The Journal of Chemical Physics

Mechanisms and topology determination of complex chemical and biological network systems from first-passage theoretical approach

Cite as: J. Chem. Phys. **139**, 144106 (2013); https://doi.org/10.1063/1.4824392
Submitted: 30 July 2013 • Accepted: 23 September 2013 • Published Online: 10 October 2013

Xin Li and Anatoly B. Kolomeisky

Mechanisms and topology determination of complex chemical and biological network systems from first-passage theoretical approach

Xin Li[1] and Anatoly B. Kolomeisky[1,2,a)]
[1]*Department of Chemistry, Rice University, Houston, Texas 77005, USA*
[2]*Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, USA*

(Received 30 July 2013; accepted 23 September 2013; published online 10 October 2013)

The majority of chemical and biological processes can be viewed as complex networks of states connected by dynamic transitions. It is fundamentally important to determine the structure of these networks in order to fully understand the mechanisms of underlying processes. A new theoretical method of obtaining topologies and dynamic properties of complex networks, which utilizes a first-passage analysis, is developed. Our approach is based on a hypothesis that full temporal distributions of events between two arbitrary states contain full information on number of intermediate states, pathways, and transitions that lie between initial and final states. Several types of network systems are analyzed analytically and numerically. It is found that the approach is successful in determining structural and dynamic properties, providing a direct way of getting topology and mechanisms of general chemical network systems. The application of the method is illustrated on two examples of experimental studies of motor protein systems. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4824392]

I. INTRODUCTION

Most chemical and biological systems can be described as complex networks of states.[1–5] To understand how these systems function one needs to determine the structure and rates in these networks. In recent years, significant experimental and theoretical advances in investigating mechanisms of complex chemical and biological systems have been achieved.[6–15] It is now possible to monitor various chemical and biological systems with a single-molecule precision and a high temporal and spatial resolutions.[6–11] Current theoretical methods also allow to calculate many molecular properties with a high accuracy.[12–15] However, a comprehensive determination of structure-function relations for many systems remains an unsolved task.

To understand the difficulty of this issue let us consider a general scheme of a chemical system that has an underlying complex network of dynamically connected states—see Fig. 1. A typical experimental measurement probes some complex event that starts from a state i and ends up at a state j (Fig. 1). As a result, experiments produce distributions of these events, and mechanisms for underlying chemical processes are proposed based on analysis of these data. However, this is an example of inverse problems that are difficult to solve since more than one mechanism could fit the measured distributions. In addition, although all information about mechanisms of complex processes is contained in these distributions, in most cases only the first moments, mean dwell times, are analyzed.

It has been recognized that studies of distributions of events for networks are associated with a first-passage problem in stochastic processes.[16,17] The mathematical analysis of first-passage processes is well developed,[18,19] however it is rarely used for understanding complex systems. A more popular approach is to utilize Monte Carlo computer simulations to generate distributions and compare them with experiments. However, some minimal information on the structure of the network is needed and a computational cost becomes prohibitive for large networks. In many cases, more productive is a matrix method that has been used in ion-channel measurements.[20–22] But it does not work for complex systems with multiple transitions. Recently, a new absorbing boundary method has been proposed for understanding dynamics of molecular motors.[23] Although the method is powerful, it requires some knowledge of the network topology and it is not efficient in obtaining average values such as mean times.[23]

In this article, we develop a new method to predict structural and dynamic information of the underlying chemical and biological networks by analyzing distributions of events *without* assuming any specific model. We propose the following hypothesis: *dwell-time distributions of events between two states as a function of time contain all information on the local structure of the network and dynamic transitions between these states*. The idea here is that at short times the distribution is dominated by events that follow the shortest pathway (in terms of the number of intermediate states) between states. At transient times the effect of other pathways starts to build up, and for large times dynamics is averaged over all possible pathways that reach the final state. Analyzing such distributions at all times for events between various states should provide a model-independent information on the whole network. In this work, the hypothesis is tested via exact analytical calculations and computer simulations for several types of networks. In addition, the method is applied for analyzing dynamics of two motor protein systems.

a)tolya@rice.edu

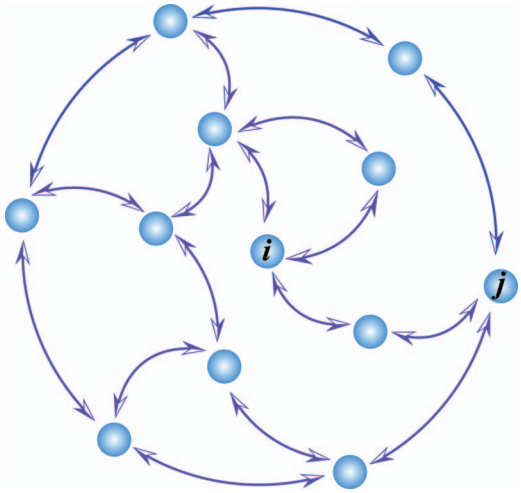

FIG. 1. A schematic view of a general chemical network. Nodes correspond to individual states and arrows describe possible transitions. Experimental measurements typically provide the information on distribution events that start at some initial state i and finish at some final state j.

II. RESULTS

A. First-passage analysis of dwell-time distributions for homogeneous One-Dimensional (1D) networks

We consider first a simplified 1D network of discrete states $j = 0, \ldots, N$, as shown in Fig. 2(a). For real processes it corresponds to dynamics of motor proteins,[12,24] polymer translocation,[7] or protein folding.[6] Suppose that at time $t = 0$ the molecule is in a state $j = m$. The system can transfer to the state $j = m + 1$ with a rate u_m, while the backward transition rate to the state $j = m - 1$ is given by w_m. We are interested in obtaining a distribution function that describes reaching the left end state ($j = 0$) or the right end state ($j = N$) for the first time. Specifically, a first-passage probability function $f_{N,m}(t)$ is defined as a probability that the system starting at time $t = 0$ in the state m will reach the right end state N at time t for the first time before reaching the left end state $j = 0$. In a similar way we define a first-passage probability function $f_{0,m}(t)$ to reach the left end. Let us concentrate on the $f_{N,m}(t)$ function, and all results for $f_{0,m}(t)$ can be obtained from symmetry arguments (changing $m \to N - m$ and $u_m \to w_m$). The temporal evolution of these probability functions can be described by *backward master equations*,[16,17]

$$\frac{\partial f_{N,m}(t)}{\partial t} = u_m f_{N,m+1}(t) + w_m f_{N,m-1}(t) - (u_m + w_m) f_{N,m}(t), \tag{1}$$

with the following boundary conditions:

$$\begin{aligned} f_{N,0}(t) &= 0, \text{ for all } t, \\ f_{N,N}(t) &= \delta(t), \\ f_{N,m}(0) &= 0, \text{ for } 1 \le m \le N-1. \end{aligned} \tag{2}$$

The main idea of our approach is to utilize Laplace transforms of distribution functions defined as

$$\tilde{f}_{N,m}(s) = \int_0^\infty e^{-st} f_{N,m}(t) dt. \tag{3}$$

Due to the boundary conditions we have

$$\tilde{f}_{N,N}(s) = 1, \quad \tilde{f}_{N,0}(s) = 0. \tag{4}$$

To illustrate the power of the method consider a simpler situation with homogeneous transition rates, $u_m = u$ and $w_m = w$ for all m. Then the backward master equation (1) in the Laplace presentation can be written as

$$s\tilde{f}_{N,m} = u\tilde{f}_{N,m+1} + w\tilde{f}_{N,m-1} - (u + w)\tilde{f}_{N,m}. \tag{5}$$

One could define an auxiliary function $a(s) = s + u + w$, and Eq. (5) simplifies into the following:

$$a\tilde{f}_{N,m} = u\tilde{f}_{N,m+1} + w\tilde{f}_{N,m-1}. \tag{6}$$

This equation can be easily solved by assuming that the general form of the solution is $\tilde{f}_{N,m} = Ax^m$ with unknown parameters A and x to be found from the substitution of this ansatz into Eq. (6). It leads to a very compact formula,

$$\tilde{f}_{N,m}(s) = \frac{x_1^m - x_2^m}{x_1^N - x_2^N}, \tag{7}$$

where

$$x_1 = \frac{a + \sqrt{a^2 - 4uw}}{2u}, \quad x_2 = \frac{a - \sqrt{a^2 - 4uw}}{2u}. \tag{8}$$

For a special case of $w = 0$ one can show that $x_1 = \frac{s+u}{u}$ and $x_2 = 0$, yielding

$$\tilde{f}_{N,m}(s) = \left(\frac{u}{s+u}\right)^{N-m}, \tag{9}$$

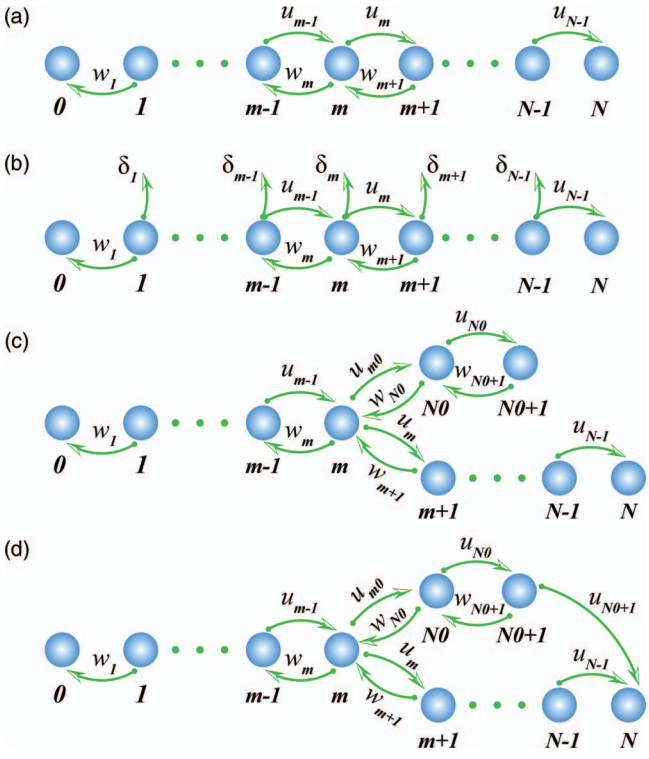

FIG. 2. A schematic view of network systems: (a) a sequential 1D network; (b) 1D network with irreversible detachments; (c) 1D network with a branch; (d) a network with two parallel segments.

which can be easily inverted to obtain the expected first-passage distribution function

$$f_{N,m}(t) = \frac{u^{N-m} t^{N-m-1} e^{-ut}}{(N-m-1)!}. \tag{10}$$

The advantage of using Laplace transforms is that all dynamic properties *at all times* can be calculated explicitly. For example, the overall probability to reach the state N, known as a splitting probability,[16,17] is given by

$$\Pi_{N,m} = \int_0^\infty f_{N,m}(t)dt = \tilde{f}_{N,m}(s=0). \tag{11}$$

The conditional mean first-passage time, which is measured in experiments as a mean dwell time, is defined as

$$\tau_{N,m} = \frac{\tau_m(1)}{\Pi_{N,m}}, \quad \text{with } \tau_m(1) = -\frac{d\tilde{f}_{N,m}(s)}{ds}|_{s=0}. \tag{12}$$

Similar expressions can be obtained for any other dynamic properties of the system.

For simple 1D networks with homogeneous transition rates in one direction $u_m = u$ and no backward transition $w_m = 0$, a simple expression for the first-passage distribution function at early times can be obtained from Eq. (10),

$$\ln f_{N,m}(t) \simeq (N-m-1)\ln t + C, \tag{13}$$

where the exponential part in Eq. (10) is close to 1 and can be neglected at $t \ll 1$ and C is a time-independent constant. This is an important result since it shows a linear relation (on log-log scale) between the first passage distribution and time, and the slope gives the number of intermediate states between the initial and final states. It suggests that dwell-time distributions can be used as a tool for determining the network structure.

To prove this idea further, computer simulations have been performed to model the homogeneous 1D network with transition rates $u_m = u = 5.0\ \text{s}^{-1}$ and $w_m = 0$. The first-passage distributions are shown in Fig. 3(a) for all times and in Fig. 3(b) at early times for different initial states $1 \le m \le 8$ with final state $N = 10$. For the special case with $m = 9$, the distribution gives an exponential decay as expected for the absence of intermediate states. As shown in Fig. 3(a) the analytical predictions from Eq. (10) [dashed lines] agree perfectly with the simulation results. The solid lines in Fig. 3(b)

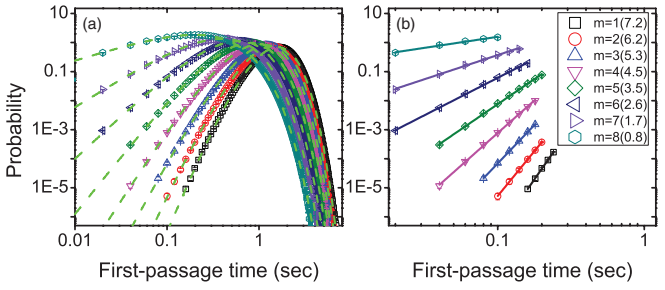

FIG. 3. First passage distributions for a homogeneous 1D network with only forward transitions. The position of the initial state m is varied from 1 to 8 and the final state is $N = 10$. (a) First-passage distributions at all times. (b) First-passage distributions at early times. Symbols correspond to computer simulations. The dashed lines in (a) are given by the analytical expression from Eq. (10), while the solid lines in (b) are from linear fittings of the simulation data. The slope for each curve in (b) is also indicated.

are given by linear fittings of simulation data at small times and the slope of the line (on log-log scale) goes down as the number of intermediate states between the initial and final states decreases. The slopes are very close to expected values $(N-m-1)$, and rounding to the upper integer for each slope gives the exact number of intermediate states. The reason for a slightly smaller value of the slope for each curve from the fittings is the effect of the exponential term that contributes more to the distribution with increasing time. Expressions similar to Eq. (10) have been proposed earlier for simple networks as in Fig. 2(a),[25] where a perturbation theory was applied with transitions from the probed states ignored and only an approximate expression was obtained. Here, in our work an exact expression for the first-passage distribution functions is derived without any assumptions, and more general and realistic networks are discussed. The early temporal behavior of the distribution function can also be obtained from the limiting $s \to \infty$ behavior of corresponding Laplace transform. Then, for the homogeneous case from Eq. (7) with $u_m = u \neq 0$, and $w_m = w \neq 0$, it can be shown that

$$\tilde{f}_{N,m}(s) \simeq \left(\frac{u}{s+u+w}\right)^{N-m}, \tag{14}$$

which corresponds to

$$f_{N,m}(t) \simeq \frac{u^{N-m} t^{N-m-1} e^{-(u+w)t}}{(N-m-1)!} \tag{15}$$

at early times. Again, the structural information of the network for simple 1D networks with homogeneous transition rate is contained in the power law dependence of the first-passage distribution which can be obtained from a simple analysis of distribution function at early times.

B. First-passage analysis of dwell-time distributions for inhomogeneous one-dimensional networks

Our method can be used for arbitrary networks, and it will be shown analytically by considering a general inhomogeneous 1D network system in Fig. 2(a). In this case, the backward master equations in the Laplace presentation are given by

$$a_m \tilde{f}_{N,m} = u_m \tilde{f}_{N,m+1} + w_m \tilde{f}_{N,m-1}, \tag{16}$$

$$a_1 \tilde{f}_{N,1} = u_1 \tilde{f}_{N,2}, \tag{17}$$

$$a_{N-1} \tilde{f}_{N,N-1} = u_{N-1} + w_{N-1} \tilde{f}_{N,N-2}, \tag{18}$$

where

$$a_m = s + u_m + w_m. \tag{19}$$

There are no known exact solutions for inhomogeneous 1D networks, but stimulated by results for homogeneous systems, we propose the following ansatz for Laplace transforms of the distribution functions:

$$\tilde{f}_{N,m}(s) = \frac{\sum_{K=0}^{m-1} \alpha_K(m) s^K}{\sum_{K=0}^{N-1} \alpha_K(N) s^K}, \tag{20}$$

where $\alpha_K(m)$ are unknown coefficients. It suggests that the Laplace transforms of distribution functions can be viewed as ratios of two polynomials with respect to the variable s. Several examples to support this are provided in the Appendix. It is important to note that the coefficients $\alpha_K(m)$ can be determined self-consistently from Eq. (1) which allows us to determine explicitly all dynamic properties of the system. Furthermore, the limiting behavior of distribution functions (for $s \to \infty$ and $s \to 0$) can be obtained knowing only few of coefficients $\alpha_K(m)$, which significantly simplifies calculations. For example, one can show (see the Appendix) that splitting probabilities and mean dwell times to reach the final state N are given by

$$\Pi_{N,m} = \frac{\alpha_0(m)}{\alpha_0(N)}, \tau_{N,m} = \frac{\alpha_1(N)}{\alpha_0(N)} - \frac{\alpha_1(m)}{\alpha_0(m)}. \quad (21)$$

For 1D networks with inhomogeneous rates [see Fig. 2(a)] the analysis yields a power law dependence of first-passage distributions at early times (see the Appendix for details),

$$f_{N,m}(t) \simeq \frac{\prod_{j=m}^{N-1} u_j}{(N-m-1)!} t^{N-m-1}. \quad (22)$$

This result fully agrees with the hypothesis, allowing to determine the number of intermediate states. Note also, that the prefactor in Eq. (22) contains the information on forward transition rates, and the exponent does not depend on transitions rates. The relative values of transitions rates only specify the range of the applicability of Eq. (22). To test the predictions from Eq. (22) computer simulations have been performed. The first-passage distributions are presented in Fig. 4(a) for all times and in Fig. 4(b) at early times for different initial states m and the final state $N = 10$. The transition rates w_m and u_m are randomly chosen. The solid lines in Fig. 4(b) come from linear fittings of the simulations at small t. Again, a decrease of the slope of the distribution function with smaller number of intermediate states is observed, as predicted by Eq. (22). Rounding the fitted values of the slope to the upper integer gives the number of intermediate states. The analysis can also be extended to another limit when $t \to \infty$, with the first-passage distribution producing (see the Appendix)

$$f_{N,m}(t) \simeq \frac{\Pi_{N,m}}{\tau_{N,m}} \exp(-t/\tau_{N,m}). \quad (23)$$

It shows how dynamic properties of networks can be obtained from large-time behavior of dwell-time distributions. It is important to note that these properties are averaged over many pathways connecting initial and final states so that the important information on the networks can also be obtained in this limit.

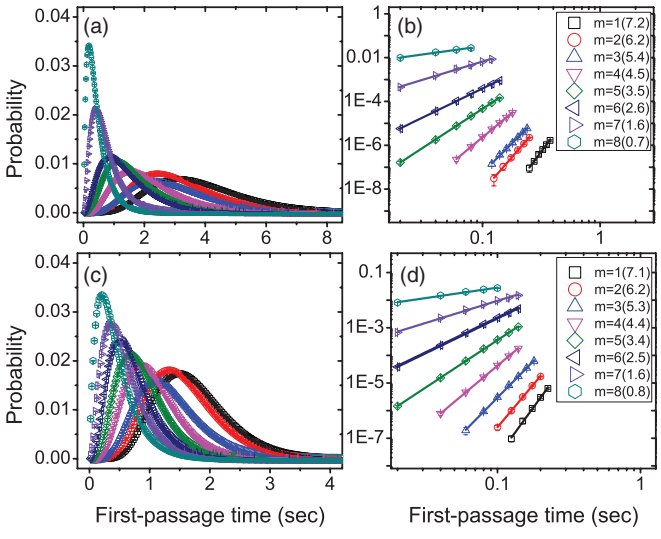

FIG. 4. First passage distributions for inhomogeneous 1D networks. The initial state m is varied from 1 to 8, and the final state is $N = 10$. Set of random rates is used for simulations. (a) and (b) Distributions for 1D inhomogeneous network shown in Fig. 2(a). (c) and (d) Distributions for 1D network with irreversible detachments as shown in Fig. 2(b). (a) and (c) First-passage distributions for all times. (b) and (d) First-passage distributions at early times. Symbols correspond to computer simulations. The solid lines in (b) and (d) are from linear fittings of the simulation data. The slope for each curve in (b) and (d) is also indicated.

C. Effect of irreversible transitions

To show the applicability of our method for more realistic systems consider a 1D network with irreversible detachments as illustrated in Fig. 2(b). This situation corresponds to fluorescence measurements of protein folding[6] when photon blinking phenomena can be viewed as irreversible dissociations from the network. Defining δ_m as an irreversible detachment rate from the state m, we can write backward master equations for evolution of first-passage probabilities $f_{N,m}(t)$ in the following form:

$$\frac{\partial f_{N,m}(t)}{\partial t} = u_m f_{N,m+1}(t) + w_m f_{N,m-1}(t) - (u_m + w_m + \delta_m) f_{N,m}(t), \quad (24)$$

with corresponding boundary conditions given by Eq. (2), as explained above.

We start first with a simple case of 1D networks with constant irreversible detachments, i.e., all detachment rates are equal $\delta_m = \delta$. The main difference of the networks with irreversible detachments in comparison with simple 1D networks without detachments is the fact the probabilities $f_{N,m}(t)$ are not conserved. This suggests the way to obtain analytic solutions by mapping the model with detachments into the model without detachments. It can be done by defining a first-passage probability function for surviving particle that is a conserved quantity,

$$F_{N,m}(t) = f_{N,m}(t) e^{\delta t}. \quad (25)$$

Substituting this into the backward master equations (24) we obtain the equation for $F_{N,m}(t)$,

$$\frac{\partial F_{N,m}(t)}{\partial t} = u_m F_{N,m+1}(t) + w_m F_{N,m-1}(t) - (u_m + w_m) F_{N,m}(t), \quad (26)$$

which is identical as Eq. (1) describing an inhomogeneous system without detachments. The Laplace transforms of the original and surviving first-passage probabilities are related as

$$\tilde{f}_{N,m}(s) = \tilde{F}_{N,m}(s + \delta). \quad (27)$$

Using this approach the model can be solved analytically and all dynamic properties can be calculated. The asymptotic behavior of first-passage distribution function for $t \to 0$ is given by

$$f_{N,m}(t) = F_{N,m}(t)e^{-\delta t} \simeq \frac{\prod_{j=m}^{N-1} u_j}{(N-m-1)!} t^{N-m-1} e^{-\delta t}$$

$$\simeq \frac{\prod_{j=m}^{N-1} u_j}{(N-m-1)!} t^{N-m-1}, \quad (28)$$

which is same as Eq. (22). Therefore, the early temporal behavior of the first-passage distributions is not modified by the irreversible dissociation processes with a constant detachment rate.

For a more general case with position dependent dissociation rates it has been shown earlier by one of us that the model with the irreversible detachments can be mapped into the model without detachments by utilizing a matrix renormalization approach.[29] Calculations indicate that the power law dependence of first-passage distribution function similar to Eq. (22) is again found at early times (see the Appendix for details). This power law dependence is also supported by computer simulations as indicated in Figs. 4(c) and 4(d). These results for 1D networks with detachments clearly support our hypothesis that the number of intermediate states in complex networks can be determined by analyzing dwell-time distributions at $t \to 0$.

D. Analysis of first-passage events for networks with different topology

In order to demonstrate the application of the method to more complex systems, two network models with different topology, as presented in Figs. 2(c) and 2(d), are also analyzed. One of the systems (Fig. 2(c)) has a set of states that branches out of the main pathway, while another system (Fig. 2(d)) has two parallel segments of states. For these two cases analytical results cannot be obtained easily, so we mainly focus on computer simulations. The first-passage distribution function for the 1D network with the branched states are shown in Figs. 5(a) and 5(b) for different initial states. The branch with two states is connected to the state $m = 5$ on the main pathway. One could see that for $1 \le m \le 5$ there are several distinct pathways. However, distribution functions again have a power law dependence at early times (see Fig. 5(b)), and the slopes are given by the number of intermediate states in the shortest pathway to the final destination. The presence of the branched states does not affect the network dynamics at early times, while the effect of the branch can be seen on long-time behavior of distribution functions (compare Figs. 4(a) and 5(a)).

For the network system with two parallel segments (Fig. 2(d)) we assumed that the loop starts at the site $m = 3$, and then two different cases of having 2 or 3 states in the upper segment have been analyzed. First-passage distribution functions for initial states $m = 2$, 3, and 4 and the final state $N = 10$ have been computed and presented in Figs. 5(c) and 5(d). As for all considered networks, the power law dependence of the distribution functions at early times

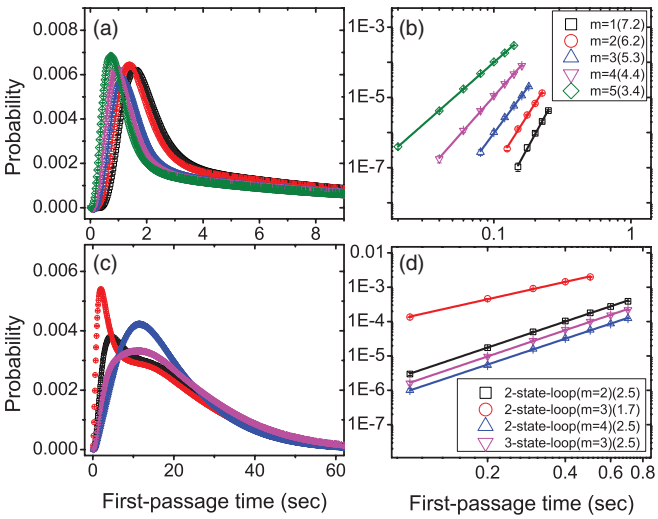

FIG. 5. First passage distributions for network models with complex topology. (a) and (b) First-passage distributions for a network with one branch as shown in Fig. 2(c). (c) and (d) First-passage distributions for a network with two parallel segments as shown in Fig. 2(d). (a) and (c) First-passage distributions in full time regime. (b) and (d) First-passage distributions at early times. The distribution functions for different initial states with $m = 1$–5 and the final state $N = 10$ are shown. For the network with two parallel segments the distribution functions for initial states with $m = 2$–4 and the final state $N = 10$ are shown. Symbols correspond to computer simulations. Set of random rates is used for simulations. The solid lines in (b) and (d) are from linear fittings of the simulation data. The slope for each curve in (b) and (d) is also indicated.

is observed (see Fig. 5(d)), and the slopes yield the number of intermediate states on the shortest pathway between initial and final states. The loop topology does not modify network dynamics at short times, while the effect of the network structure can only be seen at large times since in this case the dwell-time distributions are obtained by averaging over of all possible pathways between initial and final states. These results for both networks with complex geometry again support the hypothesis.

E. Application of the method for studying motor protein dynamics

It is important to test our method by analyzing real processes, and two motor protein systems[26,27] have been chosen as examples. Myong *et al.*[26] have investigated the DNA unwinding by virus NS3 helicase using single-molecule fluorescence analysis. Discrete steps of 3 base pairs (bp) have been observed and the corresponding dwell time histograms were also obtained with a rising phase at the beginning followed by a phase with decay. These observations indicate that a 3-bp transition is not the elementary step during the DNA unwinding. The distributions at the rising phase are shown in Fig. 6 (as demonstrated by circles and triangles). They can be fitted by linear curves (on log-log scale) as indicated in Fig. 6, supporting the power law dependence at early times. Our method predicts that the slopes of these curves give the number of intermediate states in the shortest pathway. The obtained values, 1.89 and 1.49, indicate the existence of two more hidden biochemical states during the unwinding of DNA. Then one concludes that the 3-bp transition is very likely composed of

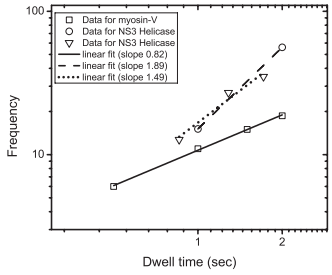

FIG. 6. Dwell time distributions for myosin-V's 74-nm steps (squares) from Ref. 27 and for steps in DNA unwinding by helicase NS3 (circles and triangles for different lengths of DNA) from Ref. 26 at early times. The lines are linear fits of experimental observations on log-log scale. The slopes for each curve are also indicated in the figure.

three elementary steps, and it is tempting to assign each step with a 1 bp length. However, one has to be careful here since this analysis only provides the number of intermediate states and not their spatial positions. Myong *et al.*[26] has proposed a similar mechanism of DNA unwinding by helicase with 1 bp per step but their analysis is based on a less realistic model that assumes irreversible equal-rate transitions. In contrast, our method is model independent.

A second example is related to studies of myosin-V molecules. Myosin-V are dimeric motor proteins that walk along actin filaments with a step size of 37 nm.[28] Two alternative stepping models have been proposed.[27] In the "inchworm" mechanism the advancement of one motor head is followed by the motion of the second one. In the "hand-over-hand" mechanism each motor head makes a 74 nm step in the alternating fashion. The dynamics of myosin-V has been studied by attaching a single fluorescent label to different parts of the protein molecule.[27] For a dye close to one motor head, the two different stepping mechanisms either give uniform 37 nm steps for the "inchworm" model or alternating 74 nm and 0 nm steps for the "hand-over-hand" model. Experiments have found only 74 nm steps and a dwell time distribution for these steps was also obtained, producing a rising phase at the beginning followed by a phase with decay. The distributions measured in Ref. 27 at the rising phase is shown as open squares in Fig. 6. Again, the distributions were fitted by a linear curve (on log-log scale) with a slope of ≈ 0.82, as shown by the solid line in Fig. 6. The slope value is close to 1, indicating the existence of one more intermediate state corresponding to the motion of the other motor protein head that is not fluorescently labeled under the "hand-over-hand" mechanisms.

III. DISCUSSION

A new theoretical approach to predict the structural and dynamic properties of complex chemical and biological networks via first-passage analysis of data on distributions of events is developed. It is based on the hypothesis that the first-passage distributions of events are fully specified by local topology and dynamics of underlying networks. Exact analytical calculations and Monte Carlo simulations are performed for several networks. A power law dependence of first-passage distribution at early times is found in all cases, and the corresponding exponent reflects the number of intermediate states in the shortest pathway between the initial and final states. At large times our approach predicts that distributions decay exponentially as a function of time with parameters that depend on the probability and the mean dwell time to reach the final state. The microscopic origin of the method can be explained using the following arguments. In the network at early times only the fastest events are recorded and they proceed via the shortest link between initial and final states. Thus the first-passage distribution at early times corresponds to trajectories along the shortest path in terms of the number of intermediate states. In contrast, the dynamics of reaching the final state at large times is averaged out over many pathways connecting two states. Using this approach, our analysis of short-time arrival dynamics from distribution functions indicates the existence of two intermediate biochemical states for each DNA unwinding step by NS3 helicase and one intermediate state for stepping of myosin-V molecules, supporting the "hand-over-hand" stepping mechanisms.

In principle, our approach is capable to determine full structural and dynamic information of networks if experimental information on distributions of events is available. To obtain the full structure many events with varying initial and final states must be analyzed. However, in the application of the method there are still questions. In this work mostly the information on short-time dynamics of distribution functions has been utilized. We argued that the behavior of distribution functions at large times provides the information on mean dwell times and probabilities to reach the final state. However it is not clear how to obtain explicit information from these average properties. In addition, more theoretical studies are needed to clarify how distribution of events depends on the underlying topology of the network. Despite these issues, it seems that the presented method might be a powerful tool for uncovering mechanisms of complex processes.

Recently, a new method to investigate similar problems, which is based on information theory approach, was proposed.[30] However, this method was only employed for very simple networks described by multi-exponential dwell time distributions and only in the limit of large times. Furthermore, only approximate values for the number of hidden states could be obtained. In contrast, our method is much simpler and it provides an exact number of hidden states from early times.

ACKNOWLEDGMENTS

This work was supported by grants from National Institutes of Health (Grant No. 1R01GM094489-01) and the Welch Foundation (Grant No. C-1559).

APPENDIX: DETAILED THEORETICAL CALCULATIONS

This appendix consists of several parts with detailed derivations of the equations and relations used in the main text.

1. Inhomogeneous 1D networks

The backward master equations for a general inhomogeneous 1D network model presented in Fig. 2(a) in the Laplace presentation are given by Eqs. (16)–(18) in the main text. In order to solve these equations we suggest the following ansatz for Laplace transform of distribution function as explained in the main text:

$$\tilde{f}_{N,m}(s) = \frac{\sum_{K=0}^{m-1} \alpha_K(m) s^K}{\sum_{K=0}^{N-1} \alpha_K(N) s^K}, \tag{A1}$$

where $\alpha_K(m)$ are some unknown coefficients. Equation (A1) satisfies boundary conditions for $m = N$. It suggests that the Laplace transforms of distribution functions can be viewed as ratios of two polynomials with respect to the variable s. For example, for the simplest case $N = 2$ and $m = 1$ it can be shown that

$$\tilde{f}_{2,1}(s) = \frac{u_1}{s + u_1 + w_1}, \tag{A2}$$

which agrees with the ansatz [see Eq. (A1)]. Inverting this expression yields the corresponding first-passage distribution function,

$$f_{2,1}(t) = u_1 \exp[-(u_1 + w_1)t]. \tag{A3}$$

Also for the case $N = 3$ with $m = 1$ or $m = 2$ one can find

$$\tilde{f}_{3,1}(s) = \frac{u_1 u_2}{s^2 + s(u_1 + u_2 + w_1 + w_2) + u_1 u_2 + w_1 u_2 + w_1 w_2}, \tag{A4}$$

and

$$\tilde{f}_{3,2}(s) = \frac{u_2 s + u_1 u_2 + w_1 u_2}{s^2 + s(u_1 + u_2 + w_1 + w_2) + u_1 u_2 + w_1 u_2 + w_1 w_2}, \tag{A5}$$

which also agree with the ansatz [see Eq. (A1)]. These expressions can also be inverted analytically.

The coefficients $\alpha_k(m)$ are determined self-consistently from backward master equations, which provides a direct way to obtain explicitly all dynamic properties of the system. For a general case with $m = 1$, the expression for Eq. (A1) is given by

$$\tilde{f}_{N,1}(s) = \frac{\alpha_0(1)}{\sum_{K=0}^{N-1} \alpha_K(N) s^K}. \tag{A6}$$

Then, combining this relation with Eq. (17) in the main text we obtain

$$\alpha_0(1)(s + u_1 + w_1) = u_1(\alpha_0(2) + \alpha_1(2)s). \tag{A7}$$

It can be proved for general conditions that $\alpha_0(1) = 1$, which simplifies many calculations. Specifically, from Eq. (A7) we derive

$$\alpha_1(2) = 1/u_1, \quad \alpha_0(2) = \frac{u_1 + w_1}{u_1}. \tag{A8}$$

Similarly, for a general m, one can obtain the following expressions:

$$\alpha_0(m) = \left(1 + \sum_{K=1}^{m-1} \prod_{j=1}^{K} \frac{w_j}{u_j}\right), \tag{A9}$$

$$\alpha_{m-1}(m) = \prod_{j=1}^{m-1} \frac{1}{u_j}. \tag{A10}$$

Using Eqs. (16) and (A1) and considering only terms proportional to s leads to

$$\alpha_0(m) + \alpha_1(m)(u_m + w_m) = u_m \alpha_1(m+1) + w_m \alpha_1(m-1). \tag{A11}$$

Let us define a new auxiliary function,

$$\Delta_m^{(1)} = \alpha_1(m+1) - \alpha_1(m), \tag{A12}$$

then it can be shown that

$$\alpha_0(m) = u_m \Delta_m^{(1)} - w_m \Delta_{m-1}^{(1)}, \tag{A13}$$

where $\alpha_0(m)$ are known from Eq. (A9). The general solution of Eq. (A13) is given by

$$\Delta_m^{(1)} = \sum_{K=1}^{m} \left(\frac{\alpha_0(K)}{u_K} \prod_{j=K+1}^{m} \frac{w_j}{u_j}\right), \tag{A14}$$

which leads to

$$\alpha_1(m) = \sum_{l=1}^{m-1} \Delta_l^{(1)} = \sum_{l=1}^{m-1} \left[\sum_{K=1}^{l} \left(\frac{\alpha_0(K)}{u_K} \prod_{j=K+1}^{l} \frac{w_j}{u_j}\right)\right]. \tag{A15}$$

Similarly, from Eqs. (16) and (A1), considering the coefficients for the terms proportional to s^K, one finds

$$\alpha_{K-1}(m) + \alpha_K(m)(u_m + w_m) = u_m \alpha_K(m+1) + w_m \alpha_K(m-1). \tag{A16}$$

Using the same procedure as above, the expression for the coefficient $\alpha_K(m)$ can be written as

$$\alpha_K(m) = \sum_{i=K}^{m-1} \left[\sum_{l=K}^{i} \left(\frac{\alpha_{K-1}(l)}{u_l} \prod_{j=l+1}^{i} \frac{w_j}{u_j}\right)\right]. \tag{A17}$$

This result is important since it indicates that the coefficient $\alpha_K(m)$ can be determined from $\alpha_{K-1}(m)$. Therefore, all coefficients $\alpha_K(m)$ can be calculated iteratively because we already determined the expressions for $\alpha_0(m)$. Then the Laplace transform $\tilde{f}_{N,m}(s)$ of distribution function can be derived from Eq. (A1) and the distribution function can be obtained from the inverse Laplace transforms. In fact, to calculate many dynamic properties, we do not need to know all coefficients $\alpha_K(m)$. For example, the splitting probability $\Pi_{N,m}$ is simply given by

$$\Pi_{N,m} = \tilde{f}_{N,m}(s=0) = \frac{\alpha_0(m)}{\alpha_0(N)} = \frac{1 + \sum_{K=1}^{m-1} \prod_{j=1}^{K} \frac{w_j}{u_j}}{1 + \sum_{K=1}^{N-1} \prod_{j=1}^{K} \frac{w_j}{u_j}}, \tag{A18}$$

which agrees with known expressions in the literature.[16] For 1D networks with homogeneous transition rates $w_m = w$ and

$u_m = m$, the splitting probability can be written in the simpler form,

$$\Pi_{N,m} = \tilde{f}_{N,m}(s=0) = \frac{1-\gamma^m}{1-\gamma^N}, \tag{A19}$$

where $\gamma = \frac{u}{w}$.

From Eq. (12) the general expression for the mean first-passage time to exit at the site N is given by

$$\tau_m(1) = \int_0^\infty t f_{N,m}(t)dt = -\frac{d\tilde{f}_{N,m}(s)}{ds}|_{s=0}$$
$$= \frac{\alpha_0(m)\alpha_1(N) - \alpha_1(m)\alpha_0(N)}{\alpha_0(N)^2}. \tag{A20}$$

And for the conditional mean first-passage time we obtain

$$\tau_{N,m} = \frac{\tau_m(1)}{\Pi_{N,m}} = \frac{\alpha_1(N)}{\alpha_0(N)} - \frac{\alpha_1(m)}{\alpha_0(m)}. \tag{A21}$$

For the special case with $w_j = 0$ from Eq. (A9) one can show that all coefficients $\alpha_0(m)$ are given by

$$\alpha_0(m) = \alpha_0(1) = 1. \tag{A22}$$

From Eq. (A18) the splitting probability is given by $\Pi_{N,m} = 1$. The coefficient $\alpha_1(m)$ can be then written as

$$\alpha_1(m) = \sum_{i=1}^{m-1} \frac{1}{u_i}, \tag{A23}$$

which gives the conditional mean first-passage time $\tau_{N,m}$,

$$\tau_{N,m} = \alpha_1(N) - \alpha_1(m) = \sum_{i=m}^{N-1} \frac{1}{u_i}. \tag{A24}$$

Let us consider another special case with $w_1 = 0$ that corresponds to a reflecting boundary condition. It can be shown that in this case the coefficient $\alpha_0(m)$ is equal to

$$\alpha_0(m) = 1, \tag{A25}$$

for all m. And the coefficient $\alpha_1(m)$ is given by

$$\alpha_1(m) = \sum_{K=1}^{m-1} \Delta_K^{(1)}, \tag{A26}$$

with

$$\Delta_K^{(1)} = \sum_{i=1}^{K} \left(\frac{1}{u_i} \prod_{j=i+1}^{K} \frac{w_j}{u_j} \right). \tag{A27}$$

Therefore, the conditional mean first-passage time $\tau_{N,m}$ can be written as

$$\tau_{N,m} = \sum_{K=m}^{N-1} \left(\sum_{i=1}^{K} \frac{1}{u_i} \prod_{j=i+1}^{K} \frac{w_j}{u_j} \right). \tag{A28}$$

From simple 1D networks with inhomogeneous transition rates, it is useful to consider the asymptotics for $\tilde{f}_{N,m}(s)$. For $s \to \infty$, which corresponds to $t \ll 1$, we have

$$\tilde{f}_{N,m}(s) \simeq \frac{\alpha_{m-1}(m)}{\alpha_{N-1}(N)} s^{m-N} = \frac{\prod_{j=1}^{m-1} \frac{1}{u_j}}{\prod_{j=1}^{N-1} \frac{1}{u_j}} \frac{1}{s^{N-m}}$$
$$= \prod_{j=m}^{N-1} u_j \frac{1}{s^{N-m}}, \tag{A29}$$

where Eq. (A10) was used to obtain the final expression above. Then, the first-passage distribution function at small $t \to 0$ is given by

$$f_{N,m}(t) \simeq \frac{\prod_{j=m}^{N-1} u_j}{(N-m-1)!} t^{N-m-1}, \tag{A30}$$

which is Eq. (22) in the main text. Again, we obtained a power law dependence of the first-passage distribution at early times. For another limit with $s \to 0$ or $t \to \infty$, we find

$$\tilde{f}_{N,m}(s) \simeq \frac{\alpha_0(m) + \alpha_1(m)s}{\alpha_0(N) + \alpha_1(N)s} \simeq \frac{\alpha_0(m)}{\alpha_0(N)} \frac{1}{1 + \left[\frac{\alpha_1(N)}{\alpha_0(N)} - \frac{\alpha_1(m)}{\alpha_0(m)} \right] s}, \tag{A31}$$

which leads to the first-passage distribution function for $t \to \infty$ as

$$f_{N,m}(t) \simeq \frac{\Pi_{N,m}}{\tau_{N,m}} e^{-\frac{t}{\tau_{N,m}}}, \tag{A32}$$

where $\Pi_{N,m}$ and $\tau_{N,m}$ are given by Eqs. (A18) and (A21), respectively. The equation above corresponds to Eq. (23) in the main text.

2. Effect of irreversible transitions

For 1D networks with irreversible detachments from the main pathway, as illustrated in Fig. 2(b), the general backward master equations for evolution of first-passage probabilities of reaching the state N for the first time at time t starting from the state m at $t = 0$ are given by Eq. (24) in the main text. For a general inhomogeneous case of 1D networks with dissociations, it has been shown earlier that the model with irreversible detachments can be mapped into the model without detachments by utilizing a matrix renormalization approach.[29] We can define a first-passage probability function $F_{N,m}(t)$ as

$$F_{N,m}(t) = f_{N,m}(t) e^{\lambda t + \gamma_m}, \tag{A33}$$

where λ and γ_m will be fixed so that $F_{N,m}(t)$ satisfies the "renormalized," *backward master equations*

$$\frac{\partial F_{N,m}(t)}{\partial t} = \tilde{u}_m F_{N,m+1}(t) + \tilde{w}_m F_{N,m-1}(t)$$
$$- (\tilde{u}_m + \tilde{w}_m) F_{N,m}(t), \tag{A34}$$

with properly "renormalized" rates $\tilde{u}_m$ and $\tilde{w}_m$. By substituting Eq. (A33) in Eq. (24) one can obtain

$$\frac{\partial F_{N,m}(t)}{\partial t} = u_m e^{\gamma_m - \gamma_{m+1}} F_{N,m+1}(t)$$
$$+ w_m e^{\gamma_m - \gamma_{m-1}} F_{N,m-1}(t)$$
$$- (u_m + w_m + \delta_m - \lambda) F_{N,m}(t). \tag{A35}$$

Comparing the corresponding terms in Eqs. (A34) and (A35), we derive the following expressions:

$$\tilde{u}_m = u_m \psi_{m+1}/\psi_m,$$
$$\tilde{w}_m = w_m \psi_{m-1}/\psi_m, \tag{A36}$$
$$-w_m \psi_{m-1} + (u_m + w_m + \delta_m)\psi_m - u_m \psi_{m+1} = \lambda \psi_m,$$

where $\psi_m = \exp(-\gamma_m)$. The expression for λ and ψ_m can be determined by solving the eigenvalue equation $\mathbf{M}\psi = \lambda \psi$, where $\mathbf{M}$ is a $N \times N$ matrix and ψ is the column vector $[\psi_m]$. Similarly, the asymptotic behavior of first-passage distribution function for $t \to 0$ is given by

$$f_{N,m}(t) = F_{N,m}(t)e^{-\lambda t}\psi_m \simeq \frac{\prod_{j=m}^{N-1} u_j}{(N-m-1)!}\psi_N t^{N-m-1}, \tag{A37}$$

which also shows a power law dependence at small t.

[1] O. N. Temkin, A. V. Zeigarnik, and D. G. Bonchev, *Chemical Reaction Networks: A Graph-Theoretical Approach* (CRC Press, New York, 1996).
[2] S. H. Strogatz, Nature **410**, 268 (2001).
[3] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
[4] S. Karalus and M. Porto, Europhys. Lett. **99**, 38002 (2012).
[5] J. Goutsias and G. Jenkinson, Phys. Rep. **529**, 199 (2013).
[6] H. S. Chung, K. McHale, J. M. Louis, and W. A. Eaton, Science **335**, 981 (2012).
[7] Q. Jin, A. M. Fleming, C. J. Burrows, and H. S. White, J. Am. Chem. Soc. **134**, 11006 (2012).
[8] Y. Sowa, A. D. Rowem, M. C. Leake, T. Yakushi, M. Homma, A. Ishijima, and R. M. Berry, Nature **437**, 916 (2005).
[9] A. R. Dunn, P. Chuan, Z. Bryant, and J. A. Spudich, Proc. Natl. Acad. Sci. U.S.A. **107**, 7746 (2010).
[10] A. Dimitrov, M. Quesnolt, S. Moutel, I. Cantaloube, C. Pous, and F. Perez, Science **322**, 1353 (2008).
[11] J. W. J. Kerssemakers, L. Munteanu, L. Laan, T. L. Noetzel, M. J. Janson, and M. Dogterom, Nature **442**, 709 (2006).
[12] A. B. Kolomeisky and M. E. Fisher, Annu. Rev. Phys. Chem. **58**, 675 (2007).
[13] J. E. Straub and D. Thirumalai, Annu. Rev. Phys. Chem. **62**, 437 (2011).
[14] J. J. Tyson and B. Novak, Annu. Rev. Phys. Chem. **61**, 219 (2010).
[15] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[16] N. G. van Kampen, *Stochastic Processes in Chemistry and Physics* (North Holland, Amsterdam, 1992).
[17] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).
[18] A. Szabo, K. Schulten, and Z. Schulten, J. Chem. Phys. **72**, 4350 (1980).
[19] A. B. Kolomeisky, E. B. Stukalin, and A. A. Popov, Phys. Rev. E **71**, 031902 (2005).
[20] D. Colquhoun and A. G. Hawkes, Proc. R. Soc., London, Ser. B **211**, 205 (1981).
[21] B. Sakmann and E. Neher, *Single-Channel Recordings* (Plenum, New York, 1995).
[22] L. S. Milescu, A. Yildiz, P. R. Selvin, and F. Sachs, Biophys. J. **91**, 1156 (2006).
[23] J.-C. Liao, J. A. Spudich, D. Parker, and S. L. Delp, Proc. Natl. Acad. Sci. U.S.A. **104**, 3171 (2007).
[24] Y. R. Chemla, J. R. Moffitt, and C. Bustamante, J. Phys. Chem. B **112**, 6025 (2008).
[25] J. Cao and R. J. Silbey, J. Phys. Chem. B **112**, 12867 (2008).
[26] S. Myong, M. M. Bruno, A. M. Pyle, and T. Ha, Science **317**, 513–516 (2007).
[27] A. Yildiz, J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin, Science **300**, 2061 (2003).
[28] A. Mehta, R. Rock, M. Rief, J. Spudich, M. Mooseker, and R. E. Cheney, Nature **400**, 590 (1999).
[29] A. B. Kolomeisky and M. E. Fisher, Physica A **279**, 1 (2000).
[30] C.-B. Li and T. Komatsuzaki, Phys. Rev. Lett. **111**, 058301 (2013).