



Article

Fractional Diffusion on Graphs: Superposition of Laplacian Semigroups Incorporating Memory

Nikita Deniskin ¹ and Ernesto Estrada ^{2,*} ¹ Faculty of Sciences, Scuola Normale Superiore, 56126 Pisa, Italy; nikita.deniskin@sns.it² Institute for Cross-Disciplinary Physics and Complex Systems (IFISC), CSIC-UIB, 07122 Palma de Mallorca, Spain

* Correspondence: estrada@ifisc.uib-csic.es

Abstract

Subdiffusion on graphs is often modeled by time-fractional diffusion equations; yet, its structural and dynamical consequences remain unclear. We show that subdiffusive transport on graphs is a memory-driven process generated by a random time change that compresses operational time, produces long-tailed waiting times, and breaks Markovianity while preserving linearity and mass conservation. While the subordination representation and complete monotonicity properties of the Mittag-Leffler function are classical, we develop a graph-based synthesis in which Mittag-Leffler dynamics admit an exact convex, mass-preserving representation as a superposition of Laplacian semigroups evaluated at rescaled times. This perspective reveals fractional diffusion as ordinary diffusion acting across multiple intrinsic time scales and enables new structural and dynamical interpretations of graphs. This framework uncovers heterogeneous, vertex-dependent memory effects and induces transport biases absent in classical diffusion, including algebraic relaxation, degree-dependent waiting times, and early-time asymmetries between sources and neighbors. These features define a subdiffusive geometry on graphs, enabling the recovery of global shortest paths, in contrast to the graph exploration of diffusive geometry, while simultaneously favoring high-degree regions. Finally, we show that time-fractional diffusion can be interpreted as a singular limit of multi-rate diffusion, in an appropriate asymptotic sense.

Keywords: subdiffusion on graphs; time-fractional diffusion; Mittag-Leffler functions; memory kernels; sum-of-exponentials; shortest paths

MSC: 26A33; 15A16; 35R11; 47D06; 45D05



Academic Editor: J. Alberto Conejero

Received: 18 March 2026

Revised: 15 April 2026

Accepted: 16 April 2026

Published: 21 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Graphs $G = (V, E)$ —also referred to as networks—provide a natural mathematical framework to represent a wide variety of complex systems arising in molecular, ecological, technological, and social contexts [1]. In this representation, the set of vertices V typically corresponds to the entities of the system, while the set of edges E encodes their interactions. A fundamental mechanism governing the transport of mass, energy, or information across such networks is diffusion [2,3]. However, due to the complexity of the environments in which many real-world networks are embedded, transport processes often deviate from classical diffusive behavior [4–7].

Subdiffusion, characterized by a slower-than-linear growth of the mean squared displacement, is especially prevalent in complex systems. For instance, subdiffusive dynamics are ubiquitous in the crowded interior of biological cells [8–11], where millions of macromolecules interact, forming intricate networks of biochemical processes. Similarly, crowding effects caused by vehicle density and driving fluctuations in urban transportation systems have been shown to induce subdiffusive traffic states [12]. Subdiffusion has also been observed in information transmission processes on online social networks, such as Twitter (now X) and Digg [13].

A standard mathematical approach to model such a crowded and heterogeneous system is to replace the classical time derivative in the diffusion equation with a fractional-time derivative [14,15]. When anomalous diffusion takes place on a graph G , this leads to the following fractional diffusion equation:

$$\begin{cases} D_t^\alpha u_\theta(t) + \theta L u_\theta(t) = 0, \\ u_\theta(0) = u_0, \end{cases} \quad (1)$$

where $\theta > 0$ denotes the diffusivity, D_t^α is the Caputo time-fractional derivative, and L is the graph Laplacian operator acting on functions defined on the vertex set. Specifically, denoting by $\mathbb{C}(V)$ the set of all complex-valued functions on V , L is the linear mapping from $\mathbb{C}(V)$ into itself given by

$$(Lf)(v) := \sum_{(v,w) \in E} (f(v) - f(w)), \quad f \in \mathbb{C}(V). \quad (2)$$

The operator D_t^α denotes the Caputo time-fractional derivative of order $0 < \alpha < 1$ [16], defined by

$$D_t^\alpha u(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{u'(\tau)}{(t-\tau)^\alpha} d\tau, \quad (3)$$

where $u'(\tau)$ is the first derivative of u evaluated at time τ . Despite their broad relevance in continuous settings, time-fractional diffusion models have been only sparsely explored on graphs and networks. To date, applications have been largely restricted to epidemiological modeling [17,18] and fractional diffusion on the human proteome, proposed as an alternative framework to account for the multi-organ damage associated with SARS-CoV-2 infection [19]. An exception is the use of (1) in an engineering context for achieving consensus in autonomous systems, where it is referred to as fractional-order consensus [20–25]. A rapidly growing line of research explores the integration of fractional derivatives into learning algorithms, leading to the development of fractional-order neural networks and related architectures, with emerging applications in machine learning and artificial intelligence [26,27].

Apart from crowding and excluded-volume effects, which reduce mobility by limiting the available space for motion, several additional mechanisms can give rise to subdiffusion in complex systems. This happens, for instance, in ecological networks, where structural disorder and heterogeneity—manifested through irregular geometries, bottlenecks, and hierarchical or fractal structures—constrain transport pathways and significantly slow down spatial exploration [28]. These features often induce trapping events and lead to broad distributions of residence times. Another characteristic of complex systems is the existence of memory. Temporal memory effects have been observed, for instance, in transcription regulator–DNA interactions in live bacterial cells [29], apart from the extensive existing evidence accumulated on single-cell experiments. As memory refers to the phenomenon where past events influence a system’s current and future states or

behaviors, the evolution of these complex systems at a given time depends on their entire past history rather than solely on their instantaneous state.

In such non-Markovian settings, classical diffusion equations fail to provide an adequate description. Instead, diffusion models with memory kernels naturally arise, capturing the nonlocal-in-time response induced by trapping, heterogeneity, and temporal correlations [30–35]. Once again, in the case of graphs, this leads to the generalized diffusion equation

$$\begin{cases} \int_0^t \gamma(t-s) \partial_s u_\theta(s) ds + \theta L u_\theta(t) = 0, \\ u_\theta(0) = u_0, \end{cases} \quad (4)$$

where $\gamma(t)$ denotes a memory kernel. Such formulations provide a unifying framework for describing subdiffusive transport and establishing a direct connection between microscopic mechanisms and macroscopic anomalous diffusion. Related non-Markovian formulations on networks arise naturally when vertex-to-vertex motion is modeled as a non-Poisson continuous-time random walk. In this setting, the node-state evolution is governed by a generalized master equation in which the graph Laplacian (or transition operator) appears entwined in time with a kernel determined by the waiting-time distribution, and memory effects are known to substantially alter diffusion and mixing properties on (temporal) networks [36–38].

While time-fractional diffusion equations and memory-kernel formulations are widely used to model subdiffusive transport in continuous settings, their role in the context of networks and graph-based diffusion has been, so far, explored mainly at a phenomenological level. Existing studies typically emphasize anomalous scaling, spectral properties, or long-time relaxation, but often treat memory as a uniform slowing down mechanism acting on otherwise classical graph diffusion [39–41]. As a result, comparatively little is known about how non-Markovian effects interact with the discrete geometry of a graph, and how they influence transport pathways, or how they modify the vertex-level and path-level behavior beyond global decay rates.

The present work is motivated by the need to clarify how subdiffusion reshapes diffusion on graphs at the structural level. By exploiting the subordination principle [42] and a mass-preserving sum-of-exponentials representation of the Mittag-Leffler operator, we connect fractional diffusion on graphs to a superposition of classical heat processes acting at different internal times. This perspective allows us to study, in a unified way, memory effects, vertex-dependent waiting times, effective distances, and path selection in subdiffusive dynamics. In doing so, we show that memory does not merely slow down diffusion uniformly, but induces heterogeneous temporal behavior across vertices and leads to well-defined subdiffusive distances and paths that differ from their classical diffusive counterparts. These results provide a concrete link between fractional models, memory kernels, and graph-based transport, and offer a more detailed understanding of what subdiffusion means when the underlying space is a network.

1.1. Subdiffusion: Physical Mechanisms and Mathematical Models

A central hallmark of anomalous diffusion is the deviation of the mean squared displacement (MSD) from the linear-in-time growth predicted by Fickian diffusion. Instead, one observes a power-law scaling [5,6,43]

$$\langle x^2(t) \rangle \sim t^\alpha, \quad 0 < \alpha < 1, \quad (5)$$

which defines subdiffusive behavior. As emphasized in [6], such subdiffusion is not a single phenomenon but rather a collective outcome of different physical mechanisms, each leading to distinct stochastic and mathematical descriptions.

One prominent mechanism underlying subdiffusion is trapping [6,14,15]. In crowded or energetically disordered environments, particles may experience long waiting times between successive displacements due to transient binding or deep potential wells. This situation is naturally described by the continuous-time random walk (CTRW) framework, where the waiting times between steps are independent random variables drawn from a heavy-tailed distribution

$$\psi(\tau) \sim \tau^{-1-\alpha}, \quad 0 < \alpha < 1, \quad (6)$$

implying a diverging mean waiting time.

In the scaling limit, CTRW dynamics lead to a fractional diffusion equation (FDE) for the probability density $p(x, t)$. Sokolov [6] stresses that the use of such fractional equations is physically justified only when the underlying trapping assumptions are valid. CTRW models generally exhibit aging, weak ergodicity breaking, and large trajectory-to-trajectory fluctuations.

Another class of subdiffusive systems arises from transport in labyrinthine or fractal structures, such as percolation clusters or tortuous channel networks [6,44–46]. In these systems, anomalous diffusion originates from geometric constraints rather than trapping times. The particle explores a space with no translational invariance and a broad distribution of path lengths.

A paradigmatic example is diffusion on critical percolation clusters or related fractal media, for which the MSD again follows a subdiffusive power law. While fractional diffusion equations may reproduce the probability density in unbounded domains, Sokolov [6] emphasizes that they generally fail to capture important properties such as first-passage times or confined-domain behavior. Consequently, geometric subdiffusion and trapping-induced subdiffusion can yield identical PDFs while corresponding to fundamentally different physical processes.

Subdiffusion may also emerge in viscoelastic environments [47–49], where the tagged particle is embedded in a complex interacting medium, such as a polymer network or the cytoskeleton. In this case, the dynamics are governed not by trapping or geometry but by long-range temporal correlations in the particle's motion.

These systems are often described by fractional Brownian motion (fBm), a Gaussian process characterized by correlated increments, or equivalently by generalized Langevin equations (GLEs) with memory kernels,

$$m\dot{v}(t) = - \int_0^t G(t-t') v(t') dt' + \zeta(t), \quad (7)$$

where the friction kernel typically follows a power law $G(t) \sim t^{-\beta}$, and $\zeta(t)$ is a correlated noise term. Depending on the noise properties, the resulting MSD scales subdiffusively. Unlike CTRW-based models, fBm and GLE dynamics are ergodic and do not exhibit aging—a distinction that is crucial for interpreting experimental data.

Another way of modeling subdiffusion is via models based on diffusion equations with time-dependent diffusion coefficients [50,51],

$$\frac{\partial p(x, t)}{\partial t} = D(t) \frac{\partial^2 p(x, t)}{\partial x^2}, \quad D(t) \sim t^{\alpha-1}. \quad (8)$$

Although such models reproduce the same MSD scaling as subdiffusive processes, they are primarily phenomenological fitting tools (see [6]). Despite yielding the same probability

density as fractional Brownian motion, they lack its correlation structure and are more closely related to mean-field descriptions of CTRW dynamics.

In closing, real systems often exhibit subdiffusion of mixed origin, combining trapping, geometric constraints, and viscoelastic effects. In such cases, different models may predict the same MSD or PDF while differing fundamentally in their aging properties, ergodicity, and trajectory statistics [6].

We emphasize that the analytical ingredients underlying fractional diffusion—such as the subordination identity and the complete monotonicity of the Mittag-Leffler function—are classical results. The novelty of the present work lies not in these foundational properties themselves, but in their synthesis and interpretation in the context of graph-based diffusion, leading to new geometric, operator-theoretic, and vertex-level insights.

1.2. Contributions of the Paper

We employ the subordination relation (Section 2) to construct a *sum-of-exponentials* (SOE) approximation for the Mittag-Leffler function. While the scalar counterpart is well understood, we work with a matrix-valued approximation, tailored to studying the time evolution governed by the graph Laplacian. Physically, we interpret the SOE as a superposition of diffusion-like processes across multiple timescales. The subordination identity and the complete monotonicity of the Mittag-Leffler function are well-established results in fractional calculus. In this work, these classical tools serve as the foundation for a new graph-based framework. Our contribution is to develop a structural and dynamical synthesis of fractional diffusion on graphs that combines operator representations, geometric constructions, and vertex-level interpretations of memory.

In Section 3, we introduce the mathematical framework of the sum-of-exponentials structure and derive appropriate tail bounds that guarantee the accuracy of the approximation as the number of terms grows to infinity (Theorem 1 and Section 3.2). In Section 4, we describe the error metrics used to assess the accuracy of the SOE approximation and discuss the influence of the fractional parameter α on the contributions of the different diffusion-like processes.

In Section 5, we present new metrizations of the graph, which assign a weight to each edge based on the diffusive, subdiffusive, or SOE-based behavior, respectively. We then introduce our main experimental tool, the *shortest paths* in these new metrics on the graph. We prove that in the small-time limit, the *subdiffusive shortest paths* coincide exactly with the shortest paths in the original metric (Theorems 2 and 3), with a preference for traversing high-degree regions (Theorem 4).

Our numerical experiments show that the shortest paths in the diffusive and subdiffusive metrics exhibit completely different behaviors: while the *diffusive shortest paths* tend to cover broader regions of the graph for different time values, the *subdiffusive shortest paths* preserve path history even over extended timescales. The *shortest paths* in the SOE metric function as a bridge between these two phenomena. We propose a memory interpretation based on the influence of previous time instants from the diffusive point of view, and on remembering the most efficient paths at smaller values of t from the metrization point of view, revealing memory-assisted navigation on a network.

Next, we study how memory affects the dynamics on a graph from a vertex-based perspective. In Section 6, we prove that a vertex may receive larger contributions from the more remote or more recent past, depending on the local temporal curvature of the solution to (1). This happens for positive temporal curvature for any $\alpha \in [0, 1)$, and for negative temporal curvature in the limit $\alpha \rightarrow 0$ (Theorem 5). In Section 7, we provide examples of how different remote or recent past biases may occur at the same time for different vertices: early-time convexity at sources and concavity at their neighbors. Furthermore, we discuss

how algebraic (rather than exponential) relaxation and degree-dependent waiting-time effects fundamentally alter transport, trapping, and path selection on networks.

Finally, in Section 8, we show how fractional dynamics can be interpreted as arising from a singular limit of multi-rate diffusion with memory, in an appropriate asymptotic sense. Time-fractional equations can be viewed as scale-free limits of a finite superposition of diffusions in the Laplace domain, equivalently describable via operator-valued Volterra memory kernels (Propositions 7 and 8). This provides a common architecture linking SOE approximations, memory equations, and fractional calculus.

In summary, the genuinely new contribution of this work lies in the graph-based synthesis of fractional diffusion, including (i) a mass-preserving, operator-level sum-of-exponentials (SOE) construction for the Mittag-Leffler propagator on graphs; (ii) the introduction of subdiffusive distances and shortest paths that define a new geometry on networks; and (iii) a vertex-resolved interpretation of memory effects, revealing heterogeneous, degree-dependent temporal behavior. These elements provide a unified perspective linking fractional dynamics, classical diffusion processes, and graph structure.

2. Preliminaries

2.1. Notation and Assumptions

We work on \mathbb{R}^n with the Euclidean norm $\|\cdot\|_2$ and the induced operator norm $\|\cdot\|_2$. We consider $G = (V, E)$ to be a simple undirected graph on n vertices with $A = (A_{ij})$ being its adjacency matrix, and $D = \text{diag}(d_i)$ with $d_i = \sum_j A_{ij}$ the matrix of vertex degree. For a finite graph, the Laplacian operator is realized by the so-called *graph Laplacian matrix* $L = D - A$. L is a real symmetric, positive semi-definite matrix satisfying $L\mathbf{1} = 0$; if G is connected, then $\ker(L) = \text{span}\{\mathbf{1}\}$. The spectrum of L is $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n =: \lambda_{\max}$; let $L = V\Lambda V^\top$ be an orthogonal diagonalization. For a bounded Borel function $f: [0, \infty) \rightarrow \mathbb{R}$ we use the spectral calculus $f(L) := Vf(\Lambda)V^\top$.

If $\{(\lambda_n, \phi_n)\}_{n=1}^N$ are the eigenpairs of L , then the solution to the abstract Cauchy problem in Equation (1) is given by

$$\begin{aligned} u(t) &= E_\alpha(-t^\alpha L) u_0 \\ &= \sum_{n=1}^N E_\alpha(-\lambda_n t^\alpha) \langle u_0, \phi_n \rangle \phi_n, \end{aligned} \quad (9)$$

where $E_\alpha(M) = E_{\alpha,1}(M)$ is the Mittag-Leffler matrix function of M , which has the following power-series definition:

$$E_\alpha(M) = \sum_{k=0}^{\infty} \frac{M^k}{\Gamma(\alpha k + 1)}, \quad (10)$$

with $\Gamma(\cdot)$ being the Euler gamma function.

2.2. Subordination Identity and the M-Wright Density

For $0 < \alpha < 1$,

$$E_\alpha(-t^\alpha L) = \int_0^\infty M_\alpha(\theta) e^{-\theta t^\alpha L} d\theta, \quad (11)$$

where $M_\alpha(\theta) \geq 0$ is the M-Wright (Mainardi) density, $\int_0^\infty M_\alpha(\theta) d\theta = 1$, and a convenient series form is [52]

$$M_\alpha(\theta) = \sum_{k=0}^{\infty} \frac{(-\theta)^k}{k! \Gamma(1 - \alpha(k+1))} \quad (\theta \geq 0). \quad (12)$$

The integral in (11) is a Bochner integral of the operator-valued map $s \mapsto e^{-sL}$ against probability measures.

2.3. Complete Monotonicity and Bernstein's Theorem

The following facts clarify why the fractional propagator is a *positive, mass-preserving* contraction.

Proposition 1. Fix $0 < \alpha < 1$. The scalar function $\lambda \rightarrow E_\alpha(-t^\alpha \lambda)$ is completely monotone on $[0, \infty)$. There is a probability density $M_\alpha(\theta)$ on $(0, \infty)$ such that

$$E_\alpha(-t^\alpha \lambda) = \int_0^\infty e^{-\theta t^\alpha \lambda} M_\alpha(\theta) d\theta,$$

and in the operator case (Bochner integral):

$$E_\alpha(-t^\alpha L) = \int_0^\infty e^{-\theta t^\alpha L} M_\alpha(\theta) d\theta.$$

Furthermore, $E_\alpha(-t^\alpha L)$ is self-adjoint, positive, and $\|E_\alpha(-t^\alpha L)\|_2 = 1$. Moreover, if $L\mathbf{1} = 0$ then $E_\alpha(-t^\alpha L)\mathbf{1} = \mathbf{1}$.

Proof. The Mittag-Leffler function is completely monotone. By Bernstein's Theorem [53], any completely monotone function can be represented as a mixture of exponentials:

$$E_\alpha(-y) = \int_0^\infty e^{-y\theta} M_\alpha(\theta) d\theta.$$

For the Mittag-Leffler function, the density is given by the Mainardi function $M_\alpha(\theta)$. For a thorough exposition of these results, see [52–55]. Using the change of variables $y = t^\alpha \lambda$, we obtain the desired scalar identity. Functional calculus then yields the following operator identity:

$$\begin{aligned} E_\alpha(L) &= V E_\alpha(\Lambda) V^\top = V \left(\int_0^\infty e^{-\theta t^\alpha \Lambda} M_\alpha(\theta) d\theta \right) V \\ &= \int_0^\infty \left(V e^{-\theta t^\alpha \Lambda} V \right) M_\alpha(\theta) d\theta = \int_0^\infty e^{-\theta t^\alpha L} M_\alpha(\theta) d\theta. \end{aligned}$$

We have positivity $E_\alpha(-t^\alpha \lambda) \geq 0$ for $0 < \alpha < 1$. From monotonicity, we obtain $E_\alpha(-t^\alpha \lambda) \leq E_\alpha(0) = 1$ for $\lambda \geq 0$. Since L is self-adjoint, the eigenvalues of L are non-negative and $\lambda_1 = 0$, so we have

$$\|E_\alpha(-t^\alpha L)\|_2 = \max_{\lambda \in \sigma(L)} |E_\alpha(-t^\alpha \lambda)| = E_\alpha(0) = 1.$$

The mass property follows from $e^{-\theta t L} \mathbf{1} = \mathbf{1}$ and $\int_0^\infty M_\alpha(\theta) d\theta = 1$. \square

If G is connected, then $\lim_{t \rightarrow \infty} E_\alpha(-t^\alpha L) u_0 = \frac{1}{n} (\mathbf{1}^\top u_0) \mathbf{1}$. For a graph with $c > 1$ components, the limit projects u_0 to the vector that is constant on each component, with the corresponding component averages.

We note that the subordination identity and the complete monotonicity properties summarized above are classical; their role here is to provide the analytical foundation for the graph-based operator constructions and interpretations developed in the subsequent sections.

2.4. Terminology: Subordination and the Random Clock

Consider the heat semigroup on the graph $\{e^{-sL}\}_{s \geq 0}$. Let $(S_\tau)_{\tau \geq 0}$ be the α -stable subordinator and $E_t = \inf\{\tau > 0 : S_\tau > t\}$ the inverse subordinator. Here, *subordination*

means that the time-fractional evolution is obtained by running the baseline (time-linear) heat dynamics with the *random clock* E_t in place of deterministic time,

$$E_\alpha(-t^\alpha L) = \int_0^\infty e^{-sL} g_\alpha(s, t) ds = \int_0^\infty M_\alpha(\theta) e^{-\theta t^\alpha L} d\theta, \quad (13)$$

with $g_\alpha(s, t) = t^{-\alpha} M_\alpha(s/t^\alpha)$; see [42,52,56]. We do *not* consider Bochner subordination by a general Bernstein function ϕ (which would lead to space-fractional $\phi(L)$). Equation (13) means that the operator $E_\alpha(-t^\alpha L)$ is the expected value of the standard diffusion operator e^{-sL} , where the $g_\alpha(s, t)$ is the probability that the random clock displays time s when the physical time that has passed is t .

2.5. The Time-Changed Process X_{E_t}

Let $\{X_s\}_{s \geq 0}$ be the continuous-time Markov process on V with generator $-L$:

$$\frac{d}{ds} p(s) = -L p(s), \quad p(s) = (\mathbb{P}[X_s = j | X_0 = i])_{i,j} = e^{-sL}.$$

Starting from node i , the holding time is an exponential random variable with decay rate $d_i = \sum_j A_{ij}$. The next node of the random walk is then chosen to be j with probability A_{ij}/d_i .

Let $E_t := \inf\{\tau > 0 : S_\tau > t\}$ be the inverse α -stable subordinator ($0 < \alpha < 1$), independent of X . The *time-changed process* is $Y_t := X_{E_t}$, i.e., the baseline diffusion observed at the random operational time E_t . Averaging over the random clock yields

$$\mathbb{P}[Y_t = j | Y_0 = i] = \int_0^\infty \mathbb{P}[X_s = j | X_0 = i] g_\alpha(s, t) ds = [E_\alpha(-t^\alpha L)]_{ij}.$$

Hence, for any initial u_0 , $u(t) = E_\alpha(-t^\alpha L) u_0$ solves the Caputo differential equation $\partial_t^\alpha u(t) = -Lu(t)$. Y_t is semi-Markov in physical time t : the holding time T_i at node i has survival $\mathbb{P}(T_i > t) = E_\alpha(-d_i t^\alpha)$, a heavy tail ($\sim t^{-\alpha}$), producing trapping and memory. Since $L\mathbf{1} = 0$ and the integral representation averages stochastic kernels, Y_t preserves total mass. On a connected graph, $\lim_{t \rightarrow \infty} Y_t$ has the same equilibrium as the baseline walk (i.e., componentwise averaging).

3. Sum of Exponentials Approximation

In this section, we present a practical, mass-preserving approximation of the Mittag-Leffler function as a sum of exponential functions from a mathematical point of view. We start with the following fundamental definition.

Definition 1. *The sum-of-exponentials (SOE) approximation is*

$$F_J(t, L) = \sum_{j=1}^J a_j e^{-b_j t^\alpha L}, \quad (14)$$

with coefficients a_j, b_j that depend only on α and not on t .

Our aim is to obtain

$$E_\alpha(-t^\alpha L) \approx \sum_{j=1}^J a_j e^{-b_j t^\alpha L},$$

This approximation is constructed as a quadrature of (11) whose window is logarithmically scaled in θ . In the rest of the section, we analyze the mathematical properties of the SOE approximation. Specifically, we perform the following:

1. Provide a log–trapezoidal construction with $a_j > 0$, $\sum_j a_j = 1$ (mass conservation),
2. Prove ensuring geometric convergence in J for a fixed window;
3. Derive explicit, uniform tail bounds from M_α to select $(\theta_{\min}, \theta_{\max})$;
4. Tabulate ready-to-use window endpoints $(\theta_{\min}, \theta_{\max})$ (and (y_{\min}, y_{\max})) for typical α, ε (see Appendix A);
5. Provide practical error metrics (relative/probe, mass error) for a posteriori assessment.

The SOE representation is obtained as a quadrature approximation of the subordination integral and should be understood primarily at the operator level. In particular, the coefficients (a_j, b_j) arise from a discretization of the integral over the M–Wright density and do not, by themselves, define a stochastic mixture or a dynamical decomposition of independent processes.

Nevertheless, the representation $E_\alpha(-t^\alpha L) \approx \sum_j a_j e^{-b_j t^\alpha L}$ may be viewed heuristically as a superposition of diffusion-like operators acting at rescaled internal times. This interpretation serves as an intuitive guide for understanding multi-scale behavior, but it is not intended as a literal probabilistic construction unless additional structure is imposed.

3.1. SOE Via Log–Trapezoidal Quadrature

Set $\theta = e^y$, $y \in \mathbb{R}$, and truncate to $[y_{\min}, y_{\max}]$. With a uniform grid $y_j = y_{\min} + (j-1)h$, $h = \frac{y_{\max} - y_{\min}}{J-1}$, define

$$b_j = \theta_j = e^{y_j}, \quad \tilde{a}_j = h M_\alpha(b_j) b_j, \quad a_j = \frac{\tilde{a}_j}{\sum_{k=1}^J \tilde{a}_k}. \quad (15)$$

Then

$$E_\alpha(-t^\alpha L) \approx \sum_{j=1}^J a_j e^{-b_j t^\alpha L}, \quad a_j > 0, b_j > 0, \sum_{j=1}^J a_j = 1. \quad (16)$$

Remark 1. The nodes b_j correspond to quadrature points in the variable θ arising from the subordination formula, and the weights a_j approximate the probability density $M_\alpha(\theta)$ in an integral sense. However, the discrete representation $\sum_j a_j e^{-b_j t^\alpha L}$ should not be interpreted as an exact probabilistic mixture of independent diffusion processes. Rather, it is a deterministic operator approximation whose coefficients inherit their meaning from the underlying integral representation. Any interpretation in terms of multiple time scales or internal clocks is therefore heuristic and serves primarily as a conceptual aid.

Convention 1. In all that follows, we fix the nodes b_j once (they depend only on α and the chosen log–window) and, for any $t > 0$, evaluate the SOE as $\sum_{j=1}^J a_j e^{-(t^\alpha b_j)L}$. Thus, the coefficients (a_j, b_j) do not depend on t or on L . We then have the following results:

Proposition 2. If $L\mathbf{1} = 0$ and $\sum_j a_j = 1$, then for all $t \geq 0$ the SOE approximation preserves mass, that is, $\mathbf{1}^\top (\sum_{j=1}^J a_j e^{-b_j t^\alpha L}) u_0 = \mathbf{1}^\top u_0$.

Proof. Since $L\mathbf{1} = 0$, $e^{-b_j t^\alpha L} \mathbf{1} = \mathbf{1}$ for each j . Thus, $(\sum_j a_j e^{-b_j t^\alpha L}) \mathbf{1} = (\sum_j a_j) \mathbf{1} = \mathbf{1}$, and left-multiplying by $\mathbf{1}^\top$ gives $\mathbf{1}^\top (\cdot) u_0 = \mathbf{1}^\top u_0$. \square

Lemma 1. For self-adjoint L and bounded Borel f, g , $\|f(L) - g(L)\|_2 = \max_{\lambda \in \sigma(L)} |f(\lambda) - g(\lambda)|$. In particular, the SOE error is

$$\|E_\alpha(-t^\alpha L) - F_J(t, L)\|_2 = \max_{\lambda \in [0, \lambda_{\max}]} |g(\lambda) - g_J(\lambda)|,$$

with $g(\lambda) = E_\alpha(-t^\alpha \lambda)$ and $g_J(\lambda) = \sum_j a_j e^{-b_j t^\alpha \lambda}$.

Proof. Diagonalize $L = V\Lambda V^\top$. Then $g(L) - g_J(L) = V(g(\Lambda) - g_J(\Lambda))V^\top$ and $\|g(L) - g_J(L)\|_2 = \|g(\Lambda) - g_J(\Lambda)\|_2 = \max_k |g(\lambda_k) - g_J(\lambda_k)|$. \square

Assuming that λ is bounded in a compact set, the error of the SOE approximation decreases geometrically in the number of terms J and uniformly in λ . This assumption holds provided that we have an upper bound on the eigenvalues of L .

Theorem 1. Let $f_\lambda(y) := M_\alpha(e^y) e^y e^{-e^y t^\alpha \lambda}$ with $t > 0$, $\lambda \in [0, \lambda_{\max}]$. For fixed $[y_{\min}, y_{\max}]$ and step $h = (y_{\max} - y_{\min}) / (J - 1)$,

$$\sup_{\lambda \in [0, \lambda_{\max}]} \left| \int_{y_{\min}}^{y_{\max}} f_\lambda(y) dy - h \sum_{j=1}^J f_\lambda(y_j) \right| \leq C_1 e^{-C_2/h},$$

with constants $C_1, C_2 > 0$ independent of h and λ . Consequently,

$$\|E_\alpha(-t^\alpha L) - \sum_{j=1}^J a_j e^{-b_j t^\alpha L}\|_2 \leq C_1 e^{-C_2(J-1)/(y_{\max} - y_{\min})}.$$

Proof (Analytic trapezoidal rule). The map $y \mapsto f_\lambda(y)$ is analytic in a strip $S_a = \{|\operatorname{Im} y| < a\}$ and decays rapidly as $\operatorname{Re} y \rightarrow \pm\infty$ (exhibiting doubly exponential decay as $\operatorname{Re} y \rightarrow +\infty$ and being integrable towards $-\infty$). For analytic, rapidly decaying functions, the (periodized) trapezoidal rule error on a finite interval is $\leq C e^{-2\pi a/h}$ via Poisson summation; see (Theorem 2.1, [57]). Uniform strip bounds over $\lambda \in [0, \lambda_{\max}]$ yield uniform constants C_1, C_2 . Finally, Lemma 1 transfers the scalar quadrature error to the operator norm bound. \square

3.2. Tail Bounds for the Subordination Integral and Window Selection

Having determined that we can approximate the integral on $[y_{\min}, y_{\max}]$, we now show how to choose the window endpoints. We start by considering the subordination identity previously defined and its scalar counterpart $F(\lambda) = \int_0^\infty M_\alpha(\theta) e^{-\theta t^\alpha \lambda} d\theta$ for $\lambda \geq 0$. For window endpoints $0 < \theta_{\min} < \theta_{\max} < \infty$ (equivalently, $y_{\min} = \log \theta_{\min}$ and $y_{\max} = \log \theta_{\max}$), we define the truncated operator

$$F_{\text{win}}(L) = \int_{\theta_{\min}}^{\theta_{\max}} M_\alpha(\theta) e^{-\theta t^\alpha L} d\theta,$$

so that the truncation error splits as the *left* and *right* tails:

$$\mathcal{T}_L(\lambda) := \int_0^{\theta_{\min}} M_\alpha(\theta) e^{-\theta t^\alpha \lambda} d\theta, \quad \mathcal{T}_R(\lambda) := \int_{\theta_{\max}}^\infty M_\alpha(\theta) e^{-\theta t^\alpha \lambda} d\theta.$$

By the spectral theorem (Lemma 1), the operator truncation error satisfies

$$\|F(L) - F_{\text{win}}(L)\|_2 = \max_{\lambda \in \sigma(L)} (\mathcal{T}_L(\lambda) + \mathcal{T}_R(\lambda)) \leq \sup_{\lambda \in [0, \lambda_{\max}]} \mathcal{T}_L(\lambda) + \sup_{\lambda \in [0, \lambda_{\max}]} \mathcal{T}_R(\lambda).$$

3.2.1. Left Tail (Small θ)

Near $\theta = 0$, the M-Wright density satisfies $M_\alpha(\theta) = \frac{1}{\Gamma(1-\alpha)} + O(\theta)$ (see [54,55]). A simple bound follows.

Proposition 3. For all $t \geq 0$, $\lambda \geq 0$ and $0 < \theta_{\min} \leq 1$,

$$\mathcal{T}_L(\lambda) \leq \frac{1}{\Gamma(1-\alpha)} \begin{cases} \frac{1 - e^{-t^\alpha \lambda \theta_{\min}}}{t^\alpha \lambda}, & \lambda > 0, \\ \theta_{\min}, & \lambda = 0. \end{cases} \quad (17)$$

Consequently, uniformly in $\lambda \in [0, \lambda_{\max}]$,

$$\sup_{\lambda \in [0, \lambda_{\max}]} \mathcal{T}_L(\lambda) \leq \frac{\theta_{\min}}{\Gamma(1-\alpha)}.$$

Proof. Use $M_\alpha(\theta) \leq \frac{1}{\Gamma(1-\alpha)}$ for $\theta \in (0, 1]$ and integrate $\int_0^{\theta_{\min}} e^{-t\lambda\theta} d\theta$ (with the obvious limit as $\lambda \downarrow 0$). \square

We remark that taking $\theta_{\min} = \varepsilon \Gamma(1-\alpha)$ forces the left-tail contribution below ε for all eigenvalues (including the zero mode).

3.2.2. Right Tail (Large θ)

As $\theta \rightarrow \infty$, the density exhibits a *stretched-exponential* decay (Chapter 4, [55]), (Chapter 2, [54]): there exist $C_\alpha, c_\alpha > 0$ and $q_\alpha := \frac{1}{1-\alpha} > 1$ such that, for all sufficiently large θ ,

$$M_\alpha(\theta) \leq C_\alpha \theta^{p_\alpha} \exp(-c_\alpha \theta^{q_\alpha}), \quad p_\alpha := \frac{\alpha - 2}{2(1-\alpha)}. \quad (18)$$

This yields two complementary bounds.

Proposition 4. Let (18) hold for all $\theta \geq \theta_0(\alpha)$. Then for any $t \geq 0$, $\lambda \geq 0$ and $\theta_{\max} \geq \theta_0(\alpha)$,

$$\mathcal{T}_R(\lambda) \leq \int_{\theta_{\max}}^{\infty} C_\alpha \theta^{p_\alpha} e^{-c_\alpha \theta^{q_\alpha}} d\theta \leq \frac{C_\alpha}{q_\alpha c_\alpha} \theta_{\max}^{p_\alpha + 1 - q_\alpha} \exp(-c_\alpha \theta_{\max}^{q_\alpha}). \quad (19)$$

In particular, the bound is independent of λ (and thus controls the zero eigenvalue).

Proof. Dropping the factor $e^{-t\lambda\theta} \leq 1$, we bound the tail of a monotone stretched-exponential via the standard inequality $\int_x^\infty u^p e^{-cu^q} du \leq \frac{1}{qc} x^{p+1-q} e^{-cx^q}$, which holds for large x . \square

Proposition 5. If the graph is connected and we restrict to $\mathbf{1}^\perp$ (thus $\lambda \geq \lambda_2 > 0$), then for any $\theta_{\max} > 0$,

$$\sup_{\lambda \in [\lambda_2, \lambda_{\max}]} \mathcal{T}_R(\lambda) \leq e^{-t^\alpha \lambda_2 \theta_{\max}} \int_{\theta_{\max}}^{\infty} M_\alpha(\theta) d\theta \leq e^{-t^\alpha \lambda_2 \theta_{\max}}. \quad (20)$$

Proof. Use $e^{-t\lambda\theta} \leq e^{-t\lambda_2\theta}$ and $\int_{\theta_{\max}}^{\infty} M_\alpha(\theta) d\theta \leq 1$. \square

We now consider the practical choices based on the previous results:

- *All modes (including $\lambda = 0$):* Pick θ_{\max} so that $c_\alpha \theta_{\max}^{q_\alpha} \geq \log(2/\varepsilon)$; then by (19), $\sup_\lambda \mathcal{T}_R(\lambda) \lesssim \frac{C_\alpha}{q_\alpha c_\alpha} \theta_{\max}^{p_\alpha + 1 - q_\alpha} e^{-c_\alpha \theta_{\max}^{q_\alpha}} \leq \varepsilon/2$ for large enough θ_{\max} .
- *Mean-zero subspace:* Set $\theta_{\max} \geq \frac{1}{t^\alpha \lambda_2} \log(2/\varepsilon)$ to guarantee $\sup_{\lambda \geq \lambda_2} \mathcal{T}_R(\lambda) \leq \varepsilon/2$ by (20).

3.2.3. Window Rules: Left and Right Together

Combining Propositions 3–5 yields explicit choices for $(\theta_{\min}, \theta_{\max})$ that ensure the total tail remains below a target tolerance ε :

Corollary 1. Given $\varepsilon \in (0, 1)$ and $0 < \alpha < 1$:

General (all modes). If we choose

$$\theta_{\min} = \frac{\varepsilon}{2} \Gamma(1 - \alpha), \quad \theta_{\max} \text{ s.t. } c_{\alpha} \theta_{\max}^{q_{\alpha}} \geq \log \frac{2}{\varepsilon},$$

then $\sup_{\lambda \in [0, \lambda_{\max}]} (\mathcal{T}_L(\lambda) + \mathcal{T}_R(\lambda)) \leq \varepsilon$.

Mean-zero subspace (connected graph). To estimate the operator on $\mathbf{1}^{\perp}$, we employ the following the choice of parameters

$$\theta_{\min} = \frac{\varepsilon}{2} \Gamma(1 - \alpha), \quad \theta_{\max} = \frac{1}{t \lambda_2} \log \frac{2}{\varepsilon}.$$

If one intends to reuse the same nodes for multiple times $t \in [t_{\min}, t_{\max}]$, it is sufficient to replace t with t_{\min} in the expression for θ_{\max} , so that the nodes (a_j, b_j) are t -independent. Then $\sup_{\lambda \in [\lambda_2, \lambda_{\max}]} (\mathcal{T}_L(\lambda) + \mathcal{T}_R(\lambda)) \leq \varepsilon$.

Remark 2. The log–trapezoidal SOE uses $y = \log \theta$ on $[y_{\min}, y_{\max}]$, with $y_{\min} = \log \theta_{\min}$ and $y_{\max} = \log \theta_{\max}$. After fixing the window by Corollary 1, increasing J then controls the discretization error inside the window with geometric rate (Theorem 1).

Remark 3. The choice of θ_{\max} via Proposition 5 with λ_2 replaced by λ_{\max} yields a conservative upper bound on the tail for high-frequency modes and motivates rules of the form $\theta_{\max} \sim C / (t\lambda)$. Empirically, setting $C \approx 32$ ensures that $\exp(-\theta_{\max} t\lambda)$ falls near machine precision and works well when the focus is on modes away from the zero eigenvalue; the rigorous alternative (19) controls all modes using only properties of M_{α} .

3.2.4. Constants in the Stretched-Exponential Bound

Explicit asymptotics for M_{α} yield $q_{\alpha} = \frac{1}{1-\alpha}$ and an exponent constant $c_{\alpha} = (1 - \alpha) \alpha^{\alpha/(1-\alpha)}$. The prefactor has a power $\theta^{p_{\alpha}}$ with $p_{\alpha} = \frac{\alpha-2}{2(1-\alpha)}$ and a multiplicative constant depending only on α (see Section 2.3, [54] and Section 4.3, [55]). Using these in (19) yields a fully explicit $\theta_{\max}(\varepsilon, \alpha)$.

4. Error of the SOE Approximation

In this section, we define several error metrics to assess the accuracy of the SOE approximation, and test them using a sample graph.

4.1. Error Metrics

Given a probe vector u_0 , let $u^*(t)$ be the solution of (1), namely

$$u^*(t) = E_{\alpha}(-t^{\alpha} L) u_0.$$

Denote by $u^{\text{SOE}}(t)$ the time evolution of u_0 obtained using the SOE approximation

$$u^{\text{SOE}}(t) = F_J(t, L) u_0 = \sum_{j=1}^J a_j e^{-b_j t^{\alpha} L} u_0.$$

We employ the following metrics to evaluate the error between $u^{\text{SOE}}(t)$ and $u^*(t)$.

4.1.1. Relative Error

$$\text{relerr}(u_0) = \frac{\|u^*(t) - u^{\text{SOE}}(t)\|_2}{\|u^*(t)\|_2}. \quad (21)$$

We use the relative error in the two-norm because it reflects the operator-level error, that is, the precision with which the operator $F_J(t, L)$ approximates $E_{\alpha}(-t^{\alpha} L)$.

4.1.2. Mass Conservation Error

$$\text{masserr}(u_0) = \left| \mathbf{1}^\top u^*(t) - \mathbf{1}^\top u^{\text{SOE}}(t) \right|. \quad (22)$$

Ideally, $\text{masserr}(u_0) = 0$; with floating-point arithmetic, it is typically at the level of machine precision due to $\sum_j a_j = 1$ and $L\mathbf{1} = 0$.

4.1.3. Scalar Error

The scalar error, intended on the spectrum of the matrix functions, is

$$\max_{\lambda \in [0, \lambda_{\max}]} |E_\alpha(-t^\alpha \lambda) - F_J(t, \lambda)|. \quad (23)$$

According to Theorem 1, and uniformly for $\lambda \in [0, \lambda_{\max}]$, the scalar error decays *geometrically* with J once the window captures the effective support of the integrand. While the raw error as a function of J need not be strictly decreasing (typical oscillations of spectrally convergent trapezoidal rules), its envelope decays until it reaches the desired accuracy [57].

In all cases, the SOE inherits the correct long-time limit and conservation properties by construction (Proposition 2). Accuracy improves with J and when the window is adapted to the relevant scale t, λ .

4.1.4. Operator Error

If an operator-level diagnostic is desired, one may estimate

$$\frac{\|E_\alpha(-t^\alpha L) - F_J(t, L)\|_2}{\|E_\alpha(-t^\alpha L)\|_2} \quad (24)$$

via power iteration on the difference operator; in practice, the probe-based relative error (21) is sufficient and cheaper.

We have constructed the SOE approximation for scalars, valid on an interval containing the eigenvalues of L , and then extended it to matrices. Therefore, we use the two-norm both for vectors (relative error) and for matrices (operator error) to provide a clear correspondence between the scalar and the operator settings. Indeed, note that in exact arithmetic, the operator error and the scalar spectral error are identical. This is because $\|E_\alpha(-t^\alpha L)\|_2 = 1$ and the two-norm of the symmetric matrix $E_\alpha(-t^\alpha L) - F_J(t, L)$ is the magnitude of its largest eigenvalue; the latter is precisely the scalar error according to the spectral theorem applied to the function $g(\lambda) = E_\alpha(-t^\alpha \lambda) - F_J(t, \lambda)$. However, matrix computations in floating-point arithmetic may introduce numerical inaccuracies, as evidenced by a comparison of Figures 1 and 2B.

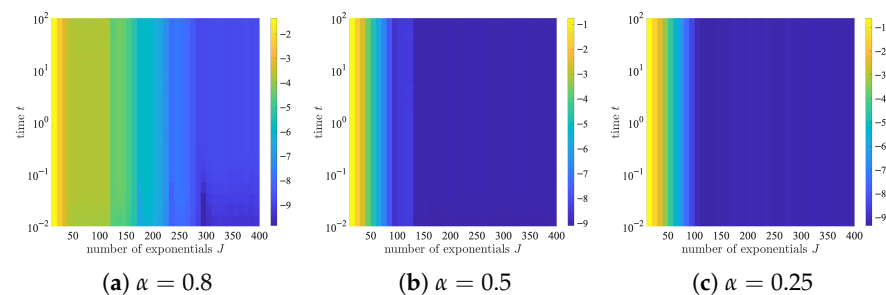


Figure 1. Heatmaps illustrating the scalar error of the SOE approximation on a base-10 logarithmic scale, plotted as a function of the number of addends J and time t . The underlying network is an Erdős–Rényi (ER) random graph with 250 vertices and 1000 edges. Results are shown for fractional orders (a) $\alpha = 0.8$, (b) $\alpha = 0.5$, and (c) $\alpha = 0.25$.

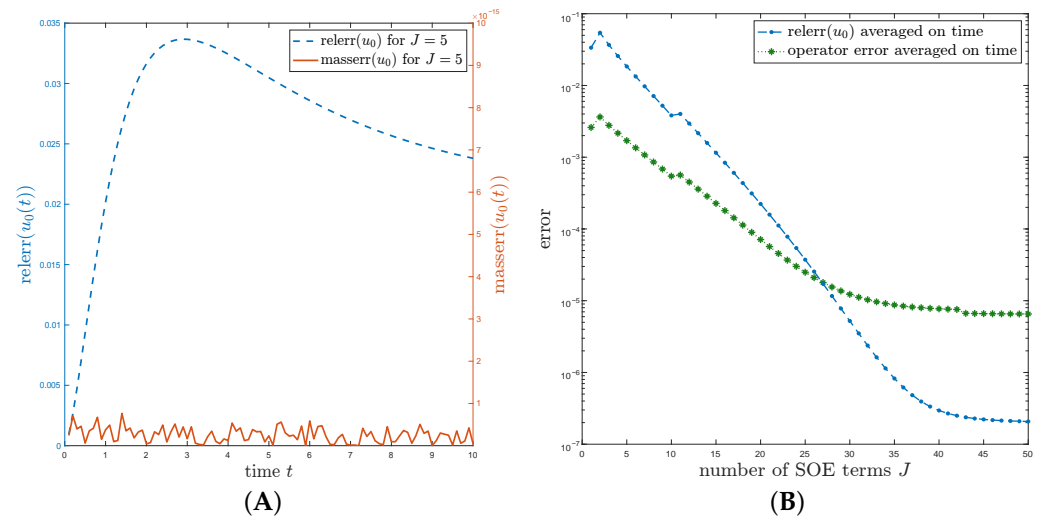


Figure 2. (A) Relative (blue) and mass (orange) errors of the probing vector $u(t)$ between the exact Mittag-Leffler solution and the SOE approximation with $J = 5$. (B) The time-averaged relative (blue) and operator (green) errors are shown as a function of J . The average is computed over 300 time steps in the interval $0.1 \leq t \leq 1000$.

4.2. Computational Results

In this section, we consider an Erdős–Rényi (ER) random graph with 250 vertices and 1000 edges. The graph is simple and connected. We compute the scalar error for three different values of α and a wide range of times, $11 \leq t \leq 1001$. The results are illustrated in Figure 1 for (a) $\alpha = 0.8$, (b) $\alpha = 0.5$, and (c) $\alpha = 0.25$. The convergence rate of the scalar error is higher for smaller values of α ; that is, fewer addends are needed to obtain a sufficiently accurate approximation.

Let us now focus on the nature of the SOE for the three different values of α , with $J = 61$. Using our approach, we obtain the operator-level approximations

$$E_{0.8}(-t^{0.8}L) \approx 0.242e^{-1.52t^{0.8}L} + 0.220e^{-1.19t^{0.8}L} + 0.148e^{-0.92t^{0.8}L} + \dots,$$

$$E_{0.5}(-t^{0.5}L) \approx 0.119e^{-1.48t^{0.5}L} + 0.115e^{-1.15t^{0.5}L} + 0.108e^{-1.90t^{0.5}L} + \dots,$$

and

$$E_{0.25}(-t^{0.25}L) \approx 0.096e^{-1.10t^{0.25}L} + 0.095e^{-1.42t^{0.25}L} + 0.091e^{-0.86t^{0.25}L} + \dots$$

The coefficients $a_j > 0$ are quadrature weights that approximate the contribution of different regions of the subordination integral, and their normalization $\sum_j a_j = 1$ reflects mass preservation at the operator level. The parameters b_j correspond to the quadrature nodes and determine the effective scaling of the semigroup factors $e^{-b_j t^\alpha L}$.

While it is convenient to view the coefficients a_j as indicating the relative contribution of different exponential components, and the b_j as associated with different effective time scales, this interpretation should be understood as heuristic. The SOE representation is not derived from a decomposition into independent stochastic processes, but from a deterministic approximation of the integral representation of the Mittag-Leffler operator.

For larger values of α , the coefficients a_j are more concentrated, and subdiffusion can be represented as a superposition of fewer operators. Indeed, in the limit $\alpha \rightarrow 1$, we exactly recover the diffusive behavior, so we expect $a_1 = 1$ and $a_j = 0$ for $j > 1$. As α decreases, the contributions of the other operators become more prominent. The three leading coefficients of the SOE for $\alpha = 0.25$ indicate that the corresponding exponential components contribute with comparable weights in the operator approximation. This means that multiple diffusive

processes are needed in the strong memory regime to effectively recover subdiffusion. We explain the connection between the scaling parameters b_j and the memory structure in the following section.

5. Subdiffusive Distance and Paths

In this section, we introduce new geometrizations of a graph related to the diffusive processes, subdiffusive processes, and the SOE approximation of the latter. We study the shortest paths under these new metrics and observe, via numerical experiments, a distinct difference in behavior: while the diffusive case explores broader regions of the graph, the subdiffusive case recovers the usual shortest paths in the standard metric.

5.1. Subdiffusive Distance

Let us consider the solution to the Caputo time-fractional diffusion equation: $u(t) = E_\alpha(-t^\alpha L) u_0$ with initial condition $u_0 = e_v$, where e_v is the vector with a one at position v and zeros elsewhere. We then evaluate the capacity of the whole graph to diffuse mass between the vertices v and w at time t . That is, we consider at time t the difference between the mass remaining at the origin, i.e., vertex v , and the mass diffused to vertex w :

$$u_{v|u_0=e_v}(t) - u_{w|u_0=e_v}(t) = (E_\alpha(-t^\alpha L))_{vv} - (E_\alpha(-t^\alpha L))_{vw}.$$

Similarly, the capacity of the whole graph to move mass from w to v conditioned to all initial mass being allocated at w is as follows:

$$u_{w|u_0=e_w}(t) - u_{v|u_0=e_w}(t) = (E_\alpha(-t^\alpha L))_{ww} - (E_\alpha(-t^\alpha L))_{wv}.$$

As we only consider undirected graphs here, the total capacity of mass diffusion between both vertices is as follows:

$$\mathcal{D}_{\alpha,t}(v, w) = (E_\alpha(-t^\alpha L))_{vv} + (E_\alpha(-t^\alpha L))_{ww} - 2(E_\alpha(-t^\alpha L))_{vw}. \quad (25)$$

Then, we have the following result.

Proposition 6. Let $\mathcal{D}_{\alpha,t}(v, w)$ be defined as before. Then $\mathcal{D}_{\alpha,t}(v, w)$ is real and non-negative. There exists an embedding of vertices $V \rightarrow \mathbb{R}^n$, $i \rightarrow x_i$ such that $\mathcal{D}_{\alpha,t}(v, w) = \|x_v - x_w\|_2^2$.

Proof. Using positivity, $E_\alpha(-t^\alpha \lambda) \geq 0$ for $0 < \alpha < 1$, we write

$$\begin{aligned} E_\alpha(-t^\alpha L) &= V E_\alpha(-t^\alpha \Lambda) V^\top \\ &= V (E_\alpha(-t^\alpha \Lambda))^{1/2} (E_\alpha(-t^\alpha \Lambda))^{1/2} V^\top \\ &= (V E_\alpha(-t^\alpha \Lambda))^{1/2} \left((V E_\alpha(-t^\alpha \Lambda))^{1/2} \right)^\top. \end{aligned}$$

Denote $X_\alpha(t) = (V E_\alpha(-t^\alpha \Lambda))^{1/2}$ and by $x_{\alpha,t}^\top(v)$ the v -th row of X . Thus,

$$(E_\alpha(-t^\alpha L))_{vw} = x_{\alpha,t}^\top(v) x_{\alpha,t}(w).$$

Then, $v \mapsto x_{\alpha,t}^\top(v)$ is an embedding of vertices in \mathbb{R}^n . Therefore

$$\mathcal{D}_{\alpha,t}(v, w) = \|x_{\alpha,t}(v) - x_{\alpha,t}(w)\|_2^2,$$

so $\mathcal{D}_{\alpha,t}(v, w)$ is a square Euclidean distance between the vertices v and w . \square

Hereafter, we call $\mathcal{D}_{\alpha,t}(v,w)$ the (squared) *subdiffusive distance* between the corresponding vertices in the graph when $\alpha < 1$. Notice that for $\alpha = 1$, the quantity $\mathcal{D}_{1,t}(v,w)$ is the diffusion distance defined by Coiffman et al. [58,59]. These distances are part of a large family of diffusion-like distances on graphs [60,61].

5.2. Subdiffusive Shortest Paths

The subdiffusive distance represents how easy it is to transmit mass between two vertices, regardless of whether they are adjacent or not. However, this distance does not directly determine which pathways are utilized for this transmission. Heuristically, paths should traverse the most efficient edges, although it is not immediately obvious how to identify them. To determine these paths in both the standard and subdiffusive regimes, we define the following weighted graph.

Definition 2. Let $\sqrt{\mathcal{D}_{\alpha,t}}$ be the matrix of (subdiffusive) distances, intended as the entry-wise square root of $\mathcal{D}_{\alpha,t}$. Define a weighted graph whose adjacency matrix is the Hadamard product $W = A \odot \sqrt{\mathcal{D}_{\alpha,t}}$, which assigns to each edge (i,j) the cost induced by the discrepancy of their diffusion profiles at time t . That is,

$$W_{\alpha,t}(v,w) = \begin{cases} \sqrt{\mathcal{D}_{\alpha,t}(v,w)}, & (v,w) \in E, \\ 0, & (v,w) \notin E. \end{cases}$$

The resulting weighted graph \tilde{G} is the geometrization of the original graph G .

The weighted graph \tilde{G} can be seen as a representation of G as a one-dimensional complex, i.e., a metric length space where each edge $e = (v,w) \in E$ is a compact one-dimensional manifold with boundary $\partial e = \{v,w\}$, parameterized as $\tilde{e}_{\alpha,t}(v,w) = [0, W_{\alpha,t}(v,w)]$ (see [62,63]).

The resulting shortest paths (computed via Dijkstra's algorithm) identify chains of vertices whose subdiffusion states remain maximally coherent at time t . We call these paths the *subdiffusive shortest paths*. These are not, in general, the paths that require the fewest edges of A , which we call *topological shortest paths* or *geodesic shortest paths*. They generalize the *shortest communicability paths* introduced in [64], which were defined using the matrix function $f(A) = \exp(\beta A)$.

Consider the subdiffusive shortest paths on $W(t)$ as time varies. When t is close to zero, the effects of diffusion are minimal. That is, each particle has explored only a small part of the graph. In this limit, the weighted metric on the graph exhibits the same behavior as the standard metric, and topological shortest paths are recovered, as the following two results show.

Theorem 2 (Shortest-path dominance for the fractional heat kernel as $t \rightarrow 0$). Let $G = (V, E)$ be a simple undirected graph with combinatorial Laplacian $L = D - A$, and let $0 < \alpha < 1$. Denote by $d(v,w)$ the graph distance between vertices $v \neq w$. Then, for every $v \neq w$,

$$(E_{\alpha}(-t^{\alpha}L))_{vw} = \frac{(-t^{\alpha})^{d(v,w)}}{\Gamma(\alpha d(v,w) + 1)} (L^{d(v,w)})_{vw} + O(t^{\alpha(d(v,w)+1)}), \quad t \downarrow 0.$$

In particular,

$$(E_{\alpha}(-t^{\alpha}L))_{vw} = O(t^{\alpha d(v,w)}),$$

and the leading-order contribution is determined by topological shortest paths between v and w . Moreover,

$$(E_\alpha(-t^\alpha L))_{vw} = \frac{t^\alpha d(v,w)}{\Gamma(\alpha d(v,w) + 1)} (A^{d(v,w)})_{vw} + O(t^{\alpha(d(v,w)+1)}),$$

where $(A^{d(v,w)})_{vw}$ is the number of shortest walks of length $d(v,w)$ from v to w .

Proof. The Mittag-Leffler function admits the operator series representation

$$E_\alpha(-t^\alpha L) = \sum_{m=0}^{\infty} \frac{(-t^\alpha)^m}{\Gamma(\alpha m + 1)} L^m,$$

which converges absolutely for all $t \geq 0$ since L is bounded (see, e.g., [5]).

Taking the (v,w) entry yields

$$(E_\alpha(-t^\alpha L))_{vw} = \sum_{m=0}^{\infty} \frac{(-t^\alpha)^m}{\Gamma(\alpha m + 1)} (L^m)_{vw}.$$

As in the classical heat-kernel case, write $L = D - A$. Any term contributing to $(L^m)_{vw}$ corresponds to a product of m factors, each equal to either D or $-A$. Diagonal factors D do not change the vertex index, while each factor A corresponds to a single edge traversal. Hence, in order for $(L^m)_{vw}$ to be nonzero, the word must contain at least $d(v,w)$ factors of A . Consequently,

$$(L^m)_{vw} = 0 \quad \text{for all } m < d(v,w).$$

For $m = d(v,w)$, the only contributing word is $(-A)^{d(v,w)}$, since any appearance of D would prevent reaching w in exactly $d(v,w)$ steps. Therefore,

$$(L^{d(v,w)})_{vw} = (-1)^{d(v,w)} (A^{d(v,w)})_{vw}.$$

Substituting into the series above, the first nonzero term occurs at $m = d(v,w)$, which proves the stated expansion. \square

Theorem 3 (Short-time selection of shortest paths). *Let G be a finite connected graph and let $X(t)$ be the time-fractional continuous-time random walk associated with the Caputo fractional diffusion equation on G (with $0 < \alpha < 1$). Let $N(t)$ be the number of jumps of $X(\cdot)$ up to time t , and let $d(v,w)$ be the graph distance between distinct vertices $v \neq w$. Then, for every $v \neq w$,*

$$\lim_{t \downarrow 0} \mathbb{P}(N(t) = d(v,w) \mid X(t) = w, X(0) = v) = 1.$$

That is, conditioned on arrival at w at very short times, the process selects a topologically shortest path with probability tending to 1.

Proof. Write $d = d(v,w)$. Let Y_m be the embedded discrete-time jump chain of $X(t)$. The fractional CTRW representation yields

$$\mathbb{P}(X(t) = w \mid X(0) = v) = \sum_{m=0}^{\infty} \mathbb{P}(N(t) = m) \mathbb{P}(Y_m = w \mid Y_0 = v),$$

see [65,66]. Let P be the one-step transition matrix of Y_m , so that $\mathbb{P}(Y_m = w \mid Y_0 = v) = (P^m)_{vw}$. By definition of graph distance, $(P^m)_{vw} = 0$ for all $m < d$. Hence,

$$\mathbb{P}(X(t) = w \mid X(0) = v) = \sum_{m=d}^{\infty} \mathbb{P}(N(t) = m) (P^m)_{vw}.$$

For the time-fractional walk, the jump-count distribution satisfies the short-time scaling

$$\mathbb{P}(N(t) = m) = \frac{t^{\alpha m}}{\Gamma(\alpha m + 1)} + O(t^{\alpha(m+1)}), \quad t \downarrow 0,$$

see [66]. Therefore, the leading contribution comes from $m = d$, and we obtain

$$\mathbb{P}(X(t) = w \mid X(0) = v) = \frac{t^{\alpha d}}{\Gamma(\alpha d + 1)} (P^d)_{vw} + O(t^{\alpha(d+1)}).$$

Similarly,

$$\mathbb{P}(N(t) = d, X(t) = w \mid X(0) = v) = \frac{t^{\alpha d}}{\Gamma(\alpha d + 1)} (P^d)_{vw} + O(t^{\alpha(d+1)}).$$

Dividing the two expansions yields

$$\mathbb{P}(N(t) = d \mid X(t) = w, X(0) = v) = 1 + O(t^\alpha),$$

which proves the claim.

Finally, on the event $\{N(t) = d, X(t) = w, X(0) = v\}$, the jump sequence (Y_0, \dots, Y_d) has length d and connects v to w ; hence, it is a topological shortest path. \square

Remark 4 (Physical interpretation of short-time path selection). *The previous results (Theorems 2 and 3) show that, for both classical and time-fractional diffusion on graphs, the short-time behavior is governed by topological constraints rather than by long-time transport mechanisms. In the fractional case, memory effects and heavy-tailed waiting times manifest themselves through the slower time scaling $t^{\alpha d(i,j)}$ of transition probabilities; however, they do not alter the mechanism by which mass first propagates across the graph. At very short times, the process has insufficient opportunity to perform redundant or backtracking moves, and any realization that reaches a vertex j from i must therefore use the minimal number of jumps permitted by the graph distance. Thus, topologically shortest paths dominate not because they are energetically or entropically preferred, but because they are the only dynamically admissible routes in the short-time regime. Memory affects when such paths become observable, but not which paths contribute to the leading-order behavior. This provides a precise mathematical explanation for the observed agreement between diffusion-based distances and graph distances at very small times, even in the presence of anomalous (fractional) temporal dynamics.*

Using Theorem 2 for two adjacent vertices v, w , we can compute the first-order term of the subdiffusive distance for $t \rightarrow 0$

$$\begin{aligned} \mathcal{D}_{\alpha,t}(v,w) &= E_\alpha(-t^\alpha L)_{vv} + E_\alpha(-t^\alpha L)_{ww} - 2E_\alpha(-t^\alpha L)_{vw} \\ &= 2 - \frac{t^\alpha}{\Gamma(\alpha + 1)} (L_{vv} + L_{ww} - 2L_{vw}) + O(t^{2\alpha}) \\ &= 2 - \frac{t^\alpha}{\Gamma(\alpha + 1)} (d_v + d_w + 2) + O(t^{2\alpha}). \end{aligned} \quad (26)$$

Note that usually there are multiple topologically shortest paths between the same pair of vertices. Theorem 4 describes which of these paths coincide with the subdiffusive shortest path as $t \rightarrow 0$. We need the following definition: for an edge (v, w) , its *edge degree* is

$$\delta_{(v,w)} = d_v + d_w - 2. \quad (27)$$

Theorem 4. Let $G = (V, E)$ be a simple, finite, undirected graph and let $v \neq w$ be two vertices. Let $P = (v_0, \dots, v_l)$ be the subdiffusive shortest path according to the metric $\mathcal{D}_{\alpha,t}$ between $v = v_0$ and $w = v_l$, passing through l edges. Then, in the limit $t \rightarrow 0$, P is a topological shortest path between v and w that maximizes the sum of the edge degrees

$$\sum_{k=0}^{l-1} \delta_{(v_k, v_{k+1})} = d_v + 2 \sum_{k=1}^{l-1} d_{v_k} + d_w - 2l. \quad (28)$$

Proof. The length of the edge (v_k, v_{k+1}) is the square root of the communicability distance $\mathcal{D}_{\alpha,t}(v_k, v_{k+1})$. In the limit $t \rightarrow 0$, from (26) we obtain

$$W_{\alpha,t}(v_k, v_{k+1}) = \sqrt{2} - \frac{\sqrt{2}}{4} \frac{t^\alpha}{\Gamma(\alpha + 1)} (\delta_{(v_k, v_{k+1})} + 4) + O(t^{2\alpha}). \quad (29)$$

Therefore, the length of P is the sum of the lengths of all its constituent edges

$$W_{\alpha,t}(P) = \sqrt{2}l \left(1 + \frac{t^\alpha}{\Gamma(\alpha + 1)} \right) - \frac{\sqrt{2}}{4} \frac{t^\alpha}{\Gamma(\alpha + 1)} \sum_{k=0}^{l-1} \delta_{(v_k, v_{k+1})} + O(t^{2\alpha}),$$

The first term depends only on l , which in the limit $t \rightarrow 0$ is the length of (any) topologically shortest path between v and w (Theorems 2 and 3). Therefore, P is the subdiffusive shortest path if and only if it minimizes the second term, or equivalently, if P has the maximum sum of the edge degrees $\sum_{k=0}^{l-1} \delta_{(v_k, v_{k+1})}$. \square

5.3. Subdiffusive Shortest Paths on a Geometric Graph

We computationally analyze our previous analytical findings on an example graph. We focus our attention on the family of large-world networks for two reasons. First, large-diameter networks make it easier to differentiate between diffusion and subdiffusion both numerically and visually. Second, in small-diameter networks, both subdiffusive and topological shortest paths comprise very few edges relative to the network's total size, making them difficult to distinguish.

Specifically, we consider a randomly generated Gabriel graph with $n = 600$ vertices and 1156 edges. We use this graph because it is geometric, in the sense that its vertices are embedded in \mathbb{R}^d , which allows for clear visual interpretation, particularly when illustrating paths. Gabriel graphs are defined as follows.

Definition 3. Let $P \subset \mathbb{R}^d$ be a finite set of points, referred to as generators. The Gabriel graph $G_G = (V_G, E_G)$ associated with P has vertex set $V_G = P$. Two distinct vertices $v, w \in P$ are connected by an edge $\{v, w\} \in E_G$ if and only if the closed ball having the segment $[vw]$ as its diameter contains no other points of P .

In this work, we restrict our attention to the planar case $d = 2$ and embed the points in a rectangle with a length-to-width ratio of 2:1. We geometrize the Gabriel graph using both the subdiffusive communicability distance $\mathcal{D}_{\alpha,t}(v, w)$ and its sum-of-exponentials (SOE) approximation. Specifically, we use

$$F_J(t, L) = \sum_{j=1}^J a_j e^{-t^\alpha b_j L},$$

and introduce the corresponding approximate squared distance

$$\tilde{\mathcal{D}}_{J,\alpha,t}(v, w) = (F_J(t, L))_{vv} + (F_J(t, L))_{ww} - 2(F_J(t, L))_{vw}. \quad (30)$$

Throughout this section, we study a low-memory regime with $\alpha = 0.85$. Using the square roots of both $\mathcal{D}_{\alpha,t}(v,w)$ and $\tilde{\mathcal{D}}_{J,\alpha,t}(v,w)$, we construct the weighted graph as in Definition 2 and compute the subdiffusive shortest paths.

Experimental Results

We begin by reporting the topological shortest paths (TSPs) between two vertices located near opposite corners of the Gabriel graph (see Figure 3a). As is typical—even for planar graphs such as Gabriel graphs—there exist multiple TSPs between a given pair of vertices. These TSPs are colored according to the *average edge degree* along the path (see Equation (27)).

We next consider the shortest paths induced by $\tilde{\mathcal{D}}_{J,\alpha,t}(v,w)$ for $J = 1$ (see Figure 3b). This case corresponds to a standard diffusive process, since it involves a single exponential term, which is the solution of the classical diffusion equation. We observe that, while at early times the diffusive paths coincide with the TSP as predicted by Theorems 2 and 3, they progressively deviate as time increases.

As the number of exponentials J in the SOE approximation increases, the corresponding shortest paths converge toward the TSP, as illustrated in panels (C)–(E) of Figure 3. The exact solution based on the Mittag-Leffler function is shown in Figure 3f, and clearly demonstrates that the subdiffusive shortest paths coincide with the TSPs. Moreover, the limiting paths correspond to those TSPs with the largest average edge degree, in agreement with the analytical results derived in Theorem 4.

The time evolution of the vector probing errors, $\text{relerr}(u_0(t))$ and $\text{masserr}(u_0(t))$, is displayed in Figure 2A. The mass error is at the level of machine precision, confirming that mass conservation of the SOE approximation is numerically guaranteed. The relative error is zero at $t = 0$ (since the two initial conditions coincide); as t grows, the error increases and then remains above a certain threshold, which indicates that the two solutions $u^*(t)$ and $u^{\text{SOE}}(t)$ are meaningfully distinct. Figure 2B illustrates how the time averages of the operator and vector probing errors decay as J increases. Both errors initially decrease before stagnating around 10^{-5} and 10^{-7} , respectively. The mass error remains on the order of machine precision and is therefore omitted.

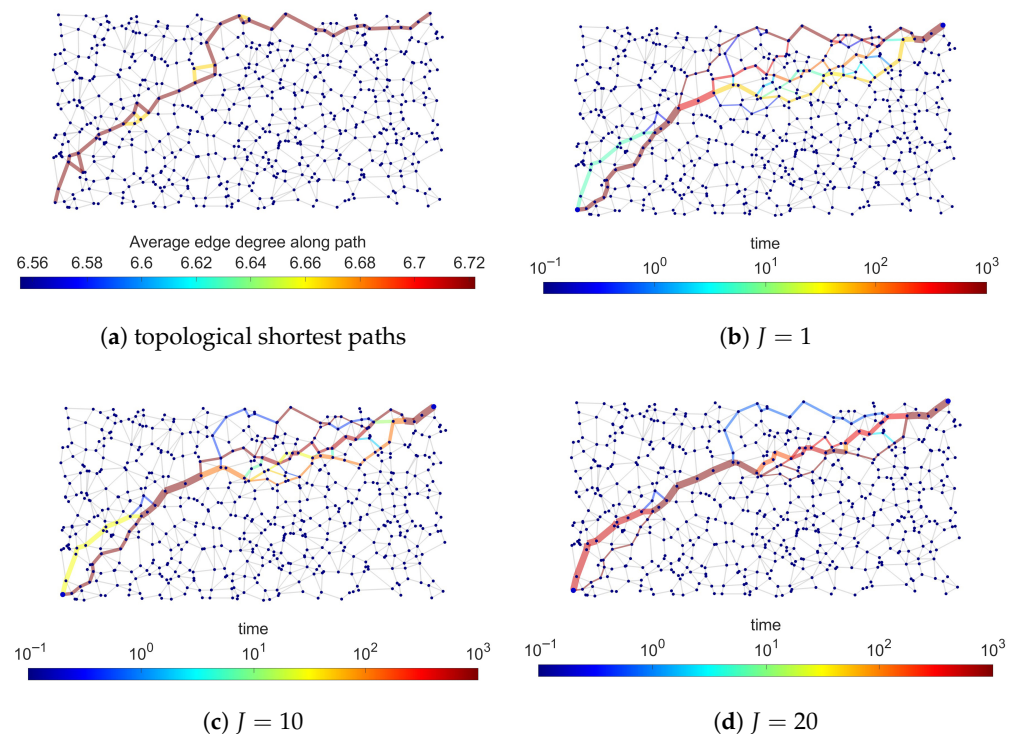


Figure 3. Cont.

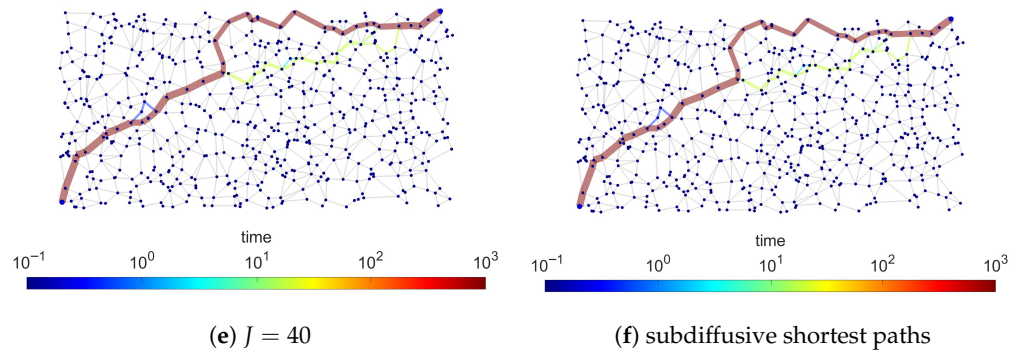


Figure 3. (a) Illustration of all topological shortest paths between a fixed pair of vertices in the example Gabriel graph, colored by average edge degree. (b–e) Shortest paths of SOE approximation for $J = 1, 10, 20, 40$, respectively, with $\alpha = 0.85$. (f) Subdiffusive shortest path between the same vertex pair derived from the Mittag-Leffler function. In (b–f), the paths are colored according to the observation time ($0.1 \leq t \leq 1000$, sampled in 300 steps). The edge thickness is proportional to the frequency with which an edge is traversed by the corresponding paths.

5.4. Physical Interpretation: Memory Reinforcement of Past States

It is clear that, from a mathematical point of view, increasing J improves the accuracy of the SOE approximation to the Mittag-Leffler function, leading to convergence of the corresponding shortest paths. To interpret this convergence from a physical perspective, recall that the SOE represents a superposition of diffusive processes. Let $b_p = \max\{b_j : 1 \leq j \leq J\}$. Then we can rewrite the SOE operator as

$$F_J(t, L) = \sum_{j=1}^J a_j e^{-b_j t^\alpha L} = \sum_{j=1}^J a_j e^{-b_p \frac{b_j}{b_p} t^\alpha L} = \sum_{j=1}^J a_j e^{-b_p t_j^\alpha L}, \tag{31}$$

where the effective time instants are given by $t_j = \left(\frac{b_j}{b_p}\right)^{\frac{1}{\alpha}} t$.

In the example graph, the maximum diffusion speed is $b_p = b_2 = 1.53$ (see Table 1). Each term $e^{-b_p t_j^\alpha L}$ therefore represents a diffusion process with speed b_p , evaluated at a slowed time t_j^α . Consequently, subdiffusion can be interpreted not only as a superposition of diffusion processes with different speeds, but also as the sampling of a single diffusion process at multiple past time instants.

Table 1. The ten largest coefficients in magnitude for the a_j of the SOE approximation in the general case of Corollary 1 with $\alpha = 0.85$. These coefficients are displayed for the SOE approximation on the example Gabriel graph.

j	1	2	3	4	5	6	7	8	9	10
a_j	0.286	0.269	0.167	0.094	0.056	0.035	0.0234	0.016	0.011	0.008
b_j	1.19	1.53	0.93	0.725	0.57	0.44	0.34	0.27	0.21	0.16

This behavior reflects the intrinsic memory property of subdiffusion. Unlike classical diffusion, whose evolution depends solely on the current state, subdiffusive dynamics depend on the entire history of the process. As shown in Figure 3, increasing J incorporates a larger number of past states, allowing the process to remember the shortest paths taken in its early-time behavior. In this sense, both the Caputo fractional derivative and subdiffusion can be viewed as time-averaged processes over the system’s past evolution. We propose a memory-based interpretation for the convergence of the subdiffusive shortest paths to the TSP shown in Figure 3: just as the SOE approximation (31) at time t recalls past time

instants of the diffusive process, the shortest paths in the SOE metric at time t are influenced by the shortest paths at earlier times $\tau < t$, which are strongly linked to the TSP as $\tau \rightarrow 0$.

The fractional order α controls the strength of this memory effect. As $\alpha \rightarrow 1$, the fractional diffusion equation reduces to the classical diffusion equation, and memory effects vanish. Conversely, as $\alpha \rightarrow 0$, greater weight is assigned to earlier times. This behavior is illustrated in Figure 4, where we plot the dissimilarities of the subdiffusive shortest paths with respect to the topological shortest path for different values of α and J . The dissimilarities are computed with the Levenshtein distance (also known as *edit distance*), which counts the minimum number of edits (insertions, deletions, or substitutions) needed to transform the edge sequence of one path into the other. We observe that smaller values of α require fewer SOE terms to recover the Mittag-Leffler shortest paths. For instance, when $J = 20$, only two distinct paths are observed for $\alpha = 0.25$, whereas multiple paths persist for $\alpha = 0.85$. This is the reason for choosing $\alpha = 0.85$ in the experiments: it is far enough from 1 to allow memory effects, but it is not small enough for memory to completely suppress the discovery of alternative paths.

The contrast between the diffusive (Figure 3a) and subdiffusive (Figure 3f) behaviors is striking. When memory is not present, the diffusion-based geometrization explores the graph, as the multiple possible paths show. The inception of memory allows one to recall the topologically shortest path, and gravitate towards it. Memory is strengthened either by lowering the parameter α , or by considering more terms (i.e., by increasing J), as each term is a recall to a previous time instant. Stronger memory means that the Levenshtein distance to the TSP is smaller, as shown in Figure 3.

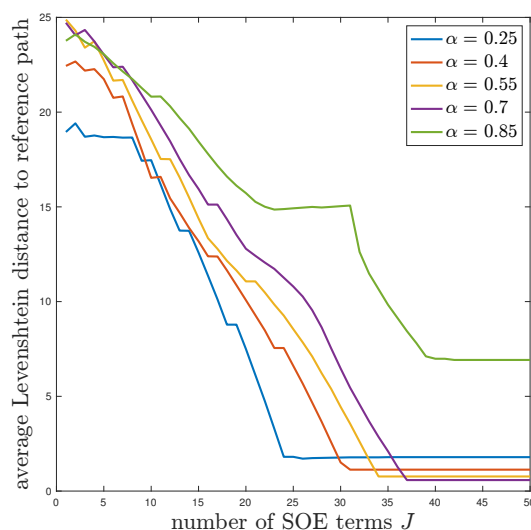


Figure 4. Comparison of the average Levenshtein distance between the shortest subdiffusive paths and the topologically shortest path in the example Gabriel graph, for five values of α . Average over 300 time instants in the interval $0.1 \leq t \leq 1000$.

6. Caputo Fractional Derivative and Memory

We have seen that memory dictates the behavior of subdiffusion and the underlying fractional-time differential equation. In this section, we analyze the memory contributions present in the Caputo derivative, distinguishing between the influence of the remote past and more recent times. We show an underlying connection between these effects and the convexity of the solution to the subdiffusion equation.

6.1. Remote Versus Recent Memory in the Caputo Fractional Derivative

Definition 4. Consider the time-fractional Caputo derivative of a function $x(t)$, as defined in (3). Divide the interval $[0, t]$ into two equal halves, and consider the contribution from each subinterval

- Remote past:

$$\mathcal{R}(\alpha) = \frac{1}{\Gamma(1-\alpha)} \int_0^{t/2} \frac{x'(\tau)}{(t-\tau)^\alpha} d\tau;$$

- Recent past (which also includes the 'present' at time $\tau = t$):

$$\mathcal{P}(\alpha) = \frac{1}{\Gamma(1-\alpha)} \int_{t/2}^t \frac{x'(\tau)}{(t-\tau)^\alpha} d\tau.$$

When $x(t)$ is the solution of the fractional diffusion Equation (1), we consider $x_i(t)$, the i -th component of the vector $x(t)$. Denote by $\mathcal{R}^{(i)}(\alpha)$ and $\mathcal{P}^{(i)}(\alpha)$ its 'remote past', and 'recent past' contributions, respectively.

Definition 5. For a fixed vertex i and fixed $t > 0$, we say that vertex i recalls its remote past more strongly than its recent past on $[0, t]$ if

$$\mathcal{R}^{(i)}(\alpha) > \mathcal{P}^{(i)}(\alpha),$$

and we say that i recalls its recent past more strongly than its remote past if

$$\mathcal{R}^{(i)}(\alpha) < \mathcal{P}^{(i)}(\alpha).$$

A vertex may exhibit both behaviors, depending on whether the derivative $x'_i(t)$ is increasing or decreasing, as shown in the following results.

Theorem 5 (Memory Bias). Let $x(t) = (x_1(t), \dots, x_n(t))^T$ be the solution of the fractional diffusion Equation (1). Fix a vertex $i \in V$ and a time $t > 0$.

- If the derivative $x'_i(\tau)$ is non-decreasing for $\tau \in [0, t]$, then for any $\alpha \in [0, 1)$

$$\mathcal{R}^{(i)}(\alpha) \leq \mathcal{P}^{(i)}(\alpha),$$

with strict inequality if $\alpha \in (0, 1)$, or if $\alpha = 0$ and $x'_i(\tau)$ is strictly increasing on $[0, t]$. In this case, vertex i recalls its recent past more strongly than its remote past on $[0, t]$.

- If the derivative $x'_i(\tau)$ is non-increasing for $\tau \in [0, t]$, then in the limit $\alpha \rightarrow 0$

$$\mathcal{R}^{(i)}(0) \geq \mathcal{P}^{(i)}(0),$$

with strict inequality if $x'_i(\tau)$ is strictly decreasing on $[0, t]$. In this case, vertex i recalls its remote past more strongly than its recent past on $[0, t]$.

Proof. The proof is purely scalar and uses only the behavior of $x'_i(t)$ on the interval $[0, t]$; the graph structure enters only through the fact that $x'_i(t)$ arises from the fractional diffusion equation.

For the first part, by assumption $x'_i(\tau)$ is non-decreasing on $[0, t]$, therefore

$$\frac{x'_i(\tau)}{(t-\tau)^\alpha} \leq \frac{x'_i(\tau + \frac{t}{2})}{(t - \frac{t}{2} - \tau)^\alpha}, \quad \text{for } \tau \in [0, t/2].$$

For $\alpha \in (0, 1)$, this inequality is strict, while for $\alpha = 0$, it is strict only if $x'_i(\tau)$ is strictly non-decreasing (i.e., increasing). By integrating over the interval $[0, t/2]$ and performing a change of variables $\tau' = \tau + \frac{t}{2}$, we obtain

$$\begin{aligned}\mathcal{R}^{(i)}(\alpha) &= \frac{1}{\Gamma(1-\alpha)} \int_0^{t/2} \frac{x'_i(\tau)}{(t-\tau)^\alpha} d\tau \\ &\leq \frac{1}{\Gamma(1-\alpha)} \int_0^{t/2} \frac{x'_i(\tau + \frac{t}{2})}{(t - \frac{t}{2} - \tau)^\alpha} d\tau \\ &= \frac{1}{\Gamma(1-\alpha)} \int_{t/2}^t \frac{x'_i(\tau')}{(t-\tau')^\alpha} d\tau' = \mathcal{P}^{(i)}(\alpha).\end{aligned}$$

For the second part, in the limit $\alpha \rightarrow 0$, the weight of the multiplying kernel becomes $\lim_{\alpha \rightarrow 0} \frac{1}{(t-\tau)^\alpha} = 1$. Thus, for $\alpha = 0$, the remote and recent past contributions are, respectively,

$$\mathcal{R}^{(i)}(0) = \int_0^{t/2} x'_i(\tau) d\tau, \quad \text{and} \quad \mathcal{P}^{(i)}(0) = \int_{t/2}^t x'_i(\tau) d\tau.$$

By the assumption that $x'_i(\tau)$ is non-increasing on $[0, t]$, the integrand of the first equation is greater than or equal to the integrand of the second; therefore, we obtain

$$\mathcal{R}^{(i)}(0) \geq \mathcal{P}^{(i)}(0).$$

The inequality is strict if $x'_i(\tau)$ is strictly non-increasing (i.e., decreasing). \square

Corollary 2. Let $x(t)$ be as in Theorem 5, and fix a vertex $i \in V$. Let $[t_1, t_2]$ and $[t_3, t_4]$ be two disjoint time intervals with $0 \leq t_1 < t_2 \leq t_3 < t_4$. Define the corresponding left and right recent contributions for $\alpha = 0$:

$$\begin{aligned}\mathcal{R}^{(i,1)}(0) &:= \frac{1}{\Gamma(1-\alpha)} \int_{t_1}^{\frac{t_1+t_2}{2}} \frac{x'_i(\tau)}{(t-\tau)^\alpha} d\tau, & \mathcal{P}^{(i,2)}(0) &:= \frac{1}{\Gamma(1-\alpha)} \int_{\frac{t_1+t_2}{2}}^{t_2} \frac{x'_i(\tau)}{(t-\tau)^\alpha} d\tau, \\ \mathcal{R}^{(i,3)}(0) &:= \frac{1}{\Gamma(1-\alpha)} \int_{t_3}^{\frac{t_3+t_4}{2}} \frac{x'_i(\tau)}{(t-\tau)^\alpha} d\tau, & \mathcal{P}^{(i,4)}(0) &:= \frac{1}{\Gamma(1-\alpha)} \int_{\frac{t_3+t_4}{2}}^{t_4} \frac{x'_i(\tau)}{(t-\tau)^\alpha} d\tau.\end{aligned}$$

Assume that

- $x'_i(t)$ is strictly decreasing on $[t_1, t_2]$;
- $x'_i(t)$ is strictly increasing on $[t_3, t_4]$.

Then

$$\mathcal{R}^{(i,1)}(0) > \mathcal{P}^{(i,2)}(0), \quad \text{and} \quad \mathcal{R}^{(i,3)}(0) < \mathcal{P}^{(i,4)}(0).$$

In particular, on $[t_1, t_2]$, vertex i recalls its more distant past more strongly than its more recent past, whereas on $[t_3, t_4]$ the bias is reversed.

Proof. Define $z_1(s) = x_i(t_1 + s)$ on $[0, t_2 - t_1]$ and $z_2(s) = x_i(t_3 + s)$ on $[0, t_4 - t_3]$. Then z'_1 is strictly decreasing and z'_2 is strictly increasing on their respective domains, and the points $s_j^{(1)}, s_j^{(2)}$ define uniform grids. Applying Theorem 5 to z_1 and z_2 yields the desired inequalities. \square

Corollary 3. Let $x(t)$ be as in Theorem 5, and consider two vertices $i, j \in V$. Fix a time $t > 0$ and, for $\alpha = 0$, consider the remote and recent past contributions of the two vertices:

$$\mathcal{R}^{(i)}(0), \mathcal{P}^{(i)}(0), \mathcal{R}^{(j)}(0), \mathcal{P}^{(j)}(0).$$

Assume that

- $x'_i(\tau)$ is strictly decreasing on $[0, t]$;
- $x'_j(\tau)$ is strictly increasing on $[0, t]$.

Then

$$\mathcal{R}^{(i)}(0) > \mathcal{P}^{(i)}(0), \quad \text{and} \quad \mathcal{R}^{(j)}(0) < \mathcal{P}^{(j)}(0).$$

Thus, over the same time interval $[0, t]$ and for the same fractional model, vertex i exhibits a remote-past memory bias, whereas vertex j exhibits a recent-past memory bias.

Proof. Apply Theorem 5 to the scalar functions $\tau \mapsto x_i(\tau)$ and $\tau \mapsto x_j(\tau)$ on the interval $[0, t]$. \square

The recent past term $\mathcal{P}(\alpha)$ incorporates the contribution of the ‘present’ instant $\tau = t$; however, in the integral formulation, the value of the function at a single point does not affect the integral. Thus, one may consider the following discretization.

Definition 6. For odd k , let the interval $[0, t]$ be partitioned into k subintervals $[t_j, t_{j+1}]$ of equal length $h = t/k$, where $j = 0, \dots, k-1$. We define the following specific regions within $[0, t]$:

- Remote past: $\mathcal{R}(k, \alpha) = h^{1-\alpha} \sum_{j=0}^{(k-1)/2} c_j(k, \alpha) x'(t_j)$;
- Recent past: $\mathcal{P}(k, \alpha) = h^{1-\alpha} \sum_{j=(k-1)/2}^{k-1} c_j(k, \alpha) x'(t_j)$;
- Present: $\mathcal{P}_{pr}(k, \alpha) = h^{1-\alpha} c_k(k, \alpha) x'(t)$;

where the coefficients $c_j(k)$ are appropriately determined by the trapezoidal rule.

The accuracy of this quadrature rule for $h \rightarrow 0$ was shown by Odibat [67].

Theorem 6. The fractional Caputo derivative $D_t^\alpha x(t)$ can be expressed as

$$D_t^\alpha x(t) = \mathcal{R}(k, \alpha) + \mathcal{P}(k, \alpha) + \mathcal{P}_{pr}(k, \alpha) + E_C(f, h, \alpha), \quad (32)$$

where $E_C(f, h, \alpha) = \mathcal{O}(h^2)$ is the error term.

In the limit $\alpha \rightarrow 1$, the contributions of $\mathcal{P}(k, \alpha)$ and $\mathcal{R}(k, \alpha)$ vanish; the only surviving contribution is $\mathcal{P}_{pr}(k, \alpha)$, which reduces to $x'(t)$.

6.2. Physical Interpretation: Memory Regimes on a Graph

The Memory Bias Theorem shows that, within the Caputo-driven fractional diffusion dynamics on a graph for $\alpha \rightarrow 0$, the relative importance of the “remote” versus “recent” parts of the past is not uniform across the network. Instead, it depends sensitively on the local temporal curvature of the solution at each vertex. That is, the sign of the second derivative $x''_i(t)$ determines whether $x'_i(t)$ is increasing or decreasing.

If the temporal derivative $x'_i(t)$ is increasing within a given window, the trajectory $x_i(t)$ is bending upwards, and older information within that window receives a smaller weight than more recent information. In this regime, vertex i is said to exhibit an *recent-past memory bias*. Our results show that the recent past dominates the remote past for any $\alpha \in [0, 1)$.

Conversely, if $x'_i(t)$ is decreasing on a given window, the trajectory is bending downwards, and the older information receives a larger weight than the more recent information. In the $\alpha \rightarrow 0$ limit, the vertex exhibits a *remote-past memory bias*.

A striking consequence is that different vertices on the same graph may simultaneously reside in opposite memory regimes, even though they are driven by the same fractional dynamics and are evaluated over the same time interval. Similarly, the *same* vertex may switch memory regimes over time, depending on the evolution of its temporal curvature. This expresses a fundamental “heterogeneity of memory” in fractional diffusion on graphs.

In summary, the fractional order determines *how much* of the past is remembered globally, but the *shape of the temporal evolution* at each vertex dictates *which* part of the past (remote versus recent) is preferentially recalled. This generates rich, spatially distributed memory patterns that reflect both the graph geometry and the initial configuration.

7. Emergence of Memory from Fractional Diffusion on a Graph

Once we have seen how convexity influences the recall of the early or late past, we now study how convexity can emerge in a network, depending on the initial mass distribution of the process. The strong influence of memory is also shown through the underlying random clock process.

7.1. Local Convexity and Concavity in Caputo Fractional Diffusion: A Mittag-Leffler and SOE-Based Analysis

We consider the Caputo fractional diffusion equation on a finite graph G with combinatorial Laplacian $L = D - A$:

$$D_t^\alpha x(t) = -Lx(t), \quad x(0) = e_v,$$

where $0 < \alpha < 1$ and e_v is the v th standard basis vector. The mild solution is given by the Mittag-Leffler matrix function

$$x(t) = E_\alpha(-t^\alpha L) e_v.$$

We are interested in the *curvature* of the time evolution at early times, i.e., the sign of $x_i''(t)$ for sufficiently small $t > 0$. This sign determines whether the temporal trajectory at a vertex is locally convex or concave. We show below that, regardless of the graph, the vertex where the mass is initially placed has an early-time *convex and decreasing* profile, while every neighbor exhibits an early-time *concave and increasing* profile. Both statements hold rigorously for every $0 < \alpha < 1$.

Theorem 7 (Local convexity/concavity from Mittag-Leffler and SOE). *Let G be a finite graph with Laplacian $L = D - A$. Consider the Caputo fractional diffusion equation*

$$D_t^\alpha x(t) = -Lx(t), \quad 0 < \alpha < 1, \quad x(0) = e_v.$$

Then:

- *At vertex v , the time evolution $x(t)$ is strictly decreasing and strictly convex for all $t \in (0, \varepsilon_v)$, for some $\varepsilon_v > 0$.*
- *If w is any neighbor of v , i.e., $A_{wv} > 0$, then $u_w(t)$ is strictly increasing and strictly concave for all $t \in (0, \varepsilon_w)$, for some $\varepsilon_w > 0$.*

In particular, there exists $\varepsilon > 0$ such that for all $t \in (0, \varepsilon)$ the excited vertex v exhibits a convex decay, while each of its neighbors exhibits a concave growth.

Proof. We use two complementary arguments: (i) the small-time expansion of the Mittag-Leffler matrix function, and (ii) a sum-of-exponentials (SOE) representation for E_α .

Step 1: Exact short-time Mittag-Leffler expansion. The matrix Mittag-Leffler function admits the convergent series

$$E_\alpha(-t^\alpha L) = I - \frac{t^\alpha}{\Gamma(\alpha + 1)}L + \frac{t^{2\alpha}}{\Gamma(2\alpha + 1)}L^2 + O(t^{3\alpha}).$$

Applying this to the initial condition $x(t) = E_\alpha(-t^\alpha L)e_v$ yields

$$x(t) = e_v - \frac{t^\alpha}{\Gamma(\alpha + 1)}Le_v + O(t^{2\alpha}).$$

At the excited vertex v we have

$$x_v(t) = 1 - \frac{\deg(v)}{\Gamma(\alpha + 1)}t^\alpha + O(t^{2\alpha}),$$

because $(Le_v)_v = \deg(v)$. Differentiating yields

$$x'_v(t) = -\frac{\deg(v)}{\Gamma(\alpha + 1)}\alpha t^{\alpha-1} + O(t^{2\alpha-1}) < 0,$$

$$x''_v(t) = -\frac{\deg(v)}{\Gamma(\alpha + 1)}\alpha(\alpha - 1)t^{\alpha-2} + O(t^{2\alpha-2}) > 0,$$

since $\alpha(\alpha - 1) < 0$ for $0 < \alpha < 1$. Note that the exponents $\alpha - 1$, $2\alpha - 2$ are negative. Thus $x_v(t)$ is strictly decreasing and strictly convex on $(0, \varepsilon_v)$.

At a neighbor w of v ,

$$x_w(t) = -\frac{t^\alpha}{\Gamma(\alpha + 1)}(Le_v)_w + O(t^{2\alpha}) = \frac{A_{wv}}{\Gamma(\alpha + 1)}t^\alpha + O(t^{2\alpha}),$$

because $(Le_v)_w = -A_{wv}$ for $w \neq v$. Thus

$$x'_w(t) = \frac{A_{wv}}{\Gamma(\alpha + 1)}\alpha t^{\alpha-1} + O(t^{2\alpha-1}) > 0,$$

$$x''_w(t) = \frac{A_{wv}}{\Gamma(\alpha + 1)}\alpha(\alpha - 1)t^{\alpha-2} + O(t^{2\alpha-2}) < 0.$$

Hence, $x_w(t)$ is strictly increasing and strictly concave on $(0, \varepsilon_w)$.

Step 2: Interpretation via SOE approximation. We have developed an SOE scheme to approximate the Mittag-Leffler function by

$$E_\alpha(-\lambda t^\alpha) \approx \sum_{m=1}^M c_m e^{-d_m t}, \quad c_m > 0, d_m > 0.$$

Applying the same SOE to the matrix L yields

$$x(t) = E_\alpha(-t^\alpha L)e_v \approx \sum_{m=1}^M c_m e^{-d_m t L} e_v.$$

Each term $e^{-d_m t L} e_v$ is the solution of a *classical* diffusion equation with rate d_m and is therefore strictly convex at the source v and strictly concave at neighbors w at early times. Because all coefficients c_m are positive, the SOE sum preserves the convexity at v and concavity at its neighbors. This matches exactly the signs obtained in the rigorous Mittag-Leffler expansion above.

Combining Steps 1 and 2 proves (i) and (ii).

Thus, over the same time interval $t \in (0, \varepsilon_v)$ and for the same fractional model, vertex v exhibits a remote-past memory bias, i.e., it recalls more the past than the present, whereas vertex w exhibits a recent-past memory bias, i.e., it recalls more the present than the past. \square

7.2. Random Time Change Induces Subdiffusion and Memory

We summarize the mechanism by which the random clock E_t slows down spreading (*subdiffusion*) and introduces *memory* into the dynamics on a graph. Under the subordination identity,

$$u(t) = E_\alpha(-t^\alpha L) u_0 = \mathbb{E}\left[e^{-E_t L}\right] u_0,$$

the state at physical time t equals the baseline heat state evaluated at the *random* operational time $s = E_t$. For $0 < \alpha < 1$, the inverse α -stable clock satisfies

$$\mathbb{E}[E_t] = \frac{t^\alpha}{\Gamma(1 + \alpha)}, \quad \text{Var}(E_t) = \frac{2 t^{2\alpha}}{\Gamma(1 + 2\alpha)} - \frac{t^{2\alpha}}{\Gamma(1 + \alpha)^2},$$

so the *typical* amount of diffusion time available by physical time t scales like t^α (rather than t). Thus, any diffusive spread measure that, at baseline, scales with s (e.g., mean-square displacement in Euclidean space, or mixing surrogates on graphs), will scale like t^α under the random clock. This is the essence of *subdiffusion*: slower-than-classical spreading ($t^\alpha/2$ instead of $t^{1/2}$ in continuum settings), and on graphs, a slower homogenization than the exponential-in- t decay of the standard heat semigroup.

In the time-changed walk $Y_t = X_{E_t}$, the *survival function* of node i is the tail of the cumulative distribution function of the holding time T_i , given for $t \geq 0$ by:

$$\mathbb{P}(T_i > t) = \int_0^\infty M_\alpha(\theta) e^{-d_i t^\alpha \theta} d\theta = E_\alpha(-d_i t^\alpha), \quad (33)$$

with $d_i = \sum_j A_{ij}$. The Mittag-Leffler tail obeys $E_\alpha(-d_i t^\alpha) \sim \frac{1}{d_i \Gamma(1 - \alpha)} t^{-\alpha}$ for $t \rightarrow \infty$, which is a power law, in contrast to the exponentially decaying tail of the diffusive case. These *rare but very long* pauses act as traps that stretch physical time relative to operational time, producing a subdiffusive spread on the graph.

The survival function can also be computed for the SOE approximation. Recall the numerical quadrature from Section 3.1, with *non-normalized* weights $\tilde{a}_j = h M_\alpha(b_j) b_j$ for $1 \leq j \leq J$. Then, for the operator, we normalize the weights as $a_j = \tilde{a}_j / \sum_{k=1}^J \tilde{a}_k$, while for scalar quantities, we use the non-normalized weights, leading to the *SOE vertex waiting times*

$$S_i^{SOE} = \sum_{j=1}^J \tilde{a}_j e^{-d_i t^\alpha b_j}. \quad (34)$$

From (34), we see that the waiting times in the SOE approximation also exhibit a heavy tail dependent upon the degree d_i ; these are modulated by the global internal-time samples b_j , whose mixtures simultaneously approximate all vertex waiting-time distributions.

The Caputo equation

$$\partial_t^\alpha u(t) = -L u(t), \quad \partial_t^\alpha u(t) = \frac{1}{\Gamma(1 - \alpha)} \int_0^t (t - \tau)^{-\alpha} u'(\tau) d\tau,$$

is explicitly *history dependent*: the instantaneous rate $u'(t)$ depends on the full past with a power-law kernel $(t - \tau)^{-\alpha}$. In renewal terms, the holding-time hazard for Mittag-Leffler waiting times is *decreasing in age*. The longer the process has been waiting, the less likely it

is to jump immediately. Hence, the future depends on how long the current wait has lasted (“aging”), which breaks the Markov property in physical time t . Conditioned on $E_t = s$, the baseline path is Markov; *after averaging over the random clock*, the observed dynamics inherit memory.

On the other hand, with $L = V\Lambda V^\top$, each mode decays as

$$E_\alpha(-t^\alpha \lambda_k) \text{ instead of } e^{-t\lambda_k}.$$

For $t \rightarrow \infty$, $E_\alpha(-t^\alpha \lambda_k) \sim \frac{1}{\lambda_k \Gamma(1-\alpha)} t^{-\alpha}$: *algebraic* decay replaces exponential decay. Hence, mixing, return probabilities, and any observable built from the heat kernel trace decay more slowly, reflecting both subdiffusion (slower spread) and long memory (long tails). The practical consequences on networks are as follows:

- *Slower homogenization*: Community imbalances, gradients, or initial heterogeneities persist longer (power-law tail).
- *Trapping in dense/central regions*: Large d_i increases the attempt rate, but the heavy-tailed clock still produces long residence episodes; dwell-time distributions are broad across nodes.
- *Non-exponential relaxation*: Observables fit Mittag-Leffler or power-law decays rather than exponentials; log-log slopes reveal α .

In summary, the random clock E_t *compresses* operational time from t to t^α on average and introduces *power-law waiting* between moves; together these yield *subdiffusive spreading* and *memory* (aging) in the network dynamics.

8. A Generalized Physico-Mathematical Context of Subdiffusion on Graphs

The fractional-time differential equation that gives rise to subdiffusion is widely known. The SOE is not only the approximation of a function: we show how the superposition of diffusions can emerge in a graph, providing a bridge to graph dynamics, as well as integrating the concepts in the generalized structure of kernel-based differential equations.

8.1. A Multiplex Diffusion

Let us consider that there are J parallel diffusion processes occurring on the graph G , each with a diffusivity constant β_j . Therefore, we can consider a representation of the graph as a *multiplex graph* formed by J layers [68,69], each of them representing the same underlying graph. Every pair of layers l_i and l_j is interconnected by means of inter-layer edges connecting each vertex $v \in l_i$ to its corresponding counterpart in l_j .

Then, we create the super-Laplacian matrix, which was introduced in [70]:

$$\mathcal{L} = \bigoplus_{j=1}^J \beta_j L + \omega I \otimes (J - I), \quad (35)$$

where \otimes is the Kronecker product, $\omega \in \mathbb{R}^+ \cup \{0\}$ is the strength coupling between the layers, J is an all-ones matrix and I is the identity matrix. Then,

$$\mathcal{L} = \begin{pmatrix} \beta_1 L & C_{12} & \cdots & C_{1J} \\ C_{21} & \beta_2 L & \cdots & C_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ C_{J1} & C_{J2} & \cdots & \beta_J L \end{pmatrix}. \quad (36)$$

Let us write the diffusion equation on the multiplex:

$$\partial_t u(t) + \mathfrak{L}u(t) = 0, \quad (37)$$

with initial condition $u(0) = \gamma \otimes \varphi$, where $\gamma = [\gamma_1 \dots \gamma_J]^T$, $\gamma_j \in \mathbb{R}$, and $\varphi \in \mathbb{R}^{n \times 1}$. The solution of the abstract Cauchy problem is then

$$u(t) = e^{-t\mathfrak{L}} u(0). \quad (38)$$

Let us consider a very weak coupling strength between pairs of layers, $\omega \ll 1$, such that we can consider the diffusion at each layer almost independently of the diffusion on other layers:

$$u(t) \cong \oplus_{j=1}^J \gamma_j e^{-t\beta_j L} \varphi. \quad (39)$$

We can then consider the sum of the concentrations at each vertex of the graph in all layers:

$$c(t) = \sum_{j=1}^h u_j(t) = \sum_{j=1}^J \gamma_j e^{-t\beta_j L} \varphi, \quad (40)$$

which is related to the SOE approximation we have found for the Mittag-Leffler Laplacian function. Therefore, the subdiffusive process on the graph can be seen as the aggregate diffusion occurring in parallel at different layers, where each layer has its own diffusivity β_j and the initial conditions on a layer are scalar multiples of the initial conditions on any other layer.

8.2. Factorized High-Order Temporal Diffusion Equation

Let us consider J independent diffusive species $v_1(t), \dots, v_J(t) \in \mathbb{R}^n$ evolving according to

$$\partial_t v_j(t) + \beta_j L v_j(t) = 0, \quad j = 1, \dots, J, \quad (41)$$

with distinct diffusion rates $\beta_j > 0$. We define the observable field

$$u(t) := \sum_{j=1}^J v_j(t). \quad (42)$$

Since L is diagonalizable, it suffices to work in a single Laplacian eigenmode with eigenvalue $\lambda \geq 0$ and corresponding eigenvector $z_\lambda \in \mathbb{R}^n$. Let $x_j(t)$ be the scalar component of $v_j(t)$ associated with the eigenmode λ ; that is, $v_j(t) = \sum_{\lambda \in \text{Sp}(L)} x_j(t) z_\lambda$. Let $y(t) = \sum_{j=1}^J x_j(t)$. The system (41) reduces to

$$\partial_t x_j(t) + \beta_j \lambda x_j(t) = 0, \quad j = 1, \dots, J,$$

whose solutions are $x_j(t) = e^{-\beta_j \lambda t} x_j(0)$. Hence,

$$y(t) = \sum_{j=1}^J e^{-\beta_j \lambda t} x_j(0).$$

Each exponential $e^{-\beta_j \lambda t}$ is annihilated by the operator $(\partial_t + \beta_j \lambda)$, and therefore

$$\prod_{m=1}^J (\partial_t + \beta_m \lambda) y(t) = 0.$$

Since this holds for every Laplacian eigenvalue λ , we obtain, in operator form, the *multiplicative diffusion equation*

$$\prod_{m=1}^J (\partial_t + \beta_m L) u(t) = 0, \quad (43)$$

with suitable initial conditions. Then, we claim that

$$u(t) = \sum_{j=1}^J \gamma_j e^{-t\beta_j L} \varphi.$$

is an exact solution of the *multiplicative diffusion equation*.

8.2.1. Verification

It suffices to check that each term

$$u_j(t) := e^{-t\beta_j L} \varphi$$

lies in the kernel of the operator $\prod_{m=1}^J (\partial_t + \beta_m L)$. First, observe that for each fixed m and j ,

$$(\partial_t + \beta_m L) u_j(t) = (-\beta_j L + \beta_m L) e^{-t\beta_j L} \varphi = (\beta_m - \beta_j) L e^{-t\beta_j L} \varphi.$$

Applying the full product, we obtain

$$\prod_{m=1}^J (\partial_t + \beta_m L) u_j(t) = \left(\prod_{m=1}^J (\beta_m - \beta_j) \right) L^J e^{-t\beta_j L} \varphi.$$

If the β_m are pairwise distinct and $j \in \{1, \dots, J\}$, then one of the factors in the product is $(\beta_j - \beta_j) = 0$. Hence,

$$\prod_{m=1}^J (\partial_t + \beta_m L) u_j(t) = 0 \quad \text{for each } j = 1, \dots, J.$$

By linearity,

$$\prod_{m=1}^J (\partial_t + \beta_m L) u(t) = \sum_{j=1}^J \gamma_j \prod_{m=1}^J (\partial_t + \beta_m L) u_j(t) = 0.$$

Thus, the function

$$u(t) = \sum_{j=1}^J \gamma_j e^{-t\beta_j L} \varphi$$

is an exact solution of (43).

8.2.2. Initial Conditions

Equation (43) is of order J in time, so one can prescribe $u(0), \partial_t u(0), \dots, \partial_t^{J-1} u(0)$. The representation

$$u(t) = \sum_{j=1}^J e^{-t\beta_j L} \psi_j$$

is the general solution of (43), and the vectors ψ_j are uniquely determined from the initial data via a Vandermonde-type linear system (mode by mode in the eigenbasis of L). Choosing

$$\psi_j = \gamma_j \varphi$$

yields precisely

$$u(t) = \sum_{j=1}^J \gamma_j e^{-t\beta_j L} \varphi.$$

Therefore, for any prescribed coefficients β_j , γ_j and profile φ , there exist initial conditions for (43) such that the unique solution is exactly

$$u(t) = \sum_{j=1}^J \gamma_j e^{-t\beta_j L} \varphi.$$

8.2.3. Physical Interpretation

Equation (43) does not describe diffusion of a single species. Rather, it is the *effective evolution equation* satisfied by the total concentration u obtained by summing k independent diffusive species with distinct diffusion rates. The product structure arises from eliminating the hidden fields v_j and encodes the presence of multiple diffusive time scales. Thus, the multiplicative diffusion Equation (43) represents the minimal closed description of a multi-rate diffusion process when only the aggregate observable is accessible.

8.3. Operator-Valued Memory and Diffusion on Graphs

8.3.1. Operator-Valued Volterra Kernels

Let X be a finite-dimensional Hilbert space and denote by $\mathcal{L}(X)$ the space of bounded linear operators on X . An *operator-valued Volterra kernel* is a strongly measurable mapping

$$K : \mathbb{R}_+ \rightarrow \mathcal{L}(X)$$

such that

$$\int_0^T \|K(t)\|_{\mathcal{L}(X)} dt < \infty \quad \text{for all } T > 0.$$

Given such a kernel, the associated Volterra convolution operator acts on sufficiently regular functions $u : \mathbb{R}_+ \rightarrow X$ by

$$(\mathcal{K}u)(t) := \int_0^t K(t-s) u(s) ds, \quad t \geq 0.$$

An evolution equation of the form

$$u'(t) + \mathcal{K}u(t) = 0 \tag{44}$$

is called a *Volterra evolution equation with operator-valued memory*. The Volterra structure enforces causality, while the operator-valued nature of K allows for mode-dependent memory effects.

Remark 5. Equation (44) describes memory acting directly on the state u . Caputo-type formulations, in which the kernel acts on u' , can be interpreted later as arising in a singular limit of this general framework.

8.3.2. Multiplicative Diffusion on Graphs

Let $L \in \mathbb{R}^{n \times n}$ be a symmetric graph Laplacian ($L \succeq 0$ and $\ker L = \text{span}\{\mathbf{1}\}$), and let $\beta_1, \dots, \beta_k > 0$ be pairwise distinct. We consider the *multiplicative diffusion equation*

$$\prod_{j=1}^k (\partial_t + \beta_j L) u(t) = 0, \quad t > 0, \tag{45}$$

supplemented with k initial conditions $u(0), u'(0), \dots, u^{(k-1)}(0)$.

Solutions of (45) admit the representation

$$u(t) = \sum_{j=1}^J e^{-\beta_j t L} \psi_j, \quad (46)$$

where the vectors $\psi_j \in \mathbb{R}^n$ are uniquely determined by the initial conditions. Consequently, the Laplace transform $\hat{u}(s) = \mathcal{L}\{u\}(s)$ satisfies

$$\hat{u}(s) = \sum_{j=1}^J (sI + \beta_j L)^{-1} \psi_j, \quad u(0) = \sum_{j=1}^J \psi_j. \quad (47)$$

8.3.3. Equivalence with an Operator-Valued Memory Equation

We now relate (45) to a Volterra evolution equation of the form

$$u'(t) + \int_0^t K(t-s) L u(s) ds = 0, \quad t > 0, \quad (48)$$

where $K : \mathbb{R}_+ \rightarrow \mathcal{L}(\mathbb{R}^n)$ is an operator-valued Volterra kernel.

Taking the Laplace transform of (48) yields

$$(sI + \hat{K}(s)L) \hat{u}(s) = u(0), \quad (49)$$

for $\Re(s)$ sufficiently large.

Assume for simplicity that $u(0) \in \text{Ran}(L)$, so that L is invertible on the dynamically relevant subspace (more generally, L^{-1} may be replaced by the Moore–Penrose inverse L^\dagger). Define

$$G^{(J)}(s) := \sum_{j=1}^J (sI + \beta_j L)^{-1} P_j, \quad (50)$$

where $P_j : \mathbb{R}^n \rightarrow \mathbb{R}^n$ are linear maps such that $\psi_j = P_j u(0)$. If $G^{(J)}(s)$ is invertible for $\Re(s)$ sufficiently large, we define

$$\hat{K}(s) := L^{-1} (G^{(J)}(s)^{-1} - sI). \quad (51)$$

Then (47) implies

$$(sI + \hat{K}(s)L) \hat{u}(s) = u(0),$$

which coincides with (49). Hence, solutions of the multiplicative diffusion Equation (45) also satisfy the Volterra Equation (48) with operator-valued kernel K .

8.3.4. Spectral Representation

Let $L = U\Lambda U^\top$ with $\Lambda = \text{diag}(\lambda_\ell)$. On each eigenmode $\lambda > 0$, the kernel is characterized by

$$\hat{K}_\lambda(s) = \frac{1}{\lambda} \left(\frac{1}{\sum_{j=1}^J \frac{p_j(\lambda)}{s + \beta_j \lambda}} - s \right), \quad (52)$$

where $p_j(\lambda)$ are the scalar multipliers induced by P_j . Thus, the memory kernel is generally mode-dependent and cannot be reduced to a single scalar kernel.

8.3.5. Fractional Diffusion as a Singular Memory Limit

In the next proposition, we show that the fractional diffusion equation is equivalent to a Volterra equation with a suitably chosen subdiffusion kernel K_S .

Proposition 7. Let $L \in \mathbb{R}^{n \times n}$ be a symmetric, positive semidefinite operator (e.g., a graph Laplacian), and let $0 < \alpha < 1$. Consider the Volterra evolution equation

$$u'(t) + \int_0^t K(t-s) L u(s) ds = 0, \quad t > 0, \quad (53)$$

with initial condition $u(0) = u_0$. If the kernel is chosen as

$$K_S(t) = \frac{1}{\Gamma(\alpha-1)} t^{\alpha-2}, \quad t > 0, \quad (54)$$

then (53) is equivalent to the Caputo fractional diffusion equation

$$D_t^\alpha u(t) = -L u(t), \quad t > 0, \quad (55)$$

where D_t^α denotes the Caputo derivative of order α . Moreover, the unique solution is given by

$$u(t) = E_\alpha(-t^\alpha L) u_0, \quad (56)$$

where E_α is the Mittag-Leffler function.

Proof. Let $\hat{u}(s) = \mathcal{L}\{u\}(s)$ denote the Laplace transform of u . Using the identities

$$\mathcal{L}\{u'\}(s) = s\hat{u}(s) - u_0, \quad \mathcal{L}\left\{\int_0^t K(t-s)Lu(s) ds\right\} = \hat{K}(s)L\hat{u}(s),$$

the Laplace transform of (53) yields

$$(sI + \hat{K}(s)L)\hat{u}(s) = u_0. \quad (57)$$

For the kernel (54), the classical Laplace-transform formula yields

$$\hat{K}_S(s) = s^{1-\alpha}, \quad \Re(s) > 0.$$

Substituting this expression into (57) yields

$$(sI + s^{1-\alpha}L)\hat{u}(s) = u_0.$$

Multiplying both sides by $s^{\alpha-1}$ yields

$$(s^\alpha I + L)\hat{u}(s) = s^{\alpha-1}u_0 \quad (58)$$

On the other hand, the Laplace transform of the Caputo derivative $D_t^\alpha u$ is

$$\mathcal{L}\{D_t^\alpha u\}(s) = s^\alpha \hat{u}(s) - s^{\alpha-1}u_0.$$

Taking the Laplace transform of (55) therefore yields

$$s^\alpha \hat{u}(s) - s^{\alpha-1}u_0 = -L\hat{u}(s),$$

which rearranges to (58). Thus, Equations (53) and (55) are equivalent.

Finally, the expression of $\widehat{u}(s)$ in terms of u_0 can be obtained from (58)

$$\widehat{u}(s) = s^{\alpha-1}(s^\alpha I + L)^{-1}u_0. \quad (59)$$

The inverse Laplace transform of $s^{\alpha-1}(s^\alpha I + L)^{-1}$ is the operator-valued Mittag-Leffler function $E_\alpha(-t^\alpha L)$, which proves (56). \square

Remark 6. We note that the Laplace-domain representation is written in a resolvent form. More generally, finite superpositions of Laplacian semigroups of the form

$$u(t) = \sum_{j=1}^J e^{-\beta_j t L} \psi_j$$

lead (via the construction in (52)) to operator-valued memory kernels whose Laplace symbols are rational in s (mode-by-mode in the spectrum of L). When the number of terms increases and the associated time scales become dense, these rational symbols may converge, in an appropriate sense, to the fractional symbol $s^{\alpha-1}$. In this resolvent sense, time-fractional diffusion can be viewed as a singular, scale-free limit of diffusion with operator-valued Volterra memory.

Note that the kernel $K_S(t)$ is negative, since $0 < \alpha < 1$ and thus $\Gamma(\alpha - 1) < 0$. The stability of the Volterra heat equation has been studied in [71], in a continuous spatial setting, with particular attention to positive and negative kernels. An example of an application with a negative kernel in a stochastic setting is given in [47].

We now formalize the connection between the SOE approximation and fractional diffusion at the level of Laplace-domain resolvents. We emphasize that the result below concerns operator-valued symbols in the Laplace domain and should not be interpreted as a full dynamical equivalence in physical time. In particular, the statement establishes a form of resolvent convergence under suitable assumptions, rather than a complete convergence of time-domain solutions or memory kernels.

Proposition 8. Let $L \in \mathbb{R}^{n \times n}$ be a symmetric graph Laplacian with spectrum $\sigma(L) \subset [0, \lambda_{\max}]$ and let $0 < \alpha < 1$. For s with $\Re(s) > 0$, define the fractional resolvent family

$$\widehat{G}_S(s) := s^{\alpha-1}(s^\alpha I + L)^{-1}. \quad (60)$$

Assume that for each $h \in \mathbb{N}$ we are given coefficients $a_j, b_j > 0$ ($j = 1, \dots, h$) and define the rational operator-valued symbol

$$\widehat{G}^{(J)}(s) := \sum_{j=1}^J a_j (sI + b_j L)^{-1}. \quad (61)$$

Suppose there exists a domain $\Omega \subset \{s \in \mathbb{C} : \Re(s) > 0\}$ and constants $M, \delta > 0$ such that for all $s \in \Omega$:

$$\|\widehat{G}_S(s)\| \leq M, \quad (62)$$

$$\|\widehat{G}_S(s)^{-1}\|_{\text{Ran}(L) \rightarrow \text{Ran}(L)} \leq M, \quad (63)$$

$$\|\widehat{G}^{(J)}(s) - \widehat{G}_S(s)\| \leq \delta, \quad \text{with } \delta M < 1. \quad (64)$$

(Here, $\widehat{G}_S(s)^{-1}$ denotes the inverse on $\text{Ran}(L)$, and the norm is the operator norm on \mathbb{R}^n .)

Then for every $s \in \Omega$, the operator $\widehat{G}^{(J)}(s)$ is invertible on $\text{Ran}(L)$ and

$$\|\widehat{G}^{(J)}(s)^{-1} - \widehat{G}_S(s)^{-1}\|_{\text{Ran}(L) \rightarrow \text{Ran}(L)} \leq \frac{M^2}{1 - \delta M} \|\widehat{G}^{(J)}(s) - \widehat{G}_S(s)\|. \quad (65)$$

Define the SOE-induced operator-valued memory symbol by

$$\widehat{K}^{(J)}(s) := (\widehat{G}^{(J)}(s)^{-1} - sI)L^\dagger, \quad s \in \Omega, \quad (66)$$

where L^\dagger is the Moore–Penrose inverse. Define analogously

$$\widehat{K}_S(s) := (\widehat{G}_S(s)^{-1} - sI)L^\dagger = s^{1-\alpha}I \quad \text{on } \text{Ran}(L). \quad (67)$$

Then

$$(sI + \widehat{K}^{(J)}(s)L)^{-1} = \widehat{G}^{(J)}(s), \quad (sI + \widehat{K}_S(s)L)^{-1} = \widehat{G}_S(s), \quad (68)$$

and moreover, the resolvents converge in operator norm:

$$\|(sI + \widehat{K}^{(J)}(s)L)^{-1} - (sI + s^{\alpha-1}L)^{-1}\| = \|\widehat{G}^{(J)}(s) - \widehat{G}_S(s)\| \xrightarrow{J \rightarrow \infty} 0, \quad s \in \Omega. \quad (69)$$

Proof. Fix $s \in \Omega$.

Step 1: Stability of invertibility on $\text{Ran}(L)$. Let $A := \widehat{G}_S(s)$ and $B := \widehat{G}^{(J)}(s)$, viewed as operators on $\text{Ran}(L)$. By assumption, A is invertible there and $\|A^{-1}\| \leq M$. Moreover,

$$\|A^{-1}(B - A)\| \leq \|A^{-1}\| \|B - A\| \leq M\delta < 1.$$

Hence, B is invertible on $\text{Ran}(L)$ and the standard resolvent identity yields

$$B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1}.$$

Using $\|B^{-1}\| \leq \|A^{-1}\|/(1 - \|A^{-1}(B - A)\|) \leq M/(1 - \delta M)$, we obtain

$$\|B^{-1} - A^{-1}\| \leq \frac{M}{1 - \delta M} \|A - B\| M = \frac{M^2}{1 - \delta M} \|B - A\|,$$

which proves (65).

Step 2: Definition of the memory symbols and resolvent identities. By definition (66),

$$\widehat{G}^{(J)}(s)^{-1} = sI + \widehat{K}^{(J)}(s)L \quad \text{on } \text{Ran}(L),$$

because $L^\dagger L$ is the identity on $\text{Ran}(L)$. Taking inverses on $\text{Ran}(L)$ yields

$$(sI + \widehat{K}^{(J)}(s)L)^{-1} = \widehat{G}^{(J)}(s) \quad \text{on } \text{Ran}(L),$$

and the same argument yields $(sI + \widehat{K}_S(s)L)^{-1} = \widehat{G}_S(s)$. This proves (68).

Step 3: identifying $\widehat{K}_S(s)$ and resolvent convergence. From (60) we have

$$\widehat{G}_S(s)^{-1} = s^{1-\alpha}(s^\alpha I + L) = sI + s^{1-\alpha}L,$$

hence, $\widehat{K}_S(s) = s^{\alpha-1}I$ as in (67) on $\text{Ran}(L)$. Therefore,

$$(sI + \widehat{K}_S(s)L)^{-1} = (sI + s^{1-\alpha}L)^{-1} = \widehat{G}_S(s).$$

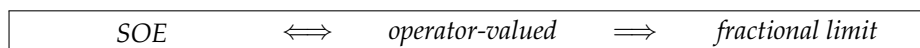
Finally, (69) follows immediately from (68) and the assumed closeness (64), and, in particular, tends to 0 if $\|\widehat{G}^{(J)}(s) - \widehat{G}_S(s)\| \rightarrow 0$ as $J \rightarrow \infty$. \square

Proposition 8 shows that, under suitable assumptions, the SOE-based resolvents approximate the fractional resolvent in operator norm on the domain Ω . This provides a

precise sense in which fractional dynamics can be approximated by finite superpositions of classical diffusions at the level of Laplace-domain operators.

An analysis of the subdiffusion equation with memory with relation to rational kernels in the Laplace domain can be found in [32].

Remark 7. Proposition 8 states that an SOE approximation does not merely approximate a solution curve, but induces an approximate evolution law in the same analytic class as the limiting fractional dynamics, namely through the Laplace-domain resolvent. It makes explicit the structural pathway



where the double arrow refers to the exact algebraic construction $\widehat{G}^{(J)} \leftrightarrow \widehat{K}^{(J)}$ via (66) and (68), while the single arrow denotes the limiting process $J \rightarrow \infty$ in the resolvent sense (69). This is the natural notion of convergence for nonlocal-in-time equations, since the resolvent family uniquely characterizes the corresponding linear evolution.

Remark 8. Proposition 8 should be interpreted strictly as a Laplace-domain (resolvent-level) approximation result. The following clarifications delimit the scope of this statement and prevent overinterpretation in terms of time-domain dynamics:

Time-domain kernel convergence. It does not claim that the inverse Laplace transforms satisfy $K^{(J)}(t) \rightarrow K_S(t)$ pointwise, in $L^1_{loc}(\mathbb{R}_+)$, or in any distributional sense.

Strong convergence of semigroup families in physical time. If a physical-time representation involves nonlinear reparametrizations (e.g., $\tau = t^\alpha$), the proposition does not claim that uniform approximation of $G^{(J)}(\tau)$ implies convergence of Laplace transforms in the physical time variable.

Equivalence to a Caputo equation for each finite J. For finite J, the induced kernel is regular (typically a finite exponential mixture in time), and the proposition does not claim that the corresponding memory equation is a Caputo fractional differential equation.

Global-in-s convergence. No statement is made about convergence uniformly for all $\Re(s) > 0$, nor about behavior near $s = 0$ or $|s| \rightarrow \infty$, unless such regions are included in Ω and the hypotheses are verified there.

In summary, the result establishes a controlled approximation at the level of Laplace-domain operators, which supports the interpretation of fractional diffusion as an effective, scale-free limit of multi-rate processes, but does not by itself constitute a full convergence result in the time domain.

9. Conclusions

In this work, we have investigated subdiffusion on graphs from structural, dynamical, and mechanistic perspectives, emphasizing the role of memory rather than modifications of spatial connectivity or transition rules. By grounding time-fractional diffusion in a random time-change framework, we have shown that subdiffusion on networks is not merely a slower version of classical diffusion, but a fundamentally non-Markovian process in which past states actively shape future evolution. Importantly, this loss of Markovianity occurs without sacrificing linearity or mass conservation, making time-fractional models both analytically tractable and physically meaningful for networked systems.

A central outcome of our analysis is the demonstration that Mittag-Leffler graph dynamics admit a convex, mass-preserving representation as a superposition of classical heat semigroups evaluated at rescaled times. This operator-level viewpoint provides the foundation for the sum-of-exponentials (SOE) approximation, which enables the efficient numerical evaluation of fractional dynamics through a finite number of matrix exponentials. From a computational perspective, this reduces the problem to repeatedly applying

classical diffusion operators, allowing one to leverage existing scalable algorithms while incorporating non-Markovian effects.

From a global perspective, the results of the paper can be understood as different manifestations of this operator-level structure. At the vertex level, the superposition of time scales induces heterogeneous memory effects: different nodes, and even the same node at different times, exhibit distinct biases toward the remote past, recent past, or present contributions, depending on the local temporal curvature of the evolving state. At the level of paths, this same structure translates into a geometric reformulation of transport: the subdiffusive distance induces a metrization of the graph that reshapes connectivity in a time-dependent manner. In particular, subdiffusive shortest paths recover topological shortest paths in the small-time limit while preferentially selecting routes through high-degree regions, revealing a form of memory-assisted navigation that contrasts with classical diffusion.

Taken together, these results show that the operator structure, vertex-level memory, and path-level geometry are not independent phenomena, but are tightly coupled aspects of a single underlying mechanism: the redistribution of diffusion across multiple intrinsic time scales. This integrated viewpoint clarifies how memory reshapes network transport in a structurally coherent way.

From a practical standpoint, the proposed framework is particularly relevant for systems in which transport is affected by memory or trapping effects, such as biological interaction networks, communication systems with delays, or transport in heterogeneous media. In such settings, the SOE approximation provides a computationally viable alternative to direct evaluation of fractional operators, enabling the simulation of subdiffusive dynamics on large graphs using standard diffusion solvers. Compared to classical diffusion, the present approach captures non-exponential relaxation, heterogeneous waiting times, and path-selection effects, while retaining computational efficiency through its reduction to a finite set of semigroup evaluations.

Finally, by connecting fractional diffusion to finite superpositions of classical diffusions and to operator-valued Volterra memory equations in the Laplace domain, we have placed time-fractional graph dynamics within a broader hierarchy of diffusion models with memory. In this context, fractional equations can be interpreted as scale-free limits of multi-rate processes, providing a unifying framework that links analytical theory, numerical approximation, and physical interpretation.

Overall, our results provide a coherent picture of subdiffusion on graphs, showing how memory, geometry, and dynamics interact across scales. This perspective not only deepens the theoretical understanding of fractional diffusion on networks, but also opens the way to practical applications and further quantitative comparisons with classical diffusion models.

Author Contributions: Conceptualization, E.E.; methodology, E.E.; software, E.E. and N.D.; validation, E.E. and N.D.; writing—original draft preparation, E.E.; writing—review and editing, E.E. and N.D.; visualization E.E. and N.D.; supervision, E.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Agencia Estatal de Investigación (AEI, MCI, Spain) MCIN/AEI/10.13039/501100011033 under grant PID2023-149473NB-I00 and by Agencia Estatal de Investigación (AEI, MCI, Spain) MCIN/AEI/10.13039/501100011033 and Fondo Europeo de Desarrollo Regional (FEDER, UE) under the María de Maeztu Program for units of Excellence in R&D, grant CEX2021-001164-M) are also acknowledged.

Data Availability Statement: The original contributions presented in this study are included in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Numerical Window Tables and Usage

We summarize practical window choices for the log–trapezoidal SOE, based on the tail bounds previously found. For the *right tail*, the stretched–exponential decay of M_α implies that

$$\theta_{\max} \geq \left(\frac{\log(2/\varepsilon)}{c_\alpha}\right)^{1/q_\alpha}, \quad c_\alpha = (1 - \alpha) \alpha^{\alpha/(1-\alpha)}, \quad q_\alpha = \frac{1}{1 - \alpha},$$

is sufficient to make $\sup_{\lambda \in [0, \lambda_{\max}]} \int_{\theta_{\max}}^\infty M_\alpha(\theta) e^{-\theta t \lambda} d\theta \lesssim \varepsilon/2$ uniformly in λ (i.e., including the zero mode); see (19). For the *left tail*, taking $\theta_{\min} = \frac{\varepsilon}{2} \Gamma(1 - \alpha)$ ensures

$$\sup_{\lambda \in [0, \lambda_{\max}]} \int_0^{\theta_{\min}} M_\alpha(\theta) e^{-\theta t \lambda} d\theta \leq \varepsilon/2 \text{ (Proposition 3).}$$

Tables A1 and A2 list these cutoffs for typical α and ε values, together with $y_{\max} = \log \theta_{\max}$ and $y_{\min} = \log \theta_{\min}$, which are the actual endpoints used by the SOE in the variable $y = \log \theta$.

Table A1. Conservative right–tail cutoff θ_{\max} ensuring $\sup_{\lambda \in [0, \lambda_{\max}]} \int_{\theta_{\max}}^\infty M_\alpha(\theta) e^{-\theta t \lambda} d\theta \lesssim \varepsilon/2$. Brackets show $y_{\max} = \log \theta_{\max}$ for direct use in the log–trapezoidal SOE. Values are independent of t and of the spectrum; they come from the uniform bound (19) using the stretched–exponential tail of M_α with constants c_α, q_α [54,55].

α	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-10}$	$\varepsilon = 10^{-12}$
0.20	14.016 [2.640]	17.474 [2.861]	20.768 [3.033]	23.936 [3.175]
0.30	11.979 [2.483]	14.529 [2.676]	16.899 [2.827]	19.134 [2.951]
0.50	7.618 [2.031]	8.744 [2.168]	9.740 [2.276]	10.644 [2.365]
0.70	4.109 [1.413]	4.464 [1.496]	4.762 [1.561]	5.023 [1.614]
0.80	2.816 [1.035]	2.976 [1.090]	3.107 [1.134]	3.219 [1.169]

Table A2. Left–tail cutoff $\theta_{\min} = \frac{\varepsilon}{2} \Gamma(1 - \alpha)$ guaranteeing $\sup_{\lambda \in [0, \lambda_{\max}]} \int_0^{\theta_{\min}} M_\alpha(\theta) e^{-\theta t \lambda} d\theta \leq \varepsilon/2$. Brackets show $y_{\min} = \log \theta_{\min}$ (Rounded to three significant digits.).

α	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-10}$	$\varepsilon = 10^{-12}$
0.20	5.82×10^{-7} [−14.357]	5.82×10^{-9} [−18.962]	5.82×10^{-11} [−23.567]	5.82×10^{-13} [−28.172]
0.30	6.49×10^{-7} [−14.248]	6.49×10^{-9} [−18.853]	6.49×10^{-11} [−23.458]	6.49×10^{-13} [−28.063]
0.50	8.86×10^{-7} [−13.936]	8.86×10^{-9} [−18.541]	8.86×10^{-11} [−23.147]	8.86×10^{-13} [−27.752]
0.70	1.50×10^{-6} [−13.413]	1.50×10^{-8} [−18.018]	1.50×10^{-10} [−22.623]	1.50×10^{-12} [−27.228]
0.80	2.30×10^{-6} [−12.985]	2.30×10^{-8} [−17.590]	2.30×10^{-10} [−22.195]	2.30×10^{-12} [−26.800]

Let us now explain how to use these Tables.

1. Fix a target tolerance ε for the *tails* (e.g., $\varepsilon = 10^{-12}$ in double precision).
2. Read off $(\theta_{\min}, y_{\min})$ from Table A2 and $(\theta_{\max}, y_{\max})$ from Table A1 for your α and ε .
3. Set the SOE window in the log variable as $[y_{\min}, y_{\max}]$; then increase J until the *in-window discretization* error decays to your overall tolerance (Theorem 1 shows geometric decay with J).

Variants and remarks.

- (*Mean-zero subspace.*) If you only act on 1^\perp and know the spectral gap $\lambda_2 > 0$, you may choose $\theta_{\max} = \frac{1}{t\lambda_2} \log(2/\varepsilon)$ (for reuse over $t \in [t_{\min}, t_{\max}]$, use t_{\min} here to keep the nodes t -independent) (Proposition 5), which can be *much smaller* than the uniform value in Table A1.

- (Heuristic rule of thumb.) For high-frequency modes it is common to take $\theta_{\max} \approx \frac{32}{t\lambda}$ so that $e^{-\theta_{\max}t\lambda} \approx 10^{-14}$ (machine precision). The uniform bound above is conservative and includes the zero mode.
- (Scaling with α .) As $\alpha \uparrow 1$ the tail constant c_α grows and the required θ_{\max} shrinks; as $\alpha \downarrow 0$ the reverse happens, hence the larger values in the top row of Table A1.

References

1. Estrada, E. *The Structure of Complex Networks: Theory and Applications*; Oxford University Press: Oxford, UK, 2012.
2. López-Pintado, D. An overview of diffusion in complex networks. In *Complex Networks and Dynamics: Social and Economic Interactions*; Springer: Cham, Switzerland, 2016; pp. 27–48.
3. Masuda, N.; Porter, M.A.; Lambiotte, R. Random walks and diffusion on networks. *Phys. Rep.* **2017**, *716*, 1–58. [[CrossRef](#)]
4. Nicolaides, C.; Cueto-Felgueroso, L.; Juanes, R. Anomalous physical transport in complex networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2010**, *82*, 055101. [[CrossRef](#)] [[PubMed](#)]
5. Metzler, R.; Klafter, J. The random walk's guide to anomalous diffusion: A fractional dynamics approach. *Phys. Rep.* **2000**, *339*, 1–77. [[CrossRef](#)]
6. Sokolov, I.M. Models of anomalous diffusion in crowded environments. *Soft Matter* **2012**, *8*, 9043–9052. [[CrossRef](#)]
7. Medina, P.; Carrasco, S.; Correa-Burrows, P.; Rogan, J.; Valdivia, J.A. Nontrivial and anomalous transport on weighted complex networks. *Commun. Nonlinear Sci. Numer. Simul.* **2022**, *114*, 106684. [[CrossRef](#)]
8. Weiss, M.; Elsner, M.; Kartberg, F.; Nilsson, T. Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophys. J.* **2004**, *87*, 3518–3524. [[CrossRef](#)]
9. Gupta, S.; Biehl, R.; Sill, C.; Allgaier, J.; Sharp, M.; Ohl, M.; Richter, D. Protein entrapment in polymeric mesh: Diffusion in crowded environment with fast process on short scales. *Macromolecules* **2016**, *49*, 1941–1949. [[CrossRef](#)]
10. Grimaldo, M.; Roosen-Runge, F.; Zhang, F.; Schreiber, F.; Seydel, T. Dynamics of proteins in solution. *Q. Rev. Biophys.* **2019**, *52*, e7. [[CrossRef](#)]
11. Basak, S.; Sengupta, S.; Chattopadhyay, K. Understanding biochemical processes in the presence of sub-diffusive behavior of biomolecules in solution and living cells. *Biophys. Rev.* **2019**, *11*, 851–872. [[CrossRef](#)]
12. Combinido, J.S.L.; Lim, M.T. Crowding effects in vehicular traffic. *PLoS ONE* **2012**, *7*, e48151. [[CrossRef](#)] [[PubMed](#)]
13. Foroozani, A.; Ebrahimi, M. Anomalous information diffusion in social networks: Twitter and Digg. *Expert Syst. Appl.* **2019**, *134*, 249–266. [[CrossRef](#)]
14. Evangelista, L.R.; Lenzi, E.K. *Fractional Diffusion Equations and Anomalous Diffusion*; Cambridge University Press: Cambridge, UK, 2018.
15. Sokolov, I.M.; Klafter, J.; Blumen, A. Fractional kinetics. *Phys. Today* **2002**, *55*, 48–54. [[CrossRef](#)]
16. Caputo, M. Linear models of dissipation whose Q is almost frequency independent. *Ann. Geophys.* **1966**, *19*, 383–393. [[CrossRef](#)]
17. Abadias, L.; Estrada-Rodriguez, G.; Estrada, E. Fractional-order susceptible-infected model: Definition and applications to the study of COVID-19 main protease. *Fract. Calc. Appl. Anal.* **2020**, *23*, 635–655. [[CrossRef](#)]
18. D'Alessandro, M.; Van Mieghem, P. Fractional derivative in continuous-time Markov processes and applications to epidemics in networks. *Phys. Rev. Res.* **2025**, *7*, 013017. [[CrossRef](#)]
19. Estrada, E. Fractional diffusion on the human proteome as an alternative to the multi-organ damage of SARS-CoV-2. *Chaos Interdiscip. J. Nonlinear Sci.* **2020**, *30*, 081104. [[CrossRef](#)]
20. Sun, W.; Li, Y.; Li, C.; Chen, Y. Convergence speed of a fractional order consensus algorithm over undirected scale-free networks. *Asian J. Control* **2011**, *13*, 936–946. [[CrossRef](#)]
21. Yan, X.; Li, K.; Yang, C.; Zhuang, J.; Cao, J. Consensus of fractional-order multi-agent systems via observer-based boundary control. *IEEE Trans. Netw. Sci. Eng.* **2024**, *11*, 3370–3382. [[CrossRef](#)]
22. Sun, F.; Han, Y.; Zhu, W.; Kurths, J. Group consensus for fractional-order heterogeneous multi-agent systems under cooperation-competition networks with time delays. *Commun. Nonlinear Sci. Numer. Simul.* **2024**, *133*, 107951. [[CrossRef](#)]
23. Huang, C.; Wang, F. Distributed Consensus Tracking of Incommensurate Heterogeneous Fractional-Order Multi-Agent Systems Based on Vector Lyapunov Function Method. *Fractal Fract.* **2024**, *8*, 575. [[CrossRef](#)]
24. Lu, J.; Shen, J.; Cao, J.; Kurths, J. Consensus of networked multi-agent systems with delays and fractional-order dynamics. In *Consensus and Synchronization in Complex Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 69–110.
25. Cao, Y.; Li, Y.; Ren, W.; Chen, Y. Distributed coordination of networked fractional-order systems. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2009**, *40*, 362–370.
26. Pang, G.; Lu, L.; Karniadakis, G.E. fPINNs: Fractional physics-informed neural networks. *SIAM J. Sci. Comput.* **2019**, *41*, A2603–A2626. [[CrossRef](#)]

27. Joshi, M.; Bhosale, S.; Vyawahare, V.A. A survey of fractional calculus applications in artificial neural networks. *Artif. Intell. Rev.* **2023**, *56*, 13897–13950. [[CrossRef](#)]
28. Kim, H.; Lawley, S.D. Cover times of many diffusive or subdiffusive searchers. *SIAM J. Appl. Math.* **2024**, *84*, 602–620. [[CrossRef](#)]
29. Jung, W.; Chen, T.Y.; Santiago, A.G.; Chen, P. Memory effects of transcription regulator-DNA interactions in bacteria. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2407647121. [[CrossRef](#)] [[PubMed](#)]
30. Sandev, T.; Tomovski, Z.; Dubbeldam, J.L.; Chechkin, A. Generalized diffusion-wave equation with memory kernel. *J. Phys. A Math. Theor.* **2018**, *52*, 015201. [[CrossRef](#)]
31. Gurtin, M.E.; Pipkin, A.C. A general theory of heat conduction with finite wave speeds. *Arch. Ration. Mech. Anal.* **1968**, *31*, 113–126. [[CrossRef](#)]
32. Ponce, R. Discrete Subdiffusion Equations with Memory. *Appl. Math. Optim.* **2021**, *84*, 3475–3497. [[CrossRef](#)]
33. Saif, S.; Malik, S. An inverse problem for a two-dimensional diffusion equation with arbitrary memory kernel. *Math. Methods Appl. Sci.* **2023**, *46*, 11007–11020. [[CrossRef](#)]
34. Trimper, S.; Zabrocki, K.; Schulz, M. Memory-controlled diffusion. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2004**, *70*, 056133. [[CrossRef](#)]
35. Toan, N.D.; Thuy, L.T. The nonclassical diffusion equations with time-dependent memory kernels and a new class of nonlinearities. *Glasg. Math. J.* **2022**, *64*, 716–733.
36. Hoffmann, T.; Porter, M.A.; Lambiotte, R. Random walks on stochastic temporal networks. In *Temporal Networks*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 295–313.
37. Lambiotte, R.; Salnikov, V.; Rosvall, M. Effect of memory on the dynamics of random walks on networks. *J. Complex Netw.* **2015**, *3*, 177–188. [[CrossRef](#)]
38. Scholtes, I.; Wider, N.; Pfitzner, R.; Garas, A.; Tessone, C.J.; Schweitzer, F. Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nat. Commun.* **2014**, *5*, 5024.
39. Kosztołowicz, T.; Dworecki, K.; Mrówczyński, S. Measuring subdiffusion parameters. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2005**, *71*, 041105.
40. Kepten, E.; Weron, A.; Sikora, G.; Burnecki, K.; Garini, Y. Guidelines for the fitting of anomalous diffusion mean square displacement graphs from single particle tracking experiments. *PLoS ONE* **2015**, *10*, e0117722.
41. Gallos, L.K.; Song, C.; Havlin, S.; Makse, H.A. Scaling theory of transport in complex biological networks. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7746–7751. [[CrossRef](#)]
42. Applebaum, D. *Lévy Processes and Stochastic Calculus*, 2nd ed.; Cambridge Studies in Advanced Mathematics; Cambridge University Press: Cambridge, UK, 2009; Volume 116. [[CrossRef](#)]
43. Bouchaud, J.P.; Georges, A. Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Phys. Rep.* **1990**, *195*, 127–293. [[CrossRef](#)]
44. Polanowski, P.; Sikorski, A. Simulation of diffusion in a crowded environment. *Soft Matter* **2014**, *10*, 3597–3607. [[CrossRef](#)]
45. Meinecke, L. Multiscale modeling of diffusion in a crowded environment. *Bull. Math. Biol.* **2017**, *79*, 2672–2695. [[CrossRef](#)]
46. Fanelli, D.; McKane, A.J. Diffusion in a crowded environment. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2010**, *82*, 021113. [[CrossRef](#)]
47. Goychuk, I. Viscoelastic subdiffusion in a random Gaussian environment. *Phys. Chem. Chem. Phys.* **2018**, *20*, 24140–24155. [[CrossRef](#)] [[PubMed](#)]
48. Chauhan, T.; Kalyanaraman, K.; Sircar, S. Quantifying macrostructures in viscoelastic sub-diffusive flows. *J. Math. Phys.* **2024**, *65*, 073101. [[CrossRef](#)]
49. Goychuk, I.; Pöschel, T. Fingerprints of viscoelastic subdiffusion in random environments: Revisiting some experimental data and their interpretations. *Phys. Rev. E* **2021**, *104*, 034125. [[CrossRef](#)]
50. Saxton, M.J. Anomalous subdiffusion in fluorescence photobleaching recovery: A Monte Carlo study. *Biophys. J.* **2001**, *81*, 2226–2240. [[CrossRef](#)]
51. Lim, S.C.; Muniandy, S.V. Self-similar Gaussian processes for modeling anomalous diffusion. *Phys. Rev. E* **2002**, *66*, 021114. [[CrossRef](#)]
52. Mainardi, F.; Mura, A.; Pagnini, G. The M-Wright Function in Time-Fractional Diffusion Processes: A Tutorial Survey. *Int. J. Differ. Equ.* **2010**, *2010*, 104505. [[CrossRef](#)]
53. Schilling, R.L.; Song, R.; Vondraček, Z. *Bernstein Functions: Theory and Applications*, 2nd ed.; De Gruyter Studies in Mathematics; De Gruyter: Berlin, Germany, 2012; Volume 37.
54. Gorenflo, R.; Kilbas, A.A.; Mainardi, F.; Rogosin, S.V. *Mittag-Leffler Functions, Related Topics and Applications*; Springer Monographs in Mathematics; Springer: Berlin, Germany, 2014.
55. Mainardi, F. *Fractional Calculus and Waves in Linear Viscoelasticity: An Introduction to Mathematical Models*; Imperial College Press: London, UK, 2010.

56. Meerschaert, M.M.; Sikorskii, A. *Stochastic Models for Fractional Calculus*; De Gruyter Studies in Mathematics; De Gruyter: Berlin, Germany, 2011; Volume 43.
57. Trefethen, L.N.; Weideman, J.A.C. The Exponentially Convergent Trapezoidal Rule. *SIAM Rev.* **2014**, *56*, 385–458. [[CrossRef](#)]
58. Coifman, R.R.; Lafon, S.; Lee, A.B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S.W. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 7426–7431. [[CrossRef](#)]
59. Coifman, R.R.; Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **2006**, *21*, 5–30. [[CrossRef](#)]
60. Estrada, E. The communicability distance in graphs. *Linear Algebra Its Appl.* **2012**, *436*, 4317–4328. [[CrossRef](#)]
61. Estrada, E.; Sanchez-Lirola, M.; De La Peña, J.A. Hyperspherical embedding of graphs and networks in communicability spaces. *Discret. Appl. Math.* **2014**, *176*, 53–77. [[CrossRef](#)]
62. Markvorsen, S. Minimal webs in Riemannian manifolds. *Geom. Dedicata* **2008**, *133*, 7–34. [[CrossRef](#)]
63. Bridson, M.R.; Haefliger, A. *Metric Spaces of Non-Positive Curvature*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 319.
64. Silver, G.; Akbarzadeh, M.; Estrada, E. Tuned communicability metrics in networks. The case of alternative routes for urban traffic. *Chaos Solitons Fractals* **2018**, *116*, 402–413. [[CrossRef](#)]
65. Baeumer, B.; Meerschaert, M.M. Stochastic solutions for fractional Cauchy problems. *Fract. Calc. Appl. Anal.* **2001**, *4*, 481–500.
66. Meerschaert, M.M.; Nane, E.; Vellaisamy, P. The fractional Poisson process and the inverse stable subordinator. *Electron. J. Probab.* **2011**, *16*, 1600–1620. [[CrossRef](#)]
67. Odibat, Z. Approximations of fractional integrals and Caputo fractional derivatives. *Appl. Math. Comput.* **2006**, *178*, 527–533. [[CrossRef](#)]
68. Kivelä, M.; Arenas, A.; Barthelemy, M.; Gleeson, J.P.; Moreno, Y.; Porter, M.A. Multilayer networks. *J. Complex Netw.* **2014**, *2*, 203–271. [[CrossRef](#)]
69. Boccaletti, S.; Bianconi, G.; Criado, R.; Del Genio, C.I.; Gómez-Gardenes, J.; Romance, M.; Sendina-Nadal, I.; Wang, Z.; Zanin, M. The structure and dynamics of multilayer networks. *Phys. Rep.* **2014**, *544*, 1–122. [[CrossRef](#)] [[PubMed](#)]
70. Gomez, S.; Diaz-Guilera, A.; Gomez-Gardenes, J.; Perez-Vicente, C.J.; Moreno, Y.; Arenas, A. Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.* **2013**, *110*, 028701. [[CrossRef](#)]
71. Li, L.; Zhou, X.; Gao, H. The stability and exponential stabilization of the heat equation with memory. *J. Math. Anal. Appl.* **2018**, *466*, 199–214. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.