View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Universidad Carlos III de Madrid e-Archivo

Generalized optimal paths and weight distributions revealed through the large deviations of random walks on networks

Ricardo Gutiérrez [1] and Carlos Pérez-Espigares [2,3]

[1] *Complex Systems Interdisciplinary Group, Department of Mathematics, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain*
[2] *Departamento de Electromagnetismo y Física de la Materia, Universidad de Granada, Granada 18071, Spain*
[3] *Institute Carlos I for Theoretical and Computational Physics, Universidad de Granada, Granada 18071, Spain*

(Received 9 November 2020; accepted 3 February 2021; published 24 February 2021)

Numerous problems of both theoretical and practical interest are related to finding shortest (or otherwise optimal) paths in networks, frequently in the presence of some obstacles or constraints. A somewhat related class of problems focuses on finding optimal distributions of weights which, for a given connection topology, maximize some kind of flow or minimize a given cost function. We show that both sets of problems can be approached through an analysis of the large-deviation functions of random walks. Specifically, a study of ensembles of trajectories allows us to find optimal paths, or design optimal weighted networks, by means of an auxiliary stochastic process (the generalized Doob transform). The paths are not limited to shortest paths, and the weights must not necessarily optimize a given function. Paths and weights can in fact be tailored to a given statistics of a time-integrated observable, which may be an activity or current, or local functions marking the passing of the random walker through a given node or link. We illustrate this idea with an exploration of optimal paths in the presence of obstacles, and networks that optimize flows under constraints on local observables.

DOI: 10.1103/PhysRevE.103.022319

I. INTRODUCTION

Shortest paths are one of the main objects of interest in network science, an interest driven by applications in transportation networks [1–3], the internet [4–8], and protein-protein interaction networks [9,10], among other systems. They also enter into the definition of key structural properties such as the closeness, the efficiency, or the betweenness centrality [11,12]. The definition of a shortest path between two nodes can be generalized to include, for example, constraints or obstacles (see, e.g., [13–16]) or to reach several targets from one or several sources (see, e.g., [17,18]). A great deal of work in the applied discrete mathematics, theoretical computer science, and statistical physics communities has dealt with the solutions of such problems, and the computational complexity of the algorithms involved to find them.

Generally speaking, this research addresses problems such as how to find the shortest path between two nodes always passing, or without ever passing, through another one, but does not consider situations where a node is visited a fraction of times. Moreover, these methods cannot be easily extended to address problems where a given path observable is required to take a specific (not necessarily extremal) value, which may be characterized in a statistical sense. For example, it is not clear how one goes about finding the path or paths that maximize the fluctuations (e.g., the variability across realizations of a stochastic process) of a given current, or the path that guarantees that a certain node is visited on average twice as frequently as another node. In more practical terms, what is the shortest path that a passenger or a data packet can take without saturating a given node or link? Or what is the choice that maximizes the number of paths taken without the overall path length exceeding a certain threshold? These can also be considered optimal paths in a generalized sense, as they are chosen to ensure that the statistics of a given observable for a particle moving across the network takes specific values or does not exceed certain bounds.

Similarly, one could think of redistributing the link weights of a network, characterized by a given adjacency matrix, so as to ensure that a set of nodes is visited with some frequency in the resulting weighted network, or that no link carries more than a certain amount of some flow. While this problem is less intensively investigated than that of optimal paths, examples can be found in the literature of, e.g., networks that are optimal for sustaining a synchronized dynamics [19–22]. Recent contributions based on a kindred dynamical-control approach show how to engineer force fields in many-particle systems so as to achieve prescribed steady-state distributions [23–25].

In this paper we propose a theoretical approach to unveil such generalized optimal paths and weight distributions by studying the statistics of trajectories using large-deviation methods [26–28]. Specifically, we analyze the large deviations of random walks on graphs [29,30]. This allows us to find paths that are optimal in the statistical sense outlined above, or weight distributions that make a network optimal for a given statistical characterization pertaining to the flow of information or physical entities. By biasing the dynamics with certain observables we obtain the random-walk stationary distribution that guarantees that such observables (e.g., the activity associated with a node or link, or a current in a specific direction in a spatial network) satisfy some statistical constraint, which may be related to its mean value,

 ©2021 American Physical Society

fluctuations, or higher-order cumulants. We then employ the auxiliary process given by the generalized Doob transform [31–34], to find the transition probabilities which give rise to the long-time statistics of the above-mentioned observable. By combining the biased stationary state with the generalized Doob transform, we extract the probability fluxes of the biased walk, which highlight the existence of optimal paths. Furthermore, the Doob-transformed process itself yields an optimal distribution of link weights in the same generalized sense.

We illustrate this versatile approach by finding generalized optimal paths in random graphs in the presence of constraints. To this end, we study the trajectories of the maximal entropy random walk (MERW) [35] in an appropriate trajectory ensemble using as observable local activities. We also study constrained optimal weight distributions for maximal current or flow on spatial networks by means of the standard random walk (SRW) [36]. These choices have been made as each is uniquely well suited to the problem under study. Once the choice of the appropriate process, observables, and trajectory ensemble is made for a given problem involving generalized optimal paths or weight distributions, the solution can be found by a few steps of linear algebra.

While the large-deviation approach lies at the basis of equilibrium statistical mechanics, and is regarded as the natural language for dealing with many problems in nonequilibrium statistical physics [28], its application to the study of networks has started only recently. We conclude this introduction with a brief summary of some significant recent research. In the first work, as far as we are aware, on large deviations of time-integrated observables of random walks on networks [29], localization and mode-switching dynamical phase transitions are revealed. More recently, in a contribution that has strongly influenced our methodology [30], such localization phenomena are explained by means of the generalized Doob transform, which is also used to shed new light on the relationship between the maximum entropy random walk and the standard random walk. Various other processes have been explored with related methodologies in publications dealing with, e.g., percolation transitions in single or multilayer networks subject to rare initial configurations, [37,38], paths leading to epidemic extinction [39,40], the connection between the rate of rare events and heterogeneity in population networks [41], or large-fluctuation-induced phase switch in majority-vote models [42]. Additionally, large-deviation and rare-event techniques have been employed in the exploration of structural properties, such as the assortativity in configuration-model networks [43], the study of ensembles of random graphs satisfying structural constraints [44,45], and the existence of a first-order condensation transition in the node degrees [46].

II. THERMODYNAMICS OF TRAJECTORIES OF RANDOM WALKS

Random walks have been studied in continuous media and discrete spaces. Among the latter, much recent work has been devoted to the study of random walks on networks [47,48]. Here we consider two different types of discrete-time random walks on networks: the SRW [36] and the MERW [35]. Both are discrete-time Markov chains the components of the probability vector of which $\boldsymbol{p}$ evolve in time as $p_j(n+1) = \sum_i \Pi_{ji}\, p_i(n)$, where the non-negative integer n is the time step, $p_i(n)$ is the probability that a random walker visits node i at time n, and Π_{ji} is the probability of a transition to j conditioned on the node being in i. As usual, $\sum_i p_i(n) = 1$ for all n, and the probabilities of all possible transitions from a given node add up to 1, $\sum_j \Pi_{ji} = 1$.

The SRW is suitable for the study of flow on networks—currents, fluid flow, goods, etc. [49]—as it considers that a particle in a node can jump to any of its neighbors with the same probability (in the case of unweighted networks) or with probabilities proportional to the link weights (for weighted networks). Given an unweighted network—this will be our starting point, though the generalization to weighted networks does not pose any difficulty—with directed adjacency matrix $\mathbf{A}$, where $A_{ji} = 1$ if there is a link pointing from i to j and is zero otherwise, the entries of the transition matrix $\mathbf{\Pi}$ are

$$\Pi_{ji}^{\mathrm{SRW}} = \frac{A_{ji}}{k_i^{\mathrm{out}}}. \tag{1}$$

The normalization by the out degree, which is defined as the number of neighbors joined by outgoing links, $k_i^{\mathrm{out}} = \sum_j A_{ji}$, ensures the conservation of probability.

The MERW, on the other hand, is most suitable for the exploration of generalized optimal paths in networks, as it assigns the same probability to all trajectories joining two given nodes that comprise the same number of steps (which the SRW does not do, see Appendix A). The transition matrix is given by

$$\Pi_{ji}^{\mathrm{MERW}} = \frac{A_{ji}}{\lambda}\frac{v_j}{v_i}, \tag{2}$$

where λ is the largest eigenvalue of the directed adjacency matrix, and $\boldsymbol{v}$ is the normalized eigenvector associated with it, $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$. For a more detailed discussion of the SRW and the MERW, as well as references treating other types of random walks on networks, see Appendix A.

For either type of random walk, a trajectory up to time τ, ω_τ, is given by the sequence of nodes visited at each step, $\omega_\tau = (i_\tau \leftarrow \cdots \leftarrow i_2 \leftarrow i_1 \leftarrow i_0)$. We consider time-extensive observables of the form $\hat{O}(\omega_\tau) = \sum_{n=1}^{\tau} \hat{o}(i_n \leftarrow i_{n-1})$, with $\hat{o}(i_n \leftarrow i_{n-1})$ being the increment of the observable at a time step, the value of which depends on the nodes i_{n-1} and i_n, which are joined by a link. As the probability assigned to the trajectory is $P(\omega_\tau) = \Pi_{i_\tau i_{\tau-1}} \cdots \Pi_{i_2 i_1} \Pi_{i_1 i_0} p_{i_0}(0)$, the probability distribution of the observable is $P_\tau(O) = \sum_{\omega_\tau} \delta[O - \hat{O}(\omega_\tau)] P(\omega_\tau)$. This distribution corresponds to an ensemble of trajectories with fixed observable O and fixed time τ. In the long-time limit, it adopts a large-deviation form $P_\tau(O) \sim e^{-\tau I(O/\tau)}$, which is here given in terms of the time-intensive observable O/τ (see Appendix B for details). The function $I(O/\tau)$ is called the rate function, which plays the role of a dynamical entropy. This ensemble of trajectories is analogous to the microcanonical ensemble of equilibrium statistical mechanics, and it is generally speaking not the most useful one to work with. Fortunately, the thermodynamic formalism of time-integrated dynamical observables developed in [26,27] shows how to study the statistics of O in more suitable ensembles.

By biasing trajectories with a parameter s—which we refer to as the *tilting parameter*—we obtain the s ensemble, where τ is fixed but O is now a fluctuating observable the probability distribution of which is given by $P^s_\tau(O) = Z^{-1}_\tau(s)\, e^{-sO} P_\tau(O)$. The normalization factor is a dynamical partition function which also acquires a large-deviation form for large τ, $Z_\tau(s) = \sum_O e^{-sO} P_\tau(O) \sim e^{\tau\theta(s)}$, where $\theta(s)$ is the so-called scaled-cumulant generating function (SCGF), and is related to the rate function by a Legendre-Fenchel transform [28]. Its role is that of a dynamical free energy, and in fact from its derivatives one can obtain the cumulants of the time-intensive observable O/τ (see Appendix B). Crucially, the partition function can be expressed as $Z_\tau(s) = \sum_{i,j}[(\mathbf{\Pi}^s)^\tau]_{ji} p_i(0)$, where $\mathbf{\Pi}^s$ is the *tilted operator*, which is akin to a transfer matrix, the elements of which are

$$\Pi^s_{ji} = e^{-s\hat{o}(j\leftarrow i)}\Pi_{ji}. \tag{3}$$

Thus, finding the SCGF $\theta(s)$—which fully characterizes the statistics of O—reduces, in the long-time limit, to an eigenvalue problem for the tilted operator (3). Actually, by spectral decomposition it is straightforward to check that, for long times, $\theta(s)$ is given by the logarithm of the largest eigenvalue of $\mathbf{\Pi}^s$. While the derivatives of $\theta(s)$ at $s = 0$ correspond to the cumulants of O in the typical (unbiased) distribution, $P_\tau(O)$, those evaluated at $s \neq 0$ provide information about the statistics of O in the tilted (biased) one, $P^s_\tau(O)$. For $s \neq 0$ such statistics is carried by the so-called rare trajectories (as they are exponentially unlikely), which are not, however, easily retrieved from the tilted operator (3), as it is not a stochastic matrix ($\sum_j \Pi^s_{ji} \neq 1$). Nevertheless, as we later explain, this unphysical tilted operator can be transformed into a proper stochastic matrix by means of the generalized Doob transform, revealing the rare trajectories of interest. The latter arise from the transition probabilities leading to the fluctuation conjugate to the tilting parameter s.

If instead of keeping fixed the duration of a trajectory τ, we fix the value that the observable O reaches for each trajectory—thus allowing τ to fluctuate—, we obtain a different statistical ensemble, namely, the x ensemble [50]. In this ensemble, we shall consider as observable O a local activity, namely, the number of times a given link (or set of links) is traversed in a trajectory, having $\hat{o}(j \leftarrow i) = 1$ whenever this occurs and zero otherwise. We denote by $P(y_O)$ the probability of a trajectory $y_O = (i_\tau \leftarrow \cdots \leftarrow i_2 \leftarrow i_1 \leftarrow i_0)$ that reaches a fixed value O in a number of steps τ. As the latter fluctuates from trajectory to trajectory, the probability distribution of the time duration for fixed O is $P_O(\tau) = \sum_{y_O} \delta[\tau - \hat{\tau}(y_O)]P(y_O)$, where the operator $\hat{\tau}$ counts the number of time steps in a trajectory. In this case, the dynamical partition function conditioned on a fixed value of O reads

$$Z_O(x) = \sum_\tau e^{-x\tau} P_O(\tau) = \sum_{i,j}[(\mathbf{\Pi}^x)^O]_{ji} p_i(0), \tag{4}$$

which again we write in terms of a tilted operator, namely,

$$\mathbf{\Pi}^x = \mathbf{\Pi}_O (e^x - \tilde{\mathbf{\Pi}})^{-1}. \tag{5}$$

Here, $\mathbf{\Pi}_O$ is a matrix which preserves the transition probabilities of $\mathbf{\Pi}$ only for those transitions (links) contributing to O, i.e., $j \leftarrow i$ such that $\hat{o}(j \leftarrow i) \neq 0$, the rest of its entries being zero, and $\tilde{\mathbf{\Pi}} = \mathbf{\Pi} - \mathbf{\Pi}_O$. For large O (which also corresponds to large τ), the grand-partition function acquires a large-deviation form $Z_O(x) \sim e^{O\varphi(x)}$, where $e^{\varphi(x)}$ is the largest eigenvalue of $\mathbf{\Pi}^x$. The cumulants of the fluctuating time between observable O updates, τ/O, can then be obtained from the derivatives of the x-ensemble SCGF $\varphi(x)$ (see Appendix B).

While the s ensemble and the x ensemble are equivalent in the $\tau \to \infty$ (hence $O \to \infty$, as O is time extensive) limit [50,51], the former is more natural for the study of time-averaged observables of the form O/τ, and the latter is more appropriate for the analysis of their reciprocal, τ/O. In the following, we will use one or the other depending on the specific problem under study.

In fact, we will also consider ensembles of biased trajectories with two different tilting parameters. The latter will be conjugate to two fluctuating time-extensive observables, or a fluctuating observable and the duration of the trajectory τ. A description of such ss and sx ensembles, as they are, respectively, referred to, as well as a more detailed characterization of the ensembles presented above can be found in Appendix B. In the ss and sx ensembles it will also be possible to obtain the relevant SCGFs by computing the largest eigenvalues of certain transfer operators, which are extensions of those given in Eqs. (3) and (5).

All such tilted operators, as explained above for the s ensemble and regardless of the ensemble under consideration, have something in common: they are not stochastic operators. However, by an application of the generalized Doob transform [31–34]—which for convenience we will just refer to as the Doob transform in the remainder of this paper—one obtains an auxiliary process the statistics of which is given by $P^s_\tau(O)$ in the long-time limit. The Doob transform of a tilted operator is discussed in Appendix C and references therein. The transformed transition matrix $\mathbf{\Pi}_{\text{Doob}}$, which unlike the tilted operator is a proper stochastic matrix, gives us the set of transition probabilities that characterize the auxiliary process. By multiplying these with the corresponding stationary probabilities $\boldsymbol{p}^{\text{st}}$, which satisfy $\mathbf{\Pi}_{\text{Doob}}\boldsymbol{p}^{\text{st}} = \boldsymbol{p}^{\text{st}}$, we obtain the probability fluxes $(\Pi_{\text{Doob}})_{ji} p^{\text{st}}_i$, which are the joint probabilities of being at node i at a given time step, and moving to node j at the next one. While the transition matrix $\mathbf{\Pi}_{\text{Doob}}$ will give us the optimal link weights in order to sustain a given statistics, the probability fluxes will visually reveal most clearly the network that results from imposing such statistics on the observables of interest, and highlight the optimal paths in it.

III. OPTIMAL PATHS FROM LARGE DEVIATIONS OF MAXIMAL ENTROPY RANDOM WALKS

A. Biased trajectories in directed rings

We first illustrate the idea of searching for optimal paths with a simple example. Consider the ring with a shortcut shown in Fig. 1(a). A particle starts from the node which has been highlighted in red, and then hops counterclockwise to the neighboring node, and then hops again always following the MERW transition probabilities. These are all identically 1, except when the particle reaches the node from which the shortcut starts, where it may continue along the ring with

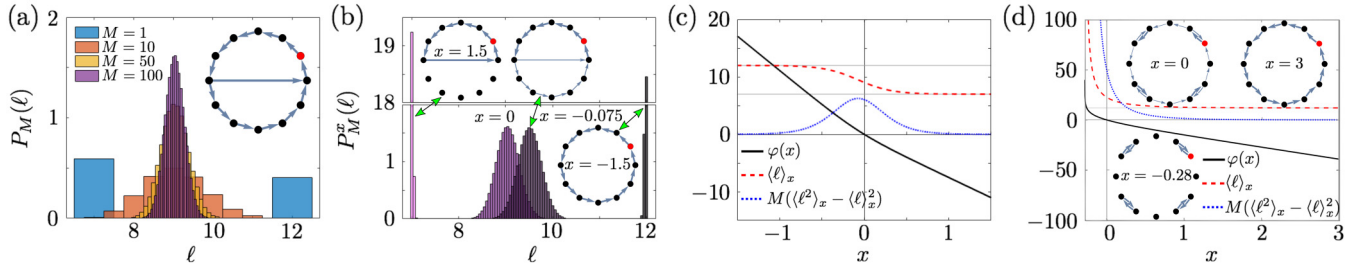

FIG. 1. Finding paths that extremize the cycle length or maximize its fluctuations in simple graphs. (a) Probability distribution of the sample mean of the cycle length ℓ (see text for the definition) of M random walk trajectories on the graph depicted on the upper right corner, starting and ending at the node highlighted in red, for $M = 1, 10, 50, 100$. The distribution follows a large deviation principle $P_M(\ell) \sim e^{-MI(\ell)}$ for large M, where only small fluctuations around the mean are Gaussian, since $I(\ell)$ is not quadratic for large fluctuations. (b) Original ℓ distribution ($x = 0$) for $M = 100$, and three tilted distributions: while $x = -0.075$ maximizes the fluctuations, $x = -1.5$ and 1.5 peak around the longest and shortest possible lengths, respectively. The graphs display the probability fluxes as given by the Doob transform (arrow thickness is proportional to the probability flux). Probability fluxes smaller than the largest value divided by 200 are not displayed for visibility reasons. (c) SCGF $\varphi(x)$ and path length mean $\langle \ell \rangle_x = -\varphi'(x)$ and (scaled) fluctuations $M(\langle \ell^2 \rangle_x - \langle \ell \rangle_x^2) = \varphi''(x)$ for large M. (d) SCGF $\varphi(x)$, $\langle \ell \rangle_x$, and $M(\langle \ell^2 \rangle_x - \langle \ell \rangle_x^2)$, and probability fluxes for $x = 3$ (shortest path) and $x = -0.28$ (localized walker), corresponding to the graph depicted in this panel for $x = 0$.

probability $p \approx 0.41$ or it may take the shortcut with probability $q = 1 - p \approx 0.59$.

We will be interested in the statistics of the length (number of time steps) of a cyclic walk that starts and ends in the red node. If the cycle is performed M times, we consider the probability distribution of the sample mean $\ell = M^{-1} \sum_{i=1}^{M} \ell^{(i)}$, where $\ell^{(i)}$ is the cycle length of a given realization. For a single cycle ($M = 1$), the walker can either follow along the ring with probability p, which gives $\ell = \ell_p = 12$, or take the shortcut with probability q, for which $\ell = \ell_q = 7$ [see Fig. 1(a)]. As M grows, ℓ takes more and more values, and for sufficiently large M, the distribution centers around the mean $p\,\ell_p + q\,\ell_q \approx 9.04$ following a large deviation principle $P_M(\ell) \sim e^{-MI(\ell)}$, with fluctuations that are approximately Gaussian only around the mean, i.e., $|\ell - \langle \ell \rangle| \sim \mathcal{O}(1/\sqrt{M})$, as expected from the law of large numbers and the central limit theorem. Fluctuations that deviate far from the average are, however, not Gaussian and they are the prime concern of large-deviation theory.

In order to unveil optimal paths, we shall focus on the probability of large deviations of ℓ. These are studied in the x ensemble, since we are interested in the fluctuations of the length (number of time steps) for a given number of cycles. The duration of the trajectory ℓM is thus fluctuating (this would correspond to τ in Sec. II) and the number of realizations M is fixed (this would correspond to O in Sec. II). The latter condition can be achieved by fixing to M the activity through the link that reaches the red node from the preceding node in the ring—this is the local observable. We thus take the probability distribution $P_M(\ell)$, corresponding to $M = 100$ cycles, and bias it using x as tilting parameter:

$$P_M^x(\ell) = e^{-xM\ell} P_M(\ell)/Z_M(x). \tag{6}$$

Here $P_M(\ell) = P_M^{x=0}(\ell)$, $Z_M(x)$ is a normalizing factor, and the large-deviation regime corresponds to large values of M. In Fig. 1(b), we show such tilted probability distribution for $M = 100$ and different values of x, namely, $x = -1.5$, -0.075, and 1.5—the untilted case $x = 0$, which was already shown in Fig. 1(a), is also included for comparison. These choices of x correspond to the largest ($x = -1.5$) and smallest ($x = 1.5$) possible $\langle \ell \rangle_x$, which are of course ℓ_p and ℓ_q, respectively, and to the value ($x = -0.075$) that maximizes the fluctuations $\langle \ell^2 \rangle_x - \langle \ell \rangle_x^2$. This information is obtained from the SCGF $\varphi(x)$ of the x ensemble of MERW trajectories, which is shown, together with its first and second derivatives, in Fig. 1(c). Such derivatives (with the appropriate signs) correspond to the mean and the (scaled) fluctuations of ℓ: $\langle \ell \rangle_x = -\varphi'(x)$ and $M(\langle \ell^2 \rangle_x - \langle \ell \rangle_x^2) = \varphi''(x)$.

In this simple directed ring, the analytical expression for the SCGF $\varphi(x)$ can be readily obtained from the large deviations of ℓ, so one does not need to compute the largest eigenvalue of the tilted operator in Eq. (5). We briefly review the main steps of the calculation—a detailed derivation can be found in Appendix D. As each $\ell^{(i)}$ is a Bernoulli trial which takes the values ℓ_p and ℓ_q with probabilities p and $q = 1 - p$, the random variable n_p, quantifying the fraction of times that the path of length ℓ_p is taken, has a binomial probability distribution. Its large-deviation form can be obtained through an application of Stirling's approximation. By a change of variable we find the distribution of $\ell = n_p \ell_p + (1 - n_p)\ell_q$, which also acquires a large-deviation form $P_M(\ell) \sim e^{-MI(\ell)}$. From the rate function $I(\ell)$, the SCGF $\varphi(x) = \lim_{M\to\infty} \log[Z_M(x)]/M$ is obtained via a Legendre transform $\varphi(x) = -\min_\ell [x\ell + I(\ell)]$, yielding

$$\varphi(x) = -x\ell_q + \log[pe^{-x(\ell_p - \ell_q)} + (1 - p)]. \tag{7}$$

The analytical expression of its first derivative shows that the average $\langle \ell \rangle_x$ is bounded between ℓ_q and ℓ_p, and approaches those bounds for large tilting-parameter (absolute) values. The asymptotic value ℓ_q is reached for positive x, while ℓ_p corresponds to negative x, as expected for a sufficiently strong bias towards shorter or longer cycles. In Fig. 1(c), for the values p, ℓ_p, and ℓ_q under consideration, such extreme values are already practically reached for $|x| \approx 1$, but this value will change if the shortcut is located elsewhere (see Appendix D).

The probability fluxes corresponding to those same values for which we show the tilted distributions, namely, $x = -1.5$, -0.075, and 1.5, are also displayed in Fig. 1(b). They highlight the optimal paths in each of the three situations considered. The extreme values of x correspond to a walk that just

moves along the ring (for negative x, which favors long paths, $\langle \ell \rangle_x \approx \ell_p$) or takes the shortcut (for positive x, which favors short paths, $\langle \ell \rangle_x \approx \ell_q$). On the other hand, the maximization of the fluctuations leads to the shortcut being taken or avoided with probability $1/2$. While one can calculate the statistics of ℓ for all x without resorting to the x-ensemble tilted generator (5)—at least for very simple systems the SCGF $\varphi(x)$ of which can be found analytically—the eigenvectors of such operator are still needed to compute the Doob transform, and therefore to obtain the probability fluxes (see Appendix C). Moreover, the distributions for $|x| = 1.5$ displayed in Fig. 1(b) are also based on the Doob-transformed process, as this allows us to circumvent the challenging task of numerically performing very strong tiltings, which result in distributions lying on regions where $P_M(\ell)$ is negligibly small—see the discussion of an analogous issue in the last section of [52].

While in the illustrative example we have just considered the resulting optimal paths are trivial, in more complex topologies our approach unveils paths and weight distributions the adequacy of which for sustaining given observable statistics is far from obvious. But before moving on to such more interesting examples, let us illustrate the point with another simple case, namely, a ring with alternate bidirectional links [see Fig. 1(d)], where the original probability fluxes are shown for $x = 0$. This allows us to discuss localization phenomena and diverging times, which will also appear later. In this case we compute the SCGF from the corresponding x-ensemble operator, and obtain from its derivatives the mean cycle length $\langle \ell \rangle_x$ and the (scaled) fluctuations $M(\langle \ell^2 \rangle_x - \langle \ell \rangle_x^2)$. For $x = 3$ the probability fluxes show the shortest possible path, as expected, which moves counterclockwise along the ring with vanishingly small fluctuations. For negative x, however, there is a vertical asymptote in the SCGF, hence the cumulants diverge. The probability fluxes for a value of x sufficiently close to the divergence show a situation where the particle becomes localized and never reaches the red node from its clockwise neighbor—see the trajectories for $x = -0.28$ in Fig. 1(d). Quite appropriately, $\langle \ell \rangle_x$ grows unboundedly as x approaches the divergence and the particle becomes more and more localized. The mathematical origin of divergences in the x ensemble is discussed in Appendix B.

B. Constrained optimal paths in random graphs

A more interesting case is considered next, namely, that of finding optimal paths in the presence of constraints in random graphs. To provide a concrete illustration, we consider a random graph of $N = 20$ nodes with $3N$ directed links distributed uniformly at random among them [see Figs. 2(c) and 2(d)—the links highlighted in colors others than gray will be discussed below]. Much larger networks with Poissonian or power-law degree distributions or networks arising from applications that do not correspond to a precise mathematical model could be similarly studied.

In this case, a particle repeatedly performs a MERW from the source node 1 and reaches the target node 20 after ℓ steps. Again, we study the statistics of ℓ, but this time we also consider whether the particle goes through an obstacle, node 15, before reaching the target node. Here, "obstacle" is used in a loose sense to indicate that some constraint, based on how

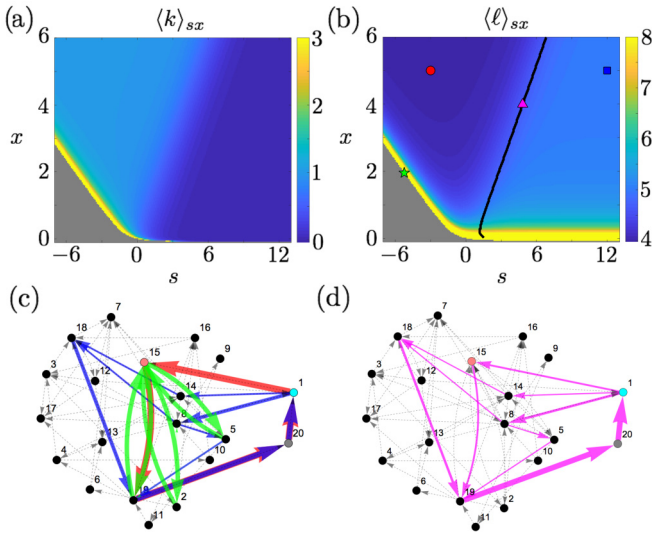

FIG. 2. Finding optimal paths in random graphs in the presence of constraints. (a) Average length $\langle l \rangle_{sx}$ of path joining nodes 1 (source) and 20 (target) in the graph shown in panels (c) and (d) as a function of the tilting parameters s and x. (b) Average activity $\langle k \rangle_{sx}$, i.e., number of times that node 15 is visited before the target is reached, as a function of the tilting parameters s and x. Highlighted points correspond to probability fluxes shown in the panels below. The black segment shows the contour line for $\langle k \rangle_{sx} = 1/3$. (c) Probability fluxes obtained from the Doob-transformed process corresponding to the red circle, the green star, and the blue square points shown in panel (b). (d) Probability fluxes obtained from the Doob-transformed process of the magenta triangle in panel (b). Probability fluxes smaller than the largest value divided by 200 are not displayed for visibility reasons.

frequently the walker goes through that node, will be applied. To this end, we employ a second observable, namely, k, which gives the activity of the obstacle, defined as the number of times the particle goes through the obstacle before reaching the target node (see [53] for a similar observable in a diffusion process). The obstacle can be completely avoided, or it can just be avoided a fraction of the times that the target node is reached, or the particle may even be biased to visit it more frequently than in the natural dynamics—it all depends on the value of the tilting parameter s conjugate to k (see below).

Before discussing the results of our analysis, we mention a technical point which may be relevant in applications. The large-deviation analysis allows us to access the average, fluctuations, and higher cumulants of k and ℓ, including correlations between them, for different tilting-parameter values. This is made possible by the existence of a link connecting node 20 with node 1, which restarts the process when the target node is reached and guarantees the time-extensivity of integrated observables (see Appendix B for details). In practice, this may already be part of the network or, if not, it should be expressly introduced for the analysis.

We consider an sx ensemble, with tilted probability distribution

$$P_M^{sx}(k, \ell) = e^{-sMk - xM\ell} P_M(k, \ell)/Z_M(s, x) \tag{8}$$

where $P_M(k,\ell) = P_M^{s=0,x=0}(k,\ell)$, $Z_M(s,x)$ is a normalizing factor, and M is assumed to be fixed and large. The fluctuating time is again $M\ell$, and Mk is an M-extensive—hence time-extensive—observable (it corresponds to K in Appendix B). The SCGF $\varphi(s,x)$ is calculated from the largest eigenvalue of the tilted operator (see Appendix B), and from its partial derivatives we obtain the average activity $\langle k\rangle_{sx} = -\partial_s\varphi(s,x)$ and the average path length $\langle \ell\rangle_{sx} = -\partial_x\varphi(s,x)$, which are shown in Figs. 2(a) and 2(b), respectively. The gray area on the lower left-hand corner corresponds to a region where the averages grow extremely rapidly and eventually diverge for reasons analogous to those given above regarding the divergence shown in Fig. 1(d). For $x = 0$ (no tilting is applied on ℓ) the physical meaning is clear: as s becomes negative, the particle goes through the obstacle more and more frequently before reaching the target node, and when s is sufficiently large and negative it never reaches it. For sufficiently large and positive s, on the other hand, the particle completely avoids the obstacle, so we have $\langle k\rangle_{sx} \approx 0$ and a finite ℓ. As x is increased (when the tilting favors shorter walks joining the source node and the target node), the particle goes more rapidly towards the target, and therefore it requires a larger negative value of s to start growing steeply and eventually diverge.

The probability fluxes derived from the Doob transform for the point $(s = -5.24, x = 1.95)$, which lies very close to the divergence and is highlighted with a green star in Fig. 2(b), are shown as green arrows in Fig. 2(c). They clearly display a trajectory where the particle moves back and forth between the obstacle and its neighbors, without ever reaching the target. This and other points discussed below are only highlighted in Fig. 2(b), where the color map shows values of $\langle \ell\rangle_{sx}$, but of course they correspond to the same parameter-space points in Fig. 2(a), where the color map shows values of $\langle k\rangle_{sx}$.

More importantly, a crossover between a region where $\langle k\rangle_{sx} \approx 0$ and a plateau where $\langle k\rangle_{sx} \approx 1$ is observed in Fig. 2(a), which corresponds to a crossover from a region where $\langle \ell\rangle_{sx}$ is larger than 4 to a region where $\langle \ell\rangle_{sx} \approx 4$ in Fig. 2(b). The probability fluxes for the point $(s = -3, x = 5)$, which is highlighted with a red disk in Fig. 2(b) and corresponds to $\langle \ell\rangle_{sx} = 4$, are shown as red arrows in Fig. 2(c). They show a trajectory where the particle moves along the shortest path between the source and the target, and gets back to the source after precisely four steps. As the obstacle lies on this path, we have $\langle k\rangle_{sx} = 1$. On the other side of the crossover and for sufficiently large s and x, for example for $(s = 12, x = 5)$, the walker chooses the shortest amongst the paths that avoid the obstacle [see the blue square in Fig. 2(b) and the blue arrows in Fig. 2(c)].

One may be interested in finding the path or combination of paths that only visit the obstacle with a given frequency while reaching the target node in the smallest possible number of steps. In contrast with the cases discussed above, this type of generalized path is, as far as we are aware, outside the reach of the graph-theoretical methods typically used for finding shortest paths, despite its clear practical interest, e.g., in transportation networks where a given station or airport may have a limited capacity that cannot be exceeded. To this end we set the activity of the obstacle to $\langle k\rangle_{sx} = 1/3$: one in three times when the target is reached the walker has passed through the obstacle. We then obtain the set of points of the (s,x) grid where $\langle k\rangle_{sx} = 1/3$ with an error of ± 0.001. The resulting segment is highlighted in black in Fig. 2(b). For $x \geqslant 4$ the average path length evaluated along the segment practically reaches an asymptotic value of $\langle \ell\rangle_{sx} \approx 4.67$, which is then the shortest time it takes to reach the target while passing through the obstacle only 1/3 of the times. The probability fluxes for the point $(s = 4.815, x = 4)$, which is highlighted with a pink triangle in Fig. 2(b), are shown in Fig. 2(d). Apart from the shortest path, which occurs with a probability 1/3 (as it should, given that it contains the obstacle), the other 2/3 are equally split among the three second shortest paths, out of a total of five, that do not cross the obstacle.

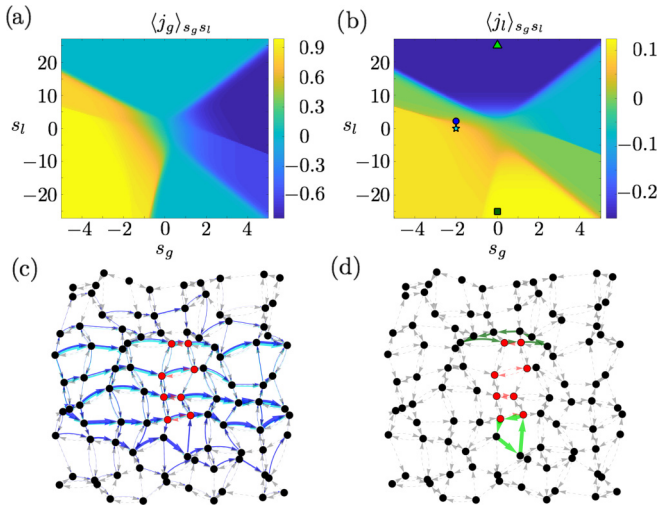

FIG. 3. Optimal weight distribution for the maximization of flows under constraints in spatial networks. (a) Average global current $\langle j_g\rangle_{s_g s_l}$ in the graph shown in panels (c) and (d) as a function of the tilting parameters s_g and s_l. (b) Average local current through the highlighted links joining two red nodes $\langle j_l\rangle_{s_g s_l}$ as a function of the tilting parameters s_g and s_l. Highlighted points correspond to probability fluxes shown in the panels below. (c) Probability fluxes obtained from the Doob-transformed process corresponding to the blue circle and the cyan star shown in panel (b). (c) Probability fluxes obtained from the Doob-transformed process corresponding to the light green triangle and the dark green square in panel (b). Probability fluxes smaller than the largest value divided by 200 are not displayed for visibility reasons.

IV. OPTIMAL WEIGHT DISTRIBUTION FROM LARGE DEVIATIONS OF STANDARD RANDOM WALKS

We next illustrate how to find optimal weight distributions that maximize flows in the presence of constraints. To do so, we consider a SRW on the spatial network shown in Figs. 3(c) and 3(d)—the links highlighted in colors other than gray are probability fluxes in certain dynamical regimes that will be discussed below. This network of $N = 100$ nodes has been generated by adding a noise term uniformly distributed in $[-0.5, 0.5]$ to the x and y coordinates of the nodes of a 10×10 square lattice of lattice constant 1. As an additional source of disorder, the links are equiprobably set to be either bidirectional or unidirectional (in the latter case, the direction is also chosen at random). Since we consider periodic

boundary conditions, these links also include those connecting one end of the network with the opposite end, both in the horizontal and vertical directions (not shown in the figure for visibility reasons). The fact that the network is spatial is important, because the definition of the flow will depend on the coordinates of the nodes that are traversed. As in the previous section, our results are meant to be illustrative and thus we focus on a moderately large and visually simple network, but the same approach can be employed on any topology without restriction.

Two observables are considered, a global observable and a local one. The global observable is the total current along the horizontal axis, which we denote j_g, and is defined as the increment in x coordinate per unit time in a random walk of τ time steps. In a jump from node i to node j, the contribution to the current is thus $x_j - x_i$, where x_i is the horizontal Cartesian coordinate of node i. As we are interested in the flow between the leftmost and rightmost nodes of a finite spatial network, those links that (due to the periodic boundary conditions) join nodes at opposite ends in the horizontal direction do not contribute to the global current j_g.

The local observable, on the other hand, is associated with the highlighted links joining the red nodes in the center of the graphs displayed in Figs. 3(c) and 3(d). We denote it j_l as it is a local current, which simply adds $+1$ if the particle jumps from a red node to another red node that its lying on the right and -1 if the jump goes towards the left. In any other case it is zero, and that includes jumps that are predominantly in the vertical direction or which connect a red node with a black node or two black nodes. In the definition of j_l we do not consider the coordinates of the red nodes, as we are just interested in limiting how frequently the walker passes through those links in a given direction. The role of predominantly horizontal links joining red nodes is in a way similar to that of the links leading to the obstacle in Fig. 2, though in this case we are considering weight distributions, not optimal paths, and currents instead of activities, and both the type of random walk and the statistical ensemble are different. While $-1 < j_g < 1$, the maximum value that j_l can achieve depends on the details of the network (see below).

As both currents, j_g and j_l, are time-averaged observables (of the form O/τ, see Sec. II), we consider the *ss* ensemble (see Appendix B), for which the tilted probability distribution is

$$P_\tau^{s_g s_l}(j_g, j_l) = e^{-s_g \tau j_g - s_l \tau j_l} P_\tau(j_g, j_l)/Z_\tau(s_g, s_l), \qquad (9)$$

where $P_\tau(j_g, j_l) = P_\tau^{s_g=0, s_l=0}(j_g, j_l)$, and $Z_\tau(s_g, s_l)$ is a normalizing factor, and the time τ is assumed to be fixed and large. The SCGF $\theta(s_g, s_l)$ is computed as the largest eigenvalue of the corresponding tilted operator (see Appendix B), and from its partial derivatives we obtain the average global current $\langle j_g \rangle_{s_g s_l} = -\partial_{s_g}\theta(s_g, s_l)$ and the average local current $\langle j_l \rangle_{s_g s_l} = -\partial_{s_l}\theta(s_g, s_l)$, which are shown in Figs. 3(a) and 3(b), respectively. Several distinct regions are identified on both color maps. While for a regular lattice there is a symmetry under the simultaneous reversal of both tilting parameters $s_g \to -s_g$, $s_l \to -s_l$, i.e., $\theta(-s_g, -s_l) = \theta(s_g, s_l)$, the presence of disorder breaks that symmetry in the finite disordered spatial network, leading to some quantitative differences. Such a symmetry would be recovered in the hydrodynamic limit, where microscopic structural details are coarse grained [54].

We focus on global currents moving from left to right, corresponding to $s_g < 0$. We first look at the case where this is the only tilting parameter, by setting $s_l = 0$ [see the ($s_g = -2$, $s_l = 0$) point highlighted as a cyan star in Fig. 3(b) and the probability fluxes depicted as cyan arrows in Fig. 3(c)]. The very large global current $\langle j_g \rangle_{s_g s_l} \approx 0.769$ [see Fig. 3(a)] is due to the existence of jumps that advance consistently to the right along the fastest possible routes. It turns out that those routes go along the links joining the red nodes [see Fig. 3(c)], and thus the local current $\langle j_l \rangle_{s_g s_l}$ is also relatively large for this point, specifically $\langle j_l \rangle_{s_g s_l} \approx 0.078$ [see Fig. 3(b)]. From the Doob-transformed transition matrix used to calculate the probability fluxes, one extracts the optimal link weights that give rise to such currents, as explained in Appendix C.

While keeping such a strong negative tilting parameter s_g so as to generate a strong global current to the right, one may want to prevent the red links from being overused. In applications, this may correspond to a transportation route or communication link that may exceed its limiting capacity. To achieve that, we set s_l to some positive value. Specifically, we select the value of s_l that gives rise to a local current $\langle j_l \rangle_{s_g s_l}$ that is half the value obtained for the same s_g and $s_l = 0$, which corresponds to the point ($s_g = -2$, $s_l = 2.19$). The resulting weight distribution finds alternative routes so that the particle goes through the red links half as frequently as for $s_l = 0$, while decreasing the total current by slightly less than 10%. See the blue circle in Fig. 3(b) and the corresponding blue arrows in Fig. 3(c).

An interesting phenomenon occurs for very strong tilting of the local current j_l. See for example the points ($s_g = 0$, $s_l = -25$) and ($s_g = 0$, $s_l = 25$), highlighted as a dark green square and a light green triangle in Fig. 3(b), respectively. They correspond to the largest and the smallest value that $\langle j_l \rangle_{s_g s_l}$ can take, which are $1/8$ and $-1/4$. These values are associated with the appearance of vortices in the trajectories [see Fig. 3(d)], the absolute value of $\langle j_l \rangle_{s_g s_l}$ being the reciprocal of the number of nodes in the cycle. An analogous vortex dynamics has been found in the simple exclusion process on general graphs [55] and in the zero-range process on a diamond lattice [56], but its appearance in random walks with periodic boundary conditions has not been previously reported, as far as we are aware. For sufficiently small (in absolute value) s_g, this vortex dynamics is preserved intact, as illustrated by the blue and yellow triangle-shaped regions in Fig. 3(b), which shows that the tilting of the local current s_l prevails over s_g there. The corresponding points in Fig. 3(a) show a zero average global current, as befits such localized cyclic motion.

To conclude the exploration of the different dynamical regimes to be found over the spatial network, we focus on positive values of s_g, which correspond to negative global currents j_g. Apart from the vortex regions for very large $|s_l|$ already discussed (where the sign of s_g becomes irrelevant), there is a crossover between a negative and a vanishing local current j_l as s_l is decreased from zero towards negative values [see Fig. 3(b)]. This reflects the fact that, while for $s_l = 0$ (or moderately positive values) the highest current is achieved

by going across the links joining the red nodes, decreasing s_l forces the local current to be positive, which entails avoiding those links. As a consequence, the absolute value of the global current also diminishes [see Fig. 3(a)].

V. DISCUSSION

We have shown how to unveil generalized optimal paths and weight distributions in networks by means of a large-deviation approach. By combining the use of ensembles of trajectories of random walks, where time-integrated observables play the role of order parameters, and the application of the Doob transform, one can visualize the paths, or obtain the transition probabilities, that yield a certain statistical characterization of one or several observables. The statistical nature of the approach presumes that the random walk is performed numerous times, either successively in time, or by many noninteracting random walkers simultaneously. On the other hand, the results are not statistical in the sense of giving information on network ensembles (e.g., all random graphs with a certain degree distribution); quite to the contrary, paths and weights are found for specific topologies (i.e., specific realizations, if one considers the network to be an element of an ensemble), which can be arbitrary—directed, undirected, weighted, unweighted, spatial, or nonspatial. Time-dependent networks could be conceivably considered too, by including a time dependence in the transition probabilities of the processes, but that problem is outside the scope of the present paper.

The most important aspect of our approach is that it shows how to find optimal paths and weight distributions with constraints, which do not have to be limited to one or a given number of observables. For example, one can find the shortest route between two nodes that does not pass through another one more than one-fifth of the times the target node is reached, while another node is visited twice as frequently; or the optimal weighted links adapted to a certain flow without exceeding some activity threshold in some set of nodes, and another one in some specific links. Paths and weights can be tailored to a given mean value or fluctuations of essentially any time-integrated observable, so we expect the approach to be widely applicable.

To our acknowledge, these are problems that are outside the reach of standard graph-theoretical algorithms and combinatorial approaches. Even when some proposals might exist to address one of the problems we have discussed—which is in fact quite possible given the large literature on the subject, spanning various fields of science and engineering—shortest-path and related algorithms are typically specific to the problem at hand, small qualitative changes in the constraints requiring important methodological modifications, while our statistical physics approach is very flexible in this regard. The examples we have shown are meant to be simple and illustrative, but more complex scenarios (in terms of network topology, observables, and constraints) can be similarly studied. All depends on a judicious choice of the appropriate process—which type of random walk, though processes involving exclusion or others could also be considered—observables, and statistical ensemble.

The computational complexity of this approach will be that of the algorithm used to extract the largest eigenvalue (the logarithm of which is the SCGF) and the associated eigenvectors of the tilted generator. While the latter is typically a sparse matrix, which may constitute a significant numerical advantage, for very large networks the numerical eigenvalue problem may be challenging. In such situations, the large-deviation function can be obtained from numerical approaches based on the cloning algorithm [57–59] or adaptive sampling [23,60]. Moreover, the intriguing possibility of finding the optimal dynamics leading to a prescribed fluctuation by approaches adapted from reinforcement learning has emerged lately [61]. Similar machine-learning formulations have been recently proposed for dealing with numerically intractable optimization problems in condensed-matter physics [62,63].

ACKNOWLEDGMENTS

We thank Juan P. Garrahan, from whom we have learned much about the thermodynamics-of-trajectories formalism used in this paper, for useful suggestions. We are also grateful to H. Touchette for reading the paper and helpful comments. C.P.E. acknowledges C. Giardinà and C. Giberti for insightful discussions. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under Marie Sklodowska-Curie Cofund Programme Athenea3I Grant No. 754446; from the European Regional Development Fund, Junta de Andalucía-Consejería de Economía y Conocimiento, Grant No. A-FQM-175-UGR18; and from MINECO, Spain (Grant No. FIS2017-84151-P). We are grateful for the computing resources and related technical support provided by PROTEUS, the supercomputing center of Institute Carlos I in Granada, Spain; and by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [64], which is funded by ENEA, the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development and by Italian and European research programs.

APPENDIX A: RANDOM WALKS ON NETWORKS

We consider two different types of discrete-time random walks on networks: the standard random walk [36] and the maximal entropy random walk [35]. Below we will explain why these choices are particularly well adapted to the problems at hand—others could conceivably be better addressed by other random-walk local rules (see, e.g., [48] and references therein). The networks are of finite size N, directed and strongly connected, i.e., a walker can reach any of the N nodes from any other node along directed links [12]. Connected undirected networks are also implicitly considered as a particular case.

1. Standard random walk

The SRW—variously known in the literature as generic, unbiased, normal, or uniform random walk, or simply *the* random walk—has been extensively studied in the last two decades [36,47,48]. A random walker moves every time step to one of the neighboring nodes that can be reached from its present location, the specific destination being chosen

(uniformly) at random among the different possibilities. Given a network with directed adjacency matrix $\mathbf{A}$, where $A_{ji} = 1$ if there is a link pointing from i to j and zero otherwise, the transition matrix of the SRW, $\mathbf{\Pi}^{\mathrm{SRW}}$, thus assigns the same probability to each of the directed links emanating from a given node:

$$\Pi_{ji}^{\mathrm{SRW}} = \frac{A_{ji}}{k_i^{\mathrm{out}}}. \tag{A1}$$

The normalization by the out degree of node i, defined as the number of neighbors joined by outgoing links $k_i^{\mathrm{out}} = \sum_j A_{ji}$, ensures the conservation of probability, $\sum_j \Pi_{ji}^{\mathrm{SRW}} = 1$. If the probability of occupying node i at time step n is denoted by $p_i(n)$, then this N-state process evolves according to $p_j(n+1) = \sum_i \Pi_{ji}^{\mathrm{SRW}} p_i(n)$. Due to general properties of finite Markov chains [65], the system evolves asymptotically towards a single stationary state, $\boldsymbol{\pi}$, $\lim_{n\to\infty}(\mathbf{\Pi}^{\mathrm{SRW}})^n \boldsymbol{p}(0) = \boldsymbol{\pi}$ [where $\boldsymbol{p}(0)$ is an arbitrary initial state], which is the right eigenvector of the transition matrix associated with eigenvalue 1, $\mathbf{\Pi}^{\mathrm{SRW}}\boldsymbol{\pi} = \boldsymbol{\pi}$. While the SRW on undirected networks is reversible and the probability of occupying a node in the stationary state is proportional to its degree, for directed networks detailed balance is generally not satisfied and the stationary state in general can only be found approximately [47]. When considering particle currents or other types of flows, the SRW appears to be the most natural choice, as illustrated in Sec. IV.

2. Maximal entropy random walk

In a regular network, where all nodes have the same out degree, every walk comprising a given number of jumps governed by Eq. (A1) occurs with the same probability. In a more general setting, however, sequences of nodes visited by the walker of the same length, joining a given source node to a given target node, are not equiprobable. In fact, the SRW trajectory $\omega_\tau = (i_\tau \leftarrow \cdots \leftarrow i_2 \leftarrow i_1 \leftarrow i_0)$ occurs with probability $P(\omega_\tau) = p_{i_0}(0)/(k_{i_\tau}^{\mathrm{out}} \cdots k_{i_1}^{\mathrm{out}} k_{i_0}^{\mathrm{out}})$, which in general depends on the out degree of the intermediate nodes.

In order to explore generalized optimal paths in networks (see Sec. III), it is necessary to start from a Markov chain that assigns the same probability to all walks of the same number of steps joining a given pair of nodes. The goal is that the contributions of specific walks to the solution only depend on the statistics of the observables under study and not on the out degree of the visited nodes. For this reason we focus on the MERW [35], the transition matrix of which is

$$\Pi_{ji}^{\mathrm{MERW}} = \frac{A_{ji}}{\lambda}\frac{v_j}{v_i}, \tag{A2}$$

where λ is the largest eigenvalue of the directed adjacency matrix, and $\boldsymbol{v}$ is the normalized eigenvector associated to it, $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$. In fact, $\boldsymbol{v}$ is the eigenvector centrality [12], so the MERW can be considered as a biased random walk [48,66,67] based on this node-centrality measure. This type of random walk assigns the same probability to each walk of τ steps between the source i_0 and the target i_τ, namely, $\frac{p_{i_0}(0)}{\lambda^\tau}\frac{v_{i_\tau}}{v_{i_0}}$.

Moreover, it can be shown that the MERW produces Shannon entropy at a rate $\ln\lambda$, which is the highest possible entropy-production rate for a discrete-time random walk [35]. While the SRW maximizes the entropy locally (i.e., in a single jump), the MERW optimizes the entropy along a trajectory, and in fact it has been studied in a quite general framework of dynamic entropy maximization with constraints [68]. For undirected networks the stationary-state probability of a given node is the square of its eigenvector centrality, while for directed networks one encounters difficulties similar to those that arise in the characterization of the stationary state of the SRW (see above). At the end of Appendix C we briefly comment on a connection between the MERW and the SRW that has been recently unveiled.

APPENDIX B: ENSEMBLES OF TRAJECTORIES

Time-integrated observables of random walks on networks are studied by means of a thermodynamic formalism of trajectories or histories [26]. Specifically, for a trajectory $\omega_\tau = (i_\tau \leftarrow \cdots \leftarrow i_2 \leftarrow i_1 \leftarrow i_0)$, involving $\tau \in \mathbb{N}$ jumps, we consider time-extensive observables of the form $\hat{O}(\omega_\tau) = \sum_{n=1}^{\tau} \hat{o}(i_n \leftarrow i_{n-1})$, where $\hat{o}(i_n \leftarrow i_{n-1})$ is the increment of the observable at each time step, the value of which depends on the nodes i_{n-1} and i_n, which are joined by a link. While the probability associated to a given trajectory is just $P(\omega_\tau) = \Pi_{i_\tau i_{\tau-1}} \cdots \Pi_{i_2 i_1}\Pi_{i_1 i_0} p_{i_0}(0)$, where the details of the transition matrix $\mathbf{\Pi}$ depend on the choice of random walk, the probability distribution of the observable is $P_\tau(O) = \sum_{\omega_\tau} \delta[O - \hat{O}(\omega_\tau)]P(\omega_\tau)$. This distribution corresponds to an ensemble of trajectories (walks) with fixed observable O and fixed time τ. If the random walk is ergodic, it acquires a large-deviation form, which in terms of the time-intensive observable O/τ is $P_\tau(O) \sim e^{-\tau I(O/\tau)}$, where the function $I(O/\tau)$ is called the rate function, and plays the role of a dynamical entropy.

This ensemble of trajectories is analogous to the microcanonical ensemble of configurations in equilibrium statistical mechanics, and, also in this context, it is less useful to work with than other ensembles that yield equivalent results in the $\tau \to \infty$ limit. In the following we will briefly review the thermodynamics of trajectories approach to the study of the large deviations of such time-integrated observables. The main theoretical framework is developed in [27], though some of the developments that we use are more recent—appropriate references are cited below.

1. s ensemble

By biasing each trajectory with a tilting parameter s conjugate to the observable O, we obtain the s ensemble $P_\tau^s(O) = Z_\tau^{-1}(s)\, e^{-sO} P_\tau(O)$, where the normalization factor is a dynamical partition function $Z_\tau(s) = \sum_O e^{-sO} P_\tau(O)$. In this ensemble, τ is still fixed, but O is not, as it can fluctuate—instead, the tilting parameter s, which plays a role similar to an inverse temperature, is fixed. For large τ, the partition function also acquires a large-deviation form, $Z_\tau(s) \sim e^{\tau\theta(s)}$, where the so-called scaled-cumulant generating function $\theta(s)$ is related to the rate function defined above by a Legendre-Fenchel transform. The SCGF can be seen as a dynamical free energy, the derivatives of which yield the cumulants of O:

$$\lim_{\tau\to\infty} \frac{\langle\langle O^p\rangle\rangle_s}{\tau} = (-1)^p \frac{d^p\theta(s)}{ds^p}, \tag{B1}$$

$\langle\langle O^p\rangle\rangle_s$ being the pth-order cumulant of O for a certain value of s. We focus on the first and second derivatives, corresponding to the average value and the fluctuations. While these derivatives evaluated at $s = 0$ yield a statistical characterization of the natural dynamics, we will also consider them for $s \neq 0$, which (as shown below) corresponds to a dynamics that does not conserve probability. In Appendix C we explain how to obtain proper stochastic dynamics that yield the same statistics for O that are found for $s \neq 0$.

It turns out that the partition function can be written as $Z_\tau(s) = \sum_{i,j}[(\mathbf{\Pi}^s)^\tau]_{ji}p_i(0)$, where the transfer matrix $\mathbf{\Pi^s}$ is the tilted operator, the matrix elements of which are

$$\Pi^s_{ji} = e^{-s\hat{o}(j\leftarrow i)}\Pi_{ji} \tag{B2}$$

for a given random-walk transition matrix Π_{ji}. The largest eigenvalue of this matrix corresponds to $e^{\theta(s)}$ [27]. Finding the statistics of O thus becomes an eigenvalue problem of the tilted generator, which is not a stochastic generator, $\sum_j \Pi^s_{ji} \neq 1$ for $s \neq 0$.

2. *ss* ensemble

If we consider two time-integrated observables instead of one, let them be denoted as O_1 and O_2, which can fluctuate in a trajectory of fixed duration τ, then we are in the *ss* ensemble, $P^{s_1s_2}_\tau(O_1, O_2) = Z^{-1}_\tau(s_1, s_2)\, e^{-s_1O_1-s_2O_2}P_\tau(O_1, O_2)$. Here s_1 and s_2 are two tilting parameters conjugate to O_1 and O_2, respectively. The *ss* ensemble is the ensemble of choice for the study of constrained flows in networks developed in Sec. IV. The large-deviation form of the partition function is in this case $Z_\tau(s_1, s_2) \sim e^{\tau\theta(s_1,s_2)}$, and it can be obtained again by computing the largest eigenvalue of the tilted operator, which adopts the form

$$\Pi^{s_1s_2}_{ji} = e^{-s_1\hat{o}_1(j\leftarrow i)-s_2\hat{o}_2(j\leftarrow i)}\Pi_{ji}. \tag{B3}$$

From the (first-, second-, or higher-order) derivatives of $\theta(s_1, s_2)$ with respect to s_1 or s_2 only, one obtains the cumulants of O_1 and O_2. The cross derivatives yield the O_1O_2 correlations:

$$\lim_{\tau\to\infty}\frac{\langle\Delta O_1\Delta O_2\rangle_{s_1s_2}}{\tau} = \frac{\partial^2\theta(s_1, s_2)}{\partial s_1\partial s_2}, \tag{B4}$$

where $\Delta O_1 = O_1 - \langle O_1\rangle_{s_1s_2}$, and ΔO_2 is defined analogously.

3. *x* ensemble

If instead of keeping fixed the duration of a trajectory τ we fix the value that the observable O reaches for each trajectory—while allowing τ to fluctuate—we obtain a different statistical ensemble, namely, the x ensemble [50]. We focus the description of this ensemble on the case in which the observable is a local activity that equals 1 if a given link (or set of links) is traversed, or zero otherwise, as this is the case that we explore in Sec. III. What follows is a discrete-time version of the discussion to be found in Sec. IV of [51].

The probability that the observable reaches a given value O at time τ is the probability that the link or links that contribute to O are traversed at times $\tau_1, \tau_2, \ldots, \tau_{O-1}, \tau_O = \tau$, where all but the last one can take any value. In terms of the operator $\mathbf{\Pi}_O$, which preserves the transition probabilities of $\mathbf{\Pi}$ only for those transitions (links) contributing to O, and $\tilde{\mathbf{\Pi}} = \mathbf{\Pi} - \mathbf{\Pi}_O$, this probability is given by

$$P_O(\tau) = \sum_{ji}\sum_{0\leqslant\tau_1\leqslant\cdots\leqslant\tau}[\mathbf{\Pi}_O\tilde{\mathbf{\Pi}}^{\tau-\tau_{O-1}-1}\cdots\mathbf{\Pi}_O\tilde{\mathbf{\Pi}}^{\tau_1-1}]_{ji}p_i(0). \tag{B5}$$

The x-ensemble probability distribution is $P^x_O(\tau) = Z_O(x)^{-1}\, e^{-x\tau}P_O(\tau)$, where the normalization factor is

$$\begin{aligned}
Z_O(x) &= \sum_{\tau=0}^{\infty} e^{-x\tau}P_O(\tau) = \sum_{ji}\sum_{\tau=0}^{\infty} e^{-x\tau}\sum_{0\leqslant\tau_1\cdots\leqslant\tau}[\mathbf{\Pi}_O\tilde{\mathbf{\Pi}}^{\tau-\tau_{O-1}-1}\cdots\mathbf{\Pi}_O\tilde{\mathbf{\Pi}}^{\tau_1-1}]_{ji}\, p_i(0)\\
&= \sum_{ji}\sum_{\Delta\tau_1=0}^{\infty}\cdots\sum_{\Delta\tau_O=0}^{\infty}\left[e^{-x}\mathbf{\Pi}_O(e^{-x}\tilde{\mathbf{\Pi}})^{\Delta\tau_O}\cdots e^{-x}\mathbf{\Pi}_O(e^{-x}\tilde{\mathbf{\Pi}})^{\Delta\tau_1}\right]_{ji}p_i(0)\\
&= \sum_{ji}\left[\left(e^{-x}\mathbf{\Pi}_O\sum_{\Delta\tau=0}^{\infty}(e^{-x}\tilde{\mathbf{\Pi}})^{\Delta\tau}\right)^O\right]_{ji}p_i(0) = \sum_{ji}[(\mathbf{\Pi}^x)^O]_{ji}p_i(0).
\end{aligned} \tag{B6}$$

The auxiliary time-increment variables $\Delta\tau_1 = \tau_1 - 1$, $\Delta\tau_2 = \tau_2 - \tau_1 - 1, \ldots, \Delta\tau_0 = \tau - \tau_{O-1} - 1$, the values of which are unrestricted due to the sum over τ, are introduced in the second line so as to emphasize that there are just O identical factors of the form $e^{-x}\mathbf{\Pi}_O\sum_{\Delta\tau=0}^{\infty}(e^{-x}\tilde{\mathbf{\Pi}})^{\Delta\tau}$. Once the sum over $\Delta\tau$ is performed (see the discussion below), the result is a dynamical grand-partition function written in terms of a transfer operator:

$$\mathbf{\Pi}^x = \mathbf{\Pi}_O\,(e^x - \tilde{\mathbf{\Pi}})^{-1}. \tag{B7}$$

This is the tilted operator for the x ensemble.

For large O, which also corresponds to large τ, the grand-partition function acquires a large-deviation form $Z_O(x) \sim e^{O\varphi(x)}$, where $e^{\varphi(x)}$ is the largest eigenvalue of $\mathbf{\Pi}^x$. The cumulants of the fluctuating time τ can then be obtained from the derivatives of $\varphi(x)$:

$$\lim_{O\to\infty}\frac{\langle\langle\tau^p\rangle\rangle_x}{O} = (-1)^p\frac{d^p\varphi(x)}{dx^p}. \tag{B8}$$

This ensemble is equivalent to the s ensemble in the large τ (large O) limit—see [50,51], the discussion of which for continuous-time dynamics can be easily adapted to discrete

time. Nevertheless, for finding optimal paths the x ensemble is more natural, as one is interested in finding the number of steps (i.e., the fluctuating time) needed to reach certain target nodes. The length of the walk between consecutive observable increments, τ/O for fixed and large number of repetitions O, is more appropriate for the task than the s ensemble O/τ, for fixed and large τ, even if one can be easily obtained from the other. A simple mapping relating the large-deviation functions of one and the other ensemble exists—see [50], as well as the pertinent discussion in Appendix C.

Since the transfer operator in Eq. (B7) results from the sum of the matrix series $\sum_{\Delta\tau=0}^{\infty}(e^{-x}\tilde{\mathbf{\Pi}})^{\Delta\tau}$, and this sum only converges for $x > \log(\lambda_{\max})$, where $\lambda_{\max}$ is the largest eigenvalue of $\tilde{\mathbf{\Pi}}$, only in that range is the SCGF well defined. As we move along the x axis towards smaller values and we get sufficiently close to $\log(\lambda_{\max})$ we see a divergence in the SCGF, and consequently in (minus) its first derivative $\langle\tau\rangle/O$. In Sec. III the physical meaning of such divergences is clarified. A divergence is clearly observed in Fig. 1(d)—and also in Figs. 2(a) and 2(b), in the case of the sx ensemble, see below—and explained in the text that discusses it. In the case of Fig. 1(c), there is no such divergence, as the largest eigenvalue of $\tilde{\mathbf{\Pi}}$ for the ring with a shortcut is $\lambda_{\max} = 0$. Physically, it is obvious that no divergence is possible there, as the red node is always reached.

4. *sx* ensemble

We next consider an ensemble involving two observables O and K in a trajectory of duration τ, like in the ss ensemble. In this case, however, O is fixed (i.e., the trajectory ends when O reaches a certain value) and K is a fluctuating observable that is a local activity. Moreover, the duration τ is also fluctuating. This is the sx ensemble, and to the best of our knowledge it has not been studied previously. The tilted probability in this ensemble has two tilting parameters s and x:

$$P_O^{sx}(K,\tau) = Z_O^{-1}(s,x)\, e^{-sK-x\tau} P_O(K,\tau) \tag{B9}$$

where $P_O(K,\tau)$ is the probability that an unbiased trajectory has a given duration τ and the fluctuating observable reaches a certain value K by the time the other observable reaches its fixed value O. The corresponding grand-partition function is

$$Z_O(s,x) = \sum_K \sum_\tau e^{-sK-x\tau} P_O(K,\tau) \sim e^{O\varphi(s,x)}, \tag{B10}$$

where the right-hand side takes the expected large-deviation form with SCGF $\varphi(s,x)$.

We again obtain the SCGF $\varphi(s,x)$ from the largest eigenvalue of a transfer operator. To do so, this time we need to split the original transition matrix into three parts:

$$\mathbf{\Pi} = \mathbf{\Pi}_O + \mathbf{\Pi}_K + \hat{\mathbf{\Pi}} \tag{B11}$$

where the three terms of the right-hand side include all transitions that result in an update of observable O, all transitions that result in an update of K, and the rest of the transitions, respectively. We obtain

$$\begin{aligned} Z_O(s,x) &= \sum_K \sum_{\tau=0}^{\infty} e^{-sK} e^{-x\tau} \sum_{ji} \sum_{0\leqslant\tau_1\leqslant\cdots\leqslant\tau} [\mathbf{\Pi}_O \hat{\mathbf{\Pi}}^{\tau-\tau_{O+K-1}-1} \mathbf{\Pi}_K \cdots \mathbf{\Pi}_O \hat{\mathbf{\Pi}}^{\tau_2-\tau_1-1} \mathbf{\Pi}_K \hat{\mathbf{\Pi}}^{\tau_1-1}]_{ji} p_i(0) \\ &= \sum_{ji} \sum_{\tau=0}^{\infty} e^{-x\tau} \sum_{0\leqslant\tau_1\leqslant\cdots\leqslant\tau} \sum_K [\mathbf{\Pi}_O \hat{\mathbf{\Pi}}^{\tau-\tau_{O+K-1}-1} (e^{-s}\mathbf{\Pi}_K) \cdots \mathbf{\Pi}_O \hat{\mathbf{\Pi}}^{\tau_2-\tau_1-1} (e^{-s}\mathbf{\Pi}_K) \hat{\mathbf{\Pi}}^{\tau_1-1}]_{ji} p_i(0) \end{aligned} \tag{B12}$$

where, apart from those sums made explicit, we are in fact also summing over all possible orders of occurrence of the K and O transitions contained in $\mathbf{\Pi}_K$ and $\mathbf{\Pi}_O$, respectively. We have split the e^{-sK} factor into K factors of the form e^{-s}, associating each of them to one occurrence of $\mathbf{\Pi}_K$.

As all possible values of K are summed over in Eq. (B12), we can combine the (tilted) transitions $e^{-s}\mathbf{\Pi}_K$ and those included in $\hat{\mathbf{\Pi}}$—we thus put together every transition that does not contribute to O, including the s tilting in those contributing to K. We are in a situation analogous to that of the first line of Eq. (B6), but replacing $\tilde{\mathbf{\Pi}}$ by $\hat{\mathbf{\Pi}} + e^{-s}\mathbf{\Pi}_K$. Then, by splitting the $e^{-x\tau}$ on the different operators, and finally summing over all possible values of the intermediate times $\tau_1, \tau_2, \ldots$, just as we did in our discussion of the x ensemble, we obtain $Z_O(s,x) = \sum_{ji}[(\mathbf{\Pi}^{sx})^O]_{ji} p_i(0)$, with a tilted operator:

$$\mathbf{\Pi}^{sx} = \mathbf{\Pi}_O (e^x - \hat{\mathbf{\Pi}} - e^{-s}\mathbf{\Pi}_K)^{-1}. \tag{B13}$$

Our remarks on the convergence of the x-ensemble operator also apply in this case. And in fact an example of a region in tilting-parameter space where the sum over τ does not converge is shown in Fig. 2.

From $\varphi(s,x) = \lim_{O\to\infty} \frac{1}{O} \log Z_O(s,x)$ one can as usual obtain mean values, fluctuations, and higher-order cumulants of K and τ by taking derivatives. In Sec. III we consider the mean values

$$\frac{\langle K\rangle}{O} = -\partial_s \varphi(s,x), \quad \frac{\langle \tau\rangle}{O} = -\partial_x \varphi(s,x), \tag{B14}$$

in the limit of large O (which also corresponds to large K and large τ, as both are O extensive). The second derivatives with respect to the same tilting parameter yield the fluctuations, and the cross derivatives are again the correlations between K and τ for fixed and large O.

APPENDIX C: DOOB TRANSFORM

The Doob transform allows us to obtain a stochastic matrix giving rise to the same statistics as the tilted operator (which is not stochastic) for some observables of interest (O, τ, etc.). From the resulting transition matrix $\mathbf{\Pi}_{\text{Doob}}$ we can obtain the

transition probabilities (link weights of the resulting weighted networks) that give rise to such statistical behavior in the long-time limit.

In Figs. 1–3 we show the Doob-transformed dynamics corresponding to some observable statistics of interest (as given by some values of the tilting parameters). More precisely, we display probability fluxes, which are the product of the Doob-transform transition probabilities and the stationary state distribution $\mathbf{\Pi}_{\text{Doob}}\boldsymbol{p}^{\text{st}}$, where $\boldsymbol{p}^{\text{st}}$ is the vector the components of which are the stationary probabilities associated with the different nodes. While the entries of $\mathbf{\Pi}_{\text{Doob}}$ are transition probabilities, $(\Pi_{\text{Doob}})_{ji}p_i^{\text{st}}$ is the joint probability of occupying node i in a given time step and node j in the next one, which is the probability flux of $j \leftarrow i$. The sizes of the colored arrows in Figs. 2(c), 2(d), 3(c), and 3(d), as well as the arrows shown for $x \neq 0$ in Figs. 1(b) and 1(d), are proportional to these probability fluxes.

We next discuss the Doob transform in each of the ensembles reviewed in Appendix B. In the s ensemble, the tilted operator $\mathbf{\Pi}^s$ is linked to the biased probability $P_\tau^s(O)$ through the largest eigenvalue $e^{\theta(s)}$. The left and right eigenvectors associated with it are $\mathbf{l}_s$ and $\mathbf{r}_s$, respectively: $\mathbf{\Pi}^s\mathbf{r}_s = e^{\theta(s)}\mathbf{r}_s$, $\mathbf{l}_s^T\mathbf{\Pi}^s = e^{\theta(s)}\mathbf{l}_s^T$, where T indicates transposition. However the tilted operator is not a proper stochastic matrix ($\sum_j \Pi_{ji}^s \neq 1$), so it does not provide the transition probabilities leading to the fluctuation of interest. In order to convert it into a proper stochastic matrix one resorts to the generalized Doob transform [31–34], which is

$$\mathbf{\Pi}_{\text{Doob}}^s = e^{-\theta(s)}\hat{\mathbf{L}}_s\mathbf{\Pi}^s\hat{\mathbf{L}}_s^{-1}, \tag{C1}$$

where $\hat{\mathbf{L}}_s$ is a diagonal matrix which has the components of $\mathbf{l}_s$ along the diagonal. The Doob transform $\mathbf{\Pi}_{\text{Doob}}^s$ conserves probability: $\sum_j(\mathbf{\Pi}_{\text{Doob}}^s)_{ji} = e^{-\theta(s)}\mathbf{l}_s^T\mathbf{\Pi}^s\hat{\mathbf{L}}_s^{-1} = \mathbf{1}^T$, where $\mathbf{1}$ is the unit vector. And it has the stationary distribution $\boldsymbol{p}^{\text{st}} = \hat{\mathbf{L}}_s\mathbf{r}_s$: $\mathbf{\Pi}_{\text{Doob}}^s\boldsymbol{p}^{\text{st}} = e^{-\theta(s)}\hat{\mathbf{L}}_s\mathbf{\Pi}^s\hat{\mathbf{L}}_s^{-1}\boldsymbol{p}^{\text{st}} = e^{-\theta(s)}\hat{\mathbf{L}}_s\mathbf{\Pi}^s\mathbf{r}_s = \hat{\mathbf{L}}_s\mathbf{r}_s = \boldsymbol{p}^{\text{st}}$.

The Doob transform can be generalized to the ss ensemble in a straightforward manner: in Eq. (C1) the matrices and the SCGF are now functions of both s_1 and s_2, but otherwise the expression remains intact,

$$\mathbf{\Pi}_{\text{Doob}}^{s_1s_2} = e^{-\theta(s_1,s_2)}\hat{\mathbf{L}}_{s_1s_2}\mathbf{\Pi}^{s_1s_2}\hat{\mathbf{L}}_{s_1s_2}^{-1}, \tag{C2}$$

and satisfies analogous properties of stochasticity and having the same stationary distribution as the tilted dynamics.

The x-ensemble tilted operator $\mathbf{\Pi}^x$ satisfies $\mathbf{\Pi}^x\mathbf{r}_x = e^{\varphi(x)}\mathbf{r}_x$, $\mathbf{l}_x^{\text{T}}\mathbf{\Pi}^x = e^{\varphi(x)}\mathbf{l}_x^{\text{T}}$, where $\mathbf{l}_x$ and $\mathbf{r}_x$ are again the eigenvectors associated with the maximum eigenvalue $e^{\varphi(x)}$. In this case, the Doob transform is

$$\mathbf{\Pi}_{\text{Doob}}^x = e^{-x}\hat{\mathbf{L}}_x\mathbf{\Pi}^{s=\varphi(x)}\hat{\mathbf{L}}_x^{-1}, \tag{C3}$$

where $\hat{\mathbf{L}}_x$ is a diagonal matrix which has the components of $\mathbf{l}_x$ along the diagonal. In fact this is just a rewriting of Eq. (C1) in terms of x-ensemble parameters, based on the ensemble equivalence whereby $\theta(s) = x$ and $\varphi(x) = s$ when the left eigenvectors in both ensembles are equal [50,51].

The Doob transform of the sx ensemble is written in terms of the left eigenvector of $\mathbf{\Pi}^{sx}$ associated with the eigenvalue $e^{\varphi(s,x)}$, $\mathbf{l}_{sx}$, more precisely on its rearrangement as a diagonal matrix $\hat{\mathbf{L}}_{sx}$. It is given by

$$\mathbf{\Pi}_{\text{Doob}}^{sx} = e^{-x}\hat{\mathbf{L}}_{sx}\mathbf{\Pi}^{s,s'=\varphi(s,x)}\hat{\mathbf{L}}_{sx}^{-1}. \tag{C4}$$

This is just a rewriting of Eq. (C2) in terms of sx-ensemble parameters, based on the ensemble equivalence whereby $\theta(s,s') = x$ and $\varphi(s,x) = s'$ when the left eigenvectors in both ensembles are equal. These relations are like those between the s and the x ensembles fixing one of the s variables (namely, s, which plays the role of s_1).

It seems pertinent to conclude by highlighting a connection between the SRW and the MERW that has been recently unveiled by means of the Doob transform. It turns out that the MERW is the Doob transform of the s-biased SRW for the observable $\sum_i \log k_i^{\text{out}}$, where the sum is taken over the nodes visited in a walk, with tilting parameter $s = -1$ [30]. While in the highly symmetric case of regular networks both random walks are equivalent, important qualitative differences in the trajectories are observed already in the presence of small deviations from regularity, including localization effects [35]. In a large-deviation framework, these are a consequence of biasing the SRW in such a way that the walker favors visiting nodes with a larger (logarithm of the) degree. Dynamical phase transitions towards localized states in biased SRWs on networks are more generally explored in [29].

APPENDIX D: LARGE DEVIATIONS OF A RANDOM WALK ON A DIRECTED RING WITH A SHORTCUT

In the case of the random walk on a directed ring with a shortcut studied in the first part of Sec. III.A, one can analytically obtain the SCGF of x ensemble $\varphi(x)$ from the large deviations of the probability distribution of the sample mean of the cyclic-walk length ℓ.

A cyclic walk that ends at the starting node is performed M times in succession, with cycle lengths $\ell^{(1)}, \ell^{(2)}, \ldots \ell^{(M)}$, where each $\ell^{(i)}$ is a Bernoulli trial which takes the values ℓ_p and ℓ_q with probabilities p and $q = 1 - p$, respectively. Let n_p be a random variable that gives the fraction of cycles of length ℓ_p that occur in the sequence, the probability distribution of which takes the following binomial distribution:

$$P_M(n_p) = \frac{M!}{(n_pM)![(1-n_p)M]!}\,p^{n_pM}(1-p)^{(1-n_p)M}. \tag{D1}$$

For large M a straightforward application of Stirling's approximation, $M! \sim M^Me^{-M}$, yields the large-deviation form $P_M(n_p) \sim e^{-MI(n_p)}$, where $I(n_p) = D_{\text{KL}}(n_p||p) = n_p\log[n_p/p] + (1-n_p)\log[(1-n_p)/(1-p)]$ is the Kullback-Leibler divergence between a Bernoulli trial with probability p and one with probability n_p. Such rate function $I(n_p)$ achieves its minimum (zero) for $n_p = p$, which corresponds to the mean value, $\langle n_p\rangle = p$. Notice that for small fluctuations around the mean, i.e., $|n_p - p| \sim \mathcal{O}(1/\sqrt{M})$, $I(n_p)$ can be approximated up to second order as $I(n_p) \approx (n_p-p)^2/[2p(1-p)] + \mathcal{O}[(n_p-p)^3]$, which corresponds to the central limit theorem prediction.

As we are interested in the statistics of $\ell = M^{-1}\sum_{i=1}^M \ell^{(i)}$, which can be rewritten as $\ell = n_p\ell_p + (1-n_p)\ell_q$, we simply change variables. The probability distribution of ℓ thus

acquires the large-deviation form

$$P_M(\ell) \sim e^{-MI(\ell)}. \tag{D2}$$

Here, $I(\ell)$ is just $I[n_p = (\ell - \ell_q)/(\ell_p - \ell_q)]$, and the Jacobian factor of the change of variable has been neglected as it does not alter the dominant exponential form of the distribution for large M.

The number of times M that the random walker returns to the starting node is fixed, but the duration of the trajectory is fluctuating. Therefore, we study the random-walk trajectories in the x ensemble. The partition function of the x-ensemble distribution $P_M^x(\ell) = Z_M^{-1}(x)e^{-xM\ell}P_M(\ell)$ takes the form

$$Z_M(x) = \sum_\ell e^{-xM\ell}P_M(\ell) \sim e^{M\varphi(x)}, \tag{D3}$$

where the SCGF $\varphi(x) = -\min_\ell[x\ell + I(\ell)]$, which is the Legendre transform of the rate function $I(\ell)$, is obtained from a saddle-point approximation.

We then minimize $x\ell + I(\ell)$, which amounts to finding the value ℓ^* such that $I'(\ell^*) = -x$, so we can write $\varphi(x) = -x\ell^* - I(\ell^*)$. The minimum is found for

$$\ell^* = \frac{pe^{-x(\ell_p-\ell_q)}\ell_p + (1-p)\ell_q}{pe^{-x(\ell_p-\ell_q)} + (1-p)}. \tag{D4}$$

In terms of $p_x = pe^{-x(\ell_p-\ell_q)}/[pe^{-x(\ell_p-\ell_q)} + (1-p)]$, $\ell^* = p_x\ell_p + (1-p_x)\ell_q$, which is the mean length for a process that avoids taking the shortcut with probability p_x. The SCGF is then

$$\varphi(x) = -x[p_x\ell_p + (1-p_x)\ell_q] - D_{\mathrm{KL}}(p_x||p), \tag{D5}$$

where $D_{\mathrm{KL}}(p_x||p)$ is the Kullback-Leibler divergence between a Bernoulli trial with probability p and one with probability p_x. This expression can be simplified so as to yield the simple form that appears in the main text:

$$\varphi(x) = -x\ell_q + \log[p\,e^{-x(\ell_p-\ell_q)} + (1-p)]. \tag{D6}$$

As expected, the same result is obtained from the analytical calculation of the largest eigenvalue of the x-ensemble transfer operator for the directed ring (B7).

The first derivative of the SCGF yields the average cyclic path length:

$$\langle\ell\rangle_x = -\varphi'(x) = \ell_q + (\ell_p - \ell_q)\frac{p}{p + (1-p)e^{x(\ell_p-\ell_q)}}. \tag{D7}$$

For large absolute values of x, $\langle\ell\rangle_x \approx \ell_q$ if $x > 0$, and $\langle\ell\rangle_x \approx \ell_p$ if $x < 0$. This is in agreement with the asymptotic values observed in Fig. 1(c), and with the intuitively expected result of a strong bias towards shorter or longer paths, respectively. As x appears multiplying the difference between path lengths in the exponent $x(\ell_p - \ell_q)$, a weaker (stronger) bias will be required to approach the asymptotic values in rings where the shortcut starts closer to the beginning (end) of the path around the ring.

[1] F. B. Zhan and C. E. Noon, Shortest path algorithms: An evaluation using real road networks, Transp. Sci. **32**, 65 (1998).

[2] W. Zeng and R. L. Church, Finding shortest paths on real road networks: The case for A, Int. J. Geogr. Inf. Sci **23**, 531 (2009).

[3] G. Li, S. D. S. Reis, A. A. Moreira, S. Havlin, H. E. Stanley, and J. S. Andrade, Jr., Towards Design Principles For Optimal Transport Networks, Phys. Rev. Lett. **104**, 018701 (2010).

[4] R. Albert, H. Jeong, and A.-L. Barabási, Diameter of the world-wide web, Nature (London) **401**, 130 (1999).

[5] H. Tangmunarunkit, R. Govindan, S. Shenker, and D. Estrin, The impact of routing policy on internet paths, in *Proc. of the IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society, Anchorage, Alaska* (IEEE, New York, 2001), Vol. 2, pp. 736–742.

[6] B. Fortz and M. Thorup, Increasing internet capacity using local search, Comput. Optim. Appl. **29**, 13 (2004).

[7] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Improved routing strategies for internet traffic delivery, Phys. Rev. E **70**, 056105 (2004).

[8] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, and B.-H. Wang, Efficient routing on complex networks, Phys. Rev. E **73**, 046108 (2006).

[9] J. R. Managbanag, T. M. Witten, D. Bonchev, L. A. Fox, M. Tsuchiya, B. K. Kennedy, and M. Kaeberlein, Shortest-path network analysis is a useful approach toward identifying genetic determinants of longevity, PLOS ONE **3**, e3802 (2008).

[10] M. Jiang, Y. Chen, Y. Zhang, L. Chen, N. Zhang, T. Huang, Y.-D. Cai, and X.-Y. Kong, Identification of hepatocellular carcinoma related genes with k-th shortest paths in a protein–protein interaction network, Mol. BioSyst. **9**, 2720 (2013).

[11] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. **424**, 175 (2006).

[12] M. E. J. Newman, *Networks: An Introduction* (Oxford University, New York, 2010).

[13] S.-Q. Zheng, J. S. Lim, and S. S. Iyengar, Finding obstacle-avoiding shortest paths using implicit connection graphs, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **15**, 103 (1996).

[14] W. M. Carlyle, J. O Royset, and R. K. Wood, Lagrangian relaxation and enumeration for solving constrained shortest-path problems, Networks **52**, 256 (2008).

[15] L. Lozano and A. L. Medaglia, On an exact method for the constrained shortest path problem, Comput. Oper. Res. **40**, 378 (2013).

[16] J. Hershberger, N. Kumar, and S. Suri, Shortest paths in the plane with obstacle violations, Algorithmica **82**, 1813 (2020).

[17] S. Knopp, P. Sanders, D. Schultes, F. Schulz, and D. Wagner, Computing many-to-many shortest paths using highway hierarchies, in *2007 Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments (ALENEX)* (SIAM, Philadelphia, PA, 2007), p. 36.

[18] G. R. Jagadeesh and T. Srikanthan, Fast computation of clustered many-to-many shortest paths and its application to map matching, ACM Trans. Spat. Algorithms Syst. **5**, 3 (2019).

[19] T. Tanaka and T. Aoyagi, Optimal weighted networks of phase oscillators for synchronization, Phys. Rev. E **78**, 046210 (2008).

[20] T. Nishikawa and A. E. Motter, Synchronization is optimal in nondiagonalizable networks, Phys. Rev. E **73**, 065106(R) (2006).

[21] R. Sevilla-Escoboza, J. M. Buldú, S. Boccaletti, D. Papo, D.-U. Hwang, G. Huerta-Cuellar, and R. Gutiérrez, Experimental implementation of maximally synchronizable networks, Physica A **100**, 113 (2016).

[22] L. Kempton, G. Herrmann, and M. Di Bernardo, Self-organization of weighted networks for optimal synchronizability, IEEE Trans. Control. Netw. Syst. **5**, 1541 (2017).

[23] T. Nemoto, R. L. Jack, and V. Lecomte, Finite-Size Scaling of A First-Order Dynamical Phase Transition: Adaptive Population Dynamics And An Effective Model, Phys. Rev. Lett. **118**, 115702 (2017).

[24] U. Ray, G. K.-L. Chan, and D. T. Limmer, Exact Fluctuations Of Nonequilibrium Steady States From Approximate Auxiliary Dynamics, Phys. Rev. Lett. **120**, 210602 (2018).

[25] R. Hurtado-Gutiérrez, F. Carollo, C. Pérez-Espigares, and P. I. Hurtado, Building Continuous Time Crystals From Rare Events, Phys. Rev. Lett. **125**, 160601 (2020).

[26] V. Lecomte, C. Appert-Rolland, and F. Van Wijland, Thermodynamic formalism for systems with Markov dynamics, J. Stat. Phys. **127**, 51 (2007).

[27] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories, J. Phys. A: Math. Theor. **42**, 075007 (2009).

[28] H. Touchette, The large deviation approach to statistical mechanics, Phys. Rep. **478**, 1 (2009).

[29] C. De Bacco, A. Guggiola, R. Kühn, and P. Paga, Rare events statistics of random walks on networks: localisation and other dynamical phase transitions, J. Phys. A: Math. Theor. **49**, 184003 (2016).

[30] F. Coghi, J. Morand, and H. Touchette, Large deviations of random walks on random graphs, Phys. Rev. E **99**, 022137 (2019).

[31] D. Simon, Construction of a coordinate Bethe ansatz for the asymmetric simple exclusion process with open boundaries, J. Stat. Mech. (2009) P07017.

[32] V. Popkov, G. M. Schütz, and D. Simon, ASEP on a ring conditioned on enhanced flux, J. Stat. Mech. (2010) P10007.

[33] R. L. Jack and P. Sollich, Large deviations and ensembles of trajectories in stochastic models, Prog. Theor. Phys. Supp. **184**, 304 (2010).

[34] R. Chetrite and H. Touchette, Nonequilibrium Markov processes conditioned on large deviations, Ann. Henri Poincaré **16**, 2005 (2015).

[35] Z. Burda, J. Duda, J.-M. Luck, and B. Waclaw, Localization Of The Maximal Entropy Random Walk, Phys. Rev. Lett. **102**, 160602 (2009).

[36] J. D. Noh and H. Rieger, Random Walks on Complex Networks, Phys. Rev. Lett. **92**, 118701 (2004).

[37] G. Bianconi, Rare events and discontinuous percolation transitions, Phys. Rev. E **97**, 022314 (2018).

[38] G. Bianconi, Large deviation theory of percolation on multiplex networks, J. Stat. Mech. (2019) P023405.

[39] J. Hindes and I. B. Schwartz, Epidemic Extinction and Control in Heterogeneous Networks, Phys. Rev. Lett. **117**, 028302 (2016).

[40] J. Hindes and I. B. Schwartz, Epidemic extinction paths in complex networks, Phys. Rev. E **95**, 052317 (2017).

[41] J. Hindes and M. Assaf, Degree Dispersion Increases The Rate Of Rare Events In Population Networks, Phys. Rev. Lett. **123**, 068301 (2019).

[42] H. Chen, C. Shen, H. Zhang, and J. Kurths, Large deviation induced phase switch in an inertial majority-vote model, Chaos **27**, 081102 (2017).

[43] H. Chen, F. Huang, G. Li, and H. Zhang, Large deviation and anomalous fluctuations scaling in degree assortativity on configuration networks, arXiv:1907.13330 (2019).

[44] F. den Hollander, M. Mandjes, A. Roccaverde, and N. J. Starreveld, Ensemble equivalence for dense graphs, Electron. J. Probab. **23**, 1 (2018).

[45] C. Giardinà, C. Giberti, and E. Magnanini, Approximating the cumulant generating function of triangles in the Erdös-Rényi random graph, J. Stat. Phys. **182**, 23 (2021).

[46] F. L. Metz and I. P. Castillo, Condensation of degrees emerging through a first-order phase transition in classical random graphs, Phys. Rev. E **100**, 012305 (2019).

[47] N. Masuda, M. A. Porter, and R. Lambiotte, Random walks and diffusion on networks, Phys. Rep. **716**, 1 (2017).

[48] A. P. Riascos and J. L. Mateos, Random walks on weighted networks: Exploring local and non-local navigation strategies, arXiv:1901.05609 (2019).

[49] R.K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms and Applications* (Prentice-Hall, Upper Saddle River, NJ, 1993).

[50] A. A. Budini, R. M. Turner, and J. P. Garrahan, Fluctuating observation time ensembles in the thermodynamics of trajectories, J. Stat. Mech. (2014) P03012.

[51] J. P. Garrahan, Simple bounds on fluctuations and uncertainty relations for first-passage times of counting observables, Phys. Rev. E **95**, 032134 (2017).

[52] F. Carollo, J. P. Garrahan, I. Lesanovsky, and C. Pérez-Espigares, Making rare events typical in Markovian open quantum systems, Phys. Rev. A **98**, 010103(R) (2018).

[53] F. Angeletti and H. Touchette, Diffusions conditioned on occupation measures, J. Math. Phys. **57**, 023303 (2016).

[54] R. Villavicencio-Sanchez and R. J. Harris, Local structure of current fluctuations in diffusive systems beyond one dimension, Phys. Rev. E **93**, 032134 (2016).

[55] T. Bodineau, B. Derrida, and J. L. Lebowitz, Vortices in the two-dimensional simple exclusion process, J. Stat. Phys. **131**, 821 (2008).

[56] R. Villavicencio-Sanchez, R. J. Harris, and H. Touchette, Current loops and fluctuations in the zero-range process on a diamond lattice, J. Stat. Mech. (2012) P07007.

[57] C. Giardinà, J. Kurchan, and L. Peliti, Direct Evaluation Of Large-Deviation Functions, Phys. Rev. Lett. **96**, 120603 (2006).

[58] C. Giardinà, J. Kurchan, V. Lecomte, and J. Tailleur, Simulating rare events in dynamical processes, J. Stat. Phys. **145**, 787 (2011).

[59] F. Carollo and C. Pérez-Espigares, Entanglement statistics in Markovian open quantum systems: A matter of mutation and selection, Phys. Rev. E **102**, 030104(R) (2020).

[60] G. Ferré and H. Touchette, Adaptive sampling of large deviations, J. Stat. Phys. **172**, 1525 (2018).

[61] D. C Rose, J. F. Mair, and J. P. Garrahan, A reinforcement learning approach to rare trajectory sampling, New J. Phys. **23**, 013013 (2021).

[62] S. Whitelam and I. Tamblyn, Learning to grow: Control of material self-assembly using evolutionary reinforcement learning, Phys. Rev. E **101**, 052604 (2020).

[63] A. Barr, W. Gispen, and A. Lamacraft, Quantum ground states from reinforcement learning, in *Proc. of the First Mathematical and Scientific Machine Learning Conference, Princeton, NJ, USA*, edited by J. Lu and R. Ward (PMLR, 2020), Vol. 107, pp. 635–653.

[64] F. Iannone, F. Ambrosino, G. Bracco, M. De Rosa, A. Funel, G. Guarnieri, S. Migliori, F. Palombi, G. Ponti, G. Santomauro, and P. Procacci, CRESCO ENEA HPC clusters: A working example of a multifabric GPFS Spectrum Scale layout, in *Proceedings of the 2019 International Conference on High Performance Computing Simulation (HPCS), Dublin, Ireland* (IEEE, 2019), pp. 1051–1052.

[65] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes* (Clarendon, Oxford, 1992).

[66] J. Gómez-Gardenes and V. Latora, Entropy rate of diffusion processes on complex networks, Phys. Rev. E **78**, 065102(R) (2008).

[67] V. Kishore, M. S. Santhanam, and R. E. Amritkar, Extreme events and event size fluctuations in biased random walks on networks, Phys. Rev. E **85**, 056120 (2012).

[68] C. Monthus, Non-equilibrium steady states: Maximization of the Shannon entropy associated with the distribution of dynamical trajectories in the presence of constraints, J. Stat. Mech. (2011) P03008.