Fractional diffusion on the human proteome as an alternative to the multi-organ damage of SARS-CoV-2 Ⓕ

Ernesto Estrada[a] ⓘD

AFFILIATIONS

Instituto Universitario de Matemáticas y Aplicaciones, Universidad de Zaragoza, 50009 Zaragoza, Spain and ARAID Foundation, Government of Aragón, 50018 Zaragoza, Spain

[a]**Author to whom correspondence should be addressed:** estrada66@unizar.es

ABSTRACT

The coronavirus 2019 (COVID-19) respiratory disease is caused by the novel coronavirus SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), which uses the enzyme ACE2 to enter human cells. This disease is characterized by important damage at a multi-organ level, partially due to the abundant expression of ACE2 in practically all human tissues. However, not every organ in which ACE2 is abundant is affected by SARS-CoV-2, which suggests the existence of other multi-organ routes for transmitting the perturbations produced by the virus. We consider here diffusive processes through the protein–protein interaction (PPI) network of proteins targeted by SARS-CoV-2 as an alternative route. We found a subdiffusive regime that allows the propagation of virus perturbations through the PPI network at a significant rate. By following the main subdiffusive routes across the PPI network, we identify proteins mainly expressed in the heart, cerebral cortex, thymus, testis, lymph node, kidney, among others of the organs reported to be affected by COVID-19.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0015626

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the new coronavirus causing the pandemic known as coronavirus disease 2019 (COVID-19). This respiratory disease is characterized by multi-organ and systemic damages in patients. The abundance of ACE2 on human organs has been claimed as responsible for such multi-organ spread of the virus damages. However, once on circulation, the virus could spread to practically every organ in the human body as ACE2 is ubiquitous on endothelia and smooth muscle cells of virtually all organs. Contrastingly, SARS-CoV-2 only damages selectively a few organs. Here, we develop the hypothesis that the effects of the SARS-CoV-2 virus can be spread through the human protein–protein interaction (PPI) network in a subdiffusive way. We then elaborate a time-fractional diffusion model on networks, which allow us to study this phenomenon. Starting the diffusion from the SARS-CoV-2 spike protein to the human PPI network, we show here that the perturbations can spread across the whole network in a very few steps. Consequently, we discover a few potential routes of propagation of these perturbations from proteins mainly expressed in the lungs to proteins mainly expressed in other different tissues, such as the heart, cerebral cortex, thymus, lymph node, testis, prostate, liver, small intestine, duodenum, kidney, among others already reported as damaged by COVID-19.

I. INTRODUCTION

Since December 2019, a new coronavirus designated SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)[1] has produced an outbreak of pulmonary disease, which has soon become a global pandemic.[2,3] The new coronavirus 2019 (COVID-19) disease is characterized by a wide range of clinical manifestations,[4,5] with an important implication of multi-organ and systemic dysfunctions,[6,7] which include heart failure, renal failure, liver damage, shock, and multi-organ failure. The new coronavirus shares about 82% of its genome with the one that produced the 2003 outbreak (SARS-CoV).[8] Both coronaviruses share the same cellular receptor, which is the angiotensin-converting enzyme 2 (ACE2) receptor.[8–10] ACE2 receptors are enriched in several tissues across the organs, such as in alveolar epithelial type II cells of lung tissues, in the heart, endothelium, kidneys, and intestines. Therefore, ACE2 has been

hypothesized as a potential cause of the major complications of the COVID-19.[11,12] However, it has been found that ACE2 has abundant expression on endothelia and smooth muscle cells of virtually all organs.[13] Therefore, it should be expected that after SARS-CoV-2 is present in circulation, it can be spread across all organs. In contrast, both SARS-CoV and SARS-CoV-2 are found specifically in some organs but not in others, as shown by *in situ* hybridization studies for SARS-CoV. This was already remarked by Hamming *et al.*[13] by stressing that it "is remarkable that so few organs become viruspositive, despite the presence of ACE2 on the endothelia of all organs and SARS-CoV in blood plasma of infected individuals." Recently, Gordon *et al.*[14] identified human proteins that interact physically with those of the SARS-CoV-2 forming a high confidence SARS-CoV-2-human protein–protein interaction (PPI) system. Using this information, Gysi *et al.*[15] discovered that 208 of the human proteins targeted by SARS-CoV-2 forms a connected component inside the human PPI network. That is, these 208 are not randomly distributed across the human proteome, but they are closely interconnected by short routes that allow moving from one to another in just a few steps. These interdependencies of protein–protein interactions are known to enable that perturbations on one interaction propagate across the network and affect other interactions.[16–19] In fact, it has been signified that diseases are a consequence of such perturbation propagation.[20–22] It has been stressed that the protein–protein interaction process requires diffusion in their initial stages.[23] The diffusive processes occur when proteins, possibly guided by electrostatic interactions, need to encounter each other many times before forming an intermediate.[24] Not surprisingly, diffusive processes have guided several biologically oriented searches in PPI networks.[25,26] Therefore, we assume here that perturbations produced by SARS-CoV-2 proteins on the human PPI network are propagated by means of diffusive processes. However, due to the crowded nature of the intra-cell space and the presence in it of spatial barriers, subdiffusive processes more than normal diffusion are expected for these protein–protein encounters.[27–29] This creates another difficulty, as remarked by Batada *et al.*,[23] which is that such (sub)diffusive processes along are not sufficient for carrying out cellular processes at a significant rate in cells.

Here, we propose the use of a time-fractional diffusion model on the PPI network of proteins targeted by SARS-CoV-2. The goal is to model the propagation of the perturbations produced by the interactions of human proteins with those of SARS-CoV-2 through the whole PPI. The subdiffusive process emerging from the application of this model to the SARS-CoV-2-human PPIs has a very small rate of convergence to the steady state. However, this process produces a dramatic increment of the probability that certain proteins are perturbed at very short times. This kind of shock wave effect of the transmission of perturbations occurs at much earlier times in the subdiffusive regime than at the normal diffusion one. Therefore, we propose here a switch and restart process in which a subdiffusive process starts at a given protein of the PPI, perturbs a few others, which then become the starting point of a new subdiffusive process and so on. Using this approach, we then analyze how the initial interaction of the SARS-CoV-2 spike protein with a human protein propagates across the whole network. We discover some potential routes of propagation of these perturbations from proteins mainly expressed in the lungs to proteins mainly expressed

in other different tissues, such as the heart, cerebral cortex, thymus, lymph node, testis, prostate, liver, small intestine, duodenum, kidney, among others.

II. MATERIALS AND METHODS

A. Settling a model

The problem we intend to model here is of a large complexity as it deals with the propagation of perturbations across a network of interacting proteins, each of which is located in a crowded intracellular space. Therefore, we necessarily have to impose restrictions and make assumptions to settle our modeling framework. As we have mentioned in Sec. I, protein encounters should necessarily occur in subdiffusive ways due to the crowded environment in which they are embedded, as well as the existence of immobile obstacles such as membranes. By a subdiffusive process, we understand that the mean square displacement of a protein scales as

$$\langle x^2(t) \rangle \sim t^\kappa, \tag{1}$$

where $0 < \kappa < 1$ is the anomalous diffusion exponent. As observed by Sposini *et al.*,[30] these anomalous diffusive processes can emerge from (i) continuous time random walk (CTRW) processes or by (ii) viscoelastic diffusion processes. In the first case, the "anomaly" is created by power-law waiting times in between motion events. This kind of processes is mainly accounted for by the generalized Langevin equation with the power-law friction kernel as well as by fractional Brownian motion (FBM). While the first processes are characterized by the stretched Gaussian displacement probability density, weak ergodicity, and aging, the second ones are ergodic processes characterized by the Gaussian density probability distribution. Therefore, our first task is to discern which of these two kinds of approaches is appropriate for the current scenario.

We start by mentioning that Weiss *et al.*[31] have analyzed data from fluorescence correlation spectroscopy (FCS) for studying subdiffusive biological processes. They have, for instance, reported that membrane proteins move subdiffusively in the endoplasmic reticulum and Golgi apparatus *in vivo*. Subdiffusion of cytoplasmatic macromolecules was also reported by Weiss *et al.*[32] using FCS. Then, Guigas and Weiss[27] simulated the way in which the subdiffusive motion of these particles should occur in a crowded intracellular fluid. They did so by assigning diffusive steps from a Weirstrass–Mandelbrot function yielding a FBM. They stated that CTRW was excluded due to its Markovian nature. In another work, Szymanski and Weiss[33] used FCS and simulations to analyze the subdiffusive motion of a protein in a simulated crowded medium. First, they reported that crowded-induced subdiffusion is consistent with the predictions from FBM or obstructed (percolation-like) diffusion. Second, they reported that CTRW does not explain the experimental results obtained by FCS and should not be appropriated for such processes.

The time resolution of FCS is in the microsecond range, i.e., 10^{-6} s.[34] However, an important question on biological subdiffusion may require higher time resolution to be solved. This is the question of how diffusive processes on short times, while the macromolecule has not felt yet the crowding of the environment, is related to the long-time diffusion. This particular problem was explored experimentally by Gupta *et al.*[35] by using state-of-the-art neutron

spin-echo (NSE) and small-angle neutron scattering (SANS), which has a resolution in the nanosecond range, i.e., 10^{-9} s. Their experimental setting was defined by the use of two globular proteins in a crowded environment formed by poly(ethylene oxide) (PEO), which mimics a macromolecular environment. In their experiments, NSE was used to tackle the fast diffusion process, which corresponds to a dynamics inside a trap built by the environment mesh. SANS captures the slow dynamics, which corresponds to the long-time diffusion at macroscopic length scales. From our current perspective, the most important result of this work is that the authors found that in a higher concentration of polymeric solutions, like in the intracellular space, the diffusion is fractional in nature. They showed this by using the fractional Fokker–Planck equation with a periodic potential. According to Gupta *et al.*,[35] this fractional nature of the crossover from fast dynamics to slow macroscopic dynamics is due to the heterogeneity of the polymer mesh in the bulk sample, which may well resemble the intra-cellular environment. As proved by Barkai *et al.*,[36] the fractional Fokker–Planck equation can be derived from the CTRW, which clearly indicates that the results obtained by Gupta *et al.* point out to the classification of the subdiffusive dynamics into the class (i). We should remark that independently of these results by Gupta *et al.*,[35] Shorten and Sneyd[37] have successfully used the fractional diffusion equation to mimic the protein diffusion in an obstructed media like within skeletal muscle. We notice in passing that the (fractional) diffusion equation can be obtained from the (fractional) Fokker–Planck equation in the absence of an external force.

In closing, because here we are interested in modeling the diffusion of proteins in several human cells, which are highly crowded, and in which we should recover the same crossover between initial fast and later slow dynamics, we will consider a modeling tool of the class (i). In particular, we will focus our modeling on the use of a time-fractional diffusion equation using Caputo derivatives. Another justification for the use of this model here is that interacting proteins can be in different kinds of cells. Thus, we consider that the perturbation of one protein is not necessarily followed by the perturbation of one of its interactors, but a time may mediate between the two processes. This is exactly the kind of processes that the time-fractional diffusion captures.

B. Time-fractional diffusion model on networks

In this work, we always consider $G = (V, E)$ to be an undirected finite network with vertices V representing proteins and edges E representing the interaction between pairs of proteins. Let us consider $0 < \alpha \leq 1$ and a function $u : [0, \infty) \rightarrow \mathbb{R}$, then we denote by $D_t^\alpha u$ the fractional Caputo derivative of u of the order α, which is given by[38]

$$D_t^\alpha u(t) = \int_0^t g_{1-\alpha}(t - \tau) u'(\tau) \, d\tau := \left(g_{1-\alpha} * u'\right)(t), \quad t > 0,$$

where $*$ denotes the classical convolution product on $(0, \infty)$ and $g_\gamma(t) \frac{t^{\gamma-1}}{\Gamma(\gamma)}$, for $\gamma > 0$, where $\Gamma(\cdot)$ is the Euler gamma function. Observe that the previous fractional derivative has sense whenever the function is derivable and the convolution is defined (for example, if u' is locally integrable). The notation g_γ is very useful in the

fractional calculus theory, mainly by the property $g_\gamma * g_\delta = g_{\gamma+\delta}$ for all $\gamma, \delta > 0$.

Here, we propose to consider the time-fractional diffusion (TFD) equation on the network as

$$D_t^\alpha x(t) = -C\mathcal{L}x(t), \tag{2}$$

with the initial condition $x(0) = x_0$, where $x_i(t)$ is the probability that the protein i is perturbed at the time t; C is the diffusion coefficient of the network, which we will set hereafter to unity; and $\mathcal{L}$ is the graph Laplacian, i.e., $\mathcal{L} = K - A$, where K is a diagonal matrix of node degrees and A is the adjacency matrix. This model was previously studied in distributed coordination algorithms for the consensus of multi-agent systems.[39–41] The use of fractional calculus in the context of physical anomalous diffusion has been reviewed by Metzler and Klafter.[42] A different approach has been developed by Riascos and Mateos.[43,44] It is based on the use of fractional powers of the graph Laplacian (see Ref. 45 and references therein). The approach has been recently formalized by Benzi *et al.*[46] This method cannot be used in the current framework because it generates only superdiffusive behaviors (see Benzi *et al.*[46]) and not subdiffusive regimes. Another disadvantage of this approach is that it can only be used to positive (semi)definite graph operators, such as the Laplacian, but not to adjacency operators such as the one used in tight-binding quantum mechanical or epidemiological approaches (see Sec. VI).

Theorem 1. *The solution of the fractional-time diffusion model on the network is*

$$x(t) = E_{\alpha,1}\left(-(tC)^\alpha \mathcal{L}\right)x_0, \tag{3}$$

where $E_{\alpha,\beta}(\gamma \mathcal{L})$ is the Mittag–Leffler function of the Laplacian matrix of a graph.

Proof. We use the spectral decomposition of the network Laplacian $\mathcal{L} = U\Lambda U^{-1}$, where $U = \left[\vec{\psi}_1 \cdots \vec{\psi}_n\right]$ and $\Lambda = diag(\mu_r)$. Then, we can write

$$D_t^\alpha x(t) = -U\Lambda U^{-1}x(t). \tag{4}$$

Let us define $y(t) = U^{-1}x(t)$ such that $D_t^\alpha x(t) = -U\Lambda y(t)$, and we have

$$U^{-1}D_t^\alpha x(t) = -\Lambda y(t),$$
$$D_t^\alpha y(t) = -\Lambda y(t). \tag{5}$$

As Λ is a diagonal matrix, we can write

$$D_t^\alpha y_i(t) = -\mu_i y_i(t), \tag{6}$$

which has the solution

$$y_i(t) = E_{\alpha,1}\left(-t^\alpha \mu_i\right)y_i(0). \tag{7}$$

We can replace $y_i(t) = U^{-1}x_i(t)$ to have

$$U^{-1}x_i(t) = E_{\alpha,1}\left(-t^\alpha \mu_i\right)U^{-1}x_i(0),$$
$$x_i(t) = UE_{\alpha,1}\left(-t^\alpha \mu_i\right)U^{-1}x_i(0), \tag{8}$$

which finally gives the result in the matrix-vector when written for all the nodes,

$$x(t) = E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right)x_0. \tag{9}$$

□

We can write $\mathcal{L} = U \Lambda U^{-1}$, where $U = \left[\vec{\psi}_1 \cdots \vec{\psi}_n \right]$ and $\Lambda = diag\left(\mu_r\right)$. Then,

$$E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right) = U E_{\alpha,1}\left(-t^\alpha \Lambda\right) U^{-1}, \tag{10}$$

which can be expanded as

$$E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right) = \vec{\psi}_1 \vec{\phi}_1^T E_{\alpha,1}\left(-t^\alpha \mu_1\right) + \vec{\psi}_2 \vec{\phi}_2^T E_{\alpha,1}\left(-t^\alpha \mu_2\right)$$
$$+ \cdots + \vec{\psi}_n \vec{\phi}_n^T E_{\alpha,1}\left(-t^\alpha \mu_n\right), \tag{11}$$

where $\vec{\psi}_j$ and $\vec{\phi}_j$ are the jth column of U and of U^{-1}, respectively. Because $\mu_1 = 0$ and $0 < \mu_2 \leq \cdots \leq \mu_n$ for a connected graph, we have

$$\lim_{t \to \infty} E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right) = \vec{\psi}_1 \vec{\phi}_1^T, \tag{12}$$

where $\vec{\psi}_1^T \vec{\phi}_1 = 1$. Let us take $\vec{\psi}_1 = \vec{1}$, such that we have

$$\lim_{t \to \infty} \vec{x}(t) = \lim_{t \to \infty} \left(E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right)\right) \vec{x}_0 = \left(\vec{1}\vec{\phi}_1^T\right) \vec{x}_0 = \left(\vec{\phi}_1^T \vec{x}_0\right) \vec{1}. \tag{13}$$

This result indicates that in an undirected and connected network, the diffusive process controlled by the TFD equation always reaches a steady state, which consists of the average of the values of the initial condition. In the case of directed networks (PPI are not directed by nature) or in disconnected networks (a situation that can be found in PPIs), the steady state is reached in each (strongly) connected component of the graph. Also, because the network is connected, μ_2 makes the largest contribution to $E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right)$ among all the nontrivial eigenvalues of $\mathcal{L}$. Therefore, it dictates the rate of convergence of the diffusion process. We remark that in practice, the steady state $\lim_{t \to \infty} \|x_v(t) - x_w(t)\| = 0, \forall v, w \in V$ is very difficult to achieve. Therefore, we use a threshold ε, e.g., $\varepsilon = 10^{-3}$, such that $\lim_{t \to \infty} \|x_v(t) - x_w(t)\| = \varepsilon$ is achieved in a relatively small simulation time.

Due to its importance in this work, we remark the structural meaning of the Mittag–Leffler function of the Laplacian matrix appearing in the solution of the TFD equation. That is, $E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right)$ is a matrix function, which is defined as

$$E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right) = \sum_{k=0}^\infty \frac{\left(-t^\alpha \mathcal{L}\right)^k}{\Gamma\left(\alpha k + 1\right)}, \alpha > 0, \tag{14}$$

where $\Gamma\left(\cdot\right)$ is the Euler gamma function as before. We remark that for $\alpha = 1$, we recover the diffusion equation on the network: $dx(t)/dt = -\mathcal{L}x(t)$ and its solution $E_{1,1}\left(-t^\alpha \mathcal{L}\right) = \exp\left(-tC\mathcal{L}\right)$ is the well-known heat kernel of the graph.

C. Time-fractional diffusion distance

We define here a generalization of the diffusion distance studied by Coifman and Lafon.[47] Let $f(\tau \mathcal{L}) = E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right)$, such that $f(\tau \mathcal{L})_{vw}$ represents the v, w entry of the matrix function $E_{\alpha,1}\left(-t^\alpha \mathcal{L}\right)$. Then, we define the following quantity:

$$D_{vw} = f(\tau \mathcal{L})_{vv} + f(\tau \mathcal{L})_{ww} - 2f(\tau \mathcal{L})_{vw}. \tag{15}$$

We have the following result.

Theorem 2. *The function $D_{vw} = f(\tau \mathcal{L})_{vv} + f(\tau \mathcal{L})_{ww} - 2f(\tau \mathcal{L})_{vw}$ is a Euclidean distance between the corresponding pair of nodes in the network.*

Proof. The matrix function $f(\tau \mathcal{L})$ can be written as $f(\tau \mathcal{L}) = U f(\tau \Lambda) U^{-1}$. Let $\vec{\varphi}_u = \left[\psi_{1,u}, \psi_{2,u}, \ldots, \psi_{n,u}\right]^T$. Then,

$$D_{vw} = (\vec{\varphi}_v - \vec{\varphi}_w)^T f(\tau \Lambda)(\vec{\varphi}_v - \vec{\varphi}_w). \tag{16}$$

Therefore, because $f(\tau \mathcal{L})$ is positively defined, we can write

$$D_{vw} = \left((\vec{\varphi}_v - \vec{\varphi}_w)^T f^{1/2}(\tau \Lambda)\right)\left(f^{1/2}(\tau \Lambda)(\vec{\varphi}_v - \vec{\varphi}_w)\right)$$
$$= \left(f^{1/2}(\tau \Lambda)\vec{\varphi}_v - f^{1/2}(\tau \Lambda)\vec{\varphi}_w\right)^T\left(f^{1/2}(\tau \Lambda)\vec{\varphi}_v - f^{1/2}(\tau \Lambda)\vec{\varphi}_w\right)$$
$$= \left(\vec{x}_v(\tau) - \vec{x}_w(\tau)\right)^T\left(\vec{x}_v(\tau) - \vec{x}_w(\tau)\right)$$
$$= \left\|\vec{x}_v(\tau) - \vec{x}_w(\tau)\right\|^2, \tag{17}$$

where $\vec{x}_v(\tau) = f^{1/2}(\tau \Lambda)\vec{\varphi}_v$. Consequently, D_{vw} is a square Euclidean distance between v and w. In this sense, the vector $\vec{x}_v(\tau) = f^{1/2}(\tau \Lambda)\vec{\varphi}_v$ is the position vector of the node v in the diffusion Euclidean space. $\square$

Because $E_{1,1}\left(-t^\alpha \mathcal{L}\right) = \exp\left(-t^\alpha \mathcal{L}\right)$, we have that D_{vw} generalizes the diffusion distance studied by Coifman and Lafon, which is the particular case when $\alpha = 1$. Let μ_j be the jth eigenvalue and ψ_{ju} the uth entry of the jth eigenvector of the Laplacian matrix. Then, we can write the time-fractional diffusion distance as

$$D_{vw} = \sum_{j=1}^n \left(\psi_{ju} - \psi_{jv}\right)^2 E_{\alpha,1}\left(-\tau \mu_j\right). \tag{18}$$

It is evident that when $\alpha = 1$, D_{vw} is exactly the diffusion distance previously studied by Coifman and Lafon.[47] The fractional-time diffusion distance between every pair of nodes in a network can be represented in a matrix form as follows:

$$\Sigma(\tau) = \left(\vec{s}(\tau)\vec{1}^T + \vec{1}\vec{s}(\tau)^T - 2f(\tau \mathcal{L})\right)^{\circ 1/2}, \tag{19}$$

where $\vec{s} = \left[f(\tau \mathcal{L})_{11}, f(\tau \mathcal{L})_{22}, \ldots, f(\tau \mathcal{L})_{nn}\right]$ is a vector whose entries are the main diagonal terms of the Mittag–Leffler matrix function, $\vec{1}$ is an all-ones vector, and $\circ$ indicates an entrywise operation. Using this matrix, we can build the diffusion distance-weighted adjacency matrix of the network,

$$W(\tau) = A \circ \Sigma(\tau) = \Sigma(\tau) \circ A. \tag{20}$$

The shortest diffusion path between two nodes is then the shortest weighted path in $W(\tau)$.

Lemma 3. *The shortest (topological) path distance between two nodes in a graph is a particular case of the time-fractional shortest diffusion path length for $\tau \to 0$.*

Proof. Let us consider each of the terms forming the definition of the time-fractional diffusion distance and apply the limit of the very small $-t^\alpha$. That is,

$$\lim_{\tau \to 0} f(\tau \mathcal{L})_{vv} = \lim_{\tau \to 0} \sum_{r=1}^n \psi_{rv}^2 f(\tau \Lambda) = 1, \tag{21}$$

and in a similar way,

$$\lim_{\tau \to 0} f(\tau \mathcal{L})_{vw} = \lim_{\tau \to 0} \sum_{r=1}^{n} \psi_{rv} \psi_{rw} f(\tau \Lambda)$$

$$= \sum_{r=1}^{n} \psi_{rv} \psi_{rw} = 0. \quad (22)$$

Therefore, $\lim_{\tau \to 0} D_{vw}(\tau) = 2$. Consequently, $\lim_{\tau \to 0} W(\tau) = 2A$, which immediately implies that the time-fractional shortest diffusion path is identical to the shortest (topological) one in the limit of very small $\tau = -t^{\alpha}$. □

D. Network of proteins targeted by SARS-CoV-2

The proteins of SARS-CoV-2 and their interactions with human proteins were determined experimentally by Gordon *et al.*[14] Gysi *et al.*[15] constructed an interaction network of all 239 human proteins targeted by SARS-CoV-2. In this network, the nodes represent human proteins targeted by SARS-CoV-2 and two nodes are connected if the corresponding proteins have been determined to interact with each other. Obviously, this network of proteins targeted by SARS-CoV-2 is a subgraph of the protein–protein interaction (PPI) network of humans. One of the surprising findings of Gysi *et al.*[15] is the fact that this subgraph is not formed by proteins randomly distributed across the human PPI, but they form a main cluster of 208 proteins and a few small isolated components. Hereafter, we will always consider this connected component of human proteins targeted by SARS-CoV-2. This network is formed by 193 proteins, which are significantly expressed in the lungs. Gysi *et al.*[15] reported a protein as being significantly expressed in the lungs if its GTEx median value is larger than 5. GTEx[48] is a database containing the median gene expression from RNA-seq in different tissues. The other 15 proteins are mainly expressed in other tissues. However, in reporting here, the tissues that were proteins are mainly expressed; we use the information reported in The Human Protein Atlas[49] where we use information not only from GTEx but also from HPA (see details at the Human Protein Atlas webpage) and FANTOM5[50] datasets.

III. RESULTS

A. Global characteristics of the time-fractional diffusion

The PPI network of human proteins targeted by SARS-CoV-2 is very sparse, having 360 edges, i.e., its edge density is 0.0167, 30% of nodes have a degree (number of connections per protein) equal to one, and the maximum degree of a protein is 14. The second smallest eigenvalue of the Laplacian matrix of this network is very small; i.e., $\mu_2 = 0.0647$. Therefore, the rate of convergence to the steady state of the diffusion processes taking place on this PPI is very slow. We start by analyzing the effects of the fractional coefficient α on these diffusive dynamics. We use the normal diffusion $\alpha = 1$ as the reference system.

To analyze the effects of changing α over the diffusive dynamics on the PPI network, we consider the solution of the TFD equation for processes starting at a protein with a large degree, i.e., PRKACA,

degree 14, and a protein with a low degree, i.e., MRPS5, degree 3. That is, the initial condition vector consists of a vector having one at the entry corresponding to either PRKACA or MRPS5 and zeroes elsewhere. In Fig. 1, we display the changes of the probability with the shortest path distance from the protein where the process starts. This distance corresponds to the number of steps that the perturbation needs to traverse to visit other proteins. For $\alpha = 1.0$, the shapes of the curves in Fig. 1 are the characteristic ones for the Gaussian decay of the probability with distance. However, for $\alpha < 1$, we observe that such decay differs from that typical shape showing a faster initial decay followed by a slower one. In order to observe this effect in a better way, we zoomed the region of distances from 2 to 4 [see Figs. 1(b) and 1(d)]. As can be seen for distances below 3, the curve for $\alpha = 1.0$ is on top of those for $\alpha < 1$, indicating a slower decay of the probability. After this distance, there is an inversion, and the normal diffusion occurs at a much faster rate than the other two for the longer distances. This is a characteristic signature of subdiffusive processes, which starts at much faster rates than a normal diffusive process and then continue at much slower rates. Therefore, here, we observe that the subdiffusive dynamics are much faster at earlier times of the process, which is when the perturbation occurs to close nearest neighbors to the initial point of perturbation.

To further investigate these characteristic effects of the subdiffusive dynamics, we study the time evolution of a perturbation occurring at a given protein and its propagation across the whole PPI network. In Fig. 2, we illustrate these results for $\alpha = 1.0$ (a), $\alpha = 0.75$ (b), and $\alpha = 0.5$ (c). As can be seen in the main plots of this figure, the rate of convergence of the processes to the steady state is much faster in the normal diffusion (a) than in the subdiffusive one (b) and (c). However, at very earlier times (see insets in Fig. 2), there is a shock wave increase of the perturbation at a set of nodes. Such kind of shock waves has been previously analyzed in other contexts as a way of propagating effects across PPI networks.[17] We have explored briefly about the possible causes of this increase in the concentration for a given subset of proteins. Accordingly, it seems that the main reason for this is the connectivity provided by the network of interactions and not a given distribution of the degrees. For instance, we have observed such "shock waves" in networks with normal-like distributions as well as with power-law ones. However, it is possible that the extension and intensity of such effects depend on the degree distribution as well as on other topological factors. The remarkable finding here is, however, the fact that such a shock wave occurs at much earlier times in the subdiffusive regimes than at the normal diffusion. That is, while for $\alpha = 1.0$, these perturbations occur at $t \approx 0.1$–0.3; for $\alpha = 0.75$, they occur at $t \approx 0.0$–0.2; and for $\alpha = 0.5$, they occur at $t \approx 0.0$–0.1. Seeing this phenomenon in the light of what we have observed in the previous paragraph is not strange due to the observation that such processes go at a much faster rate at earlier times, and at short distances, than the normal diffusion. In fact, this is a consequence of the existence of a positive scalar T for which $E_{\alpha,1}(-\gamma t^{\alpha})$ decreases faster than $\exp(-\gamma t)$ for $t \in (0, T)$ for $\gamma \in \mathbb{R}^+$ and $\alpha \in \mathbb{R}^+$ (see Theorem 4.1 in Ref. 39). Hereafter, we will consider the value of $\alpha = 0.75$ for our experiments due to the fact that it reveals a subdiffusive regime, but the shock waves observed before are not occurring in an almost instantaneous way like when $\alpha = 0.5$, which would be difficult from a biological perspective.
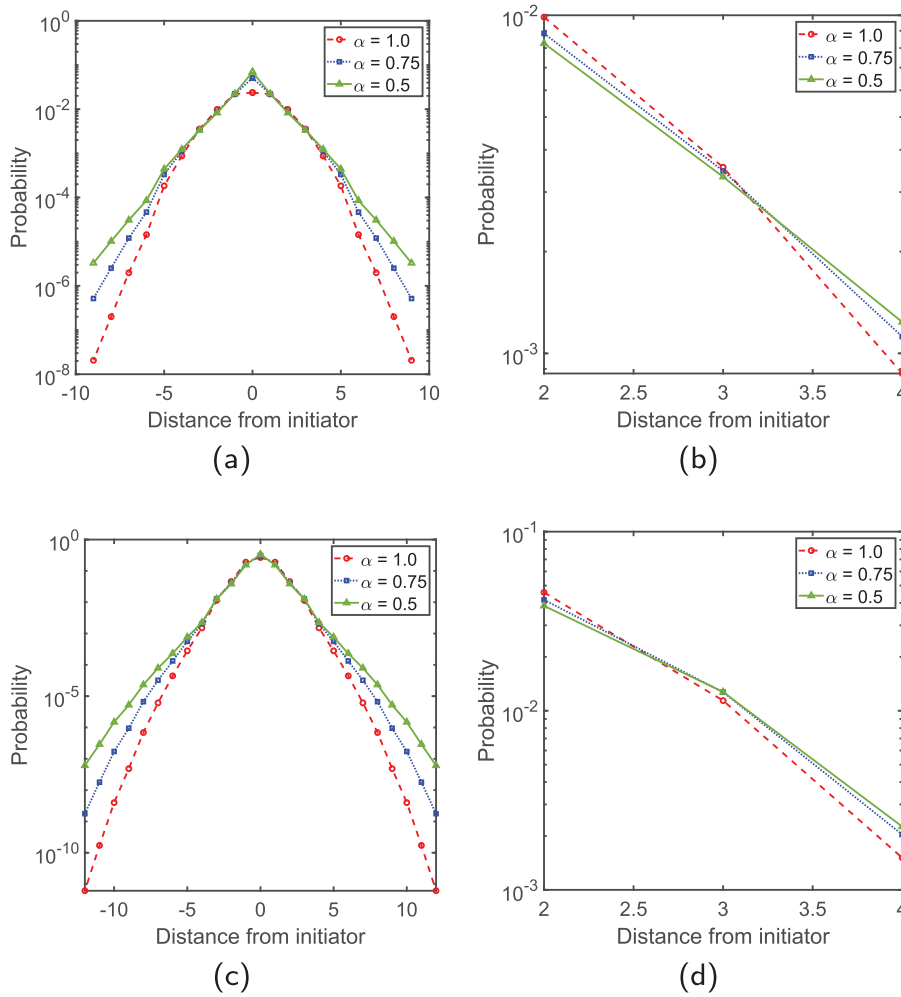
FIG. 1. Spatial propagation of perturbations. The average probability that a protein at a given distance from the perturbed protein feels the perturbation. (a) The initiators are the protein PRKACA (a) and (b) as well as MRPS5 (c) and (d). (b) and (d) zoom the scale of the processes in (a) and (c), respectively. The y axis is in a logarithmic scale. The distance (x axis) refers to the shortest path distance from the initiator, and the plots are made "artificially" symmetric for better visualization.
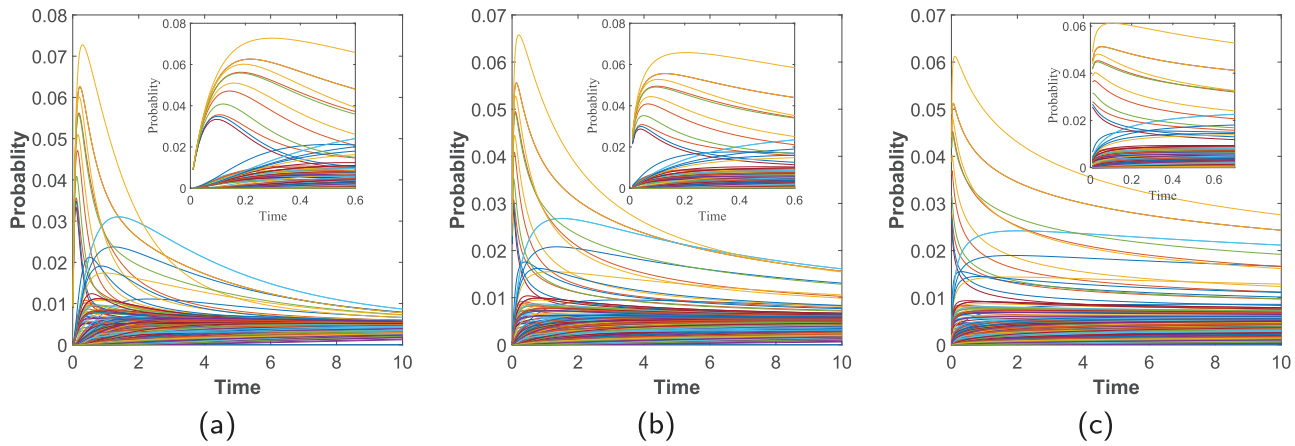

FIG. 2. Shock wave increase of perturbations across the network. Time evolution of the propagation of perturbations from the protein PRKACA to the rest of the proteins in the PPI network of human proteins targeted by SARS-CoV-2. (a) Normal diffusion $\alpha = 1.0$. (b) Subdiffusion obtained for $\alpha = 3/4$. (c) Subdiffusion obtained for $\alpha = 1/2$. The insets illustrate the shortest time evolution of the perturbations. Every curve corresponds to a protein in the PPI.

The previous results put us at a crossroads. First, the subdiffusive processes that are expected due to the crowded nature of the intra-cellular space are very slow for carrying out cellular processes at a significant rate in cells. However, the perturbation shocks occurring at earlier times of these processes are significantly faster than in normal diffusion. To sort out these difficulties, we propose a switching back and restart subdiffusive process occurring in the PPI network. That is, a subdiffusive process starts at a given protein, which is directly perturbed by a protein of SARS-CoV-2. It produces a shock wave increase of the perturbation in close neighbors of that proteins. Then, a second subdiffusive process starts at these newly perturbed proteins, which will perturb their nearest neighbors. The process is repeated until the whole PPI network is perturbed. This kind of "switch and restart processes" has been proposed for engineering consensus protocols in multiagent systems[51] as a way to accelerate the algorithms using subdiffusive regimes.

B. Time-fractional diffusion from the SARS-CoV-2 spike protein

The so-called Spike protein (S-protein) of the SARS-CoV-2 interacts with only two proteins in the human hosts, namely, ZDHHC5 and GOLGA7. The first protein, ZDHHC5, is not in the main connected component of the PPI network of SARS-CoV-2 targets. Therefore, we will consider here how a perturbation produced by the interaction of the virus S-protein with GOLGA7 is propagated through the whole PPI network of SARS-CoV-2 targets. GOLGA7 has degree one in this network, and its diffusion is mainly to close neighbors, namely, to proteins separated by two to three edges. When starting the diffusion process at the protein GOLGA7, the main increase in the probability of perturbing another protein is reached for the protein GOLGA3, which increases its probability up to 0.15 at $t = 0.2$, followed by PRKAR2A, with a small increase in its probability, 0.0081. Then, the process switch and restarts at GOLGA3, which mainly triggers the probability of the protein PRKAR2A—a major hub of the network. Once we start the process at PRKAR2A, practically, the whole network is perturbed with probabilities larger than 0.1 for 19 proteins apart from GOLGA3. These proteins are in decreasing order of their probability of being perturbed: AKAP8, PRKAR2B, CEP350, MIB1, CDK5RAP2, CEP135, AKAP9, CEP250, PCNT, CEP43, PDE4DIP, PRKACA, TUB6CP3, TUB6CP2, CEP68, CLIP4, CNTRL, PLEKHA5, and NINL. Notice that the number of proteins perturbed is significantly larger than the degree of the activator, indicating that not only nearest neighbors are activated.

An important criterion for revealing the important role of the protein PRKAR2A as a main propagator in the network of proteins targeted by SARS-CoV-2 is its average diffusion path length. This is the average number of steps that a diffusive process starting at this protein needs to perturb all the proteins in the network. We have calculated this number to be 3.6250, which is only slightly larger than the average (topological) path length, which is 3.5673. That is, in less than four steps, the whole network of proteins is activated by a diffusive process starting at PRKAR2A. Also remarkable that the average shortest diffusive path length is almost identical to the shortest (topological) one. This means that this protein mainly uses shortest (topological) paths in perturbing other proteins in the
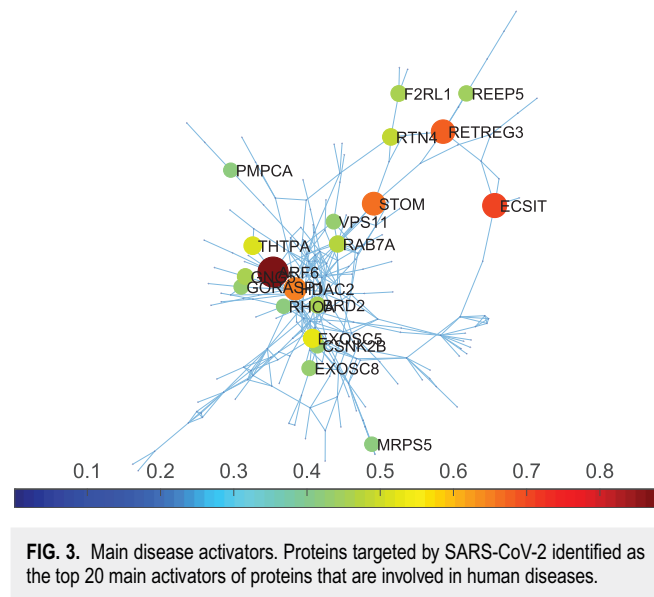
FIG. 3. Main disease activators. Proteins targeted by SARS-CoV-2 identified as the top 20 main activators of proteins that are involved in human diseases.

PPI. In other words, it is highly efficient in conducting such perturbations. We will analyze this characteristics of the PPI of human proteins targeted by SARS-CoV-2 in a further section of this work.

At this time, almost any protein in the PPI network is already perturbed. Therefore, we can switch and restart the subdiffusion from practically any protein at the PPI network. We then investigate which are the proteins with the higher capacity of activating other proteins that are involved in human diseases. Here, we use the database DisGeNet,[52] which is one of the largest publicly available collections of genes and variants associated with human diseases. We identified 38 proteins targeted by SARS-CoV-2 for which there is a "definitive" or "strong" evidence of being involved in a human disease or syndrome (see Table S1 in the supplementary material). These proteins participate in 70 different human diseases or syndromes as given in Tables S2 and S3 of the supplementary material. We performed an analysis in which a diffusive process starts at any protein of the network, and we calculated the average probability that all the proteins involved in human diseases are then perturbed. For instance, for a subdiffusive process starting at the protein ARF6, we summed up the probabilities that the 38 proteins involved in diseases are perturbed at an early time of the process $t = 0.2$. Then, we obtain a global perturbation probability of 0.874. By repeating this process for every protein as an initiator, we obtained the top disease activators. We have found that none of the 20 top activators is involved itself in any of the human diseases or syndromes considered here. They are, however, proteins that are important not because of their direct involvement in diseases or syndromes but because they propagate perturbations in a very effective way to those directly involved in such diseases/syndromes. Among the top activators, we have found ARF6, ECSIT, RETREG3, STOM, HDAC2, EXOSC5, THTPA, among others shown in Fig. 3, where we illustrate the PPI network of the proteins targeted by SARS-CoV-2 remarking the top 20 disease activators.
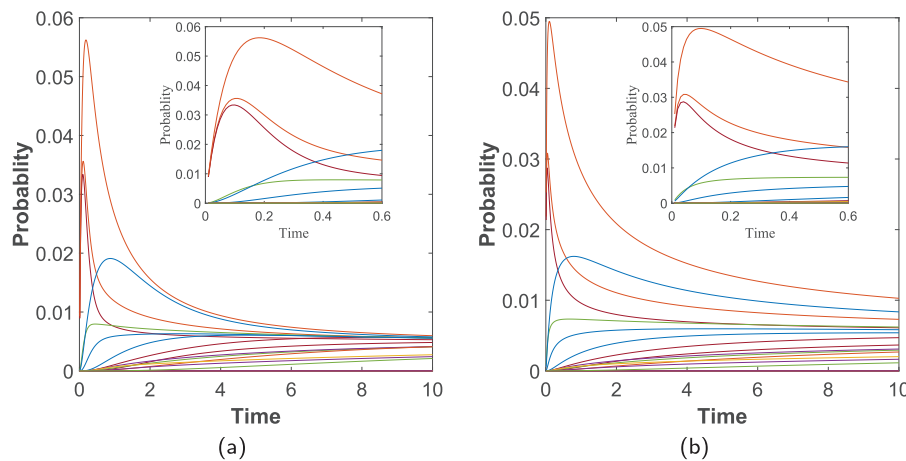
FIG. 4. Shock wave increase of perturbations in proteins outside the lungs. Time evolution of the propagation of perturbations from the protein GOLGA2 to proteins mainly expressed outside the lungs. (a) Normal diffusion $\alpha = 1.0$. (b) Subdiffusion with $\alpha = 3/4$. The insets illustrate the shortest time evolution of the perturbation. Every curve corresponds to a protein in the PPI outside the lungs.

C. Time-fractional diffusion from the lungs to other organs

We now consider how a perturbation produced by SARS-CoV-2 on a protein mainly expressed in the lungs can be propagated to proteins mainly located in other tissues (see Table S4 in the supplementary material) by a subdiffusive process. That is, we start the subdiffusive process by perturbing a given protein, which is mainly expressed in the lungs. Then, we observe the evolution of the perturbation at every one of the proteins mainly expressed in other tissues. We repeat this process for all the 193 proteins mainly expressed in the lungs. In every case, we record those proteins outside the lungs, which are perturbed at very early times of the subdiffusive process. For instance, in Fig. 4, we illustrate one example in which the initiator is the protein GOLGA2, which triggers a shock wave on proteins RBM41, TL5, and PKP2, which are expressed mainly outside the lungs. We consider such perturbations only if they occur at $t < 1$.

Not every one of the proteins expressed outside the lungs is triggered by such shock waves at a very early time of the diffusion. For instance, proteins MARK1 and SLC27A2 are perturbed in very slow processes and do not produce the characteristic high peaks in the probability at very short times. On the other hand, there are proteins expressed outside the lungs that are triggered by more than one protein from the lungs. The case of GOLGA2 is an example of a protein triggered by three proteins in the lungs. In Table I, we list some of the proteins expressed mainly in tissues outside the lungs, which are heavily perturbed by proteins in the lungs. The

TABLE I. Multi-organ propagation of perturbations. Proteins mainly expressed outside the lungs are significantly perturbed during diffusive processes that have started at other proteins expressed in the lungs. Act. is the number of lung proteins activators, p_{tot} is the sum of the probabilities of finding the diffusive particle at this protein, and t_{mean} is the average time of activation (see the text for explanations). The tissues of main expression are selected among the ones with the highest Consensus Normalized eXpression (NX) levels by combining the data from the three transcriptomics datasets (HPA, GTEx, and FANTOM5) using the internal normalization pipeline.[49] Boldface denotes the highest value in each of the columns.

| Protein	Act.	p_{tot}	t_{mean}	Tissues of the main expression
PKP2	**21**	0.498	0.29	Heart muscle
CEP43	18	0.521	0.26	Testis
CEP135	14	0.527	0.30	Skeletal muscle, heart muscle, cerebral cortex, cerebellum
TLE5	13	0.509	**0.24**	Thymus, lymph node, testis
RETREG3	6	0.390	0.43	Prostate, thymus
RBM41	6	0.209	0.49	Pancreas, T-cells, testis, retina
PRIM2	5	**0.536**	0.30	Cerebellum, parathyroid gland, testis
MIPOL1	3	0.155	0.57	Pituitary gland, testis
REEP6	1	0.175	0.48	Liver, small intestine, duodenum, testis
HOOK1	1	0.156	0.46	Liver, parathyroid gland, testis, pituitary gland
CENPF	1	0.138	0.46	Thymus, testis, bone marrow
ATP5ME	1	0.096	0.35	Skeletal muscle, dendritic cells, heart muscle, brain
TRIM59	1	0.170	0.69	Corpus callosum, brain, T-cells, testis
MARK1	1	N/A	>100	Epididymis, cerebral cortex, heart muscle, testis, cerebellum
SLC27A2	1	N/A	>100	Liver, kidney

complete list of the perturbing proteins is given in Table S5 of the supplementary material. We give three indicators of the importance of the perturbation of these proteins. They are Act., which is the number of proteins in the lungs that activate each of them; p_{tot}, which is the sum of the probabilities of finding the diffusive particle at this protein for diffusive processes that have started in their activators; and t_{mean}, which is the average time required by activators to perturb the corresponding protein. For instance, PKP2 is perturbed by 21 proteins in the lungs, which indicates that this protein, mainly expressed in the heart muscle, has a large chance of being perturbed by diffusive processes starting in proteins mainly located at the lungs. Protein PRIM2 is activated by 5 proteins in the lungs, but if all these proteins were acting at the same time, the probability that PRIM2 is perturbed will be very high, $p_{tot} \approx 0.536$. Finally, protein TLE5 is perturbed by 13 proteins in the lungs, which needs as an average $t_{mean} \approx 0.24$ to perturb TLE5. These proteins do not form a connected component among them in the network. The average shortest diffusion path between them is 5.286 with a maximum shortest subdiffusion path of 10. As an average, they are almost equidistant from the rest of the proteins in the network as among themselves. That is, the average shortest subdiffusion path between these proteins expressed outside the lungs and the rest of the proteins in the network is 5.106. Therefore, these proteins can be reached from other proteins outside the lungs in no more than six steps in subdiffusive processes like the ones considered here.

D. Subdiffusion paths

Finally, we study here how the diffusive process determines the paths that the perturbation follows when diffusing from a protein to another not directly connected to it. The most efficient way of propagating a perturbation between the nodes of a network is through the shortest (topological) paths that connect them. The problem for a (sub)diffusive perturbation propagating between the nodes

of a network is that it does not have complete information about the topology of the network as to know its shortest (topological) paths. The network formed by the proteins targeted by SARS-CoV-2 is very sparse, and this indeed facilitates that the perturbations occurs by following the shortest (topological) paths most of the time. Think, for instance, in a tree, which has the lowest possible edge density among all connected networks. In this case, the perturbation will always use the shortest (topological) paths connecting pairs of nodes. However, in the case of the PPI network studied here, a normal diffusive process, i.e., $\alpha = 1$, not always uses the shortest (topological) paths. In this case, there are 1294 pairs of proteins for which the diffusive particle uses a shortest diffusive path, which is one edge longer than the corresponding shortest (topological) path. This represents 6.11% of all total pairs of proteins that are interconnected by a path in the PPI network of proteins targeted by SARS-CoV-2. However, when we have a subdiffusive process, i.e., $\alpha = 0.75$, this number is reduced to 437, which represents only 2.06% of all pairs of proteins. Therefore, the subdiffusion process studied here through the PPI network of proteins targeted by SARS-CoV-2 has an efficiency of 97.9% relative to a process that always uses the shortest (topological) paths in hopping between proteins. In Fig. 5, we illustrate the frequency with which proteins not in the shortest (topological) paths are perturbed as a consequence that they are in the shortest subdiffusive paths between other proteins. For instance, the following is a shortest diffusive path between the two end points: RHOA-PRKACA-PRKAR2A-CEP43-RAB7A-ATP6AP1. The corresponding shortest (topological) path is RHOA-MARK2-AP2M1-RAB7A-ATP6AP1, which is one edge smaller. The proteins PRKACA, PRKAR2A, and CEP43 are those in the diffusive path that are not in the topological one. Repeating this selection process for all the diffusive paths that differs from the topological ones, we obtained the results illustrated in Fig. 5. As can be seen, there are 36 proteins visited by the shortest diffusive paths, which are not visited by the corresponding topological ones. The
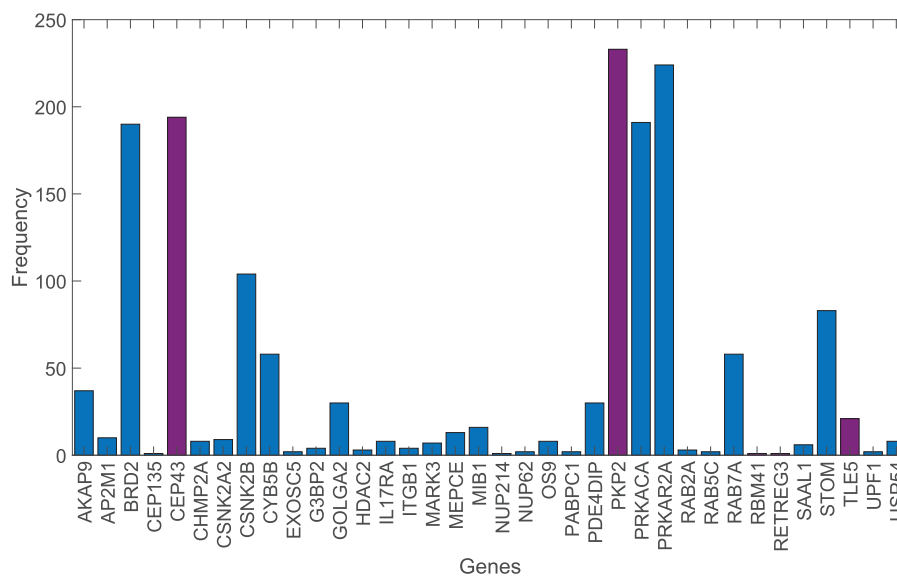
FIG. 5. Proteins in shortest subdiffusive paths. The frequency of proteins that appears in the shortest subdiffusion paths but not in the shortest (topological) paths connecting any pair of nodes in the PPI network. Bars marked in wine color are for those proteins expressed mainly outside the lungs.

average degree of these proteins is 7.28, and there is only a small positive trend between the degree of the proteins and the frequency with which they appear in these paths; e.g., the Pearson correlation coefficient is 0.46.

IV. DISCUSSION

We have presented a methodology that allows the study of diffusive processes in (PPI) networks varying from normal to subdiffusive regimes. Here, we have studied the particular case in which the time-fractional diffusion equation produces a subdiffusive regime, with the use of $\alpha = 3/4$ in the network of human proteins targeted by SARS-CoV-2. A characteristic feature of this PPI network is that the second smallest eigenvalue is very small; i.e., $\mu_2 = 0.0647$. As this eigenvalue determines the rate of convergence to the steady state, the subdiffusive process converges very slowly to that state. What it has been surprising is that even in these conditions of very small convergence to the steady state, there is a very early increase of the probability in those proteins closely connected to the initiator of the diffusive process. That is, in a subdiffusive process on a network, the time at which a perturbation is transmitted from the initiator to any of its nearest neighbors occurs at an earlier time than for the normal diffusion. This is a consequence of the fact that $E_{\alpha,1}(-\gamma t^\alpha)$ decreases very fast at small values of t^α, which implies that the perturbation occurring at a protein i at $t = 0$ is transmitted almost instantaneously to the proteins closely connected to i. This effect may be responsible for the explanation about why subdiffusive processes, which are so globally slow, can carry out cellular processes at a significant rate in cells. We have considered here a mechanism consisting in switching and restarting several times during the global cellular process. For instance, a subdiffusive process starting at the protein i perturbs its nearest neighbors at very early times, among which we can find the protein j. Then, a new subdiffusive process can be restarted again at the node j and so on.

One of the important findings of using the current model for the study of the PIN of proteins affected by SARS-CoV-2 is the identification of those proteins that are expressed outside the lungs that can be more efficiently perturbed by those expressed in the lungs (see Table I). For instance, the protein with the largest number of activators, PKP2, appears mainly in the heart muscle. It has been observed that the elevation of cardiac biomarkers is a prominent feature of COVID-19, which in general is associated with a worse prognosis.[53] Myocardial damage and heart failure are responsible for 40% of death in the Wuhan cohort (see references in Ref. 53). Although the exact mechanism involving the heart injury is not known, the hypothesis of direct myocardial infection by SARS-CoV-2 is a possibility, which acts along or in combination with the increased cardiac stress due to respiratory failure and hypoxemia, and/or with the indirect injury from the systemic inflammatory response.[53–56] As can be seen in Table I, the testis is the tissue where several of the proteins targeted by SARS-CoV-2 are mainly expressed, e.g., CEP43, TLE5, PRIM2, MIPOL1, REEP6, HOOK1, CENPF, TRIM59, and MARK1. Currently, there is no conclusive evidence about the testis damage by SARS-CoV-2.[57–60] However, the previous SARS-CoV that appeared in 2003 and which shares 82% of proteins with the current one produced testis damage and spermatogenesis, and it was concluded that orchitis was a complication

of that previous SARS disease.[57] We also detect a few proteins mainly expressed in different brain tissues, such as CEP135, PRIM2, TRIM59, and MARK1. The implication of SARS-CoV-2 and cerebrovascular diseases has been reported, including neurological manifestations as well as cerebrovascular disease, such as ischemic stroke, cerebral venous thrombosis, and cerebral hemorrhage.[61–63]

Kidney damage in SARS-CoV-2 patients has been reported,[64–66] which includes signs of kidney dysfunctions, proteinuria, hematuria, increased levels of blood urea nitrogen, and increased levels of serum creatinine. As much as 25% of an acute kidney injury has been reported in the clinical setting of SARS-CoV-2 patients. One of the potential mechanisms for kidney damage is the organ crosstalk,[64] as can be the mechanism of diffusion from proteins in the lungs to proteins in the urinary tract and kidney proposed here. A very interesting observation from Table I is the existence of several proteins expressed mainly in the thymus and T-cells, such as TLE5, RETREG3, RBM41, CENPF, and TRIM59. It has been reported that many of the patients affected by SARS-CoV-2 in Wuhan displayed a significant decrease of T-cells.[67] Thymus is an organ that displays a progressive decline with age with reduction of the order of 3%–5% a year until approximately 30–40 years of age and of about 1% per year after that age. Consequently, it was proposed that the role of thymus should be taken into account in order to explain why COVID-19 appears to be so mild in children.[67] The protein TLE5 is also expressed significantly in the lymph nodes. It was found by Feng *et al.*[68] that SARS-CoV-2 induces lymph follicle depletion, splenic nodule atrophy, histiocyte hyperplasia, and lymphocyte reductions. The proteins HOOK1 and MIPOL1 are significantly expressed in the pituitary gland. There has been some evidence and concerns that COVID-19 may also damage the hypothalamo-pituitary-adrenal axis that has been expressed by Pal,[69] which may be connected with the participation of the previously mentioned proteins.

Another surprising finding of the current work is the elevated number of subdiffusive shortest paths that coincide with the shortest (topological) paths connecting pairs of proteins in the PPI of human proteins targeted by SARS-CoV-2. This means that the efficiency of the diffusive paths connecting pairs of nodes in this PPI is almost 98% in relation to a hypothetical process that uses the shortest (topological) paths in propagating perturbations between pairs of proteins. The 437 shortest diffusive paths reported here contain one more edge than the corresponding shortest (topological) paths. The proteins appearing in these paths would never be visited in the paths connecting two other proteins if only the shortest (topological) paths were used. What is interesting to note that 6 out of the 15 proteins that are mainly expressed outside the lungs are among the ones "crossed" by these paths. They are TLE5 (thymus, lymph node, testis), PKP2 (heart muscle), CEP135 (skeletal muscle, heart muscle, cerebral cortex, cerebellum), CEP43 (testis), RBM41 (pancreas, T-cells, testis, retina), and RETREG3 (prostate, thymus). This means that the perturbation of these proteins occurs not only through the diffusion from other proteins in the lungs directly to them, but also through some "accidental" diffusive paths between pairs of proteins that are both located in the lungs.

All in all, the use of time-fractional diffusive models to study the propagation of perturbations on PPI networks seems a very promising approach. The model is not only biologically sounded but it also allows us to discover interesting hidden patterns of the

interactions between proteins and the propagation of perturbations among them. In the case of the PIN of human proteins targeted by SARS-CoV-2, our current finding may help to understand potential molecular mechanisms for the multi-organs and systemic failures occurring in many patients.

V. EXPERIMENTAL VALIDATION

After this work was completed, Qiu *et al.*[75] uploaded the manuscript entitled "Postmortem tissue proteomics reveals the pathogenesis of multiorgan injuries of COVID-19." The authors profiled the host responses to COVID-19 by means of quantitative proteomics in postmortem samples of tissues in lungs, kidney, liver, intestine, brain, and heart. They reported differentially expressed proteins (DEPs) for these organs as well as virus-host PPIs between 23 virus proteins and 110 interacting DEPs differentially regulated in postmortem lung tissues. According to their results, most DEPs (70.5%) appears in the lungs, followed by kidney (16.5%). Additionally, Qiu *et al.*[75] identified biological processes that were up- or down-regulated in the six postmortem tissue types. They found that most up-regulated processes in the lungs correspond to processes related to the response to inflammation and to immune response. However, pathways related to cell morphology, such as the establishment of endothelial barriers, were down-regulated in the lungs, which was interpreted as a confirmation that the lungs are the main focus of virus-host fights. Other fundamental processes in the six organs analyzed postmortem were significantly down-regulated, which include processes related to organ movement, respiration, and metabolism.

From the 59 proteins that we reported here as the ones with the largest effect on perturbing those 38 proteins identified in human diseases (see Table S3 in the supplementary material), 18 were found to be down-regulated in the lungs by Qiu *et al.*[75] If we make the corresponding adjustment, by considering that Qiu *et al.*[75] considered 110 instead of 209 proteins in the PPI, the previous number represents 58.1% of proteins predicted here and experimentally found as down-regulated in the lungs. From the rest of the proteins, which were not found as having the largest effect on perturbing proteins identified in human disease, only 29.1% were reported by Qiu *et al.*[75] to be down-regulated in the postmortem analysis of patients' lungs. Among the proteins reported in Table S3 of the supplementary material and by Qiu *et al.*, we have ARF6, RTN4, RAB7A, 6NG5, REEP5, VPS11, RHOA, RAB5C, among others. Finally, among the proteins mainly expressed outside the lungs that are predicted in this work to be significantly perturbed, we have five that were found by Qiu *et al.*[75] to be up-regulated in the different organs analyzed by them. From the proteins included in Table I, Qiu *et al.*[75] reported the following ones up-regulated: PKP2 (heart), REEP6 (liver), HOOK1 (several organs), ATP5ME (heart), and SLC27A2 (liver and kidney). They also reported CEP43 (reported as FGFR1OP) as down-regulated in the brain. We should remark that we have considered here many more organs than the six ones studied by Qiu *et al.*[75]

VI. FUTURE WORK

There are no doubts that in considering a diffusive propagation of perturbations among proteins in a PPI, we have made a
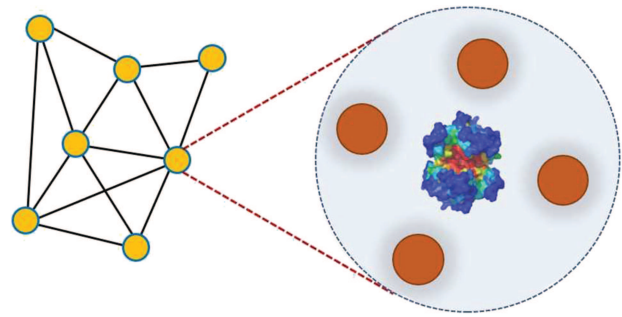
FIG. 6. PI metaplex. In the metaplex, every node of the PPI corresponds to a protein and its crowded intracellular space. There is an internal dynamics in the nodes and an external between the nodes.

few simplifications and assumptions. Every protein is embedded in an intracellular crowded environment, which drives its diffusive mechanism. Nowadays, it is well-established that this environment is conducive to molecular subdiffusive processes. As remarked by Guigas and Weiss,[27] far from obstructing cellular processes, subdiffusion increases the probability of finding a nearby target by a given protein, and therefore, it facilitates protein–protein interactions. The current approach can be improved using two recently developed theoretical frameworks: (i) metaplexes and (ii) d-path Laplacian operators on graphs.

A PPI metaplex,[70] a 4-tuple $\Upsilon = (V, E, \mathcal{I}, \omega)$, where (V, E) is a graph, $\omega = \{\Omega_j\}_{j=1}^k$ is a set of locally compact metric spaces Ω_j with Borel measures μ_j, and $\mathcal{I} : V \to \omega$ is illustrated in Fig. 6. Then, we define a dynamical $\Upsilon = (V, E, \mathcal{I}, \omega = \{\Omega_k\})$ on the metaplex as a tuple $(\mathcal{H}, \mathcal{T})$. Here, $\mathcal{H} = \{H_v : L^2(\Omega_{\mathcal{I}}(v), \mu_{\mathcal{I}}(v)) \to L^2(\Omega_{\mathcal{I}}(v), \mu_{\mathcal{I}}(v))\}_{v \in V}$ is a family of operators such that the initial value problem $\partial_t u_v = H_v(u_v)$, $u_v|_{t=0} = u_0$, is well-posed, and $\mathcal{T} = \{T_{vw}\}_{(v,w) \in E}$ is a family of bounded operators $T_{vw} : L^2(\Omega_{\mathcal{I}(v)}, \mu_{\mathcal{I}(v)}) \to L^2(\Omega_{\mathcal{I}(w)}, \mu_{\mathcal{I}(w)})$. This means that inside a node of the metaplex, we consider one protein and its crowded intracellular space. Inside the nodes, we can have a dynamics like a time-fractional diffusion equation, the fractional Fokker–Planck equation, or any other in a continuous space. The inter-nodes dynamics is then dominated by a graph-theoretic diffusive model like the one presented here.

The second possible improvement to the current model can be made by introducing the possibility of long-range interactions in the inter-nodes dynamics in the PPI metaplex. That is, instead of considering the time-fractional diffusion equation, which only accounts for subdiffusive processes in the graph, we can use the following generalization, which incorporates the d-path Laplacian operators,[71]

$$D_t^\alpha x(t) = -C \left(\sum_{d=1}^{diam} d^{-s} L_d \right) x(t), \qquad (23)$$

where L_d is a generalization of the graph Laplacian operator to account for long-range hops between nodes in a graph, d is the shortest path distance between two nodes, and $s > 0$ is a parameter. This equation has never been used before except that for the case $\alpha = 1$ where superdiffusive behavior was proved in 1- and

2-dimensional cases.[72,73] Other approaches have also been recently used for similar purposes in the literature.[74]

We then hope that the combination of metaplexes and time- and space-fractional diffusive models do capture more of the details of protein–protein interactions in crowded cellular environments.

SUPPLEMENTARY MATERIAL

See the supplementary material for a list of proteins targeted by SARS-CoV-2, which are found in the database DisGeNet as displaying "definitive" or "strong" evidence of participating in human diseases. The Disease ID is the code of the disease in DisGeNet. A list of proteins with the largest effect on perturbing those 38 proteins are identified in human diseases. p_{tot} is the sum of the probabilities that the given protein activates those identified as having "definitive" or "strong" evidence of being involved in a human disease. There is an RNA expression overview for proteins targeted by SARS-CoV-2 and mainly expressed outside the lungs. We select the top RNA expressions in the four databases reported in The Human Protein Atlas. There is a list of proteins mainly expressed outside the lungs and their major activators, which are proteins mainly expressed in the lungs.

ACKNOWLEDGMENTS

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material and also from the corresponding author upon reasonable request.

REFERENCES

[1] A. Gorbalenya, S. Baker, and R. Baric (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses), "The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2," Nat. Microbiol. **5**, 536–544 (2020).

[2] P. Zhou, X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, and H. D. Chen, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," Nature **579**, 270–273 (2020).

[3] F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, and M. L. Yuan, "A new coronavirus associated with human respiratory disease in china," Nature **579**, 265–269 (2020).

[4] F. Jiang, L. Deng, L. Zhang, Y. Cai, C. W. Cheung, and Z. Xia, "Review of the clinical characteristics of coronavirus disease 2019 (COVID-19)," J. Gen. Intern. Med. **35**, 1545–1549 (2020).

[5] B. Zhang, X. Zhou, Y. Qiu, F. Feng, J. Feng, Y. Jia, H. Zhu, K. Hu, J. Liu, Z. Liu, S. Wang, Y. Gong, C. Zhou, T. Zhu, Y. Cheng, Z. Liu, H. Deng, F. Tao, Y. Ren, B. Cheng, L. Gao, X. Wu, L. Yu, Z. Huang, Z. Mao, Q. Song, B. Zhu, and J. Wang, "Clinical characteristics of 82 death cases with COVID-19," medRxiv (2020).

[6] M. Zhou, X. Zhang, and J. Qu, "Coronavirus disease 2019 (COVID-19): A clinical update," Front. Med. **14**, 126–135 (2020).

[7] S. Zaim, J. H. Chong, V. Sankaranarayanan, and A. Harky, "COVID-19 and multi-organ response," Curr. Probl. Cardiol. **45**, 100618 (2020).

[8] J. F. Chan, K. H. Kok, Z. Zhu, H. Chu, K. K. To, S. Yuan, and K. Y. Yuen, "Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan," Emerg. Microb. Infect. **9**, 221–236 (2020).

[9] W. Li, M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, and M. Farzan, "Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus," Nature **426**, 450–454 (2003).

[10] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krueger, T. Herrler, S. Erichsen, T. S. Schiergens, G. herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten, and S. Pöhlmann, "SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor," Cell **181**, 272–280 (2020).

[11] X. Chai, L. Hu, Y. Zhang, W. Han, Z. Lu, A. Ke, J. Zhou, G. Shi, N. Fang, J. Fan, J. Cai, J. Fan, and F. Lan, "Specific ACE2 expression in cholangiocytes may cause liver damage after 2019-nCoV infection," bioRxiv (2020). 10.1101/2020.02.03.931766

[12] Y. Zhao, Z. Zhao, Y. Wang, Y. Zhou, Y. Ma, and W. Zuo, "Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov," bioRxiv (2020).

[13] I. Hamming, W. Timens, M. L. Bulthuis, A. T. Lely, G. J. Navis, and H. van Goor, "Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis," J. Pathol. **203**, 631–637 (2004).

[14] D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, M. J. O'meara, J. Z. Guo, D. L. Swaney, T. A. Tummino, R. Huttenhain, and R. M. Kaake, "A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug-repurposing," Nature **583**, 459 (2020).

[15] D. M. Gysi, I. D. Valle, M. Zitnik, A. Ameli, X. Gan, O. Varol, H. Sanchez, R. M. Baron, D. Ghiassian, J. Loscalzo, and A. L. Barabási, "Network medicine framework for identifying drug repurposing opportunities for COVID-19," arXiv:2004.07229 (2020).

[16] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," Science **296**, 910–913 (2002).

[17] M. A. Antal, C. Bode, and P. Csermely, "Perturbation waves in proteins and protein networks: Applications of percolation and game theories in signaling and drug design," Curr. Prot. Pept. Sci. **10**, 161–72 (2009).

[18] M. Santolini and A. L. Barabási, "Predicting perturbation patterns from the topology of biological networks," Proc. Natl. Acad. Sci. U.S.A. **115**, E6375–E6383 (2018).

[19] B. K. Stöcker, J. Köster, E. Zamir, and S. Rahmann, "Modeling and simulating networks of interdependent protein interactions," Integr. Biol. **10**, 290–305 (2018).

[20] A. L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: A network-based approach to human disease," Nat. Rev. Genet. **12**, 56–68 (2011).

[21] I. W. Taylor and J. L. Wrana, "Protein interaction networks in medicine and disease," Proteomics **12**, 1706–1716 (2012).

[22] L. I. Furlong, "Human diseases through the lens of network biology," Trends Genet. **29**, 150–159 (2013).

[23] N. N. Batada, L. A. Shepp, and D. O. Siegmund, "Stochastic model of protein–protein interaction: Why signaling proteins need to be colocalized," Proc. Natl. Acad. Sci. U.S.A. **101**, 6445–6449 (2004).

[24] M. Zacharias, "Accounting for conformational changes during protein-protein docking," Curr. Opin. Struct. Biol. **20**, 180–186 (2010).

[25] L. Chen, Y. H. Zhang, Z. Zhang, T. Huang, and Y. D. Cai, "Inferring novel tumor suppressor genes with a protein-protein interaction network and network diffusion algorithms," Mol. Ther. Meth. Clin. Devel. **21**, 57–67 (2018).

[26] A. Stojmirović and Y. K. Yu, "Information flow in interaction networks," J. Comput. Biol. **14**, 1115–1143 (2007).

[27] G. Guigas and M. Weiss, "Sampling the cell with anomalous diffusion—The discovery of slowness," Biophys. J. **94**, 90–94 (2008).

[28] S. Basak, S. Sengupta, and K. Chattopadhyay, "Understanding biochemical processes in the presence of sub-diffusive behavior of biomolecules in solution and living cells," Biophys. Rev. **11**, 851–872 (2019).

[29] M. Woringer and X. Darzacq, "Protein motion in the nucleus: From anomalous diffusion to weak interactions," Biochem. Soc. Trans. **46**, 945–956 (2018).

[30]V. Sposini, A. V. Chechkin, F. Seno, G. Pagnini, and R. Metzler, "Random diffusivity from stochastic equations: Comparison of two models of Brownian yet non-Gaussian diffusion," New J. Phys. **20**, 043044 (2018).

[31]M. Weiss, H. Hashimoto, and T. Nilsson, "Anomalous protein diffusion in living cells as seen by fluorescence correlation spectroscopy," Biophys. J. **84**, 4043–4052 (2003).

[32]M. Weiss, M. Elsner, F. Kartberg, and T. Nilsson, "Anomalous subdiffusion is a measure for cytoplasmic crowding in living cell," Biophys. J. **87**, 3518–3524 (2004).

[33]J. Szymanski and M. Weiss, "Elucidating the origin of anomalous diffusion in crowded fluids," Phys. Rev. Lett. **103**, 038102 (2009).

[34]M. Weiss and T. Nilsson, "In a mirror dimly: Tracing the movements of molecules in living cells," Trends Cell Biol. **14**, 267–273 (2004).

[35]S. Gupta, R. Biehl, C. Sill, J. Allgaier, M. Sharp, M. Ohl, and D. Richter, "Protein entrapment in polymeric mesh: Diffusion in crowded environment with fast process on short scales," Macromolecules **49**, 1941–1949 (2016).

[36]E. Barkai, R. Metzler, and J. Klafter, "From continuous time random walks to fractional Fokker-Planck equation," Phys. Rev. E **61**, 132–138 (2000).

[37]P. R. Shorten and J. Sneyd, "A mathematical analysis of obstructed diffusion within skeletal muscle," Biophys. J. **96**, 4764–4778 (2009).

[38]F. Mainardi, *Fractional Calculus and Waves in Linear Viscoelasticity. An Introduction to Mathematical Models* (Imperial College Press, 2010).

[39]Y. Cao, Y. Li, W. Ren, and Y. Chen, "Distributed coordination algorithms for multiple fractional-order systems," in *2008 47th IEEE Conference on Decision and Control* (IEEE, 2008), pp. 2920–2925.

[40]Y. Cao, Y. Li, W. Ren, and Y. Chen, "Distributed coordination of networked fractional-order systems," IEEE Trans. Syst. Man Cybern. B **40**, 362–370 (2009).

[41]J. Lu, J. Shen, J. Cao, and J. Kurths, "Consensus of networked multi-agent systems with delays and fractional-order dynamics," in *Consensus and Synchronization in Complex Networks* (Springer, Berlin, 2013), pp. 69–110.

[42]R. Metzler and J. Klafter, "The random walk's guide to anomalous diffusion: A fractional dynamics approach," Phys. Rep. **339**, 1–77 (2000).

[43]A. P. Riascos and J. L. Mateos, "Long-range navigation on complex networks using Lévy random walks," Phys. Rev. E **86**, 056110 (2012).

[44]A. P. Riascos and J. L. Mateos, "Fractional dynamics on networks: Emergence of anomalous diffusion and Lévy flights," Phys. Rev. E **90**, 032809 (2014).

[45]T. Michelitsch, A. P. Riascos, B. Collet, A. Nowakowski, and F. Nicolleau, *Fractional Dynamics on Networks and Lattices* (John Wiley and Sons, Inc., 2019).

[46]M. Benzi, D. Bertaccini, F. Durastante, and I. Simunec, "Nonlocal network dynamics via fractional graph Laplacians," J. Compl. Net. (in press) (2020).

[47]R. R. Coifman and S. Lafon, "Diffusion maps," Appl. Comput. Harm. Anal. **21**, 5–30 (2006).

[48]J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, and B. Foster, "The genotype-tissue expression (GTEx) project," Nat. Genet. **45**, 580–585 (2013).

[49]M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. Al-Khalili Szigyarto, J. Odeberg, D. Djureinovic, J. Ottosson Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Pontén, "Tissue-based map of the human proteome," Science **347**, 1260419 (2015).

[50]The FANTOM Consortium and the RIKEN PMI and CLST (DGT), "A promoter-level mammalian expression atlas," Nature **507**, 462–470 (2014).

[51]W. Sun, Y. Li, C. Li, and Y. Chen, "Convergence speed of a fractional order consensus algorithm over undirected scale-free networks," Asian J. Control **13**, 936–46 (2011).

[52]J. Piñero, J. M. Ramírez-Anguita, J. Saüch-Pitarch, F. Ronzano, E. Centeno, F. Sanz, and L. I. Furlong, "The DisGeNET knowledge platform for disease genomics: 2019 update," Nucleic Acids Res. **48**, D845–D855 (2020).

[53]A. Akhmerov and E. Marbán, "COVID-19 and the heart," Circulation Res. **126**, 1443–1455 (2020).

[54]R. M. Inciardi, L. Lupi, G. Zaccone, L. Italia, M. Raffo, D. Tomasoni, D. S. Cani, M. Cerini, D. Farina, E. Gavazzi, and R. Maroldi, "Cardiac involvement in a patient with coronavirus disease 2019 (COVID-19)," JAMA Cardiol. **5**(7), 819–824 (2020).

[55]Y. Zheng, Y. Ma, J. Zhang, and X. Xie, "COVID-19 and the cardiovascular system," Nat. Rev. Cardiol. **17**, 259–260 (2020).

[56]K. J. Clerkin, J. A. Fried, J. Raikhelkar, G. Sayer, J. M. Griffin, A. Masoumi, S. S. Jain, D. Burkhoff, D. Kumaraiah, L. Rabbani, and A. Schwartz, "Coronavirus disease 2019 (COVID-19) and cardiovascular disease," Circulation **141**, 1648–1655 (2020).

[57]W. D. Maya, S. S. Du Plessis, and P. A. Velilla, "SARS-CoV-2 and the testis: Similarity to other viruses and routes of infection," Reproduct. BioMed. Online **40**, 763–764 (2020).

[58]F. Chen and D. Lou, "Rising concern on damaged testis of COVID-19 patients," Urology (published online 2020).

[59]S. Wang, X. Zhou, T. Zhang, and Z. Wang, "The need for urogenital tract monitoring in COVID-19," Nat. Rev. Urol. **17**, 314–315 (2020).

[60]C. Fan, K. Li, Y. Ding, W. L. Lu, and J. Wang, "ACE2 expression in kidney and testis may cause kidney and testis damage after 2019-nCoV infection," medRxiv (2020).

[61]J. M. Saavedra, "COVID-19, angiotensin receptor blockers, and the brain," Cell. Mol. Neurobiol. **40**, 667–674 (2020).

[62]N. Lushina, J. S. Kuo, and H. A. Shaikh, "Pulmonary, cerebral, and renal thromboembolic disease associated with COVID-19 infection," Radiology (published online 2020).

[63]B. Zhou, J. She, Y. Wang, and X. Ma, "A case of coronavirus disease 2019 with concomitant acute cerebral infarction and deep vein thrombosis," Front. Neurol. **11**, 296 (2020).

[64]C. Ronco and T. Reis, "Kidney involvement in COVID-19 and rationale for extracorporeal therapies," Nat. Rev. Nephrol. **16**, 308–310 (2020).

[65]V. Fanelli, M. Fiorentino, V. Cantaluppi, L. Gesualdo, G. Stallone, C. Ronco, and G. Castellano, "Acute kidney injury in SARS-CoV-2 infected patients," Critical Care **24**, 155 (2020).

[66]Z. Li, M. Wu, J. Yao, J. Guo, X. Liao, S. Song, J. Li, G. Duan, Y. Zhou, X. Wu, and Z. Zhou, "Caution on kidney dysfunctions of COVID-19 patients," medRxiv (2020).

[67]A. Ruggiero, G. Attinà, and A. Chiaretti, "Additional hypotheses about why COVID-19 is milder in children than adults," Acta Paediatr. **69**, 1690 (2020).

[68]Z. Feng, B. Diao, R. Wang, G. Wang, C. Wang, Y. Tan, L. Liu, C. Wang, Y. Liu, Y. Liu, and Z. Yuan, "The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) directly decimates human spleens and lymph nodes," medRxiv (2020).

[69]R. Pal, "COVID-19, hypothalamo-pituitary-adrenal axis and clinical implications," Endocrine **68**, 251–252 (2020).

[70]E. Estrada, G. Estrada-Rodriguez, and H. Gimperlin, "Metaplex networks: Influence of the exo-endo structure of complex systems on diffusion," SIAM Rev. (in press).

[71]E. Estrada, "Path Laplacian matrices: Introduction and application to the analysis of consensus in networks," Lin. Algebra Appl. **436**, 3373–3391 (2012).

[72]E. Estrada, E. Hameed, N. Hatano, and M. Langer, "Path Laplacian operators and superdiffusive processes on graphs. I. One-dimensional case," Lin. Algebra Appl. **523**, 307–334 (2017).

[73]E. Estrada, E. Hameed, M. Langer, and A. Puchalska, "Path Laplacian operators and superdiffusive processes on graphs. II. Two-dimensional lattice," Lin. Algebra Appl. **555**, 373–397 (2018).

[74]M. Asllani, T. Carletti, F. Di Patti, D. Fanelli, and F. Piazza, "Hopping in the crowd to unveil network topology," Phys. Rev. Lett. **120**, 158301 (2018).

[75]M. Qiu, D. Wu, W. Ning, J. Zhang, T. Shu, C. Huang, R. Chen, M. Huang, J. Xu, Q. Yang, and R. Li, Postmortem tissue proteomics reveals the pathogenesis of multiorgan injuries of COVID-19 (2020).