Supplementary Materials
Temporal heterogeneity shapes diffusion dynamics in complex networks

Cheng Luo（罗成）[1,2], Renaud Lambiotte[3], Peng Ji*（纪鹏）[1,2,4]

[1]Interdisciplinary Research Centre for Complex Systems,
Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China

[2]Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Ministry of Education, Shanghai 200433, China

[3]Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, United Kingdom

[4]MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China

April 9, 2026

Corresponding author: pengji@fudan.edu.cn

Contents

1 Diffusion equation

The movement of walkers on a network can be macroscopically regarded as diffusion. In order to have a deeper understanding of the mechanism of the diffusion, we cannot ignore the details of the movement of a single random walker between different nodes in the network and only focus on the connections between network nodes. We use two matrix functions to describe the dynamics.

1.1 Waiting time matrix

We use the matrix $\boldsymbol{\rho}$ to describe the distribution of the time spent by a walker on a node before a jump. More precisely, let a walker arrive on a node, say i, the waiting time distribution determines the time before the next jump. Note that it is assumed that the waiting time depends only on the node where the node stays, i, and not on the destination. We denote the waiting time distribution function as $\rho_i(t)$ and as the i-th element on the diagonal matrix $\boldsymbol{\rho}(t)$. We denote diagonal matrix $\boldsymbol{\rho}(t)$ as Waiting Time Matrix, which can be written as:

$$\begin{bmatrix} \rho_1(t) & 0 & \cdots & 0 \\ 0 & \rho_2(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_n(t) \end{bmatrix}.$$

In many cases, matrix $\boldsymbol{\rho}(t)$ can be written as the identity matrix times a function, so that the waiting time distributions of different nodes are the same. In the simplest case where the succession of steps is a Poisson process, the waiting time distribution takes an exponential form. In general, and as observed in a variety of empirical systems, each node is characterized by a different, non-exponential distribution. For example, in neuroscience, if we study the transmission of nerve signals between neurons, the time that the signal stays on different neurons depends on the structure and function of nerve cells.

1.2 Transition matrix

We use $\boldsymbol{P}_{ij}$ to represent the probability that the random walker chooses to move from i to j. We define the i,j-th component of the Transition Matrix $\boldsymbol{P}$ as $\boldsymbol{P}_{ij}$, which can be written as:

$$\begin{bmatrix} 0 & \boldsymbol{P}_{12} & \cdots & \boldsymbol{P}_{1n} \\ \boldsymbol{P}_{21} & 0 & \cdots & \boldsymbol{P}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{P}_{n1} & \boldsymbol{P}_{n2} & \cdots & 0 \end{bmatrix}.$$

1.3 State matrix of the diffusion system

We assume that a random walker starts from node i at time 0, and the probability that it appears at j at time t is $\boldsymbol{X}_{ij}(t)$. We define the State Matrix of the diffusion system $\boldsymbol{X}(t)$ as:

$$\begin{bmatrix} \boldsymbol{X}_{11}(t) & \boldsymbol{X}_{12}(t) & \cdots & \boldsymbol{X}_{1n}(t) \\ \boldsymbol{X}_{21}(t) & \boldsymbol{X}_{22}(t) & \cdots & \boldsymbol{X}_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}_{n1}(t) & \boldsymbol{X}_{n2}(t) & \cdots & \boldsymbol{X}_{nn}(t) \end{bmatrix}.$$

It is determined by $\boldsymbol{\rho}(t)$ and $\boldsymbol{P}$.

Now we want to derive an expression for $\boldsymbol{X}_{ij}(t)$. We divide the state of a random walker into two cases, the first case is that the random walker does not move before t, we note this event as A. In this case, we can write

the conditional probability: $P(i \to j|A) = \delta_{ij} \int_t^\infty \rho_i(\tau)d\tau$, $\delta_{ij} = 1$ when $i = j$, $\delta_{ij} = 0$ when $i \neq j$. The second case is that the first transition of the random walker occurred at time τ before t, we denote this event as B. The random walker transfers to another node k other than i for the first time, and the probability of appearing at node j after the first transition is $\int_0^t \rho_i(\tau)\boldsymbol{P}_{ik}\boldsymbol{X}_{kj}(t-\tau)d\tau$. In this case, we can write the conditional probability: $P(i \to j|B) = \int_0^t \sum_{k\neq i} \rho_i(\tau)\boldsymbol{P}_{ik}\boldsymbol{X}_{kj}(t-\tau)d\tau$. Finally, we can write the hole probability that a random walker starts from i and be found at j after t:

$$\begin{aligned}\boldsymbol{X}_{ij}(t) &= P(i \to j|A) + P(i \to j|B) \\ &= \delta_{ij} \int_t^\infty \rho_i(\tau)d\tau + \int_0^t \sum_{k\neq i} \rho_i(\tau)\boldsymbol{P}_{ik}\boldsymbol{X}_{kj}(t-\tau)d\tau.\end{aligned} \tag{1}$$

We can write this equation in matrix:

$$\boldsymbol{X}(t) = I \int_t^\infty \boldsymbol{\rho}(\tau)d\tau + \int_0^t \boldsymbol{\rho}(\tau)\boldsymbol{P}\boldsymbol{X}(t-\tau)d\tau. \tag{2}$$

What we are most curious about this model is how the speed of diffusion on the network is affected by the complex mechanism of the network, we will discuss this in the next section.

1.4 Simplified case of exponential waiting time distributions

Before that, let's first discuss the case where the waiting time distribution is exponential. In this case, the waiting time is memoryless, and the diffusion process will be a Markov process. the equation (2) can be written as:

$$\boldsymbol{X}(t) = \int_t^\infty \begin{bmatrix} a_1e^{-a_1\tau} & 0 & \cdots & 0 \\ 0 & a_2e^{-a_2\tau} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_ne^{-a_n\tau} \end{bmatrix} d\tau + \int_0^t \begin{bmatrix} a_1e^{-a_1\tau} & 0 & \cdots & 0 \\ 0 & a_2e^{-a_2\tau} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_ne^{-a_n\tau} \end{bmatrix} \boldsymbol{P}\boldsymbol{X}(t-\tau)d\tau. \tag{3}$$

Taking out the term containing t from the integral term, we get:

$$\boldsymbol{X}(t) = \begin{bmatrix} e^{-a_1t} & 0 & \cdots & 0 \\ 0 & e^{-a_2t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{-a_nt} \end{bmatrix} (\boldsymbol{I} + \int_0^t \begin{bmatrix} a_1e^{-a_1t+a_1\tau} & 0 & \cdots & 0 \\ 0 & a_2e^{-a_2t+a_2\tau} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_ne^{-a_nt+a_n\tau} \end{bmatrix} \boldsymbol{P}X(\tau)d\tau). \tag{4}$$

Multiplying both sides of the equation by a diagonal matrix gives:

$$\begin{bmatrix} e^{a_1t} & 0 & \cdots & 0 \\ 0 & e^{a_2t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{a_nt} \end{bmatrix} \boldsymbol{X}(t) = \boldsymbol{I} + \int_0^t \begin{bmatrix} a_1e^{-a_1t+a_1\tau} & 0 & \cdots & 0 \\ 0 & a_2e^{-a_2t+a_2\tau} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_ne^{-a_nt+a_n\tau} \end{bmatrix} \boldsymbol{P}\boldsymbol{X}(\tau)d\tau. \tag{5}$$

Taking the derivative of both sides of the equation, we can get:

$$\begin{aligned}\begin{bmatrix} e^{a_1 t} & 0 & \cdots & 0 \\ 0 & e^{a_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{a_n t} \end{bmatrix} \frac{d}{dt}\boldsymbol{X}(t) + \begin{bmatrix} e^{a_1 t} & 0 & \cdots & 0 \\ 0 & e^{a_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{a_n t} \end{bmatrix} \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix} \boldsymbol{X}(t) \\ = \begin{bmatrix} e^{a_1 t} & 0 & \cdots & 0 \\ 0 & e^{a_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{a_n t} \end{bmatrix} \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix} \boldsymbol{P}\boldsymbol{X}(t). \end{aligned} \tag{6}$$

Simplify the above equation, and replace the parameter a_i with $\frac{1}{\mu_i}$(μ_i is the mean of $\rho_i(t) = a_i \exp(-a_i t)$), we can get:

$$\frac{d}{dt}\boldsymbol{X}(t) = \begin{bmatrix} \frac{1}{\mu_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\mu_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\mu_n} \end{bmatrix} \boldsymbol{L}\boldsymbol{X}(t), \tag{7}$$

where $\boldsymbol{L} = \boldsymbol{P} - \boldsymbol{I}$ is the Laplacian of the random walk. We find that the diffusion equation can be simplified to an ordinary differential equation. This shows that the diffusion equation based on the Markov process is a special case of equation (2).

2 Mixing Time

2.1 Steady state

When we discuss the speed of diffusion, we essentially assume that the solution to this equation will converge to a stable state as time goes to infinity. The steady-state solution $\boldsymbol{X}(\infty)$ for the semi-Markov process described by the integral equation can be derived as follows. We apply the Laplace transform to both sides. Although because $\boldsymbol{X}(t)$ does not converge to 0, we are formally performing a Laplace transform. The Laplace transform of a function $f(t)$ is defined as

$$\hat{f}(s) = \int_0^\infty e^{-st} f(t)\, dt.$$

Although because $\boldsymbol{X}(t)$ does not converge to $\mathbf{0}$, we are formally performing a Laplace transform. Taking the Laplace transform of both sides gives

$$\hat{\boldsymbol{X}}(s) = I \int_0^\infty e^{-st} \left(\int_t^\infty \boldsymbol{\rho}(\tau)\, d\tau \right) dt + \int_0^\infty e^{-st} \left(\int_0^t \boldsymbol{\rho}(\tau) \boldsymbol{P} \boldsymbol{X}(t-\tau)\, d\tau \right) dt.$$

The first term simplifies by changing the order of integration:

$$\int_0^\infty e^{-st} \left(\int_t^\infty \boldsymbol{\rho}(\tau)\, d\tau \right) dt = \int_0^\infty \boldsymbol{\rho}(\tau) \left(\int_0^\tau e^{-st} dt \right) d\tau = \int_0^\infty \boldsymbol{\rho}(\tau) \frac{1 - e^{-s\tau}}{s} d\tau = \frac{1 - \hat{\boldsymbol{\rho}}(s)}{s},$$

where $\hat{\boldsymbol{\rho}}(s) = \int_0^\infty e^{-s\tau} \boldsymbol{\rho}(\tau)\, d\tau$ is the Laplace transform of $\rho(t)$. Since ρ is diagonal, so is $\hat{\boldsymbol{\rho}}(s)$. The second term is a convolution, so its Laplace transform becomes

$$\hat{\boldsymbol{\rho}}(s) \cdot \boldsymbol{P} \cdot \hat{\boldsymbol{X}}(s).$$

Thus, the transformed equation becomes

$$\hat{\boldsymbol{X}}(s) = \frac{1 - \hat{\boldsymbol{\rho}}(s)}{s} \cdot \boldsymbol{I} + \hat{\boldsymbol{\rho}}(s) \cdot \boldsymbol{P} \cdot \hat{\boldsymbol{X}}(s).$$

We rearrange this to isolate $\hat{\boldsymbol{X}}(s)$:

$$(\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}) \hat{\boldsymbol{X}}(s) = \frac{1 - \hat{\boldsymbol{\rho}}(s)}{s} \cdot \boldsymbol{I},$$

so

$$\hat{\boldsymbol{X}}(s) = \frac{1 - \hat{\boldsymbol{\rho}}(s)}{s} \cdot (\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P})^{-1}.$$

To find the long-time limit $\boldsymbol{X}(\infty) = \lim_{t\to\infty} \boldsymbol{X}(t)$, we recall that in Laplace space this corresponds to

$$\boldsymbol{X}(\infty) = \lim_{s \to 0} s \hat{\boldsymbol{X}}(s).$$

Substituting the expression for $\hat{\boldsymbol{X}}(s)$, we obtain

$$\boldsymbol{X}(\infty) = \lim_{s \to 0} (1 - \hat{\boldsymbol{\rho}}(s))(\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P})^{-1}.$$

To evaluate the stationary distribution matrix $\boldsymbol{X}(\infty)$, we consider the Laplace-domain expression

$$\hat{\boldsymbol{X}}(s) = [\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}]^{-1} \cdot (\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)).$$

Assuming that each waiting time density $\rho_i(t)$ has a finite first moment $m_i = \int_0^\infty t \rho_i(t)\, dt$, the Laplace transform can be expanded as

$$\hat{\boldsymbol{\rho}}(s) = \boldsymbol{I} - s\boldsymbol{D} + o(s), \quad \text{as } s \to 0,$$

with $\boldsymbol{D} = \mathrm{diag}(m_1, \ldots, m_n)$. Substituting this into the expression for $[\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}]^{-1}$, we obtain

$$[\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}]^{-1} = [\boldsymbol{I} - (\boldsymbol{I} - s\boldsymbol{D} + o(s))\boldsymbol{P}]^{-1} = [\boldsymbol{I} - \boldsymbol{P} + s\boldsymbol{D}\boldsymbol{P} + o(s)]^{-1}.$$

Since $\boldsymbol{P}$ is a stochastic matrix, $\boldsymbol{I} - \boldsymbol{P}$ is singular with rank $n-1$, and the left null-space is spanned by the stationary distribution π satisfying $\pi\boldsymbol{P} = \pi$, $\pi\mathbf{1} = 1$. Under the theory of singular matrix perturbation, it is known that as $s \to 0$, the matrix $[\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}]^{-1}$ admits the asymptotic form

$$[\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}]^{-1} \sim \frac{1}{s\pi D\mathbf{1}}\,\mathbf{1}\pi, \quad \text{as } s \to 0,$$

which is a rank-one approximation projecting onto the steady-state subspace. Next, we expand the right factor:

$$\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s) = s\boldsymbol{D} + o(s).$$

Multiplying with $[\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}]^{-1}$, we find the limiting behavior of $s\hat{\boldsymbol{X}}(s)$:

$$s\hat{\boldsymbol{X}}(s) = [\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}]^{-1} \cdot (s\boldsymbol{D} + o(s)) \sim \frac{1}{s\pi D\mathbf{1}}\mathbf{1}\pi(s\boldsymbol{D} + o(s)) = \frac{1}{\pi D\mathbf{1}}\mathbf{1}\pi D + o(s).$$

Thus, taking the limit $s \to 0$, which corresponds to the long-time behavior $t \to \infty$, we obtain

$$\boldsymbol{X}(\infty) = \lim_{t\to\infty} \boldsymbol{X}(t) = \lim_{s\to 0} s\hat{\boldsymbol{X}}(s) = \frac{1}{\pi D\mathbf{1}}\mathbf{1}\pi D.$$

This result shows that $\boldsymbol{X}(\infty)$ is a rank-one matrix where each row is identical and given by the normalized vector $\frac{1}{\pi D\mathbf{1}}\pi D$. This vector reflects the stationary distribution of the semi-Markov process, which incorporates both the topological structure of the underlying Markov chain (via π) and the heterogeneity in mean waiting times (via D). In particular, this distribution differs from the usual Markovian case unless all m_i are equal. We have:

$$\boldsymbol{X}(\infty) = \begin{bmatrix} \frac{\pi_1 m_1}{\sum_k \pi_k m_k} & \frac{\pi_2 m_2}{\sum_k \pi_k m_k} & \cdots & \frac{\pi_n m_n}{\sum_k \pi_k m_k} \\ \frac{\pi_1 m_1}{\sum_k \pi_k m_k} & \frac{\pi_2 m_2}{\sum_k \pi_k m_k} & \cdots & \frac{\pi_n m_n}{\sum_k \pi_k m_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\pi_1 m_1}{\sum_k \pi_k m_k} & \frac{\pi_2 m_2}{\sum_k \pi_k m_k} & \cdots & \frac{\pi_n m_n}{\sum_k \pi_k m_k} \end{bmatrix}.$$

This result aligns with semi-Markov process theory and ensures consistency between the integral equation and probabilistic interpretation. If a node's waiting time is infinite, the diffusion steady state occurs when all walkers eventually enter that node and do not leave thereafter. Conversely, when the waiting time is zero, the probability of walkers appearing at that node is zero.

2.2 Exponential decay

When the entire system reaches stability, it essentially means that each component function of the matrix $\boldsymbol{X}(t)$ converges to a certain value. That is, each component function of $\boldsymbol{Y}(t) = \boldsymbol{X}(t) - \boldsymbol{X}(\infty)$ converges to 0. We need to come up with a way to compare how quickly functions converge to 0. A more intuitive idea is to perform logarithmic linear regression on the function, and use the slope to represent the speed of convergence. But this is meaningless. We cannot calculate an analytical solution to the equation, so we cannot understand the impact of the diffusion mechanism of the network on the diffusion speed. But this idea reminds us that we may be able to equate the function with the exponential function, and then use the coefficients of the equivalent exponential function to define the convergence speed of the function. We decompose the function $f(t)$ into exponential tail and heavy tail terms:

$$f(t) = f_0(t)e^{-t/\tau},$$

where $f_0(t)$ satisfies:$\int_0^\infty e^{ut} f_0(t)dt = \infty$ for all $u > 0$. We use the value of τ to define the speed at which $f(t)$ converges to 0. We perform Laplace transformation on $f(t)$ to get $f(s)$, and then calculate the pole s_0 of $f(s)$. We discover that $\tau = \frac{-1}{s_0}$. For n^2 elements of the $\boldsymbol{Y}(t) = \boldsymbol{X}(t) - \boldsymbol{X}(\infty)$, we can get up to n^2 different τ_{ij}'s:

$$\tau_{ij} = \sup\{a| \int_0^\infty \exp(t/a)(X_{ij}(t) - X_{ij}(\infty))dt\}. \tag{8}$$

We want to find the largest τ_{ij}, which is the speed of the element with the slowest convergence, to represent the convergence speed of the entire system. This requires a systematic analysis of the convergence of $\boldsymbol{Y}(t)$.

2.3 Remove the influence of static items

Based on Eq. (2), we can get:

$$(\boldsymbol{Y}(t) + \boldsymbol{X}(\infty)) = I \int_t^\infty \boldsymbol{\rho}(\tau)d\tau + \int_0^t \boldsymbol{\rho}(\tau)\boldsymbol{P}(\boldsymbol{Y}(t-\tau) + \boldsymbol{X}(\infty))d\tau. \tag{9}$$

We perform Laplace transformation on both sides of Eq. (9) to get:

$$(\hat{\boldsymbol{Y}}(s) - s^{-1}\boldsymbol{X}(\infty)) = \frac{1}{s}(\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)) + \hat{\boldsymbol{\rho}}(t)\boldsymbol{P}(\hat{\boldsymbol{Y}}(s) - s^{-1}\boldsymbol{X}(\infty)). \tag{10}$$

Thus we can get the expression for $\boldsymbol{Y}(s)$:

$$\hat{\boldsymbol{Y}}(s) = (I - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}dt)^{-1}\frac{1}{s}(\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)) - \boldsymbol{X}(\infty)\frac{1}{s}. \tag{11}$$

Here we need a statement. Although the choice of $\rho_i(t)$ is random, if it does not have an exponential tail, such as $\rho_i(t) = \frac{1}{(t+1)^2}$, then it is meaningless to solve the exponential tail of the matrix $\boldsymbol{Y}(t)$ through Laplace transform. In the subsequent discussion, this corresponds to the case where the exponential tail takes an infinite value.

2.4 Argument for the pole at 0

We already know that $\boldsymbol{X}(t)$ gradually converges to $\boldsymbol{X}(\infty)$, that is, $\boldsymbol{Y}(t)$ gradually converges to the $\mathbf{0}$ matrix. This shows that when calculating the Laplace transform for any element of $\boldsymbol{Y}(t)$, there cannot be a pole when $s > 0$. We can calculate $\hat{\boldsymbol{Y}}(s)$ in a simple situation: $\boldsymbol{\rho}(t) = e^{-t}I$. We put the above conditions into Eq. (11) to get:

$$\begin{aligned}
\hat{\boldsymbol{Y}}(s) &= (I - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}dt)^{-1}\frac{1}{s}(\boldsymbol{I} - \frac{1}{s+1}I) - \boldsymbol{P}^\infty\frac{1}{s} \\
&= \left(I - \frac{1}{s+1}\boldsymbol{P}\right)^{-1}\frac{1}{s+1} - \boldsymbol{P}^\infty\frac{1}{s} \\
&= \left(I - \frac{1}{s+1}\boldsymbol{F}^{-1}\Sigma\boldsymbol{F}\right)^{-1}\frac{1}{s+1} - (\boldsymbol{F}^{-1}\Sigma\boldsymbol{F})^\infty\frac{1}{s} \\
&= \boldsymbol{F}^{-1}\left[\left(I - \frac{1}{s+1}\Sigma\right)^{-1}\frac{1}{s+1} - \Sigma^\infty\frac{1}{s}\right]\boldsymbol{F},
\end{aligned} \tag{12}$$

where $\Sigma = diag\{\lambda_0(P) = 1, \lambda_1(\boldsymbol{P}), \cdot, \lambda_{n-1}(P)\}$ is the matrix after diagonalization of $\boldsymbol{P}$, and $\boldsymbol{F}$ is the matrix representing the linear transformation that diagonalizes $\boldsymbol{P}$, composed of the eigenvectors of $\boldsymbol{P}$. We can get:

$$\hat{\boldsymbol{Y}}(s) = \boldsymbol{F}^{-1}\left(\begin{bmatrix} \frac{s+1}{s+1-\lambda_0} & 0 & \cdots & 0 \\ 0 & \frac{s+1}{s+1-\lambda_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{s+1}{s+1-\lambda_{n-1}} \end{bmatrix}\frac{1}{s+1} + \begin{bmatrix} \lambda_0^\infty & 0 & \cdots & 0 \\ 0 & \lambda_1^\infty & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1}^\infty \end{bmatrix}\frac{1}{s}\right)\boldsymbol{F}$$

$$= \boldsymbol{F}^{-1}\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{s+1}{s+1-\lambda_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{s+1}{s+1-\lambda_{n-1}} \end{bmatrix}\boldsymbol{F}. \tag{13}$$

It can be seen from Eq. (13) that all poles of $\hat{\boldsymbol{Y}}(s)$ are less than 0. This shows that after subtracting the stable state, each element of $\hat{\boldsymbol{Y}}(s)$ has convergence with an exponential tail. So we can find the maximum pole, which is $\lambda_1 - 1$, and get the convergence speed of the entire system $\frac{1}{1-\lambda_1}$.

Now we are trying to apply this to the model in Eq. (11). We have:

$$\hat{\boldsymbol{Y}}(s) = (\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}dt)^{-1}\frac{1}{s}(\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)). - \boldsymbol{X}(\infty)\frac{1}{s}.$$

Expanding $\hat{\boldsymbol{\rho}}(s)$ at $s = 0$ yields

$$\hat{\boldsymbol{\rho}}(s) = \boldsymbol{I} - s\boldsymbol{D} + O(s^2), \quad \boldsymbol{D} = \mathrm{diag}(m_1, \ldots, m_n), \quad m_i = \int_0^\infty t\rho_i(t)dt.$$

Hence,

$$\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)P = (\boldsymbol{I} - \boldsymbol{P}) + sDP + O(s^2),$$

and using perturbation theory for singular matrices, we obtain

$$(\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)P)^{-1} = \frac{1}{sc}\mathbf{1}\pi + O(1), \quad \text{where} \quad c = \sum_i \pi_i m_i,$$

with π the stationary distribution of $\boldsymbol{P}$ and $\mathbf{1}$ the all-ones column vector. Combining these, the singular behavior near $s = 0$ becomes

$$\hat{\boldsymbol{Y}}(s) = \frac{1}{s}\left[\frac{1}{c}\mathbf{1}(\pi D) - \mathbf{1}\pi\right] + O(1) = \frac{1}{s}\mathbf{1}\left(\frac{\pi D}{c} - \pi\right) + O(1).$$

This shows that $\hat{\boldsymbol{Y}}(s)$ generally has a simple pole at $s = 0$ unless $\pi D = c\pi$, which occurs only if all mean waiting times m_i are identical.

While the function $\hat{\boldsymbol{Y}}(s)$ generally possesses a simple pole at $s = 0$, this singularity does not provide meaningful information about the convergence speed of $\boldsymbol{X}(t)$ toward its stationary limit $\boldsymbol{X}(\infty)$. There are several reasons for this conclusion. First, the pole at $s = 0$ merely reflects the fact that the process converges in the long-time limit, i.e., that $\boldsymbol{Y}(t) = \boldsymbol{X}(t) - \boldsymbol{X}(\infty)$ vanishes as $t \to \infty$. This singularity ensures the existence of a nonzero total mass in the time domain but does not indicate the rate of decay. In contrast, the convergence rate is determined by how rapidly $\boldsymbol{Y}(t)$ approaches zero, not simply by whether it does. Second, the presence of the singularity at $s = 0$ is generic in the semi-Markov setting. Unless all waiting-time distributions have equal means, the asymptotic expansion of $\hat{\boldsymbol{Y}}(s)$ will include a $1/s$ term. This ubiquity renders the $s = 0$ pole uninformative for comparing different systems or quantifying decay speeds. Third, the $s = 0$ singularity

corresponds to asymptotic behavior at infinite time but not to exponential decay characteristics. In the Laplace domain, exponential convergence is linked to the location of singularities in the left-half complex plane away from the origin. Specifically, the dominant singularity with the largest real part (excluding $s = 0$) determines the leading-order decay rate of $\boldsymbol{Y}(t)$ as $t \to \infty$. For these reasons, although the singularity at $s = 0$ is mathematically valid and essential for capturing steady-state behavior, it lacks analytical value in characterizing convergence speed.

Therefore, to quantify convergence, we must identify the dominant singularity of $\hat{\boldsymbol{Y}}(s)$ in the complex s-plane *excluding* $s = 0$. Specifically, the rightmost pole of $\hat{\boldsymbol{Y}}(s)$ in the left half-plane (i.e., with maximal real part) governs the asymptotic decay of $\boldsymbol{Y}(t)$ via the inverse Laplace transform:

$$\boldsymbol{Y}(t) \sim \text{constant} \cdot e^{s_{\text{mix}} t}, \quad \text{with } s_{\text{mix}} = \sup\{Re(\lambda) : \lambda \in S(\hat{\boldsymbol{Y}}(s)) \setminus \{0\}\},$$

where $S(\hat{\boldsymbol{Y}}(s))$ denotes the set of poles. The value s_{mix}, often referred to as the *spectral gap*, represents the true decay rate. In the following analysis, we focus on locating s_{mix} and its dependence on the waiting-time statistics and network topology.

2.5 Definition of the mixing time

There is a simple calculation example showing the relationship between poles and convergence speed. We consider exponential diffusion on a network of three nodes connected to each other and this indicates:

$$\hat{\rho(t)} = \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{-t} \end{bmatrix}, P = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

By diagonalizing P, we can get $P = F^{-1}\Sigma F$, where Σ is a diagonal matrix $diag\{1, -1/2, -1/2\}$. We can thus obtain a the Laplace transform of state matrix:

$$\begin{aligned} \hat{\boldsymbol{X}}(s) &= F^{-1}\left(I - \frac{1}{s+1}\Sigma\right)^{-1} \frac{1}{s+1} F \\ &= F^{-1} \begin{bmatrix} 1/s & 0 & 0 \\ 0 & \frac{1}{s+3/2} & 0 \\ 0 & 0 & \frac{1}{s+3/2} \end{bmatrix} F. \end{aligned} \tag{14}$$

As can be seen from Eq. (14), all poles of $\hat{\boldsymbol{X}}(s)$ are 0, -3/2. From the discussion before, we know that all the poles of $\hat{\boldsymbol{Y}}(s) = \int_0^\infty \exp(-st)(\boldsymbol{X}(t) - \boldsymbol{X}(\infty))\mathrm{d}t$ are -3/2. We perform the inverse Laplace transform on $\hat{\boldsymbol{X}}(s), \hat{\boldsymbol{Y}}(s)$ to get:

$$\boldsymbol{X}(t) = F^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-3t/2} & 0 \\ 0 & 0 & e^{-3t/2} \end{bmatrix} F,$$

$$\boldsymbol{Y}(t) = F^{-1} \begin{bmatrix} 0 & 0 & 0 \\ 0 & e^{-3t/2} & 0 \\ 0 & 0 & e^{-3t/2} \end{bmatrix} F.$$

After the above analysis, we find that the poles of $\hat{\boldsymbol{Y}}(s)$ coincides with the negative poles of $\hat{\boldsymbol{X}}(s)$. But in fact, the convergence domain of $\hat{\boldsymbol{X}}(s)$ is positive, so in order to discuss the extreme point of negative numbers,

we must remove the stationary term of $\boldsymbol{X}(t)$, that is, to study $\hat{\boldsymbol{Y}}(s)$. we consider $\rho_i(t) = \rho_i^0(t)\exp(-t/\tau_{\text{tail}}^i)$, where $\rho_i^0(t)$ is a fat-tailed distribution, i.e. whose tail decreases slower than an exponential for large values of t, and where $\exp(-t/\tau_{\text{tail}}^i)$ ensures an exponential decay for the asymptotics of the distribution. We define that τ_{tail}^i is the exponential tail of $\rho_i(t)$. Observe the first term of Eq. (11). If we want to determine the largest pole whose real part is negative, we need to find the poles of $(\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}dt)^{-1}$ and $(\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s))$. The pole of the second term is obviously $\{\tau_{tail}^i\}$. We define the first term as the inverse matrix of $\hat{\boldsymbol{Q}}(s) = \boldsymbol{I} - \int_0^\infty e^{-st}\boldsymbol{\rho}(t)\boldsymbol{P}dt$. We now analyze the poles of each component of $[\hat{\boldsymbol{Q}}(s)]^{-1}$ using the following theorem.

Theorem 1. *s_0 is a pole of an element of $\hat{\boldsymbol{Q}}^{-1}(s)$ if and only if s_0 is a zero of $|\hat{\boldsymbol{Q}}(s)|$.*

Proof. 1.Suppose s_0 is a pole of an element of $(\hat{\boldsymbol{Q}}(s))^{-1}$, if $|\hat{\boldsymbol{Q}}(s_0)| \neq 0$, $\hat{\boldsymbol{Q}}(s_0)$is an invertible matrix, which means that $(\hat{\boldsymbol{Q}}(s_0))^{-1}$ is bounded, which against the assumption s_0 is a pole of an element of $(\hat{\boldsymbol{Q}}(s))^{-1}$, so $|\hat{\boldsymbol{Q}}(s_0)| = 0$.

2.Suppose $|\hat{\boldsymbol{Q}}(s_0)| = 0$, if $\lim_{s\to s_0}(\hat{\boldsymbol{Q}}(s))_{ij}^{-1}$ exists for all i, j; we have $\lim_{s\to s_0}[(\hat{\boldsymbol{Q}}(s))^{-1} \cdot \hat{\boldsymbol{Q}}(s)] = I$ and $\lim_{s\to s_0}|(\hat{\boldsymbol{Q}}(s))^{-1} \cdot \hat{\boldsymbol{Q}}(s)| = 1$. Due to the continuity of the determinant function $|\cdot|$, we have $|\hat{\boldsymbol{Q}}(s_0)| \cdot |\lim_{s\to s_0}(\hat{\boldsymbol{Q}}(s))^{-1}| = 1$, which against the assumption $|\hat{\boldsymbol{Q}}(s_0)| = 0$. □

Using the above theorem, we transform the problem of calculating the poles of each element of $\hat{\boldsymbol{X}}(s)$ into calculating the zero points of $|\hat{\boldsymbol{Q}}(s)|$. When we determine that a certain point s_0 is the zero point of $|\hat{\boldsymbol{Q}}(s)|$, we bring in $\hat{\boldsymbol{X}}(s)$, get the position i, j of the pole element $\hat{\boldsymbol{X}}_{ij}(s)$, and compare it with the pole of $\rho_i(s)$, the larger value corresponds to the convergence speed of $\hat{\boldsymbol{X}}_{ij}(t)$. Excluding the few cases where $\hat{\boldsymbol{\rho}}(t)P$ can be diagonalized, when s is a zero point of $|\hat{\boldsymbol{Q}}(s)|$, it is also the pole of element of $\hat{\boldsymbol{X}}(s)^{-1}$, so the pole of element of $\hat{\boldsymbol{X}}(s)$, for example $\hat{\boldsymbol{X}}_{ij}(s)$, should be smaller than the pole of $\rho_i(s)$. We note the pole of $\rho_i(s)$ as $\frac{-1}{\tau_{tail}^i}$, where τ_{tail}^i denotes the exponential tail of $\boldsymbol{\rho}_i(t)$. Therefore, we can establish the calculation interval as $s \in \left(\max(\frac{-1}{\tau_{tail}^i}), 0\right)$. We define $s_{\text{mix}} = Re\left(\max\left\{s \in \left(\max(\frac{-1}{\tau_{tail}^i}), 0\right) : |\hat{\boldsymbol{Q}}(s)| = 0\right\}\right)$, and define the mixing time of the system as:

$$\tau_{\text{mix}} = \frac{-1}{s_{\text{mix}}}. \tag{15}$$

If there are no root within this interval satisfying $|\hat{\boldsymbol{Q}}(s)| = 0$, the mixing time of the system cannot be computed. Our subsequent calculations of mixing time are based on finding the roots of the following equation:

$$|\boldsymbol{I} - \text{diag}\{\hat{\rho}_1(s), \cdots, \hat{\rho}_n(s)\}\boldsymbol{P}| = 0. \tag{16}$$

The definition of τ_{mix} can be verified by numerical simulation. In fact, if we perform numerical simulation on Eq. (2) and calculate $\max_j \max_{ik} |X_{ij}(t) - X_{kj}(t)|$, we will get a function that decays to 0. Fitting its slope through logarithmic linear regression, we can get:

$$\lim_{t\to\infty}\left(\max_j \max_{ik} |X_{ij}(t) - X_{kj}(t)|\right) \sim \text{constant} \cdot \exp(-t/\tau_{\text{mix}}). \tag{17}$$

We do not define the mixing time through this equation, but we need to find the corresponding mixing time in the real system. If the τ_{mix} we define is meaningful, then it must be clearly reflected in $\boldsymbol{X}(t)$ (that is, Eq. (17)). At the same time, Eq. (17) can also be used as a tool for numerical simulation to verify theoretical derivation.

2.6 Simplified case of exponential waiting time distributions

We consider the case where the waiting time distribution is exponential, then

$$\boldsymbol{\rho}(t) = \mathrm{diag}\{\mu_1^{-1}\exp(-t/\mu_1), \cdots, \mu_n^{-1}\exp(-t/\mu_n)\}.$$

We can write Eq. (16) as:

$$\begin{aligned}
|\hat{\boldsymbol{Q}}(s)| &= |\boldsymbol{I}_n - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}| \\
&= \left|\begin{bmatrix} \frac{1}{\hat{\rho}_1(s)} & 0 & \cdots & 0 \\ 0 & \frac{1}{\hat{\rho}_2(s)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\hat{\rho}_n(s)} \end{bmatrix} - \boldsymbol{P}\right| \\
&= \left|\begin{bmatrix} \mu_1 s + 1 & 0 & \cdots & 0 \\ 0 & \mu_2 s + 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n s + 1 \end{bmatrix} - \boldsymbol{P}\right| \\
&= \left|\begin{bmatrix} s + \frac{1}{\mu_1} & 0 & \cdots & 0 \\ 0 & s + \frac{1}{\mu_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s + \frac{1}{\mu_n} \end{bmatrix} - \begin{bmatrix} \frac{1}{\mu_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\mu_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\mu_n} \end{bmatrix} \boldsymbol{P}\right| \\
&= \left|s\boldsymbol{I}_n - \begin{bmatrix} \frac{1}{\mu_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\mu_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\mu_n} \end{bmatrix} \boldsymbol{L}\right|. \qquad (18)
\end{aligned}$$

We find that the mixing time in this case is $-\frac{1}{\lambda_1(\tilde{\boldsymbol{L}})}$. This is consistent with the result corresponding to Eq. (7), because the slowest eigenmode of $\boldsymbol{X}(t) = \exp(-\boldsymbol{L}t)$ corresponds to $\lambda_1(\boldsymbol{L})$.

2.7 Numerical simulation method

We develop a numerical solution method for the original diffusion equation, beginning with a reformulation of the convolution term within the governing equation:

$$\begin{aligned}
\boldsymbol{X}(t) &= \int_t^\infty \boldsymbol{\rho}(\tau)d\tau + \int_0^\infty \boldsymbol{\rho}(\tau)\boldsymbol{P}X(t-\tau)d\tau \\
&= \int_t^\infty \boldsymbol{\rho}(\tau)d\tau + \int_0^\infty \boldsymbol{\rho}(t-\tau)\boldsymbol{P}\boldsymbol{X}(\tau)d\tau,
\end{aligned} \qquad (19)$$

where $\boldsymbol{X}(t), P$ are generally two $n \times n$ square matrices, $\boldsymbol{\rho}(t)$ is a diagonal matrix, and we shall use $\boldsymbol{I}$ for the identity matrix. Define

$$t_j = j \times \Delta, with j = 0, 1, 2, \cdots, \tag{20}$$

$$X_j = \boldsymbol{X}(t_j), \tag{21}$$

where Δ denotes the iteration step size. We then have

$$\begin{aligned} X_{j+1} &= \int_{t_{j+1}}^{\infty} \boldsymbol{\rho}(\tau) d\tau + \int_0^{t_{j+1}} \boldsymbol{\rho}(t_{j+1} - \tau) \boldsymbol{P} \boldsymbol{X}(\tau) d\tau \\ &= \int_{t_{j+1}}^{\infty} \boldsymbol{\rho}(\tau) d\tau + \sum_{k=0}^{j} \int_{t_k}^{t_{k+1}} \boldsymbol{\rho}(t_{j+1} - \tau) \boldsymbol{P} \boldsymbol{X}(\tau) d\tau. \end{aligned} \tag{22}$$

We shall approximate this equation in two different ways.

$$X_{j+1} \approx \int_{t_{j+1}}^{\infty} \boldsymbol{\rho}(\tau) d\tau + \sum_{k=0}^{j} \int_{t_k}^{t_{k+1}} \boldsymbol{\rho}(t_{j+1} - \tau) d\tau \boldsymbol{P} X_k, \tag{23}$$

and

$$X_{j+1} \approx \int_{t_{j+1}}^{\infty} \boldsymbol{\rho}(\tau) d\tau + \sum_{k=0}^{j} \int_{t_k}^{t_{k+1}} \boldsymbol{\rho}(t_{j+1} - \tau) d\tau \boldsymbol{P} X_{k+1}. \tag{24}$$

We average Eqs. (23) and (24) to obtain

$$X_{j+1} \approx \int_{t_{j+1}}^{\infty} \boldsymbol{\rho}(\tau) d\tau + \sum_{k=0}^{j} \int_{t_k}^{t_{k+1}} \boldsymbol{\rho}(t_{j+1} - \tau) d\tau \boldsymbol{P} \frac{X_{k+1} + X_k}{2}. \tag{25}$$

We need to treat $j = 0$ differently to $j > 0$. For $j = 0$ we have

$$\left(\boldsymbol{I} - \frac{1}{2} \int_0^{t_1} \boldsymbol{\rho}(t_1 - \tau) d\tau \boldsymbol{P} \right) X_1 \approx \left(\int_{t_1}^{\infty} \boldsymbol{\rho}(\tau) d\tau + \frac{1}{2} \int_0^{t_1} \boldsymbol{\rho}(t_1 - \tau) d\tau \boldsymbol{P} \right). \tag{26}$$

For $j \geq 1$ we have

$$\begin{aligned} X_{j+1} \approx & \left(\int_{t_{j+1}}^{\infty} \boldsymbol{\rho}(\tau) d\tau + \frac{1}{2} \int_0^{t_1} \rho(t_1 - \tau) d\tau \boldsymbol{P} \right) \\ & + \frac{1}{2} \left(\sum_{k=1}^{j} \int_{t_k}^{t_{k+1}} \rho(t_{j+1} - \tau) d\tau \boldsymbol{P} X_k + \sum_{k=0}^{j} \int_{t_k}^{t_{k+1}} \rho(t_{j+1} - \tau) d\tau \boldsymbol{P} X_{k+1} \right), \end{aligned} \tag{27}$$

which can be written as

$$
\begin{aligned}
&(\boldsymbol{I}-\frac{1}{2}\int_{t_j}^{t_{j+1}}\rho(t_{j+1}-\tau)d\tau\boldsymbol{P})X_{j+1}\\
&\approx\left(\int_{t_{j+1}}^{\infty}\boldsymbol{\rho}(\tau)d\tau+\frac{1}{2}\int_{0}^{t_1}\rho(t_1-\tau)d\tau\boldsymbol{P}\right)\\
&+\frac{1}{2}\left(\sum_{k=1}^{j}\int_{t_k}^{t_{k+1}}\rho(t_{j+1}-\tau)d\tau\boldsymbol{P}X_k+\sum_{k=1}^{j}\int_{t_{k-1}}^{t_k}\rho(t_{j+1}-\tau)d\tau\boldsymbol{P}X_k\right),
\end{aligned}
\tag{28}
$$

or

$$
\begin{aligned}
&(\boldsymbol{I}-\frac{1}{2}\int_{t_j}^{t_{j+1}}\rho(t_{j+1}-\tau)d\tau\boldsymbol{P})X_{j+1}\\
&\approx\left(\int_{t_{j+1}}^{\infty}\boldsymbol{\rho}(\tau)d\tau+\frac{1}{2}\int_{0}^{t_1}\rho(t_1-\tau)d\tau\boldsymbol{P}\right)\\
&+\frac{1}{2}\sum_{k=1}^{j}\int_{t_{k-1}}^{t_{k+1}}\rho(t_{j+1}-\tau)d\tau\boldsymbol{P}X_k.
\end{aligned}
\tag{29}
$$

Finally, we have the following iterative approach to the determination of $X(t_j)$:
1.Start with $X(0)=I$.
2.$X(\delta)$ is determined by

$$
\begin{aligned}
X_1\approx&\left(\boldsymbol{I}-\frac{1}{2}\int_{0}^{t_1}\rho(t_1-\tau)d\tau\boldsymbol{P}\right)^{-1}\\
&\times\left(\int_{t_1}^{\infty}\boldsymbol{\rho}(\tau)d\tau+\frac{1}{2}\int_{0}^{t_1}\rho(t_1-\tau)d\tau\boldsymbol{P}\right).
\end{aligned}
$$

3.X_{j+1} for $j=1,2,3,\cdots$ are determined by

$$
\begin{aligned}
X_1\approx&\left(\boldsymbol{I}-\frac{1}{2}\int_{t_j}^{t_{j+1}}\rho(t_{j+1}-\tau)d\tau\boldsymbol{P}\right)^{-1}\\
&\times\left(\int_{t_{j+1}}^{\infty}\boldsymbol{\rho}(\tau)d\tau+\frac{1}{2}\int_{0}^{t_1}\rho(t_1-\tau)d\tau\boldsymbol{P}\right)\\
&+\left(\boldsymbol{I}-\frac{1}{2}\int_{t_j}^{t_{j+1}}\rho(t_{j+1}-\tau)d\tau\boldsymbol{P}\right)^{-1}\times\frac{1}{2}\sum_{k=1}^{j}\int_{t_{k-1}}^{t_{k+1}}\rho(t_{j+1}-\tau)d\tau\boldsymbol{P}X_k.
\end{aligned}
$$

Through this numerical algorithm, we can calculate $\boldsymbol{X}(t)$ with a certain accuracy from Eq. (2) and verify it with the theoretical results. At the same time, the mixing time τ_{mix} can also be estimated through Eq. (17).

3 Acceleration and Deceleration

3.1 Homogeneous waiting time

We first consider the case where the waiting time distributions of all nodes are the same. In this case, $\hat{\boldsymbol{\rho}}(s) = \hat{\rho}(s)\boldsymbol{I}$. Substituting into Eq. (16) we can get:

$$\begin{aligned} |\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{P}| &= |\boldsymbol{I} - \hat{\boldsymbol{\rho}}(s)\boldsymbol{F}^{-1}\Sigma\boldsymbol{F}| \\ &= |\boldsymbol{I} - \hat{\rho}(s)\Sigma| \\ &= \prod_{i=0}^{n-1}(1 - \hat{\rho}(s)\lambda_i(\boldsymbol{P})). \end{aligned} \tag{30}$$

Considering the eigenvalues, $\lambda_0(\boldsymbol{P})$ corresponds to a stationary distribution, so the mixing time corresponds to $\lambda_1(\boldsymbol{P})$. By performing Pade approximation on $\hat{\rho}(s)$, we can obtain:

$$\hat{\rho}(s) \approx \frac{2\mu + (\sigma^2 - \mu^2)s}{2\mu + (\sigma^2 + \mu^2)s}. \tag{31}$$

We apply this approximation to $1 - \hat{\rho}(s)\lambda_1(\boldsymbol{P}) = 0$, and we can approximate the root s_{mix}, and then obtain $\tau_{\text{mix}} = \frac{-1}{s_{\text{mix}}}$:

$$\tau_{\text{mix}} \approx \max\{\frac{\mu}{1 - \lambda_1(\boldsymbol{P})} + \frac{\sigma^2 - \mu^2}{2\mu}, \tau_{\text{tail}}\}. \tag{32}$$

This result demonstrates that the asymptotic dynamics depends crucially on the network structure through the spectral gap $1 - \lambda_1(\boldsymbol{P})$, and the temporal statistics via μ, σ^2, and τ_{tail}. The competition between these factors determines the dominant timescale of the system.

3.2 Critical analysis of the (1,1) Padé approximant

In this section, we provide a critical analysis of the (1,1) Padé approximant invoked in Eq. (31). We justify its application for root-finding in the negative half-plane ($s < 0$), explore its physical implications in the time domain, and discuss the relationship between the approximated and actual tails.

3.2.1 Motivation

Our theoretical requirement is to estimate the spectral root s_{mix} of the master equation $1 - \hat{\rho}(s)\lambda_1(P) = 0$ by utilizing the primary statistical descriptors of the waiting time distribution: the mean μ and the variance σ^2. Given that the root s_{mix} resides in the negative half-plane ($s < 0$), the choice of approximation for $\hat{\rho}(s)$ is crucial for maintaining accuracy beyond the immediate neighborhood of $s = 0$. Compared to a second-order Taylor expansion $\hat{\rho}(s) \approx 1 - \mu s + \frac{1}{2}\langle t^2\rangle s^2$, the (1,1) Padé approximant $\hat{\rho}(s) \approx \frac{1+as}{1+bs}$ is preferred for the following reasons:

- Exactness for exponential distributions: For a Poisson process ($\rho(t) = \frac{1}{\mu}e^{-t/\mu}$), the Padé form recovers $\hat{\rho}(s) = \frac{1}{1+\mu s}$ exactly, yielding the precise root $s_{mix} = -(1-\lambda_1)/\mu$. In contrast, Taylor expansions introduce truncation errors that may shift the root unphysically.
- Superior convergence in the negative half-plane: While Taylor polynomials are local approximations that diverge rapidly as s moves away from the origin, rational functions like the Padé approximant generally offer a much larger radius of convergence. By effectively mimicking the pole structure of $\hat{\rho}(s)$, the Padé form provides higher approximation precision at $s < 0$, which is essential for accurately locating the spectral root s_{mix}.

- Algebraic tractability: Substituting the Padé form into the master equation reduces the transcendental problem to a linear equation in s. This allows for the derivation of the analytical expression for τ_{mix} in Eq. (32), providing a direct mapping between temporal moments and network relaxation.

3.2.2 Time-domain interpretation

The (1,1) Padé approximant in Eq. (31) is not merely a curve-fitting tool but implies a specific mixture distribution in the time domain. Here we provide the complete derivation of its inverse Laplace transform. We start with the general form of the (1,1) Padé approximant for the Laplace transform $\hat{\rho}(s)$:

$$\hat{\rho}(s) \approx \frac{1+as}{1+bs}. \tag{33}$$

To ensure this approximation captures the bulk dynamics, we require it to match the first two moments of the original distribution $\rho(t)$: the mean $\mu = -\hat{\rho}'(0)$ and the second moment $\langle t^2 \rangle = \hat{\rho}''(0)$. Taking the derivatives of the Padé form:

$$\hat{\rho}'(s) = \frac{a(1+bs) - b(1+as)}{(1+bs)^2} = \frac{a-b}{(1+bs)^2}, \tag{34}$$

$$\hat{\rho}''(s) = \frac{-2b(a-b)}{(1+bs)^3}. \tag{35}$$

Evaluating at $s = 0$ and setting the constraints:

$$\hat{\rho}'(0) = a - b = -\mu \implies b - a = \mu, \tag{36}$$

$$\hat{\rho}''(0) = -2b(a-b) = 2b(b-a) = 2b\mu = \langle t^2 \rangle. \tag{37}$$

Solving Eqs. (36) and (37) for a and b, we obtain:

$$b = \frac{\langle t^2 \rangle}{2\mu} = \frac{\mu^2 + \sigma^2}{2\mu} = \tau_e, \quad a = b - \mu = \frac{\sigma^2 - \mu^2}{2\mu}. \tag{38}$$

Substituting a and b back into the Padé form yields Eq. (31):

$$\hat{\rho}(s) \approx \frac{2\mu + (\sigma^2 - \mu^2)s}{2\mu + (\sigma^2 + \mu^2)s}. \tag{31}$$

To perform the inverse Laplace transform, we decompose $\hat{\rho}(s)$ into a constant term and a simple pole term:

$$\hat{\rho}(s) \approx \frac{a}{b} + \frac{1 - a/b}{1+bs}. \tag{39}$$

Let $w = a/b$. Using the expressions for a and b derived above:

$$w = \frac{a}{b} = \frac{(\sigma^2 - \mu^2)/2\mu}{(\sigma^2 + \mu^2)/2\mu} = \frac{\sigma^2 - \mu^2}{\sigma^2 + \mu^2}. \tag{40}$$

The remaining weight is $1 - w = 1 - \frac{\sigma^2 - \mu^2}{\sigma^2 + \mu^2} = \frac{2\mu^2}{\sigma^2 + \mu^2}$. Thus, the Laplace domain expression becomes:

$$\hat{\rho}(s) \approx w + (1-w)\frac{1}{1+\tau_e s}. \tag{41}$$

Applying the inverse Laplace transform operator $\mathcal{L}^{-1}$ to each term:

1. $\mathcal{L}^{-1}\{w\} = w\delta(t)$, where $\delta(t)$ is the Dirac delta function.
2. $\mathcal{L}^{-1}\left\{\frac{1-w}{1+\tau_e s}\right\} = \frac{1-w}{\tau_e}\mathcal{L}^{-1}\left\{\frac{1}{s+1/\tau_e}\right\} = \frac{1-w}{\tau_e}e^{-t/\tau_e}$.

Summing these components, we obtain the effective time-domain distribution $\rho_{\text{eff}}(t)$:

$$\rho_{\text{eff}}(t) = w\delta(t) + \frac{1-w}{\tau_e}e^{-t/\tau_e}. \tag{42}$$

This derivation shows that the (1,1) Padé approximant yields a two-component time-domain representation of an arbitrary waiting-time distribution. When $w \in [0,1]$ (equivalently, $\sigma^2 \geq \mu^2$), this representation admits a probabilistic interpretation as a mixture of two distinct processes; when $\sigma^2 < \mu^2$ (so $w < 0$), it should instead be viewed as a formal (signed) decomposition corresponding to the rational approximation of $\hat{\rho}(s)$, rather than a nonnegative density:

- An instantaneous jump component $w\delta(t)$, representing a fraction of transitions that occur with zero waiting time. This term is essential for matching the variance when $\sigma^2 > \mu^2$.
- A continuous relaxation component with characteristic timescale τ_e, capturing the exponential decay that dominates the late-stage spreading process.

By construction, this effective distribution $\rho_{\text{eff}}(t)$ matches the mean μ and variance σ^2 of the original distribution exactly, thereby providing a robust analytical approximation for the network-driven mixing time.

3.2.3 Relationship between effective and asymptotic tails and the safeguard mechanism

We now investigate the relationship between the Padé-derived effective tail τ_e and the actual asymptotic tail τ_{tail} (where $\rho(t) \sim e^{-t/\tau_{tail}}$ as $t \to \infty$). This comparison reveals the conditions under which the Padé approximation is most accurate and how the framework remains robust when discrepancies arise.

- Small deviation regime: For distributions close to the exponential family, such as the **Gamma distribution** (with shape k and scale θ), the true tail is $\tau_{tail} = \theta$, while the Padé-derived tail is $\tau_e = \frac{k+1}{2}\theta$. In the near-exponential case ($k \approx 1$), τ_e and τ_{tail} are closely aligned, ensuring that the spectral root s_{mix} accurately captures the relaxation.
- Large deviation and safeguard: In systems with high temporal heterogeneity ($\sigma^2 \gg \mu^2$), the Padé-derived τ_e is "stretched" by the second moment to account for the impact of long-waiting events. However, if the approximation significantly deviates from the true physical tail, our final formula employs the $\max(\cdot)$ operator as a theoretical safeguard:

$$\tau_{mix} \approx \max\{\frac{\mu}{1-\lambda_1(\boldsymbol{P})} + \frac{\sigma^2-\mu^2}{2\mu}, \tau_{tail}\}. \tag{43}$$

  By explicitly comparing the spectral term (derived from Padé moment-matching) with the intrinsic tail τ_{tail}, the operator ensures that the slowest physical process always dominates the bound. If the Padé approximation underestimates the tail decay, the τ_{tail} term provides the necessary correction, preventing the model from yielding unphysical results.

3.2.4 Limitations on heavy-tailed distributions and boundary conditions

It is essential to clarify that our framework, and specifically the Padé-based spectral derivation, is strictly inapplicable to heavy-tailed distributions that lack an exponential decay (e.g., pure power-law distributions).

- Non-analyticity and divergence: For distributions where $\rho(t) \sim t^{-(1+\alpha)}$ with $\alpha \leq 2$, the second moment $\langle t^2 \rangle$ diverges, making the (1,1) Padé approximant ill-defined. Furthermore, the Laplace transform $\hat{\rho}(s)$ for such distributions is typically non-analytic at $s = 0$ (containing branch cuts), which precludes the search for a simple spectral root $s_{mix} < 0$ in the negative half-plane.
- Boundary completion via the max operator: While these cases fall outside the spectral research framework, the $\max(\cdot)$ operator naturally completes the boundary conditions of our model. In regimes where an exponential-like spectral relaxation cannot be established (effectively setting the spectral contribution to a lower bound or zero), the formula $\tau_{mix} \approx \max\{\ldots, \tau_{tail}\}$ will be entirely determined by the intrinsic tail term $\tau_{tail} = \infty$. This ensures that even when the moment-matching approximation reaches its mathematical limit, the framework consistently points to the tail-driven bottleneck as the dominant factor for system mixing.

3.3 Modifying waiting time of one node

Next, we consider the impact of changing the waiting time distribution of a node (from $\rho(t)$ to $\rho'(t)$) on the mixing time based on the case where the waiting time distribution of all nodes is homogeneous. Symbolically, we denote the mixing time of the homogeneous case (32) as τ_{mix}, and the changed mixing time as τ'_{mix}. We are more concerned about this change, that is, the sign and magnitude of $\Delta\tau_{\text{mix}} = \tau'_{\text{mix}} - \tau_{\text{mix}}$. In the following derivation, in order to distinguish unit matrices of different orders, we use subscripts to mark the order, such as: $\boldsymbol{I}_n, \boldsymbol{I}_{n-1}$. We first derive $|\boldsymbol{I}_n - \int_0^\infty e^{-st}\boldsymbol{\rho}(t)\boldsymbol{P}dt| = 0$:

$$
\begin{aligned}
|\boldsymbol{I}_n - \operatorname{diag}(\hat{\rho}_1(s), \ldots, \hat{\rho}'(s), \ldots, \hat{\rho}(s))\boldsymbol{P}| &= \frac{\hat{\rho}'(s)}{\hat{\rho}(s)} \left| \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\hat{\rho}(s)}{\hat{\rho}'(s)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} - \hat{\rho}(s)\boldsymbol{P} \right| \\
&= \frac{\hat{\rho}'(s)}{\hat{\rho}(s)} \left| \boldsymbol{I}_n - \hat{\rho}(s)\boldsymbol{P} + \begin{bmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\hat{\rho}(s)}{\hat{\rho}'(s)} - 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \right| \\
&= \frac{\hat{\rho}'(s)}{\hat{\rho}(s)} \left(|\boldsymbol{I}_n - \hat{\rho}(s)\boldsymbol{P}| + \left(\frac{\hat{\rho}(s)}{\hat{\rho}'(s)} - 1 \right) (I_{n-1} - \hat{\rho}(s)\boldsymbol{M}_{kk}) \right) \\
&= \frac{\hat{\rho}'(s)}{\hat{\rho}(s)} \left(\prod_{i=0}^{n-1} (1 - \hat{\rho}(s)\lambda_i(\boldsymbol{P})) + \left(\frac{\hat{\rho}(s)}{\hat{\rho}'(s)} - 1 \right) \prod_{i=1}^{n-1} (1 - \hat{\rho}(s)\lambda_i(\boldsymbol{M}_{kk})) \right) \\
&= 0. \qquad (44)
\end{aligned}
$$

Consequently, $|\boldsymbol{I}_n - \int_0^\infty e^{-st}\boldsymbol{\rho}(t)\boldsymbol{P}dt| = 0$ is equivalent to:

$$
\frac{\hat{\rho}(s)}{\hat{\rho}'(s)} = 1 - \frac{\prod_{i=0}^{n-1}(1 - \hat{\rho}(s)\lambda_i(\boldsymbol{P}))}{\prod_{i=1}^{n-1}(1 - \hat{\rho}(s)\lambda_i(\boldsymbol{M}_{kk}))}. \qquad (45)
$$

Through the Cauchy alternation theorem, we can get $\lambda_{n-1}(\boldsymbol{P}) \leq \lambda_{n-1}(\boldsymbol{M}_{kk}) \leq \cdots \leq \lambda_1(\boldsymbol{P}) \leq \lambda_1(\boldsymbol{M}_{kk}) \leq \lambda_0(\boldsymbol{P})$.

3.4 Qualitative analysis

Next, for further qualitative analysis, we define:

$$F(s) = \frac{\hat{\rho}(s)}{\hat{\rho}'(s)}, \tag{46}$$

$$G(s) = 1 - \frac{\prod_{i=0}^{n-1}(1-\hat{\rho}(s)\lambda_i(\boldsymbol{P}))}{\prod_{i=1}^{n-1}(1-\hat{\rho}(s)\lambda_i(\boldsymbol{M}_{kk}))}. \tag{47}$$

It is obvious that $F(0) = G(0) = 1$. The largest negative real part (noted by s'_{mix}) of the intersections between $F(s)$ and $G(s)$ corresponds to the diffusion speed. $G(s)$ in our discussion is fixed for $\hat{\rho}'(s)$. We consider Gamma distributions:

$$\rho(t) = \frac{t^{\mu/\tau_{\text{tail}}-1}}{\Gamma(\mu/\tau_{\text{tail}})\tau_{\text{tail}}^{\mu/\tau_{\text{tail}}}} e^{-t/\tau_{\text{tail}}}, \tag{48}$$

$$\rho'(t) = \frac{t^{\mu'/\tau'_{\text{tail}}-1}}{\Gamma(\mu'/\tau'_{\text{tail}})\tau'^{\mu'/\tau'_{\text{tail}}}_{\text{tail}}} e^{-t/\tau'_{\text{tail}}}. \tag{49}$$

Therefore,

$$\hat{\rho}(s) = \int_0^\infty e^{-st}\rho(t)dt = \frac{1}{(\tau_{\text{tail}}s+1)^{\mu/\tau_{\text{tail}}}}, \tag{50}$$

$$\hat{\rho}'(s) = \int_0^\infty e^{-st}\rho'(t)dt = \frac{1}{(\tau'_{\text{tail}}s+1)^{\mu'/\tau'_{\text{tail}}}}. \tag{51}$$

Derivative of $\hat{\rho}(s)$ gives:

$$\frac{d}{ds}\hat{\rho}(s) = \frac{-\mu}{\tau_{\text{tail}}s+1}\hat{\rho}(s). \tag{52}$$

The exponential distribution is a special case of the Gamma distribution, so when $\rho(t) = \mu^{-1}e^{-t/\mu}$, we can get:

$$\frac{d}{ds}\hat{\rho}(s) = \frac{-\mu}{\mu s+1}\hat{\rho}(s). \tag{53}$$

Based on the previous discussion, we can get:

$$\begin{aligned}\frac{d}{ds}F(s) &= \frac{\frac{d}{ds}\hat{\rho}(s)\hat{\rho}'(s) - \hat{\rho}(s)\frac{d}{ds}\hat{\rho}'(s)}{(\hat{\rho}'(s)^2)} \\ &= \frac{\frac{-\mu}{\tau_{\text{tail}}s+1}\hat{\rho}(s)\hat{\rho}'(s) - \hat{\rho}(s)\frac{-\mu'}{\tau'_{\text{tail}}s+1}\hat{\rho}'(s)}{(\hat{\rho}'(s)^2)}.\end{aligned} \tag{54}$$

The only possible negative zero of $\frac{d}{ds}F(s)$ is

$$s = \frac{\mu-\mu'}{\tau_{\text{tail}}\mu' - \tau'_{\text{tail}}\mu}. \tag{55}$$

This shows that $F(s)$ only have one vertex. Through the analysis of the intersection of functions $F(s)$ and $G(s)$, we search for the possible location of the root of $F(s) = G(s)$. The root of $G(s) = 1$ corresponds to

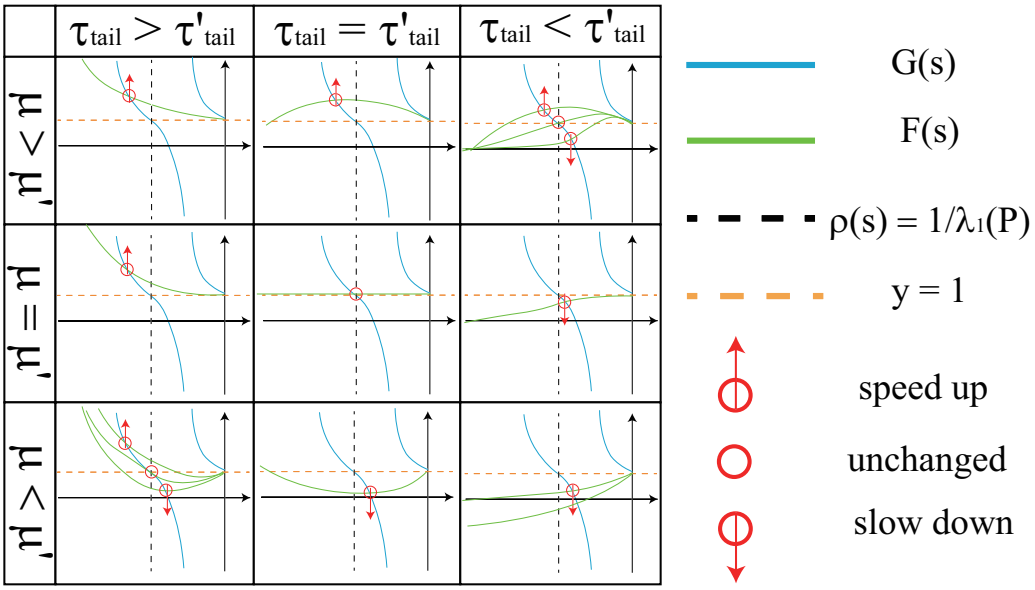

Fig. S1. The size relationship between the mean (μ', μ) and the exponential tail $(\tau'_{\text{tail}}, \tau_{tail})$ of $\hat{\rho}'(s)$ and $\hat{\rho}(s)$ will affect the qualitative analysis of $\text{sign}(\Delta\tau_{\text{mix}})$. We draw a schematic diagram of the 9 possible situations of $F(s)$ and $G(s)$, where the blue line represents $G(s)$; the green line represents $F(s)$; The black line represents $\hat{\rho}(s) = 1/\lambda_1(\boldsymbol{P})$, which is $s = s_{\text{mix}}$; The yellow line represents $y = 1$, which is $F(s)$ when the waiting time distribution is homogeneous.

s_{mix}, which corresponds to the mixing time when the waiting time distributions of all nodes are $\rho(t)$. The root of $G(s) = F(s)$ corresponds to s'_{mix} , which corresponds to the mixing time when modify the waiting time distribution of node k to $\rho'(t)$. Therefore, when $s'_{\text{mix}} < s_{\text{mix}}$, we can get $\tau'_{\text{mix}} < \tau_{\text{mix}}$, it represents the acceleration of diffusion, that is, $\text{sign}(\Delta\tau_{\text{mix}}) < 0$. When $s'_{\text{mix}} > s_{\text{mix}}$, we can get $\tau'_{\text{mix}} > \tau_{\text{mix}}$, it represents the deceleration of diffusion, that is, $\text{sign}(\Delta\tau_{\text{mix}}) > 0$. we classify the cases in which diffusion is accelerated and decelerated, and these situations are shown in Fig. S1. The content in Fig. S1 is a display of various situations, not real data. In Fig. S2, S3, S4, we take $\tau_{\text{tail}} = \mu = 1$, take different values for τ'_{tail} and μ', and then draw images of $F(s)$ and $G(s)$, verifying the situation we classified in Fig. 1.

Case. 1 If $\mu > \mu'$ and $\tau_{\text{tail}} > \tau'_{\text{tail}}$, the derivative of $F(s)$ at 0 is positive, and $F(s)$ tends to positive infinity at

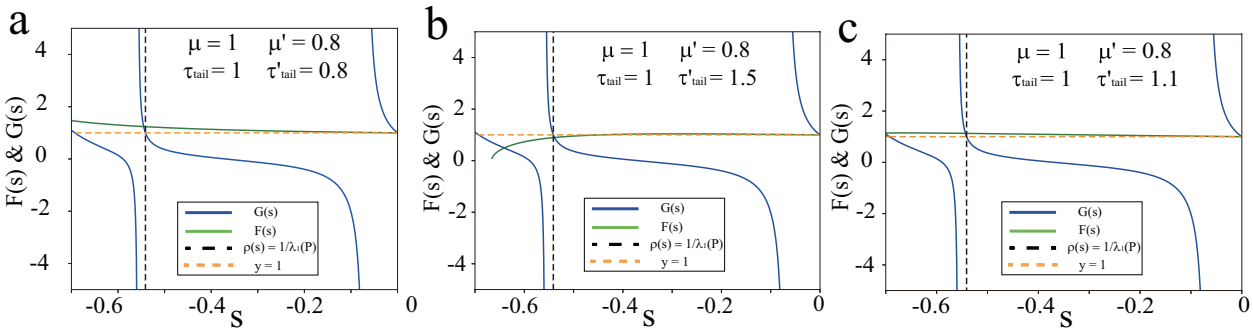

Fig. S2. a, for the case where $\mu > \mu'$, $\tau_{\text{tail}} > \tau'_{\text{tail}}$, $\Delta\tau_{\text{mix}}$ takes a negative value. b-c, for the first case where $\mu > \mu'$, $\tau_{\text{tail}} < \tau'_{\text{tail}}$, $\Delta\tau_{\text{mix}}$ takes a positive (b) or negative value (c).

the left endpoint of its domain, which occurs at $-1/\tau_{\text{tail}}$. We have:

$$\begin{aligned}
\frac{d}{ds}G(0) &= -\frac{\sum_{i=0}^{n-1}(\prod_{j\neq i}(1-\rho(0)\lambda_j(P)))\cdot(-\lambda_i(\boldsymbol{P}))\frac{d}{ds}\rho(0)(\prod_{i=1}^{n-1}(1-\rho(0)\lambda_i(\boldsymbol{M}_{kk})))}{(\prod_{i=1}^{n-1}(1-\rho(0)\lambda_i(\boldsymbol{M}_{kk})))^2} \\
&\quad - \frac{\sum_{i=1}^{n-1}(\prod_{j\neq i}(1-\rho(0)\lambda_j(\boldsymbol{M}_{kk})))\cdot(-\lambda_i(\boldsymbol{M}_{kk}))\frac{d}{ds}\rho(0)(\prod_{i=0}^{n-1}(1-\rho(0)\lambda_i(\boldsymbol{P})))}{(\prod_{i=1}^{n-1}(1-\rho(0)\lambda_i(\boldsymbol{M}_{kk})))^2} \\
&= \frac{\prod_{j\neq 0}(1-\lambda_j)E(\rho)\prod_{i=1}^{n-1}(1-\lambda_i(\boldsymbol{M}_{kk}))-0}{\prod_{i=1}^{n-1}(1-\lambda_i(\boldsymbol{M}_{kk}))^2} \\
&= \frac{\prod_{i=1}^{n-1}(1-\lambda_i(\boldsymbol{P}))E(\rho)}{\prod_{i=1}^{n-1}(1-\lambda_i(\boldsymbol{M}_{kk}))} \\
&> E(\rho) \\
&> E(\rho) - E(\rho') \\
&= \frac{d}{ds}F(0).
\end{aligned} \tag{56}$$

Thus, the derivative of $G(s)$ at 0 is larger than that of $F(s)$. As a result, $F(s)$ does not intersect with the first branch of $G(s)$, but instead intersect with the second branch of $G(s)$, where $G(s) = F(s) > 1$. In this case, $\Delta\tau_{\text{mix}} < 0$ and $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) > 1$, the diffusion is accelerated.

We plot $F(s)$ and $G(s)$ for selected parameters and distributions to verify the qualitative analysis (Fig. S2 a).

Case. 2 If $\mu > \mu'$ and $\tau_{\text{tail}} < \tau'_{\text{tail}}$, the derivative of $F(s)$ at 0 is positive, and $F(s) = 0$ at the left endpoint of its domain, which occurs at $s = -1/\tau'_{\text{tail}}$. This indicates that the general trend of $F(s)$ is first to increase and then decrease. By applying a similar analytical approach as in Case 1, we can also arrive at analogous conclusions:

(1) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) > 1$, then $\Delta\tau_{\text{mix}} < 0$, and in this case, the diffusion is accelerated.

(2) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) < 1$, then $\Delta\tau_{\text{mix}} > 0$, and the diffusion is decelerated.

By organizing these conclusions, we derive that if $\rho'^{-1}(1/\lambda_1(\boldsymbol{P})) < \hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))$, the diffusion is accelerated; if $\rho'^{-1}(1/\lambda_1(\boldsymbol{P})) > \hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))$, the diffusion is decelerated. In numerical experiments, we plot $F(s)$ and $G(s)$ for selected parameters and distributions. Notice that the curve of $F(s)$ changes more gradually compared to $G(s)$, which rules out the possibility of extreme scenarios suggested by the

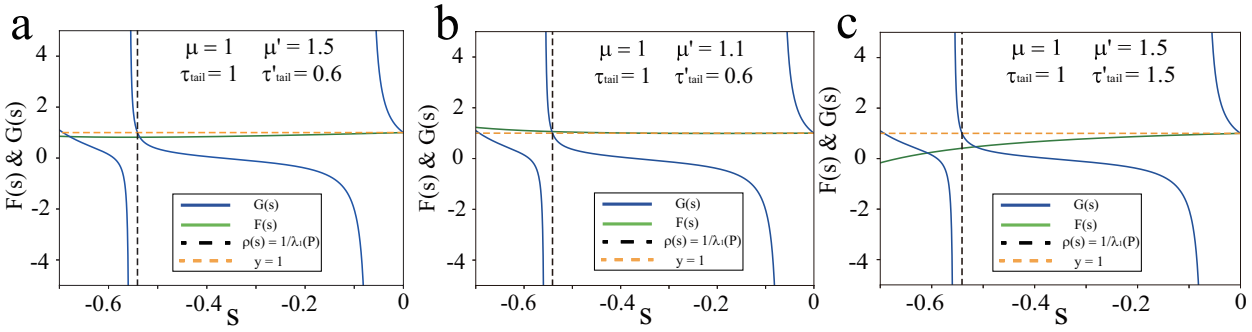

Fig. S3. a-b, for the first case where $\mu < \mu'$, $\tau_{\text{tail}} > \tau'_{\text{tail}}$, $\Delta\tau_{\text{mix}}$ takes a positive (a) or negative value (b). c, for the case where $\mu < \mu'$, $\tau_{\text{tail}} < \tau'_{\text{tail}}$, $\Delta\tau_{\text{mix}}$ takes a positive value.

qualitative analysis. The acceleration case is shown in Fig. S2 b, and the deceleration case is shown in Fig. S2 c.

Case. 3 If $\mu < \mu'$ and $\tau_{\text{tail}} > \tau'_{\text{tail}}$, the derivative of $F(s)$ at 0 is less than zero, and $F(s)$ tends to infinity at the left endpoint of its domain, which occurs at $s = -1/\tau_{\text{tail}}$. This indicates that the general trend of $F(s)$ is to first decrease and then increase. If $\rho(t) = \rho'(t)$, we obtain $F(s) = 1$. The intersection of $F(s)$ and $G(s)$ is the intersection of $G(s)$ and $y = 1$, which occurs at $s = \hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))$. Thus, when $\Delta\tau_{\text{mix}} < 0$, the intersection shifts to the left; when $\Delta\tau_{\text{mix}} > 0$, the intersection shifts to the right. In our earlier derivation, we established that $F(s)$ has only one extreme in its domain, so by comparing $F(s)$ with 1, we can determine the direction of the intersection's movement:

(1) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) > 1$, then $\Delta\tau_{\text{mix}} < 0$, and the diffusion is accelerated.

(2) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) < 1$, then $\Delta\tau_{\text{mix}} > 0$, and the diffusion is decelerated.

We plot $F(s)$ and $G(s)$ for selected parameters and distributions to verify the qualitative analysis. The acceleration case is shown in Fig. S3 a, and the deceleration case is shown in Fig. S3 b.

Case. 4 If $\mu < \mu'$ and $\tau_{\text{tail}} < \tau'_{\text{tail}}$, $F(s)$ will intersect with the second branch of $G(s)$, where $G(s) = F(s) < 1$. In this case, $\Delta\tau_{\text{mix}} > 0$, $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) < 1$, the diffusion is decelerated. We plot $F(s)$ and $G(s)$ for selected parameters and distributions to verify the qualitative analysis (Fig. S3 c).

Case. 5 If $\mu > \mu'$ and $\tau_{\text{tail}} = \tau'_{\text{tail}}$, the derivative of $F(s)$ at 0 is positive, and $F(s) = 1$ at the left endpoint of its domain, This indicates that the general trend of $F(s)$ is first to increase and then decrease to 1. which occurs at $-1/\tau_{\text{tail}}$. As we know in case 1, the derivative of $G(s)$ at 0 is larger than that of $F(s)$. Thus, $F(s)$ intersects with the second branch of $G(s)$, where $G(s) = F(s) > 1$. In this case, $\Delta\tau_{\text{mix}} < 0$, indicating that the diffusion is accelerated. We plot $F(s)$ and $G(s)$ for selected parameters and distributions to verify the qualitative analysis (Fig. S4 a).

Case. 6 If $\mu = \mu'$ and $\tau_{\text{tail}} > \tau'_{\text{tail}}$, the derivative of $F(s)$ at 0 is zero, and $F(s)$ tends to infinity at the left endpoint of its domain, which occurs at $s = -1/\tau_{\text{tail}}$. It is uncertain whether $F(s)$ is increasing or decreasing in the neighborhood of $s = 0$. However, given that $F(s)$ has at most one extreme, the possible shape of the $F(s)$ curve is similar to the cases in Case 1 and Case 3. If $F(s)$ is monotonically increasing, then $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) > 1$, indicating that $\Delta\tau_{\text{mix}} < 0$; diffusion is accelerated. If $F(s)$ first decreases and then increases, the value of $F(s)$ at $s = \hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))$ can still be used to assess the change in $\Delta\tau_{\text{mix}}$:

(1) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) > 1$, then $\Delta\tau_{\text{mix}} < 0$, and in this case, the diffusion is accelerated.

(2) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) < 1$, then $\Delta\tau_{\text{mix}} > 0$, and the diffusion is decelerated.

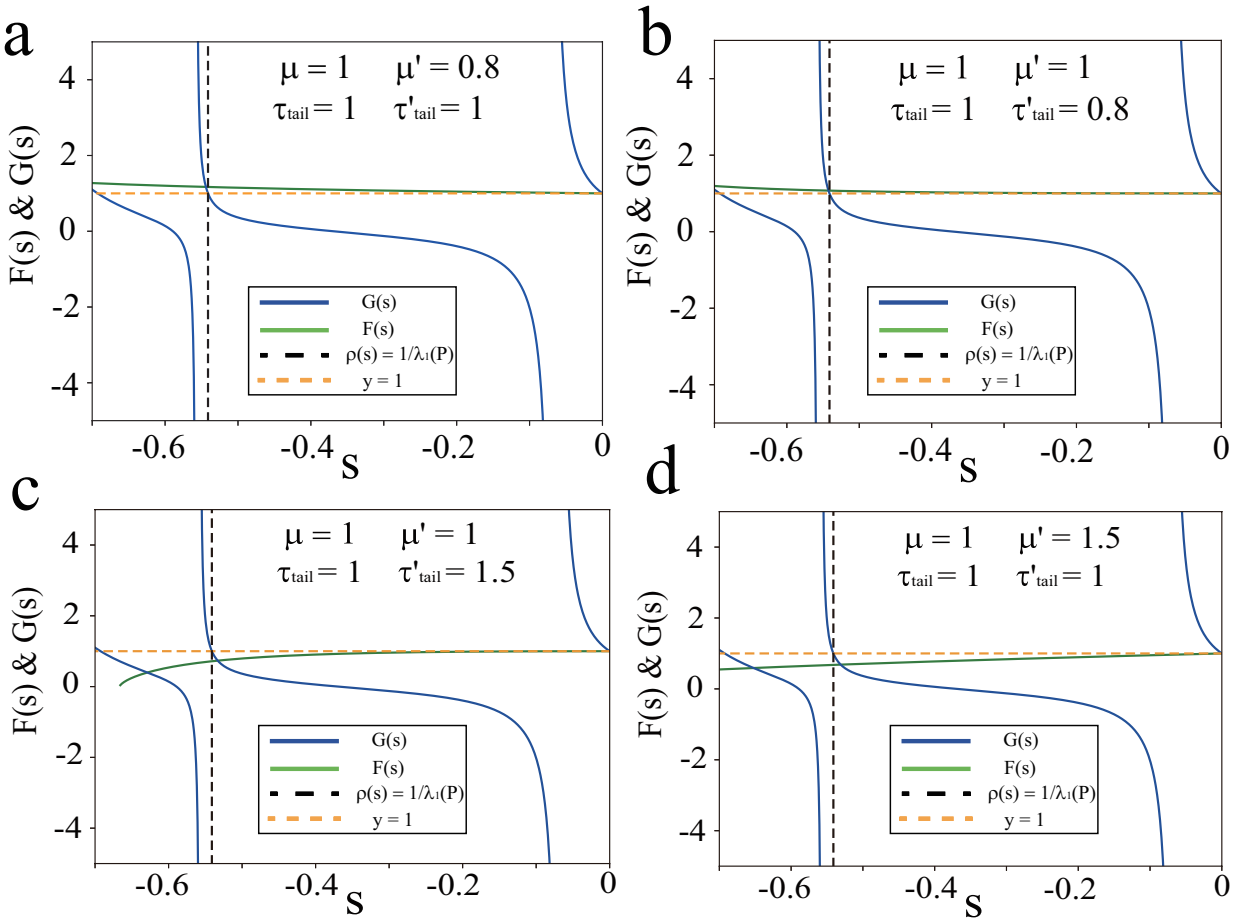

Fig. S4. a, for the case where $\mu > \mu'$, $\tau_{\text{tail}} = \tau'_{\text{tail}}$, $\Delta\tau_{\text{mix}}$ takes a negative value. b, for the case where $\mu = \mu'$, $\tau_{\text{tail}} > \tau'_{\text{tail}}$, $\Delta\tau_{\text{mix}}$ takes a negative value. c, for the case where $\mu = \mu'$, $\tau_{\text{tail}} < \tau'_{\text{tail}}$, $\Delta\tau_{\text{mix}}$ takes a positive value. d, for the case where $\mu = \mu'$, $\tau_{\text{tail}} < \tau'_{\text{tail}}$, $\Delta\tau_{\text{mix}}$ takes a positive value.

We plot $F(s)$ and $G(s)$ for selected parameters and distributions to verify the qualitative analysis (Fig. S4 b).

Case. 7 If $\mu = \mu'$ and $\tau_{\text{tail}} < \tau'_{\text{tail}}$, the derivative of $F(s)$ at 0 is zero, and $F(s) = 0$ at the left endpoint of its domain, which occurs at $s = -1/\tau'_{\text{tail}}$. It is uncertain whether $F(s)$ is increasing or decreasing in the neighborhood of $s = 0$. However, given that $F(s)$ has at most one extreme, the possible shape of the $F(s)$ curve is similar to the cases in Case 2 and Case 4. If $F(s)$ is monotonically decreasing, then $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) < 1$, indicating that $\Delta\tau_{\text{mix}} > 0$; diffusion is accelerated. If $F(s)$ first increases and then decreases, the value of $F(s)$ at $s = \hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))$ can still be used to assess the change in $\Delta\tau_{\text{mix}}$:

(1) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) > 1$, then $\Delta\tau_{\text{mix}} < 0$, and in this case, the diffusion is accelerated.

(2) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) < 1$, then $\Delta\tau_{\text{mix}} > 0$, and the diffusion is decelerated.

We plot $F(s)$ and $G(s)$ for selected parameters and distributions to verify the qualitative analysis (Fig. S4 c).

Case. 8 If $\mu < \mu'$ and $\tau_{\text{tail}} = \tau'_{\text{tail}}$, the derivative of $F(s)$ at 0 is negative, and $F(s) = 1$ at the left endpoint of its domain, This indicates that the general trend of $F(s)$ is to first increase and then decrease to 1. This occurs at $-1/\tau_{\text{tail}}$. $F(s)$ intersects with the second branch of $G(s)$, where $G(s) = F(s) < 1$. In this case, $\Delta\tau_{\text{mix}} > 0$, indicating that the diffusion is accelerated. We plot $F(s)$ and $G(s)$ for selected parameters and distributions to verify the qualitative analysis (Fig. S4 d).

Case. 9 If $\mu = \mu'$ and $\tau_{\text{tail}} = \tau'_{\text{tail}}$, then $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) = G((\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P})))) = 1$ and $\Delta\tau_{\text{mix}} = 0$.

Based on the analysis and validation of the previous nine cases, we derive the following criteria:
(1) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) > 1$, then $\Delta\tau_{\text{mix}} < 0$, and in this case, the diffusion is accelerated.
(2) If $F(\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))) < 1$, then $\Delta\tau_{\text{mix}} > 0$, and the diffusion is decelerated.
We can get the following formula:

$$\text{sign}\left(\Delta\tau_{\text{mix}}\right) = \text{sign}\left(\hat{\rho}'(\frac{1}{\lambda_1(\boldsymbol{P})}) - \hat{\rho}(\frac{1}{\lambda_1(\boldsymbol{P})})\right). \tag{57}$$

It is important to note that $\lambda_1(\boldsymbol{P})$ is independent of the choice of node k whose waiting time distribution is modified. Applying the Padé approximation to $\hat{\rho}(s)$ and $\hat{\rho}'(s)$, respectively, yields:

$$\hat{\rho}^{-1}(\frac{1}{\lambda_1(\boldsymbol{P})}) \approx \frac{\mu}{1-\lambda_1(\boldsymbol{P})} + \frac{\sigma^2-\mu^2}{2\mu}, \tag{58}$$

$$\rho'^{-1}(\frac{1}{\lambda_1(\boldsymbol{P})}) \approx \frac{\mu'}{1-\lambda_1(\boldsymbol{P})} + \frac{\sigma'^2-\mu'^2}{2\mu'}, \tag{59}$$

where σ^2 and σ'^2 are the variances of $\rho(t)$ and $\rho'(t)$, respectively. Since both are Gamma distributions, we have $\sigma^2 = \mu\tau_{\text{tail}}$ and $\sigma'^2 = \mu'\tau'_{\text{tail}}$. Substituting the above equations into Eq. (57) gives:

$$\begin{aligned}
\text{sign}\left(\Delta\tau_{\text{mix}}\right) &\approx \text{sign}\left(\frac{\mu'}{1-\lambda_1(\boldsymbol{P})} + \frac{\sigma'^2-\mu'^2}{2\mu'} - \frac{\mu}{1-\lambda_1(\boldsymbol{P})} - \frac{\sigma^2-\mu^2}{2\mu}\right) \\
&= \text{sign}\left(\frac{\mu'}{1-\lambda_1(\boldsymbol{P})} + \frac{\tau'_{\text{tail}}-\mu'}{2} - \frac{\mu}{1-\lambda_1(\boldsymbol{P})} - \frac{\tau_{\text{tail}}-\mu}{2}\right) \\
&= \text{sign}\left(2\mu' + (1-\lambda_1(\boldsymbol{P}))(\tau'_{\text{tail}}-\mu') - 2\mu - (1-\lambda_1(\boldsymbol{P}))(\tau_{\text{tail}}-\mu)\right) \\
&= \text{sign}\left((1+\lambda_1(\boldsymbol{P}))(\mu'-\mu) + (1-\lambda_1(\boldsymbol{P}))(\tau'_{\text{tail}}-\tau_{\text{tail}})\right).
\end{aligned} \tag{60}$$

The line separating acceleration and deceleration of the mixing process can be expressed as

$$\tau'_{\text{tail}} = \frac{\lambda_1(\boldsymbol{P})+1}{\lambda_1(\boldsymbol{P})-1}(\mu'-\mu) - \tau_{tail}. \tag{61}$$

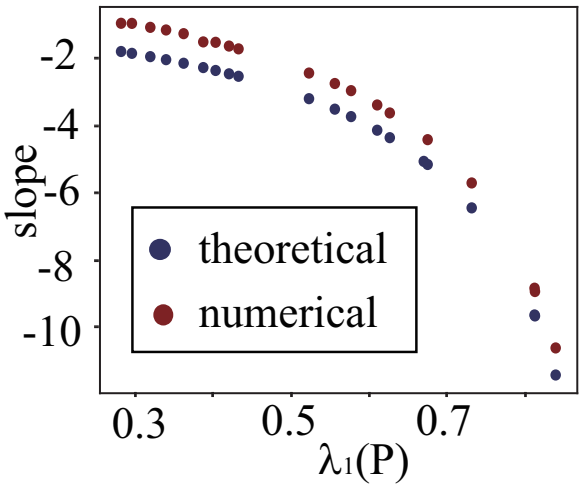

Fig. S5. we use Eq. (45) to calculate $\Delta\tau_{\text{mix}} = 0$ for different networks. We fit the slope of τ'_{tail} about μ' (red dots) and compare it with the theoretical value (blue dots) from Eq. (61). The networks are constructed as follows: starting with a 20-node ring, random edges are added to create a new network, the calculations are performed on a group of networks constructed in this way.

Note that in the case of exponential waiting time distributions, where $\mu = \tau_{tail}$, we find that the line separating acceleration and deceleration is simply $\tau_{tail} = \tau'_{\text{tail}}$, as expected. We found that the network's spectral gap is the balance between the mean (μ', μ) and the exponential tail $(\tau'_{\text{tail}}, \tau_{tail})$. In order to verify the influence of the spectral gap on the slope in Eq. (61), we conducted numerical experiments and calculated the theoretical value of $\Delta\tau_{\text{mix}}$ for a group of networks (the networks are constructed as follows: starting with a 20-node ring, random edges are added to create a new network.) when $\tau'_{\text{tail}}, \mu'$ takes values in $[0.5, 1.5]$ through Eq. (45). Based on this, we used the least squares method to fit the slope of $\Delta\tau_{\text{mix}} = 0$ (red dots). Then we compared the estimated value of the slope (blue dots) calculated by Eq. (61). We found that the two are in good agreement (Fig. S5).

4 Upper bounds and Lower bounds

After considering the influencing factors of the $\Delta\tau_{\text{mix}}$ sign, we found that when the waiting time distributions are Gamma distribution, the qualitative conclusion of whether the perturbed waiting time distribution is accelerated or decelerated diffusion is the same for different nodes. The magnitude of acceleration and deceleration, that is, the difference in $|\Delta\tau_{\text{mix}}|$, corresponds to the difference between different nodes. Our subsequent discussion does not need to assume that the waiting time distribution is a *Gamma* distribution. The previous qualitative analysis made this assumption to ensure that $F(s)$ does not have too weird image features.

4.1 Waiting time distribution with single parameter variation

Consider a family of waiting time distributions $\rho'(t,a)$ that vary along a single parameter a. The waiting time distributions of other nodes in the network is $\rho(t)$. The waiting time distribution of the selected node k is $\rho'(t,a)$. We add conditions to $\rho'(t,a)$: Define $\hat{\rho'}(s,a)$ as the Laplace transform of $\rho'(t,a)$. For any s, $\hat{\rho'}(s,a)$ varies monotonically with respect to a, and as a increases, the mean μ' of $\hat{\rho'}(s,a)$ also increases. The family of waiting time distributions $\rho'(t,\mu')$ that varies along the mean μ' is a special case of the single parameter family defined above. Under such conditions, we can obtain the monotonically changing family of curves $F(s,a) = \frac{\hat{\rho'}(s,a)}{\hat{\rho}(s)}$, and it can be proved that for any $a_1 > a_2 > 0$, if s_1 is the intersection of $F(s,a_1)$ and $G(s)$, and s_2 is the intersection of $F(s,a_2)$ and $G(s)$, then $s_1 > s_2$ must hold, so that the mixing time of the system determined by a_1 is larger than that of a_2: $\tau^1_{\text{mix}} > \tau^2_{\text{mix}}$. We prove the above discussion through the following theorem:

Definition 1. *If the Laplace transform $\hat{\rho'}(s,a)$ of the waiting time distribution $\rho'(t,a)$ satisfies that $\hat{\rho'}(s,a)$ is a monotonic function with respect to the parameter a for any fixed $s > \frac{-1}{\tau^i_{tail}}(i = 1, \cdots, n)$, then the distribution $\rho'(t,a)$ is said to be monotonic with respect to the parameter a.*

Theorem 2. *If $\rho'(t,a)$ is monotonic with respect to the parameter a, then for the diffusion system*

$$\boldsymbol{\rho}(t) = diag\{\rho(t), \cdots, \rho'(t,a), \cdots, \rho(t)\},$$

where $\rho'(t,a)$ is the waiting time distribution of node k, the mixing time τ_{mix} monotonically varies with a.

Proof. We already know that for any fixed $s > \frac{-1}{\tau^i_{\text{tail}}}(i = 1, \cdots, n)$, for all $a_1 > a_2 > 0$, it holds that $\hat{\rho'}(s,a_1) > \hat{\rho'}(s,a_2) > 0$, thus, $F(s,a_1) < F(s,a_2)$. Suppose $F(s,a_i)$ and $G(s)$ have their first non-zero intersection points at s_i, if $\tau^{a_1}_{\text{mix}} \leq \tau^{a_2}_{\text{mix}}$, we have $s_1 \leq s_2$. Define $H_i(s) = F(s,a_i) - G(s)$. Since $G(\rho^{-1}(\frac{-1}{\lambda_1(\boldsymbol{M}_{kk})})) = -\infty$, it follows that $H_1(s_1) = 0$ and $H_2(s_1) \leq 0$. This contradicts $F(s,a_1) < F(s,a_2)$. Hence, the assumption $\tau^{a_1}_{\text{mix}} \leq \tau^{a_2}_{\text{mix}}$ does not hold, $\tau^{a_1}_{\text{mix}} > \tau^{a_2}_{\text{mix}}$. As the parameter a increases, the mixing time increases. □

We perform numerical simulations on $\rho(t) = e^{-t}, \rho'(t,a) = \mu'^{-1}e^{-t/\mu'}$ in a randomly generated network and plot the images of $F(s)$ and $G(s)$ (see Fig. S6). It can be seen that when $a \to 0$, $F(s,a) \to \hat{\rho}(s)$, the mixing time at this time is the intersection of $\hat{\rho}(s)$ and $G(s)$. This is the upper limit of network diffusion that can be accelerated by changing the waiting time distribution of the selected node along the parameter a, which is also the lower bound of the mixing time. When a is greater than a certain value, $F(s)$ may not have an intersection with $G(s)$ within the definition domain $(-\frac{1}{a}, 0)$. At this time, the mixing time can no longer be obtained by the definition of $|\hat{\boldsymbol{Q}}(s)| = 0$. But in fact, when a is large enough, the increase in mixing time has converged. If we can calculate the mixing time when $a \to \infty$, we can also approximate the upper bound of the mixing time.

In our previous definition, as a increases, the mean μ' of $\hat{\rho'}(s,a)$ also increases, so the mixing time corresponding to $a \to 0, a \to \infty$ and $\mu' \to 0, \mu' \to \infty$ is consistent. Now we directly calculate the mixing time for the cases of $\mu' \to 0$ and $\mu' \to \infty$. Note that: when $\mu' \to 0$, $\hat{\rho'}(s) = 1$; when $\mu' \to \infty$, $\hat{\rho'}(s) = \infty$. For the case of $\mu' \to 0$, we make the average waiting time of the walker reach the lower bound (0), and prove that this case

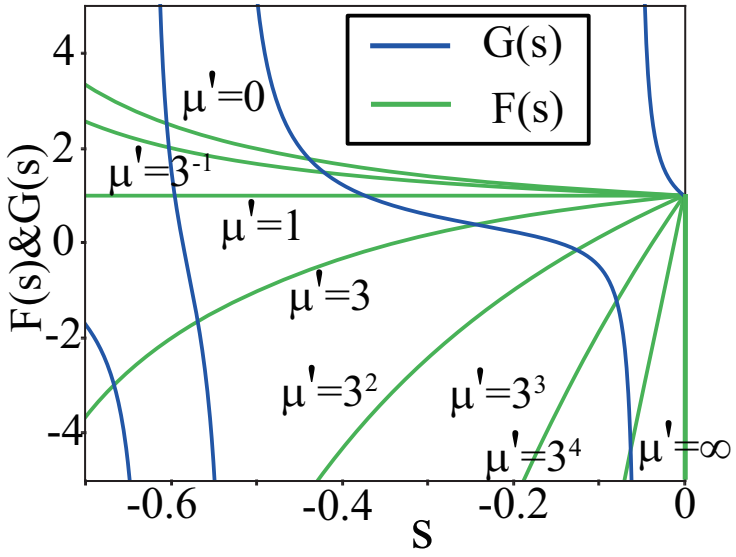

Fig. S6. The figure shows the intersection of $F(s)$ (green line) and $G(s)$ (blue line) for different mean values of $\rho'(t)$. It can be seen that as μ' increases from 0 to ∞, the intersection of $F(s)$ and $G(s)$ continues to move to the right.

reaches the lower bound of the mixing time (through the above theorem). Obviously, in this case, $\tau'_{\text{tail}} \to 0$. So we use the lower bound to describe the state and the mixing time at this time. For the case of $\mu' \to \infty$, we make the average waiting time of the walker reach the upper bound (∞), and prove that this case reaches the upper bound of the mixing time. Obviously, in this case, $\tau'_{\text{tail}} \to \infty$. So we use the upper bound to describe the state and the mixing time at this time. In fact, $\Delta\tau_{\text{mix}}$ and τ_{mix} are synchronously taken to the upper and lower bounds. When we care about the change of the mixing time, we discuss the upper and lower bounds of $\Delta\tau_{\text{mix}}$, and when we care about the mixing time, we discuss the upper and lower bounds of τ_{mix}. For the case where the waiting time distribution of each node is different (Eq. 16), the two extreme cases of waiting time distribution, $\mu' \to 0$ and $\mu' \to \infty$, can also be calculated. Although we cannot prove that the mixing time can reach the lower and upper bounds in these two extreme cases, we still use the lower bound and upper bound (for the average waiting time) to refer to these two cases. At this time, we are mainly concerned with the value of τ_{mix} in extreme cases, rather than $\Delta\tau_{\text{mix}}$. This is exactly what we do in the real data analysis later. In order to simplify the representation in matrix operations and without loss of generality, we may as well assume that the waiting time of node n is adjusted to the extreme case.

4.2 Upper bound of mixing time

When the waiting time distribution $\rho'(t)$ of node n decelerates infinitely, causing $\mu' \to \infty$ and $\tau'_{\text{tail}} \to \infty$. In this case, it takes an infinite time for a walker to leave node n after entering it; in other words, node n acts as a black hole. Although we may not find a root for $|\boldsymbol{Q}(s)| = 0$ within its domain when μ' is huge, i.e., for any $s > -1/min(\tau^i_{\text{tail}})$, $|\boldsymbol{Q}(s)| \neq 0$, we can directly calculate τ_{mix} when $\mu' \to \infty$ to study the mixing time of long

waiting times. In this case, equation (16) can be written as:

$$|\boldsymbol{Q}(s)| = |\boldsymbol{I}_n - \begin{bmatrix} \hat{\rho_1}(s) & 0 & \cdots & 0 \\ 0 & \hat{\rho_2}(s) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{P}| = 0. \tag{62}$$

For the situation discussed at the beginning of Section 2: the waiting time distributions of all nodes are the same, modify the waiting time distribution of the nth node. We can get:

$$|\boldsymbol{Q}(s)| = |\boldsymbol{I}_n - \begin{bmatrix} \hat{\rho}(s) & 0 & \cdots & 0 \\ 0 & \hat{\rho}(s) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{P}| = 0 \tag{63}$$

$$\Downarrow$$

$$|I_{n-1} - \begin{bmatrix} \hat{\rho}(s) & 0 & \cdots & 0 \\ 0 & \hat{\rho}(s) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\rho}(s) \end{bmatrix} \boldsymbol{M}_{nn}| = 0, \tag{64}$$

where $\boldsymbol{M}_{nn}$ denotes the algebraic cofactor matrix of P_{nn}. The mixing time in this case is:

$$\tau_{\text{mix}} = -\frac{1}{\rho^{-1}(1/\lambda_1(\boldsymbol{M}_{nn}))}. \tag{65}$$

Intuitively, Eq. (65) represents the scenario where Waiting time tends to be infinite, and $\Delta\tau_{\text{mix}}$ should also reach its upper bound:

$$\Delta\tau_{\text{mix}} = -\frac{1}{\rho^{-1}(1/\lambda_1(\boldsymbol{M}_{nn}))} + \frac{1}{\rho^{-1}(1/\lambda_1(\boldsymbol{P}))}. \tag{66}$$

4.3 Lower bound of mixing time

When the waiting time distribution $\rho'(t)$ of node n undergoes infinite acceleration, causing $\mu' \to 0$, $\tau'_{\text{tail}} \to 0$, in this case, $\rho'(s) = 1$, $|\boldsymbol{Q}(s)| = 0$ can be expressed as:

$$|\boldsymbol{I}_n - \begin{bmatrix} \hat{\rho_1}(s) & 0 & \cdots & 0 \\ 0 & \hat{\rho_2}(s) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \boldsymbol{P}| = 0. \tag{67}$$

For the situation discussed at the beginning of Section 2: the waiting time distributions of all nodes are the same, modify the waiting time distribution of the nth node. We can get:

$$\begin{aligned}|\boldsymbol{Q}(s)| &= |\boldsymbol{I}_n - \begin{bmatrix} \hat{\rho}(s) & 0 & \cdots & 0 \\ 0 & \hat{\rho}(s) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \boldsymbol{P}| \\ &= |I_{n-1} - \hat{\rho}(s)\boldsymbol{P}^{\text{lower bound}}| \\ &= 0,\end{aligned}$$

with

$$P^{\text{lower bound}} = \begin{bmatrix} \boldsymbol{P}_{1,n}\boldsymbol{P}_{n,1} & \boldsymbol{P}_{1,2} + \boldsymbol{P}_{1,n}\boldsymbol{P}_{n,2} & \cdots & \boldsymbol{P}_{1,n-1} + \boldsymbol{P}_{1,n}\boldsymbol{P}_{n,n-1} \\ \boldsymbol{P}_{2,1} + \boldsymbol{P}_{2,n}\boldsymbol{P}_{n,1} & \boldsymbol{P}_{2,n}\boldsymbol{P}_{n,2} & \cdots & \boldsymbol{P}_{2,n-1} + \boldsymbol{P}_{2,n}\boldsymbol{P}_{n,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{P}_{n-1,1} + \boldsymbol{P}_{n-1,n}\boldsymbol{P}_{n,1} & \boldsymbol{P}_{n-1,2} + \boldsymbol{P}_{n-1,n}\boldsymbol{P}_{n,2} & \cdots & \boldsymbol{P}_{n-1,n}\boldsymbol{P}_{n,n-1} \end{bmatrix}. \tag{68}$$

From the above equations, it can be observed that when $\rho'(t)$ is accelerated to infinity, node n acts as a hub. This implies that signals pass through this node without delay; we can rewrite $\boldsymbol{P}^{\text{lower bound}}$ as $\boldsymbol{M}_{nn} + \boldsymbol{P}_{n\cdot} \cdot \boldsymbol{P}_{\cdot n}^T$, where $\boldsymbol{P}_{n\cdot}$ is obtained as an $(n-1)$-dimensional vector by removing $\boldsymbol{P}_{n,n}$ from the n-th row of matrix $\boldsymbol{P}$, $\boldsymbol{P}_{\cdot n}$ is derived as an $(n-1)$-dimensional vector by removing $\boldsymbol{P}_{n,n}$ from the n-th column of $\boldsymbol{P}$. Thus,

$$\tau_{\text{mix}} = -\frac{1}{\rho^{-1}(1/\lambda_1(\boldsymbol{M}_{nn} + \boldsymbol{P}_{n\cdot} \cdot \boldsymbol{P}_{\cdot n}^T))}. \tag{69}$$

Intuitively, Eq. (65) represents the scenario where Waiting time tends to be zero, and $\Delta\tau_{\text{mix}}$ should also reach its lower bound:

$$\Delta\tau_{\text{mix}} = -\frac{1}{\rho^{-1}(1/\lambda_1(\boldsymbol{M}_{nn} + \boldsymbol{P}_{n\cdot} \cdot \boldsymbol{P}_{\cdot n}^T))} + \frac{1}{\rho^{-1}(1/\lambda_1(\boldsymbol{P}))}. \tag{70}$$

4.4 Upper bound of nodes in configuration model

We consider the approximate relationship between the upper bound of a node and the node degree in a configuration model with a large number of nodes. Let G be an undirected graph generated by the configuration model, with $\boldsymbol{P}$ denoting the random walk transition matrix. For each node i, the transition probability to each of its neighbors is $1/d_i$, where d_i is the degree of node i. The matrix $\boldsymbol{P}$ thus satisfies:

- Each row sums to 1;
- $P_{ii} = 0$ (no self-loops);
- $\boldsymbol{P}_{ij} = 1/d_i$ if nodes i and j are connected.

Let $\boldsymbol{M}_{kk}$ denote the *principal submatrix* of $\boldsymbol{P}$ obtained by removing the k-th row and column. Let $\lambda_1(\boldsymbol{M}_{kk})$ denote the largest eigenvalue of $\boldsymbol{M}_{kk}$. We aim to show that, in the configuration model, the following approximation holds:

$$\lambda_1(\boldsymbol{M}_{kk}) \approx 1 - \frac{d_k}{\sum_i d_i}.$$

By the Perron–Frobenius theorem, the largest eigenvalue of $\boldsymbol{P}$ is $\lambda_1(\boldsymbol{P}) = 1$, which is simple, and its corresponding left eigenvector is the stationary distribution:

$$\pi_i = \frac{d_i}{\sum_j d_j}.$$

When node k is removed, the resulting matrix $\boldsymbol{M}_{kk}$ becomes substochastic (rows corresponding to neighbors of k now sum to less than 1), hence its spectral radius strictly satisfies $\lambda_1(\boldsymbol{M}_{kk}) < 1$. Let $\boldsymbol{A}$ be the adjacency matrix of G, and let $\boldsymbol{D} = \mathrm{diag}(d_1, d_2, \ldots, d_n)$. Define the symmetric normalized adjacency matrix $\boldsymbol{S} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$, which is similar to $\boldsymbol{P}$ and hence shares the same eigenvalues. Consider the vector $x \in \mathbb{R}^{n-1}$ defined by $x_i = \sqrt{d_i}$ for all $i \neq k$. Then the Rayleigh quotient of x with respect to $\boldsymbol{S}_{kk}$ (the $(n-1) \times (n-1)$ matrix obtained from S by removing the k-th row and column) is:

$$R(x) = \frac{x^\top \boldsymbol{S}_{kk} x}{x^\top x} = \frac{\sum_{i\neq k, j\neq k} \sqrt{d_i} \frac{A_{ij}}{\sqrt{d_i d_j}} \sqrt{d_j}}{\sum_{i\neq k} d_i} = \frac{\sum_{i,j\neq k} A_{ij}}{\sum_{i\neq k} d_i} \tag{71}$$

$$= \frac{2(E - d_k)}{\sum_i d_i - d_k} = 1 - \frac{d_k}{\sum_i d_i - d_k}, \tag{72}$$

where $\boldsymbol{E} = \sum_i d_i/2$ is the total number of edges in G. This gives a lower bound for the largest eigenvalue:

$$\lambda_1(\boldsymbol{M}_{kk}) \geq 1 - \frac{d_k}{\sum_i d_i - d_k}.$$

Noting that $\lambda_1(\boldsymbol{M}_{kk}) < 1$, and the above lower bound is tight when degrees are homogeneous, we conclude the approximation:

$$\lambda_1(\boldsymbol{M}_{kk}) \approx 1 - \frac{d_k}{\sum_i d_i}.$$

We now quantify the error by expanding the denominator:

$$1 - \frac{d_k}{\sum_i d_i - d_k} = 1 - \frac{d_k}{\sum_i d_i} \cdot \frac{1}{1 - \frac{d_k}{\sum_i d_i}} \tag{73}$$

$$= 1 - \frac{d_k}{\sum_i d_i} - \frac{d_k^2}{(\sum_i d_i)^2} + O\left(\frac{d_k^3}{(\sum_i d_i)^3}\right). \tag{74}$$

Therefore, the approximation error is of order $O(d_k^2/(\sum_i d_i)^2)$. In sparse configuration models where node degrees remain bounded or have finite mean, the total degree $\sum_i d_i = O(n)$, while $d_k = O(1)$ for typical nodes. It follows that

$$\lambda_1(\boldsymbol{M}_{kk}) = 1 - \frac{d_k}{\sum_i d_i} + O\left(\frac{1}{n^2}\right),$$

so the approximation becomes asymptotically exact as $n \to \infty$. We conclude that, under the configuration model, the largest eigenvalue of the principal submatrix $\boldsymbol{M}_{kk}$ satisfies

$$\lambda_1(\boldsymbol{M}_{kk}) = 1 - \frac{d_k}{\sum_i d_i} + O\left(\frac{1}{n^2}\right).$$

For large networks, we observe that $\lambda_1(\boldsymbol{M}_{kk}) \approx 1$, which implies $\frac{1}{\lambda_1(\boldsymbol{M}_{kk})} \approx 1$. Consequently, the solution s satisfying Eq. $\hat{\rho}(s) = \frac{1}{\lambda_1(\boldsymbol{M}_{kk})}$ lies near $s = 0$. This justifies the approximation $\hat{\rho}(s) \approx 1 - \mu s$. In this case, the upper bound (66) can be further approximated as:

$$\Delta\tau_{\mathrm{mix}} \approx \mu \frac{\sum_i d_i - d_k}{d_k} + \frac{1}{\hat{\rho}^{-1}(1/\lambda_1(\boldsymbol{P}))}. \tag{75}$$

4.5 Lower bound of nodes in large network

Now we consider the approximate value of the lower bound when the number of network nodes is large. We demonstrate that the lower bound asymptotically approaches zero in the context of large-scale networks. Given that

$$\Delta\tau_{\text{mix}} = \frac{-1}{\rho^{-1}\left(1/\lambda_1(\boldsymbol{M}_{kk} + \boldsymbol{P}_{k\cdot} \cdot \boldsymbol{P}_{\cdot k}^T)\right)} + \frac{1}{\rho^{-1}\left(1/\lambda_1(\boldsymbol{P})\right)}, \tag{76}$$

the $\Delta\tau_{\text{mix}} \approx 0$ is equivalent to the condition $\lambda_1(\boldsymbol{P}) \approx \lambda_1(\boldsymbol{M}_{kk} + \boldsymbol{P}_{k\cdot} \cdot \boldsymbol{P}_{\cdot k}^T)$, and the following theorem will prove it. Without loss of generality and for analytical simplicity, we assume $k = n$, which can be achieved through proper node relabeling.

Theorem 3. *Let $\boldsymbol{P}$ be an $n \times n$ Markov transition matrix for a random walk on a network, where: 1. $\boldsymbol{P}_{ii} = 0$ for all i, 2. Each row sum of $\boldsymbol{P}$ equals 1.*

Define the $(n-1) \times (n-1)$ matrix H by:

$$\boldsymbol{H}_{ij} = \boldsymbol{P}_{ij} + \boldsymbol{P}_{in}\boldsymbol{P}_{nj} \quad \text{for } 1 \le i, j \le n-1.$$

Assume that as $n \to \infty$, the transition probabilities to/from node n satisfy:

$$\max_{1 \le i \le n-1} \boldsymbol{P}_{in} \approx 0 \quad \text{and} \quad \max_{1 \le j \le n-1} \boldsymbol{P}_{nj} \approx 0.$$

Then, the second-largest eigenvalues of $\boldsymbol{P}$ and H satisfy:

$$\lambda_1(\boldsymbol{P}) \approx \lambda_1(\boldsymbol{H}) \quad \text{as } n \to \infty.$$

Proof. We partition $\boldsymbol{P}$ into submatrices:

$$P = \begin{bmatrix} \boldsymbol{M}_{nn} & \boldsymbol{P}_{\cdot n} \\ \boldsymbol{P}_{n\cdot}^\top & 0 \end{bmatrix},$$

where $\boldsymbol{M}_{nn}$ is the $(n-1) \times (n-1)$ submatrix excluding the last row and column, $\boldsymbol{P}_{n\cdot}, \boldsymbol{P}_{\cdot n}$ are the column vectors. By definition, $H = \boldsymbol{M}_{nn} + \boldsymbol{P}_{n\cdot}\boldsymbol{P}_{\cdot n}^\top$. Since $\boldsymbol{P}$ is a Markov matrix, H is also a Markov matrix:

$$\sum_{j=1}^{n-1} \boldsymbol{H}_{ij} = \sum_{j=1}^{n-1} (\boldsymbol{P}_{ij} + \boldsymbol{P}_{in}\boldsymbol{P}_{nj}) = (1 - \boldsymbol{P}_{in}) + \boldsymbol{P}_{in}(1 - \boldsymbol{P}_{nn}) = 1 \quad (\text{since } \boldsymbol{P}_{nn} = 0).$$

The perturbation matrix $\boldsymbol{E} = \boldsymbol{P}_{n\cdot}\boldsymbol{P}_{\cdot n}^\top$ has entries:

$$\boldsymbol{E}_{ij} = \boldsymbol{P}_{in}\boldsymbol{P}_{nj}.$$

Under the assumption $\max_i \boldsymbol{P}_{in} \approx 0$ and $\max_j \boldsymbol{P}_{nj} \approx 0$, the spectral norm of E satisfies:

$$\begin{aligned} \|\boldsymbol{E}\|_2 \le \|\boldsymbol{P}_{n\cdot}\|_2 \|\boldsymbol{P}_{\cdot n}\|_2 &= \sqrt{\sum_{i=1}^{n-1} \boldsymbol{P}_{in}^2} \cdot \sqrt{\sum_{j=1}^{n-1} \boldsymbol{P}_{nj}^2} \\ &\le \sqrt{(n-1) \cdot \left(\max_i \boldsymbol{P}_{in}\right)^2} \cdot \sqrt{(n-1) \cdot \left(\max_j \boldsymbol{P}_{nj}\right)^2} \\ &= (n-1) \cdot \max_i \boldsymbol{P}_{in} \cdot \max_j \boldsymbol{P}_{nj} \approx 0. \end{aligned}$$

By the Bauer-Fike theorem, if $\boldsymbol{M}_{nn}$ is diagonalized as $\boldsymbol{M}_{nn} = V\Lambda V^{-1}$, then for any eigenvalue μ of H:

$$\min_i |\lambda_i(\boldsymbol{H}) - \lambda_i(\boldsymbol{M}_{kk})| \le \kappa(V)\|\boldsymbol{E}\|_2,$$

where $\kappa(V) = \|V\|_2\|V^{-1}\|_2$ is the condition number of V, it remains bounded, and we have $\lambda_1(\boldsymbol{H}) \approx \lambda_2(\boldsymbol{M}_{nn})$. The eigenvalues of $\boldsymbol{P}$ satisfy:

$$\det(P - \lambda \boldsymbol{I}) = \det\left(\begin{bmatrix} \boldsymbol{M}_{nn} - \lambda \boldsymbol{I} & \boldsymbol{P}_{\cdot n} \\ \boldsymbol{P}_{n\cdot}^\top & -\lambda \end{bmatrix}\right) = -\lambda \det(\boldsymbol{M}_{nn} - \lambda \boldsymbol{I}) - \boldsymbol{P}_{n\cdot}^\top \operatorname{adj}(\boldsymbol{M}_{nn} - \lambda \boldsymbol{I}) \boldsymbol{P}_{\cdot n} = 0,$$

where $\operatorname{adj}\{\boldsymbol{M}_{nn} - \lambda \boldsymbol{I}\}$ denotes the adjugate matrix of $\boldsymbol{M}_{nn} - \lambda \boldsymbol{I}$. As $n \to \infty$, $\|\boldsymbol{P}_{\cdot n}\|, \|\boldsymbol{P}_{n\cdot}\| \approx 0$, so the term $\boldsymbol{P}_{n\cdot}^\top \operatorname{adj}(\boldsymbol{M}_{nn} - \lambda \boldsymbol{I}) \boldsymbol{P}_{\cdot n}$ becomes negligible. Thus, the non-zero eigenvalues of $\boldsymbol{P}$ approximate those of $\boldsymbol{M}_{nn}$:

$$\lambda \det(\boldsymbol{M}_{nn} - \lambda \boldsymbol{I}_{n-1}) \approx 0 \implies \lambda_i(P) \approx \lambda_{i+1}(\boldsymbol{M}_{nn}),$$

in particular, $\lambda_1(\boldsymbol{P}) \approx \lambda_2(\boldsymbol{M}_{nn})$. Since both $\lambda_1(\boldsymbol{P})$ and $\lambda_1(\boldsymbol{H})$ approximate $\lambda_2(\boldsymbol{M}_{nn})$, we conclude:

$$\lambda_1(\boldsymbol{P}) \approx \lambda_1(\boldsymbol{H}) \quad \text{as } n \to \infty.$$

□

4.6 Node-level spectral and dynamical approximations

A central approximation used in the main text is the node-level eigenvalue estimate $\lambda_1(\mathbf{M}_{kk}) \approx 1 - d_k/(2m)$ and the resulting $\Delta\tau_{\text{mix}}$ in Eqs. (10–11) of the main text. To strengthen the evidence for these approximations in finite-size networks, this section provides a quantitative assessment of two potential sources of deviation: (i) finite network size N, and (ii) nonzero clustering C, which introduces higher-order correlations not explicitly retained by the configuration-model baseline. We performed systematic numerical experiments on ensembles of Erdős–Rényi (ER) and Barabási–Albert (BA) networks. All networks were generated with fixed mean degree $\langle k \rangle = 6$, and network sizes were varied over $N \in \{20, 50, 100, 150, 200, 300, 400\}$. For each network type and each value of N, we generated 50 independent realizations and retained only connected graphs. For ER networks, graphs were constructed using the $G(N, m)$ model with $m = \text{round}(N\langle k \rangle/2)$. For BA networks, an initial preferential-attachment graph with parameter $m_{\text{BA}} = \max(1, \langle k \rangle/2)$ was generated and, when necessary, additional edges were added uniformly at random to match the target edge count while preserving connectivity.

To probe the influence of clustering at fixed degree sequence, we additionally performed degree-preserving rewiring sweeps at fixed $N = 100$. Starting from each baseline realization, we apply repeated double-edge swaps (two candidate pairings per trial) that preserve every node degree exactly and reject any move that disconnects the graph. For the increase sweep, we retain only moves that strictly increase the best-so-far global transitivity, producing networks with progressively larger achieved clustering coefficients. For the decrease sweep, we retain only moves that strictly decrease the best-so-far transitivity (with the additional constraint that clustering never exceeds its baseline value), thereby exploring networks with suppressed triangle density while keeping the original degree distribution. In all cases, the achieved global clustering coefficient C is measured a posteriori and used for analysis. Here C denotes the **global transitivity** (global clustering coefficient). For each generated graph and for every node k, we compute the exact leading eigenvalue $\lambda_1(\mathbf{M}_{kk})$ of the node-deleted transition matrix. We then form the degree-based approximation

$$\lambda_1^{\text{approx}}(\mathbf{M}_{kk}) \approx 1 - \frac{d_k}{2m},$$

(using d_k consistently, and writing the degree sum as $2m$). We quantify spectral accuracy via the node-level relative error

$$\epsilon_k^\lambda = \left|\lambda_1^{\text{exact}}(\mathbf{M}_{kk}) - \lambda_1^{\text{approx}}(\mathbf{M}_{kk})\right| \Big/ \left|\lambda_1^{\text{exact}}(\mathbf{M}_{kk})\right|.$$

For the dynamical quantity in Eqs. (10–11) of the main text, we use the exact and approximate forms

$$\Delta\tau_{\text{mix}}^{\text{exact}} = \frac{1}{1 - \lambda_1^{\text{exact}}(\mathbf{M}_{kk})}, \qquad \Delta\tau_{\text{mix}}^{\text{approx}} = \frac{2m}{d_k} - 1,$$

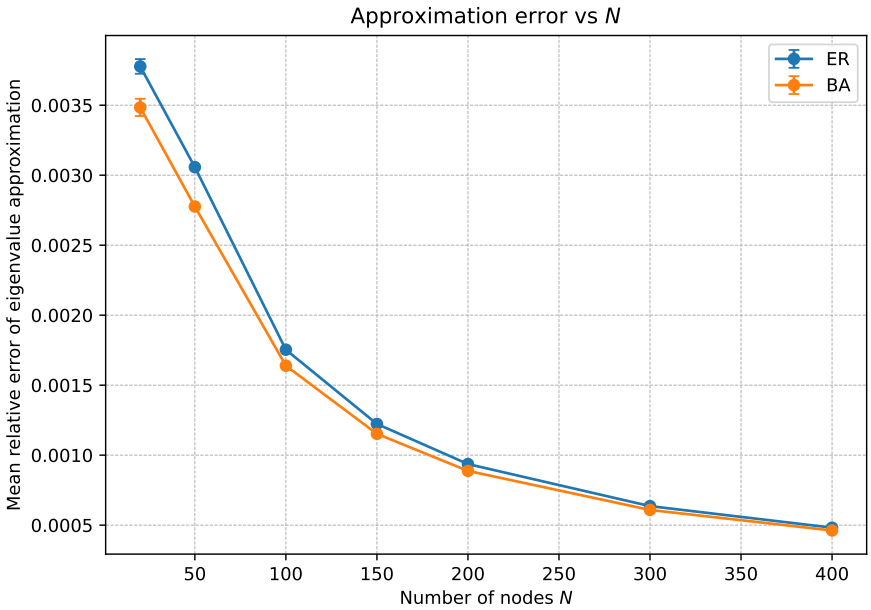

Fig. S7. Mean relative error of the node-level eigenvalue approximation versus network size N. Error bars indicate the standard error over 50 independent realizations, showing systematic improvement with increasing N.

and define the corresponding node-level relative error

$$\epsilon_k^{\Delta\tau} = \left|\Delta\tau_{\text{mix}}^{\text{exact}} - \Delta\tau_{\text{mix}}^{\text{approx}}\right| / \left|\Delta\tau_{\text{mix}}^{\text{exact}}\right|.$$

For each realization, we record the mean of ϵ_k^{λ} and $\epsilon_k^{\Delta\tau}$ over nodes. For the network-size scan (Figs. S7 and S9), we then average these per-realization means over 50 independent realizations for each (type, N), with error bars indicating the standard error over realizations. For the clustering sweeps at fixed $N = 100$ (Figs. S8 and S10), we start from 10 baseline realizations per network type and aggregate the resulting rewired samples within each C-bin. In the clustering sweeps ($N = 100$), we report *binned* averages as a function of C using bins centered at 0.01ℓ with half-width 0.005 (as in the plotted figures), where $\ell \in \mathbb{Z}_{\geq 0}$ (i.e., $\ell = 0, 1, 2, \ldots$) is the bin index. Each circular marker corresponds to the mean relative error of all sampled networks whose achieved clustering falls in that bin, and the vertical error bar indicates the standard error of the mean within the bin. Baseline realizations (i.e., the original ER/BA networks before any rewiring) are overlaid as cross markers to indicate the starting low-clustering regime from which the degree-preserving sweeps depart. When the within-bin variability is small, the error bars can be visually shorter than the marker size.

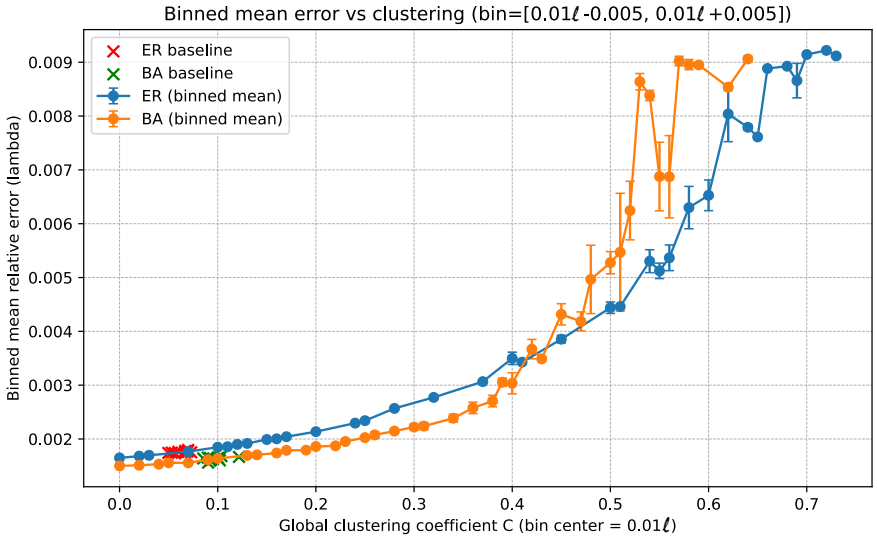

Fig. S8. Binned mean relative error of the eigenvalue approximation versus global clustering coefficient C at fixed $N = 100$ (degree-preserving rewiring sweeps). Circular markers report the mean within each C-bin (center 0.01ℓ, half-width 0.005) and error bars indicate the standard error within the bin. Cross markers show the baseline ER/BA realizations prior to rewiring.

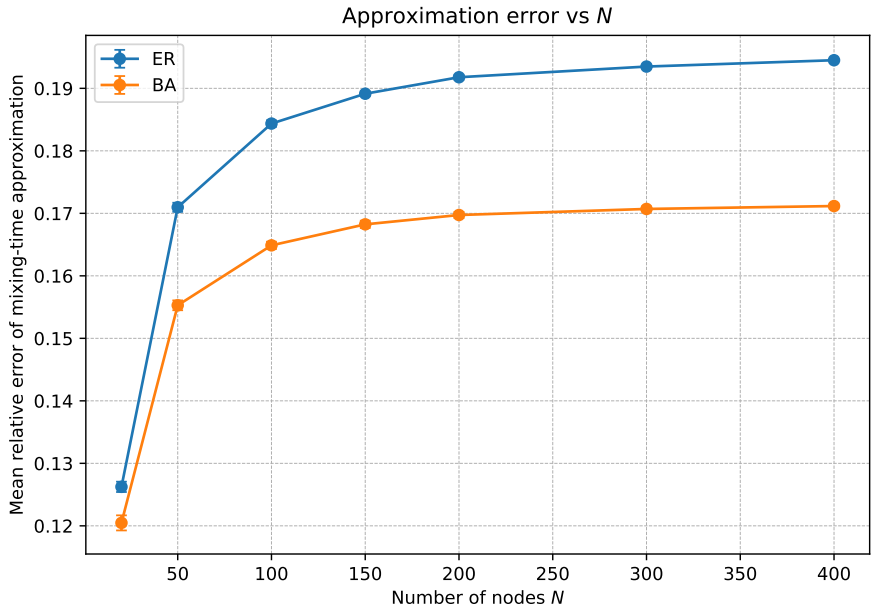

Fig. S9. Mean relative error of the $\Delta\tau_{\mathrm{mix}}$ approximation (Eqs. (10–11) in the main text) versus network size N (ER and BA, $\langle k\rangle = 6$). Error bars indicate the standard error over 50 independent realizations. Although the eigenvalue approximation improves with N, the nonlinear mapping from λ_1 to $\Delta\tau_{\mathrm{mix}}$ can amplify residual discrepancies, yielding a moderate error that varies only weakly at larger N.

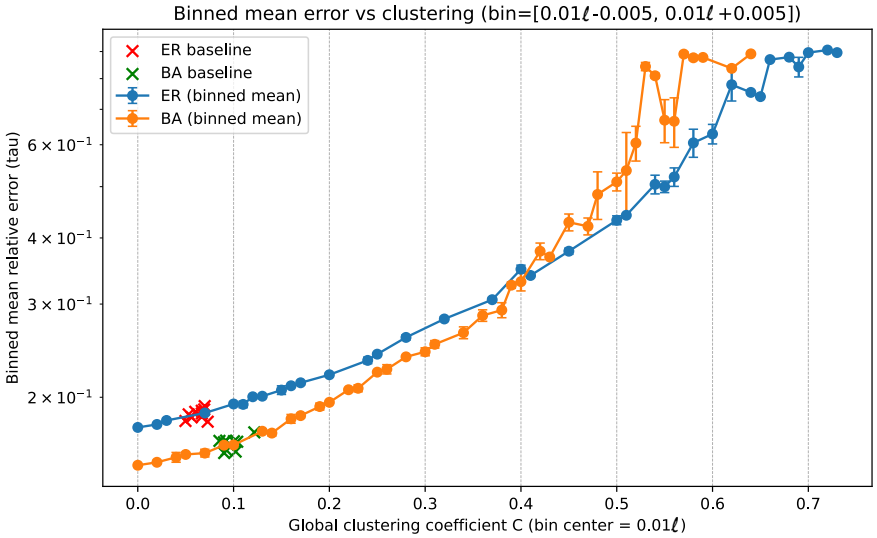

Fig. S10. Binned mean relative error of the $\Delta\tau_{\mathrm{mix}}$ approximation versus global clustering coefficient C at fixed $N = 100$ (degree-preserving rewiring sweeps). Markers and error bars are defined as in Fig. S8, with cross markers indicating the baseline networks before rewiring. The error increases rapidly at large C, quantifying the loss of accuracy in highly clustered regimes.

Fig. S7 shows that the mean node-level relative error for $\lambda_1(\mathbf{M}_{kk})$ exhibits a clear decreasing trend with increasing N for both ER and BA ensembles. In particular, for the $N = 100$ regime used in the main text, the error is already small and continues to contract as N grows. Fig. S8 isolates the impact of clustering at fixed $N = 100$ and fixed degree sequence: as C is increased by degree-preserving rewiring, the spectral approximation error rises in a systematic manner. This trend is consistent with the interpretation that the estimate $\lambda_1(\mathbf{M}_{kk}) \approx 1 - d_k/(2m)$ is a configuration-model baseline that is most accurate when triangle-induced correlations remain weak. The overlaid baseline points (cross markers) cluster in the low-C region, making explicit that the original ER/BA ensembles at $N = 100$ occupy the parameter range where the configuration-model baseline is most accurate. Although ER and BA networks have vanishing clustering in the large-N limit, this visualization shows that the unrewired $N = 100$ realizations used in the main text already lie in a weakly clustered regime, whereas the degree-preserving sweeps probe how accuracy degrades when C is driven far beyond that baseline range. This clarifies why ER/BA provide an appropriate low-clustering testbed for the configuration-model baseline at the system sizes considered in the main text.

The mixing-time increment depends on the inverse spectral gap, which amplifies small deviations in $\lambda_1(\mathbf{M}_{kk})$ when λ_1 is close to unity. Accordingly, Fig. S9 shows that the relative error for $\Delta\tau_{\mathrm{mix}}$ does not necessarily decrease with N even though the spectral approximation improves: the nonlinear transformation $\Delta\tau_{\mathrm{mix}} = 1/(1 - \lambda_1)$ can magnify residual eigenvalue-level discrepancies. At the same time, the finite-size trend becomes weak once N is moderately large: across the larger-N range in Fig. S9, the mean relative error remains confined to a relatively narrow band for both ER and BA ensembles, indicating that the dominant source of dynamical error is the gap-amplification mechanism rather than strong additional N-dependent bias. At fixed $N = 100$, Fig. S10 further demonstrates that increasing clustering substantially degrades the dynamical approximation, with a pronounced rise in error as C becomes large. This provides a direct, quantitative characterization of the regime in which triangle-driven correlations are no longer secondary corrections to the configuration-model baseline. In contrast, the same plots show that once the degree-preserving rewiring drives C far beyond the baseline regime, the dynamical approximation can deteriorate sharply, thereby providing an operational

indication of the regime in which Eqs. (10–11) remain accurate.

Together, these experiments validate the intended domain of the node-level approximations used in the main text. The spectral estimate $\lambda_1(\mathbf{M}_{kk}) \approx 1 - d_k/(2m)$ converges with network size and remains accurate in weakly clustered networks. The corresponding $\Delta\tau_{\mathrm{mix}}$ formula in Eqs. (10–11) is inherently more sensitive due to spectral-gap amplification, and the clustering sweeps make explicit that large C can drive a clear loss of accuracy. Importantly, the baseline ER/BA ensembles at $N = 100$ and $\langle k \rangle = 6$ lie in the low-clustering regime where the configuration-model baseline provides a meaningful first-order description of the node-level spectral and dynamical quantities analyzed in the paper.

5 Applying to Mice brain

Equation (2) can describe the complex connectivity patterns and significant heterogeneity of networks in different environments such as the Internet, society, and biology. From information to viral transmission, a wide range of real-world dynamics are related to diffusion. Existing works use a simplified version of Eq. (2) - a system of linear ordinary differential equations to model the spread of α-synuclein pathology through the brain connectome [1], the spread of tau pathology in the mouse brain [2], and Disease Progression in Dementia [3]. The cases considered in these works correspond to the case where the waiting time distribution is exponentially distributed, but the spread of microscopic pathology does not necessarily follow the memoryless property. Although these works have successfully modeled the spread of pathology in the mouse and human brain, considering more sophisticated dynamics can help us understand the pathology spread more deeply.

Gamma distribution is suitable for describing multi-stage cumulative random processes and is widely used in biological systems. Existing works[4] [5] use Gamma distribution to model the distribution of apparent diffusion coefficient (ADC), which more accurately characterizes the non-Gaussian characteristics of DWI signal attenuation than single exponential distribution and provides interpretable biophysical parameters for tissue microstructure. Another work [6] used Gamma distribution to model multi-compartment microstructure in brain tissue T2 relaxometry MRI data, and achieved a tissue differentiation effect that is more consistent with neuropathological findings than exponential distribution.

Our idea is to first fit the diffusion model of exponential and Gamma waiting time distribution respectively through the data, then compare the fitting effect of the model, and finally compare the upper bound and lower bound of the nodes of different models. Note that the upper and lower bound here refer to the mixing time when the waiting time of the nodes (brain regions) tend to infinity (upper bound) and tend to 0 (lower bound).

5.1 Exponential model

We first consider the case where the waiting time distribution follows an exponential distribution, and the diffusion equation (2) can be rewritten as

$$\frac{d\boldsymbol{x}(t)}{dt} = \text{diag}\{\frac{1}{\mu_1}, \cdots, \frac{1}{\mu_n}\}(\boldsymbol{P} - \boldsymbol{I})\boldsymbol{x}(t), \tag{77}$$

where the vector $\boldsymbol{x}(t)$ corresponds to the magnitude of α-synuclein pathology, the initial injection of α-synuclein is on the s-th region, i.e., the s-th element of $x(0)$ equal to 1; otherwise, it is 0. There are $n = 116$ brain regions. The strength of the connections between regions is quantified by fluorescence intensity measurements from retrograde viral tract tracing, which provides the diffusion transition matrix P. The corresponding dataset is available at `https://github.com/ejcorn/connectome_diffusion`.

From experiments, the regional pathology is averaged across all mice at 1, 3 and 6 Months Post-Injection(MPI) as observed data. To maximize the correlation between the simulated pathology and the observed pathology, we optimize the values of the elements of the proportional coefficients $\text{diag}\{\frac{1}{\mu_1}, \cdots, \frac{1}{\mu_n}\}$. The logarithmic plot of the experimental data and the model results is shown in Fig. S11. The diffusion model when the waiting time distributions are exponential corresponds to Opti-M in the main text.

5.2 Baseline Model: Synu-M

The Synu-M model replicates the framework of Henderson et al. [1], serving as our biologically informed baseline. In this model, the network topology is represented by the transition matrix P, which is obtained by normalizing the raw structural connectivity matrix of the mouse brain. To integrate regional biological heterogeneity, the mean waiting time μ_i for each region is not only determined by the inverse of mRNA expression, but also adjusted to compensate for the total incoming/outgoing connection weights of that region in the raw connectome. Specifically, μ_i is calculated by scaling the endogenous mRNA concentration with the regional

connectivity strength, ensuring that the parameters of the waiting-time distribution account for both the local biochemical environment and the physical scale of the structural links. This method provides a null model that honors the original study's weighting without additional free parameter fitting.

5.3 Gamma model

Next, we consider the case where the waiting time distributions follows Gamma distributions, and the diffusion equation is rewritten as

$$\boldsymbol{x}(t) = \int_t^\infty \boldsymbol{\rho}(\tau)x(0)d\tau + \int_0^t \boldsymbol{\rho}(\tau)\boldsymbol{P}x(t-\tau)d\tau, \tag{78}$$

where $\boldsymbol{\rho}(t) = \{\rho_1(t), \cdots, \rho_n(t)\}$, $\rho_i(t) = \frac{t^{\mu_i/\tau_i - 1}}{\Gamma(\mu_i/\tau_i)\tau_i^{\mu_i/\tau_i}} e^{-t/\tau_i}$. Numerically we solve the above integral equation to obtain $\boldsymbol{x}(t)$. Then, we optimize the parameters μ_i, τ_i of the waiting time distribution to achieve the highest correlation between the simulated and the observed pathology, as shown in Fig. S11. The diffusion model when the waiting time distributions are Gamma distributions corresponds to Gamma-M in the main text.

5.4 Optimization and calibration details

To calibrate the parameters in the high-dimensional space of Opti-M (58 independent parameters) and Gamma-M (116 independent parameters), we implement a greedy coordinate-wise optimization (hill-climbing) procedure. Since the experimental data provides pathology measurements at $1, 3$, and 6 Months Post-Injection (MPI), our optimization objective is to maximize the average Pearson correlation coefficient across these three time points:

$$\bar{r} = \frac{1}{3} \sum_{t \in \{1,3,6\}} r_t. \tag{79}$$

The optimization procedure is implemented as follows:

1. **Initialization:** All parameters (μ_i and τ_i) are initially set to 1.0.
2. **Iterative Update for μ_i:** For each of the 58 symmetric region pairs, we test an incremental change of step size $\Delta\mu = 0.05$. In each iteration, we evaluate the gradient of $\bar{r}$ with respect to each region pair and update the specific pair that provides the maximum increase in $\bar{r}$. Because the global time scale is determined by selecting the best-fitting simulated time point, the optimization focuses on the relative regional heterogeneity of μ_i.
3. **Iterative Update for τ_i (Gamma-M):** For the Gamma model, we similarly test both increasing and decreasing the tail parameters by $\Delta\tau = 0.05$ (subject to the constraint $\tau_i > 0$). The update that maximizes the increment in $\bar{r}$ is applied.
4. **Stopping Criterion:** The optimization process terminates when the improvement in the average correlation between successive iterations, $\Delta\bar{r}$, falls below the threshold of 0.001.

5.5 Lower and upper bounds

Numerical results show that the diffusion model based on the Gamma distribution provides a better fit than the exponential distribution, and can therefore be used as a powerful tool to predict pathological protein spread. Additionally, to investigate the role of network topology in the diffusion, we further calculate the upper and lower bounds for each brain region based on both the exponential distribution and Gamma distribution. We then compare the upper and lower bound rankings of nodes between the two models (Fig. S12) and find that the upper bound rankings of the nodes are similar, while the lower bound rankings show significant differences.

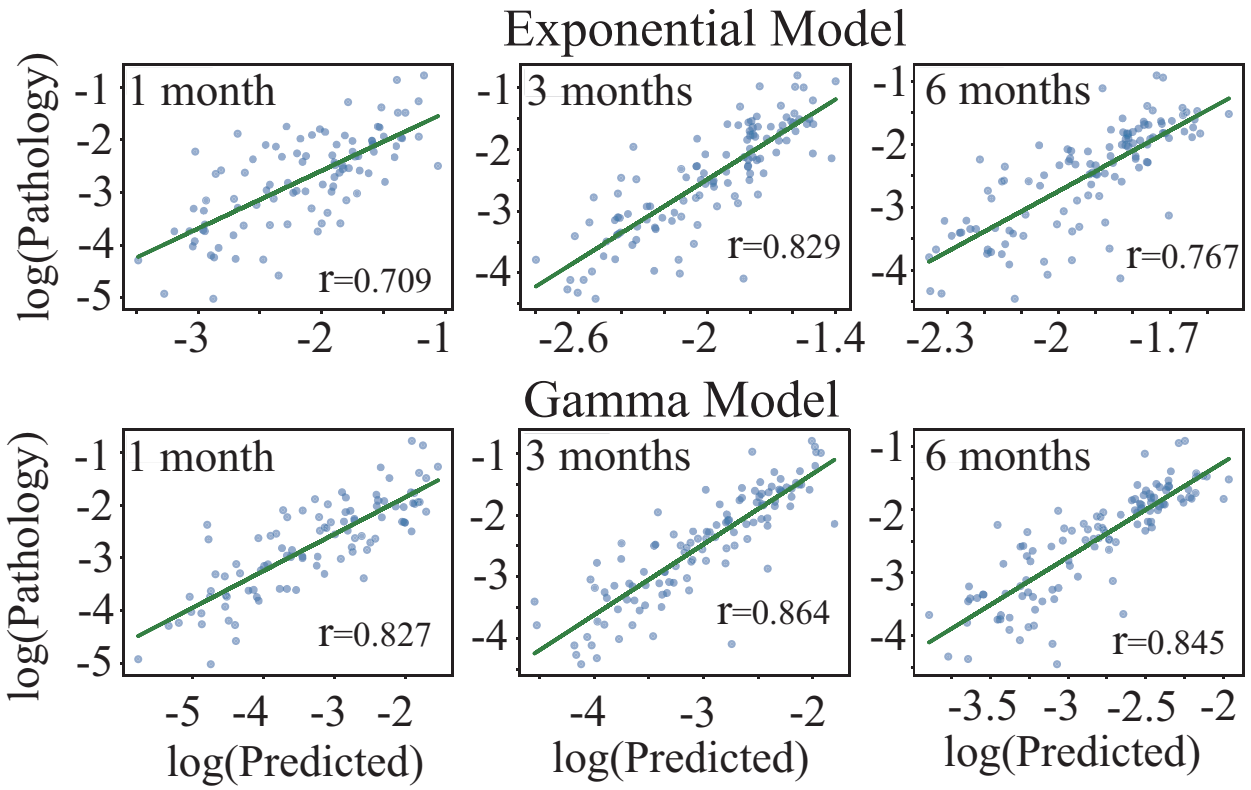

Fig. S11. Scatterplots and Pearson's correlation coefficients (r) of log predicted pathology based on exponential and Gamma models versus log actual pathology values for each region are shown for 1 month, 3 months and 6 months.

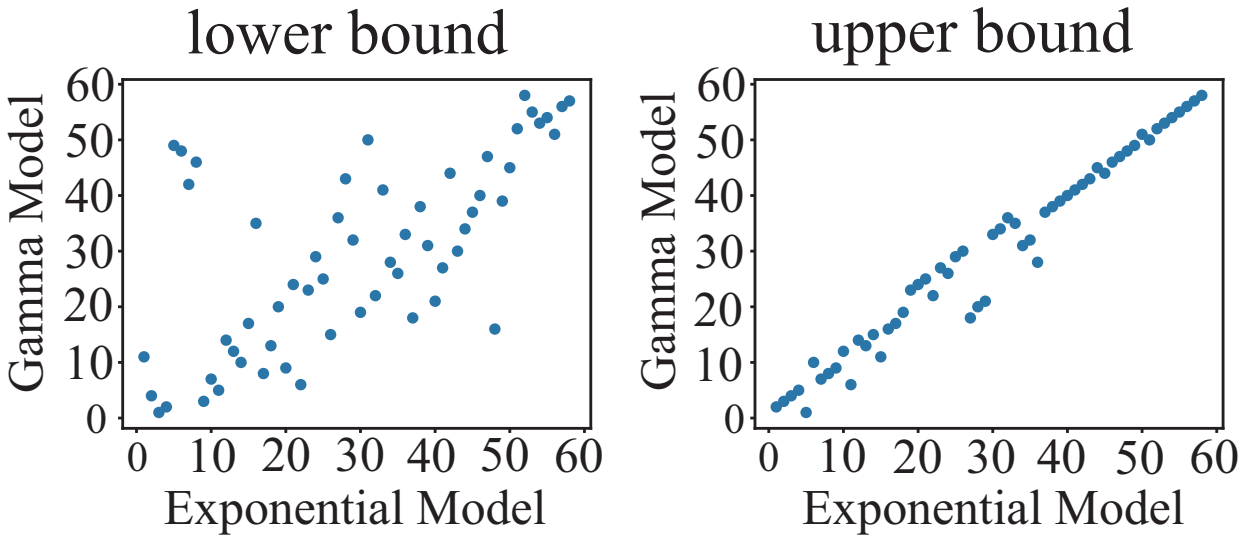

Fig. S12. Based on the exponential model and the Gamma model, nodes are sorted from small to large according to lower bound and upper bound, respectively, and the ranking of different models are used as the horizontal and vertical coordinates.

References

[1] M. X. Henderson, E. J. Cornblath, A. Darwich, B. Zhang, H. Brown, R. J. Gathagan, R. M. Sandler, D. S. Bassett, J. Q. Trojanowski, and V. M. Lee, Nature neuroscience **22**, 1248 (2019).

[2] E. J. Cornblath, H. L. Li, L. Changolkar, B. Zhang, H. J. Brown, R. J. Gathagan, M. F. Olufemi, J. Q. Trojanowski, D. S. Bassett, V. M. Lee, *et al.*, Science Advances **7**, eabg6677 (2021).

[3] A. Raj, A. Kuceyeski, and M. Weiner, Neuron **73**, 1204 (2012).

[4] K. Oshio, H. Shinmoto, and R. V. Mulkern, Magnetic Resonance in Medical Sciences **13**, 191 (2014).

[5] Z. Soleimani, M. Saboori, I. Abedi, M. Irannejad, and S. Khanbabapour, Medicine **103**, e39593 (2024).

[6] S. Chatterjee, O. Commowick, O. Afacan, S. K. Warfield, and C. Barillot, in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (IEEE, 2018) pp. 141–144.