Bulletin of Mathematical Biology (2024) 86:55
https://doi.org/10.1007/s11538-024-01283-3

ORIGINAL ARTICLE

Understanding and Quantifying Network Robustness to Stochastic Inputs

Hwai-Ray Tung[1] **· Sean D. Lawley**[1]

Received: 25 December 2023 / Accepted: 18 March 2024

Abstract

A variety of biomedical systems are modeled by networks of deterministic differential equations with stochastic inputs. In some cases, the network output is remarkably constant despite a randomly fluctuating input. In the context of biochemistry and cell biology, chemical reaction networks and multistage processes with this property are called robust. Similarly, the notion of a forgiving drug in pharmacology is a medication that maintains therapeutic effect despite lapses in patient adherence to the prescribed regimen. What makes a network robust to stochastic noise? This question is challenging due to the many network parameters (size, topology, rate constants) and many types of noisy inputs. In this paper, we propose a summary statistic to describe the robustness of a network of linear differential equations (i.e. a first-order mass-action system). This statistic is the variance of a certain random walk passage time on the network. This statistic can be quickly computed on a modern computer, even for complex networks with thousands of nodes. Furthermore, we use this statistic to prove theorems about how certain network motifs increase robustness. Importantly, our analysis provides intuition for why a network is or is not robust to noise. We illustrate our results on thousands of randomly generated networks with a variety of stochastic inputs.

Keywords Homeostasis · Robustness · Pharmacokinetics · Pharmacodynamics · Medication nonadherence · Medication adherence

1 Introduction

Homeostasis is a pillar of biology (Michael 2007). Coined in the 1930s and stemming from Greek words meaning "standing still" (Cannon 1932), homeostasis refers to the

✉ Sean D. Lawley
lawley@math.utah.edu

Hwai-Ray Tung
ray.tung@utah.edu

1 Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

Published online: 12 April 2024

processes that living systems use to maintain stable conditions in the face of fluctuating environments. Homeostasis or biological "robustness" is seen across many scales in biology, from ecosystems, to organisms, to organs, to cells, to subcellular systems (Félix and Barkoulas 2015).

A quantitative understanding of such robustness is often challenging due to the very large number of interacting constituents. For example, biochemical reaction networks in cells often involve many different chemical species, where each species may interact with many other species and be involved in a variety of cellular processes. A key question is to understand the functional consequences and emergent properties of such complicated interactions.

A mathematical framework for studying robustness is to consider how stochastic perturbations propagate through a deterministic network of differential equations (Anderson et al. 2007; Anderson and Mattingly 2007; Browning et al. 2023). Consider the following system of linear ordinary differential equations (ODEs) with stochastic forcing,

$$\frac{\mathrm{d}}{\mathrm{d}t}c = Rc + I(t)\mathbf{e}_1. \tag{1}$$

In (1), $c(t) = \{c_i(t)\}_{i=1}^m$ is a vector of concentrations in different compartments (or "states" or "nodes"), $R \in \mathbb{R}^{m\times m}$ is a reaction rate matrix describing interactions between compartments, and $I(t) \in \mathbb{R}$ is some stochastic input into compartment $i = 1$ ($\mathbf{e}_1 \in \mathbb{R}^{m\times 1}$ denotes the first standard basis vector). How does the stochastic input $I(t)$ flow through the network? When does the network dampen the stochastic input so that the "output" of the system (either the concentration in a given node of interest or the flux out of the system) is nearly constant? How does this depend on network properties such as the size, topology, and rate constants? How does this depend on properties of the stochastic input? The purpose of this paper is to address these questions. See Fig. 1 for an illustration.

Motivated by biochemical reaction networks, these mathematical questions were posed by Anderson, Mattingly, Nijhout, and Reed in 2007. In that very interesting work, the authors proved that if the network consists of a single chain of irreversible reactions (see Fig. 1b), then the variance of the flux out of each compartment strictly decreases down the chain. These authors also studied the effects of side chains and feedback loops. These authors proved their results under very general assumptions on the stochastic input, and they also considered the case that the input is white noise and used the resulting Gaussianity of the system to make explicit calculations.

More recently, Browning et al. (2023) studied this problem. These authors were motivated by a variety of multistage processes in biology, including viral replication (Louten 2016), bacteriophage replication (Campbell 2003), progression through the cell cycle (Morgan 2007), cascade reactions at the molecular level (Huang and Ferrell 1996), transport through discrete layers (Carr and Simpson 2019), and social processes such as queuing (Liu et al. 2004). Similar to Anderson et al. (2007), these authors focused on networks consisting of a single chain of irreversible reactions and also considered the effects of feedback and feedforward loops. These authors assumed

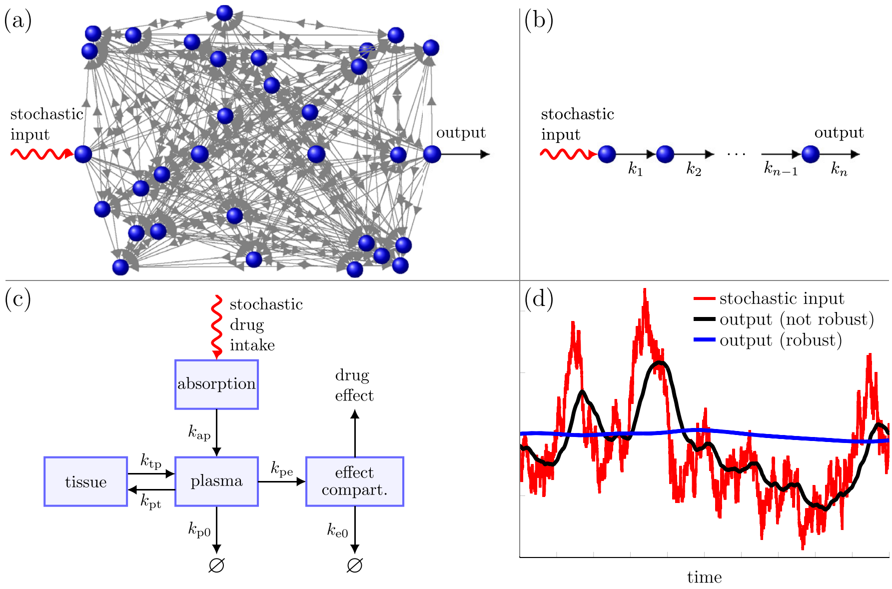

Fig. 1 **a** Complex network. **b** Irreversible linear chain network. **c** Pharmacokinetic/pharmacodynamic example network with a stochastic drug intake modeling medication nonadherence. **d** Noisy input (red curve) and noisy output (black curve) for a network that is not robust compared to a nearly constant output (blue curve) for a robust network (Color figure online)

that the stochastic input is a Ornstein–Uhlenbeck process and used the resulting Gaussianity of the system to make explicit calculations.

An alternative motivation for these mathematical questions comes from the problem of medication nonadherence. Medication adherence is the extent to which patients take medications as prescribed by their physicians. Medication *non*adherence is a major problem, resulting in over 100,000 preventable deaths and over \$100 billion in preventable health care costs per year in the United States alone (Osterberg and Blaschke 2005). In fact, the World Health Organization has claimed that improving adherence may have a far greater impact on public health than any improvement in specific medical treatments (Sabaté and Sabaté 2003; Haynes et al. 2002). To combat nonadherence, it is often recommended to prescribe so-called “forgiving” drugs, which maintain their effect despite lapses in patient adherence (missed doses, late doses, etc.) (Osterberg et al. 2010). Mathematically, the pharmacokinetics and pharmacodynamics of a drug are often described by a network of ODEs as in (1) (Gibaldi and Perrier 1982; Rosenbaum 2016). The different components of $c(t) = \{c_i(t)\}_{i=1}^m$ model drug concentrations or effects in different compartments of the body, the matrix R models the transfer rates between compartments, and the stochastic input $I(t)$ models the patient’s imperfect adherence to the prescribed dosing regimen (which would be deterministic for a perfectly adherent patient) (Li and Nekka 2007, 2009; Lévy-Véhel and Lévy-Véhel 2013; Fermín and Lévy-Véhel 2017; Counterman and Lawley 2021, 2022; McAllister and Lawley 2022; Clark and Lawley 2022). See Fig. 1c for an illustration. In this framework, drug forgiveness is a network property (i.e. depending on the topology and kinetics of the ODEs) in which the output (drug effect) is robust to stochastic perturbations of the input (i.e. late doses, missed doses, etc). Hence, drug forgiveness in pharmacology is akin to network robustness in the biochemical and cell biology applications described above.

In this paper, we propose a simple summary statistic to characterize network robustness. This summary statistic is the variance of a certain random "network time" T whose probability distribution is defined by the network R (we denote the variance of T by $\mathrm{Var}(T)$). We show that more variable network times (i.e. larger values of $\mathrm{Var}(T)$) correspond to more robust networks. Robustness means that the variability in either the flux out of the system or the variability in a given compartment of interest is much less than the variability of the stochastic input (see Sect. 2.5 for a discussion of different notions of robustness). Importantly, this result holds for (i) very general networks (including large and complex networks such as illustrated in Fig. 1a) and (ii) very general stochastic inputs. We prove that this network time T is a certain passage time of a random walk on the network. Specifically, if one is interested in the variability in the flux out of the system, then T is the first passage time to exit the network. If one is interested in the variability in a given compartment n, then T is the last passage time to compartment n conditioned on arrival to compartment n.

We now highlight four implications of our analysis. First, $\mathrm{Var}(T)$ can be quickly computed on a modern computer, even for complex networks with thousands of nodes. Hence, $\mathrm{Var}(T)$ is an easily computable measure of network robustness that is independent of the noisy input and can be used to compare the robustness of different networks. Indeed, we demonstrate via numerical simulations on thousands of randomly generated networks that larger values of $\mathrm{Var}(T)$ indeed correspond to more robust networks.

Second, our analysis predicts that a network is robust to a given noisy input if the standard deviation of the network time T (denoted $\mathrm{SD}(T) := \sqrt{\mathrm{Var}(T)}$) is large compared to the correlation time $\tau \in (0, \infty)$ of the noisy input, i.e.

$$\mathrm{SD}(T) \gg \tau.$$

Third, having established that $\mathrm{Var}(T)$ measures network robustness, we determine how certain network structures and motifs affect robustness by proving how they affect $\mathrm{Var}(T)$. For example, we prove that under quite generic circumstances, appending an arbitrary network to an existing network must increase $\mathrm{Var}(T)$. This result therefore shows how increasing the size and complexity of a network can increase its robustness.

Fourth, and perhaps most importantly, our analysis provides intuition for why a network is or is not robust to noise. That is, it is *a priori* rather mysterious why some networks are robust or why adding or deleting a certain edge in a network may make it more or less robust. However, characterizing network robustness by variability in T gives insight into why certain network features increase or decrease robustness. In particular, if the time for a random walk to traverse the network is highly variable, then the network is robust. Hence, a network with many different paths tends to be robust. Further, appending a network to an existing network can increase robustness because a random walk on the network may or may not visit this appended network which increases variability in T.

The rest of the paper is organized as follows. In Sect. 2, we show how network robustness depends on variability in T. In Sect. 3, we verify via numerical simulations on randomly generated networks that $\mathrm{Var}(T)$ does indeed characterize network robustness. In Sect. 4, we find representations for T in terms of random walk passage times on the network. We then use these stochastic representations of T to prove results about how certain network structures can increase or decrease variability in T. We conclude with a brief discussion. We collect the proofs in an appendix.

2 Robustness by Variability in Network Time *T*

In this section, we show that network robustness depends on the variability of a random time T whose probability distribution is defined by the network. As we detail below, the definition of T depends on our specific notion of robustness.

2.1 Model Setup

The following definition characterizes the class of reaction rate matrices $R \in \mathbb{R}^{m\times m}$ that we consider in this paper.

Definition 1 A matrix $R = \{R_{ji}\}_{j,i=1}^{m} \in \mathbb{R}^{m\times m}$ for $m \geq 1$ is called admissible if the following three conditions hold.

(a) $R_{ji} \geq 0$ if $i \neq j$ and $d_i := -\sum_{j=1}^{m} R_{ji} \geq 0$ for all $i \in \{1, \dots, m\}$.
(b) For each $n \in \{2, 3, \dots, m\}$, there exists a sequence of $k \geq 2$ distinct states $i_1 = 1, i_2, \dots, i_k = n$ so that $R_{i_2,i_1} R_{i_3,i_2} \cdots R_{i_k,i_{k-1}} > 0$.
(c) For each $i \in \{1, \dots, m\}$ with $d_i = 0$, there exists a $j \in \{1, \dots, m\}$ with $d_j > 0$ and a sequence of $k \geq 2$ distinct states $i_1 = i, i_2, \dots, i_k = j$ so that $R_{i_2,i_1} R_{i_3,i_2} \cdots R_{i_k,i_{k-1}} > 0$.

We interpret R_{ji} as the transfer rate from node i to node $j \neq i$ and $d_i \geq 0$ as the decay rate from node i. Property (a) of Definition 1 thus ensures that the ODE in (1) conserves mass in the sense that mass can only move between compartments or decay. In particular, property (a) allows us to define a continuous-time Markov chain associated to R.

Definition 2 If a matrix $R = \{R_{ji}\}_{j,i=1}^{m} \in \mathbb{R}^{m\times m}$ satisfies property (a) of Definition 1, then we say that the continuous-time Markov chain (Norris 1998),

$$X = \{X(t)\}_{t\geq 0} \text{ with state space } \{\varnothing, 1, 2, \dots, m\}, \tag{2}$$

is a random walk on R if X jumps from state $i \in \{1, 2, \dots, m\}$ to state $j \in \{1, 2, \dots, m\}$ at rate R_{ji} for $i \neq j$, jumps from state $i \in \{1, 2, \dots, m\}$ to state $\varnothing$ at rate

$$d_i := -\sum_{j=1}^{m} R_{ji} \geq 0, \tag{3}$$

and never leaves state $\varnothing$.

Table 1 Notation introduced in Sect. 2.1

$c_i(t)$	Concentration in compartment i at time t
$c(t)$	Vector whose entries are c_i
m	The total number of compartments
R	The reaction rate matrix
$I(t)$	The noisy input into compartment 1 at time t
$\xi(t)$	The fluctuation in noisy input about the average ($\xi(t) := I(t) - \mathbb{E}[I(t)]$)
$\mathbf{e}_i$	The i^{th} standard basis vector
d_i	The decay rate from compartment i (see (3))
X	Continuous time Markov chain induced by R (see Definition 2)

Property (b) of Definition 1 avoids trivial cases by ensuring that input into node 1 may reach any node $n \in \{2, 3, \ldots, m\}$. Property (c) ensures that any input into the system eventually decays. The following proposition makes "eventual decay" precise.

Proposition 1 *If R is admissible, then its eigenvalues have strictly negative real parts.*

Having defined the admissible class of "reaction rate matrices" or "networks" $R \in \mathbb{R}^{m \times m}$, consider the following system of linear ODEs with stochastic input,

$$\frac{\mathrm{d}}{\mathrm{d}t} c(t) = Rc(t) + I(t)\mathbf{e}_1, \quad t \in \mathbb{R}. \tag{4}$$

In (4), $c(t) = \{c_i(t)\}_{i=1}^m \in \mathbb{R}^{m \times 1}$ is a vector of concentrations in different compartments, $I(t) \in \mathbb{R}$ is some noisy input to compartment $i = 1$, $\mathbf{e}_1 \in \mathbb{R}^{m \times 1}$ is the first standard basis vector, and $R \in \mathbb{R}^{m \times m}$ is admissible. Let $\xi(t)$ denote the fluctuations of the input about the average,

$$\xi(t) := I(t) - \mathbb{E}[I(t)],$$

so that $\mathbb{E}[\xi(t)] = 0$. Assume that $\mathbb{E}[I(t)] + \mathbb{E}[\xi^2(t)] \leq B$ for all $t \in \mathbb{R}$ for some $B \in (0, \infty)$ and that $\mathbb{E}[I(t)]$ and $\xi(t)$ are piecewise continuous and take only finitely many jumps on any finite time interval. The notation introduced in this section is collected in Table 1 for convenience.

2.2 Network Times T_n, T_{abs} and a Pedagogical Example

We introduce two random times corresponding to different notions of robustness. We refer to these random times as "network times" since they depend only on the network R (and thus not on the noisy input I). By using the variance of these network times as summary statistics, we aim to quantify the contribution of R to robustness under general stochastic inputs I.

One notion of robustness is the variability of the flux out of the system.

Definition 3 For a given reaction rate matrix R, let the flux out of the system J be

$$J(t) := \sum_{i=1}^{m} d_i c_i(t) = \mathbf{d}^\top c(t), \tag{5}$$

where $\mathbf{d}^\top = -\mathbf{1}^\top R \in \mathbb{R}^{m\times 1}$ is the vector of decay rates $d_1, \ldots, d_m$.

We show below that the variance of J can be understood in terms of the following network time T_{abs}.

Definition 4 For a given reaction rate matrix R, let T_{abs} be a random time with probability density function given by

$$p_{\text{abs}}(t) = \mathbf{d}^\top e^{Rt}\mathbf{e}_1, \quad t \geq 0, \tag{6}$$

where $\mathbf{d}^\top = -\mathbf{1}^\top R \in \mathbb{R}^{m\times 1}$ is the vector of decay rates $d_1, \ldots, d_m$.

Rather than the flux J, an alternative notion of robustness concerns the variability in the concentration c_n in some given node of interest $n \in \{1, 2, \ldots, m\}$. In this case, we show below that the variability of c_n can be understood in terms of the following network time T_n.

Definition 5 For a given reaction rate matrix R, let T_n be a random time with probability density function given by

$$p_n(t) := \frac{1}{t_n}\mathbb{P}(X(t) = n \,|X(0) = 1) = \frac{1}{t_n}\mathbf{e}_n^\top e^{Rt}\mathbf{e}_1, \quad t \geq 0, \tag{7}$$

where

$$t_n := \int_0^\infty \mathbb{P}(X(t) = n \mid X(0) = 1)\,\mathrm{d}t = -(R^{-1})_{n1} \in (0, \infty). \tag{8}$$

Note that the definitions of T_{abs} and T_n depends solely on the network R and not on the stochastic input I. In later sections, we use T to refer to both T_{abs} and T_n depending on context. Furthermore, we show in Sect. 4 that T_{abs} and T_n can be interpreted as first passage times and last passage times. In particular, the "abs" in T_{abs} stands for absorbing state since T_{abs} can be interpreted as the first passage time to the absorbing state.

To illustrate the connection between our network times and robustness, we give a simple example regarding J and T_{abs}, which can be easily modified for c_n and T_n. Consider the single ODE describing exponential decay at rate $d > 0$ with a stochastic input $I(t)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}c = -dc + I(t).$$

Table 2 Notation introduced in Sect. 2.2

$\mathbf{d}^\top$	$-\mathbf{1}^\top R \in \mathbb{R}^{m\times 1}$. vector of decay rates $d_1, \dots, d_m$
$J(t)$	$\mathbf{d}^\top c(t)$. flux out of the system
$p_{\text{abs}}(t)$	$\mathbf{d}^\top e^{Rt}\mathbf{e}_1$ where $t \geq 0$. A probability density
T_{abs}	Random network time with density p_{abs}
t_n	Expected time a random walk from 1 spends in compartment n
$p_n(t)$	$\frac{1}{t_n}\mathbf{e}_n^\top e^{Rt}\mathbf{e}_1$ where $t \geq 0$. A probability density
T_n	Random network time with density $p_n(t)$
T	Random network time. Could be T_{abs} or T_n

In this simple case, $m = 1$ and $R = -d$. The flux out of the system is

$$J(t) = dc(t) = d\int_0^t e^{-d(t-s)} I(s)\,\mathrm{d}s.$$

Upon changing variables $s' = t - s$, we can write this output as the following conditional expectation,

$$J(t) = \int_0^\infty de^{-ds'} I(t-s')\,\mathrm{d}s' = \mathbb{E}[I(t - T_{\text{abs}}) \mid \{I(s)\}_{s\leq t}], \tag{9}$$

where, for this example, T_{abs} is exponentially distributed with rate $d > 0$. This agrees with Definition 4. In words, (9) says that the output J is a weighted average of the past input I, where the weights are p_{abs}, the probability density function of T_{abs}.

As d grows, p_{abs} approaches a Dirac delta function, T_{abs} becomes highly concentrated, and J at time t becomes determined by I at time t. As d vanishes, T_{abs} becomes highly variable, and J at time t approaches an average with similar weights for any past inputs. Reminiscent of the law of large numbers, we thus expect fluctuations in input to average each other out if d is small, resulting in $J(t)$ being almost deterministic. Summarizing, J becomes less variable as the variability in T_{abs} increases.

This intuition of a negative correlation between robustness and variability in T extends beyond the simple exponential decay case. Indeed, one can show that

$$J(t) = \mathbb{E}[I(t - T_{\text{abs}}) \mid \{I(s)\}_{s\leq t}]$$

holds for all admissible networks, and we collect that calculation at the beginning of the proof of Theorem 1 in the appendix. It is also straightforward to apply this intuition to c_n and T_n. The remainder of Sect. 2 gives rigorous results where the variance of our network times serve as a summary statistic for robustness robustness. The notation introduced in this section is collected in Table 2.

2.3 Decomposing Network Variability

The two factors determining the dynamics of the system (4) are (i) the stochastic input $I = \{I(t)\}_{t\in\mathbb{R}}$ and (ii) the topology and rates of the network R. The following result shows how (i) and (ii) contribute to the variability in the flux out of the system. Throughout this paper,

$$\mathbb{E}[Z], \quad \mathrm{Var}(Z), \quad \mathrm{SD}(Z), \quad \text{and } \mathrm{CV}(Z)$$

respectively denote the mean, variance, standard deviation, and coefficient of variation of a random variable Z (note that $\mathrm{CV}(Z) = \mathrm{SD}(Z)/\mathbb{E}[Z]$).

Theorem 1 *Suppose $c = \{c(t)\}_{t\in\mathbb{R}}$ satisfies (4) for some admissible $R \in \mathbb{R}^{m\times m}$. Then, for any time $t \in \mathbb{R}$, the variance of the flux J in (5) is*

$$Var(J(t)) = \mathbb{E}[K(t - T_{abs}, t - T'_{abs})], \tag{10}$$

where

$$K(t, s) := \mathbb{E}[\xi(t)\xi(s)]$$

is the correlation function of the noisy input and T_{abs}, T'_{abs} are independent and identically distributed (iid) realizations of a random variable with probability density in (6).

Similarly, when interested in the variability in some given node of interest $n \in \{1, 2, \dots, m\}$, we have the following result.

Theorem 2 *Suppose $c = \{c(t)\}_{t\in\mathbb{R}}$ satisfies (4) for some admissible $R \in \mathbb{R}^{m\times m}$. Then, for any time $t \in \mathbb{R}$ and compartment $n \in \{1, 2, \dots, m\}$, we have that*

$$Var(c_n(t)) = t_n^2 \mathbb{E}[K(t - T_n, t - T'_n)], \tag{11}$$

where T_n, T'_n are iid realizations of a random variable with probability density in (7).

Theorems 1–2 thus decompose the variance of J and c_n into (i) properties of the noisy input ξ (through its correlation function $K(\cdot, \cdot)$) and (ii) properties of the network (through T_{abs}, T_n, and t_n).

The following definition characterizes the networks for which $p_{\mathrm{abs}} = p_n$ (and thus $T_{\mathrm{abs}} = T_n$).

Definition 6 A network $R \in \mathbb{R}^{m\times m}$ is called $\underline{n\text{−}terminal}$ if it is admissible and

$$d_n = -\sum_{j=1}^{m} R_{jn} > 0 \quad \text{for some } n \in \{1, \dots, m\},$$

$$\text{and} \quad d_i = -\sum_{j=1}^{m} R_{ji} = 0 \quad \text{for all } i \neq n.$$

In words, a network is n-terminal if compartment n is the only compartment with nonzero decay. The following result follows from a quick calculation (collected in the appendix with the rest of the proofs).

Proposition 2 *If R is n-terminal, then $p_{abs} = p_n$ and t_n in (8) is given by*

$$t_n = 1/d_n > 0.$$

2.4 Stationary Noisy Input

If the input $I(t) = \mathbb{E}[I(t)] + \xi(t)$ is wide-sense stationary (WSS) (Pavliotis 2014), then $\mathbb{E}[I(t)] = \mathbb{E}[I]$ is constant in time and the correlation function simplifies to

$$K(t, s) = \mathbb{E}[\xi(t - s)\xi(0)] =: K(t - s),$$

and must satisfy (Pavliotis 2014)

$$|K(t)| \leq K(0) = \mathrm{Var}(\xi) = \mathrm{Var}(I), \quad t \in \mathbb{R}.$$

In the WSS case, the conclusions of Theorems 1–2 simplify, and we state these in the following two corollaries.

Corollary 1 *If the input $I(t)$ is WSS and not constant in time, then the variance of the flux out of the system in (5) satisfies*

$$Var(J) = \mathbb{E}[K(T_{abs} - T'_{abs})] < Var(I). \tag{12}$$

Corollary 2 *If the input $I(t)$ is WSS, then the mean of compartment $n \in \{1, 2, \dots, m\}$ is*

$$\mathbb{E}[c_n] = t_n \mathbb{E}[I].$$

If the input $I(t)$ is WSS, not constant in time, and satisfies $\mathbb{E}[I] \neq 0$, then the coefficient of variation of compartment n satisfies

$$\mathrm{CV}(c_n) := \frac{\mathrm{SD}(c_n)}{\mathbb{E}[c_n]} = \frac{\sqrt{\mathbb{E}[K(T_n - T'_n)]}}{\mathbb{E}[I]} < \frac{\sqrt{\mathrm{Var}(I)}}{\mathbb{E}[I]}. \tag{13}$$

The bound $\mathrm{Var}(J) < \mathrm{Var}(I)$ in (12) implies that the variance of the output is always less than the variance of the input. That is, any linear system must strictly dampen a noisy input. The bound $\mathrm{CV}(c_n) < \sqrt{\mathrm{Var}(I)}/\mathbb{E}[I]$ in (13) was previously shown in Theorem 3.1 in Anderson et al. (2007).

2.5 Network Robustness from Variability in *T*

We are interested in determining under what conditions a network is robust to a WSS noisy input $I(t)$ with $\mathbb{E}[I] > 0$ and $\mathrm{Var}(I) > 0$. There are two natural notions of "robust" which we consider, where either notion might be more relevant depending on the specific application. The first notion of robustness is that the variance of the output J in (5) is much less than the variance of the input I,

$$\mathrm{Var}(J) \ll \mathrm{Var}(I). \tag{14}$$

Since $\mathbb{E}[J] = \mathbb{E}[I]$ (this calculation is collected in Proposition 4 in the appendix), note that (14) is equivalent to a statement about coefficients of variation, i.e. $\mathrm{CV}(J) \ll \mathrm{CV}(I)$.

Rather than the output, in some applications one is more interested in the concentration in some specific node $n \in \{1, \ldots, m\}$. Hence, the second notion of robustness is that the coefficient of variation of the concentration in node n is much less than the coefficient of variation of the input,

$$\mathrm{CV}(c_n) \ll \mathrm{CV}(I). \tag{15}$$

We note that $\mathrm{Var}(c_n)$ and $\mathrm{Var}(I)$ cannot be compared directly since they have different units (concentration versus concentration per time).

Assuming merely that the noisy input has a decaying correlation function (i.e. $K(t) \to 0$ as $|t| \to \infty$), the notion of network robustness in (14) corresponds to variability in the random variable T_{abs} with probability density in (6) since Corollary 1 implies

$$\frac{\mathrm{Var}(J)}{\mathrm{Var}(I)} = \frac{\mathbb{E}[K(T_{\mathrm{abs}} - T'_{\mathrm{abs}})]}{K(0)} < 1.$$

That is, if T_{abs} is highly variable (compared to the decay rate of K), then $|T_{\mathrm{abs}} - T'_{\mathrm{abs}}|$ is likely large and thus $K(T_{\mathrm{abs}} - T'_{\mathrm{abs}})$ is likely small. By the same argument, the notion of network robustness in (15) corresponds to variability in the random variable T_n with probability density in (7) since Corollary 2 implies

$$\frac{\mathrm{CV}(c_n)}{\mathrm{CV}(I)} = \sqrt{\frac{\mathbb{E}[K(T_n - T'_n)]}{K(0)}} < 1.$$

Therefore, we propose that measures of the variability of T_{abs} and T_n can serve as summary statistics of the robustness of the network (where one uses T_{abs} for the robustness notion in (14) and T_n for (15)). A natural measure of the variability of T_{abs} or T_n which can be quickly computed via the matrix R is its variance. Specifically, a quick calculation using (6) and (7) (see Proposition 5 in the appendix) shows that the variance of T_{abs} is

$$\mathrm{Var}(T_{\mathrm{abs}}) = -2\sum_{i=1}^{m} d_i (R^{-3})_{i1} - \left(\sum_{i=1}^{m} d_i (R^{-2})_{i1}\right)^2, \tag{16}$$

where $d_i := -\sum_{j=1}^{m} R_{ji} \geq 0$, and the variance of T_n is

$$\mathrm{Var}(T_n) = \frac{2(R^{-3})_{n1}}{(R^{-1})_{n1}} - \left(\frac{(R^{-2})_{n1}}{(R^{-1})_{n1}}\right)^2. \tag{17}$$

Hence, by setting either

$$T = T_{\mathrm{abs}}, \quad \text{or} \quad T = T_n,$$

depending on the notion of robustness (either (14) or (15)), the formulas (16)–(17) allow one to quickly estimate the robustness of a network R in terms of a single timescale,

$$\mathrm{SD}(T) := \sqrt{\mathrm{Var}(T)},$$

where large values of $\mathrm{SD}(T)$ correspond to more robust networks. Indeed, (16)–(17) can be computed in seconds on a modern computer for networks with thousands of nodes. We note that $(R^{-j})_{i1}$ is often most efficiently computed by solving the linear system $R^j v = \mathbf{e}_1$ for $v \in \mathbb{R}^{m\times 1}$ and setting

$$(R^{-j})_{i1} = \mathbf{e}_i^\top v.$$

Furthermore, to estimate the robustness of a given network R to a given WSS noisy input I, we can compare $\mathrm{SD}(T)$ to the correlation time of I. That is, if the noisy input has a finite, nonzero correlation time defined by

$$\tau := \frac{1}{K(0)} \int_0^\infty K(t)\,\mathrm{d}t, \tag{18}$$

then our analysis suggests that the network R is robust to I if

$$\mathrm{SD}(T) \gg \tau.$$

We note that in the case of white noise input (which has zero correlation time and infinite variance), more variable network times T still correspond to more robust networks (see section 3 below).

3 Numerical Simulations

We argued in section 2 that variability in the network time T is a proxy for network robustness. We now illustrate this point via numerical simulations. In particular, for a variety of randomly generated networks and stochastic inputs, we show that $\mathrm{Var}(J)$ is inversely related to $\mathrm{SD}(T_{\mathrm{abs}})$, and $\mathrm{CV}(c_n)$ is inversely related to $\mathrm{SD}(T_n)$. The results of the simulations are seen in Fig. 2. Section 3.1 summarizes and interprets Fig. 2, while Sects. 3.2–3.4 detail how the simulation results were computed.

3.1 Simulation Results

For WSS noisy inputs, we proved in Sect. 2 that the only property of the noisy input that affects the variance in the network is the correlation function $K(t) := \mathbb{E}[\xi(t)\xi(0)]$. In the simulations in this section, we consider two choices of the correlation function. We suppose that the correlation function is either (i) a decaying exponential,

$$K(t) = \sigma^2 e^{-|t|/\tau}, \tag{19}$$

where $\tau > 0$ is the correlation time and $\mathrm{Var}(\xi) = \sigma^2 > 0$, or (ii) a Dirac delta function,

$$K(t) = 2\tau\sigma^2\delta(t), \tag{20}$$

where $\tau > 0$ and $\sigma\tau > 0$ are characteristic time and concentration scales of the noisy input, respectively. As we detail below, (19) corresponds to a variety of noisy inputs, including that $\xi(t)$ is an Ornstein–Uhlenbeck process or a telegraph process. Further, (20) corresponds to the case that $\xi(t)$ is white noise.

In the top panels of Fig. 2, we plot the standard deviation of the flux out of the system,

$$\mathrm{SD}(J(t)) := \sqrt{\mathrm{Var}\left(\sum_{i=1}^{m} d_i c_i(t)\right)} = \sqrt{\mathbb{E}[K(T_{\mathrm{abs}} - T'_{\mathrm{abs}})]}, \tag{21}$$

against $\mathrm{SD}(T_{\mathrm{abs}})$ for 10^4 different randomly generated networks. In the bottom panels of Fig. 2, we plot the coefficient of variation for the nth compartment,

$$\mathrm{CV}(c_n(t)) := \frac{\mathrm{SD}(c_n(t))}{\mathbb{E}[c_n(t)]} = \frac{\sqrt{\mathbb{E}[K(T_n - T'_n)]}}{\mathbb{E}[I]}, \tag{22}$$

against $\mathrm{SD}(T_n)$ for 10^4 different randomly generated networks. The final equalities in (21)–(22) are due to Corollaries 1–2 and we set $\mathbb{E}[I] = 1$ and $\tau = \sigma^2 = 1$ in (19)–(20). The left and right panels of Fig. 2 are for noisy inputs with correlation functions in (19) and (20), respectively. We detail how the networks were generated in Sect. 3.4, but here we note that the number of nodes in each network ranges from $m = 1$ to $m = 100$, and approximately half of the networks are linear chains (half of which are reversible and half irreversible) and half are Erdos-Renyi random networks. Further, all of the rate constants in the networks were chosen randomly.

Figure 2 demonstrates that $\mathrm{SD}(T) = \sqrt{\mathrm{Var}(T)}$ does indeed characterize network robustness (where one sets $T = T_{\mathrm{abs}}$ or $T = T_n$ depending on the notion of robustness, see Sect. 2.5). While close inspection of Fig. 2 shows that it is possible for a network to have a larger value of $\mathrm{SD}(T)$ and yet be less robust than another network, the overwhelming majority of the networks in Fig. 2 show that a larger value of $\mathrm{SD}(T)$ entails a more robust network. Furthermore, the regions $\mathrm{SD}(T) \ll \tau$ and $\mathrm{SD}(T) \gg \tau$ indeed respectively characterize not robust and robust networks.

Another feature of Fig. 2 is that the data points for the randomly generated networks agree closely with the red curves. The red curve in each plot is simply for the case

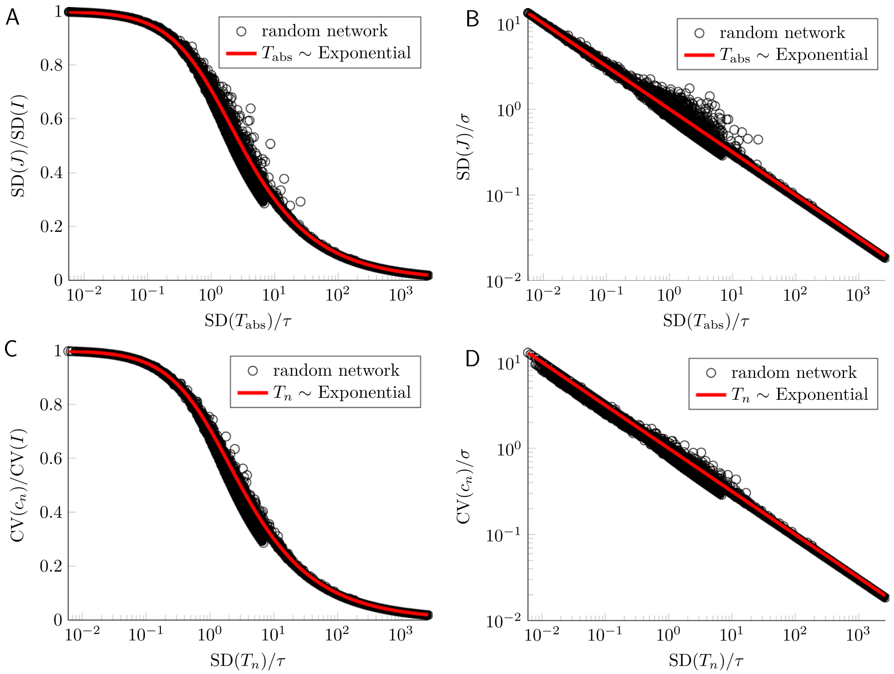

Fig. 2 Variability in network time T characterizes network robustness. The circle markers are for 10^4 randomly generated networks and the red curves depict the case that T is exponentially distributed. **A** Standard deviation of flux out of the system as a function of the standard deviation of $T_{\rm abs}$ for noisy inputs with exponentially decaying correlation functions. **B** Standard deviation of the flux out of the system as a function of the standard deviation of $T_{\rm abs}$ for white noise input. **C** Coefficient of variation of compartment of interest c_n as a function of the standard deviation of T_n for noisy inputs with exponentially decaying correlation functions. **D** Coefficient of variation of compartment of interest c_n as a function of the standard deviation of T_n for white noise input. In all plots, the mean input is $\mathbb{E}[I] = 1$ (Color figure online)

that T is exactly exponentially distributed. In the case of the exponentially decaying correlation function in (19), we show below that this is

$$\sqrt{\mathbb{E}[K(T-T')]} = \frac{\sigma}{\sqrt{1+\mathrm{SD}(T)/\tau}} \qquad \text{if } T \sim \text{Exponential}. \tag{23}$$

In the case of the Dirac delta function correlation function in (20), we show below that this is

$$\sqrt{\mathbb{E}[K(T-T')]} = \frac{\sigma}{\sqrt{\mathrm{SD}(T)/\tau}} \qquad \text{if } T \sim \text{Exponential}. \tag{24}$$

The result that the circle markers tend to lie on the red curves in Fig. 2 suggests that the network time T is approximately exponentially distributed for many of the random networks. This result is perhaps not surprising in light of the matrix exponential forms of the probability density function of T in (6)–(7). Furthermore, we show in section 4 below that T is equivalent to a certain first passage time on an associated network, and first passage times on networks are known to be approximately exponentially distributed in many scenarios (Meyer et al. 2011).

3.2 Exponentially Decaying Correlation Function

A variety of WSS noisy inputs have an exponentially decaying correlation function as in (19). For example, suppose $\xi(t)$ is an Ornstein–Uhlenbeck process, and thus satisfies the following stochastic differential equation,

$$d\xi = -\frac{\xi}{\tau}\,dt + \sqrt{\frac{2\sigma^2}{\tau}}\,dW, \tag{25}$$

where W is a standard Brownian motion. In this case, the correlation function of ξ is in (19) with correlation time $\tau > 0$ and variance $\sigma^2 > 0$.

Another WSS noisy input with the exponentially decaying correlation function in (19) is a telegraph process (also called dichotomous Markov noise (Bena 2006)), which is simply a two-state, continuous-time Markov chain. In particular, suppose $\xi(t)$ switches between values ξ_0 and ξ_1 according to

$$\xi_0 \underset{(1-q)/\tau}{\overset{q/\tau}{\rightleftharpoons}} \xi_1. \tag{26}$$

That is, ξ jumps from state ξ_0 to state ξ_1 at rate q/τ and jumps from state ξ_1 to state ξ_0 at rate $(1-q)/\tau$ for $q \in (0, 1)$ and $\tau > 0$. In this case, the correlation function of ξ is in (19) with correlation time $\tau > 0$ and variance

$$\mathrm{Var}(\xi) = \sigma^2 := (\xi_0 - \xi_1)^2 q(1-q) > 0.$$

We note that though both the Ornstein–Uhlenbeck process in (25) and the telegraph process in (26) share the same correlation function in (19), the Ornstein–Uhlenbeck process is Gaussian but the telegraph process is not.

The data points in Fig. 2 are computed for this exponentially decaying correlation function (19) via

$$\begin{aligned}\mathbb{E}[K(T-T')] &= \sigma^2 \int_0^\infty \int_0^\infty p(s)p(s')e^{-|s-s'|/\tau}\,ds'\,ds\\ &= 2\sigma^2 \int_0^\infty \int_0^s p(s)p(s')e^{-(s-s')/\tau}\,ds'\,ds,\end{aligned} \tag{27}$$

and computing the double integral in (27) numerically (where $T = T_{\mathrm{abs}}$ and $p = p_{\mathrm{abs}}$ or $T = T_n$ and $p = p_n$). Note that (23) follows immediately from (27) upon setting $p(t) = \lambda e^{-\lambda t}$ and $\mathrm{SD}(T) = 1/\lambda$.

3.3 White Noise Input

The Dirac delta function correlation function in (20) corresponds to a white noise input. As is well-known (Pavliotis 2014), white noise can be obtained from the Ornstein–Uhlenbeck process in (25) by taking the correlation time to zero and the variance to infinity while keeping their ratio fixed. Specifically, if we rescale the correlation time and variance in (25) according to $\tau \to \varepsilon\tau$ and $\sigma^2 \to \sigma^2/\varepsilon$ and take $\varepsilon \to 0$ we obtain

$$\lim_{\varepsilon\to 0} K(t) = \lim_{\varepsilon\to 0} \mathbb{E}[\xi(t)\xi(0)] = \lim_{\varepsilon\to 0} \frac{\sigma^2}{\varepsilon} e^{-|t|/(\varepsilon\tau)} = 2\tau\sigma^2\delta(t),$$

which implies the heuristic

$$\lim_{\varepsilon\to 0} \xi(t) = \sqrt{2\tau\sigma^2}\,\frac{\mathrm{d}W}{\mathrm{d}t},$$

where $\frac{\mathrm{d}W}{\mathrm{d}t}$ denotes white noise. In this white noise limit, τ is no longer the correlation time of ξ and σ^2 is no longer the variance of ξ (since white noise has zero correlation time and infinite variance), but we retain the parameters $\tau \in (0,\infty)$ and $\sigma \in (0,\infty)$ since they represent characteristic scales of ξ.

Having taken the white noise limit, we then obtain

$$\begin{aligned}\mathbb{E}[K(T-T')] &= 2\tau\sigma^2 \int_0^\infty \int_0^\infty p(s)p(s')\delta(s-s')\,\mathrm{d}s\,\mathrm{d}s' \\ &= 2\tau\sigma^2 \int_0^\infty (p(s))^2\,\mathrm{d}s = 2\tau\sigma^2\|p\|_2^2 = 2\tau\sigma^2\mathbb{E}[p(T)],\end{aligned} \tag{28}$$

where $T = T_{\mathrm{abs}}$ and $p = p_{\mathrm{abs}}$ or $T = T_n$ and $p = p_n$. Note that $\mathbb{E}[p(T)]$ is a measure of the variability of T (with small values of $\mathbb{E}[p(T)]$ indicating a more variable T), which is akin to the Simpson index (Simpson 1949). Note that (24) follows immediately from (28) upon setting $p(t) = \lambda e^{-\lambda t}$ and $\mathrm{SD}(T) = 1/\lambda$.

3.4 Random Network Construction

We now describe how the 10^4 random networks were constructed for Fig. 2. These networks were created to show that variability in T corresponds to network robustness across a large variety of networks (i.e. across networks that vary considerably in their size, connectivity or topology, and rate constants).

First, the size of the network was chosen according to

$$m = \mathrm{round}(10^{2U}),$$

where $U \in [0,1]$ is uniformly distributed and round$(\cdot)$ merely rounds to the nearest integer. Next, a random variable Θ was drawn uniformly on $\{1,2,3,4\}$ to determine the type of network.

If $\Theta = 1$, then the network is an irreversible linear chain with compartment of interest given by the terminal node $n = m$. Further, the rate constants were taken to be

$$R_{i+1,i} = 1 + E_i, \quad i \in \{1,2,\ldots,m-1\},$$

and $d_m = 1 + E_m$, where $\{E_i\}_{i=1}^m$ are iid unit rate exponential random variables.

If $\Theta = 2$, then the network is a reversible linear chain created identically to the case $\Theta = 1$, except rather than setting $R_{i,i+1} = 0$, we set $R_{i,i+1} = R_{i+1,i}$ for $i \in \{1, 2, \ldots, m-1\}$.

If $\Theta = 3$, then the network was taken to be of Erdos-Renyi type. Specifically, we let connection probability $p \in [0, 1]$ be uniformly distributed on $[0, 1]$ and set the transition rate from i to $j \neq i$ to be

$$R_{j,i} = Z_{i,j}(1 + E_{i,j}), \quad i, j \in \{1, 2, \ldots, m\}, \ i \neq j,$$

and set the decay rate from state i to be $d_i = Z_{0,i}(1 + E_{0,i})$, where $\{Z_{i,j}\}_{i,j}$ are iid with

$$\mathbb{P}(Z_{i,j} = 1) = p, \quad \mathbb{P}(Z_{i,j} = 0) = 1 - p,$$

and $\{E_{i,j}\}_{i,j}$ are iid unit rate exponential random variables. The compartment of interest n was taken to be m.

If $\Theta = 4$, we first create a network R' identically to the case $\Theta = 3$ and then we set the network to be $R_{ji} = (R'_{ji} + R'_{ij})/2$ so that $R_{ij} = R_{ji}$ for all $i, j \in \{1, 2, \ldots, m\}$.

Finally, to ensure that there is a path from state 1 to state n and that all the eigenvalues of R have strictly negative real parts for cases $\Theta = 3$ and $\Theta = 4$, if the random construction of the network resulted in an R such that the condition number of R^3 exceeded 10^{16}, then the connection probability p, the connections $\{Z_{i,j}\}_{i,j}$, and the jump rates $\{E_{i,j}\}_{i,j}$ were resampled.

4 Stochastic Analysis of Network Time *T*

In Sects. 2–3 above, we showed that the robustness of a network depends on the variability of a certain random time T whose probability distribution is defined by the network R. In this section, we study representations of T in terms of a random walk on the network (Sects. 4.1–4.2). We then use these representations of T to study how certain network structures increase or decrease network robustness (Sects. 4.3–4.5). In particular, by giving a stochastic interpretation of T in terms of random walks on the network, we obtain (i) understanding for why certain network structures increase robustness and (ii) a way to study robustness in terms of stochastic analysis of random walks.

4.1 Stochastic Representation of Network Time *T*

To define a random walk on the network, let $X = \{X(t)\}_{t \geq 0}$ be the continuous-time Markov chain in (2). The next result follows from an elementary calculation (collected in the appendix with the rest of the proofs).

Theorem 3 *The first passage time of X to the absorbing state,*

$$T^X_{abs} := \inf\{t \geq 0 : X(t) = \varnothing\}, \tag{29}$$

has the same probability density p_{abs} in (6) as T_{abs} if we condition that $X(0) = 1$.

We use the superscript X in (29) because T_{abs}^X is defined by paths of X, whereas T_{abs} is defined simply by its probability density p_{abs}.

The next result is that the probability density p_n in (7) describes the time that X leaves node n for the last time (the so-called last passage time (Bao and Jia 2004; Comtet et al. 2020)), conditioned that X reaches node n starting from node 1.

Theorem 4 *The last passage time of X to node $n \in \{1, 2, \dots, m\}$,*

$$T_n^X := \sup\{t \geq 0 : X(t) = n\}, \tag{30}$$

has the same probability density p_n in (7) as T_n if we condition that $X(0) = 1$ and $X(t) = n$ for some $t > 0$.

Although it may not be intuitively clear why the probability densities of T_n and T_n^X match (and similarly for T_{abs} and T_{abs}^X), it is intuitive that the last passage time is relevant to robustness. Recall from the intuition given in Sect. 2.2 that T_n is of interest since c_n is a weighted average of past inputs I whose weights are p_n. A highly variable T_n means that inputs across a wide range in time contribute to c_n. Similarly, a highly variable last passage time T_n^X implies that particles from inputs across a wide range in time could be in compartment n at time t.

The proof of Theorem 4 follows from showing that T_n^X has the same distribution as the first passage time of a modified network, then using a similar elementary calculation as in the proof of Theorem 3 to show that they have probability density p_n. We now discuss the construction of the modified network.

We first relate p_n in (7) to the first passage time to the absorbing state $\varnothing$ of a suitably modified random walk denoted $\widetilde{X} = \{\widetilde{X}(t)\}_{t\geq 0}$. In addition to its use in proving Theorem 4, reframing the last passage time as a first passage time can also be helpful for leveraging the much vaster literature on first passage times to analyze specific networks. To define $\widetilde{X}$, let A be the event that X reaches compartment n,

$$A = \{X(t) = n \text{ for some } t \geq 0\}, \tag{31}$$

and let $\mathbb{P}_i$ denote the probability measure conditioned that $X(0) = i$. Without loss of generality, we assume

$$\mathbb{P}_j(A) > 0 \quad \text{for all } j \in \{1, 2, \dots, m\},$$

since if $\mathbb{P}_j(A) = 0$, then we can replace compartment j with the absorbing state $\varnothing$ without affecting p_n in (7) (or T_n^X in Theorem 4). Define the following modified reaction rate matrix $\widetilde{R} = \{\widetilde{R}_{ij}\}_{i,j=1}^m \in \mathbb{R}^{m\times m}$ via

$$\widetilde{R}_{ij} = R_{ij}\mathbb{P}_i(A)/\mathbb{P}_j(A), \tag{32}$$

and note that $\widetilde{R}_{ii} = R_{ii}$. Define $\widetilde{X} = \{\widetilde{X}(t)\}_{t\geq 0}$ to be the continuous-time Markov chain associated with $\widetilde{R}$ (analogous to X in (2) associated with R). In words, $\widetilde{R}$ is a re-weighted version of R so that $\widetilde{X}$ is sure to hit node n.

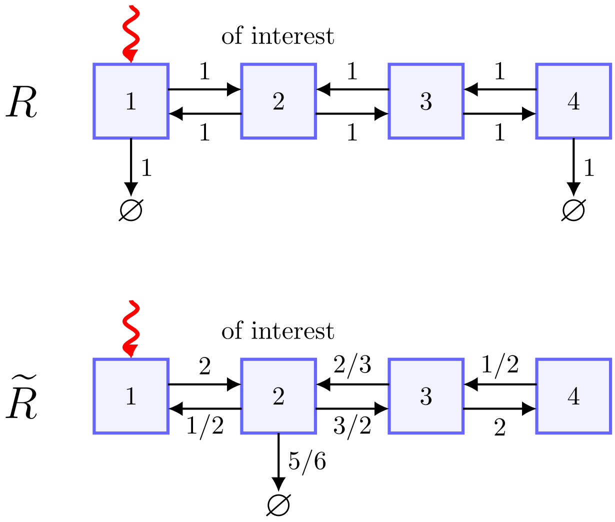

Fig. 3 A network R (top) and its corresponding $\widetilde{R}$ network (bottom) as defined in (32)

Theorem 5 *The first passage time of $\widetilde{X}$ to the absorbing state,*

$$T_{abs}^{\widetilde{X}} := \inf\{t \geq 0 : \widetilde{X}(t) = \varnothing\}, \tag{33}$$

has probability density p_n in (7) if we condition that $\widetilde{X}(0) = 1$.

Recall that a network is called n-terminal if node n is the only compartment with nonzero decay (see Definition 6). If R is n-terminal, then it follows immediately that $R = \widetilde{R}$ and $T_{\text{abs}}^{X} = T_{\text{abs}}^{\widetilde{X}} = T_n^X$. The following result (proof collected in Lemma 2 in the appendix) states that $\widetilde{R}$ is always n-terminal.

Proposition 3 *The network $\widetilde{R}$ in (32) is n-terminal.*

4.2 An Illustration of $\widetilde{R}$

In this section, we find the n-terminal $\widetilde{R}$ associated with R in Fig. 3 where $n = 2$ is the node of interest. To start, we compute the probabilities that the random walk X starting at compartment i reaches compartment $n = 2$. It is clear that $\mathbb{P}_2(A) = 1$ and $\mathbb{P}_\varnothing(A) = 0$. Since all rates are 1, the other $\mathbb{P}_i(A)$ must be the average of $\mathbb{P}_j(A)$ where j are the neighbors of i. As such

$$\mathbb{P}_1(A) = \frac{1}{2}, \quad \mathbb{P}_2(A) = 1, \quad \mathbb{P}_3(A) = \frac{2}{3}, \quad \mathbb{P}_4(A) = \frac{1}{3}.$$

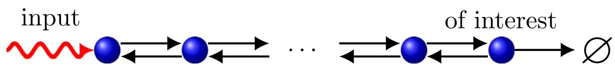

Fig. 4 A reversible linear chain

Using the definition $\widetilde{R}_{ij} = R_{ij}\mathbb{P}(A_i)/\mathbb{P}(A_j)$, we find that

$$R = \begin{pmatrix} -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \end{pmatrix}, \quad \widetilde{R} = \begin{pmatrix} -2 & 1/2 & 0 & 0 \\ 2 & -2 & 3/2 & 0 \\ 0 & 2/3 & -2 & 2 \\ 0 & 0 & 1/2 & -2 \end{pmatrix}.$$

Notice that most edges in R are still in $\widetilde{R}$. Edges only disappear when moving along that edge prevents the particle from ever returning to the compartment of interest, which in this case are the edges that lead to decay. The only edge that can be created is from the compartment of interest to decay. An advantage of using $\widetilde{R}$ over R is that it is n-terminal, and therefore T_n can be formulated as in (33) as the first hitting time of $\widetilde{X}$ to $\varnothing$.

4.3 Linear Chains

We first consider an irreversible, linear chain of $m = n$ compartments. Suppose particles move from compartment $i \neq n$ to compartment $i+1$ at rate $R_{i+1,i} = k_i > 0$, except compartment n where particles decay to state $\varnothing$ at rate $d_n = k_n > 0$. A visual representation is seen in Fig. 1b.

Since R is n-terminal, we have $t_n = 1/d_n$ and the network time T_n is simply the first passage time to state $\varnothing$, which is

$$T_n = \sum_{i=1}^{n} \frac{E_i}{k_i}, \tag{34}$$

where $\{E_i\}_{i=1}^n$ are iid unit rate exponential random variables. Therefore,

$$\mathrm{Var}(T_n) = \sum_{i=1}^{n} \frac{1}{k_i^2}. \tag{35}$$

Equation (35) implies that the longer the chain is, the larger the variance of T_n and therefore suggests that more compartments make the network more robust to stochastic input. This result agrees with the result of Anderson et al. (2007), which showed that the variances of the fluxes strictly decrease down an irreversible linear chain. We also note that (34) implies that T_n is unaffected by any permutation of the rates and (35) implies that decreasing any of the rates $k_1, \ldots, k_n$ increases $\mathrm{Var}(T_n)$ and thus tends to increase network robustness.

Suppose we now make the chain reversible by supposing that particles move from compartment $i \in \{2, \ldots, n\}$ back to compartment $i-1$ at rate $R_{i-1,i} > 0$, as seen in

Fig. 4. Since R is still n-terminal, we have $t_n = 1/d_n$ and T_n is the first passage time to $\varnothing$. Since this reversible linear chain is a birth-death process, it is known that this first passage time is (Fill 2009)

$$T_n = \sum_{i=1}^{n} \frac{E_i}{\lambda_i},$$

where $\{E_i\}_{i=1}^n$ are iid unit rate exponential random variables and $\{\lambda_i\}_{i=1}^n$ are the (necessarily positive) eigenvalues of $-R$.

4.4 Adding a Side Network

We now consider the effect of adding a general "side" network onto a general network (i.e. not necessarily a linear chain network). We define a side network as a network attached onto an existing network R at compartment $s \in \{1, 2, \ldots, m\}$ such that random walks can enter the side network only through compartment s and must eventually leave the side network through compartment s. Define T_{abs} and T_n as in Theorems 1–2 for the existing network without the added side network, and define T^*_{abs} and T^*_n analogously for the network with the added side network.

Theorem 6 *If a side network is attached to a node $s \in \{1, 2, \ldots, m\}$ whose removal would disconnect paths from node 1 to the absorbing state $\varnothing$, then*

$$Var(T^*_{abs}) > Var(T_{abs}).$$

If a side network is attached to a node $s \in \{1, 2, \ldots, m\}$ whose removal would disconnect paths from node 1 to node $n \in \{1, 2, \ldots, m\}$, then

$$Var(T^*_n) > Var(T_n).$$

Theorem 6 thus suggests that adding a side network at compartments that paths must cross increases robustness. Hence, as a general principle, this result shows that network robustness can typically be increased by increasing the network size and complexity.

We point out that the conclusion of Theorem 6 that adding a side network increases the variance of the network time may not hold if the side network is attached to a node s whose removal would not disconnect paths as in Theorem 6. Indeed, an example where a side chain decreases the variance can be seen in Fig. 5a. A quick calculation for this example with node of interest $n = 5$ shows that

$$\mathrm{Var}(T^*_{\mathrm{abs}}) = \mathrm{Var}(T^*_n) = 70.25 < 72 = \mathrm{Var}(T_{\mathrm{abs}}) = \mathrm{Var}(T_n).$$

This example does not contradict Theorem 6 because the side chain is added at node $s = 3$, and paths from node 1 can reach $\varnothing$ and node $n = 5$ without hitting node $s = 3$.

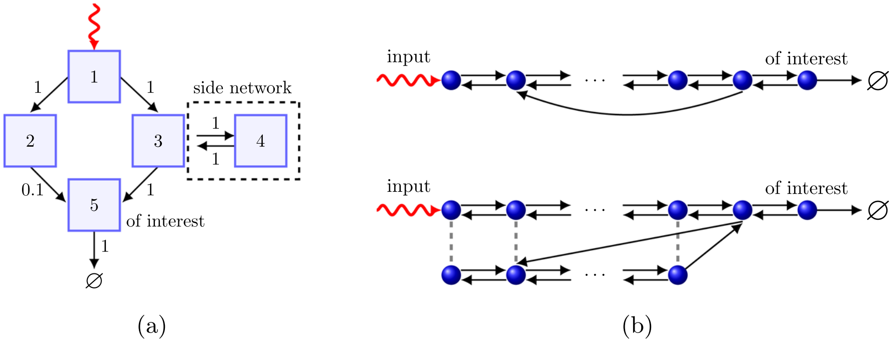

Fig. 5 **a** An example network where adding a side network (seen in dotted box) reduces the variance of the network time. **b** A linear chain with feedback loop (top) has the same T as a linear chain with side network (bottom). The gray dashed lines indicate that the nodes (and rates) in the side network are copied from the nodes in the linear chain

4.5 Linear Chain with Feedback

Theorem 6 allows us to study the effect of feedback loops on a linear chain. Specifically, if the network is a linear chain (either reversible or irreversible) with a feedback loop, then we can create a network with an equivalent network time by adding a side network to a linear chain (see Fig. 5b). This equivalent network is formed through the same idea as the proof of Theorem 5.3 in Anderson et al. (2007). As the new network is a reversible linear chain with a side network, the following result is a corollary of Theorem 6.

Corollary 3 *Suppose R is a linear chain (either reversible or irreversible). If R^* is formed by adding a feedback loop from node $s \in \{1, 2, \ldots, n\}$ to node $j \in \{1, 2, \ldots, s\}$, then*

$$Var(T_n^*) > Var(T_n).$$

5 Discussion

In this paper, we have argued that the robustness of a network to a stochastic input can be understood and quantified in terms of the variability of a certain "network time" T. The probability distribution of the network time is defined by the network (see (6)–(7)), and the network time can also be defined in terms of certain passage times of a random walk on the network (see Theorems 3, 4, and 5). The network time T is independent of the noisy input, and $\mathrm{Var}(T)$ can be quickly computed on a modern computer, even for complex networks with thousands of nodes. Further, we can estimate the robustness of a network to a given noisy input by comparing $\mathrm{SD}(T) = \sqrt{\mathrm{Var}(T)}$ to the correlation time of the noisy input. Computational tests on many complex networks demonstrate the validity of using $\mathrm{Var}(T)$ to estimate network robustness (see Sect. 3). Importantly, linking network robustness with variability in T gives intuition for why certain network structures increase robustness (see Sects. 4.3–4.5). Indeed, our analysis shows that large, complex networks (perhaps with many feedback loops) generally increase robustness by increasing the variability in certain random walk passage times on the network.

Of course, it is impossible for a single statistic to encapsulate all the properties of how a network (a possibly very high-dimensional object) modifies all stochastic inputs (an infinite dimensional space). Nevertheless, we propose that $\mathrm{Var}(T)$ is an effective summary statistic for network robustness. That is, similar to how the variance of a random variable is a useful measure of the variability of a probability distribution (an infinite dimensional space), the variance of the network time is an easily computable and useful statistic to understand and measure network robustness.

As described in the Introduction section, the basic mathematical problem of network robustness was previously studied in Anderson et al. (2007) and Browning et al. (2023). Anderson et al. 2007 was motivated by biochemical reaction networks and Browning et al. 2023 was motivated by a wide variety of multistage processes in biology. An alternative motivation stems from models of medication nonadherence, in which the ODE network describes a drug's pharmacokinetics/pharmacodynamics and the stochastic input models the drug dosing behavior of an imperfectly adherent patient. In this pharmacometric context, a "robust network" is akin to a "forgiving drug" (Osterberg et al. 2010; McAllister and Lawley 2022), where we use the notion of network robustness in (15) with the compartment of interest given by the plasma compartment or perhaps the effect compartment in the case of a pharmacodynamic model (Felmlee et al. 2012).

Most of our analysis focused on stationary stochastic inputs (specifically, WSS stochastic inputs), which is natural for applications to biochemical reaction networks (Anderson et al. 2007) and various multistage biological processes (Browning et al. 2023). We now make four comments on how the assumption of stationary stochastic (drug) inputs relate to models of medication nonadherence. First, several prior models of medication nonadherence assume stationary drug inputs (Li and Nekka 2007; Lévy-Véhel and Lévy-Véhel 2013; Fermín and Lévy-Véhel 2017), and thus our analysis is directly applicable to these prior models. Second, though a stationary drug input is a mathematically convenient assumption, more realistic models allow the drug input statistics to vary in time relative to the prescribed dosing interval. For example, if the patient is prescribed to take a medication every morning, it is reasonable to suppose that they are more likely to take dose in the morning compared to the evening (or the middle of the night). Third, stationary inputs can sometimes approximate non-stationary inputs for models of medication nonadherence. To illustrate, consider the common model of medication nonadherence (Lévy-Véhel and Lévy-Véhel 2013; Fermín and Lévy-Véhel 2017; Counterman and Lawley 2021, 2022; McAllister and Lawley 2022), which assumes that at every scheduled dosing time (i.e. at integer multiples of the prescribed dosing interval τ), the patient takes a dose with probability $p \in (0, 1)$ independently of their prior behavior. The stochastic drug input is then

$$I(t) = D \sum_{n \in \mathbb{Z}} \xi_n \delta(t - n\tau), \tag{36}$$

where $D > 0$ is the ratio of a dose size to the apparent volume of distribution, $\{\xi_n\}_{n \in \mathbb{Z}}$ is an iid sequence of Bernoulli random variables with $\mathbb{P}(\xi_n = 1) = p = 1 - \mathbb{P}(\xi_n - 0)$, and δ denotes the Dirac delta function. Pharmacokinetic models often assume that an ingested dose initially enters an "absorption" compartment and then moves irreversibly to the "plasma" compartment at a constant absorption rate $k > 0$ (Gibaldi and Perrier 1982). Hence, the input (36) into the absorption compartment is equivalent to the following input into the plasma compartment,

$$I_0(t) = \int_{-\infty}^{t} ke^{-k(t-s)} I(s)\,\mathrm{d}s = Dk \sum_{n \le t/\tau} e^{-k(t-n\tau)} \xi_n. \tag{37}$$

Now, observe that I_0 in (37) evaluated at integer multiples of the dosing interval τ satisfies the following recurrence relation,

$$I_0(n\tau + \tau) = e^{-k\tau} I_0(n\tau) + Dk\xi_{n+1}, \quad n \in \mathbb{Z}. \tag{38}$$

Further, (38) is an Euler-Maruyama approximation of the following Ornstein–Uhlenbeck process,

$$\mathrm{d}I_{\mathrm{ou}} = \frac{1 - e^{-k\tau}}{\tau}\Big(\frac{Dkp}{1 - e^{-k\tau}} - I_{\mathrm{ou}}\Big)\,\mathrm{d}t + Dk\sqrt{\frac{p(1-p)}{\tau}}\,\mathrm{d}W, \tag{39}$$

on the discrete time grid $\{n\tau\}_{n\in\mathbb{Z}}$ (Kloeden and Platen 1992). In particular, the Ornstein–Uhlenbeck process (39) approximates the solution to the recurrence relation (38) if $k\tau \ll 1$. Summarizing, the non-stationary input in (36) can be approximated by the stationary input in (39) for drugs which are absorbed slowly compared to their prescribed dosing interval. Fourth and finally, the decompositions obtained in Theorems 1–2 do not assume stationary inputs. Thus, we expect that our that basic result that network robustness is characterized by variability in T extends to non-stationary inputs, but we leave a detailed investigation of this point for future work.

Appendix: Proofs

Proofs for Section 2

Proof of Proposition 1 Define the first passage time of X in (2) to the absorbing state $\varnothing$,

$$T_{\mathrm{abs}}^{X} := \inf\{t > 0 : X(t) = \varnothing\}.$$

We claim that

$$\mathbb{P}(T_{\mathrm{abs}}^{X} < \infty \,|\, X(0) = i) = 1, \quad \text{for each } i \in \{1, \dots, m\}. \tag{40}$$

To show (40), we first recall some Markov chain definitions and theory.

Following Following Norris (1998), we write $i \to j$ if there is a directed path from i to j or if $i = j$ (meaning, either $i = j$ or there exists $k \ge 2$ distinct states

$i_1 = i, \dots, i_k = j$ with $R_{i_2,i_1} \cdots R_{i_k,i_{k-1}} > 0$). We write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$. It is straightforward to check that $\leftrightarrow$ is an equivalence relation and thus partitions $\{1, 2, \dots, m\}$ into equivalence classes called communicating classes. We say that a communicating class $C \subseteq \{1, 2, \dots, m\}$ is closed if $i \in C$ and $i \to j$ imply $j \in C$. In words, a closed class is one from which there is no escape.

Now, the absorbing state $\varnothing$ is a closed class. If there was another closed class $C \subseteq \{1, \dots, m\}$, then we must have $d_i = 0$ for all $i \in C$ by definition of closed. However, this would break property (c) in Definition 1 in our assumption that R is admissible. Therefore, all communicating classes in $\{1, \dots, m\}$ must not be closed and therefore must be transient (see Theorems 1.5.3, 1.5.5, and 3.4.1 in Norris (1998)), which proves (40).

Equation (40) implies

$$\lim_{t\to\infty} \mathbb{P}(T_{\text{abs}}^X > t \mid X(0) = i) = 0, \quad \text{for each } i \in \{1, 2, \dots, m\}.$$

Since the survival probability of T_{abs} is given in terms of the matrix exponential of R (Norris 1998),

$$\mathbb{P}(T_{\text{abs}}^X > t \mid X(0) = i) = \mathbf{1}^\top e^{Rt}\mathbf{e}_i, \quad i \in \{1, 2, \dots, m\},$$

where $\mathbf{1} \in \mathbb{R}^{m\times 1}$ is a column vector of all 1 s, it follows that all eigenvalues of R have strictly negative real parts. □

Proof of Theorem 1 It is immediate that the solution to (4) is

$$c(t) = \int_{-\infty}^{t} e^{R(t-s)}\mathbf{e}_1 I(s)\,\mathrm{d}s. \tag{41}$$

Therefore, the flux out of the system is

$$J(t) := \mathbf{d}^\top c(t) = \mathbf{d}^\top \int_{-\infty}^{t} e^{R(t-s)}\mathbf{e}_1 I(s)\,\mathrm{d}s = \int_0^\infty p_{\text{abs}}(s) I(t-s)\,\mathrm{d}s,$$

where we have changed variables $s \to t - s$ and used the definition of $p_{\text{abs}}(s) = \mathbf{d}^\top e^{Rs}\mathbf{e}_1$ in the final equality. Using that $\mathbb{E}[\xi(t)] = 0$ then yields

$$\begin{aligned}
\operatorname{Var}(J(t)) &= \mathbb{E}\Big[\Big(\int_0^\infty p_{\text{abs}}(s)\xi(t-s)\,\mathrm{d}s\Big)^2\Big] \\
&= \mathbb{E}\Big[\int_0^\infty \int_0^\infty p_{\text{abs}}(s)p_{\text{abs}}(s')\xi(t-s)\xi(t-s')\,\mathrm{d}s\,\mathrm{d}s'\Big] \\
&= \int_0^\infty \int_0^\infty p_{\text{abs}}(s)p_{\text{abs}}(s')K(t-s, t-s')\,\mathrm{d}s\,\mathrm{d}s' \\
&= \mathbb{E}[K(t - T_{\text{abs}}, t - T'_{\text{abs}})],
\end{aligned}$$

where $K(s, s') := \mathbb{E}[\xi(s)\xi(s')]$ denotes the correlation function of ξ (also called the covariance, autocovariance, or autocorrelation function (Pavliotis 2014)), and

$T_{\text{abs}}, T'_{\text{abs}}$ are iid realizations of a random variable with probability density p_{abs}. Note that we are assured that p_{abs} is a probability density by Theorem 3. $\square$

Proof of Theorem 2 From (41), we have that

$$c_n(t) = \mathbf{e}_n^\top c(t) = \mathbf{e}_n^\top \int_{-\infty}^{t} e^{R(t-s)}\mathbf{e}_1 I(s)\,\mathrm{d}s = t_n \int_0^\infty p_n(s) I(t-s)\,\mathrm{d}s, \tag{42}$$

where we have changed variables $s \to t-s$ and used the definition of $p_n(s) = t_n^{-1}\mathbf{e}_n^\top e^{Rs}\mathbf{e}_1$ in the final equality. Using that $\mathbb{E}[\xi(t)] = 0$ then yields

$$\begin{aligned} t_n^{-2}\mathrm{Var}(c_n(t)) &= \mathbb{E}\Big[\Big(\int_0^\infty p_n(s)\xi(t-s)\,\mathrm{d}s\Big)^2\Big] \\ &= \mathbb{E}\Big[\int_0^\infty \int_0^\infty p_n(s)p_n(s')\xi(t-s)\xi(t-s')\,\mathrm{d}s\,\mathrm{d}s'\Big] \\ &= \int_0^\infty \int_0^\infty p_n(s)p_n(s')K(t-s,t-s')\,\mathrm{d}s\,\mathrm{d}s' \\ &= \mathbb{E}[K(t-T_n, t-T'_n)], \end{aligned}$$

where $K(s,s') := \mathbb{E}[\xi(s)\xi(s')]$ and T_n, T'_n are iid realizations of a random variable with density p_n. $\square$

Proof of Proposition 2 Since R is n-terminal, $\mathbf{d} = d_n\mathbf{e}_n$. Since $p_{\text{abs}}(t) = \mathbf{d}^\top e^{Rt}\mathbf{e}_1 = d_n\mathbf{e}_n^\top e^{Rt}\mathbf{e}_1$ and $p_n(t) = t_n^{-1}\mathbf{e}_n^\top e^{Rt}\mathbf{e}_1$ are both probability densities of nonnegative random variables, we must have

$$\int_0^\infty d_n\mathbf{e}_n^\top e^{Rt}\mathbf{e}_1\,\mathrm{d}t = 1 = \int_0^\infty t_n^{-1}\mathbf{e}_n^\top e^{Rt}\mathbf{e}_1\,\mathrm{d}t,$$

and therefore $d_n = 1/t_n$. $\square$

Proof of Corollary 1 The result

$$\mathrm{Var}(J) = \mathbb{E}[K(T_{\text{abs}} - T'_{\text{abs}})] \le \mathrm{Var}(I), \tag{43}$$

follows immediately from Theorem 1 and the general result that the correlation function of any WSS process satisfies $|K(t)| \le K(0) = \mathrm{Var}(I)$ for all $t \in \mathbb{R}$ (Pavliotis 2014). Furthermore, if $K(s) = K(0)$ for some time $s \neq 0$, then $\xi(t)$ is periodic with period s. Therefore, since $\xi(t)$ is not constant in time and T_{abs} has a continuous probability distribution, the inequality in (43) must be strict. $\square$

Proof of Corollary 2 Taking the expectation of (42) yields $\mathbb{E}[c_n] = t_n\mathbb{E}[I]$. The result

$$\mathrm{CV}(c_n) := \frac{\mathrm{SD}(c_n)}{\mathbb{E}[c_n]} = \frac{\sqrt{\mathbb{E}[K(T_n - T'_n)]}}{\mathbb{E}[I]} \le \frac{\sqrt{\mathrm{Var}(I)}}{\mathbb{E}[I]}, \tag{44}$$

follows immediately from Theorem 2 and the general result that the correlation function of any WSS process satisfies $|K(t)| \leq K(0) = \mathrm{Var}(I)$ for all $t \in \mathbb{R}$ (Pavliotis 2014). The result that the inequality in (44) must be strict follows from the same argument in the proof of Corollary 1 since T_n has a continuous probability distribution. □

Proposition 4 *If the input $I(t)$ is WSS, then $\mathbb{E}[J] = \mathbb{E}[I]$.*

Proof of Proposition 4 Taking the mean of (41) and using that I is WSS yields

$$\mathbb{E}[c(t)] = \int_{-\infty}^{t} e^{R(t-s)} \mathbf{e}_1 \mathbb{E}[I] \, \mathrm{d}s = -R^{-1} \mathbf{e}_1 \mathbb{E}[I].$$

Since $J = \mathbf{d}^\top c$ and $\mathbf{d}^\top = -\mathbf{1}^\top R$, we obtain

$$\mathbb{E}[J] = \mathbf{d}^\top \mathbb{E}[c] = -\mathbf{d}^\top R^{-1} \mathbf{e}_1 \mathbb{E}[I] = \mathbf{1}^\top R R^{-1} \mathbf{e}_1 \mathbb{E}[I] = \mathbb{E}[I],$$

which completes the proof. □

Proposition 5 *Suppose $R \in \mathbb{R}^{m \times m}$ is invertible, $u, v \in \mathbb{R}^{m \times 1}$, and*

$$p(t) = u^\top e^{Rt} v, \quad t \geq 0,$$

is a probability density of a nonnegative random variable T. Then

$$\mathbb{E}[T] = u^\top R^{-2} v, \quad \mathbb{E}[T^2] = -2u^\top R^{-3} v.$$

Proof of Proposition 5 Integrating by parts yields

$$\begin{aligned} \mathbb{E}[T] = \int_0^\infty t p(t) \, \mathrm{d}t &= \int_0^\infty t u^\top e^{Rt} v \, \mathrm{d}t \\ &= -\int_0^\infty u^\top R^{-1} e^{Rt} v \, \mathrm{d}t = u^\top R^{-2} v. \end{aligned}$$

Similarly, integrating by parts twice yields

$$\begin{aligned} \mathbb{E}[T^2] = \int_0^\infty t^2 p(t) \, \mathrm{d}t &= \int_0^\infty t^2 u^\top e^{Rt} v \, \mathrm{d}t \\ &= -\int_0^\infty 2t u^\top R^{-1} e^{Rt} v \, \mathrm{d}t \\ &= \int_0^\infty 2u^\top R^{-2} e^{Rt} v \, \mathrm{d}t = -2u^\top R^{-3} v, \end{aligned}$$

which completes the proof. □

Proofs for Section 4

Proof of Theorem 3 The survival probability of the absorption time conditioned that $X(0) = 1$ is (Norris 1998)

$$\mathbb{P}_1(T_{\text{abs}}^X > t) = \mathbf{1}^\top e^{Rt}\mathbf{e}_1, \quad t > 0, \tag{45}$$

where $\mathbf{1} \in \mathbb{R}^{m\times 1}$ is a column vector of all ones. Since $d_i = -\sum_{j=1}^m R_{ji} \geq 0$, we have that

$$\mathbf{1}^\top R = -\mathbf{d}^\top.$$

Therefore, taking the derivative of (45) yields

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}_1(T_{\text{abs}}^X > t) = -\mathbf{1}^\top Re^{Rt}\mathbf{e}_1 = \mathbf{d}^\top e^{Rt}\mathbf{e}_1 = p_{\text{abs}}(t),$$

which completes the proof. □

Proof of Theorem 5 Theorem 3 implies that $T_{\text{abs}}^{\widetilde{X}}$ has density

$$\widetilde{p}_{\text{abs}} := \mathbf{d}^\top e^{\widetilde{R}t}\mathbf{e}_n,$$

if we condition that $\widetilde{X}(0) = 1$, where $\mathbf{d}^\top := -\mathbf{1}^\top \widetilde{R}$ where $\mathbf{1} \in \mathbb{R}^{m\times 1}$ is the vector of all 1 s. Further, Proposition 3 implies that $\widetilde{R}$ is n-terminal, and therefore Proposition 2 implies that

$$\widetilde{p}_{\text{abs}}(t) := \mathbf{d}^\top e^{\widetilde{R}t}\mathbf{e}_n = \widetilde{p}_n(t) := \widetilde{d}_n\mathbf{e}_n^\top e^{\widetilde{R}t}\mathbf{e}_1,$$

where $\widetilde{d}_n = \mathbf{d}^\top\mathbf{e}_n$ is the nth element of $\mathbf{d}$. Therefore, $T_{\text{abs}}^{\widetilde{X}}$ has density $\widetilde{p}_n$ if we condition that $\widetilde{X}(0) = 1$.

It thus remains to show that $\widetilde{p}_n = p_n$. If we define the diagonal matrix $\Lambda = \text{diag}(\mathbb{P}_1(A), \mathbb{P}_2(A), \ldots, \mathbb{P}_m(A))$, then (32) yields

$$\widetilde{R} = \Lambda R\Lambda^{-1}.$$

Noting that $\mathbf{e}_n^\top\Lambda^{-1} = \mathbf{e}_n^\top\frac{1}{\mathbb{P}_n(A)} = \mathbf{e}_n^\top$ and $\Lambda\mathbf{e}_1 = \mathbb{P}_1(A)\mathbf{e}_1$, we compute

$$\begin{aligned}
t_n p_n(t) &= \mathbf{e}_n^\top e^{Rt}\mathbf{e}_1 \\
&= \mathbf{e}_n^\top\Lambda^{-1}e^{\widetilde{R}t}\Lambda\mathbf{e}_1 \\
&= \mathbf{e}_n^\top\frac{1}{\mathbb{P}_n(A)}e^{\widetilde{R}t}\mathbb{P}_1(A)\mathbf{e}_1 \\
&= \mathbb{P}_1(A)\mathbf{e}_n^\top e^{\widetilde{R}t}\mathbf{e}_1 = \mathbb{P}_1(A)\widetilde{t}_n\widetilde{p}_n(t).
\end{aligned}$$

Hence, p_n and $\widetilde{p}_n$ are equal up to a scalar multiple, but since $\int_0^\infty p_n(t)\,\mathrm{d}t = \int_0^\infty \widetilde{p}_n(t)\,\mathrm{d}t = 1$, we must have $p_n = \widetilde{p}_n$. □

We now aim to prove Theorem 4. The next lemma proves Theorem 4 in the special case that R is n-terminal.

Lemma 1 *If R is n-terminal, then the last passage time of X to node $n \in \{1, 2, \ldots, m\}$,*

$$T_n^X := \sup\{t \geq 0 : X(t) = n\},$$

has probability density p_n in (7) if we condition that $X(0) = 1$.

Proof of Lemma 1 Note that n-terminal implies that a particle leaves compartment n for the last time by immediately decaying. As such, if a particle is still in a compartment $1, 2, \ldots, m$, then it has not left compartment n for the last time. Mathematically, this means

$$\mathbb{P}_1(T_n^X > t) = \mathbb{P}_1(X(t) \in \{1, 2, \ldots, m\}) = \mathbf{1}^\top e^{Rt}\mathbf{e}_1,$$

where $\mathbf{1}$ is a column vector with all 1 s. Taking the derivative yields

$$f_n(t) := -\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}_1(T_n^X > t) = \mathbf{1}^\top(-R)e^{Rt}\mathbf{e}_1.$$

Since R is n-terminal, $\mathbf{1}^\top R = -d_n\mathbf{e}_n^\top$. Therefore, (7) implies

$$f_n(t) = d_n\mathbf{e}_n^\top e^{Rt}\mathbf{e}_1 = d_n t_n p_n(t). \tag{46}$$

Noting that $\int_0^\infty f_n(t)\,\mathrm{d}t = \int_0^\infty p_n(t)\,\mathrm{d}t = 1$, we deduce from (46) that $t_n = 1/d_n$ and $f_n(t) = p_n(t)$ which completes the proof. $\square$

To prove Theorem 4 in the general case that R may not be n-terminal, we first need the alternative representation for $\widetilde{R}$ in (32).

Lemma 2 *If B_i^R is the event that X moves to state i on its first jump, then $\widetilde{R}$ in (32) satisfies*

$$\widetilde{R}_{ij} = \begin{cases} R_{jj} & i = j, \\ -R_{jj}\mathbb{P}_j(B_i^R|A) & i \neq j,\ j \neq n, \\ R_{in}\mathbb{P}_i(A) & i \neq j,\ j = n. \end{cases} \tag{47}$$

Furthermore, $\widetilde{R}$ is n-terminal.

Proof of Lemma 2 We need to show that when $i \neq j$ and $j \neq n$,

$$-R_{jj}\mathbb{P}_j(B_i^R|A) = R_{ij}\frac{\mathbb{P}_i(A)}{\mathbb{P}_j(A)}.$$

The other cases are trivial; the two definitions match when $i = j$ and noting that $\mathbb{P}_n(A) = 1$ resolves the last case. To start, we use Bayes rule to find

$$-R_{jj}\mathbb{P}_j(B_i^R|A) = -R_{jj}\frac{\mathbb{P}_j(A|B_i^R)\mathbb{P}_j(B_i^R)}{\mathbb{P}_j(A)}.$$

The Markov property implies $\mathbb{P}_j(A|B_i^R) = \mathbb{P}_i(A)$, and therefore

$$-R_{jj}\mathbb{P}_j(B_i^R|A) = -R_{jj}\mathbb{P}_j(B_i^R)\frac{\mathbb{P}_i(A)}{\mathbb{P}_j(A)}.$$

Using that $-R_{jj}\mathbb{P}_j(B_i^R) = R_{ij}$, we find that the representations of $\widetilde{R}$ in (32) and (47) are equivalent.

Using the representation in (47), we can easily show $\widetilde{R}$ is n-terminal. When $j \neq n$, the sum of the elements of column j of $\widetilde{R}$ is

$$\sum_{i=1}^{m} \widetilde{R}_{ij} = R_{jj} - R_{jj} \sum_{i=1,i\neq j}^{m} \mathbb{P}_j(B_i^R|A) = R_{jj} - R_{jj} = 0,$$

since $\sum_{i=1,i\neq j}^{m} \mathbb{P}_j(B_i^R|A) = 1$ because conditioning on A means the particle cannot immediately decay after leaving compartment j and must move into some compartment. When $j = n$, the sum is

$$\sum_{i=1}^{m} \widetilde{R}_{in} = R_{nn} + \sum_{i=1,i\neq n}^{m} R_{in}\mathbb{P}_i(A) \leq R_{nn} + \sum_{i=1,i\neq n}^{m} R_{in} \leq 0.$$

Notice that

$$\sum_{i=1}^{m} \widetilde{R}_{in} = R_{nn} + \sum_{i=1,i\neq n}^{m} R_{in}\mathbb{P}_i(A) = R_{nn} + \sum_{i=1,i\neq n}^{m} R_{in} = 0 \tag{48}$$

if any only if both of the following two conditions hold

$$\mathbb{P}_i(A) = 1 \quad \text{for all } i \text{ such that } R_{in} > 0, \tag{49}$$

$$R_{nn} + \sum_{i=1,i\neq n}^{m} R_{in} = 0. \tag{50}$$

For the sake of contradiction, assume (48) holds and therefore (49)–(50) both hold. Equation (50) implies that R does not permit decay from compartment n, and (49) implies that particles leaving compartment n are guaranteed to arrive at n again. Together, (49)–(50) imply that particles that have arrived at compartment n never decay, but this is not possible given our assumptions on R in section 2.1. As such, (49)–(50) cannot both hold and therefore (48) cannot hold and thus

$$\sum_{i=1}^{m} \widetilde{R}_{in} < 0,$$

proving that $\widetilde{R}$ is n-terminal. □

Proof of Theorem 4 Define the probability densities,

$$f_n(t) := \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(T_n^X \le t \mid X(0) = 1,\ A),$$
$$\widetilde{f}_n(t) := \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(T_n^{\widetilde{X}} \le t \mid \widetilde{X}(0) = 1).$$

To start, we show $f_n = \widetilde{f}_n$.

Let S be the time it takes X leave compartment n for the first time conditioned on X eventually arriving at compartment n. Let U be the time it takes X to leave compartment n after reaching compartment n some time after the first jump conditioned on X eventually arriving at compartment n. Let C be the event that X reaches compartment n sometime after the first jump.

Then, the number of times a particle visits compartment n follows a geometric distribution, and

$$\mathbb{P}_1(T_n^X > t \mid A) = (1 - \mathbb{P}_n(C)) \sum_{\ell=0}^{\infty} \mathbb{P}_n(C)^{\ell}\, \mathbb{P}\left(S + \sum_{m=0}^{\ell} U_m > t\right)$$
$$\mathbb{P}_1\left(T_n^{\widetilde{X}} > t\right) = (1 - \mathbb{P}_n(\widetilde{C})) \sum_{\ell=0}^{\infty} \mathbb{P}_n(\widetilde{C})^{\ell}\, \mathbb{P}\left(\widetilde{S} + \sum_{m=0}^{\ell} \widetilde{U}_m > t\right),$$

where the U_m are iid copies of U, and $\widetilde{C}$, $\widetilde{S}$, and $\widetilde{U}_m$ are defined analogously to C, S, and U_m but $\widetilde{X}$ rather than X. Hence, we can show

$$\mathbb{P}_1(T_n^X > t \mid A) = \mathbb{P}_1(T_n^{\widetilde{X}} > t)$$

by showing

$$\mathbb{P}_n(C) = \mathbb{P}_n(\widetilde{C}), \quad S =_{\mathrm{d}} \widetilde{S}, \quad U =_{\mathrm{d}} \widetilde{U},$$

where $=_{\mathrm{d}}$ denotes equality in distribution.

First, we show $S =_{\mathrm{d}} \widetilde{S}$. We prove this by showing sample paths for S and $\widetilde{S}$ have the same rates for moving from compartment j to i. This immediately follows from looking at the case $i \neq j$ and $j \neq n$ in Lemma 2. We can disregard $j = n$ because S and $\widetilde{S}$ do not depend on where the particle goes after leaving compartment n.

Next, to show $\mathbb{P}_n(C) = \mathbb{P}_n(\widetilde{C})$, observe that

$$\mathbb{P}_n(C) = \sum_{i=1, i\neq n}^{m} \mathbb{P}_n(B_i^R)\mathbb{P}_i(A) = \sum_{i=1, i\neq n}^{m} \frac{R_{in}}{-R_{nn}}\mathbb{P}_i(A).$$

From Lemma 2, we recall $\widetilde{R}_{in} = R_{in}\mathbb{P}_i(A)$ and $R_{nn} = \widetilde{R}_{nn}$. Then,

$$\sum_{i=1, i\neq n}^{m} \frac{R_{in}}{-R_{nn}}\mathbb{P}_i(A) = \sum_{i=1, i\neq n}^{m} \frac{\widetilde{R}_{in}}{-\widetilde{R}_{nn}} = \sum_{i=1, i\neq n}^{m} \mathbb{P}_n(B_i^{\widetilde{R}}). \tag{51}$$

Because $\widetilde{R}$ is n-terminal, a particle leaving compartment n and immediately going to any other compartment would eventually return to compartment n. As such, the right hand side of (51) is $\mathbb{P}_n(\widetilde{C})$ and thus $\mathbb{P}_n(C) = \mathbb{P}_n(\widetilde{C})$.

Lastly, we show $U =_{\mathrm{d}} \widetilde{U}$. To do so, we need to confirm that when $i \neq j$,

$$-R_{jj}\mathbb{P}_j(B_i^R|C) = -\widetilde{R}_{jj}\mathbb{P}_j(B_i^{\widetilde{R}}|\widetilde{C}) \tag{52}$$

When $j \neq n$, conditioning on C is the same as conditioning on A_j and conditioning on $\widetilde{C}$ is unnecessary since n-terminal systems guarantee arrival to compartment n from any other compartment. This means we can repeat the proof for $S = \widetilde{S}$. As such, we just need to examine when $j = n$. Using Bayes' rule and the Markov property yields

$$\begin{aligned} -R_{nn}\mathbb{P}_n(B_i^R|C) = -R_{nn}\frac{\mathbb{P}_n(C \mid B_i^R)\mathbb{P}_n(B_i^R)}{\mathbb{P}_n(C)} &= -R_{nn}\frac{\mathbb{P}_i(A)\mathbb{P}_n(B_i^R)}{\mathbb{P}_n(C)} \\ &= R_{in}\frac{\mathbb{P}_i(A)}{\mathbb{P}_n(C)} = \frac{\widetilde{R}_{in}}{\mathbb{P}_n(\widetilde{C})}. \end{aligned}$$

Since $\widetilde{R}$ is n-terminal,

$$-\widetilde{R}_{nn}\mathbb{P}_n(B_i^{\widetilde{R}}|\widetilde{C}) = -\widetilde{R}_{nn}\frac{\mathbb{P}_n(B_i^{\widetilde{R}} \cap \widetilde{C})}{\mathbb{P}_n(\widetilde{C})} = -\widetilde{R}_{nn}\frac{\mathbb{P}_n(B_i^{\widetilde{R}})}{\mathbb{P}_n(\widetilde{C})} = \frac{\widetilde{R}_{in}}{\mathbb{P}_n(\widetilde{C})}.$$

Hence, (52) is confirmed. As such, $\mathbb{P}(T_n^X > t \mid A) = \mathbb{P}(T_n^{\widetilde{X}} > t)$ and thus $f_n = \widetilde{f}_n$.

Next, we note that $\widetilde{f}_n = \widetilde{p}_n$ follows immediately from Lemma 1 since $\widetilde{R}$ is n-terminal.

Lastly, we have that $\widetilde{p}_n = p_n$ by Theorems 3 and 5. Therefore, we have

$$f_n = \widetilde{f}_n = \widetilde{p}_n = p_n,$$

which completes the proof. □

Proof of Theorem 6 We prove the theorem for the case that $T = T_n$ and $T^* = T_n^*$. The case for T_{abs} and T_{abs}^* is nearly identical. Let $X(t)$ be a path on R, let α_i be the i^{th} time $X(t)$ reaches s, and let β_i be the i^{th} time $X(t)$ leaves s. Similarly, let $X^*(t)$ be a path on R^* (i.e. the network the additional side network), let α_i^* be the i^{th} time $X^*(t)$ enters compartment s from a compartment outside the side network, and let β_i^* be the i^{th} time $X^*(t)$ leaves s for a compartment not in the side network. Let H be a random variable denoting the largest index for β, or equivalently the number of times $X(t)$ hits s. For notational convenience, we also set $\beta_0^* = \beta_0 = 0$, $\alpha_{H+1} = \sup_t\{X(t) = n\}$

and $\alpha^*_{H+1} = \sup_t\{X^*(t) = n\}$. Then we can express

$$T = \sum_{i=1}^{H+1}(\alpha_i - \beta_{i-1}) + \sum_{i=1}^{H}(\beta_i - \alpha_i),$$
$$T^* = \sum_{i=1}^{H+1}(\alpha^*_i - \beta^*_{i-1}) + \sum_{i=1}^{H}(\beta^*_i - \alpha^*_i).$$

Note that the sum of $\alpha_i - \beta_{i-1}$ and the sum of $\alpha^*_i - \beta^*_{i-1}$ have the same distribution, which we call T_{-s}. We define T_s as equal in distribution to $\beta_i - \alpha_i$ and similarly T^*_s as equal in distribution to $\beta^*_i - \alpha^*_i$ and let $\{T_s^{(i)}\}_{i\geq 1}$ and $\{T_s^{*(i)}\}_{i\geq 1}$ be iid copies of T_s and T^*_s, respectively. This gives

$$T = T_{-s} + \sum_{i=1}^{H} T_s^{(i)},$$
$$T^* = T_{-s} + \sum_{i=1}^{H} T_s^{*(i)}.$$

Writing $K = -R_{ss}$ and k_s as the rate from s into the side network, we have

$$T_s = \mathrm{Exp}(K),$$
$$T^*_s = \mathrm{Exp}(K + k_s) + \sum_{j=1}^{G}\left(T_s^{(j)} + \mathrm{Exp}(K + k_s)\right),$$

where G is a geometric variable representing the number of entries from s to the side network,

$$\mathbb{P}(G = i) = \left(\frac{K}{k_s + K}\right)\left(\frac{k_s}{k_s + K}\right)^i, \quad i \in \{0, 1, 2, \dots\},$$

and T^*_s is broken up into the time in the side network, i.e. T_s, and the time spent at compartment s, i.e. $\mathrm{Exp}(K + k_s)$. It is well known that

$$\mathbb{E}[G] = \frac{k_s}{K},$$
$$\mathrm{Var}[G] = \frac{k_s(k_s + K)}{K^2}.$$

Using this, we find

$$
\begin{aligned}
\mathbb{E}[T_s^*] &= \frac{1}{K+k_s} + \frac{k_s}{K}\left(\mathbb{E}[T_s] + \frac{1}{K+k_s}\right) \\
&= \frac{1}{K} + \frac{k_s}{K}\mathbb{E}[T_s] \\
&> \frac{1}{K} = \mathbb{E}[T_s],
\end{aligned}
$$

which implies

$$
\mathbb{E}[T^*|A] - \mathbb{E}[T|A] > 0. \tag{53}
$$

We also determine

$$
\begin{aligned}
\mathrm{Var}(T_s^*) &= \left(\frac{1}{K+k_s}\right)^2 + \frac{k_s}{K}\left(\mathrm{Var}(T_s) + \left(\frac{1}{K+k_s}\right)^2\right) \\
&\quad + \frac{k_s(k_s+K)}{K^2}\left(\mathbb{E}[T_s] + \frac{1}{K+k_s}\right) \\
&= \frac{1}{K^2} + \frac{k_s}{K}\mathrm{Var}(T_s) + \frac{k_s(k_s+K)}{K^2}\mathbb{E}[T_s] \\
&> \frac{1}{K^2} = \mathrm{Var}(T_s).
\end{aligned}
$$

Lastly, note that since $\alpha_1 - \beta_0$ and $\alpha_{H+1} - \beta_H$ are independent of H and T_s^*, we have

$$
\begin{aligned}
\mathrm{Cov}\left(T_{-s}, \sum_{i=1}^{H} T_s^{*(i)}\right) &= \mathrm{Cov}\left(\sum_{i=1}^{H-1}(\alpha_{i+1} - \beta_i), \sum_{i=1}^{H} T_s^{*(i)}\right) \\
&= \mathbb{E}[\alpha_{i+1} - \beta_i]\mathbb{E}[T_s^*]\mathrm{Var}(H) \\
&> \mathbb{E}[\alpha_{i+1} - \beta_i]\mathbb{E}[T_s]\mathrm{Var}(H) = \mathrm{Cov}\left(T_{-s}, \sum_{i=1}^{H} T_s^{(i)}\right).
\end{aligned}
$$

Putting everything together gives

$$
\begin{aligned}
\mathrm{Var}(T^*|A) &= \mathrm{Var}(T_{-s}|A) + \mathrm{Var}\left(\sum_{i=1}^{H} T_s^{*(i)}\right) + \mathrm{Cov}\left(T_{-s}, \sum_{i=1}^{H} T_s^{*(i)}\right) \\
&> \mathrm{Var}(T_{-s}|A) + \mathrm{Var}\left(\sum_{i=1}^{H} T_s^{(i)}\right) + \mathrm{Cov}\left(T_{-s}, \sum_{i=1}^{H} T_s^{(i)}\right),
\end{aligned}
$$

which yields

$$
\mathrm{Var}(T^*|A) - \mathrm{Var}(T|A) > 0,
$$

the desired result for proving Theorem 6. □

Acknowledgements SDL was supported by the National Science Foundation (Grant Nos. CAREER DMS-1944574, DMS-1814832, and DMS-2325258).

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Anderson DF, Mattingly JC (2007) Propagation of fluctuations in biochemical systems, II: nonlinear chains. IET Syst Biol 1(6):313–325

Anderson DF, Mattingly JC, Frederik Nijhout H, Reed MC (2007) Propagation of fluctuations in biochemical systems, I: Linear SSC networks. Bull Math Biol 69(6):1791–1813

Bao J-D, Jia Y (2004) Determination of fission rate by mean last passage time. Phys Rev C 69(2):027602

Bena I (2006) Dichotomous markov noise: exact results for out-of-equilibrium systems. Int J Mod Phys B 20(20):2825–2888

Browning AP, Jenner AL, Baker RE, Maini PK (2023) Smoothing in linear multicompartment biological processes subject to stochastic input. arXiv preprint arXiv:2305.09004

Campbell A (2003) The future of bacteriophage biology. Nat Rev Genet 4(6):471–477

Cannon WB (1932) The wisdom of the body. Norton & Co., New York

Carr EJ, Simpson MJ (2019) New homogenization approaches for stochastic transport through heterogeneous media. J Chem Phys 150(4)

Clark ED, Lawley SD (2023) How drug onset rate and duration of action affect drug forgiveness. J Pharmacokin Pharmacodyn **(in press)**

Clark ED, Lawley SD (2022) Should patients skip late doses of medication? A pharmacokinetic perspective. J Pharmacokinet Pharmacodyn 49(4):429–444

Comtet A, Cornu F, Schehr G (2020) Last-passage time for linear diffusions and application to the emptying time of a box. J Stat Phys 181(5):1565–1602

Counterman ED, Lawley SD (2021) What should patients do if they miss a dose of medication? A theoretical approach. J Pharmacokinet Pharmacodyn 48(6):873–892

Counterman ED, Lawley SD (2022) Designing drug regimens that mitigate nonadherence. Bull Math Biol 84(1):1–36

Félix M-A, Barkoulas M (2015) Pervasive robustness in biological systems. Nat Rev Genet 16(8):483–496

Felmlee MA, Morris ME, Mager DE (2012) Mechanism-based pharmacodynamic modeling. In: Computational toxicology, pp 583–600. Springer

Fermín LJ, Lévy-Véhel J (2017) Variability and singularity arising from poor compliance in a pharmacokinetic model ii: the multi-oral case. J Math Biol 74(4):809–841

Fill JA (2009) The passage time distribution for a birth-and-death chain: strong stationary duality gives a first stochastic proof. J Theor Probab 22(3):543–557

Gibaldi M, Perrier D (1982) Pharmacokinetics, 2nd edn. Marcelly Dekker, New York

Haynes RB, McDonald HP, Garg A, Montague P (2002) Interventions for helping patients to follow prescriptions for medications. Cochrane Database Syste Rev 2

Huang C, Ferrell JE (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci USA 93(19):10078–10083

Kloeden PE, Platen E (1992) Numerical solution of stochastic differential equations, corrected edition. Springer, Berlin

Lévy-Véhel P-E, Lévy-Véhel J (2013) Variability and singularity arising from poor compliance in a pharmacokinetic model i: the multi-iv case. J Pharmacokinet Pharmacodyn 40(1):15–39

Li J, Nekka F (2007) A pharmacokinetic formalism explicitly integrating the patient drug compliance. J Pharmacokinet Pharmacodyn 34(1):115–139

Li J, Nekka F (2009) A probabilistic approach for the evaluation of pharmacological effect induced by patient irregular drug intake. J Pharmacokinet Pharmacodyn 36(3):221–238

Liu L, Liu X, Yao DD (2004) Analysis and optimization of a multistage inventory-queue system. Manag Sci 50(3):365–380

Louten J (2016) Virus structure and classification. In: Essential human virology, pp 19

McAllister N, Lawley S (2022) A pharmacokinetic and pharmacodynamic analysis of drug forgiveness. J Pharmacokinet Pharmacodyn 49:1–17

Meyer B, Chevalier C, Voituriez R, Bénichou O (2011) Universality classes of first-passage-time distribution in confined media. Phys Rev E 83(5):051116

Michael J (2007) Conceptual assessment in the biological sciences: a national science foundation-sponsored workshop

Morgan DO (2007) The cell cycle: principles of control. New Science Press, Beijing

Norris JR (1998) Markov chains, statistical & probabilistic mathematics. Cambridge University Press, Cambridge

Osterberg L, Blaschke T (2005) Adherence to medication. N Engl J Med 353(5):487–497

Osterberg LG, Urquhart J, Blaschke TF (2010) Understanding forgiveness: minding and mining the gaps between pharmacokinetics and therapeutics. Clin Pharmacol Therapeutics 88(4):457–459

Pavliotis GA (2014) Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations, vol 60. Springer, Berlin

Rosenbaum SE (2016) Basic pharmacokinetics and pharmacodynamics: an integrated textbook and computer simulations. Wiley, Hoboken

Sabaté E, Sabaté E et al (2003) Adherence to long-term therapies: evidence for action. World Health Organization, Geneva

Simpson EH (1949) Measurement of diversity. Nature 163(4148):688–688