcal{O}(E+k V \log (E / V))$ for a network comprising $V$ nodes and $E$ edges. The algorithm is empirically observed to outperform alternative general $k$ shortest paths algorithms that have superior asymptotic time complexity, such as those of Eppstein [time complexity $\mathcal{O}(E+V+k \log k)],{ }^{92,93}$ Azevedo et al. [time complexity $\mathcal{O}(k E)],{ }^{94,95}$ and Martins, Pascoal, and dos Santos [time complexity $\mathcal{O}(k V \log V)],{ }^{96-98}$ because the REA is associated with a comparatively small computational overhead. ${ }^{70}$ In the following informal derivation of the REA, we assume that all nodes are reachable from all nonabsorbing nodes, i.e. the set $\mathcal{S} \backslash \mathcal{A}$ is transient.

The REA formulates the general single-source node, single-sink node $a \in \mathcal{A} \leftarrow b \in \mathcal{B} k$ shortest paths problem as a set of Bellman equations, ${ }^{99}$ which are solved recursively. ${ }^{100}$ Let the $k$-th shortest path to node $j$ be denoted $\xi^{k}(j)$, with associated cost $\mathcal{R}\left[\xi^{k}(j)\right]$, and the set of nodes with direct transitions to node $j$ be denoted $\mathcal{D}(j)$. The first stage of the REA constructs the shortest path tree for the transitions from the single initial node to all alternative nodes using any appropriate procedure, such as Dijkstra's algorithm [worst case time complexity $\mathcal{O}(E+V \log V)] .{ }^{101-103}$ The REA exploits the fact that the $k$-th shortest path to node $j$ can be written in the form $\xi^{k}(j) \equiv \xi^{k^{\prime}}(i) \cup\{j \leftarrow i\}$, where $i \in \mathcal{D}(j)$ and $k^{\prime} \leq k$. At the $(k-1)$-th iteration of the REA, the next $(k$-th) shortest path to the absorbing node $a$ can therefore be selected from a list $\mathcal{M}(a)$ of such candidate paths. For each node $i \in \mathcal{D}(a)$, only the candidate path $\xi^{k^{\prime}}(i) \cup\{a \leftarrow i\}$ for which $\xi^{k^{\prime}}(i)$ has the lowest cost, and which has not already been chosen as a previous shortest path to the $a$-th node, needs to be considered. Hence, there are at $\operatorname{most}|\mathcal{D}(a)|$ candidates for the next shortest path to node $a$; one for each node with a transition to $a$. Here, we have noted that ties may be broken arbitrarily, and that there is no more than one edge connecting any pair of nodes in
a Markovian network. The REA maintains an array of candidate paths $\mathcal{M}(j)$ for all nodes $j$ of the network, and an array of the $k$-th shortest paths to each node. At each iteration of the REA, a function is called to determine the next shortest path to the target node $a$. Recursive calls to this function are used to ensure that candidate paths are assigned, and that with lowest cost selected, for the shortest paths to preceding nodes, as required. The pseudocode for this procedure applied to determine the highest-probability first passage paths in a finite Markov chain, employing path costs $\mathcal{R}[\xi] \equiv-\ln \mathcal{W}[\xi]$, is presented in Algorithm 1.

## E. Transition flux-paths

The transition probabilities are a local measure of the probability flux, and $-\ln T_{i j}$ represents only one possible choice for the $i \leftarrow j$ edge costs to extract dynamical information from shortest paths algorithms. ${ }^{87}$ As we show in Sec. III, the $k$ shortest paths for this choice of edge costs are very closely related for Markov chains exhibiting metastability, and may together account for only a small proportion of the total $\mathcal{A} \leftarrow \mathcal{B}$ probability flux. Hence, for metastable Markov chains, the number of shortest paths that can be feasibly determined by the REA is typically insufficient to converge the pathwise sum for the MFPR (Eq. 15).

An alternative choice, which may be especially useful in the metastable regime, is to use edge costs that represent a global measure of the probability flux. The contribution of a transition path $\xi \equiv\left\{a \in \mathcal{A} \equiv i_{n+1} \leftarrow\right.$ $\left.i_{n} \leftarrow \ldots \leftarrow i_{1} \leftarrow b \in \mathcal{B}\right\}$, where $i_{1}, \ldots, i_{n} \notin \mathcal{A} \cup \mathcal{B}$, to the total reactive steady-state ${ }^{104}$ flux $\mathcal{J}_{\mathcal{A B}}$ is ${ }^{105}$

$$
\begin{equation*}
\mathcal{J}[\xi]=f_{i_{1} b}^{+} \prod_{k=1}^{n} \frac{f_{i_{k+1} i_{k}}^{+}}{f_{i_{k}}^{+}} \tag{16}
\end{equation*}
$$

Here,

$$
f_{i j}^{+}= \begin{cases}\pi_{j} T_{i j}\left(q_{i}^{+}-q_{j}^{+}\right), & \text {if } q_{i}^{+}>q_{j}^{+}  \tag{17}\\ 0, & \text { otherwise }\end{cases}
$$

is the net $\mathcal{A} \leftarrow \mathcal{B}$ reactive flux along the $i \leftarrow j$ edge, ${ }^{71}$ $f_{j}^{+}=\sum_{\gamma} f_{\gamma j}^{+}=\sum_{\gamma} f_{j \gamma}^{+}, \pi_{j}$ is the stationary probability for the $j$-th node, ${ }^{83,106}$ and $q_{j}^{+}$is the forward committor probability for the $j$-th node, ${ }^{107-109}$ i.e. the probability that a trajectory initialized at node $j$ hits the target set of nodes $\mathcal{A}$ before hitting the initial state $\mathcal{B} .{ }^{31}$ Eq. 16 implies the following definition for the $i \leftarrow j$ edge costs:

$$
\mathcal{R}[\{i \leftarrow j\}]= \begin{cases}-\ln \frac{f_{i j}^{+}}{f_{j}^{+}}, & \text {if } \quad j \notin \mathcal{B}  \tag{18}\\ -\ln f_{i j}^{+}, & \text {otherwise }\end{cases}
$$

Unlike the local edge costs based on transition probabilities, i.e. $-\ln T_{i j}$, the global edge costs based on reac-
tive fluxes (Eq. 18) represent the $\mathcal{A} \leftarrow \mathcal{B}$ transition mechanism when the system has reached a steady-state. ${ }^{71}$ That is, $f_{i j}^{+}$(Eq. 17) is the net productive flux along the $i \leftarrow j$ edge for the equilibrium FPPE, ${ }^{104}$ and therefore $\mathcal{J}_{\mathcal{A B}}$ is the total steady-state reactive $\mathcal{A} \leftarrow \mathcal{B}$ flux. ${ }^{105}$ This stationary flux directly relates to the steady-state rate constant, which is the dynamical observable associated with the equilibrium FPPE. ${ }^{85}$ Note that it is also possible to define net reactive fluxes along individual edges, and hence edge costs to determine flux-paths (cf. Eq. 18), for the nonequilibrium FPPE, ${ }^{62,104}$ for which the MFPT is the associated dynamical observable.

When using edge costs given by Eq. 18, the weighted network representing the Markov chain in the shortest paths algorithm has unidirectional edges, which are directed such that paths are forced to proceed productively through a series of isocommittor cuts in the network. ${ }^{110}$ Hence, there are no loops in the network (i.e. all paths based on the choice of edge costs representing the net reactive flux are simple ${ }^{89}$ ), and the sum over transition flux-paths to obtain the total reactive flux, $\mathcal{J}_{\mathcal{A B}}=$ $\sum_{\xi \in\{\mathcal{A} \leftarrow \mathcal{B}\}} \mathcal{J}[\xi]$, is finite. Since the set of nodes $\mathcal{S} \backslash \mathcal{A}$ is not transient when the edge costs are given by Eq. 18, the REA as presented in Algorithm 1 must be adapted to account for the situation where a candidate path does not exist, as outlined in the original description of the REA (see Ref. 70). With this minor modification, which is also required when using local edge costs $-\ln T_{i j}$ for Markov chains that are reducible, ${ }^{3}$ the REA can be used to obtain the complete set of reactive $\mathcal{A} \leftarrow \mathcal{B}$ flux-paths and their contributions to the total reactive flux.

For ill-conditioned Markov chains, evaluation of the edge costs in Eq. 18 is highly susceptible to numerical error. ${ }^{37-47}$ Hence, in the metastable regime, the stationary probabilities $\left\{\pi_{j}\right\}$ required in Eq. 17 should be determined by a numerically stable method, for example using the GTH algorithm ${ }^{83,106}$ or an uncoupling-coupling procedure. ${ }^{77,111-113}$ In Ref. 104, we report a state reduction algorithm for the efficient and robust computation of the committor probabilities $\left\{q_{j}^{+}\right\}$.

## III. NUMERICAL RESULTS

We illustrate our methodology with results for the model eight-state CTMC considered in Ref. 115, for which the disconnectivity graph ${ }^{114}$ is shown in Fig. 1. The system corresponds to a coarse-grained representation of an energy landscape, ${ }^{57}$ with a discrete set of states connected via energy barriers. The internode transition rates have an Arrhenius form, ${ }^{116}$ dependent on the temperature $T,{ }^{35}$ and characterize the Markovian network dynamics in terms of branching probabilities ${ }^{73}$ and mean waiting times for transitions from nodes. ${ }^{30}$ A complete specification of this model system is given in Appendix A.

The variation in the heights of energy barriers for transitions in the system induces a separation of timescales that increases with decreasing temperature. Both the


FIG. 1. Disconnectivity graph ${ }^{114}$ representing the energy landscape of the model eight-state CTMC, at a threshold energy increment of $\Delta E=2$. The branches of the tree terminate at the energies of the corresponding nodes. A fork indicates that there exists a path between the corresponding sets of nodes via a highest-energy transition state that lies in between the neighbouring energy thresholds. The branches corresponding to the absorbing and initial nodes, which constitute the sets $\mathcal{A}$ and $\mathcal{B}$, are colored red and blue, respectively.
$\mathcal{A} \leftarrow \mathcal{B}$ MFPT and the steady-state $\mathcal{A} \leftarrow \mathcal{B}$ reactive flux vary by around twenty orders of magnitude in the range of inverse temperature $1 / T$ from 0.1 to 2 (Fig. 2). At low temperatures, conventional linear algebra methods to determine MFPTs fail owing to numerical instability. ${ }^{36}$ This CTMC therefore provides a useful benchmark problem, since Markov chains representing realistic dynamical processes are frequently metastable ${ }^{117-124}$ and therefore ill-conditioned. ${ }^{37-48}$ We consider a single source node and a single sink node. Thus there is no contribution to the path probability from the initial node occupation probability distribution, and therefore $\mathcal{W}[\xi] \equiv \mathcal{P}[\xi]$.

Fig. 3 shows the probability distribution for the path action in the FPPE obtained from kinetic path sampling ${ }^{68}$ (kPS) simulations, and the mean path action computed by the nodewise iterative formulation of the generalized GT algorithm (Eqs. 1 and 14), for the eight-state CTMC at an inverse temperature of $1 / T=2$. At low temperatures, where the model Markov chain is metastable, the use of the standard kinetic Monte Carlo algorithm ${ }^{125}$ to sample $\mathcal{A} \leftarrow \mathcal{B}$ first passage paths is unfeasibly inefficient. ${ }^{52}$ Kinetic path sampling ${ }^{68,69}$ (kPS) can instead be used to sample the numbers of individual $i \leftarrow j$ transitions along $\mathcal{A} \leftarrow \mathcal{B}$ paths, and hence the probability distribution for first passage path rewards.


FIG. 2. Variation in the $\mathcal{A} \leftarrow \mathcal{B}$ mean first passage time, $\mathcal{T}_{\mathcal{A B}}$, and the steady-state $\mathcal{A} \leftarrow \mathcal{B}$ reactive flux, $\mathcal{J}_{\mathcal{A} \mathcal{B}}$, with inverse temperature, for the eight-state CTMC (Fig. 1).


FIG. 3. Probability distribution of the path action for the ensemble of $\mathcal{A} \leftarrow \mathcal{B}$ first passage paths, obtained from 100000 kinetic path sampling ${ }^{68}$ iterations, and mean path action obtained using the generalized GT algorithm (pink), for the eight-state CTMC (Fig. 1) at an inverse temperature of $1 / T=2$. At this temperature, the CTMC is strongly metastable.

The path action distribution at this low value of the temperature is fat-tailed. ${ }^{126-128}$ That is, there is a small but appreciable proportion of probability mass at extreme values, which makes a substantial contribution to the mean, and thus the second and higher central moments of the distribution are significant. Hence, reliable estimation of the mean path action in the metastable regime by sampling paths requires a very large number of observations, even for this low-dimensional system.

It is common in dynamical models of realistic systems for first passage time distributions associated with transitions between two endpoint states to be fat-tailed. ${ }^{52,126-128}$ One approach to examine this phenomenon is to compute the proportion of the $\mathcal{A} \leftarrow \mathcal{B}$
probability flux that can be attributed to the dominant first passage paths, and to examine the convergence of the sum for the expectation of the first passage time (cf. Eq. 15) when an increasing number of paths are included.

In Fig. 4, we compare the mean path action computed using GT to the values associated with the highestprobability paths determined by the REA, ${ }^{70}$ for the eight-state CTMC (Fig. 1) at an inverse temperature of $1 / T=2$. Fig. 4 also illustrates the convergence of the pathwise sum (Eq. 15) for the MFPT, $\mathcal{T}_{\mathcal{A B}}$. The dominant first passage paths are highly atypical, and are associated with values for the path action and time that are several standard deviations smaller than the means of the respective distributions. Because these distributions are fat-tailed, the 100000 highest-probability paths account for a fraction of only around $5 \times 10^{-7}$ of the to$\operatorname{tal} \mathcal{A} \leftarrow \mathcal{B}$ probability flux, and the pathwise sum for the MFPT is far from converged. The probabilities associated with the 100000 dominant first passage paths are close to uniform at this low temperature, suggesting that the paths determined by the REA are all closely related. Indeed, the small number of paths (around 5000) with the very highest probabilities are very similar, involving a small number of transitions via the lowest energy barriers. However, subsequent shortest paths can be divided into two families: longer paths involving only the most favourable transitions, and short paths proceeding via one or more alternative, less favourable, transitions. Our analysis of this simple model demonstrates that, while examination of the properties of the highest-probability first passage paths is insightful, this analysis alone may be misleading, and it is crucial to calculate the expectation for the path property of interest.

The extent to which the first passage time and path action distributions are fat-tailed depends strongly on the metastability of the Markov chain. Fig. 5 shows the evolution of the cumulative sum of probabilities for the 100000 dominant $\mathcal{A} \leftarrow \mathcal{B}$ first passage paths at varying temperature. Notably, the convergence of the path probability sum follows the same pattern at all temperatures. There are a very small number (roughly 10-100) of first passage paths with relative probabilities that are particularly high, and the profile for the cumulative $\mathcal{A} \leftarrow \mathcal{B}$ path probability then reaches a plateau. Thus, even for the model system with a small state space and in the high-temperature limit, it is unfeasible to obtain the set of paths that account for the majority (say, $>90 \%$ ) of the $\mathcal{A} \leftarrow \mathcal{B}$ first passage path probability by shortest paths algorithms, which would require an exceptionally large number (more than $10^{10}$ ) of paths to be determined. Nonetheless, at high temperatures, where there is no significant separation of characteristic timescales, almost $50 \%$ of the total $\mathcal{A} \leftarrow \mathcal{B}$ probability flux is accounted for by the 100000 dominant paths, and therefore it is feasible to obtain a representative picture of the global dynamics using the REA with edge costs that represent a local measure of the probability flux. At low temper-


FIG. 4. Values of the path action (blue) and cumulative pathwise sum (Eq. 15) for the MFPT (pink) for the 100000 highest-probability $\mathcal{A} \leftarrow \mathcal{B}$ first passage paths in the eightstate CTMC (Fig. 1) at an inverse temperature of $1 / T=2$. At this temperature, the CTMC is strongly metastable. The value of the MFPT, computed using GT, is indicated by a dashed pink line. The paths were determined using the recursive enumeration algorithm (Algorithm 1). ${ }^{70}$ Note the small scale for the path action axis.


FIG. 5. Cumulative sum of $\mathcal{A} \leftarrow \mathcal{B}$ first passage path probabilities at varying temperature for the eight-state CTMC (Fig. 1), accounting for an increasing number of the highestprobability paths determined by the recursive enumeration algorithm (Algorithm 1). ${ }^{70}$ The annotations denote the value of the inverse temperature, i.e. $1 / T$.
atures, however, the set of shortest paths accounts for a negligible proportion of the total probability flux, and is therefore not kinetically relevant.

To provide quantitative information on the relative importance of alternative families of first passage paths, we use the REA employing the edge costs given in Eq. 18, which are based on the net reactive flux along individual edges (Eq. 17). The decomposition of the total reactive $\mathcal{A} \leftarrow \mathcal{B}$ steady-state flux $\mathcal{J}_{\mathcal{A B}}$ (Eq. 16) into contributions
from individual simple transition flux-paths, at varying temperature, is shown in Fig. 6. The observed order of committor probabilities for nodes, which does not change with temperature for this system, yields a pattern of unidirectional net reactive fluxes associated with a total of $36 \mathcal{A} \leftarrow \mathcal{B}$ simple flux-paths. Evidently, the reactive flux becomes increasingly localized among a small subset of transition flux-paths with decreasing temperature (increasing metastability). In the high-temperature regime $(1 / T=0.1)$, the single flux-path associated with the largest contribution to the pathwise sum for $\mathcal{J}_{\mathcal{A B}}$ contributes around a third of the total reactive flux, and the 15 dominant simple paths are required to account for almost all $(>99 \%)$ of the total reactive flux. Conversely, in the low-temperature regime $(1 / T=2)$, the highestflux simple path contributes more than half of the total reactive flux, and the vast majority of the total $\mathcal{A} \leftarrow \mathcal{B}$ flux is associated with the four dominant flux-paths.

Because the set of simple flux-paths is finite, decomposition of the total reactive flux $\mathcal{J}_{\mathcal{A B}}$ into additive contributions from transition flux-paths (cf. Eq. 16) provides a representative picture of the global dynamics even in the metastable regime, where the set of highest-probability first passage paths accounts for a negligible proportion of the total path probability (Fig. 5). By effectively grouping together paths that are related by unproductive flickering, ${ }^{52}$ a quantitative comparison of the kinetic relevance of different competing $\mathcal{A} \leftarrow \mathcal{B}$ transition mechanisms is recovered. This shortest paths analysis is an alternative to the augmenting paths algorithm of Ref. 71, which distinguishes families of simple flux-paths on the basis of their associated dynamical bottleneck edges. ${ }^{105}$ Moreover, computation of the committor probabilities ${ }^{109}$ by a GT-like algorithm ${ }^{104}$ or alternative linear algebra methods, ${ }^{108}$ and determination of the shortest paths using the REA, scales favourably with dimensionality of the Markov chain. Hence, the complete set of simple flux-paths can be feasibly computed using the REA for sparse networks comprising several tens of thousands of nodes.

## IV. CONCLUSIONS

We have derived a general expression for renormalized rewards (Eq. 3) associated with arbitrary paths on a censored Markov chain. ${ }^{16,77-82}$ We have also derived numerically stable iterative (Eqs. 1 and 14) and block (Eqs. 12 and 13) graph transformation ${ }^{28-32}$ (GT) procedures to compute the mean reward for the ensemble of first passage paths, ${ }^{57,62}$ i.e. the MFPR for a transition from an initial set of nodes $\mathcal{B}$ to an absorbing set of nodes $\mathcal{A}$. These formulations are applicable to both discrete- and continuous-time finite Markov chains. ${ }^{35}$ If the system is not metastable, so that the transition probability matrix is well-conditioned, then MFPRs for transitions from all nonabsorbing nodes can be computed simultaneously using a single matrix inversion operation (Eq. 8). ${ }^{85}$


FIG. 6. Cumulative sum of relative contributions (cf. Eq. 16) to the total $\mathcal{A} \leftarrow \mathcal{B}$ steady-state reactive flux, $\mathcal{J}_{\mathcal{A B}}$, from alternative simple flux-paths, for the eight-state CTMC (Fig. 1) at varying temperature. There are 36 simple flux-paths in total, but no more than 15 transition flux-paths are required to account for the vast majority ( $>99 \%$ ) of the total reactive flux for all temperatures shown. The annotations denote the value of the inverse temperature, i.e. $1 / T$. The flux-paths were determined using the recursive enumeration algorithm ${ }^{70}$ (Algorithm 1) with edge costs based on net reactive fluxes (Eq. 18).

Knowledge of the expectation for the probability distributions of path properties in the FPPE is useful for assessing the convergence when sampling these distributions, which are frequently observed to be fattailed ${ }^{126-128}$ owing to the existence of rare events in dynamical models for realistic systems. ${ }^{124}$

The mean values for first passage path properties can also be compared to the values associated with the highest-probability paths ${ }^{74,91}$ determined by the recursive enumeration algorithm (REA), ${ }^{70}$ to assess the extent to which the characteristics of the dominant first passage paths ${ }^{87}$ are typical or otherwise. The shortest paths analysis allows us to evaluate the dominant terms in the pathwise sum (Eq. 15) for the expectation of a first passage path property, such as the $\mathcal{A} \leftarrow \mathcal{B}$ MFPT. ${ }^{36}$ Even for low-dimensional Markov chains that do not feature a separation of characteristic timescales, a substantial proportion of the $\mathcal{A} \leftarrow \mathcal{B}$ probability flux is attributable to an exceptionally large number of paths each associated with a very small probability. Hence, the set of shortest paths alone typically accounts for only a fraction of the pathwise sum for the MFPR.

In the metastable regime, low-probability paths comprising a very large number of transitions account for the overwhelming majority of the total first passage path probability, and the set of shortest paths alone is therefore not kinetically relevant. Alternative edge costs reflecting the global dynamics (Eq. 18) can be employed to exactly decompose the reactive steady-state $\mathcal{A} \leftarrow \mathcal{B}$ flux, $\mathcal{J}_{\mathcal{A B}}$, into a sum of contributions from simple flux-paths.

This formulation is exact, ${ }^{105}$ and provides a complementary viewpoint to the typical approach of decomposing $\mathcal{J}_{\mathcal{A B}}$ into additive contributions (cf. Eq. 17) from members of a set of edges that together constitute an $\mathcal{A}-\mathcal{B}$ cut in the network. ${ }^{71}$ The latter framework effectively groups together transition paths that share the same dynamical bottleneck edge of the cut set, and is therefore best suited to compare the relative importance of individual edges comprising a chosen $\mathcal{A}-\mathcal{B}$ cut. ${ }^{129}$ The fluxpathwise analysis proposed in the current work provides a more detailed analysis that can be used to quantitatively understand the characteristic features of the whole $\mathcal{A} \leftarrow \mathcal{B}$ transition mechanism.

The GT and REA procedures scale favourably and can be applied to complex networks with state spaces comprising several hundred thousand nodes. ${ }^{31,91}$ The methodology described herein will therefore provide fundamental insight into a variety of first passage processes in stochastic models. For instance: what is the single most probable route for the extinction of a species in a population dynamics ${ }^{16-18}$ process? What are the most probable paths that together account for a specified proportion of the first passage probability flux, and what is the collective contribution of these paths to the MFPT?

## V. SUPPLEMENTARY INFORMATION

The graph transformation, kinetic path sampling, and recursive enumeration algorithms are implemented in DISCOTRESS. DISCOTRESS is a highly general C++ program for simulation and analysis of arbitrary discrete- and continuous-time Markov chains, including metastable models that are numerically challenging. DISCOTRESS is freely available software under the GNU General Public License. Code, documentation, and tutorials are provided at https://github.com/danieljsharpe.

## VI. ACKNOWLEDGEMENTS

DJS gratefully acknowledges the Cambridge Commonwealth, European, and International Trust for a PhD scholarship. DJW gratefully acknowledges support from the EPSRC.

## Appendix A: Description of the model system

Here we provide a complete specification of the model eight-state CTMC (for which the disconnectivity graph is shown in Fig. 1) that was employed to demonstrate our proposed methodology in Sec. III. Let the diagonal element $E_{j j}$ of the matrix $\mathbf{E}$ represent the energy of the $j$ th node of the Markov chain, and the off-diagonal element $E_{i j}$ (for $i \neq j$ ) represent the energy of the transition state connecting nodes $i$ and $j$, so that $E_{i j}-E_{j j}$ is the energy
barrier for the $i \leftarrow j$ transition. The matrix $\mathbf{E}$ is:

$$
\mathbf{E}=\left(\begin{array}{cccccccc}
0 & 28 & 103 & \infty & \infty & \infty & 18 & \infty  \tag{A1}\\
\vdots & 8 & 20 & 25 & 30 & \infty & 22 & \infty \\
\vdots & \ddots & 10 & 35 & 25 & \infty & 83 & \infty \\
\vdots & \ddots & \ddots & 9 & 20 & 125 & \infty & 24 \\
\vdots & \ddots & \ddots & \ddots & 7 & 26 & \infty & 36 \\
\vdots & \ddots & \ddots & \ddots & \ddots & 1 & \infty & 19 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 1 & \infty \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 2
\end{array}\right),
$$

where off-diagonal entries $E_{i j}$ equal to $\infty$ indicate that a direct connection between the $i$ and $j$ nodes does not exist. Note that the matrix $\mathbf{E}$ is symmetric about the diagonal, hence only the upper triangular elements are specified in Eq. A1. In the numerical results of Sec. III, we defined the initial state to be $\mathcal{B}=\{1\}$ and the absorbing state to be $\mathcal{A}=\{8\}$.

The $i \leftarrow j$ transition rate is then given by the Arrhenius expression ${ }^{116}$

$$
\begin{equation*}
K_{i j}=\exp \left(-\frac{E_{i j}-E_{j j}}{T}\right) \quad \forall i \neq j \tag{A2}
\end{equation*}
$$

where $T$ is an effective temperature and we have set all the pre-exponential factors to unity for simplicity. The calculations in Sec. III used the Markov chain parameterized by the branching probability matrix as defined in Sec. II A. The eight-state model system provides an ideal benchmark to test the numerical stability of algorithms, since the extent to which the Markov chain is more or less ill-conditioned can be tuned by the value of the parameter $T(c f$. Fig. 2).

[^1]${ }^{14}$ E. Seneta, Linear Algebra Appl. 34, 259-267 (1980).
${ }^{15}$ D. P. Heyman, J. Appl. Probab. 32, 893-901 (1995).
${ }^{16}$ T. Dayar, H. Hermanns, D. Spieler, and V. Wolf, Numer. Linear Algebra Appl. 18, 931-946 (2011).
${ }^{17}$ E. Renshaw, Modelling Biological Populations in Space and Time (Cambridge University Press, Cambridge, UK, 1991).
${ }^{18}$ J. Hofbauer and K. Sigmund, Evolutionary Games and Population Dynamics (Cambridge University Press, Cambridge, UK, 1998).
${ }^{19}$ J. Goutsias, J. Chem. Phys. 122, 184102 (2005).
${ }^{20}$ Y. Cao, D. T. Gillespie, and L. R. Petzold, J. Chem. Phys. 122, 014116 (2005).
${ }^{21}$ W. E, D. Liu, and E. Vanden-Eijnden, J. Chem. Phys. 123, 194107 (2005).
${ }^{22}$ W. E, D. Liu, and E. Vanden-Eijnden, J. Comput. Phys. 221, 158-180 (2007).
${ }^{23}$ H.-W. Kang and T. G. Kurtz, Ann. Appl. Probab. 23, 529-593 (2013).
${ }^{24}$ J. Goutsias and G. Jenkinson, Phys. Rep. 529, 199-264 (2013).
${ }^{25}$ B. Munsky and M. Khammash, J. Chem. Phys. 124, 044104 (2006).
${ }^{26}$ K. N. Dinh and R. B. Sidje, Phys. Biol. 13, 035003 (2016).
${ }^{27}$ R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. (MIT Press, Cambridge, Massachusetts, 2018).
${ }^{28}$ S. A. Trygubenko and D. J. Wales, Mol. Phys. 104, 1497-1507 (2006).
${ }^{29}$ S. A. Trygubenko and D. J. Wales, J. Chem. Phys. 124, 234110 (2006).
${ }^{30}$ D. J. Wales, J. Chem. Phys. 130, 204111 (2009).
${ }^{31}$ J. D. Stevenson and D. J. Wales, J. Chem. Phys. 141, 041104 (2014).
${ }^{32}$ R. S. MacKay and J. D. Robinson, Phil. Trans. Roy. Soc. A 376, 20170232 (2018).
${ }^{33}$ S. Redner, A Guide to First-Passage Processes (Cambridge University Press, Cambridge, UK, 2012).
${ }^{34}$ R. Metzler, G. Oshanin, and S. Redner, First-Passage Phenomena and Their Applications (World Scientific, Singapore, 2014).
${ }^{35}$ D. Kannan, D. J. Sharpe, T. D. Swinburne, and D. J. Wales, J. Chem. Phys. 153, 244108 (2020).
${ }^{36}$ J. J. Hunter, Linear Algebra Appl. 549, 100-122 (2018).
${ }^{37}$ W. Grassmann and D. A. Stanford, "Matrix analytic methods," in Computational Probability, edited by W. Grassmann (Springer, New York, 2000) pp. 153-203.
${ }^{38}$ J. R. Koury, D. F. McAllister, and W. J. Stewart, SIAM J. Alg. Discr. Meth. 5, 164-186 (1984).
${ }^{39}$ D. P. Heyman, SIAM J. Alg. Discr. Meth. 8, 226-232 (1987).
${ }^{40}$ D. P. Heyman and A. Reeves, ORSA J. Comp. 1, 52-60 (1989).
${ }^{41}$ B. Philippe, Y. Saad, and W. J. Stewart, Oper. Res. 40, 11561179 (1992).
${ }^{42}$ C. A. O'Cinneide, Numer. Math. 65, 109-120 (1993).
${ }^{43}$ C. D. Meyer Jr, SIAM J. Matrix Anal. Appl. 15, 715-728 (1994).
${ }^{44}$ C. A. O'Cinneide, Numer. Math. 73, 507-519 (1996).
${ }^{45}$ D. P. O'Leary and Y.-J. J. Wu, SIAM J. Matrix Anal. Appl. 17, 470-488 (1996).
${ }^{46}$ D. J. Hartfiel and C. D. Meyer Jr, Linear Algebra Appl. 272, 193-203 (1998).
${ }^{47}$ J. L. Barlow, SIAM J. Matrix Anal. Appl. 22, 230-241 (2000).
${ }^{48}$ M. Benzi, Linear Algebra Appl. 386, 27-49 (2004).
${ }^{49}$ O. Valsson, P. Tiwary, and M. Parrinello, Annu. Rev. Phys. Chem. 67, 159-184 (2016).
${ }^{50}$ B. Trendelkamp-Schroer and F. Noé, Phys. Rev. X 6, 011009 (2016).
${ }^{51}$ M. A. Novotny, Phys. Rev. Lett. 74, 1-5 (1995).
${ }^{52}$ D. J. Sharpe and D. J. Wales, J. Chem. Phys. 153, 024121 (2020).
${ }^{53}$ M. Manhart and A. V. Morozov, Phys. Rev. Lett. 111, 088102 (2013).
${ }^{54}$ U. Seifert, Rep. Prog. Phys. 75, 126001 (2012).
${ }^{55}$ M. V. S. Bonança and C. Jarzynski, Phys. Rev. E 93, 022101 (2016).
${ }^{56}$ C. Maes and K. Netočný, J. Stat. Phys. 110, 269-310 (2003).
${ }^{57}$ M. Manhart, W. Kion-Crosby, and A. V. Morozov, J. Chem. Phys. 143, 214106 (2015).
${ }^{58}$ S. X. Sun, Phys. Rev. Lett. 96, 210602 (2006).
${ }^{59}$ B. Harland and S. X. Sun, J. Chem. Phys. 127, 104103 (2007).
${ }^{60}$ T. Oppelstrup, V. V. Bulatov, A. Donev, M. H. Kalos, G. H. Gilmer, and B. Sadigh, Phys. Rev. E 80, 066701 (2009).
${ }^{61}$ S. Hwang, D.-S. Lee, and B. Kahng, Phys. Rev. Lett. 109, 088701 (2012).
${ }^{62}$ M. von Kleist, C. Schütte, and W. Zhang, J. Stat. Phys. 170, 809-843 (2018).
${ }^{63}$ P. Dixit, A. Jain, G. Stock, and K. A. Dill, J. Chem. Theory Comput. 11, 5464-5472 (2015).
${ }^{64}$ P. Dixit, J. Wagoner, C. Weistuch, S. Pressé, K. Ghosh, and K. A. Dill, J. Chem. Phys. 148, 010901 (2018).
${ }^{65}$ C. Van den Broeck, "Stochastic thermodynamics: A brief introduction," in Proceedings of the International School of Physics "Enrico Fermi" Vol. 184: Physics of Complex Colloids, edited by C. Bechinger, F. Sciortino, and P. Ziherl (IOS Press, Amsterdam, 2013) pp. 155-193.
${ }^{66}$ U. Seifert, Phys. Rev. Lett. 95, 040602 (2005).
${ }^{67}$ M. J. Goldberg and S. Kim, Appl. Comput. Harmon. Anal. 30, 37-46 (2011).
${ }^{68}$ M. Athènes and V. V. Bulatov, Phys. Rev. Lett. 113, 230601 (2014).
${ }^{69}$ M. Athènes, S. Kaur, G. Adjanor, T. Vanacker, and T. Jourdan, Phys. Rev. Materials 3, 103802 (2019).
${ }^{70}$ V. M. Jiménez and A. Marzal, "Computing the $k$ shortest paths: a new algorithm and experimental comparison," in Algorithm Engineering: 3rd International Workshop, WAE'99, London, $U K$, edited by J. S. Vitter and C. D. Zaroliagis (Springer Berlin, Heidelberg, 1999) pp. 15-29.
${ }^{71}$ P. Metzner, C. Schütte, and E. Vanden-Eijnden, Multiscale Model. Simul. 7, 1192-1219 (2009).
${ }^{72}$ J. G. Kemeny and J. L. Snell, Theory Prob. Its Appl. 6, 101-105 (1961).
${ }^{73}$ S. A. Serebrinsky, Phys. Rev. E 83, 037701 (2011).
${ }^{74}$ D. J. Sharpe and D. J. Wales, J. Chem. Phys. 151, 124101 (2019).
${ }^{75}$ J. K. Weber, D. Shukla, and V. S. Pande, Proc. Natl. Acad. Sci. USA 112, 10377-10382 (2015).
${ }^{76}$ S. Vaikuntanathan, T. R. Gingrich, and P. L. Geissler, Phys. Rev. E 89, 062108 (2014).
${ }^{77}$ C. D. Meyer Jr., SIAM Rev. 31, 240-272 (1989).
${ }^{78}$ Y. Q. Zhao and D. Liu, J. Appl. Probab. 33, 623-629 (1996).
${ }^{79}$ E. Meerbach, C. Schütte, and A. Fischer, Linear Algebra Appl. 398, 141-160 (2005).
${ }^{80}$ N. Pekergin, T. Dayar, and D. N. Alparslan, Eur. J. Oper. Res. 165, 810-825 (2005).
${ }^{81}$ J.-M. Fourneau, N. Pekergin, and S. Younès, "Censoring Markov chains and stochastic bounds," in EPEW 2007: Formal Methods and Stochastic Models for Performance Evaluation, edited by K. Wolter (Springer Berlin, Heidelberg, 2007) pp. 213-227.
${ }^{82}$ A. Milias-Argeitis and J. Lygeros, J. Chem. Phys. 138, 184109 (2013).
${ }^{83}$ W. K. Grassmann, M. I. Taksar, and D. P. Heyman, Oper. Res. 33, 1107-1116 (1985).
${ }^{84}$ D. Kannan, D. J. Sharpe, T. D. Swinburne, and D. J. Wales, "Dimensionality reduction of finite markov chains by renormalization," (2020), (unpublished).
${ }^{85}$ T. D. Swinburne and D. J. Wales, J. Chem. Theory Comput. 16, 2661-2679 (2020).
${ }^{86}$ G. R. Bowman, L. Meng, and X. Huang, J. Chem. Phys. 139, 121905 (2013).
${ }^{87}$ S. Viswanath, S. M. Kreuzer, A. E. Cardenas, and R. Elber, J. Chem. Phys. 139, 174105 (2013).
${ }^{88}$ D. R. Shier, Networks 9, 195-214 (1979).
${ }^{89}$ E. Q. V. Martins, Eur. J. Op. Res. 18, 123-130 (1984).
${ }^{90}$ E. Q. V. Martins and M. M. B. Pascoal, 4OR 1, 121-133 (2003).
${ }^{91}$ J. M. Carr and D. J. Wales, "The energy landscape as a computational tool," in Latest Advances in Atomic Cluster Collisions: Structure and Dynamics from the Nuclear to the Biological Scale, edited by A. Solov'yov and J.-P. Connerade (Imperial College Press, London, 2008) pp. 321-330.
${ }^{92}$ D. Eppstein, SIAM J. Comput. 28, 652-673 (1999).
${ }^{93}$ V. M. Jiménez and A. Marzal, "A lazy version of Eppstein's $k$ shortest paths algorithm," in International Workshop on Experimental and Efficient Algorithms, WEA 2003, edited by K. Jansen, M. Margraf, M. Mastrolilli, and J. D. P. Rolim (Springer Berlin, Heidelberg, 2003) pp. 179-191.
${ }^{94}$ J. A. Azevedo, M. E. O. S. Costa, J. J. R. E. S. Madeira, and E. Q. V. Martins, Eur. J. Op. Res. 69, 97-106 (1993).
${ }^{95}$ J. A. Azevedo, J. J. R. E. S. Madeira, E. Q. V. Martins, and F. M. A. Pires, Eur. J. Op. Res. 73, 188-191 (1994).
${ }^{96}$ E. Q. V. Martins, M. M. B. Pascoal, and J. L. E. dos Santos, Int. J. Found. Comp. Sci. 10, 247-261 (1999).
${ }^{97}$ E. Q. V. Martins and J. L. E. dos Santos, Investigação Operacional 20, 47-62 (2000).
${ }^{98}$ E. Q. V. Martins, M. M. B. Pascoal, and J. L. E. dos Santos, Investigação Operacional 21, 47-60 (2001).
${ }^{99}$ R. E. Bellman and S. E. Dreyfus, Applied Dynamic Programming (Princeton University Press, New Jersey, USA, 1962).
${ }^{100}$ S. E. Dreyfus, Oper. Res. 17, 395-412 (1969).
${ }^{101}$ E. W. Dijkstra, Numer. Math. 1, 269-271 (1959).
${ }^{102}$ M. L. Fredman and R. E. Tarjan, J. Assoc. Comput. Mach. 34, 596-615 (1987).
${ }^{103}$ R. K. Ahuja, K. Mehlhorn, J. B. Orlin, and R. E. Tarjan, J. Assoc. Comput. Mach. 37, 213-223 (1990).
${ }^{104}$ D. J. Sharpe and D. J. Wales, "Numerical analysis of first passage processes in metastable finite Markov chains," (2020), (unpublished).
${ }^{105}$ A. Berezhkovski, G. Hummer, and A. Szabo, J. Chem. Phys. 130, 205102 (2009).
${ }^{106}$ T. J. Sheskin, Oper. Res. 33, 228-235 (1985).
${ }^{107}$ P. Metzner, C. Schütte, and E. Vanden-Eijnden, J. Chem. Phys. 125, 084110 (2006).
${ }^{108}$ J.-H. Prinz, M. Held, J. C. Smith, and F. Noé, Multiscale Model. Simul. 9, 545-567 (2011).
${ }^{109}$ A. M. Berezhkovskii and A. Szabo, J. Chem. Phys. 150, 054106 (2019).
${ }^{110}$ M. K. Cameron and E. Vanden-Eijnden, J. Stat. Phys. 156, 427-454 (2014).
${ }^{111}$ M. Haviv, SIAM J. Numer. Anal. 22, 952-966 (1987).
${ }^{112}$ T. Dayar and W. J. Stewart, SIAM J. Sci. Comput. 17, 287-303 (1996).
${ }^{113}$ Y.-J. J. Wu, "A parallel implementation of the block-GTH algorithm," Tech. Rep. TR-3347 (University of Maryland, 1994).
${ }^{114}$ O. M. Becker and M. Karplus, J. Chem. Phys. 106, 1495-1517 (1997).
${ }^{115}$ T. D. Swinburne, D. Kannan, D. J. Sharpe, and D. J. Wales, J. Chem. Phys. 153, 134115 (2020).
${ }^{116}$ A. Kells, V. Koskin, E. Rosta, and A. Annibale, J. Chem. Phys. 152, 104108 (2020).
${ }^{117}$ M. Weber, S. Kube, L. Walter, and P. Deuflhard, Multiscale Model. Simul. 6, 396-416 (2007).
${ }^{118}$ A. Dickson and A. R. Dinner, Annu. Rev. Phys. Chem. 61, 441-459 (2010).
${ }^{119}$ J. T. Berryman and T. Schilling, J. Chem. Phys. 133, 244101 (2010).
${ }^{120}$ C. Giardina, J. Kurchan, V. Lecomte, and J. Tailleur, J. Stat. Phys. 145, 787-811 (2011).
${ }^{121}$ C. Hartmann, R. Banisch, M. Sarich, T. Badowski, and C. Schütte, Entropy 16, 350-376 (2014).
${ }^{122}$ M. Sarich, R. Banishc, C. Hartmann, and C. Schütte, Entropy 16, 258-286 (2014).
${ }^{123}$ M. K. Cameron, J. Chem. Phys. 141, 184113 (2014).
${ }^{124}$ C. Pérez-Espigares and P. I. Hurtado, Chaos 29, 083106 (2019).
${ }^{125}$ A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. 17, 10-18 (1975).
${ }^{126}$ T. Hoffmann, M. A. Porter, and R. Lambiotte, Phys. Rev. E 86, 046102 (2012).
${ }^{127}$ J.-C. Delvenne, R. Lambiotte, and L. E. C. Rocha, Nat. Com-
mun. 6, 7366 (2015).
${ }^{128}$ E. F. dos Reis, A. Li, and N. Masuda, Phys. Rev. E 102, 052303 (2020).
${ }^{129}$ D. Nagel, A. Weber, and G. Stock, J. Chem. Theory Comput. 16, 7874-7882 (2020).
input : network $\mathcal{G}$ representing a finite Markov chain with state space $\mathcal{S}$, and $i \leftarrow j$ edge
$$
\text { costs }-\ln T_{i j} \forall i, j \in \mathcal{S}
$$
initial (source) node $b \in \mathcal{B}$ and absorbing (sink) node $a \in \mathcal{A}$
total number of highest-probability paths to compute, $k_{\text {tot }}$
output: $k_{\text {tot }}$ highest-probability $a \leftarrow b$ first passage paths; $\xi^{k}(a)$ for $1 \leq k \leq k_{\text {tot }}$

```
\(\xi^{k}(j) \leftarrow \emptyset \quad \forall 1 \leq k \leq k_{\mathrm{tot}}, j \in \mathcal{S} ;\)
\(\mathcal{M}(j) \leftarrow \emptyset \quad \forall j ;\)
\(\xi^{1}(j) \forall j \neq b \leftarrow \operatorname{Dijkstra}(\mathcal{G}) ;\)
/* main loop of REA */
for \(k=2, \ldots, k_{\text {tot }}\) do
    \(\xi^{k}(a) \leftarrow \operatorname{NextPath}(k, a) ;\)
return \(\xi^{k}(a) \quad\) for \(1 \leq k \leq k_{\mathrm{tot}}\);
/* one-to-all shortest path algorithm
function Dijkstra(G) */
    return \(\xi^{1}(j) \leftarrow \arg \min _{\xi(j)} \mathcal{R}\left[\xi(j) \equiv \xi^{1}(i) \cup\{j \leftarrow i\}\right] \quad \forall j \in \mathcal{S} \backslash b ;\)
function \(\operatorname{Pred}(k, j)\)
    return \(k^{\prime}, i \quad\) where \(\xi^{k}(j) \equiv \xi^{k^{\prime}}(i) \cup\{j \leftarrow i\} ;\)
/* function to determine the \(k\)-th most probable first passage path to node \(j\), given that the
        \(1, \ldots,(k-1)\)-th most probable first passage paths are known
        */
function NextPath \((k, j)\)
    if \(k==2\) then
        /* Initialize set of candidates for the next most probable path to node \(j\) */
        \(\mathcal{M}(j) \leftarrow\left\{\xi^{1}(i) \cup\{j \leftarrow i\} \quad \forall i \in \mathcal{D}(j)\right.\) and \(\left.1, i \neq \operatorname{Pred}(1, j)\right\} ;\)
        if \(j==b\) then
            goto selectpath;
    \(k^{\prime}, i \leftarrow \operatorname{Pred}(k-1, j)\);
    /* the ( \(k^{\prime}+1\) )-th shortest path to node \(i\) is a viable parent segment of a candidate path to
        node \(j\). If unknown, compute this path with a recursive call to the NextPath function */
    if \(\xi^{\left(k^{\prime}+1\right)}(i) \equiv \emptyset\) then
        \(\xi^{\left(k^{\prime}+1\right)}(i) \leftarrow \operatorname{NextPath}\left(k^{\prime}+1, i\right) ;\)
    \(\mathcal{M}(j) \leftarrow \mathcal{M}(j) \cup\left\{\xi^{\left(k^{\prime}+1\right)}(i) \cup\{j \leftarrow i\}\right\} ; / /\) add candidate path to list
    selectpath:
        \(\xi^{k}(j) \leftarrow \arg \min _{\xi \in \mathcal{M}(j)} \mathcal{R}[\xi] ; / /\) assign candidate path with lowest cost
        \(\mathcal{M}(j) \leftarrow \mathcal{M}(j) \backslash \xi^{k}(j) ; / /\) remove assigned candidate path from list
    return \(\xi^{k}(j)\);
```

Algorithm 1: Recursive enumeration algorithm ${ }^{70}$ (REA) to compute the $k$ highest-probability first passage paths from an initial node $b$ to an absorbing node $a$ in an irreducible finite Markov chain. $\xi^{k}(j)$ denotes the $k$-th most probable $a \leftarrow b$ path to node $j$. The cost associated with the path $\xi$ is $\mathcal{R}[\xi]=-\sum_{(i \leftarrow j) \in \xi} \ln T_{i j} . \mathcal{D}(j)$ denotes the set of nodes for which a direct $i \leftarrow j$ connection exists. $\mathcal{M}(j)$ denotes the set of candidates for the next most probable path to the $j$-th node. The ordered sequence of transitions along the $k$-th highest-probability path can be obtained by tracing the shortest paths array using the Pred function. For reducible Markov chains, or when using edge costs based on net reactive fluxes (Eq. 18), the NextPath function may encounter the situation where a candidate path to a node cannot be found. In this case, the next shortest path to the node does not exist. If the node in question is the target node $a$, then the main loop of the REA is exited, the complete set of $a \leftarrow b$ paths having been found.


[^0]:    ${ }^{\text {a) }}$ Electronic mail: dw34@cam.ac.uk

[^1]:    ${ }^{1}$ J. R. Norris, Markov Chains (Cambridge University Press, New York, USA, 1997).
    ${ }^{2}$ N. Masuda, M. A. Porter, and R. Lambiotte, Phys. Rep. 716717, 1-58 (2017).
    ${ }^{3}$ J. G. Kemeny and J. L. Snell, Finite Markov Chains (Van Nostrand, New Jersey, USA, 1960).
    ${ }^{4}$ 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).
    ${ }^{5}$ F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, Proc. Natl. Acad. Sci. USA 106, 19011-19016 (2009).
    ${ }^{6}$ 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).
    ${ }^{7}$ B. E. Husic and V. S. Pande, J. Am. Chem. Soc. 140, 2386-2896 (2018).
    ${ }^{8}$ A. Mardt, L. Pasquali, H. Wu, and F. Noé, Nat. Commun. 9, 5 (2018).
    ${ }^{9}$ R. G. Mantell, C. E. Pitt, and D. J. Wales, J. Chem. Theory Comput. 12, 6182-6191 (2016).
    ${ }^{10}$ D. J. Wales, Mol. Phys. 100, 3285-3305 (2002).
    ${ }^{11}$ D. J. Wales, Int. Rev. Phys. Chem. 25, 237-282 (2006).
    ${ }^{12}$ J. A. Joseph, K. Röder, D. Chakraborty, R. G. Mantell, and D. J. Wales, Chem. Commun. 53, 6974-6988 (2017).
    ${ }^{13}$ D. J. Wales, Annu. Rev. Phys. Chem. 69, 401-425 (2018).

