Accepted Manuscript

Random walks and diffusion on networks

Naoki Masuda, Mason A. Porter, Renaud Lambiotte

PII: S0370-1573(17)30294-6
DOI: http://dx.doi.org/10.1016/j.physrep.2017.07.007
Reference: PLREP 1972

To appear in: *Physics Reports*

Accepted date : 25 July 2017

Please cite this article as: N. Masuda, M.A. Porter, R. Lambiotte, Random walks and diffusion on networks, *Physics Reports* (2017), http://dx.doi.org/10.1016/j.physrep.2017.07.007

Random walks and diffusion on networks

Naoki Masuda

Department of Engineering Mathematics, University of Bristol, Bristol, UK

Mason A. Porter

Department of Mathematics, University of California Los Angeles, Los Angeles, USA

Mathematical Institute, University of Oxford, Oxford, UK

CABDyN Complexity Centre, University of Oxford, Oxford, UK

Renaud Lambiotte

Department of Mathematics/Naxys, University of Namur, Namur, Belgium

Mathematical Institute, University of Oxford, Oxford, UK

Abstract

Random walks are ubiquitous in the sciences, and they are interesting from both theoretical and practical perspectives. They are one of the most fundamental types of stochastic processes; can be used to model numerous phenomena, including diffusion, interactions, and opinions among humans and animals; and can be used to extract information about important entities or dense groups of entities in a network. Random walks have been studied for many decades on both regular lattices and (especially in the last couple of decades) on networks with a variety of structures. In the present article, we survey the theory and applications of random walks on networks, restricting ourselves to simple cases of single and non-adaptive random walkers. We distinguish three main types of random walks: discrete-time random walks, node-centric continuous-time random walks, and edge-centric continuous-time random walks. We first briefly survey random walks on a line, and then we consider random walks on various types of networks. We extensively discuss applications of random walks, including ranking of nodes (e.g., PageRank), community detection, respondent-driven sampling, and opinion models such as voter models.

Contents

Email address: naoki.masuda@bristol.ac.uk (Naoki Masuda)

Random walks and diffusion on networks

Naoki Masuda

Department of Engineering Mathematics, University of Bristol, Bristol, UK

Mason A. Porter

Department of Mathematics, University of California Los Angeles, Los Angeles, USA

Mathematical Institute, University of Oxford, Oxford, UK

CABDyN Complexity Centre, University of Oxford, Oxford, UK

Renaud Lambiotte

Department of Mathematics/Naxys, University of Namur, Namur, Belgium

Mathematical Institute, University of Oxford, Oxford, UK

Keywords: random walk, network, diffusion, Markov chain, point process

1. Introduction

Random walks (RWs) are popular models of stochastic processes with a very rich history [1–5]. [1] The term "random walk" was coined by Karl Pearson [6], and the study of RWs dates back to the "Gambler's Ruin" problem analyzed by Pascal, Fermat, Huygens, Bernoulli, and others [7]. Additionally, Albert Einstein formulated stochastic motion (in the form of "Brownian motion") of particles in continuous time due to their collisions with atoms and molecules [8]. Theoretical developments have involved mathematics (especially probability theory), computer science, statistical physics, operations research, and more. RW models have also been applied in various domains, ranging from locomotion and foraging of animals [9–12], the dynamics of neuronal firing [13, 14] and decision-making in the brain [15, 16] to population genetics [17], polymer chains [18, 19], descriptions of financial markets [20, 21], evolution of research interests (through RWs on problem space) [22], ranking systems [23], dimension reduction and feature extraction from high-dimensional data (e.g., in the form of "diffusion maps") [24, 25], and even sports statistics [26, 27]. RW theory can also help predict arrival times of diseases spreading on networks [28]. There exist several monographs and review papers on RWs. Many of them treat RWs

Email address: naoki.masuda@bristol.ac.uk (Naoki Masuda)

[1] See https://www.youtube.com/watch?v=stgYW6M5o4k for an introduction to random walks for a public audience from the U.S. Public Broadcasting Service (PBS).

on classical network topologies, such as regular lattices (e.g., $\mathbb{Z}^d$) and Cayley trees (i.e., trees in which each node has the same number of neighboring nodes, which we henceforth call the node "degree") [4, 29–35]. Other monographs and surveys focus on RWs on fractal structures, revealing diffusion properties that are "anomalous" compared to RWs on regular lattices or Euclidean spaces (i.e., $\mathbb{R}^d$) [32, 36–40]. Other literature treats RWs on finite networks, which are equivalent to a finite Markov chain (in the discrete-time case) [1, 32, 41, 42] and are at the core of several stochastic algorithms.

In parallel, "network science" has emerged in recent years as a central approach to the study of complex systems [43–46]. Networks are a natural representation of systems composed of interacting elements and allow one to examine the impact of structure on the dynamics and function of a system (as well as the impact of dynamics and function on network structure). Examples include friendship networks, international relationships, gene-regulatory networks, food webs, airport networks, the internet, and myriad more. In each case, one can represent the system's connectivity structure as a set of nodes (representing the entities in the system) and edges (representing interactions among those entities). The study of networks is highly interdisciplinary, and it integrates theoretical and computational tools from subjects such as applied mathematics, statistical physics, computer science, engineering, sociology, economics, biology, and other domains. Many networks exhibit complex yet regular patterns that are explainable (sometimes arguably) by simple mechanisms. Network science has also had a strong impact on the understanding of dynamical processes because of the critical role of structure on spreading processes, synchronization, and others [47–49]. As with RWs, numerous books and review papers have been written on networks, including textbooks [44, 45, 50–52], general review articles [46, 53], and more specialized reviews on topics such as dynamical processes on networks [48, 49, 54], connections to statistical physics [55, 56], temporal networks [57–59], multilayer networks [60–62], and community structure [63–65].

The main purpose of the present review is to bring together two broad subjects — RWs and networks — by discussing their many interconnections and their ensuing applications. RWs are often used as a model for diffusion, and there has been intense research on the impact of network architecture on the dynamics of RWs. Moreover, nontrivial network structure paves the way for different definitions of RWs, and different definitions can be "natural" from some perspective, while leading to different diffusive processes on the same network. Finally, RWs are at the core of several algorithms to uncover structural properties in networks. We will discuss these points further in the next three paragraphs.

First, RWs are often used as a model for diffusion, and there has been intense research on the impact of network architecture on the dynamics of RWs. The finiteness of a network — along with properties such as degree heterogeneity, community structure, and others — can make diffusion on networks both quantitatively and even qualitatively different from diffusion on regular or infinite lattices. RWs on networks are an example of a Markov chain in which the set of nodes is the state space and the transition probabilities depend on the exis-

tence and weights of the edges between nodes. In this review, we will include a summary of results on the dependence of dynamical properties — including stationary distribution and mean first-passage time — on structural properties of an underlying network.

Second, the irregularity of underlying network structure opens the door for different definitions of RWs. Each is "natural" from some perspective, but they lead to different diffusive processes even when considering the same network. For example, it is useful to distinguish between discrete-time and continuous-time RWs. On networks in which degree (i.e., the number of neighbors) is heterogeneous (i.e., it depends on the node), one needs to subdivide continuous-time RWs further into two major types, depending on whether the random events that induce walker movement are generated on nodes or edges and corresponding to different types of propagators (normalized versus unnormalized Laplacian matrices). Different literatures use different variants of RWs, often implicitly. We distinguish different types of RWs and clarify the relationship between them, and we discuss formulations and results that are informed by empirical networks (such as networks with heavy-tailed degree distributions, multilayer networks, and temporal networks).

Finally, RWs lie at the core of many algorithms to uncover various types of structural properties of networks. Consider the notion of identifying "central" nodes, edges, or other substructures in networks [44]. A powerful set of diagnostics (e.g., PageRank [23, 66] and eigenvector centrality [67]) are derived based on recursive arguments of the type "a node is important if it is connected to many important nodes", and such derivations often rely on the trajectories of random walkers. Similarly, flow-based algorithms, based on trajectories of dynamical processes (e.g., RWs) being trapped within certain sets of node for a long time, are helpful for discovering mesoscale patterns in networks [65, 68]. These techniques and algorithms open a wealth of applications that go well beyond classical applications of RWs. Their design benefits both explicitly and implicitly from developing an understanding of how RW dynamics are influenced by network structure and how different types of RWs behave on the same network.

There has been a vast amount of research on RWs on networks, and it is scattered across disparate corners of the scientific literature. It is impossible to cover everything, and we choose specific subsets of it to make our review cohesive, although we will occasionally include pointers to other parts of the landscape. First, we focus on the most standard types of RWs, in which a random walker moves to a neighbor with a probability proportional to edge weight, and their very close relatives. We only very rarely mention some of the numerous other types of RWs, which include correlated RWs [69], self-avoiding RWs [4, 70, 71], zero-range processes [72], multiplicative random processes [73, 74], adaptive RWs (including reinforced RWs [75]), branching RWs [76], Lévy flights [34, 35], elephant RWs [77], quantum walks [78, 79], intermittent RWs [80], persistent RWs [81], starving RWs [82–84], mortal RWs [85], and so on. These processes are of course fascinating, and many of the different flavors of RWs are often developed with specific motivation from an application (e.g., a

Pac-Man-like "hungry RW" [86] has been used as a model for chemotaxis in a porous medium), are often inspired by applications, such as animal movement [10, 12] or financial markets [21], and one can find discussions of different flavors of RWs in Refs. [4, 34, 35]. Second, we will not cover many results for RWs on particular generative models of networks, except that we do give extensive attention to first-passage times for fractal and pseudo-fractal network models (see Section 3.2.5). Third, we will not discuss various important, rigorous results from mathematics and theoretical computer science. For such results, see [1, 4, 30, 41, 42]. We focus instead on results that we believe give physical insight on RW processes and their applications.

As a final warning, we focus exclusively on diffusive processes in which the total number of walkers (or, equivalently, the total probability of observing a walker) is a conserved quantity [2]. The only exception is in Section 5.7, where we use "coalescing RWs" as an analytical tool. As we will see, this conservation rule translates into certain properties of the operator that drives the RW process. When transposed, the operator leads naturally to linear models for consensus dynamics (see Sections 5.7 and 5.8). Among notable non-conservative processes, which we do not cover in this review, are classical epidemic processes [48, 49, 89, 90], in which the number of entities (e.g., viruses or infected individuals) varies over time. In the linear regime, corresponding to a small number of infected nodes, the propagator of infection events in simple epidemic processes such as susceptible–infected (SI) and susceptible–infected–recovered (SIR) models are the adjacency matrix [91, 92]. In contrast, a propagator of an RW is a type of Laplacian matrix, as we will discuss in detail in Section 3. If all nodes have the same degree, these Laplacian and adjacency matrices are related linearly, and their dynamics are essentially the same [59, 93]. However, they are generically different for heterogeneous networks, such as when degree depends on node identity. Therefore, the difference between conservative dynamics (described by a Laplacian matrix) and non-conservative dynamics (described by the adjacency matrix) tends to be more striking for heterogeneous than for homogeneous networks. Other spreading models that are also beyond the scope of this work include threshold models of social contagions [49, 94] (e.g., for modeling adoption of behaviors) and reaction–diffusion dynamics [95].

The rest of our review proceeds as follows. In Section 2, we discuss RWs on the line. In Section 3, we give a lengthy presentation of RWs on networks. We then discuss RWs on multilayer networks in Section 4.1 and RWs on temporal networks in Section 4.2. We discuss applications in Section 5, and we conclude in Section 6.

[2] We thus consider "conservative" processes, though non-conservative processes are also interesting [87, 88].

2. Random walks on the line

In this section, we review some basic properties of RW processes on one-dimensional space (i.e., the infinite line). This section serves as a primer to later sections, in which we examine RWs on general networks. In this and later sections, we carefully distinguish between discrete-time and continuous-time models.

2.1. Discrete time

Consider a discrete-time RW (DTRW) process on the infinite line, which we identify with $\mathbb{R}^1 \equiv \mathbb{R}$. There is a single walker. At each discrete time step, it moves from some point to some other point, including the case of moving from a point to itself. The length and direction of the move are both random variables. We assume that the probability that a walker located at x moves to the interval $[x+r, x+r+\Delta r]$ in one step is equal to $f(r)\Delta r$. The normalization is $\int_{-\infty}^{\infty} f(r)\mathrm{d}r = 1$, and we assume that moves at different times are independent.

Let's derive the probability density $p(x;n)$ that a random walker is located at a point $x \in \mathbb{R}$ after n steps. (For emphasis, we sometimes use the term "discrete time" or "event time" for n.) The master equation is given by

$$p(x;n) = \int_{-\infty}^{\infty} f(x-x')p(x';n-1)\mathrm{d}x' \,. \tag{1}$$

It is convenient to solve Eq. (1) for general x and n in the Fourier domain. We define the Fourier transform by

$$\hat{p}(k;n) \equiv \int_{-\infty}^{\infty} p(x;n)e^{-ikx}\mathrm{d}x \tag{2}$$

and the inverse Fourier transform by

$$p(x;n) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{p}(k;n)e^{ikx}\mathrm{d}k \,. \tag{3}$$

Note that $\hat{p}(-k;n)$ is the "characteristic function" of a random variable x with probability density $p(x;n)$. The Fourier transform $\hat{f}(k)$ of $f(x)$ is sometimes called the "structure function" of the RW. The Taylor expansion of $\hat{p}(k;n)$ around $k=0$ yields

$$\begin{aligned}\hat{p}(k;n) =& \langle e^{-ikx}\rangle \\ =& 1 - ik\langle x\rangle - \frac{1}{2}k^2\langle x^2\rangle + O(k^3)\,,\end{aligned} \tag{4}$$

where $\langle\cdot\rangle$ is the expectation unless we state otherwise. One can thereby obtain moments of $p(x;n)$ from the derivatives of $\hat{p}(k;n)$ at $k=0$.

The Fourier transform maps a convolution, such as Eq. (1), to a product; and Eq. (1) thus yields

$$\hat{p}(k;n) = \hat{f}(k)\hat{p}(k;n-1)\,. \tag{5}$$

If a random walker is located initially at $x = 0$, we obtain $p(x; 0) = \delta(x)$, where $\delta(x)$ is the Dirac delta function, which has Fourier transform $\hat{p}(k; 0) = 1$. We thereby obtain

$$\hat{p}(k; n) = \left[\hat{f}(k)\right]^n . \tag{6}$$

Using the inverse Fourier transform in Eq. (3), we obtain a formal solution for $p(x; n)$ in the time domain:

$$p(x; n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\hat{f}(k)\right]^n e^{ikx} \mathrm{d}k \, . \tag{7}$$

The qualitative behavior of the solution in Eq. (7) depends on the details of the structure function $\hat{f}(k)$. However, the asymptotic behavior of the RW as $n \to \infty$ depends only on some of the properties of $\hat{f}(k)$. When the first two moments of $\hat{f}(k)$ are finite, the solution converges to the Gaussian profile

$$p(x; n) = \frac{1}{(2\pi Dn)^{1/2}} e^{-\frac{(x-vn)^2}{4Dn}} , \tag{8}$$

where $v \equiv \langle r \rangle$ and $D \equiv \langle (r - \langle r \rangle)^2 \rangle / 2$. Equation (8) implies that the variance of x grows linearly with time. This result is the "central limit theorem" for the sum of the sizes of the moves, which are independent random variables. This asymptotic regime is well-defined because the underlying space (i.e., the line) is infinitely large. One can derive these results in a similar manner when the underlying space is discrete (e.g., a one-dimensional lattice) [2, 4, 30, 31]. In situations in which the second moment of the structure function diverges, the process exhibits superdiffusion and the probability profile converges to so-called "Lévy distributions" [34, 35].

2.2. Continuous time

In this section, we consider continuous-time RWs (CTRWs), which incorporate the timing of moves [4, 5, 30, 34, 35, 96]. We assume that a walker waits betweegn two moves for a duration τ that independently obeys the probability density function $\psi(\tau)$. In other words, the move events are generated by a renewal process [3]. If $\tau = 1$ with probability 1, the CTRW reduces to the DTRW described in Section 2.1. In a standard CTRW, one assumes that the time of a move event and the selection of a destination in a given move are independent. Therefore, a combination of $\psi(\tau)$ and $f(r)$, where r is the displacement in a single move, completely determines the dynamical properties of a random walker.

Let t_n denote the time of the nth move. By definition, $t_n = \sum_{i=1}^{n} \tau_i$, where each τ_i is independent and identically distributed (i.i.d.) and drawn from some distribution $\psi(\tau)$. Additionally, we can write

$$p(x; t) = \sum_{n=0}^{\infty} p(x; n) p(n, t) \, , \tag{9}$$

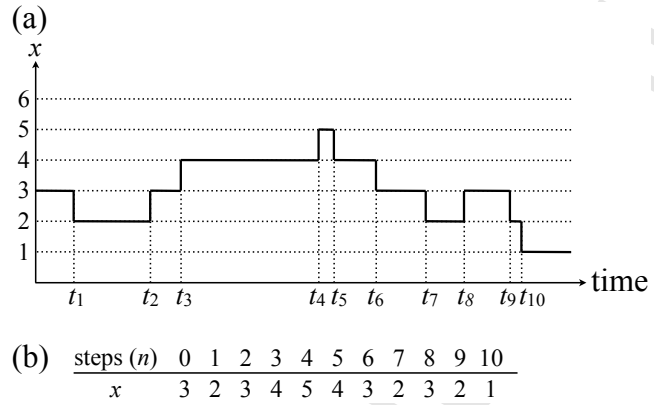

Figure 1: Schematic of the standard continuous-time random walk (CTRW) on a one-dimensional lattice. (a) The position x of the walker in physical time t is described by $p(x;t)$. Note that t_n represents the time of the nth move. (b) The position of the walker after n moves is described by $p(x;n)$.

where $p(x;t)$ is the probability that the walker is located at x at time t, the quantity $p(x;n)$ is the probability that the walker is located at x after n steps, and $p(n,t)$ is the probability density that the walker has moved n times at time t. Note that it is crucial to distinguish $p(x;t)$ and $p(x;n)$, and we illustrate the difference between these probabilities with a schematic in Fig. 1. Equation (9) reflects the fact that a walker can visit x at time t after some number n of steps.

The probability $p(x;n)$ is given by the same solution, Eq. (7), as for the DTRW. To obtain $p(x;t)$ from Eq. (9), we need to examine $p(n,t)$, and we thus need to consider a renewal process generated by $\psi(\tau)$. According to the elementary renewal theorem [97], the mean of n at time t is

$$\langle n \rangle = \frac{t}{\langle \tau \rangle} . \tag{10}$$

Equation (10) indicates that $n(t)$ grows linearly with time on average, irrespective of the details of the distribution $\psi(\tau)$. However, realized values of n are random, inducing heterogeneity in the length of the RW "trajectory" (i.e., the walk measured in terms of the number of moves) observed at a given time t.

When the CTRW is driven by a Poisson process, $\psi(\tau)$ is the exponential distribution (i.e., $\psi(\tau) = \beta e^{-\beta\tau}$). In this case, n obeys the Poisson distribution with mean βt. That is,

$$p(n,t) = \frac{(\beta t)^n}{n!} e^{-\beta t} . \tag{11}$$

It requires some effort to derive $p(n,t)$ when $\psi(\tau)$ is a general distribution. To calculate the time of the nth event or the number of events in a given time interval, we need to sum i.i.d. variables that obey $\psi(\tau)$. The duration $\tau \geq 0$ is

nonnegative, so we take a Laplace transform

$$\hat{\psi}(s) = \int_0^\infty \psi(\tau) e^{-s\tau} \mathrm{d}\tau \equiv \langle e^{-s\tau} \rangle \,. \tag{12}$$

The Taylor expansion of Eq. (12) is given by

$$\hat{\psi}(s) = \sum_{n=0}^{\infty} (-1)^n \frac{\langle \tau^n \rangle s^n}{n!} \tag{13}$$

and implies that $\hat{\psi}(s)$ generates the moments of $\psi(\tau)$ if they exist. One computes the inverse Laplace transform by integrating in the complex plane:

$$\psi(\tau) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \hat{\psi}(s) e^{s\tau} \mathrm{d}s \,, \tag{14}$$

where c is a real constant that is larger than the real part of all singularities of $\hat{\psi}(s)$.

The probability that no event has occurred up to time t is

$$p(0,t) = \int_t^\infty \psi(t') \mathrm{d}t' \,, \tag{15}$$

whose Laplace transform is

$$\hat{p}(0,s) = \frac{1 - \hat{\psi}(s)}{s} \,. \tag{16}$$

The probability that one event occurs in $[0,t]$ is

$$p(1,t) = \int_0^t \psi(t') p(0, t-t') \mathrm{d}t' \,. \tag{17}$$

By Laplace-transforming Eq. (17) and applying Eq. (16), we obtain

$$\hat{p}(1,s) = \hat{\psi}(s) \frac{1 - \hat{\psi}(s)}{s} \,. \tag{18}$$

By the same arguments, the probability density that n events occur at times t_1, t_2, …, t_n but at no other times in $[0,t]$ is given by $\psi(t_1)\psi(t_2 - t_1) \cdots \psi(t_n - t_{n-1}) p(0, t - t_n)$. This yields [97, 98]

$$\hat{p}(n,s) = \left[\hat{\psi}(s) \right]^n \frac{1 - \hat{\psi}(s)}{s} \,. \tag{19}$$

In the analysis of RWs, Eq. (19) relates two ways to count time: one is in terms of the number of moves (n), and the other is in terms of the physical time (t).

For a CTRW driven by a Poisson process, we obtain

$$\hat{\psi}(s) = \int_0^\infty \beta e^{-\beta\tau} e^{-s\tau} \mathrm{d}\tau = \frac{\beta}{s + \beta} \,. \tag{20}$$

Substituting Eq. (20) into Eq. (19) yields

$$\hat{p}(n,s) = \left(\frac{\beta}{s+\beta}\right)^n \frac{1}{s+\beta}\,. \tag{21}$$

By taking the Fourier transform of Eq. (9) with respect to x and the Laplace transform of Eq. (9) with respect to t and then using Eqs. (6) and (19), we obtain

$$\hat{p}(k;s) = \hat{p}(k;n)\hat{p}(n,s) \tag{22}$$

$$= \frac{1-\hat{\psi}(s)}{s} \sum_{n=0}^{\infty} \hat{f}(k)^n \hat{\psi}(s)^n$$

$$= \frac{1-\hat{\psi}(s)}{s} \frac{1}{1-\hat{f}(k)\hat{\psi}(s)}\,. \tag{23}$$

This result is central to the theory of CTRWs [96], and we will extend it to the case of general networks in Section 3.3. Taking the inverse transform of Eq. (23) with respect to both time and space yields $p(x;t)$, and we can examine the behavior of the RW for large t by expanding $\hat{p}(k;s)$ or $\hat{p}(x;s)$ for small s.

3. Random walks on networks

3.1. Notation

For our discussions, we assume that our networks are finite. However, to estimate how certain quantities scale with the number N of nodes, we sometimes examine the $N \to \infty$ limit. We allow our networks to have self-edges and multi-edges. We assume that the edge weights are nonnegative, so our networks are unsigned. For now, we assume that our networks are ordinary graphs (i.e., the best-studied types of networks), but we will consider multilayer networks in Section 4.1 and temporal networks in Section 4.2. Because introducing edge weights does not usually complicate RW problems, we assume that our networks are weighted unless we state otherwise, and we consider unweighted networks to be a special case of weighted networks. We also assume that our networks are directed unless we state otherwise. We summarize our main notation in Table 1.

An undirected network is called "regular" if all nodes have the same degree. Notably, many mathematical results for RWs on networks are restricted to regular graphs [1, 42, 99]. In this review, we are interested in networks with heterogeneous degree distributions, which tend to be the norm rather than the exception in empirical networks in numerous domains [100].

In our discussions, we assume that undirected networks are connected networks and that directed networks are "weakly connected" (i.e., that they are connected when one ignores the directions of the edges). It is clear (in the absence of jumps such as "teleportation" [23] to augment the RW) that a random walker is confined in the component in which it starts, and the analysis of RWs is then reduced to analysis within each component. See [44] for extensive discussions of components and weakly connected components.

Table 1: Main notation.

N	number of nodes
M	number of edges
v_i	the ith node (where $i \in \{1, \dots, N\}$)
A	The $N \times N$ weighted adjacency matrix of the network; the matrix component $A_{ij} \geq 0$ represents the weight of the edge from node v_i to node v_j. In an undirected network, $A_{ij} = A_{ji}$ (where $i, j \in \{1, \dots, N\}$). In an unweighted network, $A_{ij} \in \{0, 1\}$ (again with $i, j \in \{1, \dots, N\}$).
L	combinatorial Laplacian matrix
L'	RW normalized Laplacian matrix
s_i	The strength of node v_i in an undirected network; it is defined by $s_i \equiv \sum_{j=1}^{N} A_{ij} = \sum_{j=1}^{N} A_{ji}$. In an undirected and unweighted network, s_i is equal to the degree of v_i, which we denote by k_i.
s_i^{in}	In-strength of v_i; it is defined by $s_i^{\text{in}} = \sum_{j=1}^{N} A_{ji}$. In an unweighted network, s_i^{in} is equal to the in-degree of v_i, which we denote by k_i^{in}.
s_i^{out}	Out-strength of v_i; it is defined by $s_i^{\text{out}} = \sum_{j=1}^{N} A_{ij}$. In an unweighted network, s_i^{out} is equal to the out-degree of v_i, which we denote by k_i^{out}.
$\langle k \rangle$	mean degree, which is given by $\langle k \rangle = \sum_k kp(k)$ and indicates the sample mean of the degree for a network
D	The $N \times N$ diagonal matrix whose (i,i)th element is equal to s_i^{out} (where $i \in \{1, \dots, N\}$). In an undirected network, the (i,i)th element of D is equal to s_i.
n	discrete time
t	continuous time
p_i	probability that a random walker visits v_i
p_i^*	stationary density of a random walker at v_i
$\approx$	approximately equal to
$\propto$	proportional to

3.2. Discrete time

3.2.1. Definition and temporal evolution

Consider a DTRW on a directed network. We suppose that there is a single walker, which moves during each time step. When the walker is located at v_i, it moves to the out-neighbor v_j with a probability proportional to A_{ij}. The transition-probability matrix T has elements T_{ij}, which give the probability that the walker moves from v_i to v_j, of

$$T_{ij} = \frac{A_{ij}}{s_i^{\text{out}}}\,, \tag{24}$$

where we assume that $s_i^{\text{out}} > 0$. Other choices of T, informed by the adjacency matrix A, are also possible. One example is a "degree-biased RW" in unweighted (and usually undirected) networks [101–106]; in this case, $T_{ij} \propto k_j^{\alpha}$, where α is a constant. If $A_{ij} = A_{ji} = (k_i k_j)^{\alpha}$, then T given by Eq. (24) gives this degree-

biased RW. Another example of a biased transition-probability matrix T is a "maximum entropy RW" [107–111].

Because a random walker must go somewhere — including perhaps the current node — in a given move, the following conservation condition holds:

$$\sum_{j=1}^{N} T_{ij} = 1\,. \tag{25}$$

A DTRW on a finite network is a Markov chain on N states. There is a huge literature (both pedagogical and more advanced) on Markov chains in general and for RWs in particular. This is especially true for finite state spaces (corresponding to finite networks) and for stationary Markov chains in which the transition probability does not depend on discrete time n [1, 112–120]. We draw from this literature to explain several properties of DTRWs in the rest of this section.

Let $p_i(t)$ denote the probability that node v_i is visited at discrete time n. This probability evolves according to

$$p_j(n+1) = \sum_{i=1}^{N} p_i(n) T_{ij} \quad (j \in \{1, \ldots, N\})\,. \tag{26}$$

Additionally,

$$\sum_{i=1}^{N} p_i(n) = 1 \tag{27}$$

for any n if Eq. (27) holds for $n = 0$. Equation (26) is equivalent to

$$\boldsymbol{p}(n+1) = \boldsymbol{p}(n) T\,, \tag{28}$$

where $\boldsymbol{p}(t) = (p_1(n)\,, \ldots\,, p_N(n))$. From Eq. (28), we see that

$$\boldsymbol{p}(n) = \boldsymbol{p}(0) T^n\,. \tag{29}$$

3.2.2. Stationary density

Consider the stationary density (i.e., the so-called "occupation probability") $\boldsymbol{p}^* = (p_1^*, \ldots, p_N^*)$, where $p_i^* = \lim_{n\to\infty} p_i(n)$ (with $i \in \{1, \ldots, N\}$). Substituting $p_i(n) = p_i(n+1) = p_i^*$ into Eq. (28) yields

$$\boldsymbol{p}^* = \boldsymbol{p}^* T\,. \tag{30}$$

Therefore, the stationary density is the left eigenvector of T with eigenvalue 1. The corresponding right eigenvector is $(1\ , \ldots,\ 1)^\top$, where $\top$ represents transposition.

For a directed network that is "strongly connected" (i.e., a walker can travel from any node v_i to any other node v_j along directed edges [44]), $\boldsymbol{p}^*$ is unique. In undirected networks, one just needs a network to be connected, which we have assumed.

In undirected networks, we obtain the central result

$$p_i^* = \frac{s_i}{\sum_{\ell=1}^{N} s_\ell} \quad (i \in \{1, \ldots, N\}) , \tag{31}$$

which one can verify by substituting Eq. (31) into Eq. (30). For unweighted networks, Eq. (31) reduces to $p_i^* = k_i/2M$. Regardless of other structural properties of a network, the stationary density is determined solely by strength (and thus by degree for unweighted networks). Equation (31) also holds for directed networks that satisfy $s_i \equiv s_i^{\text{in}} = s_i^{\text{out}}$ (where $i \in \{1, \ldots, N\}$). Such directed networks are sometimes called "balanced" [1].

In undirected networks,

$$p_i^* T_{ij} = p_j^* T_{ji} . \tag{32}$$

In other words, for each edge, the flow of probability in each direction must equal each other at equilibrium. This property, called "detailed balance" in statistical physics [121] and "time reversibility" in mathematics [1, 42], does not generally hold for directed networks.

Let's consider a generalization of the degree-biased RW to weighted networks (i.e., a strength-biased RW) in which the probability that a random walker located at node v_i or v_j traverses the edge (v_i, v_j) is proportional to $(s_i s_j)^\alpha$. It follows that

$$T_{ij} = \frac{(s_i s_j)^\alpha}{\sum_{\ell=1}^{N} (s_i s_\ell)^\alpha} = \frac{s_j^\alpha}{\sum_{\ell; v_\ell \in \mathcal{N}_i} s_\ell^\alpha} , \tag{33}$$

where $\mathcal{N}_i$ is the neighborhood of v_i. A strength-biased RW is equivalent to an RW on a modified undirected network whose weighted adjacency matrix is given by $A'_{ij} = (s_i s_j)^\alpha$ (see Fig. 2 for an example). The strength of node v_i in this modified network is given by $s'_i = \sum_{j=1}^{N} A'_{ij} = s_i^\alpha \sum_{j; v_j \in \mathcal{N}_i}^{N} s_j^\alpha$. By substituting s'_i into Eq. (31) in place of s_i, we obtain the stationary density

$$p_i^* = \frac{s_i^\alpha \sum_{v_j \in \mathcal{N}_i} s_j^\alpha}{\sum_{i'=1}^{N} s_{i'}^\alpha \sum_{v_{j'} \in \mathcal{N}_{i'}} s_{j'}^\alpha} . \tag{34}$$

For an unweighted network constructed using a "configuration model" [122], a standard model of random networks, we obtain $p_i^* \approx k_i^{\alpha+1} / \sum_{\ell=1} k_\ell^{\alpha+1}$ [123–125]. In particular, we obtain $p_i^* = 1/N$ for all nodes when $\alpha = -1$. Therefore, in general, we expect that a node with a large strength tends to have a large p_i^* when $\alpha > -1$ (including for the unweighted case $\alpha = 0$) and that the same node tends to have a small p_i^* when $\alpha < -1$. For nodes with a large strength, we expect p_i^* to increase as α increases.

For directed networks in general, one can write a first-order approximation to the stationary density from Eq. (30). We assume that we do not possess any information about the neighbors of v_i, so we replace p_j^* and s_j^{out} by their mean values:

$$p_i^* = \sum_{j=1}^{N} p_j^* \frac{A_{ji}}{s_j^{\text{out}}} \approx (\text{const}) \times \sum_{j=1}^{N} A_{ji} \propto s_i^{\text{in}} . \tag{35}$$

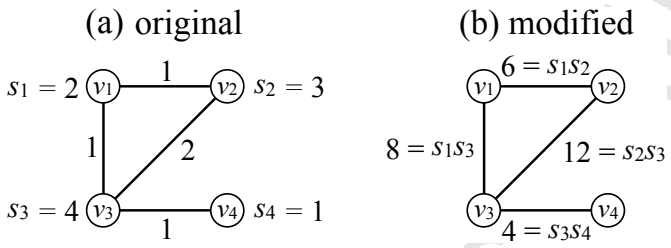

Figure 2: Strength-biased RW. (a) An original undirected network, whose weighted adjacency matrix is given by A. (b) The modified undirected network, whose weighted adjacency matrix is given by A'. The numbers attached to the edges represent the edge weight. We set $\alpha = 1$.

On both synthetic and empirical networks, Eq. (35) is reasonably accurate in some cases but not in others [126–133].

3.2.3. Relaxation time

To determine the relaxation time to the stationary state, it is instructive to project the solution, Eq. (29), onto an appropriate basis of vectors and to represent it in terms of its modes. The procedure, which is analogous to taking a Fourier transform [see Eq. (2)], is sometimes called a "graph Fourier transform" [134, 135] and will be explained in this section [see Eqs. (43)–(45)].

For simplicity, we consider undirected networks. In general, the transition probability matrix T is asymmetric even for undirected networks, except for regular graphs. However, one can derive its eigenvalues and eigenvectors from those of the symmetric matrix

$$\tilde{A}_{ij} = \frac{A_{ij}}{\sqrt{s_i s_j}}\,, \tag{36}$$

which we can decompose as follows:

$$\tilde{A}_{ij} = \sum_{\ell=1}^{N} \lambda_\ell \boldsymbol{u}_\ell \boldsymbol{u}_\ell^\top\,, \tag{37}$$

where λ_ℓ is the ℓth eigenvalue of $\tilde{A}$ and $\boldsymbol{u}_\ell$ is the corresponding normalized eigenvector (so that $\langle \boldsymbol{u}_\ell, \boldsymbol{u}_{\ell'} \rangle = \delta_{\ell\ell'}$, where $\langle\,,\rangle$ is the inner product), and δ is the Kronecker delta. Because $\tilde{A}$ is symmetric, each eigenvalue λ_ℓ is real.

Because $T_{ij} = \sqrt{s_j}\tilde{A}_{ij}/\sqrt{s_i}$, we have the following similarity relationship between T and A [1, 136]:

$$T = D^{-1/2}\tilde{A}D^{1/2}\,, \tag{38}$$

where we defined D (a matrix whose nonzero entries lie only on the diagonal) in Section 3.1. Equation (38) implies that T and $\tilde{A}$ have the same eigenvalues. In particular, all eigenvalues of T are real-valued, because that is the case for

$\tilde{A}$. The left and right eigenvectors of T corresponding to the eigenvalue λ_ℓ are, respectively,

$$\boldsymbol{u}_\ell^{\mathrm{L}} = \boldsymbol{u}_\ell^\top D^{1/2} = ((u_\ell)_1\sqrt{s_1},\ \dots\ ,(u_\ell)_N\sqrt{s_N}) \tag{39}$$

and

$$\boldsymbol{u}_\ell^{\mathrm{R}} = D^{-1/2}\boldsymbol{u}_\ell = ((u_\ell)_1/\sqrt{s_1},\ \dots\ ,(u_\ell)_N/\sqrt{s_N})^\top\ . \tag{40}$$

One can verify Eqs. (39) and (40) using Eq. (38) and the relation $\tilde{A}\boldsymbol{u}_\ell = \lambda_\ell \boldsymbol{u}_\ell$. Using

$$\begin{aligned} T^n &= D^{-1/2}\tilde{A}^n D^{1/2} \\ &= D^{-1/2}\sum_{\ell=1}^{N}\lambda_\ell^n \boldsymbol{u}_\ell \boldsymbol{u}_\ell^\top D^{1/2} \\ &= \sum_{\ell=1}^{N}\lambda_\ell^n \boldsymbol{u}_\ell^{\mathrm{R}}\boldsymbol{u}_\ell^{\mathrm{L}}\,, \end{aligned} \tag{41}$$

we obtain the following mode expansion of the solution of the RW:

$$\boldsymbol{p}(n) = \boldsymbol{p}(0)T^n = \sum_{\ell=1}^{N}\lambda_\ell^n \boldsymbol{u}_\ell^{\mathrm{L}}\langle \boldsymbol{p}(0), \boldsymbol{u}_\ell^{\mathrm{R}}\rangle\,. \tag{42}$$

That is,

$$p_i(n) = \sum_{\ell=1}^{N} a_\ell(n)(u_\ell^{\mathrm{L}})_i\,, \tag{43}$$

where

$$a_\ell(n) = \lambda_\ell^n a_\ell(0)\,, \tag{44}$$

$$a_\ell(0) \equiv \langle \boldsymbol{p}(0), \boldsymbol{u}_\ell^{\mathrm{R}}\rangle\,, \tag{45}$$

and $a_\ell(n)$ is the projection onto the ℓth eigenmode. Equations (43)–(45) map the state vector $\boldsymbol{p}(n)$, which is defined on the nodes, to a vector $(a_1(n),\ \dots\ ,a_N(n))$ of eigenvector amplitudes (i.e., their coefficients). This transform, called the "graph Fourier transform", generalizes the standard Fourier transform of an RW [see Eqs. (3) and (7)], and the eigenvectors of the transition-probability matrix T play the role of the Fourier modes e^{ikx}.

For the matrix T and $\tilde{A}$, the eigenvalues λ_ℓ each satisfy $-1 \le \lambda_\ell \le 1$ [1, 42]. Except in the special cases of multipartite graphs, the strict inequality $\lambda_\ell > -1$ also holds. In this case, the mode with $\lambda_\ell = 1$ corresponds to the stationary density, and we thus write $\boldsymbol{u}_\ell^{\mathrm{L}} = \boldsymbol{p}^*$. The right eigenvector that corresponds to this mode is $\boldsymbol{u}_\ell^{\mathrm{R}} \propto (1,\ \dots\ ,1)^\top$. All modes for which $-1 < \lambda_\ell < 1$ decay to 0. The eigenvalue $\lambda_\ell = 1$ is the largest-magnitude eigenvalue, and the Perron–Frobenius theorem guarantees that all elements of $\boldsymbol{u}_\ell^{\mathrm{L}}$ and $\boldsymbol{u}_\ell^{\mathrm{R}}$ are positive.

Similar results hold for directed networks, although we cannot take advantage of the symmetric structure of the matrix $\tilde{A}$ in general. In directed networks, the eigenvalues satisfy $|\lambda_\ell| \leq 1$. When $|\lambda_\ell| < 1$ holds for all but one eigenvalue, which is the case except for directed variants of multipartite graphs with an even number of components, the mode with $\lambda_\ell = 1$ corresponds to the stationary density. In this case, we obtain $\boldsymbol{u}_\ell^{\mathrm{L}} = \boldsymbol{p}^*$ and $\boldsymbol{u}_\ell^{\mathrm{R}} \propto (1, \ \ldots \ , 1)^\top$. Again, the Perron–Frobenius theorem guarantees that all elements of $\boldsymbol{u}_\ell^{\mathrm{L}}$ are positive.

By letting $n \to \infty$ in Eq. (42), we obtain $\boldsymbol{p}^* = \boldsymbol{u}_{\max}^{\mathrm{L}} \langle \boldsymbol{p}(0), \boldsymbol{u}_{\max}^{\mathrm{R}} \rangle$, where the subscript "max" indicates the mode corresponding to the dominant eigenvalue (which is equal to 1). Because $\boldsymbol{u}_{\max}^{\mathrm{R}} \propto (1, \ \ldots \ , 1)^\top$, it follows that $\langle \boldsymbol{p}(0), \boldsymbol{u}_{\max}^{\mathrm{R}} \rangle = 1$ regardless of the initial condition $\boldsymbol{p}(0)$. This is consistent with the fact that $\boldsymbol{u}_{\max}^{\mathrm{L}}$ gives the stationary density. By letting n be large but finite, we obtain

$$\boldsymbol{p}(n) \approx \boldsymbol{u}_{\max}^{\mathrm{L}} \langle \boldsymbol{p}(0), \boldsymbol{u}_{\max}^{\mathrm{R}} \rangle + \lambda_2^n \boldsymbol{u}_2^{\mathrm{L}} \langle \boldsymbol{p}(0), \boldsymbol{u}_2^{\mathrm{R}} \rangle \, , \tag{46}$$

where λ_2 is the second-largest (in magnitude) eigenvalue of T. In deriving Eq. (46), we only kept two terms, because $|\lambda_\ell|^n \ll |\lambda_2|^n$ for all eigenvalues λ_ℓ with $\ell > 2$, assuming that $|\lambda_\ell| < |\lambda_2|$ (where $\ell \in \{3, \ldots, N\}$). Equation (46) indicates that the second-largest eigenvalue of T governs the relaxation time. More generally, the relaxation speed is determined by the ratio between $|\lambda_2|$ and $\lambda_{\max} = 1$. The difference $1 - \lambda_2$ is often called the "spectral gap". A large spectral gap (i.e., a small-magnitude for λ_2) entails fast relaxation.

The "Cheeger inequality" gives useful bounds on λ_2 [137]. The "Cheeger constant", which is also called "conductance", is defined by

$$h = \min_S \left\{ \frac{(\text{number of edges that connect } S \text{ and } \overline{S})}{\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}} \right\} , \tag{47}$$

where S is a set of nodes in a network, $\overline{S}$ is the complementary set of the nodes (i.e., $S \cap \overline{S} = \emptyset$ and $S \cup \overline{S}$ is the complete set of the N nodes), and $\mathrm{vol}(S) \equiv \sum_{i=1; v_i \in S}^{N} s_i$. In the minimization in Eq. (47), we seek a bipartition of a network such that the two parts are the most sparsely connected. (In other words, we want a minimum cut.) The denominator in the right-hand side of Eq. (47) prevents the selection of a very uneven bipartition, which would easily yield a small value for the numerator. The Cheeger inequality is

$$\frac{h^2}{2} < 1 - |\lambda_2| \leq 2h \, , \tag{48}$$

so a small Cheeger constant h implies a small spectral gap $1 - |\lambda_2|$ and hence slower relaxation. This result is intuitive, because one can partition a network with a small value of h into two well-separated communities such that it is difficult for random walkers to cross from one community to the other. Note that there are various versions of Cheeger constants and inequalities. They give qualitatively similar — but quantitatively different — results [1, 42, 54, 138–140]. As discussed in Ref. [68] and references therein, such results are important considerations for community detection.

A fact related to the relaxation time is that the power method is a practical method to calculate the stationary density of an RW in a directed network [141]. Suppose that we start with an arbitrary initial vector $\boldsymbol{p}(0)$, excluding one that is orthogonal to $\boldsymbol{p}^*$, and repeatedly left-multiply it by T. After many iterations, we obtain an accurate estimate of $\boldsymbol{p}^*$. Because any $\boldsymbol{p}(0)$ that is orthogonal to $\boldsymbol{p}^*$ includes a negative entry, one can start iterations with any probability vector $\boldsymbol{p}(0)$. In practice, one may have to normalize $\boldsymbol{p}(n)$ after each iteration (or after some number of iterations) to avoid the elements of $\boldsymbol{p}(n)$ becoming too large or small.

3.2.4. Exit probability

One is often interested in the probability that a random walker terminates at a particular node, which is then called an "absorbing state". Upon reaching an absorbing state, a stochastic process cannot escape from it. A node v_i is "absorbing" if and only if $T_{ii} = 1$, which implies that $T_{ij} = 0$ (for $j \neq i$). A set of nodes is an "ergodic" set if (1) it is possible to go from v_i to v_j for any nodes in the set and (2) the process does not leave the set once it has been reached. An absorbing node is an ergodic set that consists of a single node. A state in a Markov chain is said to be a "transient state" if it does not belong to an ergodic set.

When an RW is composed of N_1 transient-state nodes and N_2 absorbing-state nodes, there are $N_1 + N_2 = N$ nodes in total. Without loss of generality, we relabel the nodes such that v_1, ..., v_{N_1} are transient and v_{N_1+1}, ..., v_N are absorbing. The transition-probability matrix T then has the following form:

$$T = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} , \tag{49}$$

where Q is an $N_1 \times N_1$ matrix that describes transitions between transient-state nodes, R is an $N_1 \times N_2$ matrix that describes transitions from transient-state nodes to absorbing-state nodes, and I is the $N_2 \times N_2$ identity matrix that corresponds to individual absorbing-state nodes. Taking powers of Eq. (49) yields

$$T^n = \begin{pmatrix} Q^n & R + QR + \cdots + QR^{n-1} \\ 0 & I \end{pmatrix} . \tag{50}$$

Suppose that we start from transient-state node v_i and want to calculate the mean number of visits to transient-state node v_j before reaching an absorbing-state node. This number of visits is equal to the (i, j)th element of the matrix

$$W = \sum_{n=0}^{\infty} Q^n = (I - Q)^{-1} , \tag{51}$$

because the (i, j)th element of Q^n is equal to the probability that a random walker starting from v_i visits v_j at discrete time n. The matrix W is called the "fundamental matrix" associated with Q. The matrix on the right-hand side of

Eq. (51) is called the "resolvent" of Q. Similar considerations arise in the study of "central" (i.e., important) nodes in networks [142].

The "exit probability" (i.e., the "first-passage-time probability") is defined as the probability U_{ij} that the walker terminates at an absorbing state v_j when it starts from a transient state v_i. When there are multiple absorbing-state nodes, it is nontrivial to determine the exit probability. The probability that the walker reaches v_j after exactly n steps is given by the (i,j)th element of $Q^{n-1}R$. Therefore, we obtain the exit probability in matrix form as follows:

$$U = \sum_{n=1}^{\infty} Q^{n-1}R = WR\,. \tag{52}$$

3.2.5. Mean first-passage and recurrence times

When does a random walker starting from a certain source node arrive at a target node for the first time? The answer to this question is known as the "first-passage time" (or "first-hitting time") if the source and target nodes are different and is known as the "recurrence time" (or the "first-return time") when the source and target nodes are identical. Let m_{ij} (with $i \neq j$) denote the mean first-passage time (MFPT) from node v_i to node v_j. The mean recurrence time is m_{ii}. For directed networks, we assume strongly connected networks throughout this section to guarantee that $m_{ij} < \infty$ (for $i, j \in \{1, \ldots, N\}$). For reviews on first-passage problems on networks and other media, see [31, 40].

General networks: Let's first consider some general results. The following identity holds [1, 112, 113, 115]:

$$m_{ij} = 1 + \sum_{\ell=1;\ell\neq j}^{N} T_{i\ell} m_{\ell j}\,. \tag{53}$$

In its first step, a random walker moves from node v_i to node v_ℓ, which produces the 1 on the right-hand side of Eq. (53). If $\ell = j$, then the walk terminates at v_ℓ, resulting in a first-passage time of 1. Otherwise, we seek the first-passage from node v_ℓ (with $\ell \neq j$) to node v_j. This produces the second term on the right-hand side. Note that Eq. (53) is also valid when $i = j$.

In matrix notation, we write Eq. (53) as

$$M = J + T(M - M_{\mathrm{dg}})\,, \tag{54}$$

where $M = (m_{ij})$, all of the elements of the matrix J are equal to 1, and M_{dg} is the diagonal matrix whose diagonal elements are equal to m_{ii}. By left-multiplying Eq. (54) by $\boldsymbol{p}^*$ and using $\boldsymbol{p}^* J = (1, \ \ldots \ , 1)$ and $\boldsymbol{p}^* T = \boldsymbol{p}^*$, we obtain the mean recurrence time

$$m_{ii} = \frac{1}{p_i^*}\,. \tag{55}$$

Equation (55) is called "Kac's formula" [1, 118, 119].

There are several different ways to evaluate the MFPT m_{ij} (with $i \neq j$), and it is insightful to discuss different approaches.

One method is simply to iterate Eq. (53) [115].

A second method to calculate the MFPT, for a given j, is to rewrite Eq. (53) as

$$\overline{\boldsymbol{m}}^{(j)} = \mathbf{1} + \overline{T}^{(j)} \overline{\boldsymbol{m}}^{(j)} \,, \tag{56}$$

where $\overline{\boldsymbol{m}}^{(j)} = (m_{1j}, \ \ldots \ , m_{j-1,j} \ , m_{j+1,j} \ \ldots \ , m_{Nj})^{\top}$ and $\mathbf{1} = (1, \ \ldots \ , 1)^{\top}$ are $(N-1)$-dimensional column vectors and $\overline{T}^{(j)}$ is the $(N-1) \times (N-1)$ submatrix of T that excludes the jth row and jth column [124]. The formal solution of Eq. (56) is

$$\overline{\boldsymbol{m}}^{(j)} = \left(\overline{L}^{(j)} \right)^{-1} \overline{D}^{(j)} \mathbf{1} \,, \tag{57}$$

where $\overline{D}^{(j)}$ is the submatrix of D that excludes the jth row and jth column and $\overline{L}^{(j)} = \overline{D}^{(j)} - \overline{A}^{(j)}$, where $\overline{A}^{(j)}$ is the submatrix of A that excludes the jth row and jth column. The matrix $\overline{L}^{(j)}$ is sometimes called a "grounded Laplacian matrix" [143] (although it is not a Laplacian matrix), and it is invertible because we assumed strongly connected networks. One can derive and solve Eq. (57) separately for each j.

A third method to calculate the MFPT is to take advantage of relaxation properties of RWs [144]. Let $p_{ij}(n)$ denote the probability that a walker starting at node v_i visits node v_j after n moves. The master equation is

$$p_{ij}(n+1) = \sum_{\ell=1}^{N} p_{i\ell}(n) T_{\ell j} \,. \tag{58}$$

Let $F_{ij}(n)$ denote the probability that the walker starting from v_i arrives at v_j for the first time after n moves. We obtain

$$p_{ij}(n) = \delta_{n0}\delta_{ij} + \sum_{n'=0}^{n} F_{ij}(n') p_{jj}(n-n') \,. \tag{59}$$

Using a discrete-time Laplace transform (see, e.g., [145] for an extensive discussion of such generating functions), defined by

$$\hat{p}_{ij}(s) \equiv \sum_{n=0}^{\infty} e^{-sn} p_{ij}(n) \tag{60}$$

and

$$\hat{F}_{ij}(s) \equiv \sum_{n=0}^{\infty} e^{-sn} F_{ij}(n) \,, \tag{61}$$

we transform Eq. (59) to

$$\hat{p}_{ij}(s) = \delta_{ij} + \hat{F}_{ij}(s) \hat{p}_{jj}(s) \tag{62}$$

and thereby obtain

$$\hat{F}_{ij}(s) = \frac{\hat{p}_{ij}(s) - \delta_{ij}}{\hat{p}_{jj}(s)} \,. \tag{63}$$

Using Eq. (63) then yields

$$\begin{aligned} m_{ij} &= \sum_{n=0}^{\infty} n F_{ij}(n) = -\hat{F}'_{ij}(0) \\ &= \frac{-\hat{p}'_{ij}(0)\hat{p}_{jj}(0) + \hat{p}'_{jj}(0)\left[\hat{p}_{ij}(0) - \delta_{ij}\right]}{\hat{p}_{jj}(0)^2} \,. \end{aligned} \tag{64}$$

To evaluate Eq. (64), we define

$$R_{ij}^{(m)} \equiv \sum_{n=0}^{\infty} n^m \left[p_{ij}(n) - p_j^*\right] \,. \tag{65}$$

Equation (65) quantifies the relaxation speed at which $p_{ij}(n)$ approaches the stationary density. To write the Laplace transform, we multiply both sides of Eq. (65) by $(-1)^m s^m/m!$ and sum over m. We thereby obtain

$$\begin{aligned} \sum_{m=0}^{\infty} R_{ij}^{(m)} (-1)^m \frac{s^m}{m!} &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} n^m (-1)^m \frac{s^m}{m!} \left[p_{ij}(n) - p_j^*\right] \\ &= \sum_{n=0}^{\infty} e^{-sn} \left[p_{ij}(n) - p_j^*\right] \\ &= \tilde{p}_{ij}(s) - \frac{p_j^*}{1 - e^{-s}} \,. \end{aligned} \tag{66}$$

Substituting Eq. (66) into Eq. (63) then yields

$$\begin{aligned} \hat{F}_{ij}(s) &= \frac{\frac{p_j^*}{s+o(s)} + \sum_{m=0}^{\infty} R_{ij}^{(m)} (-1)^m \frac{s^m}{m!} - \delta_{ij}}{\frac{p_j^*}{s+o(s)} + \sum_{m=0}^{\infty} R_{jj}^{(m)} (-1)^m \frac{s^m}{m!}} \\ &= \frac{p_j^* + R_{ij}^{(0)} s - \delta_{ij} s + o(s)}{p_j^* + R_{jj}^{(0)} s + o(s)} \\ &= 1 + \frac{R_{ij}^{(0)} - R_{jj}^{(0)} - \delta_{ij}}{p_j^*} s + o(s) \,, \end{aligned} \tag{67}$$

where $o(s)$ represents a quantity that is much smaller than s in the relevant asymptotic limit ($s \to 0$ in the present case). Consequently,

$$m_{ij} = -\hat{F}'_{ij}(0) = \begin{cases} \frac{1}{p_j^*} & (j = i) \,, \\ \frac{R_{jj}^{(0)} - R_{ij}^{(0)}}{p_j^*} & (j \neq i) \,, \end{cases} \tag{68}$$

which is consistent with Kac's formula [see Eq. (55)]. For undirected networks, substituting $p_j^* = s_j / \sum_{\ell=1}^N s_\ell$ into Eq. (68) yields

$$m_{ij} = \begin{cases} \frac{\sum_{\ell=1}^N s_\ell}{s_j} & (j = i)\,, \\ \frac{\sum_{\ell=1}^N s_\ell}{s_j} \left(R_{jj}^{(0)} - R_{ij}^{(0)} \right) & (j \neq i)\,. \end{cases} \tag{69}$$

A fourth method to examine the MFPT is to estimate m_{ij} using a mean-field approximation [146–148]. Regardless of the source node v_i, the target node v_j is reached with an approximate probability of p_j^* in each time step. Therefore,

$$m_{ij} \approx \sum_{n=1}^{\infty} n p_j^* (1 - p_j^*)^{n-1} = \frac{1}{p_j^*} = m_{jj}\,. \tag{70}$$

Equation (70) is a rather coarse approximation, and m_{ij} can deviate considerably from $m_{jj} = 1/p_j^*$. More sophisticated mean-field approaches can likely do better, especially for networks with structures that are well-suited to the employed approximation.

There have been many studies of MFPTs for various network models using both analytical and numerical approaches [31, 149–151, 151–153]. We will discuss some examples of undirected and unweighted networks. We focus mainly on the MFPT between different nodes, although it is of course also interesting to calculate recurrence times.

Regular networks: For a complete graph, m_{ij} (with $i \neq j$) is independent of i and j because of the symmetry of the network. Therefore, Eq. (53) reduces to

$$m_{ij} = \frac{1}{N-1} + \frac{N-2}{N-1}(1 + m_{ij})\,, \tag{71}$$

which yields $m_{ij} = N - 1$ for $i \neq j$. Kac's formula [see Eq. (55)] implies that $m_{ii} = N$.

For regular lattices $\mathbb{Z}^d$ of any dimension d, Eq. (55) implies that $m_{ii} \propto N$ because $p_i^* \propto k_i = 2d$ for any i. Define $m_{\bullet j}$ to be the MFPT averaged over all source nodes v_i ($i \neq j$) [154]. For $\mathbb{Z}^d$, it satisfies the scalings $m_{\bullet j} \propto N^2$ for $d = 1$, $m_{\bullet j} \propto N \ln N$ for $d = 2$, and $m_{\bullet j} \propto N$ for $d = 3$.

Erdős–Rényi (ER) random graphs: Consider an ER random graph $G(N, p)$, where p denotes the (independent) probability that each node pair has an edge. Assuming that the mean degree $\langle k \rangle$ is kept constant (i.e., $p = \langle k \rangle / (N-1) \propto 1/N$), we obtain $m_{ii} \propto N$ and $m_{ij} \propto N^{3/2}$ (with $i \neq j$) as $N \to \infty$ [155] for the "giant component" (i.e., a largest connected component that scales linearly with the number N of network nodes as $N \to \infty$ [44]). Now suppose that we assume instead that $p > \ln N / N$, so that all nodes belong to a single component (in the $N \to \infty$ limit) and thus m_{ij} (for $i, j \in \{1, \ldots, N\}$) is well-defined. It then follows that m_{ij} averaged over all source and target nodes is equal to $N - 1$, independently of p [156, 157]. In other words, for a sufficiently dense ER random graph, the MFPT is the same as that for the complete graph. The MFPT is

much longer for directed ER graphs than for undirected ones, because random walkers do not backtrack on directed networks [158].

Other network models with random features: Much effort in studying RWs on networks has considered first-passage times on Watts–Strogatz (WS) small-world networks [149, 159–164]. As expected, given that WS networks interpolate between regular lattices and ER networks [3], these studies have found that the behavior of an RW on WS networks lies somewhere between that on a regular lattice and that on ER graphs.

Equation (69) has also been elaborated further for "scale-free" networks, which are defined as networks with a power-law degree distribution $p(k) \propto k^{-\gamma}$, where $p(k)$ is the degree distribution. Let's consider scale-free networks that are generated by a "configuration model" [122], so there are no degree–degree correlations. We examine the mean of the MFPT m_{ij} over the position of the source node v_i (with $i \neq j$), which we select according to the stationary density. We use $\tilde{m}_{\bullet j}$ to denote this weighted mean of the MFPT over i. This mean is distinct from the unweighted mean $m_{\bullet j}$. For scale-free networks constructed using a configuration model, we obtain for large N that [166]

$$\tilde{m}_{\bullet j} \propto \begin{cases} N^{2/d_\mathrm{s}} & (d_\mathrm{s} < 2)\,, \\ N k_j^{(1-2/d_\mathrm{s})(\gamma-1)} & (2 < d_\mathrm{s} < 2(\gamma-1)/(\gamma-2))\,, \\ N k_j^{-1} & (d_\mathrm{s} > 2(\gamma-1)/(\gamma-2))\,, \end{cases} \tag{72}$$

where $d_\mathrm{s} \equiv 2d_\mathrm{f}/d_\mathrm{w}$ is the "spectral dimension" of the network; the "fractal dimension" d_f is defined as the exponent of the scaling relation $N_r \propto r^{d_\mathrm{f}}$, where N_r is the number of nodes within distance r from a source node; and the "walk dimension" d_w is defined from the scaling relation $\langle r^2 \rangle \propto t^{2/d_\mathrm{w}}$, where r is the distance between the current position of the walker and the source node [36, 39]. In practice, one calculates the walk dimension as the scaling exponent for the time t_exit for a random walker to exit from a sphere of radius r from the source node (so that $t_\mathrm{exit} \propto r^{d_\mathrm{w}}$) [167]. For regular lattices, $d_\mathrm{w} = 2$, and the diffusion is thus called "normal". If $d_\mathrm{w} \neq 2$, the diffusion is called "anomalous" [39]. For the "compact exploration" case of $d_\mathrm{s} < 2$, Eq. (72) suggests that the asymptotic scaling of $\tilde{m}_{\bullet j}$ with N does not depend on the target node at leading order. However, if $d_\mathrm{s} > 2$ (the second and the third cases in Eq. (72)), nodes with higher degrees are reached faster. In particular, for networks that satisfy the "small-world property" (i.e., the mean path length between nodes scales proportionally to $\ln N$ or even more slowly) [165], including popular scale-free network models (such as ones generated by a configuration model), one obtains $d_\mathrm{s} = \infty$ (and d_s is very large for many empirical networks). Therefore, the third case in Eq. (72) applies.

Fractal and pseudo-fractal networks: There are various deterministic mechanisms to grow networks in a recursive manner. Depending on the mode, these

[3] Technically, it is a variant of WS networks with edge rewiring (rather than edge addition) that interpolates between regular lattices and ER networks [165].

Table 2: The term "hierarchical network" has been used (sometimes in a misleading way) to describe various network structures. To help readers, we provide a short summary of three common uses.

Hierarchical modularity	A hierarchical network can indicate the presence of "hierarchical modularity", in which dense modules are themselves composed of dense submodules in the recursive manner of a "Russian doll" [174].
Status theory	One can also understand a hierarchy in the context of "status theory", in which certain nodes have a higher status than others, and a directed edge indicates a difference of status [175]. This notion leads naturally to trees that are dominated by a root and, more generally, to acyclic networks [176].
Pseudo-fractal networks	Some models of pseudo-fractal networks are sometimes called hierarchical networks. Ravasz and Barabási proposed to characterize such "hierarchical" structure by examining a scaling relation between clustering coefficient and node degree [169, 170].

algorithms yield "pseudo-fractal" scale-free networks [168] (also called "hierarchical networks" [169, 170] or "transfractals" [171]; see Table 2 for different meanings of the term "hierarchical network" that exist in the literature), which have a highly symmetric structure and satisfy the small-world property; fractal networks that do not satisfy the small-world property [171–173]; or classical fractals [39]. These objects are defined and studied in the limit $N \to \infty$. For such models, it is often possible to exploit their deterministic and recursive nature to exactly calculate the MFPT, and generating functions again can be helpful.

Let's start by looking at fractals that do not have a heavy-tailed degree distribution. In a recursive process of generating a fractal structure from a model of a fractal, we stop the process in each iteration and regard any intersection with more than one edges as a node. In this way, we define a network corresponding to each iteration. The recursive process generates a series of networks, where the number N of nodes becomes larger as one iterates further. We are interested in how the MFPT scales in such networks as a function of N. For example, consider a network constructed from the Sierpinski gasket [177]. When the target node is located at the apex of the gasket, the MFPT averaged over a uniform distribution of the source node is $m_{\bullet j} \propto N^{\ln 5/\ln 3} \approx N^{1.46}$ [39, 155, 178]. Another example is the so-called "T-graph", which is produced by the initial condition of two nodes connected by an edge and recursive replacement of each edge by a star composed of four nodes to produce a fractal [179, 180]. For the T-graph, the MFPT when the target is the unique central node and the source node is distributed uniformly over the $N-1$ remaining nodes is $m_{\bullet j} \propto N^{\ln 6/\ln 3} \approx N^{1.63}$ [181]. Yet another example are so-called "Vicsek

fractals", which are produced by the initial condition of a star having $f+1$ nodes and recursive addition of f replicas of the current network, such that each replica network is connected to the current network by one edge between leaves (i.e., between a node with degree 1 in a replica and a node with degree 1 in the current network) [182, 183]. For Vicsek fractals, the MFPT averaged over all pairs of source and target nodes, chosen from all possible pairs and denoted by $m_{\bullet\bullet}$, scales as $m_{\bullet\bullet} \propto N^{\ln(3f+3)/\ln(f+1)}$ [184]. Similar scaling results have also been studied in other deterministic and stochastic fractals and heterogeneous media [31, 39, 180, 185].

Now let's consider fractal networks that have a power-law degree distribution. One generates a so-called "(u,v)-flower", where u and v are integers, by starting with two nodes connected by an edge and replacing each edge by two parallel paths of length u and v in each generation. This model produces fractal and scale-free networks for $u, v \geq 2$ [171, 186]. The degree distribution of a (u,v)-flower is $p(k) = k^{-\gamma}$, where $\gamma = 1 + \ln(uv)/\ln 2$. For this network, the MFPT between so-called "hubs" (which, in this context, are defined as nodes that are present in the same finite generation and whose degree thus becomes infinite as $N \to \infty$) scales as $m_{ij} \propto N^{\frac{\ln(uv)}{\ln(u+v)}}$ [171]. Consistent with this result, when $u = v$, the MFPT, averaged over source-node position (which is distributed according to the stationary density), to the node with the largest degree (i.e., one of the two nodes that exist initially) is given by $\tilde{m}_{\bullet j} \propto N^{2\ln u/\ln(2u)}$ [187]. A tree-like network model, called the "(u,v)-tree", is produced if, in each generation, one replaces every edge by a path of u edges and add two new paths of $v/2$ edges that start from each end point of the already-added path of u edges and have a loose end. (If v is odd, one adds two paths of $(v \pm 1)/2$ edges.) When $u \geq 2$, the (u,v)-tree model produces fractal and scale-free networks with $\gamma = 1 + \ln(u+v)/\ln 2$ [171, 173]. For such networks, the MFPT between hubs (which here too are defined as nodes that are present in the same finite generation) scales as $m_{ij} \propto N^{\frac{\ln[u(u+v)]}{\ln(u+v)}}$ [171, 188].

All of the above results on fractals and fractal scale-free networks are consistent with a known scaling law for the MFPT: it scales proportionally to $N^{2/d_s} = N^{d_w/d_f}$ [155]. There are known analytical expressions for d_f and d_w for the fractals and fractal scale-free networks whose MFPT we discussed above. The spectral dimension is $d_s = \ln 9/\ln 5 \approx 1.37$ for the Sierpinski gasket [37], $d_s = \ln 9/\ln 6 \approx 1.23$ for the T-graph [179], $d_s = 2\ln(f+1)/\ln(3f+3)$ for the Vicsek fractals [183], $d_s = 2\ln(u+v)/\ln(uv)$ for the fractal (u,v)-flowers [171, 189], and $d_s = 2\ln(u+v)/\ln u(u+v)$ for the fractal (u,v)-trees [171, 189].

As we mentioned in the beginning of this section, there are also scale-free network models that are constructed deterministically and recursively. The resulting networks are not fractals [168–171, 190–193] and are sometimes called "pseudo-fractals" [168]. In the literature, fractal and pseudo-fractal networks are usually distinguished as follows. By definition, pseudo-fractal networks satisfy the small-world property, as they have a small mean path length (which scales as $\log N$ or smaller [165]) between pairs of nodes, possibly due to the creation of shortcuts during the generation of the network. In contrast, the fractal network

models discussed above, as well as conventional fractals, have large worlds, as the mean path length scales as a power of N [172]. Similar to the case of fractal networks, it is possible to exactly calculate the MFPT for a variety of pseudo-fractals by exploiting the recursive nature of their definitions.

Before general (u, v)-flowers were proposed in Ref. [171], the special case with $u = 1$ and $v = 2$ had already been studied [168]. A $(1, 2)$-flower has degree distribution $p(k) \propto k^{-\gamma}$, where $\gamma = 1 + \ln 3/\ln 2 \approx 2.59$ [168]. A (u, v)-flower has a small mean path length and is non-fractal when u or v is equal to 1 [171]. In a $(1, 2)$-flower, the MFPT for an arbitrary pair of nodes (present in a particular finite generation of the network) scales as $m_{ij} \propto N$ [155]. For the same network, m_{ij} averaged over a uniformly distributed location of the source node scales as $m_{\bullet j} \propto N^{\ln 2/\ln 3} \approx N^{0.63}$ when the target node v_j is the largest hub (whose degree $k \approx N^{\ln 2/\ln 3}$) [194]. For a $(1, v)$-flower for general v, the MFPT between hubs (i.e., nodes that are present in the same finite generation, so their degree becomes infinite as $N \to \infty$) scales as $m_{ij} \propto N^{\ln v/\ln(v+1)}$, which is consistent with the results in Ref. [194] that we explained above. For a $(1, v)$-tree for general v, which produces non-fractal scale-free networks [171], the MFPT between hubs (i.e., nodes present in the same finite generation) scales as $m_{ij} \propto N$ and that between non-hub nodes (i.e., nodes of finite degree) scales as $m_{ij} \propto N \ln N$ [171]. The MFPT to the most connected hub v_j (i.e., the node that is present initially) averaged over the position of the uniformly distributed source node v_i (with $i \neq j$) scales as $m_{\bullet j} \propto N$ [188]. Consider a different scale-free tree model, in which, in each generation, m new nodes are connected to each of the already existing nodes. This model produces a power-law degree distribution with $\gamma = 1 + \ln(2m + 1)/\ln(m + 1)$ [191]. For this network model, the MFPT averaged over all pairs of source and target nodes selected uniformly at random scales as $m_{\bullet\bullet} \propto N \ln N$ [195]. The MFPT when the target node is selected from the stationary density of an RW is also proportional to $N \ln N$ as $N \to \infty$ for an arbitrary source node [196]. Similar results have also been derived for pseudo-fractal scale-free networks that include loops. In one such network model, one starts from a single node and, in each generation, adds two replicas of the present network and connects some nodes in each replica to the initially-present single node. This model produces scale-free networks with loops and with $\gamma = \ln 3/\ln 2 \approx 1.59$ [190]. For this model, the MFPT from the largest-degree hub (i.e., the initially-existing node) to a low-degree node created in the latest generation in the growth (and the corresponding MFPT in the reverse direction) scales as $m_{ij} \propto N^{1-\ln 2/\ln 3} \approx N^{0.37}$ [197]. The MFPT to the largest-degree hub starting from a uniformly distributed source node (where the position of the source node is selected with the equal probability from the $N-1$ nodes excluding the target hub node) also scales as $m_{\bullet j} \propto N^{1-\ln 2/\ln 3}$ [197]. One obtains a related pseudo-fractal scale-free network model by starting the recursive growth process of a network from an $N_{\rm init}$-node connected network in which one root node is specified [169, 170]. In each generation, one adds $N_{\rm init} - 1$ replicas $N_{\rm init} \geq 3$) and connects them to the root node by some edges. This model produces a scale-free network with $\gamma = 1 + \ln N_{\rm init}/\ln(N_{\rm init} - 1)$. For this network model, the MFPT to the root node, which has the largest degree,

starting from a source node, selected with equal probability from all nodes but the root, scales as $m_{\bullet j} \propto N^{1-\ln(N_{\rm init}-1)/\ln N_{\rm init}}$ [198]. Because $N_{\rm init} \geq 3$, the MFPT scales no faster than $N^{1-\ln 2/\ln 3} \approx N^{0.37}$. Finally, a so-called "Apollonian network" is defined through an Apollonian packing (i.e., a space-filling packing of spheres) and produces a power-law degree distribution with $\gamma = 1 + \ln 3/\ln 2 \approx 2.58$ [192, 193]. For Apollonian networks, the MFPT to the node with the largest degree, where the source node is selected with the equal probability from all but the target node, is given by $m_{\bullet j} \propto N^{2-\ln 5/\ln 3} \approx N^{0.54}$ [199].

In the results in the above paragraph for pseudo-fractal scale-free (but non-fractal) networks, the MFPT scales at most proportional to $N \ln N$ and mostly scales sublinearly in N. The MFPT is smaller than for fractals and fractal scale-free networks for which m_{ij} (or its mean over source or target nodes) scales superlinearly (i.e., in proportion to $N^{2/d_{\rm s}}$, where $d_{\rm s} < 2$). Because $d_{\rm s} = \infty$ for the aforementioned pseudo-fractal scale-free networks, which satisfy the small-world property, the MFPT does not scale in proportion to $N^{2/d_{\rm s}}$. These results are consistent qualitatively with the third case in Eq. (72), although Eq. (72) was derived for a source node whose location satisfies the stationary density, and many of the aforementioned theoretical results were derived for specific source — target pairs or a source node selected with equal probability from all nodes (excluding the target node). Note that the largest degree in the aforementioned pseudo-fractal scale-free networks (including the $(1, v)$-flowers and $(1, v)$-trees) scales as a sublinear power of N [168–171, 190–193]. Therefore, the third line of Eq. (72) suggests sublinear power-law scaling of the MFPT with respect to N for these networks.

Unsurprisingly, the MFPT can depend on the distance between source and target nodes. The results in Ref. [144] have been extended to the case of networks such as fractal and pseudo-fractal networks in a way that takes into account the distance between the source and target [167, 200]. The MFPT is

$$m_{ij} \propto \begin{cases} N(A + Br^{d_{\rm w}-d_{\rm f}}) & (d_{\rm f} < d_{\rm w}\,;\ \text{i.e., } d_{\rm s} < 2)\,, \\ N(A + B\ln r) & (d_{\rm w} = d_{\rm f}\,;\ \text{i.e., } d_{\rm s} = 2)\,, \\ N(A - Br^{d_{\rm w}-d_{\rm f}}) & (d_{\rm w} > d_{\rm f}\,;\ \text{i.e., } d_{\rm s} > 2)\,, \end{cases} \tag{73}$$

where r is the distance between nodes v_i and v_j, and A and B are constants. For example, the Sierpinski gasket has $d_{\rm f} = \ln 3/\ln 2$ and $d_{\rm w} = \ln 5/\ln 2$. Therefore, Eq. (73) implies that $m_{ij} \propto Nr^{(\ln 5-\ln 3)/\ln 2}$. The pseudo-fractal scale-free networks that we discussed above satisfy the small-world property, so $d_{\rm f} = \infty$ because the number N_r of nodes within radius r grows exponentially in r [172]. Additionally, Eq. (73) still holds if we replace $d_{\rm f}$ by the box-counting dimension $d_{\rm B}$. The box-counting dimension is defined by the scaling relation $N_{\rm B}/N \propto \ell_{\rm B}^{-d_{\rm B}}$, where $N_{\rm B}$ is the number of non-overlapping boxes of linear size $\ell_{\rm B}$ (e.g., the length of a side for a square) that are necessary to cover an entire fractal (and, in the present context, an entire network). For fractals without a heavy-tailed degree distribution, $d_{\rm B} = d_{\rm f}$ [172].

For discussion of scaling theory based on renormalization theory for first-

passage time and other quantities on networks, see Refs. [152, 201]. For other approaches to first-passage times and return times on networks, see Refs. [150, 202, 203].

3.2.6. Cover time

"Cover time" is defined as the time required for a random walker to visit all nodes [1, 42]. It has been proven that the expected cover time c, maximized with respect to the source node, scales approximately as $c \ln [c/(c-1)] N \ln N$ in an Erdős–Rényi random graph in which each pair of nodes is adjacent with a probability of approximately $c(\ln N)/N$ [204]. For a Barabási–Albert scale-free network, the expected cover time scales as $2m/(m-1)N \ln N$, where m is the number of edges in each new node [205]. These results hold with high probability in the limit of infinite network size (i.e., with probability tending to 1 as $N \to \infty$). For arbitrary networks, researchers have developed a universal form of the distribution of cover times [206] and a method for accurately calculating the mean cover time for networks on which RWs relax rapidly [207].

In practice, exactly covering all nodes tends to be a rather strong requirement. In contrast to the above and other rigorous mathematical results on exact cover time, physicists have tended to instead examine "coverage" $C(n)$ in terms of the number of distinct nodes visited at least once within n steps [36, 96, 105, 149, 208–212]. For a complete graph, one can calculate that

$$C(n) = \sum_{i=1}^{N} [1 - (1 - p_i^*)^n] \ . \tag{74}$$

because each node is visited with probability $p_i^* = 1/N$ in a single step. In some situations, one can also expect Eq. (74) to hold approximately as a mean-field calculation. The "edge coverage" (i.e., the number of distinct edges visited at least once within n steps) has also been examined for various networks [208, 213].

3.3. Continuous-time random walks (CTRWs)

Similar to the case of RWs on a line, CTRWs on networks have two main components: the statistics of a walker's trajectory in terms of the number of steps and the statistics of the times at which events take place. By combining these two components, one can specify the probability that a random walker visits a specified node at a specified time. For RWs on networks, the dynamics of a walker are affected not only by the statistical properties of temporal events, but also by the type of network unit in which a temporal process is defined. First, we distinguish between node-centric CTRWs and edge-centric CTRWs [1, 136, 214, 215]. For dynamical processes in general, there are often substantial differences between node-based dynamics and edge-based dynamics [49], so it is crucial to distinguish between these situations. A second delineation is between active and passive CTRWs, depending on whether a walker passively follows edges when available or actively initializes them as it travels. This second distinction becomes crucial for temporal process other than Poisson process. One can

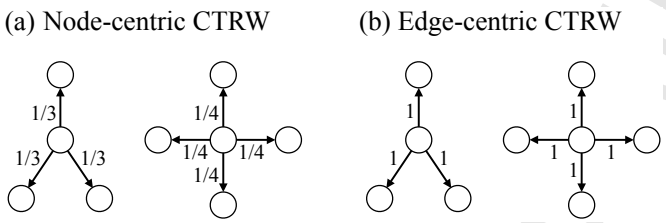

Figure 3: Schematic of two types of continuous-time random walks (CTRWs) on networks: (a) a node-centric CTRW and (b) an edge-centric CTRW. In each case, a walker is visiting either a degree-3 node or a degree-4 node in a network, which we assume is unweighted for simplicity. We show the transition rates for each edge. In panel (a), the walker travels at a unit rate and moves to one of its out-neighbors with equal probability for each choice. Therefore, the transition rate for each edge is the reciprocal of the out-degree of the node that the walker is visiting. In panel (b), however, the transition rate on each edge is equal to 1. Therefore, on average, a walker visiting the node with out-degree 4 leaves the node earlier than a walker visiting the node with out-degree 3.

combine the above components to consider various types of walks (e.g., node-centric active CTRWs).

3.3.1. Node-centric versus edge-centric random walks

In a CTRW, a walker waits until the next move for a time τ, where τ is a random variable. For the sake of simplicity, let's start with a scenario in which moves occur as independent Poisson processes. In other words, τ is distributed according to the exponential distribution with parameter λ. We can safely normalize λ to 1, because λ only sets the time scale. In a node-centric CTRW, a walker moves from node v_i when it becomes active, and it selects one of the out-neighbors, which we denote by v_j, as the destination with a probability proportional to A_{ij} [see Fig. 3(a)]. This assumption is the same as that for a DTRW.

The master equation for the Poissonian node-centric CTRW on a network is

$$\frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = \boldsymbol{p}(t)(-I+T) = -\boldsymbol{p}(t)D^{-1}L\,, \tag{75}$$

where

$$L \equiv D - A \tag{76}$$

is the ("combinatorial") "Laplacian matrix" of the network. The process is driven by the "random-walk normalized Laplacian"

$$L' \equiv D^{-1}L = I - T\,. \tag{77}$$

That is, $(L')_{ij} = \delta_{ij} - (A_{ij}/s_i^{\mathrm{out}})$. If we examine the node-centric CTRW in terms of the number n of moves, the trajectories are statistically the same as those of the DTRW in Eq. (26). Consistent with this observation, node-centric CTRWs are also called the "continuization" of the DTRW [1]. In particular, the

stationary density of the node-centric CTRW is the same as that of the DTRW. By setting the left-hand side of Eq. (75) to 0, we obtain $\boldsymbol{p}^*(-I+T) = 0$, so that $\boldsymbol{p}^* = \boldsymbol{p}^* T$. If the network is undirected, $p_i^* = s_i / \sum_{\ell=1}^{N} s_\ell$. Node-centric CTRWs have been used in, for example, some empirical-data-driven metapopulation disease-spreading models [216, 217]. In those models, a network consists of subpopulations of individuals, and individuals move from one subpopulation to another through a mobility rule. The simplest mobility rule, which has been used widely, is that individuals move according to a Poissonian node-centric CTRW. (For a discussion of mobility models, see Ref. [59].)

Another type of CTRW is an edge-centric CTRW, in which each edge (rather than a node) is activated independently according to a renewal process [see Fig. 3(b)]. By definition, once an edge is activated, it becomes available, and a random walker can use it to move to the associated adjacent node. This RW model has also been called the "fluid model" [1].

When a Poisson process with a rate proportional to the edge weight is assigned independently to each edge, the master equation is

$$\frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = \boldsymbol{p}(t)(-D+A) = -\boldsymbol{p}(t)L\,. \tag{78}$$

The Poissonian edge-centric CTRW is associated with the unnormalized (i.e., combinatorial) Laplacian L. Equation (78) implies that the transition rate at node v_i is equal to s_i^{out}. A walker leaves a node with a large out-strength (such a node may be a network "hub") more quickly than a node with a small out-strength. This situation contrasts with the aforementioned node-centric CTRW, for which the transition rate of a walker is the same for all nodes.

The stationary density for Eq. (78) is

$$\boldsymbol{p}^* L = 0\,. \tag{79}$$

Equation (79) is equivalent to $p_i^* s_i^{\mathrm{out}} - \sum_{j=1}^{N} p_j^* A_{ji} = 0$ (for $i \in \{1, \ldots, N\}$), which indicates that the in-flow of the probability (i.e., $\sum_{j=1}^{N} p_j^* A_{ji}$) and the out-flow of the probability (i.e., $p_i^* s_i^{\mathrm{out}}$) are balanced at each node. Equation (79) also indicates that $\boldsymbol{p}^*$ is a left eigenvector of L with eigenvalue 0. In connected undirected networks, the 0 eigenvalue, which we denote by $\lambda_1 = 0$, is an isolated eigenvalue. Its associated eigenvector is

$$\boldsymbol{p}^* = \frac{1}{N}(1\,, \ldots,\, 1)\,. \tag{80}$$

For a directed network, the right eigenvector corresponding to $\lambda_1 = 0$ is still given by $(1, \ \ldots \ , 1)^\top / N$, but the left eigenvector (i.e., $\boldsymbol{p}^*$) is different in general. Equation (79) is equivalent to $\boldsymbol{p}^* D = (\boldsymbol{p}^* D)\left(D^{-1}A\right) = \boldsymbol{p}^* D T$, where (as usual) T is the transition-probability matrix of the DTRW. Therefore, $\boldsymbol{p}^* D$ is the stationary density for the DTRW (and hence for the above node-centric CTRW) in general directed networks. In other words, for the edge-centric CTRW, p_i^* is given by the expression for p_i^* for the node-centric CTRW divided by s_i^{out}

and properly normalized. Using this relationship, we divide Eq. (35) by s_i^{out} to derive the first-order approximation [132, 218]:

$$p_i^* \approx (\text{const}) \times \frac{s_i^{\text{in}}}{s_i^{\text{out}}} \,. \tag{81}$$

For Poissonian node-centric CTRWs and Poissonian edge-centric CTRWs (and also for DTRWs), one can express the stationary density for directed networks by enumerating spanning trees. We present this technique now because it is easier to understand this approach using L rather than L'. The "(i, j) cofactor" of L is defined by

$$\text{Co}\,(i,j) \equiv (-1)^{i+j} \det \overline{L}^{(i,j)} \,, \tag{82}$$

where $\overline{L}^{(i,j)}$ is the $(N-1)\times(N-1)$ matrix obtained by deleting the ith row and the jth column of L. (Previously, we used $\overline{L}^{(i)}$ to denote the $(N-1)\times(N-1)$ matrix obtained by deleting the ith row and column from L (see Section 3.2.5), and here we use the notation $\overline{L}^{(i,j)}$ without ambiguity. Taking $i = j$ yields $\overline{L}^{(i,i)} \equiv \overline{L}^{(i)}$.) Because $\sum_{j=1}^{N} L_{ij} = 0$ (with $i \in \{1, \ldots, N\}$), the value of $\text{Co}\,(i, j)$ is independent of j. Using Eq. (82) and the fact that L is singular because of the 0 eigenvalue, we obtain

$$\begin{aligned} \sum_{i=1}^{N} \text{Co}(i,i) L_{ij} &= \sum_{i=1}^{N} \text{Co}(i,j) L_{ij} \\ &= \det L = 0 \end{aligned} \tag{83}$$

for any j. This yields

$$p_i^* \propto \text{Co}\,(i,i) = \det \overline{L}^{(i,i)} \,. \tag{84}$$

From the matrix–tree theorem (i.e., Kirchhoff's theorem), $\det \overline{L}^{(i,i)}$ is equal to the sum of the weights of all possible directed spanning trees rooted at v_i (called "arbolescence") [219, 220]. One thereby obtains p_i^* from weighted spanning trees in a formula called the "Markov-chain tree formula" [1]. The "weight" of a spanning tree is defined as the product of the weight of the $N - 1$ edges that form the tree. For unweighted networks, the weight of a spanning tree is 1, and $\det \overline{L}^{(i,i)}$ is equal to the number of spanning trees rooted at v_i. When we apply Eq. (84) to a node-centric CTRW (or to a DTRW), we replace L by L'. In doing this, we must be aware of the weight of spanning trees even for unweighted networks because L' is the combinatorial Laplacian for the weighted adjacency matrix $D^{-1}A$, where A is a binary (i.e., unweighted) adjacency matrix.

Equation (84) is useful for exacting calculating p_i^* for some directed networks, including a variant of Watts–Strogatz small-world networks and multipartite networks [221], and for approximately calculating p_i^* for some types of directed networks with community structure [222].

Although the stationary density differs for node-centric and edge-centric CTRWs, their trajectories (and also those of the DTRW) are statistically the same and are determined by the transition-probability matrix T [see Eq. (24)] for Poisson processes. For edge-centric CTRWs, this is true because the probability that a Poisson process on the edge $(v_i,\ v_j)$ occurs first among the Poisson processes on all edges $(v_i,\ v_\ell)$ (where $\ell \in \{1, \dots, N\}$) is proportional to the rate of the process on the edge $(v_i,\ v_j)$ (i.e., it is proportional to A_{ij}). Let $\boldsymbol{p}(n) = (p_1(n),\ \dots\ , p_N(n))$ denote the distribution of the random walker, where $p_i(n)$ is the probability that the walker visits v_i after exactly n moves. In the Poissonian case, the master equations for the DTRW, the node-centric CTRW, and the edge-centric CTRW in terms of n are each given by Eq. (28). However, the temporal properties along these trajectories are in general different for the two Poissonian CTRWs. In the Poissonian node-centric CTRW, moves are triggered by a Poisson process at a constant rate, so the probability $p(n,t)$ of having performed n steps at time t is given by a Poisson distribution. In the Poissonian edge-centric CTRW, however, $p(n,t)$ depends on a walker's trajectory. When a walker is at a node v_i, the time to the next event is drawn from the exponential distribution with mean $1/s_i^{\text{out}}$. If a trajectory includes many nodes with large out-strengths, the number n of moves at a given time t tends to be larger than for trajectories that traverse many nodes with small out-strengths.

The combinatorial Laplacian L of a connected, undirected network includes exactly one 0 eigenvalue, so $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N$, where λ_ℓ is its ℓth smallest eigenvalue. The combinatorial Laplacian of a directed network satisfies an analogous relationship, $0 = \lambda_1 < \mathrm{Re}(\lambda_2) \leq \cdots \leq \mathrm{Re}(\lambda_N)$, provided the network is strongly connected or has just one strongly connected component from which all other nodes can be reached by a directed path [54, 220, 223]. In the latter case, we call such a strongly connected component the "root component" (including the case of a single node, which is then a "root node"). If there are multiple components in an undirected network or multiple root components, then there are multiple 0 eigenvalues in L, although we do not consider such situations in the present article. The spectral gap (and thus λ_2) governs the relaxation time. The corresponding eigenvector $\boldsymbol{u}_2$ is called the "Fiedler vector". For details of spectral properties of networks, see Refs. [44, 51, 54, 93, 137, 139, 140, 224, 225].

When a network is undirected, one can also construct Eq. (78) as a type of deterministic, linear synchronization or coordination dynamics in which $p_i(t)$ is the state of node v_i and nodes v_i and v_j attract each other with a coupling strength of A_{ij} [54]. The only difference between CTRW dynamics and linearized synchronization dynamics is that $p_i(t)$ is confined between 0 and 1 and normalized in CTRWs, whereas it is not in synchronization dynamics. Therefore, various theoretical results on linear synchronization dynamics on networks are applicable to edge-centric CTRWs. In particular, methods to estimate the relaxation time via the spectral gap of L are useful for understanding relaxation properties of RWs [54, 226, 227].

3.3.2. Active versus passive random walks

In Section 3.3.1, we assumed that temporal events are determined from Poisson processes. In that case, it was not necessary to specify if temporal events are defined on the walker or on the network. However, for non-Poisson processes, it is crucial to specify these properties. In this section, we assume that temporal events are generated by renewal processes with arbitrary distributions of inter-event times. Various empirical data sets related to human activity support heavy-tailed (and hence non-exponential) distributions [57, 228]. See Ref. [229] for a discussion of how to estimate such distributions from empirical data.

One type of model arises when a renewal process describes the timings of the moves of a random walker. In other words, the walker carries its own clock and re-initializes it after each move. The CTRW is then said to be active, which may be appropriate components of models of human or animal trajectories.

A second model consists of assuming that it is the timings at which nodes or edges become active that are generated by a renewal process. In such scenarios, the node or the edge (rather than a walker) carries a clock, and the arrival of a walker does not modify it. The random walker is thus a passive entity that follows edges when they become available [214, 215]. Passive RWs are often used in models of spreading of a virus on a time-dependent contact network or in the spreading of information on a communication network.[4] Active and passive walks model different types of situations. One can interpret active walks as a continuous-time process that can take place on a fixed network architecture. One can then construe the resulting flickering of edges induced by a walker as components of a temporal network. In contrast, passive walks are event-driven processes that take place on a temporal network, which has its own intrinsic dynamics. As we will see, the two types of walks have radically different mathematical properties.

Node-centric active CTRWs. When the inter-event time between two moves obeys a distribution $\psi(\tau)$ that is not exponential, the RW dynamics are non-Markovian. In a non-Markovian setting, the rate at which a walker moves depends on the time since the last move. To analyze this scenario, we consider the extension of Eq. (9) to the case of general networks and write

$$\boldsymbol{p}(t) = \sum_{n=0}^{\infty} \boldsymbol{p}(n) p(n,t)\,, \tag{85}$$

where we recall that $p(n,t)$ is the probability that a walker has moved n times at time t. By taking the Laplace transform of Eq. (85) and using Eqs. (19), we obtain

$$\hat{\boldsymbol{p}}(s) = \frac{1-\hat{\psi}(s)}{s} \sum_{n=0}^{\infty} \boldsymbol{p}(n) \hat{\psi}(s)^n\,. \tag{86}$$

[4] However, spreading processes are typically non-conservative, so one needs to be careful about using RWs in these situations.

We then substitute $\boldsymbol{p}(n) = \boldsymbol{p}(0)T^n$ [see Eq. (29)] into Eq. (86), where T is the transition-probability matrix of the DTRW, to obtain

$$\hat{\boldsymbol{p}}(s) = \frac{1-\hat{\psi}(s)}{s}\boldsymbol{p}(0)\left[I - T\hat{\psi}(s)\right]^{-1} . \tag{87}$$

Equation (87) is a generalization to arbitrary networks of results by Montroll and Weiss [96]. We have implicitly taken a node-centric perspective, as the waiting time (i.e., the time to the next event) of the walker does not depend on the node degree; when the walker is ready for a move, it chooses one of the node's edges uniformly at random and traverses it. The inverse Laplace transform of Eq. (87) gives the probability $p_i(t)$ that the walker visits v_i at time t.

For a Poisson process (i.e., when $\psi(\tau) = \beta e^{-\beta\tau}$), substituting $\hat{\psi}(s) = \beta/(s+\beta)$ [see Eq. (20)] in Eq. (87) yields

$$s\hat{\boldsymbol{p}}(s) - \boldsymbol{p}(0) = \beta\hat{\boldsymbol{p}}(s)(-I+T) \tag{88}$$

after some calculations. Because the inverse Laplace transform of $s\hat{\boldsymbol{p}}(s) - \boldsymbol{p}(0)$ is equal to $\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}(t)$, Eq. (88) leads to Eq. (75) up to a multiplicative constant β.

To understand how the form of $\psi(\tau)$ affects diffusive processes, let's work in the graph-Fourier domain. That is, we work in terms of the amplitude of the eigenmodes, and we examine how the relaxation of different eigenmodes deviates from the situation for Poisson processes [230]. Combining Eqs. (43)–(45) and (86) yields

$$\hat{\boldsymbol{p}}(s) = \frac{1-\hat{\psi}(s)}{s}\sum_{\ell=1}^{N}\frac{a_\ell(0)}{1-\lambda_\ell\hat{\psi}(s)}\boldsymbol{u}_\ell^{\mathrm{L}} , \tag{89}$$

where λ_ℓ is an eigenvalue of T and $\boldsymbol{u}_\ell^{\mathrm{L}}$ is the corresponding left eigenvector. By taking the inner product of both sides of Eq. (89) with the right eigenvector $\boldsymbol{u}_\ell^{\mathrm{R}}$ of T for a particular value ℓ, we obtain

$$\hat{a}_\ell(s) = \frac{1-\hat{\psi}(s)}{s\left[1-\lambda_\ell\hat{\psi}(s)\right]}a_\ell(0) . \tag{90}$$

For CTRWs driven by Poisson processes, an eigenmode relaxes exponentially in time. However, relaxation dynamics can be rather different when $\psi(t)$ is not an exponential distribution. For simplicity, we assume that $\psi(t)$ has finite mean and finite variance. (When these moments are not defined, one can examine dynamical processes using the framework of fractional calculus [231].) We substitute a small-s expansion

$$\hat{\psi}(s) = 1 - \langle\tau\rangle s + \frac{1}{2}\langle\tau^2\rangle s^2 + o(s^2) \tag{91}$$

into Eq. (90). For the ℓth mode, where $\lambda_\ell \neq 1$, one can calculate that

$$a_\ell(s) = \frac{\langle\tau\rangle}{1-\lambda_\ell}\left[1 - s\left(\frac{\lambda_\ell\langle\tau\rangle}{1-\lambda_\ell} + \frac{\langle\tau^2\rangle}{2\langle\tau\rangle}\right)\right] . \tag{92}$$

This leads to a characteristic time $t_{\rm cha}$ of

$$t_{\rm cha} = \frac{\lambda_\ell \langle \tau \rangle}{1-\lambda_\ell} + \frac{\langle \tau^2 \rangle}{2\langle \tau \rangle} = \langle \tau \rangle \left(\frac{1}{\epsilon_\ell} + \beta_{\rm burst} \right) , \tag{93}$$

where $\epsilon_\ell = 1 - \lambda_\ell$ is the eigenvalue of the random-walk normalized Laplacian L' and

$$\beta_{\rm burst} = \frac{\sigma_\tau^2 - \langle \tau \rangle^2}{2\langle \tau \rangle^2} , \tag{94}$$

where $\sigma_\tau^2 = \langle \tau^2 \rangle - \langle \tau \rangle^2$ is the variance of τ. The quantity $\beta_{\rm burst} \in [-1/2, \infty)$ is a measure of burstiness. Poisson processes have $\beta_{\rm burst} = 0$, and $\beta_{\rm burst} = -1/2$ when $\psi(\tau)$ is distributed as a delta function. A heavy-tailed distribution, implying bursty activity of nodes, generates a large value of $\beta_{\rm burst}$.

Let's consider the slowest-decaying mode associated with the spectral gap ϵ_ℓ (i.e., the smallest nonzero eigenvalue of L'). The corresponding characteristic decay time $t_{\rm cha}$ indicates the relaxation time of the CTRW towards equilibrium. Equation (93) includes competition between two factors. When the spectral gap is small relative to $1/\beta_{\rm burst}$, the first term on the right-hand side of Eq. (93) is dominant. In this case, $t_{\rm cha}$ is determined primarily by structural bottlenecks in a network (e.g., through the existence of sets of densely-connected nodes called "communities" (see Section 5.3), which are connected weakly to each other) [68, 137, 140]. When the spectral gap is larger or when an event sequence is bursty (in the sense of a large variation in inter-event times), the second term dominates the right-hand side of Eq. (93). In this case, $t_{\rm cha}$ is determined primarily by the properties of $\psi(\tau)$ rather than by network structure.

Because the inter-event time and the number of moves in a RW are statistically independent, the stationary density of the node-centric CTRW with a general $\psi(\tau)$ is the same as those for a DTRW or a Poissonian node-centric CTRW. One can thus calculate the recurrence time and first-passage time of a node-centric CTRW by multiplying the corresponding results for the DTRW (see Section 3.2.5) by $\langle \tau \rangle$.

Edge-centric active CTRWs. One can define other types of active RWs that have qualitatively different behaviors of the stationary density and first-passage times. For instance, consider the following edge-centric active RW: when a walker arrives at a node, it considers each edge and takes the first edge available for transport. The time at which each edge appears is independently drawn from the same distribution $\psi(\tau)$ where, as before, the clock on each edge is re-initialized upon the arrival of a walker at an incident node. Because only the first edge to appear is taken by the walker, there is a competition between different edges. The probability density that a random walker moves from node v_i to node v_j at time τ since the walker arrived at v_i is

$$f(\tau; j \leftarrow i) = \psi(\tau) \left[\int_\tau^\infty \psi(\tau') {\rm d}\tau' \right]^{k_i - 1} . \tag{95}$$

Some calculations yield

$$p_i^* = \frac{\langle \min_{\ell=1,\dots,k_i} \tau_\ell \rangle k_i}{\sum_{j=1}^N \langle \min_{\ell=1,\dots,k_j} \tau_\ell \rangle k_j}\,, \tag{96}$$

where the factors of τ_ℓ are independent copies of inter-event times that are drawn from the distribution $\psi(\tau)$. Because

$$\left\langle \min_{\ell=1,\dots,k_i} \tau_\ell \right\rangle = \int_0^\infty \left[\int_{\tau'}^\infty \psi(\tau') \mathrm{d}\tau' \right]^{k_i} \mathrm{d}\tau' \tag{97}$$

depends only on k_i, Eqs. (96) and (97) imply that p_i^* depends only on k_i. Note that the stationary density for the active RW is not proportional to k_i unless τ is constant, which reduces the model to the DTRW. The mean recurrence time for node v_i is

$$m_{ii} = \frac{\sum_{j=1}^N \left\langle \min_{\ell=1,\dots,k_j} \tau_\ell \right\rangle k_j}{k_i} \propto \frac{1}{k_i}\,. \tag{98}$$

Equations (96) and (98) indicate that Kac's formula [see Eq. (55)] is not satisfied unless the network is regular.

Edge-centric passive CTRWs. Passive RWs differ from active ones in that properties of a network (rather than a random walker) evolves as a renewal process. We start with edge-centric passive RWs, which have attracted considerable attention because of their many applications (e.g., diffusion on temporal networks). We thus assume that each edge is governed by an independent renewal process, which we assume for simplicity is the same distribution $\psi(\tau)$ for each edge. A first important difference from active walks arises from the "waiting-time paradox" (which is also called the "bus paradox") [3, 232]. In this paradox, a walker arrives at node v_i from node v_ℓ. The waiting time before edge (v_i, v_j) (with $j \neq \ell$) is activated is typically longer than the naive expected value $\langle \tau \rangle / 2$. Let $\psi^{\mathrm{w}}(\tau^{\mathrm{w}})$ denote the distribution of waiting times τ^{w} on edge $(v_i,\, v_j)$ after a walker has arrived at node v_i from node v_ℓ (where $\ell \neq j$). See Fig. 4 for a schematic. One can calculate $\psi^{\mathrm{w}}(\tau^{\mathrm{w}})$ from $\psi(\tau)$ when the arrival of a walker to v_i and the activation of edge $(v_i,\, v_j)$ are statistically independent processes. In that situation, the probability density for the time at which a walker moves from v_ℓ to v_i lies in an interval of length τ satisfies

$$f(\tau) = \frac{\tau \psi(\tau)}{\int_0^\infty \tau' \psi(\tau') \mathrm{d}\tau'} = \frac{\tau \psi(\tau)}{\langle \tau \rangle}\,. \tag{99}$$

Conditioned on the walker's arrival time to v_i lying in an interval of length τ, the probability density for the waiting time to be equal to τ^{w} is

$$g(\tau^{\mathrm{w}}|\tau) = \begin{cases} 1/\tau & (0 \leq \tau^{\mathrm{w}} \leq \tau)\,, \\ 0 & (\tau > \tau^{\mathrm{w}})\,. \end{cases} \tag{100}$$

Equations (99) and (100) yield

$$\psi^{\mathrm{w}}(\tau^{\mathrm{w}}) = \int_{\tau^{\mathrm{w}}}^\infty f(\tau) g(\tau^{\mathrm{w}}|\tau) \mathrm{d}\tau = \frac{1}{\langle \tau \rangle} \int_{\tau^{\mathrm{w}}}^\infty \psi(\tau) \mathrm{d}\tau\,. \tag{101}$$

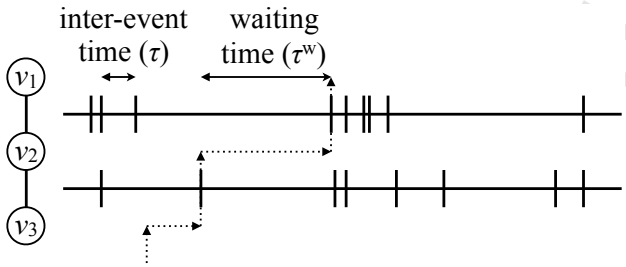

Figure 4: Schematic illustrating the concept of waiting time. We show a trajectory of a random walker using dotted arrows. The walker moves from node v_3 to node v_2, and it then moves to node v_1. This example corresponds to $j = 1$, $i = 2$, and $\ell = 3$ in the main text. (See the $j \neq \ell$ case in Eq. (102).)

In particular, the mean waiting time is given by $\int_0^\infty \tau^{\mathrm{w}} \psi^{\mathrm{w}}(\tau^{\mathrm{w}}) \mathrm{d}\tau^{\mathrm{w}} = \langle \tau^2 \rangle / (2\langle \tau \rangle)$. If $\psi(\tau)$ is heavy-tailed, $\langle \tau^2 \rangle$ is much larger than $\langle \tau \rangle$, so a typical waiting time is very long. For example, if $\psi(\tau) \propto \tau^{-\gamma}$, with $\gamma \in (2, 3]$, the mean inter-event time is finite, whereas the mean waiting time diverges because $\langle \tau^2 \rangle$ diverges.

A second difference is that one can only derive approximate master equations for edge-centric passive CTRWs, whereas they are exact for active CTRWs. When a random walker moves from node v_ℓ to node v_i at time t, the waiting time (i.e., the time to the next event) on edge (v_i, v_j), where $j \neq \ell$ (we will consider the case $j = \ell$ in the next paragraph), is estimated by the distribution ψ^{w} However, if a random walker has already traversed edge (v_i, v_j) in the past — let's suppose that the last traversal time occurred at t' — the independence assumption that is required to derive Eq. (101) is not satisfied, and the waiting time on (v_i, v_j) is not given exactly by the distribution ψ^{w}, unless the process is Poissonian and ψ is an exponential distribution. The deviation between the waiting-time distribution and ψ^{w} increases when t' approaches t. In the remainder of the present section, we ignore any modification of the distribution of the subsequent waiting time caused by past events on (v_i, v_j); this corresponds to assuming that $t' = -\infty$. To our knowledge, the impact of such a memory effect (i.e., finite t') has not been considered in detail in the literature.

A third difference stems from the possibility of non-Markovian trajectories for random walkers. To explain this point, consider the case of backtracking moves (i.e., $v_\ell \to v_i \to v_\ell$). For such backtracking moves, the waiting time on the edge (v_i, v_ℓ) is distributed according to ψ, rather than ψ^{w}, as the waiting-time paradox does not apply. The existence of different waiting times for backtracking and non-backtracking moves has impacts the motion of a walker. For a walker to move to node v_j at time τ^{w} since the walker moved from node v_ℓ to node v_i, there cannot be any events on any edges emanating from v_i in $[0, \tau^{\mathrm{w}}]$, and then an event must occur on the edge (v_i, v_j) at time τ^{w}. Let $f(\tau^{\mathrm{w}}; j \leftarrow i | i \leftarrow \ell)$ denote the probability density of the event that a walker that has moved from

v_ℓ to v_i moves to node v_j at time $\tau^{\rm w}$. We obtain

$$f(\tau^{\rm w}; j \leftarrow i | i \leftarrow \ell) \approx \begin{cases} \psi(\tau^{\rm w}) \left[\int_{\tau^{\rm w}}^{\infty} \psi^{\rm w}(\tau') {\rm d}\tau'\right]^{k_i - 1} & (j = \ell)\,, \\ \psi^{\rm w}(\tau^{\rm w}) \left[\int_{\tau^{\rm w}}^{\infty} \psi^{\rm w}(\tau') {\rm d}\tau'\right]^{k_i - 2} \int_{\tau^{\rm w}}^{\infty} \psi(\tau') {\rm d}\tau' & (j \neq \ell)\,. \end{cases} \tag{102}$$

Equation (102) indicates that where a walker moves depends not only on its current position but also on the edge that it used to arrive to that position. For trajectories of RWs, one can construe this situation as a special case of the "memory networks" that we will discuss in Section 4.2.2.

Unless ψ is an exponential distribution, $f(\tau^{\rm w}; \ell \leftarrow i | i \leftarrow \ell)$ is not equal to $f(\tau^{\rm w}; j \leftarrow i | i \leftarrow \ell)$ (with $j \neq \ell$) in general, so the trajectory of an RW (i.e., the walk measured in terms of the number of moves) is non-Markovian. In particular, if ψ is a heavy-tailed distribution, the mean waiting time is larger than the mean inter-event time. Therefore, a walker tends to backtrack (i.e., there are sequences of moves of the form $v_\ell \to v_i \to v_\ell$), and diffusion dynamics are slowed down. This slowing down is caused entirely by the modification of trajectories in non-exponential distributions, and, in particular, it does not arise from a competition between structural and temporal factors (in contrast to Eq. (94)). If ψ has lighter tails than an exponential distribution, a walker tends to avoid backtracking. (We briefly discuss non-backtracking RWs in Section 6.) When ψ is not an exponential distribution, trajectories of the edge-centric passive CTRW are different from those of active CTRWs or DTRWs.

We now evaluate the stationary density and recurrence time of non-Poissonian edge-centric passive CTRWs [215]. Let $q_{j \leftarrow i}(t)$ denote the rate at which a random walker moves from node v_i to node v_j at time t. This quantity satisfies the following approximate self-consistency equation:

$$q_{j \leftarrow i}(t) \approx \sum_{\ell \in \mathcal{N}_i} \left[\int_0^t f(t - t'; j \leftarrow i | i \leftarrow \ell) q_{i \leftarrow \ell}(t') {\rm d}t' \right] + p_{j \leftarrow i}(0) \delta(t)\,, \tag{103}$$

where we recall that $\mathcal{N}_i$ is set of the neighbors of v_i. The initial condition satisfies

$$\sum_{j \in \mathcal{N}_i} p_{i \leftarrow j}(0) = p_i(0)\,. \tag{104}$$

Equation (104) implies that one needs to specify an initial condition that includes not only the current position of the walker but also its previous location. More generally, the transition probability of a move depends on the previous move. The master equation is given by

$$\frac{d}{dt} p_i(t) = \sum_{j \in \mathcal{N}_i} \left[q_{i \leftarrow j}(t) - q_{j \leftarrow i}(t) \right]\,. \tag{105}$$

To derive the stationary density, we work in terms of $q_{i\leftarrow j}(t)$ rather than $p_i(t)$. We take the Laplace transform of Eq. (103) to obtain

$$\hat{q}_{j\leftarrow i}(s) \approx \sum_{\ell\in\mathcal{N}_i} \left[\hat{f}(s; j \leftarrow i | i \leftarrow \ell)\hat{q}_{i\leftarrow\ell}(s)\right] + p_{j\leftarrow i}(0)\,. \tag{106}$$

Note that $\hat{q}_{j\leftarrow i}(s) \neq \hat{q}_{i\leftarrow j}(s)$ in general even for undirected networks. Equation (106) is a set of linear equations with $2M$ unknowns. We solve $\hat{q}_{j\leftarrow i}(s)$ and then calculate the stationary value of $q_{j\leftarrow i}(t)$ (i.e., $q^*_{j\leftarrow i} \equiv \lim_{t\to\infty} q_{j\leftarrow i}(t)$ as $\hat{q}_{j\leftarrow i}(0)$). We thereby obtain p^*_i as a weighted sum of $q^*_{i\leftarrow j}$ terms, where $j \in \mathcal{N}_i$. In fact, $q^*_{j\leftarrow i}$ does not depend on i or j, and the final result is

$$p^*_i = \frac{1}{N} \quad (i \in \{1, \ldots, N\})\,. \tag{107}$$

Therefore, the stationary density is the uniform density, independent of the network structure and the form of $\psi(\tau)$. The mean recurrence time is

$$m_{ii} \approx \frac{N\langle\tau\rangle}{k_i}\,. \tag{108}$$

Equation (108) indicates that the mean recurrence time is essentially independent of $\psi(\tau)$, as it depends only on the mean $\langle\tau\rangle$, which gives the trivial normalization of time. Equations (107) and (108) imply that Kac's formula [see Eq. (55)] is not satisfied by any edge-centric passive CTRW except in regular networks.

Node-centric passive CTRWs. To conclude our taxonomy of CTRWs on networks, we mention a fourth combination: passive node-centric RWs. We are not aware of studies of node-centric passive RWs, though they may be relevant for situations in which the activity of a temporal network is driven by node dynamics more than by interactions between nodes. Node-centric passive CTRWs are also subject to the bus paradox, but they are substantially simpler mathematically than edge-centric active walks, because non-Markovian trajectories do not arise when the renewal processes on the nodes are independent.

4. Random walks on generalized networks

4.1. Multilayer networks

A multilayer network includes different "layers" and allows one to explicitly incorporate different types of subsystems and/or different types of ties between edges [60, 61]. The latter case, which is often called a "multiplex" network, occurs when there are different types of interactions between individuals, different modes of transportation, and so on. If there are $\ell_{\max}$ layers, one can represent a multilayer network as an ordinary (i.e., "monolayer") network with $\ell_{\max}N$ nodes, where there are $\ell_{\max}$ replicates of each node if each entity (represented by a node) exists on every layer. How strongly different layers are connected to each other (and which interlayer edges are present) has an enormous effect

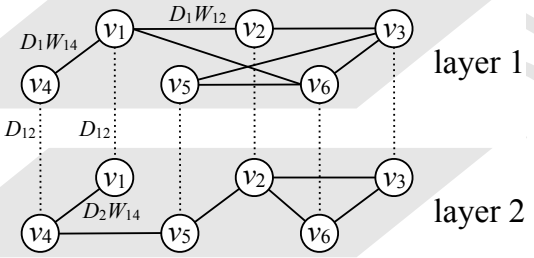

Figure 5: Schematic of a Poissonian edge-centric CTRW on a multilayer network with $\ell_{\max} = 2$ layers. The values on the edges represent edge weights.

on diffusive dynamics in multilayer networks [61, 62, 233]. It thereby affects anything else, such as various community-detection methods, that are based on RWs (see Section 5.3) [234–236].

Let's consider Poissonian edge-centric CTRWs. For simplicity, we also assume undirected multilayer networks in which each intra-layer network is a connected network [237–239] and each node is present on every layer (though of course this need not be true in general). We also assume that inter-layer edges occur only between the same entity in different layers (i.e., so-called "diagonal" coupling) and that there is only a single type (i.e., "aspect") of layering [61]. (For example, a single-aspect multilayer network can be a multiplex network, but it cannot be both multiplex and time-dependent.) Let $A^\alpha = (A^\alpha_{ij})$ denote the adjacency matrix for the αth layer. One needs to think about both diffusion within layers and diffusion between layers (see Fig. 5). Let D_α denote the intra-layer diffusion constant in the αth layer, and let $D_{\alpha\beta}$ (with $\alpha, \beta \in \{1, \dots, \ell_{\max}\}$) denote the inter-layer diffusion constant between the αth and βth layers. Such constants set the edge weights between pairs of nodes that represent the same entity in different layers, and the corresponding nodes in the αth and βth layers are connected by an edge on which there is a Poisson process with rate $D_{\alpha\beta}$. The master equation is given by

$$\frac{\mathrm{d}p_i^\alpha(t)}{\mathrm{d}t} = D_\alpha \sum_{j=1}^{N} A_{ij}^\alpha \left[p_j^\alpha(t) - p_i^\alpha(t)\right] + \sum_{\alpha'=1}^{\ell_{\max}} D_{\alpha\beta} \left[p_i^{\alpha'}(t) - p_i^\alpha(t)\right] , \qquad (109)$$

where $p_i^\alpha(t)$ is the probability that a random walker visits the ith node in the αth layer. The normalization is given by $\sum_{\alpha=1}^{\ell_{\max}} \sum_{i=1}^{N} p_i^\alpha(t) = 1$.

Consider the case of two layers and $D_{\mathrm{x}} \equiv D_{12} = D_{21}$ [237, 239]. Equation (109) is written concisely as

$$\frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = -\boldsymbol{p}(t)\mathcal{L} , \qquad (110)$$

where $\boldsymbol{p}(t) = (p_1^1(t),\ p_2^1(t),\ \dots\ , p_N^1(t),\ p_1^2(t),\ p_2^2(t),\ \dots\ , p_N^2(t))$, and

$$\mathcal{L} = \begin{pmatrix} D_1 L_1 + D_{\mathrm{x}} I & -D_{\mathrm{x}} I \\ -D_{\mathrm{x}} I & D_2 L_2 + D_{\mathrm{x}} I \end{pmatrix} \qquad (111)$$

is the (combinatorial) "supra-Laplacian", where L_1 and L_2 are the (combinatorial) Laplacian matrices for the intra-layer network. Because this RW is an edge-centric CTRW on an undirected network, the stationary density is $(p_i^\alpha)^* = 1/(2N)$ (with $i \in \{1, \dots, N\}$ and $\alpha \in \{1, 2\}$).

The supra-Laplacian matrix $\mathcal{L}$ has a 0 eigenvalue that corresponds to the stationary density. The relaxation time is governed by the smallest positive eigenvalue (i.e., the spectral gap) λ_2 of $\mathcal{L}$. One of the nonzero eigenvalues is $2D_\mathrm{x}$ and has a corresponding eigenvector of $(1, \ \dots \ , 1, \ -1, \ \dots \ , -1)$. If the inter-layer diffusion constant D_x is small, then $\lambda_2 = 2D_\mathrm{x}$, so the inter-layer hopping is a bottleneck for diffusion in the entire multilayer network. In the opposite limit ($D_\mathrm{x} \gg 1$), one can examine diffusion properties using a perturbative analysis [237]. The quantity $2D_\mathrm{x}$ is still an eigenvalue, but it diverges to infinity in the limit $D_\mathrm{x} \to \infty$, and there are N copies of the same eigenvalue in this limit. Another important quantity is $\lambda_\mathrm{s}/2$, the eigenvalue of $(L_1 + L_2)/2$; and there are also N copies of this eigenvalue. Therefore, $\lambda_2 = \lambda_\mathrm{s}/2$. Note that $L_1 + L_2$ is the (combinatorial) Laplacian for the monolayer network obtained by adding the intra-layer edge weights for each intra-layer edge and ignoring the inter-layer edges. We obtain

$$\frac{\lambda_\mathrm{s}}{2} \geq \frac{\lambda_2^{\alpha=1} + \lambda_2^{\alpha=2}}{2} \geq \min(\lambda_2^{\alpha=1}, \lambda_2^{\alpha=2}) \,, \tag{112}$$

where λ_2^α is the second-smallest eigenvalue (i.e., the spectral gap) of L_α, so it specifies the speed at which an RW on the network consisting only of the αth layer (so there are no inter-layer edges) relaxes to the stationary density. Equation (112) implies that above diffusion in the two-layer network is faster than diffusion in the slower layer. For some multilayer networks, however, diffusion can occur faster than in each layer considered individually [237, 238].

The small-D_x and $D_\mathrm{x} \gg 1$ regimes are connected by a discontinuous (i.e., "first-order") phase transition [239]. More precisely, there exists a threshold value D_x^* of D_x, such that $\lambda_2 = 2D_\mathrm{x}$ for $D_\mathrm{x} \leq D_\mathrm{x}^*$ and $\lambda_2 \leq \lambda_\mathrm{s}/2$ for $D_\mathrm{x} \geq D_\mathrm{x}^*$. Note that $D_\mathrm{x} \to \lambda_\mathrm{s}/2$ as $D_\mathrm{x} \to \infty$. The first derivative of λ_2 with respect to D_x is discontinuous at $D_\mathrm{x} = D_\mathrm{x}^*$. The transition point has an upper bound given by $D_\mathrm{x}^* \leq \lambda_\mathrm{s}/4$.

Reference [240] investigated the so-called "coverage" time of different types of CTRWs in multilayer networks by calculating the mean fraction of distinct nodes that are visited at least once (in any layer) in some time period by a walk (which can start from any node in a network). Reference [240] then examined coverage as a function of time when some nodes are deleted and used it to consider the resilience of multilayer networks to random node failures. In their paper, node failure is defined with respect to the removal of nodes in individual layers (rather than, e.g., removal from all layers), such as a failure of a station in a single transportation mode (i.e., a single layer) in a transportation network.

See Refs. [60–62] and references therein for further discussion of diffusion processes in multilayer networks. For example, RWs have been employed to estimate the number of layers in multilayer networks [241]. The investigation of RWs in multilayer networks is a very active area of research.

4.2. Temporal networks

Many empirical networks vary over time, and one can describe them as temporal networks [57, 58]. CTRWs with non-exponential distributions of inter-event times (see Section 3.3) are often discussed in the context of temporal networks, because non-Poissonian distributions of inter-event times are a fundamental property of most empirical temporal networks [57, 228].

In this section, we discuss some situations in which a temporal network is given in the form of a sequence of static networks (which are called "snapshots" in [59])[5]. In this type of example, one time-independent network corresponds to a single observation (with a time stamp) of a temporal network, whose time resolution may correspond to that imposed by a recording period (e.g., every 20 secs). One can then consider an RW on a (temporal) sequence of adjacency matrices:

$$\mathcal{A} = \{A(1), A(2), \dots, A(n_{\max})\}\,, \tag{113}$$

where $(A(n))_{ij}$ encodes the activation of edge $(v_i,\, v_j)$ at discrete time n (with $n \in \{1, \dots, n_{\max}\}$). See the review [58] for a discussion of several models of RWs on temporal networks in addition to the ones that we will discuss in the following sections.

4.2.1. Activity-driven model

RWs on temporal networks have been examined both analytically and computationally. One useful approach is to examine RWs on an "activity-driven model" of temporal networks [147].

The simplest type of activity-driven model generates a sequence of uncorrelated time-independent networks [242]. First, we associate each node v_i (with $i \in \{1\ , \dots,\ N\}$) with a random variable a_i, called the "activity potential", drawn from a given distribution $F(a)$ (with $a \geq 0$). Second, at each discretized time t, each node v_i is independently active with probability $a_i \Delta t < 1$ and inactive with probability $1 - a_i \Delta t$, where Δt is the time difference (which we assume to be homogeneous) between two consecutive time points. Third, at each t, each activated node generates m undirected edges that connect to m other nodes uniformly at random. When nodes v_i and v_j are both active and each connects to the other with an edge at time t, we suppose that there is exactly one unweighted edge $(v_i,\, v_j)$ at t. In practice, we suppose that $a_i \Delta t$ is sufficiently small to prevent such mutual edge creation to occur too often. We regard the network at each t as an undirected and unweighted network, and we repeat this procedure independently to generate a time-independent network for the time interval Δt.

Consider the aggregation of a temporal network into a time-independent network, which we construct by summing the edge weights across some time

[5] There are also other types of temporal networks [57, 58], and it is important to consider the time scales of both network evolution and the evolution of dynamical processes on a network to determine appropriate frameworks for network analysis [49].

window for each edge. The aggregated network neglects any temporal information contained in the temporal network during that window. If we aggregate observed time-independent networks over some time — which cannot be too long, or else the aggregated network might be a complete weighted graph — the aggregated (and sometimes called "annealed") adjacency matrix is given by

$$A^*_{ij} \approx \frac{m\,(a_i + a_j)}{N}\,, \tag{114}$$

where we neglect $o(1/N)$ terms. The degree distribution of the aggregated network is

$$p(k^*) \approx \frac{1}{m} F\left(\frac{k}{m} - \langle a\rangle\right)\,, \tag{115}$$

where $\langle a\rangle = \int aF(a)\mathrm{d}a$ is the ensemble average of a. Therefore, a heterogeneous distribution $F(a)$ yields a comparably heterogeneous degree distribution in the aggregated network.

When we observe a temporal network with a fine temporal resolution, the network at each time point is very sparse[6]. This also occurs for the above activity-driven model if $a_i\Delta t$ and m are sufficiently small. A walker has to remain at a node if the node is isolated at the present time t, and this fact has a substantial effect on RW dynamics. In the above activity-driven model, there are two ways for a walker located at node v_i to move to node v_j in a network at time t [147, 244]. The first way is to combine the following three independent events: (i) v_i is activated with probability $a_i\Delta t$, (ii) node v_i is connected to v_j with probability m/N, and (iii) the edge $(v_i,\, v_j)$ is traversed with probability $1/(m+m\langle a\rangle\Delta t)$. Note that the mean degree of v_i in a time-independent network at an arbitrary time t when v_i is activated is equal to $m + m\langle a\rangle\Delta t$, because v_i has $m\langle a\rangle\Delta t$ edges from the activation of other nodes. The second way is to combine the following four independent events: (i) node v_i is not activated with probability $1 - a_i\Delta t$, (ii) node v_j is activated with probability $a_j\Delta t$, (iii) v_j is connected to v_i with probability m/N, and (iv) the edge $(v_i,\, v_j)$ is traversed with probability $1/(1 + m\langle a\rangle\Delta t)$. By adding these contributions and assuming that Δt is small, we obtain a transition-probability matrix T with elements

$$\begin{aligned} T_{ij} &\approx a_i\Delta t\frac{m}{N}\frac{1}{m + m\langle a\rangle\Delta t} + (1 - a_i\Delta t)a_j\Delta t\frac{m}{N}\frac{1}{1 + m\langle a\rangle\Delta t} \\ &\approx \frac{\Delta t}{N}\,(a_i + ma_j) \quad (j \neq i)\,. \end{aligned} \tag{116}$$

Note that $T_{ii} = 1 - \sum_{j=1;j\neq i}^{N} T_{ij}$.

We aggregate all nodes with the same value of a into one group, and we regard a as continuous. Let $p_a(t)$ denote the probability that a single node with

[6] We use the term "sparse" to indicate the presence of an extremely small number of edges rather than in a conventional graph-theoretic sense, in which a sparse network still typically has a large number of edges (but with an edge density that scales sufficiently slowly as the number N of nodes becomes large) [44, 243].

activity potential a is visited at time t. The normalization is $\int p_a(t)F(a)\mathrm{d}a = 1$, and the master equation in the $\Delta t \to 0$ limit is

$$\frac{\mathrm{d}p_a(t)}{\mathrm{d}t} = \int a' p_{a'}(t)F(a')\mathrm{d}a' - a p_a(t) + ma\frac{1}{N} - m\langle a\rangle p_a(t)\,. \tag{117}$$

The first and second terms on the right-hand side of Eq. (117) account, respectively, for the in-flows and out-flows of probability driven by $(\Delta t/N)a_i$ on the right-hand side of Eq. (116). The third and fourth terms account, respectively, for the in-flows and out-flows driven by $(\Delta t/N)ma_j$ in Eq. (116). This RW is a Poissonian node-centric CTRW whose general master equation is given by Eq. (75).

The stationary density of Eq. (117) is

$$p_a^* = \frac{\frac{ma}{N} + \phi}{a + m\langle a\rangle}\,, \tag{118}$$

where

$$\phi = \int a p_a^* F(a)\mathrm{d}a \tag{119}$$

is the mean probability flow from active nodes at equilibrium. By combining Eqs. (118) and (119), we obtain the following self-consistency equation:

$$\phi = \int a\frac{\frac{ma}{N} + \phi}{a + m\langle a\rangle}F(a)\mathrm{d}a\,. \tag{120}$$

Because we are considering a Poissonian node-centric CTRW in an undirected network, the stationary density for a time-independent, aggregated network has components that are proportional to node degree. Equation (114) implies that p_a^* for the aggregated network is proportional to $m(a + \langle a\rangle)$. However, the stationary density for the CTRW on the activity-driven temporal network model, obtained by numerically solving Eq. (120) for a given heterogeneous $F(a)$, is rather different from the time-independent case [147]. In particular, in the activity-driven model, p_a^* saturates as the degree (or, equivalently, a) increases.

The MFPT is also different in the temporal and aggregated networks. At equilibrium, the probability that a walker moves to node v_j in each discrete step of time Δt is $\xi_j = \sum_{i=1;i\neq j}^{N} p_i^* T_{ij}$. The probability that the walker arrives at v_j for the first time after n steps is thus given by $\xi_j(1-\xi_j)^{n-1}$ under the mean-field approximation in Eq. (70). One can then calculate that the MFPT for the above activity-driven model is

$$m_{ij} \approx \sum_{n=1}^{\infty}\Delta t n\xi_j(1-\xi_j)^{n-1} = \frac{\Delta t}{\xi_j} = \frac{N}{ma_j + \sum_{\ell=1}^{N} a_\ell p_\ell^*}\,. \tag{121}$$

This result is different from the aggregated (time-independent) network case, in which $m_{ij} \approx 1/p_j^*$ under the mean-field approximation in Eq. (70). A crucial

difference between RW dynamics in the temporal and aggregated cases is that a walker in the activity-driven model can be trapped for some time in an isolated node v_i and is temporarily unable to travel to a different node. At a later time, v_i becomes connected to another node, and the walker can then move away from v_i. This phenomenon never happens in a time-independent (i.e., aggregated) network, as edges are always present. These results were recently extended to RWs on an extended activity-driven model in which each node is assigned an attractiveness value in addition to an activity potential [245].

One can also define RWs on empirical temporal networks. For example, given a sequence of time-independent networks, one can use each time-independent network to induce one time step of a DTRW [148]. (Another approach is to construct a multilayer representation of such a temporal network, and examine an RW on the resulting multilayer network [61, 234].) In Ref. [148], the authors compared properties of RWs on empirical temporal networks to those on randomized temporal networks, which included ones in which the times of activating edge (v_i, v_j) are redistributed uniformly over time while keeping the weight of each edge in the aggregated network the same as that in the original temporal network. In comparison to such randomized temporal networks, the numerical computations in Ref. [148] suggest that empirical temporal networks tend to slow down RW processes, as the MFPT is large and the coverage at a given time is small. See Refs. [230, 231, 246–249] for discussions of the effects of temporal networks on the speed of diffusion on networks.

Note that if the time-independent network at each time point is sparse, the trajectory of a random walker may not be as random as the terminology RW might suggest. For example, if the degree of v_i equals 1 at a certain time t, then the walker located at v_i must move to its one neighbor. If v_i is isolated at time t, then the walker does not move at t. In the extreme case in which each node is adjacent to just one node or is isolated at all times, the trajectory of the "random" walk is deterministic. For example, in the temporal network on $N = 4$ nodes in Fig. 6, a walker starting from node v_1 always visits node v_4 after three time steps, so there is no randomness. In a CTRW, this situation always occurs in some sense: if $\psi(\tau)$ is a continuous distribution, then multiple events occur at the same time with probability 0 because of the continuous-time nature of the stochastic dynamics. However, because the event times themselves are determined from a random process, we safely regard CTRWs as RWs. This situation is not shared by RWs on temporal networks when a network is given by a single realization of empirical or numerical data. Fortunately, there are at least two (imperfect) ways out of this conundrum. One solution is to aggregate a sequence of time-independent networks with a sufficiently large time window to make them sufficiently dense. Another solution is to allow walkers to wait at the current node with some probability even if an edge is available for it to move to another node.

4.2.2. Memory networks

By definition, a DTRW is a (stationary) Markov chain such that the transition probability does not depend on the past trajectory. Poissonian CTRWs and

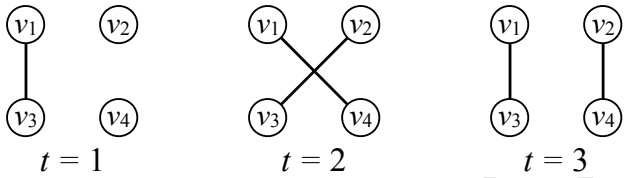

Figure 6: A temporal network with three time points and $N = 4$ nodes.

non-Poissonian active CTRWs (either node-centric or edge-centric) also share this property. However, many real temporal networks have correlations in edge activations [57–59]. Therefore, one does not expect a trajectory of RWs on an empirical temporal network to be a Markov chain, as certain trajectories are favored and others are discouraged or even forbidden. Such trajectories are poorly reproduced by the first-order Markov chains that we have considered thus far. In this situation, using higher-order Markov chains may be helpful [248, 250], and it is also important to explore non-Markovian stochastic processes.

To consider the above issue with empirical data in the context of temporal networks, we first map time series of edge activations in a sequence of time-independent networks to trajectories of walkers [250]. We assume that a walker is located initially at a uniformly randomly selected node v_i. (The choice of initial condition can matter if RW trajectories simulated in the following are short.) A walker waits there until at least one edge is available for it to move. When at least one edge becomes available, the walker leaves the node with probability $1-q$ and does not move with probability q. As usual, the destination node v_j is selected with probability $A_{ij}(t)/\sum_{\ell=1}^{N} A_{i\ell}(t)$. We repeat this procedure several times and thereby generate multiple trajectories starting at $n = 1$ and finishing at $n = n_{\max}$. When $q = 0$, the walker always moves to a different node using the first available edge [148, 230]. When $q \in (0, 1)$, some randomness is introduced into the trajectories [251], preventing spurious effects such as a strong tendency for backtracking [252]. However, for sufficiently large q, the effect of temporal correlations between edges at short time scales becomes unimportant, which may dilute the impact of the temporality of the data. If trajectories are statistically independent of the past locations of a walker, it is sufficient to use a first-order Markov chain. In this case, the transition-probability matrix $T = (T_{ij})$ constructed from an aggregated network, in which the weight of edge (v_i, v_j) is equal to the sum of $(A(t))_{ij}$ over time, is sufficient for describing the RWs. We denote a first-order Markov chain on an aggregated network by $\mathcal{M}_1$. See the top right panel of Fig. 7.

In general, the probability that a random walker visits node v_i after the $(n+1)$th step depends on the entire history of a stochastic process. To partially take into account temporal correlations between edge activations, one can use a second-order Markov chain. We define a process, which we denote by $\mathcal{M}_2$, using an expanded transition probability tensor, whose element $T_{i'ij}$ represents the probability that a walker moves from node v_i to node v_j given that the

previous position is node $v_{i'}$. Another representation of the process $\mathcal{M}_2$ is to use a memoryless RW (i.e., a first-order Markov chain) between directed edges of the original network. In this representation, the probability that directed edge $\overrightarrow{v_i v_j}$ is visited depends on $\overrightarrow{v_{i'} v_i}$ rather than only on node v_i, as in the first-order Markov chain $\mathcal{M}_1$. For simplicity, for the rest of the present discussion, we use the shorthand notation $\vec{ij}$ for a directed edge $\overrightarrow{v_i v_j}$. For this representation, we regard the state space (i.e., the set of directed edges) as the nodes of a new network, which we call the "$\mathcal{M}_2$ network" or "(second-order) memory network". One construes the original network as a "physical network", and the state space of $\mathcal{M}_2$ is the so-called "directed line graph" of the original network [253]. The memory network has $2M$ nodes whether the original network is directed or undirected. We sometimes use the term "memory nodes" for the nodes of a memory network. Even for undirected networks, we must assign two memory nodes $\vec{ij}$ and $\vec{ji}$ to each pair of adjacent nodes v_i and v_j in the original network, because a memory node encodes the time ordering of visits. The number of edges in a memory network is proportional to $\langle k^2 \rangle N$ [254].

To improve accuracy, one can also examine memory networks in the form of higher-order Markov chains. For example, in a third-order Markov chain, the transition probability depends on the currently visited node v_i and two previously visited nodes v_j and v_ℓ. A memory node is then specified by $\overrightarrow{v_\ell v_j v_i}$. However, going beyond second-order Markov chains is not always practical. First, a second-order memory network is conceptually simpler than higher-order counterparts, as the memory nodes are given by edges of the original network rather than by higher-order structures. Second, one may only obtain marginal gains by considering higher orders [250]. Third, higher-order memory networks require a lot of data, because the number of memory nodes and transition probabilities to be estimated increases exponentially with the order of the Markov chain.

One encodes the dynamics of a second-order Markov chain by a transition-probability matrix on the network with $2M$ nodes whose elements are given by $p(\vec{ij} \to \vec{jk})$ (see Fig. 7). In practice, one estimates $p(\vec{ij} \to \vec{jk})$ with

$$p(\vec{ij} \to \vec{jk}) = \frac{(\text{number of transitions } \vec{ij} \to \vec{jk})}{\sum_{\ell=1}^{N} (\text{number of transitions } \vec{ij} \to \vec{j\ell})} , \tag{122}$$

where one counts the number of transitions in the RW trajectories generated by the sequence of time-independent networks. One interprets the transitions as movements between directed edges. The normalization is given by $\sum_{\ell=1}^{N} p(\vec{ij} \to \vec{j\ell}) = 1$. In situations in which one can measure RW trajectories in empirical data, they can be used directly to estimate Eq. (122) [250].

In a first-order Markov chain $\mathcal{M}_1$ (i.e., a DTRW) on an unweighted network, we obtain

$$p(\vec{ij} \to \vec{j\ell}) = \begin{cases} 1/k_j & (v_\ell \text{ is a neighbor of } v_j) , \\ 0 & (\text{otherwise}) . \end{cases} \tag{123}$$

In general second-order Markov chains, the probability that a walker visits node

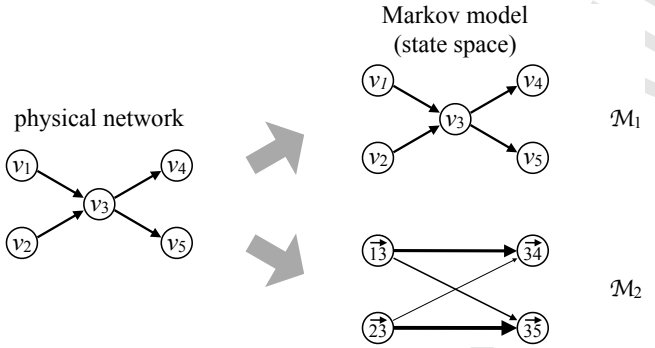

Figure 7: Memory networks (of order 2). The network on the left shows a part of a directed network (a "physical network"). The width of each edge represents edge weight. In the present example, we assume for simplicity that the physical network is unweighted. In the first-order Markov chain $\mathcal{M}_1$, a state is a node of the physical network. In the second-order Markov chain $\mathcal{M}_2$ (of which we show a part), a state is a directed edge of the physical network. The state space is the directed line graph of the physical network. If the process that occurs on the physical network is Markovian, transitions in $\mathcal{M}_2$ are uniform in the following sense. Suppose, as indicated in the figure, that node v_3 has two in-edges and two out-edges in the physical network. One then should be able to reach node $\overrightarrow{34}$ with equal probability from nodes $\overrightarrow{13}$ and $\overrightarrow{23}$, yielding the same weight for edges $\overrightarrow{13} \to \overrightarrow{34}$ and $\overrightarrow{23} \to \overrightarrow{34}$. In the part of $\mathcal{M}_2$ (determined from, for example, a temporal network) that we show in this figure has edge weights that are different from the expectation of the first-order Markov chain $\mathcal{M}_1$. In other words, a move from node v_3 to node v_4 is more likely to occur when a walker arrives at v_3 from v_1 than from v_2. Therefore, the process represented by $\mathcal{M}_2$ network is not Markovian on the physical network.

$\overrightarrow{j\ell}$ after $n+1$ steps is given by

$$p(\overrightarrow{j\ell}; n+1) = \sum_{i=1}^{N} p(\overrightarrow{ij}; n) p(\overrightarrow{ij} \to \overrightarrow{j\ell})\,. \tag{124}$$

Edge-centric passive CTRWs with a non-exponential distribution $\psi(\tau)$ of inter-event times are one example of a situation that is appropriate to model using a second-order Markov chain rather than a first-order chain. Equation (102) implies that $p(\overrightarrow{\ell i} \to \overrightarrow{ij})$ depends on whether $j = \ell$ or $j \neq \ell$. In particular, if $\psi(\tau)$ is a heavy-tailed distribution, then $p(\overrightarrow{\ell i} \to \overrightarrow{i\ell})$ (i.e., the probability to backtrack) is larger than is expected in a first-order Markov chain. All other $p(\overrightarrow{\ell i} \to \overrightarrow{ij})$ $(j \neq \ell)$ values are the same. In contrast, if $\psi(\tau)$ is a lighter-tailed distribution than an exponential distribution, $p(\overrightarrow{\ell i} \to \overrightarrow{i\ell})$ is smaller than expected in a first-order Markov chain, and random walkers tend to avoid backtracking. The extreme case of the latter situation is a non-backtracking RW [248, 250, 255, 256]. In such an RW, a walker performs an RW, except that it is not allowed to backtrack [257, 258], so $p(\overrightarrow{ij} \to \overrightarrow{ji}) = 0$ and $p(\overrightarrow{ij} \to \overrightarrow{j\ell}) = 1/(s_j^{\text{out}} - A_{ji})$ (with $\ell \neq i$).

A network's associated non-backtracking matrix, which is a $2M \times 2M$ adjacency matrix for the $\mathcal{M}_2$ network, has been used recently in several applications,

including percolation [259, 260], network centralities [261], community detection [262–264], and efficient "immunization" algorithms [265]. More generally, we also note that non-backtracking matrices help with "message passing" and "belief propagation" approaches to network analysis.

To quantify the difference between a first-order Markov chain $\mathcal{M}_1$ and a second-order Markov chain $\mathcal{M}_2$, we compare their entropy rates. "Entropy rate" quantifies the uncertainty of the next state given the current state, weighted by the stationary density. For $\mathcal{M}_1$, the entropy rate is

$$H_1 = -\sum_{i,j=1}^{N} p_i^* T_{ij} \log T_{ij} \,. \tag{125}$$

In $\mathcal{M}_2$, one calculates the entropy rate for a first-order Markov chain on the memory network and thereby obtains

$$H_2 = -\sum_{i,j,\ell=1}^{N} p_{\vec{ij}}^* p(\vec{ij} \to \vec{j\ell}) \log p(\vec{ij} \to \vec{j\ell}) \,, \tag{126}$$

where $p_{\vec{ij}}^*$ is the stationary density at node $\vec{ij}$ in the memory network. In many empirical temporal networks, H_2 is considerably smaller than H_1, implying that one cannot neglect memory effects [248, 250] (also see [266, 267] for similar measurements). The first-order Markov chain $\mathcal{M}_1$ tends to overestimate the number of available neighbors around the current node of a random walker compared to its higher-order counterparts.

The observation that $H_2 < H_1$ can influence RW dynamics, other dynamical processes on networks, and how one wants to calculate certain structural features of networks. For example, communities of networks found by second-order Markov chains (see Section 5.3.1) tend to contain edges that are activated at the same time [255]. Such communities are undetectable using first-order models (such as the usual RWs). Memory also affects the relaxation time of an RW or other Markov processes towards a stationary state [247].

The eigenvalue λ_2 of T with the second-largest absolute value influences network community structure and determines the relaxation time of RWs [230]. (See Section 5.3 for more discussions of community structure.) Temporal correlations can either increase or decrease λ_2, depending on how temporal correlations are introduced [247]. If memory increases $|\lambda_2|$, a random walker in a second-order Markov process tends to be confined in a certain part of the original network (i.e., the $\mathcal{M}_1$ network) than is suggested by network structure alone. In the corresponding $\mathcal{M}_2$ network, a random walker tends to be trapped in a community. In this case, memory has slowed down relaxation to a steady state. However, if memory decreases $|\lambda_2|$, a walker moves from one community to another faster than is suggested by the original network. In this case, memory accelerates relaxation to a steady state. Moreover, non-Markovian pathways in a network without community structure can still create community structure in the associated $\mathcal{M}_2$ network [59]. As a simple example (see Fig. 8), consider an undirected 3-clique (i.e., a triangle).

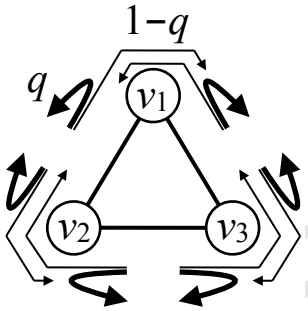

Figure 8: A second-order Markov chain on a 3-clique. The widths of the arrows represent (schematically) the transition probabilities in a second-order Markov chain. For example, a walker that has moved from node v_2 to node v_1 moves back to v_2 with probability q and moves to v_3 with probability $1-q$ in the next move. Because $q > 1/2$ in this figure, random walkers tend to backtrack.

The transition-probability matrix of the usual DTRW (i.e., the $\mathcal{M}_1$ process) is

$$T = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} , \tag{127}$$

which yields $\lambda_2 = -1/2$. On the triangle network, consider the second-order Markov chain process defined by

$$p(\overrightarrow{12} \to \overrightarrow{21}) = p(\overrightarrow{21} \to \overrightarrow{12}) = p(\overrightarrow{13} \to \overrightarrow{31}) = p(\overrightarrow{31} \to \overrightarrow{13}) = p(\overrightarrow{23} \to \overrightarrow{32}) = p(\overrightarrow{32} \to \overrightarrow{23}) = q\,, \tag{128}$$

$$p(\overrightarrow{12} \to \overrightarrow{23}) = p(\overrightarrow{21} \to \overrightarrow{13}) = p(\overrightarrow{13} \to \overrightarrow{32}) = p(\overrightarrow{31} \to \overrightarrow{12}) = p(\overrightarrow{23} \to \overrightarrow{31}) = p(\overrightarrow{32} \to \overrightarrow{21}) = 1-q\,, \tag{129}$$

where $q \in [1/2, 1)$ (see Fig. 8). This RW backtracks the edge traversed in the previous step with probability q. If we order the nodes in the $\mathcal{M}_2$ network as $\overrightarrow{12}$, $\overrightarrow{21}$, $\overrightarrow{13}$, $\overrightarrow{31}$, $\overrightarrow{23}$, and $\overrightarrow{32}$, the transition-probability matrix is

$$T = \begin{pmatrix} 0 & q & 0 & 0 & 1-q & 0 \\ q & 0 & 1-q & 0 & 0 & 0 \\ 0 & 0 & 0 & q & 0 & 1-q \\ 1-q & 0 & q & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-q & 0 & q \\ 0 & 1-q & 0 & 0 & q & 0 \end{pmatrix} . \tag{130}$$

The eigenvalues of T are 1, $1-2q$, and $\left[-1+q \pm \sqrt{(1-q)^2 + 4(2q-1)}\right]/2$. The last eigenvalues (for each of $\pm$) have multiplicity two. The relaxation time is governed by $\lambda_2 = \left[-1+q-\sqrt{(1-q)^2+4(2q-1)}\right]/2 < 0$. When $q = 1/2$, we obtain $\lambda_2 = -1/2$, which is consistent with the memoryless case. When $q > 1/2$,

we see that λ_2 decreases monotonically towards -1, which one obtains in the limit $q \to 1$. A large value of q makes $|\lambda_2|$ large and hence makes the spectral gap small, so a random walker tends to spend a long time in a community in the $\mathcal{M}_2$ network. In this situation, each of the three edges constitutes a community, and it is difficult for the walker to leave any edge.

Storing the stationary density of a second-order Markov chain (i.e., $p^*_{\vec{ij}}$) may be prohibitive, particularly for a network that is not sparse, because the $\mathcal{M}_2$ network has $2M$ nodes. A space-friendly alternative is to introduce an approximation $p^*_{\vec{ij}} \approx \hat{p}^*_i \hat{p}^*_j$ (with $i, j \in \{1, \ldots, N\}$) and estimate $\hat{p}^*_i$ [268]. The estimated $\hat{p}^*_i$ is the stationary density of a modified second-order-like Markov chain called a "spacey RW" [269]. In a spacey RW, a walker visiting node v_j forgets the last node v_i that it has visited. The walker then draws the fictive last position v_i uniformly at random from the list of the nodes visited in the past. (The probability that each node is selected is weighted by the number of past visits to the node.) The walker then moves to v_ℓ according to the probability $p(\vec{ij} \to \vec{j\ell})$. Spacey RWs are a type of "reinforced RW", in which nodes or edges (nodes in the present case) visited frequently in the past are also visited more frequently in subsequent steps [75]. Spacey RWs have such a richer-get-richer mechanism embedded in the process to select the fictive last position v_i.

The formalism in this section allows one to examine how temporal correlations in a network affect spreading processes [255, 267]. It can also be used to directly exploit knowledge of the trajectories of diffusing entities (so-called "trajectory data") when they can be observed and collected. For instance, the trajectory of a traveler between different airports is rather different from a first-order Markov process, so it is important to consider higher-order Markov processes or even non-Markovian dynamics [250]. Similar conclusions arise when studying animal movements [270], Website traffic [271], and other applications.

Although trajectory data are becoming increasingly available, it is difficult to measure trajectories for the vast majority of systems. Moreover, even when they can be measured, a high-order Markov model or non-Markovian model may be unnecessarily complicated to extract the most salient features of a system. Consequently, researchers have proposed simple models of second-order Markov dynamics based on the distinction between different types of transitions on networks. In practice, one can calibrate the model parameters in systems in which trajectories can be measured and then use these models to simulate trajectories in similar systems for which data on trajectories are not available.

In [250], Rosvall et al. enumerated three different types of transitions:

1. A *return step*, in which a walker coming from $\vec{ij}$ jumps to $\vec{ji}$. In other words: a walker coming from node i to j returns to node i.
2. A *triangular step*, in which a walker coming from $\vec{ij}$ moves to edge $\vec{j\ell}$, where $\ell \neq i$ is a neighbor of node i.
3. An *exploratory step*, in which a walker moves from $\vec{ij}$ to an edge $\vec{j\ell'}$ whose end point ℓ' is neither node i nor any of i's neighbors.

To each of above types of transition, one then assigns a positive weight (denoted r_2, r_3, and $r_{>3}$, respectively) to account for their relative contributions. One

can recover several existing types of processes for specific choices of parameters. For example, $r_2 = r_3 = r_{>3}$ yields a first-order Markov process and $r_2 = 0$, $r_3 = r_{>3} > 0$ yields a non-backtracking RW.

5. Applications

5.1. Search on networks

People are often interested in finding a resource, service, or piece of information that is available only at some nodes in a network [44]. If network structure is completely known to a user or a designer, a shortest path from the initially visited node to a destination node provides the most efficient way of searching, although it may be sensible to plan a detour if one expects congestion from traffic somewhere along a shortest path.

If a searcher has partial information about his/her destination (e.g., the geographic distance to it), one can of course use such information to inform search paths [272]. In contrast, if one does not have any information about network structure or has only local information (such as the degrees of neighbors), RWs provide a viable approach for searching in networks. One context in which this idea has been investigated and implemented are decentralized peer-to-peer networks [273, 274]. A node that sends a query emits N_{rw} packets to neighbors selected uniformly at random. Each packet behaves as a random walker, which travels until it finds the item or reaches a prescribed lifetime $n_{\max}$, which is the maximum number of steps it is allowed to take before it is removed from the network. Search overhead is determined by $N_{\mathrm{rw}} n_{\max}$, which is a measure of the number of walkers, averaged over time, that are wandering in a network. One expects larger $N_{\mathrm{rw}} n_{\max}$ to yield better search efficiency (i.e., a higher probability that an item is found). Therefore, there is a trade-off between search overhead and search efficiency. RW search methods are comparable with flooding search methods in various networks and scenarios [273]. In a flooding method, first used by Gnutella, a node with a query asks all of its neighbors, each of which in turn asks all of its unvisited neighbors, and so on [275].

Most empirical networks are highly heterogeneous in node degree [44]. If a node that is making or passing on a query knows the degrees of its neighbors, one can enhance search efficiency by sending the query to high-degree neighbors [276]. The main limitation of such an approach is that most queries are forwarded to hubs, potentially causing overloading at such nodes (depending on their capacity).

5.2. Ranking

In the study of networks, one often seeks to rank nodes, edges, or other structures based on their relative importances (i.e., "centralities"). There are myriad ways to measure centralities in networks, especially for ranking nodes [44, 277], and new ones are published at a very rapid pace. Many methods for computing node centralities are based on eigenvectors of matrices and are derived from various types of RWs or other walks. These include "Katz centrality" [278] and

related measures (such as "communicability") [142], "eigenvector centrality" [67], "PageRank" [23], "hubs" and "authorities" [279], "non-backtracking centrality" [261], and many others. By considering RWs on multilayer and temporal networks, one can also generalize such notions of centrality [240, 251, 280–285].

5.2.1. PageRank

The most famous centrality measure is probably "PageRank", which was introduced originally for ranking web pages. In this context, it was introduced by Brin and Page [66] (see also [286]), although an equivalent formulation had already existed for two decades [287]. (Brin and Page's discovery was independent of Ref. [287].)

PageRank is discussed thoroughly in many review papers and monographs [23, 288–292], and it has been used (and generalized) for numerous applications — including ranking of academic journals and papers, professional sports, disease-gene identification, discovery of correlated genes and proteins, systemic risk in financial networks, anomaly detection in distributed engineered systems, ordering of the most important functions in Linux, prediction of traffic flow and human movement, recommendation systems in online marketplaces, image search engines, identifying community structure in networks, and much more [23]. We indicate a few fascinating applications in passing. For example, seven new genes that predict the survival of patients in a type of pancreatic cancer were identified using PageRank [293]. PageRank has also been used to rank professional tennis players [294], and PageRank and other RW-based ranking methods have been used for ranking teams in U.S. college football [295, 296] and ranking players in Major League Baseball [297]. PageRank and other eigenvector-based centrality measures have also been used to rank universities [298], mathematics research programs [284, 299], baby names [300], and many other things.

The PageRank vector is defined as the stationary density of a DTRW on a network that is a modification of an original network to guarantee that the stationary density always exists. For the original network, the temporal evolution of the probability $\boldsymbol{p}(n)$ that node v_i (with $i \in \{1, \ldots, N\}$) is visited at time n is governed by Eq. (26) (or, equivalently, by Eq. (28)). The essential idea of PageRank is to use the stationary density in Eq. (30) as a centrality measure. Equation (30) implies that node v_i is central if many edges enter node v_i (i.e., it has a large in-degree), the source node of the edge that enters v_i is a central node, and the source node v_j of the edge that enters v_i has a small out-degree. The last condition ensures that the total centrality of v_j is shared among its out-neighbors. This recursive relationship (i.e., a node is central if it is adjacent to central nodes) leads to an eigenvalue problem. Other centrality measures — including eigenvector centrality, Katz centrality, the hyperlink-induced topic search algorithm (which uses "hubs" and "authorities"), and many others — are based on the same basic idea [44]. In PageRank, the eigenvalue problem corresponds specifically to the stationary density of a DTRW.

In an empirical directed network, one cannot typically use a transition-probability matrix T without modification to measure centralities, because such networks are not usually strongly connected. Consequently, there are transient

nodes with stationary density equal to 0, and the stationary density need not be unique, as it depends on the initial condition of an RW when there are multiple absorbing states. To overcome these problems, we allow walkers to "teleport" (e.g., uniformly at random) to other nodes to construct an effective network that is strongly connected. The master equation for the altered RW is

$$p_i(t+1) = \alpha \sum_{j=1}^{N} p_j(t) T_{ji} + (1-\alpha) u_i \,, \tag{131}$$

where the "preference vector" $(u_1, \ \ldots \ , u_N)$, which satisfies the constraint $\sum_{i=1}^{N} u_i = 1$, determines the conditional probability that a walker teleports to node v_i when it teleports. At any node with at least one out-edge, a walker teleports with probability $1-\alpha$. To prevent the transition probability in Eq. (24) from being ill-defined, it is standard to ensure that a walker teleports with probability 1 (rather than with probability $1-\alpha$) when it visits a so-called "dangling node" (which have no out-edges, so $s_i^{\text{out}} = 0$ for a dangling node v_i). Mathematically, we set $T_{ij} = u_j$ (with $j \in \{1, \ldots, N\}$) for any dangling node v_i. For web browsing, one interprets teleportation as a move to a new web page without following a hyperlink on the web page that is currently being visited. If $u_i > 0$ (with $i \in \{1, \ldots, N\}$), any $\alpha \in (0, 1)$ renders the altered RW ergodic, and Eq. (131) thus converges to a unique stationary density. The PageRank vector is the stationary state of Eq. (131), and it is equal to the normalized eigenvector corresponding to the largest positive eigenvalue of the matrix T' with elements $T'_{ij} = \alpha T_{ij} + (1-\alpha) u_j$.

Power iteration of T' converges rapidly if the spectral gap of T' is large (or, equivalently, if the second-largest eigenvalue of T' has small magnitude). The second-largest (in magnitude) eigenvalue of T' is equal to $\alpha\lambda_2$, where λ_2 is the second-largest (in magnitude) eigenvalue of T [288]. Therefore, power iteration converges towards the PageRank vector at a rate that is proportional to $1/\alpha$ [23]. However, a small value of α, which corresponds to a large teleportation probability, dilutes the effect of the original network structure (which is encoded in the transition-probability matrix T). A rule of thumb is to set α near 1 to suppress the effect of teleportation, but to also make sure that it is not too close. A popular choice is to let $\alpha = 0.85$ and use a preference vector of $u_i = 1/N$ (with $i \in \{1, \ldots, N\}$) so that one teleports to nodes uniformly at random. An alternative choice is a "personalized PageRank" [23, 288–291, 301–304], in which the preference vector is localized around one node or a small number of nodes (which can be helpful for applications to community detection [68]). One can also examine other teleportation strategies [305].

The stationary density of Eq. (131) has components

$$p^*_{i;\alpha} = (1-\alpha) \sum_{j=1}^{N} u_j \left[(I - \alpha T)^{-1} \right]_{ji} \,, \tag{132}$$

and we note that we explicitly include the dependence on α in our notation.

The Taylor expansion of Eq. (132) yields [306, 307]

$$p^*_{i;\alpha} \approx u_i + \sum_{\ell=1}^{\infty} \alpha^\ell \sum_{j=1}^{N} u_j \left(T^\ell_{ji} - T^{\ell-1}_{ji}\right) . \tag{133}$$

Equation (133) includes terms for walks of all lengths ℓ, and it thereby reveals the non-local nature of PageRank. When the value of α is large, a lot of credit is given to long walks. (See Ref. [142] for similar discussions in the context of centrality measures such as communicability.) In fact, the stationary density can change drastically as a function of α [288]. Let's set $u_i = 1/N$ (with $i \in \{1, \ldots, N\}$) and rewrite Eq. (133) as

$$p^*_{i;\alpha} = \frac{1}{N} + \sum_{\ell=1}^{\infty} \frac{\alpha^\ell}{N} \sum_{j,j'=1}^{N} \left(\frac{s^{\text{in}}_{j'} - s^{\text{out}}_j}{s^{\text{in}}_{j'}} \right) T_{jj'} T^{\ell-1}_{j'i} . \tag{134}$$

The leading contribution for small α makes the PageRank vector uniform across all nodes. Heterogeneity arises as α increases. Equation (134) indicates that the contribution of each length-ℓ walk is proportional to $s^{\text{in}}_{j'} - s^{\text{out}}_j$. Each term on the right-hand side of Eq. (134) vanishes when a network is regular in the weighted sense (i.e., when $s^{\text{in}}_i = s^{\text{out}}_i = s$, where $i \in \{1, \ldots, N\}$). This yields $p^*_{i;\alpha} = 1/N$ for any value of α.

A strategy to minimize the dependence of the PageRank vector on α is to carefully choose the preference vector. One choice is $u_i = s^{\text{in}}_i / \sum_{\ell=1}^{N} s^{\text{in}}_\ell$ [305], inspired by the observation that the in-strength of a node is often correlated positively with p^*_i for a DTRW on the original network (see Section 3.2.2). With this choice of u_i, one uniformly randomly selects an edge rather than a node. One then teleports, uniformly at random, to one of the two end points of the selected edge. Substituting this preference vector into Eq. (133) yields

$$p^*_{i;\alpha} = \frac{s^{\text{in}}_i}{\sum_{\ell=1}^{N} s^{\text{in}}_\ell} + \sum_{\ell=1}^{\infty} \frac{\alpha^\ell}{\sum_{\ell=1}^{N} s^{\text{in}}_\ell} \sum_{j=1}^{N} \left(s^{\text{in}}_j - s^{\text{out}}_j\right) T^\ell_{ji} , \tag{135}$$

which differs from Eq. (134) in several respects. As $\alpha \to 0$, the components of the PageRank vector in Eq. (135) are given by the in-strength of the nodes. (The simplest — and a rather popular — measure of centrality in networks is simply to calculate node degrees and/or node strengths.) The ℓth-order contribution consists of a weighted mean of the walks of length ℓ. One expresses their contribution to the PageRank vector in terms of the source node of a walk (i.e., v_j) in Eq. (135). This contrasts with Eq. (134), where one instead expresses the contribution in terms of edges $(v_j, v_{j'})$. A node v_j that is the source of more probability flow than it receives as a destination (i.e., $s^{\text{in}}_j > s^{\text{out}}_j$) makes a positive contribution to the PageRank vector, and a node v_j with $s^{\text{in}}_j < s^{\text{out}}_j$ makes a negative contribution. Equation (135) is independent of α when a network is balanced. (Recall from Section 3.2.2 that a directed network is balanced when $s^{\text{in}}_i = s^{\text{out}}_i$ for each i.) In a balanced network, Eq. (135) reduces to $p^*_i = s^{\text{in}}_i / \sum_{\ell=1}^{N} s^{\text{in}}_\ell$.

Chung proposed a variant of PageRank called "heat-kernel PageRank" (which is defined for strongly connected networks) [308, 309]. It is the probability density of a Poissonian node-centric CTRW at time t, where t is the only parameter and it plays the role of α from the original PageRank. One uses a preference vector as an initial condition. Heat-kernel PageRank tends to the stationary density of a DTRW as $t \to \infty$. (For undirected networks, the components of the limiting stationarity density are thus proportional to the node strengths.)

We also note that various versions of PageRank and similar RW-based centralities for multilayer networks have been proposed [281–283, 310–312].

5.2.2. Laplacian centrality

PageRank is essentially the stationary density of a DTRW. The stationary density of the Poissonian edge-centric CTRW has also been employed as a centrality measure for directed networks (and, in fact, it has a longer history than PageRank [219, 313–315]). For strongly connected networks, such a "Laplacian centrality" is defined by the left eigenvector corresponding to the 0 eigenvalue of the (combinatorial) Laplacian L. That is, it is given by $\boldsymbol{p}^*$ in Eq. (79). This Laplacian centrality has been used, for example, to rank football teams [316], baseball players [297], and neurons [222]. It has also been used in population ecology as a "reproductive value" [317, 318].

5.2.3. TempoRank

One can extend the DTRW to temporal networks by using sequences $\{A(1), A(2), \ldots\}$ of adjacency matrices (see Section 4.2). Therefore, one can also extend PageRank to temporal networks. One such generalization is called "TempoRank" [251], and Katz centrality [280, 319] and all eigenvector-based centralities [284] have been generalized to such temporal networks.

In this section, we discuss TempoRank. We consider an undirected temporal network whose edge weights at each discrete time have (nonnegative) integer values. The latter assumption corresponds to a situation in which an event is an unweighted edge and each node pair can experience multiple events during the time window corresponding to a given matrix in the sequence. One can also image a sequence of networks, in which one has a time-independent view (or approximation) of a temporal network at a given instant in time. This weighting assumes that a random walker at node v_i that moves at discrete time n selects each available edge (i.e., event) with the same probability and then traverse the chosen edge. Because we consider DTRWs, the walker moves at most once per time step. To avoid using a multilayer-network formalism, we also assume that there are no inter-layer edges between different matrices in the sequence.

To make the walk random even when just a single edge is available to a walker in a time period, we assume that, in each time period, a walker resists moving from node v_i with probability q per unit weight of an edge connected to v_i. For example, if v_i is adjacent to a node with two events (i.e., edge weight equal to two) and to another node with three events at discrete time n, a walker visiting v_i stays at the same node with probability q^5 at time n. A large q entails slow diffusion, and the parameter q allows one to explore situations in

which diffusion is slower than the time scale of the dynamics of the network. We define the transition probability from node v_i to node v_j at discrete time n as

$$T_{ij}(n) = \begin{cases} \delta_{ij} & (s_i(n) = 0\,, j \in \{1, \ldots, N\})\,, \\ q^{s_i(n)} & (s_i(n) \geq 1\,, i = j)\,, \\ (A(n))_{ij}(1 - q^{s_i(n)})/s_i(n) & (s_i(n) \geq 1\,, i \neq j)\,, \end{cases} \tag{136}$$

where $s_i(n) = \sum_{j=1}^{N} (A(n))_{ij}$ is the strength of v_i at time n. Note that $\sum_{j=1}^{N} T_{ij}(n) = 1$. From Eq. (136), we see that a walker does not move with probability $q^{s_i(n)}$. Otherwise, it moves to a neighbor with a uniform probability of $1/s_i(n)$. By setting the probability of not moving to $q^{s_i(n)}$, one ensures that the probability of not moving from v_i is unaffected by whether multiple edges are present simultaneously in a time period or if they are distributed over multiple times. For example, if v_i is connected simultaneously to three other nodes by unweighted edges at time $n = 1$ but isolated at times $n = 2$ and $n = 3$, the probability that a walker visiting v_i does not move during $n = 1$, $n = 2$, and $n = 3$ is equal to q^3. The probability is the same if v_i is connected to one node at each of $n = 1$, $n = 2$, and $n = 3$. Note that one can derive the former case (i.e., three edges simultaneously connected to v_i) from the latter case (i.e., one edge connected to v_i at each time) by coarse-graining the temporal network (e.g., by regarding $A(3n-2) + A(3n-1) + A(3n)$ as a new adjacency matrix at a rescaled discrete time n). Our formulation mitigates the effect of temporal resolution (and time-window size) by equating the probability of not moving in the two cases.

The transition probability depends on time. When there are $n_{\max}$ time windows, the transition probability for one "cycle" (i.e., one time through the full time period in the temporal sequence of adjacency matrices) is defined as

$$T^{\rm tp} \equiv T(1)T(2)\cdots T(n_{\max})\,. \tag{137}$$

Using periodic boundary conditions (i.e., by having the last adjacency matrix $A(n_{\max})$ loop back to $A(1)$), the "stationary density" at node v_i is given by the ith element of $\boldsymbol{u}(1)$, where

$$\boldsymbol{u}(1) = \boldsymbol{u}(1)T^{\rm tp}\,. \tag{138}$$

There is no stationary density in the present RW process in the conventional sense, because the network is changing in time. Due to the periodic boundary conditions, the stationary density of walkers at each node differs across time periods. The vector $\boldsymbol{u}(1)$ represents the stationary density when the RW is observed right after time $n_{\max}$ (and before time 1) in each cycle. One defines the TempoRank vector based on the running mean of the stationary density over all time periods. That is, it is given by $\boldsymbol{u}^{\rm avg} \equiv \sum_{n=1}^{n_{\max}} \boldsymbol{u}(n)/n_{\max}$, where $\boldsymbol{u}(n)$ is the stationary density when the observation is made right after time $n-1$ (and before time n).

5.2.4. Random-walk betweenness centrality

In our discussions of ranking methods, we have discussed centrality measures (e.g., PageRank) that are derived from RWs. RWs are also useful for deriving

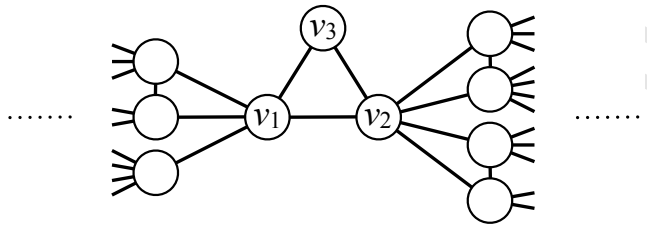

Figure 9: A network with two clearly distinguished communities.

variants of other familiar centrality measures, such as "betweenness centrality".

Shortest-path betweenness centrality (i.e., geodesic betweenness centrality) of a node is defined from a normalized count of the shortest paths that pass through a focal node for all pairs of distinct source and target nodes in a network [44, 320]. Specifically, the shortest-path betweenness of node v_i is

$$b_i^{\mathrm{geo}} = \sum_{i_{\mathrm{s}} \neq i_{\mathrm{t}}} \frac{(\text{number of shortest paths from } v_{i_{\mathrm{s}}} \text{ to } v_{i_{\mathrm{t}}} \text{ that pass through } v_i)}{N(N-1) \times (\text{number of shortest paths from } v_{i_{\mathrm{s}}} \text{ to } v_{i_{\mathrm{t}}})}, \tag{139}$$

where the nodes v_i, $v_{i_{\mathrm{s}}}$, and $v_{i_{\mathrm{t}}}$ are all distinct. However, restricting to strictly shortest paths can be problematic [321]. For example, consider the network in Fig. 9 that includes two communities of densely-connected nodes. Nodes v_1 and v_2 have large betweenness-centrality values because any shortest path connecting one node in each community must pass through both v_1 and v_2. However, because such a shortest path does not pass through v_3, the shortest-path betweenness of node v_3 is 0, yet v_3 may be more important than most other nodes in connecting different parts of the network (albeit to a lesser extent than v_1 and v_2). One can capture this intuition by allowing paths that are longer than the strictly shortest ones to contribute to the value of a betweenness centrality. One way to do this is to use RWs [144, 321].

We now explain the "RW betweenness centrality" introduced in Ref. [321]. Consider an undirected network. Similar to the definition of shortest-path betweenness centrality, we specify the starting node $v_{i_{\mathrm{s}}}$ and terminal node $v_{i_{\mathrm{t}}}$ of an RW. Intuitively, RW betweenness centrality of a node v_i measures the number of times that a random walker starting from $v_{i_{\mathrm{s}}}$ passes through v_i before reaching $v_{i_{\mathrm{t}}}$. If we do not specify $v_{i_{\mathrm{t}}}$, a walker wanders forever in the network, and the centrality of v_i is proportional to s_i [see Eq. (31)]. In RW betweenness centrality, one still discounts long walks, because a walk terminates once a walker reaches $v_{i_{\mathrm{t}}}$.

The RW betweenness centrality of node v_i as

$$b_i^{\mathrm{rw}} \propto \sum_{i_{\mathrm{s}}=1}^{N} \sum_{i_{\mathrm{t}}=1}^{i_{\mathrm{s}}-1} (\text{number of times that a walker starting at } v_{i_{\mathrm{s}}} \text{ and terminating at } v_{i_{\mathrm{t}}} \text{ "effectively" visits } v_i) \tag{140}$$

Note that the "effective" number of transitions between nodes v_i and $v_j \in \mathcal{N}_i$ is equal to the difference (in absolute value) between the number of times that a walker moves from node v_i to node v_j and the number of times that it moves from node v_j to node v_i. An effective transition from v_ℓ to v_i and then to a different node v_j (with $j \neq \ell$) completes an effective visit to v_i. Therefore, the number of effective visits to v_i on the right-hand side of Eq. (140) is given by $\sum_{j\in\mathcal{N}_i}$ (number of effective transitions between v_i and v_j)/2.

Because an RW on a network is related to a corresponding electric circuit on the same network [1, 35, 41, 44, 118, 119], we also discuss a centrality based on electric circuits and then relate it to RW betweenness centrality b_i^{rw}. Consider an electric circuit in which one injects a unit current at node v_{i_s} and drains it at v_{i_t}. Suppose that each edge has a conductance of A_{ij}, and let V_i denote the voltage at node v_i. Kirchhoff's current law at each v_i implies that

$$\sum_{j=1}^{N} A_{ij}(V_i - V_j) = \delta_{i,i_\mathrm{s}} - \delta_{i,i_\mathrm{t}}\,. \tag{141}$$

The left-hand side of Eq. (141) represents the current that flows from node v_i to node v_j for each $j \in \{1,\ldots,N\}$. Because

$$\sum_{j=1}^{N} A_{ij} = s_i\,, \tag{142}$$

we rewrite Eq. (141) as

$$(D-A)\boldsymbol{V} = L\boldsymbol{V} = \boldsymbol{I}^{\mathrm{curr}}\,, \tag{143}$$

where $\boldsymbol{V} = (V_1,\ \ldots\ , V_N)^\top$, the quantity $\boldsymbol{I}^{\mathrm{curr}}$ is the column vector of size N given by

$$I_i^{\mathrm{curr}} = \begin{cases} 1\,, & (i = i_\mathrm{s})\,, \\ -1\,, & (i = i_\mathrm{t})\,, \\ 0\,, & (i \notin \{i_\mathrm{s}\,, i_\mathrm{t}\})\,, \end{cases} \tag{144}$$

and we recall that L is the combinatorial Laplacian matrix.

Because L does not have full rank, Eq. (143) does not have N independent solutions, even though it consists of a set of N linear equations with unknowns V_i (with $i \in \{1,\ldots,N\}$). Therefore, we delete an arbitrary i_0th row from L, corresponding to setting $V_{i_0} = 0$, without loss of generality. As in Section 3.2.5, we also delete the i_0th row and column from D and A to yield $(N-1)\times(N-1)$ matrices $\overline{D}^{(i_0)}$ and $\overline{A}^{(i_0)}$, respectively. Similarly, we remove the i_0th element from $\boldsymbol{V}$ and $\boldsymbol{I}^{\mathrm{curr}}$ to obtain $(N-1)$-dimensional vectors $\overline{\boldsymbol{V}}^{(i_0)}$ and $\overline{\boldsymbol{I}}^{\mathrm{curr}(i_0)}$, respectively. Equation (143) is thus equivalent to

$$(\overline{D}^{(i_0)} - \overline{A}^{(i_0)})\overline{\boldsymbol{V}}^{(i_0)} = \overline{\boldsymbol{I}}^{\mathrm{curr}(i_0)}\,. \tag{145}$$

For a connected network, the matrix $\overline{D}^{(i_0)} - \overline{A}^{(i_0)}$ has full rank, and we obtain

$$\overline{\boldsymbol{V}}^{(i_0)} = (\overline{D}^{(i_0)} - \overline{A}^{(i_0)})^{-1} \overline{\boldsymbol{I}}^{\mathrm{curr}(i_0)} . \tag{146}$$

We now reinsert the i_0th row and column of $(\overline{D}^{(i_0)} - \overline{A}^{(i_0)})^{-1}$ by filling them with 0s, and we denote the resulting $N \times N$ matrix by $\overline{R} = (\overline{R}_{ij})$. Substituting Eq. (144) into Eq. (146) then yields

$$V_i = \overline{R}_{i,i_\mathrm{s}} - \overline{R}_{i,i_\mathrm{t}} . \tag{147}$$

Note that Eq. (147) satisfies the condition $V_{i_0} = 0$. The total current that flows through node v_i is

$$\mathrm{Current}_i^{i_\mathrm{s},i_\mathrm{t}} = \begin{cases} \dfrac{1}{2} \displaystyle\sum_{j=1}^{N} A_{ij} \left| V_i - V_j \right| = \frac{1}{2} \sum_{j=1}^{N} A_{ij} \left| \overline{R}_{i,i_\mathrm{s}} - \overline{R}_{i,i_\mathrm{t}} - \overline{R}_{j,i} + \overline{R}_{j,i} \right| & (i \notin \{i_\mathrm{s}, i_\mathrm{t}\}) , \\ 1 & (i \in \{i_\mathrm{s}, i_\mathrm{t}\}) . \end{cases} \tag{148}$$

The division by 2 in the first case of Eq. (148) arises from the fact the same current is counted twice when it flows into and out of v_i.

One can show that RW betweenness centrality is equal to

$$b_i^{\mathrm{rw}} = \frac{\displaystyle\sum_{i_\mathrm{s}=1}^{N} \sum_{i_\mathrm{t}=1}^{i_\mathrm{s}-1} \mathrm{Current}_i^{i_\mathrm{s},i_\mathrm{t}}}{N(N-1)/2} . \tag{149}$$

That is, it is the normalized frequency that a random walker visits node v_i before it reaches v_{i_t}. To verify Eq. (149), let's consider a DTRW with an absorbing boundary at v_{i_t}. The transition-probability matrix consists of the elements

$$T'_{ij} = \begin{cases} \frac{A_{ij}}{s_i} & (i \neq i_\mathrm{t}) , \\ \delta_{i_\mathrm{t} j} & (i = i_\mathrm{t}) . \end{cases} \tag{150}$$

The matrix T' is equal to the transition-probability matrix of a DTRW with an absorbing boundary, so T' is equal to $D^{-1}A$ except in the i_tth row. We remove the i_tth row and column from T', D^{-1}, and A to obtain

$$\overline{T}'^{(i_\mathrm{t})} = \left(\overline{D}^{(i_\mathrm{t})} \right)^{-1} \overline{A}^{(i_\mathrm{t})} . \tag{151}$$

Whenever the row sum of $\overline{T}'$ is less than 1, the walk is absorbed at v_{i_t} with the residual probability.

Consider an RW that starts from node v_{i_s}. The probability that a random walker visits v_i (with $i \neq i_\mathrm{t}$) after n steps is given by the (i_s, i)th element of $\left(\overline{T}'^{(i_\mathrm{t})} \right)^n$. (For clarity, we use the indices $1, \ldots, i_\mathrm{t}-1, i_\mathrm{t}+1, \ldots, N$ rather than $1, \ldots, N-1$ for the elements of $\overline{T}'$.) Conditioned on this event, the probability that

the walker moves to node v_j in the next step is equal to $1/k_i$. The expected number of times that the walker steps from node v_i to a neighboring node $v_j \in \mathcal{N}_i$ is

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{\left(\left(\overline{T}'^{(i_t)}\right)^n\right)_{i_s i}}{k_i} &= \frac{\left(\left[I - \overline{T}'^{(i_t)}\right]^{-1}\right)_{i_s i}}{k_i} \\ &= i\text{th element of } \left(\overline{\boldsymbol{I}}^{\mathrm{curr}(i_t)}\right)^{\top} \left[I - \left(\overline{D}^{(i_t)}\right)^{-1} \left(\overline{A}^{(i_t)}\right)\right]^{-1} \left(\overline{D}^{(i_t)}\right)^{-1} \\ &= i\text{th element of } \left(\overline{\boldsymbol{I}}^{\mathrm{curr}(i_t)}\right)^{\top} \left(\overline{D}^{(i_t)} - \overline{A}^{(i_t)}\right)^{-1} . \qquad (152)\end{aligned}$$

Because $\overline{D}^{(i_t)}$ and $\overline{A}^{(i_t)}$ are symmetric matrices, the left-hand side of Eq. (152) is also equal to the ith element of $\left[\left(\overline{\boldsymbol{I}}^{\mathrm{curr}(i_t)}\right)^{\top} \left(\overline{D}^{(i_t)} - \overline{A}^{(i_t)}\right)^{-1}\right]^{\top} = \left(\overline{D}^{(i_t)} - \overline{A}^{(i_t)}\right)^{-1} \overline{\boldsymbol{I}}^{\mathrm{curr}(i_t)}$. Therefore, Eq. (146) guarantees that the quantity $\sum_{n=0}^{\infty}([(\overline{T}'^{(i_t)})^n]_{i_s i}/k_i)$ is equal to voltage V_i when $v_{i_0} = v_{i_t}$. Finally, the "effective" number of transitions — i.e., the difference between the number of times that a walker moves from node v_i to node v_j and the number of times that it moves from node v_j to node v_i — is equal to $|V_i - V_j|$.

We now consider "RW centrality" [144], another a variant of RW betweenness centrality. This centrality quantifies the speed at which a walker starting from node v_i reaches other nodes compared to the speed at which a walker starting from an arbitrary node reaches v_i. To formalize this idea, we use Eq. (69), which gives the MFPT m_{ij} from node v_i to node v_j, and we focus on undirected networks. One measures the importance of node v_i relative to node v_j by calculating

$$m_{ij} - m_{ji} = \left(\sum_{\ell=1}^{N} s_\ell\right) \times \left[\left(\frac{R_{jj}^{(0)}}{s_j} - \frac{R_{ii}^{(0)}}{s_i}\right) - \left(\frac{R_{ij}^{(0)}}{s_j} - \frac{R_{ji}^{(0)}}{s_i}\right)\right] . \qquad (153)$$

For undirected networks, the following detailed balance, which extends Eq. (32), holds [144]:

$$\begin{aligned} s_i p_{ij}(n) &= s_i \sum_{\ell_1,\ell_2,\ldots,\ell_{n-1}=1}^{N} \frac{A_{i\ell_1}}{s_i} \frac{A_{\ell_1\ell_2}}{s_{\ell_1}} \times \frac{A_{\ell_{n-1}j}}{s_{\ell_{n-1}}} \\ &= \sum_{\ell_1,\ell_2,\ldots,\ell_{n-1}=1}^{N} \frac{A_{i\ell_1}}{s_{\ell_1}} \frac{A_{\ell_1\ell_2}}{s_{\ell_2}} \times \frac{A_{\ell_{n-1}j}}{s_j} s_j = s_j p_{ji}(n) . \qquad (154)\end{aligned}$$

Substituting Eq. (154) into Eq. (65) yields

$$\begin{aligned}\frac{R_{ij}^{(0)}}{s_j} &= \frac{\sum_{n=0}^{\infty}\left[p_{ij}(n)-p_j^{\infty}\right]}{s_j} \\ &= \frac{\sum_{n=0}^{\infty}\left[\frac{s_j p_{ji}(n)}{s_i}-\frac{s_j}{\sum_{\ell=1}^{N}s_\ell}\right]}{s_j} \\ &= \frac{\sum_{n=0}^{\infty}\left[p_{ji}(n)-\frac{s_i}{\sum_{\ell=1}^{N}s_\ell}\right]}{s_i} = \frac{R_{ji}^{(0)}}{s_i}\,. \end{aligned} \tag{155}$$

We then apply Eq. (155) to Eq. (153) to obtain

$$m_{ij}-m_{ji} = C_{\rm rw}(j)^{-1} - C_{\rm rw}(i)^{-1}\,, \tag{156}$$

where

$$\begin{aligned} C_{\rm rw}(i) &\equiv \frac{s_i}{R_{ii}^{(0)}\sum_{\ell=1}^{N}s_\ell} \\ &= \frac{s_i}{\sum_{n=0}^{\infty}\left[p_{ii}(n)-\frac{s_i}{\sum_{\ell=1}^{N}s_\ell}\right]\sum_{\ell=1}^{N}s_\ell} \end{aligned} \tag{157}$$

is defined to be the RW centrality.

5.2.5. Discrete-choice models

Discrete-choice models describe decisions between distinct alternatives [322, 323]. Examples of discrete choices occur in everyday life; for example, one can choose to shop at a given store, use a specific mode of transportation, or root for the Los Angeles Dodgers instead of some other baseball team. In many applications, one faces the problem of "rank aggregation" [324], as it is necessary to aggregate preferences about an item over a set of alternatives, which one observes for different individuals, who have different subsets of alternatives. For example, the Bradley–Terry–Luce (BTL) model defines the probability to select alternative i (where $i \in \{1,\ldots,N\}$) over alternative j in a pairwise comparison as

$$p_{ij} = \frac{\gamma_i}{\gamma_i+\gamma_j}\,, \tag{158}$$

where $\gamma_i > 0$ is a latent parameter that encodes the attractiveness of alternative i [325, 326].

The pairwise-choice Markov chain (PCMC) model is a discrete-choice model that uses the stationary density of a CTRW as the probability to select i among several alternatives [327]. In the PCMC model, one considers a Poissonian edge-centric CTRW on an N-node directed and weighted network. An individual can choose an item from a subset S of the N alternatives (i.e., nodes). Instead of using the network's adjacency matrix A to construct a transition-rate matrix for a CTRW on the entire network (see Eq. (78)), the PCMC model uses A to define a transition-rate matrix $Q_S = (q_{ij})$ on S. The rows and columns of

Q_S are indexed by the elements in S, and they are defined by $q_{ij} = A_{ij}$ (for $j \neq i$) and $q_{ii} = -\sum_{j \in S \setminus i} q_{ij}$. For any set S, note that Q_S does not require the diagonal elements of A, so we assume that they are 0. The PCMC model uses the stationary density of the CTRW on S as the probability that an individual chooses i when S is the set of alternatives. One can then estimate the matrix A from, for example, empirical-choice data.

A generalization of the BTL model is the multinomial logit model (also called the Plackett–Luce moel) [326, 328, 329], which treats the case of a choice among more than two alternatives. The multinomial logit model defines the probability p_{iS} to choose i from S as

$$p_{iS} = \frac{\gamma_i}{\sum_{j \in S} \gamma_j} \,. \tag{159}$$

This model is a PCMC model, where the adjacency matrix is determined by the BTL model, so $A_{ji} = \gamma_i/(\gamma_i + \gamma_j)$. A large γ_i value makes A_{ji} large, which in turn results in a large probability in-flow to the ith node and an increased probability that an individual chooses i. In fact, the vector $\boldsymbol{p}^* = (p_{iS})$, with $i \in S$, is the stationary density of the CTRW on S, because

$$\begin{aligned}(\boldsymbol{p}^* Q_S)_i =& \frac{1}{\sum_{\ell \in S} \gamma_\ell} \left(\sum_{j \in S; j \neq i} \gamma_j A_{ji} - \gamma_i \sum_{j \in S; j \neq i} A_{ij} \right) \\ =& \frac{\gamma_i}{\sum_{\ell \in S} \gamma_\ell} \left(\sum_{j \in S; j \neq i} \frac{\gamma_j}{\gamma_i + \gamma_j} - \sum_{j \in S; j \neq i} \frac{\gamma_j}{\gamma_i + \gamma_j} \right) = 0 \quad (i \in \{1, \ldots, N\}) \,.\end{aligned} \tag{160}$$

Consider a data set given in the form of $\mathcal{D} = \{(i_\ell, S_\ell) | \ell = 1, \ldots, \ell_{\max}\}$, where S_ℓ is the set of the items presented in the ℓth choice, $i_\ell \in S_\ell$ is the item chosen in the ℓth choice, and $\ell_{\max}$ is the number of choices. The PCMC in which the parameters (i.e., entries of A) are estimated by a maximum-likelihood method yields a better predictive performance than benchmark discrete-choice models on two empirical data sets [327].

One can also derive the maximum-likelihood estimator of the multinomial logit model as the stationary density of a Poissonian edge-centric CTRW [330]. The likelihood $\tilde{L}$ of the parameters $\boldsymbol{\gamma} \equiv \{\gamma_1, \ldots, \gamma_N\}$ given data $\mathcal{D}$ is

$$\tilde{L}\left(\boldsymbol{\gamma} | \mathcal{D}\right) = \prod_{\ell=1}^{\ell_{\max}} \frac{\gamma_{i_\ell}}{\sum_{i' \in S_\ell} \gamma_{i'}} \,. \tag{161}$$

By maximizing the log likelihood, one obtains

$$\begin{aligned}\frac{\partial(\log \tilde{L})}{\partial \hat{\gamma}_i} =& \frac{\partial}{\partial \hat{\gamma}_i} \sum_{\ell=1}^{\ell_{\max}} \left(\log \hat{\gamma}_{i_\ell} - \log \sum_{i' \in S_\ell} \hat{\gamma}_{i'} \right) \\ =& \sum_{\ell=1;\ell \in \breve{W}_i}^{\ell_{\max}} \left(\frac{1}{\hat{\gamma}_i} - \frac{1}{\sum_{i' \in S_\ell} \hat{\gamma}_{i'}} \right) - \sum_{\ell=1;\ell \in \breve{L}_i}^{\ell_{\max}} \frac{1}{\sum_{i' \in S_\ell} \hat{\gamma}_{i'}} \\ =& 0 \quad (i \in \{1, \dots, N\}) \,, \end{aligned} \tag{162}$$

where $\breve{W}_i = \{\ell | i \in S_\ell \text{ and } i \text{ is chosen}\}$, $\breve{L}_i = \{\ell | i \in S_\ell \text{ and } i \text{ is not chosen}\}$, and $\hat{\gamma}_i$ (with $i \in \{1, \dots, N\}$) is the maximum-likelihood estimator. By multiplying $\hat{\gamma}_i$ by Eq. (162), one obtains

$$\sum_{\ell=1;\ell \in \breve{W}_i}^{\ell_{\max}} \frac{\sum_{j \in S_\ell; j \neq i} \hat{\gamma}_j}{\sum_{i' \in S_\ell} \hat{\gamma}_{i'}} - \sum_{\ell=1;\ell \in \breve{L}_i}^{\ell_{\max}} \frac{\hat{\gamma}_i}{\sum_{i' \in S_\ell} \hat{\gamma}_{i'}} = 0 \quad (i \in \{1, \dots, N\}) \,. \tag{163}$$

Because $\breve{L}_i = \cup_{j=1; j \neq i}^{N} (\breve{W}_j \cap \breve{L}_i)$, one can rewrite Eq. (163) as

$$\sum_{j=1; j \neq i}^{N} \left[\sum_{\ell=1;\ell \in \breve{W}_i \cap \breve{L}_j}^{N} \frac{\hat{\gamma}_j}{\sum_{i' \in S_\ell} \hat{\gamma}_{i'}} - \sum_{\ell=1;\ell \in \breve{W}_j \cap \breve{L}_i}^{N} \frac{\hat{\gamma}_i}{\sum_{i' \in S_\ell} \hat{\gamma}_{i'}} \right] = 0 \quad (i \in \{1, \dots, N\}) \,. \tag{164}$$

One rewrites Eq. (164) as

$$\sum_{j=1; j \neq i}^{N} \hat{\gamma}_i f(\mathcal{D}_{j \succ i}, \hat{\boldsymbol{\gamma}}) = \sum_{j=1; j \neq i}^{N} \hat{\gamma}_j f(\mathcal{D}_{i \succ j}, \hat{\boldsymbol{\gamma}}) \quad (i \in \{1, \dots, N\}) \,, \tag{165}$$

where

$$f(\mathcal{D}', \hat{\boldsymbol{\gamma}}) = \sum_{S \in \mathcal{D}'} \frac{1}{\sum_{i' \in S} \hat{\gamma}_{i'}} \,, \tag{166}$$

$\mathcal{D}' \subset \mathcal{D}$ is a subset of the observation set $\mathcal{D}$, and $\mathcal{D}_{i \succ j} = \{(i_\ell, S_\ell) \in \mathcal{D} | \ell \in \breve{W}_i \cap \breve{L}_j\} \subset \mathcal{D}$ is the set of observations in which i is preferred to j. Equation (165) implies that the maximum-likelihood estimator is the stationary density of the CTRW whose transition rate from the jth to the ith node is given by $f(\mathcal{D}_{i \succ j}, \hat{\boldsymbol{\gamma}})$. One interprets $f(\mathcal{D}_{i \succ j}, \hat{\boldsymbol{\gamma}}) = \sum_{S \in \mathcal{D}_{i \succ j}} \left(1 / \sum_{i' \in S} \hat{\gamma}_{i'}\right)$ as the number of times i is chosen over j (taken into accounted by the sum $\sum_{S \in \mathcal{D}_{i \succ j}}$), weighted by the strength of the alternatives in each observation (which is taken into account with the term $1 / \sum_{i' \in S} \hat{\gamma}_{i'}$). Taking advantage of this relationship between the CTRW and the maximum-likelihood estimator of the multinomial logit model has resulted in inference algorithms for the multinomial logit model that is faster and more accurate than previous methods for several data sets [330].

For other methods of rank aggregation based on RWs, see Refs. [324, 331, 332].

5.3. Community detection

A useful approach for studying networks is to examine mesoscale structures, of which the best-known type is "community structure" [63–65]. There are numerous methods to algorithmically detect communities (and many applications in which communities can be insightful), which are sets of densely connected nodes such that connections between different communities are relatively sparse. RWs provide a theoretical basis for understanding community structure and practical algorithms for detecting them. The main idea is that, if a given network has community structure, a random walker should be trapped within a community for a relatively long time before leaving it. This arises from the high density of edges within communities and the sparse connections across communities. Therefore, RWs that are observed on a short time scale should reveal intra-community structure in a network, and RWs that are observed on a long time scale should reveal global structure about the same network.

In this section, we introduce some algorithms for community detection that are based on RWs. For other RW-based algorithms and theoretical underpinnings, see papers such as Refs. [68, 236, 333–342].

5.3.1. Markov-stability formulation of modularity

It is common to use the "modularity" objective function Q to quantify the quality of a partition of a network into nonoverlapping communities, and many community-detection methods are based on maximizing Q [65]. Consider a partition of an undirected network into N_{CM} communities. Let CM_c denote the cth community (with $c \in \{1, 2, \ldots, N_{\mathrm{CM}}\}$). We use a variant (sometimes called the "Newman–Girvan null model") of an undirected configuration model [122] that is defined as a random graph with a specified strength s_i at each node. For this configuration model, the probability that nodes v_i and v_j are adjacent is approximately $P_{ij} \equiv s_i s_j/(2M')$, where $M' = \sum_{i=1}^{N} s_i/2$ is the sum of the edge weight over all edges [44]. (Technically, P_{ij} is a probability only for sufficiently small edge weights; otherwise, it is an expectation.) Note that $M' = M$ for an unweighted network, where we recall that M is the number of edges. Modularity is defined by

$$\begin{aligned} Q &= \frac{1}{2M'} \sum_{c=1}^{N_{\mathrm{CM}}} \left[\sum_{\substack{i,j=1; \\ v_i, v_j \in \mathrm{CM}_c}}^{N} \left(A_{ij} - \frac{s_i s_j}{2M'} \right) \right] \\ &= \frac{1}{2M'} \sum_{i,j=1}^{N} \left(A_{ij} - \frac{s_i s_j}{2M'} \right) \delta(g_i, g_j) \,, \end{aligned} \tag{167}$$

where g_i is the community to which node v_i has been assigned, and $\delta(g_i, g_j) = 1$ if $g_i = g_j$ and $\delta(g_i, g_j) = 0$ otherwise. The quantity P_{ij} gives the elements of a null-model matrix, and a wide variety of different versions of the matrix $P = (P_{ij})$ have been examined [343, 344]. More precisely, P is not a "null

model" but rather a "null network" (which is a network generated from a null model) [344].

Methods based on modularity maximization suffer from the fact that Q has a resolution limit, so using Eq. (167) does not allow one to detect dense communities of nodes that are smaller than a certain scale [345, 346] (though some null models attempt to address this issue). Modularity maximization also implicitly favors communities of a particular size that depend on the size of the entire network (not only its internal structure), and methods based on maximizing Q also have various other problematic features [65].

One can use RWs to gain insights into modularity and its resolution issues. Modularity is closely related to "Markov stability", which quantifies the tendency for a random walker to stay inside a community for a long time. The Markov stability of a partition of a network is defined as the probability that a walker is in the same community at time 0 and time t in the equilibrium of the Poissonian node-centric CTRW [347–350]. See Refs. [350, 351] for a version of Markov stability derived from a DTRW.

The master equation is

$$\frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = -\boldsymbol{p}(t)L' \,, \tag{168}$$

where we recall that L' is the random-walk normalized Laplacian matrix [see Eq. (77)]. The stationary density is given by Eq. (31).

Consider a pair of nodes, v_i and v_j, that belong to the same community. Equation (168) implies that, in the stationary state, the probability that a random walker visits v_i and then v_j after time t is equal to $p_i^*(e^{-tL})_{ij}$. As with modularity maximization, one needs to compare this quantity with a null model. For Markov stability $R(t)$, the standard null model is given by the probability that a walker visits node v_i at $t = 0$ and node v_j at $t = \infty$. This yields a null probability of $p_i^* p_j^*$. One thereby obtains a Markov stability of

$$R(t) = \sum_{i,j=1}^{N} \left[\left(p_i^* e^{-tL'} \right)_{ij} - p_i^* p_j^* \right] \delta(g_i, g_j) \,. \tag{169}$$

Because of the exponential factor e^{-tL}, Markov stability combines walks of various lengths between two nodes. The time t acts as a resolution parameter, enabling one to zoom in and out to unravel multiscale structure in a network. A large value of t gives large weightings to long walks and yields a small number of communities. In the limit $t \to \infty$, Markov stability is optimized by the bipartition given by the signs of the elements of the Fiedler vector (i.e., a type of spectral partitioning) if the corresponding eigenvalue is not degenerate [338]. More generally, spectral partitioning is related to RWs on networks because it uses the eigenvectors of matrices such as the combinatorial Laplacian matrix or a modularity matrix [88, 352].

Because it is computationally expensive to calculate $e^{-tL'}$ for large networks, we use a linear approximation $e^{-tL'} \approx I - tL'$. To simplify our exposition, we

now assume the case of undirected networks for the rest of this section [350]. By substituting $p_i^* = s_i/(2M')$ and $p_j^* = s_j/(2M')$ into Eq. (169), we obtain

$$R(t) = \frac{1}{2M'} \sum_{i,j=1}^{N} \left[tA_{ij} + (1-t)\delta_{ij} s_i + \frac{s_i s_j}{2M'} \right] \delta(g_i, g_j) \,. \tag{170}$$

Because $\sum_{i,j=1}^{N} (1-t)\delta_{ij} s_i \delta(g_i, g_j) = \sum_{i=1}^{N} s_i$ does not depend on the partitioning of a network, maximizing $R(t)$ is equivalent to maximizing

$$Q(\gamma) = \frac{1}{2M'} \sum_{i,j=1}^{N} \left(A_{ij} - \gamma \frac{s_i s_j}{2M'} \right) \delta(g_i, g_j) \,, \tag{171}$$

where $\gamma \equiv 1/t$. We ignore the constraint that t is small (which is admittedly naughty mathematically) and thereby allow general values for γ when maximizing $Q(\gamma)$. We also note that $Q(\gamma)$ was derived originally using the perspective of a Potts spin glass [353], and recently it has been related to maximum-likelihood methods [354].

When $\gamma = 1$, Eq. (171) coincides with Eq. (167). Therefore, modularity is an approximate variant of Markov stability. A large value of γ emphasizes the penalty for classifying nodes into the same community and results in many communities. The choice of the natural resolution parameter γ is an important practical issue [352, 355], and it can be examined from a maximum-likelihood approach [354].

5.3.2. Walktrap

In the Walktrap algorithm, one defines a measure of similarity between nodes based on DTRWs and uses it for community detection [356]. (See Ref. [357] for a similar method that uses DTRWs.) Consider an undirected and unweighted network. Define the RW-based distance between two nodes, v_i and v_j, by

$$r_{ij} = \sqrt{\sum_{\ell=1}^{N} \frac{(T_{i\ell}^n - T_{j\ell}^n)^2}{k_\ell}} \,, \tag{172}$$

where n is the number of steps in a DTRW. The distance r_{ij} is small when a pair of random walkers — one starting from v_i and the other starting from v_j — visit each node with similar probabilities after n steps. The denominator k_ℓ discounts the fact that a walker visits v_ℓ with a probability proportional to k_ℓ at equilibrium. Note that n needs to be large enough for random walkers to be able to travel to any node. However, n should not be too large, because $\lim_{n\to\infty} T_{i\ell}^n = \lim_{n\to\infty} T_{j\ell}^n = p_\ell^*$ implies that r_{ij} is very close to 0 for all $i, j \in \{1, \ldots, N\}$ when n is large [64].

We expect that a pair of nodes, v_i and v_j, that are separated by a small distance r_{ij} are likely to belong to the same community. One uses a standard agglomerative and hierarchical clustering algorithm on the distance matrix

$r = (r_{ij})$. One starts from the partition composed of N single-node communities and joins a pair of communities (so-called "tentative communities") with the smallest distance, one pair at time, to produce a series of partitions until the entire network is in a single community. In the merging process, one measures the distance between two communities CM_c and $\mathrm{CM}_{c'}$ by the r_{ij} value, normalized in some way, between $v_i, v_j \in \mathrm{CM}_c \cup \mathrm{CM}_{c'}$. This agglomerative clustering algorithm is similar to a greedy algorithm to maximize modularity across partitionings with different numbers of communities [358]. In Walktrap, one merges a pair of communities under the restriction that they can be merged only when they are adjacent to each other by at least one edge.

Other community-detection methods also rely on defining a similarity measure between nodes. An interesting approach is based on the concept of mean first-passage time m_{ij} of a random walker (see Section 3.2.5) and its symmetrization $m_{ij} + m_{ji}$ (the so-called "mean commute time") [359]. The square root of the mean commute time has the desirable property of being a Euclidian distance between nodes. In this context, it is called the "Euclidian commute-time distance". It decreases when the number of paths between two nodes increases or when the length of any path between the two nodes decreases, and it can be derived from the pseudo-inverse of the combinatorial Laplacian matrix L [360].

5.3.3. InfoMap

InfoMap is another algorithm for community detection based on RWs [361]. It is very popular and has been extended to the case of hierarchical algorithms [362], memory networks [250], and multilayer networks [235]. In this section, we discuss the basic version of InfoMap.

Consider a DTRW on a network, which can be directed or weighted. If the network has meaningful community structure, a random walker tends to be trapped within a community for a long time before traveling to a different community. A trajectory of the RW is a sequence of the visited nodes (e.g., v_3, v_6, v_3, v_1, v_8, ...). Let's encode each node into a finite binary sequence (i.e., "a code word") and concatenate the code words to encode the trajectory of a random walker. For example, if v_1, v_2, v_3, v_4, v_5, ... are encoded into $000, 001, 010, 011, 100, \ldots$, then the trajectory v_3, v_6, v_3, v_1, v_8, ... is encoded into $010101010000111\cdots$. For unique decoding, one needs a "prefix-free" coding scheme. In other words, a code word cannot be a "prefix" (i.e., an initial segment) of another code word. For instance, if v_1 and v_2 are coded as 000 and 0001, respectively, then one's code is not prefix-free, because 000 is an initial segment of 0001.

The "Huffman code" is a popular prefix-free code that encodes individual symbols (i.e., nodes v_i) separately and tends to yield short binary sequences [363]. It assigns a short code word to a frequently visited node. In a stationary state, the mean code word length per step of an RW is $\sum_{i=1}^{N} p_i^* \times \mathrm{len}(i)$, where $\mathrm{len}(i)$ denotes the length of the code word assigned to v_i.

If symbols (such as v_i in our context) appear independently in each step of an RW, the Huffman code yields a mean code word length in each step that is

close to the theoretical lower bound set by the Shannon entropy

$$H = -\sum_{i=1}^{N} p_i^* \log p_i^* \,. \tag{173}$$

However, the sequence of symbols is correlated in time, because it is produced by an RW. Consequently, a different coding scheme can yield a mean code length that is smaller than the Shannon entropy. InfoMap exploits community structure and uses a two-layer variant of the Huffman code to achieve this goal. Because there are fewer nodes in a community than in an entire network, one can express a trajectory within each community using a shorter, different Huffman code that is local to individual communities. In practice, one constructs the two-layer Huffman code as follows:

1. When a random walker enters the cth community, one issues the (predetermined) code word that corresponds to entering community CM_c.
2. The walker moves around within community CM_c for some time. One records the trajectory during this period by the sequence of code words that corresponds to the sequence of visited nodes. One concatenates these code words, and they appear after the code word (obtained in the previous step) that corresponds to the entry to community CM_c.
3. The walker eventually exits CM_c. This event is represented by a special code word, which one places after the sequence of code words that one has obtained thus far.
4. The exit from CM_c implies an immediate entry to a different community, which we denote by $\mathrm{CM}_{c'}$. Therefore, we concatenate the code word corresponding to the entry to $\mathrm{CM}_{c'}$ to the end of the sequence of code words that we have obtained thus far.
5. One uses the code words that are local to $\mathrm{CM}_{c'}$ to record the trajectory until the walker exits $\mathrm{CM}_{c'}$. Note that one can use the same code word to represent a node in CM_c and a node in $\mathrm{CM}_{c'}$. This fact does not cause any problems, because one determines the current coding table from the entry and exit code words.
6. Repeat steps 3–5.

Let's consider the network in Fig. 10. The InfoMap algorithm partitions the network into four communities, whose boundaries we show with the dotted lines. The binary sequence at each node represents the local code word within the corresponding community. When a random walker enters or exits a community, one uses the corresponding "in" and "out" code word, respectively. For example, the trajectory indicated by the red arrows is encoded into 11 111 10 01 00 00 10 01 110. The first "11" indicates that the RW starts in the top left community, the subsequent "111" indicates that the walk starts at node "111" in this community, the "00 00" in the middle indicates that the walk exits this community (because of the first "00") and simultaneously enters the community to the right (because of the second "00").

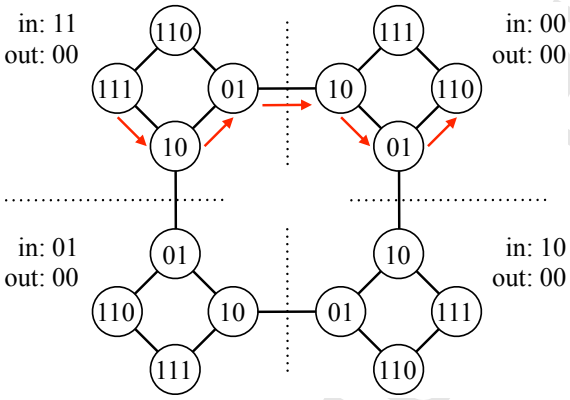

Figure 10: Optimal partitioning from the InfoMap algorithm along with its resulting code words. We draw this example from a demonstration applet available at [364].

In contrast to the original Huffman code, we need $2N_{\mathrm{CM}}$ additional code words to encode entry to and exit from communities. However, we can use a smaller code length when a random walker travels within a community because the code words local to a community are generally shorter than the code words of the original Huffman code. If a network has strong community structure, one expects that an RW within a community occupies a majority of steps if one optimally partitions the network into communities. Consequently, one expects the mean code length to be smaller using InfoMap than by using a straightforward Huffman code in networks with community structure. In practice, InfoMap optimizes a quality function, called the "map equation" (where the word "equation" is a misnomer), instead of constructing the optimized coding scheme. The map equation generalizes Eq. (173). The resulting quality function provides a theoretical limit of how concisely one can encode an RW using a given partition. One can optimize this function using some computational heuristic.

5.3.4. Local community detection

Another approach to community detection is to use local algorithms. For example, given a node v_i of interest, one can use a local algorithm to identify a relatively small community around v_i by examining only the nodes that are adjacent to nodes that have been examined before. Local algorithms are particularly useful when a network is huge, and it is thus costly to apply a partitioning algorithm to the entire network. As discussed in Ref. [68] (and in several references therein), they also provide a means to studying overlapping communities and to incorporate dynamical processes and seed sets into community detection.

Nibble is a local community-detection algorithm based on DTRWs [365–367]. The idea is to examine nodes that are visited frequently by a random walker that starts from a node v_i. Specifically, Nibble uses the transition-probability

matrix

$$T_{\text{Nibble}} = \frac{D^{-1}A + I}{2} \,. \tag{174}$$

Equation (174) implies that a random walker obeys the usual DTRW with probability 1/2 and does not move with probability 1/2 in each time step. For each of the nodes, Nibble also reduces the probability of a visit to it to 0 in each time step if it is smaller than some threshold. Therefore, the probability that the random walker is still present in the network decreases in time. The probability reduction ensures that the detected community does not become too large in a small number n of steps. One terminates the DTRW after a certain number of steps according to a stopping criterion, which guarantees that the discovered set of nodes has a low conductance (see Eq. (47)) and is neither too small nor too large. Nibble can also be used as a building block for network-partitioning algorithms that run in $O(M)$ time [365, 367]. (Recall that M denotes the number of edges (see Table 1).)

In the "seed-set expansion problem", one seeks to discover a local community that emanates from a small subset S of a network's nodes. One expands the seed set to estimate the rest of a community by ranking the nodes outside S. Variants of personalized PageRank and heat-kernel PageRank are popular approaches for studying seed-set expansion [368–370]. Like Nibble, one starts a DTRW from a node $v_i \in S$, and one then examines T_{ij}^n, which gives the probability that a walker starting from v_i visits node v_j after n steps. The score for v_j is given by a weighted sum of T_{ij}^n over different lengths of walks. That is, the score is $\sum_{n=1}^{\infty} w_n T_{ij}^n$, where w_n is the weight assigned to walks of length n [370].

5.3.5. Multilayer modularity

One can generalize Markov stability to multilayer networks to derive modularity functions for such networks, including temporal networks given in the form of a sequence of adjacency matrices (with interlayer edges that connect corresponding nodes in the sequence) [234, 344].

As in Section 4.1, consider a multilayer network in the (supra-adjacency) form of a weighted network on $N\ell_{\max}$ nodes, where $\ell_{\max}$ is the number of layers. One specifies a node by the pair (v_i, ℓ), where $i \in \{1, \dots, N\}$ indexes an entity and $\ell \in \{1, \dots, \ell_{\max}\}$ indicates a layer. The adjacency matrix in each layer ℓ (which can be, e.g., an aggregation over some time window of a temporal network) is $A(\ell)$, which we assume to be undirected for simplicity. The weight of the interlayer edge between nodes (v_i, ℓ) and (v_i, ℓ') is $C_{i\ell'\ell}$. We consider a multilayer network in which only nodes with the same index i can be adjacent to each other, though multilayer networks also allow much more general structures [61]. (Note that an entity v_i need not exist on all layers [234].) For a multilayer network that represents a temporal network, the simplest choice is to connect the corresponding nodes (i.e., nodes with the same index i) across the adjacent layers symmetrically and uniformly, so $\omega = C_{i\ell\ell'} = C_{i\ell'\ell} > 0$ when $\ell' = \ell + 1$ for $\ell \in \{1, \dots, L-1\}$ and $C_{i\ell\ell'} = 0$ for $\ell' \neq \ell \pm 1$.

To derive an expression for multilayer modularity for these "multislice" networks, we generalize the RW interpretation of modularity for time-independent

networks (see Section 5.3.1) to the case of multilayer networks [234]. Random walkers are allowed to move either between layers or within a layer. Consider a Poissonian node-centric CTRW on a multilayer network with $N\ell_{\max}$ nodes. The master equation is given by

$$\frac{\mathrm{d}p_{i\ell}(t)}{\mathrm{d}t} = \sum_{\ell'=1}^{\ell_{\max}} \sum_{j=1}^{N} \frac{[A_{ij}(\ell')\delta_{\ell\ell'} + \delta_{ij}C_{j\ell\ell'}]\, p_{j\ell'}(t)}{\kappa_{j\ell'}} - p_{i\ell}(t)\,, \tag{175}$$

where $\kappa_{j\ell'} = k_{j\ell'} + c_{j\ell'}$ is the strength of the jth node in the ℓ'th layer, $k_{j\ell'} = \sum_{i=1}^{N} A_{ij}(\ell)$ is the intra-layer strength of the jth node in the ℓ'th layer, and $c_{j\ell'} = \sum_{\ell''=1}^{\ell_{\max}} C_{j\ell'\ell''}$ is the inter-layer strength of the same node. The summand on the right-hand side of Eq. (175) represents the rate at which a random walker moves from node (v_j, ℓ') to node (v_i, ℓ). A move to (v_i, ℓ) is possible from the nodes (v_j, ℓ) in the same layer at a rate of $A_{ij}(\ell)/\kappa_{j\ell'}$ and from the ith node in a different layer ℓ' at a rate of $C_{j\ell\ell'}/\kappa_{j\ell'}$. If $C_{i\ell\ell'} = C_{i\ell'\ell}$ (with $i \in \{1, \ldots, N\}$ and $\ell, \ell' \in \{1, \ldots, \ell_{\max}\}$), the stationary density is given by

$$p^*_{i\ell} = \frac{\kappa_{i\ell}}{\sum_{\ell'=1}^{\ell_{\max}} \sum_{i'=1}^{N} \kappa_{i'\ell'}} \equiv \frac{\kappa_{i\ell}}{2\mu}\,. \tag{176}$$

In the same manner as with monolayer networks, we examine the probability that a random walker visits node (v_j, ℓ') at time $t = 0$ and node (v_i, ℓ) at a small time Δt. Within the small time Δt, a walker initially at (v_j, ℓ') can make at most a single step. Based on Eq. (175), the probability that the walker visits node (v_j, ℓ') at time 0 and node (v_i, ℓ) at small time Δt is

$$\left[\delta_{ij}\delta_{\ell\ell'} + \Delta t \left(\frac{A_{ij}(\ell)\delta_{\ell\ell'} + \delta_{ij}C_{j\ell\ell'}}{\kappa_{j\ell'}} - \delta_{ij}\delta_{\ell\ell'}\right)\right] \frac{\kappa_{j\ell'}}{2\mu}\,. \tag{177}$$

Under the independence assumption, which sets the null model, the situation remains the same, but each intra-layer network is now replaced by a Newman–Girvan (NG) null network whose degree distribution is determined by the original set of adjacencies of the same layer [344]. The inter-layer transition probability, determined by $C_{j\ell\ell'}$, remains the same. Under the independence assumption, the probability that a walker visits node (v_j, ℓ') at time $t = 0$ and node (v_i, ℓ) after a single move is

$$\left(\frac{k_{i\ell}}{2M_\ell}\frac{k_{j\ell'}}{\kappa_{j\ell'}}\delta_{\ell\ell'} + \delta_{ij}\frac{C_{j\ell\ell'}}{c_{j\ell'}}\frac{c_{j\ell'}}{\kappa_{j\ell'}}\right)\frac{\kappa_{j\ell'}}{2\mu}\,, \tag{178}$$

where $M_\ell = \sum_{j=1}^{N} k_{j\ell}$. In Eq. (178), $\kappa_{j\ell'}/(2\mu)$ is the probability that the random walker visits (v_j, ℓ') at time 0 at equilibrium. The quantity in parentheses represents the conditional probability that a walker visits node (v_i, ℓ) after a single move starting from node (v_j, ℓ') at time 0. A move occurs within the ℓ'th layer with probability $k_{j\ell'}/\kappa_{j\ell'}$. If an intra-layer move occurs, the walker moves to the ith node in the same layer with probability $k_{i\ell'}/(2M_{\ell'})$ according to the NG null model. Alternatively, the walker moves to a different layer with

probability $c_{j\ell'}/\kappa_{j\ell'} = 1 - k_{j\ell'}/\kappa_{j\ell'}$. If an inter-layer move occurs, the walker moves to the jth node in the ℓth layer with probability $C_{j\ell\ell'}/c_{j\ell'}$.

By subtracting Eq. (178) from Eq. (177) and then summing over nodes (v_i, ℓ) and (v_j, ℓ') that belong to the same community, we obtain

$$Q = \frac{1}{2\mu} \sum_{i,j,\ell,\ell'} \left[(1-\Delta t)\delta_{ij}\delta_{\ell\ell'} + \Delta t A_{ij}(\ell)\delta_{\ell\ell'} - \frac{k_{i\ell}k_{j\ell'}}{2M_\ell}\delta_{\ell\ell'} + (\Delta t - 1)\delta_{ij}C_{j\ell\ell'} \right] \times \delta(g_{i\ell}, g_{j\ell'}) \,, \tag{179}$$

where $g_{i\ell}$ is the community to which node (v_i, ℓ) has been assigned. Because $\sum_{i,j,\ell,\ell'} \delta_{ij}\delta_{\ell\ell'}\delta(g_{i\ell}, g_{j\ell'}) = N\ell_{\max}$ is independent of the partitioning of the multilayer network and thus does not affect the maximization of Q, we ignore the first term on the right-hand side of Eq. (179). By rescaling $C_{j\ell\ell'}$ by a multiplicative factor of $(\Delta t - 1)/\Delta t$, we can also ignore $(\Delta t - 1)$ in the fourth term. If we allow $\gamma \equiv 1/\Delta t$ to depend on the layer (see [234] for the justification), corresponding to different diffusion rates in different layers, we obtain the following formula for multilayer modularity:

$$Q = \frac{1}{2\mu} \sum_{i,j,\ell,\ell'} \left[A_{ij}(\ell) - \gamma(\ell)\frac{k_{i\ell}k_{j\ell'}}{2M_\ell}\delta_{\ell\ell'} + \delta_{ij}C_{j\ell\ell'} \right] \delta(g_{i\ell}, g_{j\ell'}) \,. \tag{180}$$

For simplicity, suppose that the inter-layer edge weight is uniform; that is, $\omega = C_{i\ell\ell'}$ for any i, ℓ, and ℓ' whenever entity v_i exists in both layers. If an entity v_i does not exist in a layer, its associated interlayer edges have weight 0 because they do not exist. If $\omega = 0$, the different layers are independent networks. If ω is sufficiently large, all existing copies (v_i, ℓ) of each node v_i (with $\ell \in \{1, \ldots, \ell_{\max}\}$) are assigned to the same community because the third term on the right-hand side of Eq. (180) dominates the others. More generally, a large value of ω tends to yield a smaller number of communities. In contrast, a large $\gamma(\ell)$ value tends to yield a large number of communities. See Refs. [343, 344, 355, 371] for illustrations and discussions.

5.4. Core–periphery structure

It is often insightful to decompose a network into one or more densely-connected cores along with sparsely-connected peripheral nodes. By definition, nodes in a core are heavily interconnected and also tend to be well-connected to peripheral nodes. By contrast, peripheral nodes are sparsely connected (or, ideally, not adjacent at all) to other peripheral nodes and tend to be adjacent predominantly to core nodes. This idea, whose intuition draws somewhat on the notion of pealing an onion (especially in the case of a single core), is also a mesoscale network structure, but it has a rather different character from community structure. See Ref. [372] for a review of core–periphery, and see the introduction of Ref. [373] for a brief survey.

There is an RW-based algorithm to extract core–periphery structure from networks [374]. The idea is that if a random walker is located at a peripheral

node, it is very unlikely to visit another peripheral node in the next time step in a DTRW. One defines a "persistence probability" α_S for a set of nodes S by

$$\alpha_S = \frac{\sum_{i,j\in S} p_i^* T_{ij}}{\sum_{i\in S} p_i^*} , \tag{181}$$

where we recall that p_i^* is the stationary density at node v_i, and T_{ij} is the transition probability from v_i to v_j in a single move. Equation (181) is the steady-state probability that a DTRW starting from a node in S remains in S in the next time step. For an undirected network, we substitute $p_i^* = s_i / \sum_{\ell=1}^{N} s_\ell$ to reduce Eq. (181) to

$$\alpha_S = \frac{\sum_{i,j\in S} A_{ij}}{\sum_{i\in S} s_i} . \tag{182}$$

Ideally, one obtains $\alpha_S = 0$ for any set S of nodes that includes only peripheral nodes. This condition is trivially satisfied when S consists of a single node, and it becomes very difficult to satisfy as S becomes large. Reference [374] used the following greedy algorithm. Start from a node with the smallest total node strength $s_i^{\text{in}} + s_i^{\text{out}}$. If there are multiple such nodes, we select one of them uniformly at random. For undirected networks, this reduces to selecting a node with the minimum node strength. The set S is composed of a single node. One then adds one node to the set S so that adding this node yields the smallest value of α_S. Again, if there are multiple candidate nodes, we break the tie by selecting one of them uniformly at random. One continues this procedure and sequentially adds nodes to try to keep α_S small. One then assigns each node v_i a coreness value of α_i, which one sets as the value of α_S when v_i is added. Nodes with larger values of α_i are deeper into a network core. One also defines a network's "α-periphery" as the set of nodes that satisfy $\alpha_i \leq \alpha$. Although the algorithm has randomness in it because of the tie-breakers, Ref. [374] reported that the randomness had negligible effects on their results for empirical networks.

5.5. Diffusion maps

Dimension reduction is a type of compression that has numerous practical applications in data mining, image processing, visualization, and many other subjects [375]. Its aim is to find a transformation of a set of data points into a low-dimensional space in a way that preserves quantities of interest, such as distances between any pair of data points, preferably with a small number of free parameters. "Diffusion maps" are a framework of RW-based dimension reduction and encompass a wide variety of methods, such as kernel eigenmap methods, as special cases [24, 25]. Diffusion maps are also useful for identifying synchronous clusters of nodes in synchronization dynamics [376].

Consider a DTRW on an undirected, weighted network constructed from a given set of data points, which one identifies with nodes. The edge weight between nodes v_i and v_j is $A_{ij} = A_{ji}$, and it is given by a similarity value

between the ith and jth data points. In our terminology, the "diffusion distance" is defined by

$$\overline{d}_{ij}(n) = \sqrt{\sum_{\ell=1}^{N} \frac{\left(T_{i\ell}^n - T_{j\ell}^n\right)^2}{p_\ell^*}} = \sqrt{\sum_{\ell=1}^{N} \frac{\left(T_{i\ell}^n - T_{j\ell}^n\right)^2}{s_\ell} \times \sum_{\ell=1}^{N} s_\ell}\,, \tag{183}$$

which is the same as the distance measure used in the Walktrap algorithm, except for the normalization (see Eq. (172)). Because $\overline{d}_{ij}(n)$ involves the summation of all walks of length n starting from v_i and the summation of such walks starting from v_j, Refs. [24, 25] suggested that it is more robust to noise in data than when using A_{ij} as a similarity or distance measure for dimension reduction.

Substituting Eq. (41) into Eq. (183) yields

$$\begin{aligned}\overline{d}_{ij}(n) &= \sqrt{\frac{\sum_{\ell=1}^{N}\left[\sum_{\ell'=1}^{N}\lambda_{\ell'}^n\left(\frac{(u_{\ell'})_i}{\sqrt{s_i}} - \frac{(u_{\ell'})_j}{\sqrt{s_j}}\right)(u_{\ell'})_\ell\sqrt{s_\ell}\right]^2}{s_\ell} \times \sum_{\ell=1}^{N} s_\ell} \\ &= \sqrt{\sum_{\ell=1}^{N}\left[\sum_{\ell'=1}^{N}\lambda_{\ell'}^n\left(\frac{(u_{\ell'})_i}{\sqrt{s_i}} - \frac{(u_{\ell'})_j}{\sqrt{s_j}}\right)(u_{\ell'})_\ell\right]^2 \times \sum_{\ell=1}^{N} s_\ell}\,,\end{aligned} \tag{184}$$

where $\boldsymbol{u}_{\ell'}$ is the eigenvector corresponding to the ℓ'th eigenvalue of $\tilde{A}$ (see Eq. (36)) and $\lambda_{\ell'}$ is the ℓ'th largest eigenvalue of $\tilde{A}$ in terms of absolute value. Note that $\lambda_1 = 1$. Using $\langle \boldsymbol{u}_{\ell'}, \boldsymbol{u}_{\ell''}\rangle = \delta_{\ell'\ell''}$, Eq. (184) reduces to

$$\begin{aligned}\overline{d}_{ij}(n) &= \sqrt{\sum_{\ell'=1}^{N}\lambda_{\ell'}^{2n}\left(\frac{(u_{\ell'})_i}{\sqrt{s_i}} - \frac{(u_{\ell'})_j}{\sqrt{s_j}}\right)^2 \times \sum_{\ell=1}^{N} s_\ell} \\ &= \sqrt{\sum_{\ell'=2}^{N}\lambda_{\ell'}^{2n}\left(\frac{(u_{\ell'})_i}{\sqrt{s_i}} - \frac{(u_{\ell'})_j}{\sqrt{s_j}}\right)^2 \times \sum_{\ell=1}^{N} s_\ell}\,.\end{aligned} \tag{185}$$

To derive the last line in Eq. (185), we used $\boldsymbol{u}_1 = (\sqrt{s_1}, \ldots, \sqrt{s_N})^\top$, corresponding to the stationary density (see Section 3.2.3). By neglecting eigenmodes whose contributions are much smaller than the largest eigenmode in Eq. (185) (i.e., $\boldsymbol{u}_2$), one defines a diffusion map by

$$\Psi(i;n) = \frac{1}{\sqrt{s_i}}\begin{pmatrix}\lambda_2^n(u_2)_i \\ \vdots \\ \lambda_{\tilde{\ell}}^n(u_{\tilde{\ell}})_i\end{pmatrix}, \tag{186}$$

where $\tilde{\ell}$ is the largest index ℓ' such that $|\lambda_{\ell'}|^n > \overline{\delta}\,|\lambda_2|^n$, and $\overline{\delta}$ is a parameter. Each component of $\Psi(i;n)$ is called a "diffusion coordinate". Equations (185) and (186) imply that, in $\mathbb{R}^{\tilde{\ell}-1}$, the Euclidean distance between two data points i and j is equal to the diffusion distance $\overline{d}_{ij}(n)$ with a tolerance of $\overline{\delta}$.

The properties of diffusion maps depend on the parameters n and $\overline{\delta}$. A large value of $\overline{\delta}$ yields a small value of $\tilde{\ell}$ and hence results in a large dimension reduction. A diffusion map with a larger value of n extracts geometry on a more global scale than one with a smaller value of n, so a collection of diffusion maps for different values of n allows one to describe a data set with multliscale geometric properties.

5.6. Respondent-driven sampling

One often is interested in estimating a population mean of certain quantities, such as the fraction of infected individuals, the fraction of people who have a particular opinion, or demographics such as age. If a population is large, which is typical in the context of social surveys, it is impossible to record all individuals. In such situations, a common challenge is how to sample individuals in as unbiased manner as possible.

"Respondent-driven sampling" (RDS) is a popular sampling method that uses edge-tracing in a social network [377, 378]. In RDS, one starts from a seed individual (i.e., a seed node). The seed individual recruits his/her neighbors to a survey by passing a coupon to each of them. The successfully recruited individuals then participate in the survey and in turn pass coupons to their neighbors who have not yet participated. To try to promote participation, individuals who participate are rewarded financially. One takes a weighted mean of the samples to derive an estimate of the quantity of interest (e.g., mean age of a population).

It is necessary to take a weighted mean because the probability of being recruited depends on the position of a person in a network. The so-called "RDS II estimator" is an efficient and realistic estimator [379]. Consider the case in which each respondent passes a single coupon to one of its uniformly randomly selected neighbors. One can then describe the recruitment process as a DTRW if one allows sampling with replacement for simplicity (i.e., if the same individual can be sampled more than once). Again for simplicity, let's also assume that the network is undirected and unweighted. The essential idea of the RDS II estimator is that one should discount the effect of a sampled node v_i by a factor of its degree k_i, because v_i is visited with probability $p_i^* \propto k_i$. Note that respondents have to report k_i to be able to calculate this estimator, although empirically it is difficult to accurately collect the k_i values of respondents [380, 381].

We are interested in estimating the mean $\langle y \rangle$ of a quantity y_i assigned to node v_i. We denote the set of sampled nodes by S and the number of samples (i.e, the size of S) by N_S. The estimator $\langle \hat{y} \rangle$ of $\langle y \rangle$ is

$$\langle \hat{y} \rangle = \frac{1}{N_S} \sum_{v_i \in S} \frac{y_i}{N \hat{p}_i^*}\,, \tag{187}$$

where $\hat{p}_i^*$ is the estimate of the stationary density p_i^*. We set the discount factor on the right-hand side of Eq. (187) to be $N\hat{p}_i^*$, because it is normalized so that $\langle N\hat{p}_i^* \rangle = 1$. By assuming that we do not have access to the mean degree $\langle k \rangle$ of the entire network, we estimate it by calculating

$$\hat{p}_i^* = \frac{k_i}{N\langle \hat{k} \rangle}\,, \tag{188}$$

where $\langle \hat{k} \rangle$ is an estimate of $\langle k \rangle$. We use

$$\langle \hat{k} \rangle = \frac{\sum_{v_i \in S} \frac{k_i}{N p_i^*}}{\sum_{v_i \in S} \frac{1}{N p_i^*}} = \frac{N_S}{\sum_{v_i \in S} (k_i)^{-1}}\,. \tag{189}$$

Combining Eqs. (187), (188), and (189) yields

$$\langle \hat{y} \rangle = \frac{\sum_{v_i \in S} (k_i)^{-1} y_i}{\sum_{v_i \in S} (k_i)^{-1}}\,. \tag{190}$$

The estimated quantity y can be either continuous-valued or discrete-valued. Alternatively, one can estimate the proportion of nodes P_A that have a discrete type A (e.g., an infected state) by setting y_i to the indicator function (i.e., $y_i = 1$ when v_i is of type A and $y_i = 0$ otherwise). In this case, we obtain

$$\hat{P}_A = \frac{\sum_{v_i \in A \cap S} (k_i)^{-1}}{\sum_{v_i \in S} (k_i)^{-1}}\,. \tag{191}$$

Note that, even if one controls for the effect of p_i^* in this manner, the estimator $\langle y \rangle$ is statistically biased in practice. For example, the estimator is inaccurate when networks have community structure [382] or have multiple connected components [383]. Additionally, different techniques are required for directed networks, because Eq. (188) (or, more succinctly, $p_i^* \propto k_i$) does not hold for directed networks [384, 385]. Furthermore, actual sampling trajectories are non-backtracking, and one can incorporate this feature into RDS estimators [386].

A strategy other than RDS II or other estimators of unbiased sampling of nodes is to use a "Metropolis–Hasting RW" [387]. In such sampling, one modifies the edge weight of the original network to guarantee that the stationary density is the uniform density. This method has been used for sampling in peer-to-peer (P2P) and online social networks [42, 388, 389].

5.7. Consensus probability and time of voter models

Voter models are a prototypical family of models of opinion formation that are often defined in terms of a Markov process on a network [1, 31, 33, 49, 390–392]. In traditional voter models, each node assumes one of two opinions, which we call opinion 0 and and opinion 1, and the nodes' opinions evolve stochastically in time. If two adjacent nodes have the opposite opinion, a local consensus of

opinion 0 or opinion 1 between the two nodes occurs at some rate. We suppose that the local consensus dynamics on each edge obeys an independent Poisson process, so the nodes update their opinions asynchronously. For example, if a local consensus on the edge (v_i, v_j) in an undirected network occurs according to a Poisson process at rate $\propto A_{ij}$, we say that voter dynamics obeys "edge dynamics" (ED) (see Fig. 11) [393, 394]. (Note that people often use the term "link dynamics" (LD), because it is common in physics to use the term "link" for "edge".) On finite networks, the final state of a network is the perfect "consensus" of either opinion 0 or opinion 1 for every node. These two consensus configurations are the only absorbing states of the stochastic process. Note that consensus is sometimes also called "fixation" or "coordination".

The best-studied phenomena in voter models include the probability for a network to achieve consensus of a particular opinion and the mean time to achieve consensus. The consensus probability is the probability that a consensus of one opinion (e.g., opinion 0) is reached. With the complementary probability, a finite network achieves a consensus of the other opinion (e.g., opinion 1). When computing mean consensus time, one conditions on the consensus being reached. Both consensus probability and mean consensus time depend both on the initial configuration of opinions and on network structure.

The duality relationship between voter models and "coalescing RWs" (which are non-conservative) makes analysis of RWs a powerful approach for calculating consensus probability and mean consensus time [1, 390, 391, 395]. By definition, a coalescing RW [396] starts by placing a random walker on each node in a network, and the walkers perform independent Poissonian edge-centric CTRWs. If different walkers meet at a node, they coalesce into one and continue as a single random walker. On a finite network, all walkers eventually coalesce into a single random walker.

When examining the dual process, we invert the time and direction of edges [1, 390, 391, 395]. When proceeding backwards in time, two individuals sometimes "collide" in the dual process. Such a coalescence event corresponds to two individuals sharing a common ancestor in the original opinion-formation process. After two individuals coalesce in the dual process, they behave as a single individual.

The duality relationship guarantees that the consensus probability F_i for opinion 0 when node v_i initially has opinion 0 and the other $N-1$ nodes initially have opinion 1 is given by the stationary density of the coalescing RW on the network that one obtains by reversing all edges in an original network. Because all walkers eventually coalesce into a single walker, F_i is given by the stationary density of the usual RW on the edge-reversed network. If initially there are multiple nodes with opinion 0, then the consensus probability for opinion 0 is equal to the sum of F_i over the nodes with initial opinion 0. The mean consensus time is equal to the mean time needed for all walkers to coalesce into one walker. This equality is useful for evaluating the mean consensus time for some networks, because the latter quantity is roughly approximated by the mean time for the first meeting of two independent walkers whose initial location is selected uniformly at random [397–399]. Similar to the MFPT, the mean time

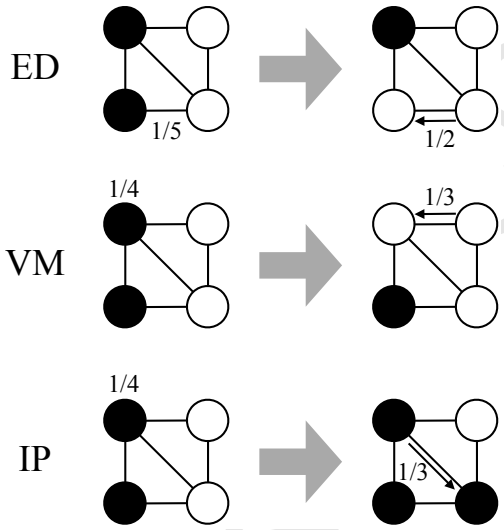

Figure 11: Three updating rules for variants of the classical voter model on a network. For illustration, assume that we have an undirected and unweighted network. With edge-dynamics (ED), one first selects one of the $M = 5$ edges with equal probability (i.e., with probability $1/5$ each). One then selects one of the two directions of the edge with equal probability $1/2$, and then one performs an opinion-updating step. In the most traditional voter model (VM), which has node dynamics, one selects one of the $N = 4$ nodes with equal probability $1/4$. One then determines uniformly at random the neighbor from which the selected node imports its opinion. In the invasion process (IP), one first selects one of the $N = 4$ nodes with equal probability $1/4$ (as in the VM). One then determines uniformly at random the neighbor to which the selected node exports its opinion.

for two random walkers to meet is relatively easy to calculate.

Consider a directed network. As a convention, we assume that the directed edge from v_i to v_j indicates that v_i can coax v_j into v_i's opinion. Even if the network is undirected, one has to distinguish three rules of opinion updating unless the network is regular [393, 394] (see Fig. 11). We evaluate the consensus probability for these three types of voter dynamics using the duality relationship [132, 395].

First, let's consider a variant of the voter model that focuses on the dynamics of edges [393, 394]. Under these "edge dynamics" (ED), one selects a directed edge $v_i \to v_j$ (i.e., from node v_i to node v_j) with probability $A_{ij}/\sum_{i',j'=1}^{N} A_{i'j'}$ in each step, and then node v_j copies v_i's opinion with probability 1. One then advances time by $1/N$, so each node is updated once per unit time on average. The dynamics are equivalent to opinion dynamics in which each edge has a Poisson process with rate $NA_{ij}/\sum_{i',j'=1}^{N} A_{i'j'}$, and an event induces a local consensus event. The dual process for ED is a coalescing RW on the edge-reversed network in continuous time. (In fact, it is a Poissonian edge-centric CTRW.) By modifying Eq. (78), a single random walker satisfies the following

master equation:

$$\frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = \boldsymbol{p}(t)(-D^{\mathrm{rev}} + A^{\top}) = -\boldsymbol{p}(t)L^{\mathrm{rev}}\,, \tag{192}$$

where $A^{\top}$ is the adjacency matrix of the edge-reversed network, D^{rev} is the diagonal matrix whose (i,i)th element is s_i^{in}, and L^{rev} is the combinatorial Laplacian of the edge-reversed network. The consensus probability F_i^{ED} for each node is given by the equilibrium of Eq. (192). That is,

$$(F_1^{\mathrm{ED}},\ \dots\ , F_N^{\mathrm{ED}})L^{\mathrm{rev}} = 0\,. \tag{193}$$

We can obtain an intuitive understanding of Eq. (193) by writing a recursive equation for the consensus probability when the process starts from a single node v_i with opinion 0 (i.e., for F_i^{ED}). We obtain

$$F_i^{\mathrm{ED}} = \sum_{j=1}^{N} \frac{A_{ij}}{\sum_{i',j'=1}^{N} A_{i'j'}} F_{\{i,j\}}^{\mathrm{ED}} + \frac{\sum_{j=1}^{N} A_{ji}}{\sum_{i',j'=1}^{N} A_{i'j'}} \times 0 + \frac{\sum_{i',j'=1;i'\neq i,j'\neq i}^{N} A_{i'j'}}{\sum_{i',j'=1}^{N} A_{i'j'}} F_i^{\mathrm{ED}}\,, \tag{194}$$

where $F_{\{i,j\}}^{\mathrm{ED}}$ is the probability that one reaches the consensus of opinion 0 starting from the configuration in which v_i and v_j but no other nodes have opinion 0. To prove that $F_{\{i,j\}}^{\mathrm{ED}} = F_i^{\mathrm{ED}} + F_j^{\mathrm{ED}}$, imagine that there are N different opinions rather than two, and suppose that node v_i (with $i \in \{1,\ \dots\ , N\}$) holds opinion i. One can express the probability that opinion i or j eventually occupies the entire network either as $F_{\{i,j\}}^{\mathrm{ED}}$ or as $F_i^{\mathrm{ED}} + F_j^{\mathrm{ED}}$, so it follows that $F_{\{i,j\}}^{\mathrm{ED}} = F_i^{\mathrm{ED}} + F_j^{\mathrm{ED}}$. By substituting the latter relationship into Eq. (194), we obtain

$$\sum_{j=1}^{N} A_{ij} F_j^{\mathrm{ED}} = F_i^{\mathrm{ED}} \sum_{j=1}^{N} A_{ji}\,, \tag{195}$$

and we note that Eq. (195) is equivalent to Eq. (193).

The quantity F_i^{ED} is the stationary density of the Poissonian edge-centric CTRW on the edge-reversed network. If the network is undirected, we obtain $L^{\mathrm{rev}} = L$ and $p_i^* = F_i^{\mathrm{ED}} = 1/N$ (with $i \in \{1, \dots, N\}$). Therefore, the likelihood of propagating an opinion does not depend on which node is the seed of the opinion. If the network is directed, we obtain a first-order approximation to the consensus probability of a node by applying Eq. (81) for the edge-reversed network [132]:

$$F_i^{\mathrm{ED}} \approx (\mathrm{const}) \times \frac{s_i^{\mathrm{out}}}{s_i^{\mathrm{in}}}\,. \tag{196}$$

Equation (196) is intuitive, because an out-edge indicates that v_i can enforce its opinion on another node, and an in-edge indicates that v_i listens to neighboring nodes.

In the traditional node-based "voter model" (VM) updating rule, one selects a node v_i uniformly at random (i.e., with equal probability $1/N$) in each time

step. One then selects an in-neighbor v_j of v_i with a probability that is proportional to the weight of the in-edge from that node (i.e., $= A_{ji}/s_i^{\text{in}}$), and v_i copies the opinion of v_j with probability 1. One then advances time by $1/N$ so that on average one node experiences one opinion update per unit time. One can map the dynamics of the VM updating rule to ED dynamics with a modified weighted adjacency matrix $A(D^{\text{rev}})^{-1}$, whose (i,j)th element is equal to A_{ij}/s_j^{in}. The master equation for a single random walker on the edge-reversed network is thus

$$\frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = \boldsymbol{p}(t)(-I + (D^{\text{rev}})^{-1}A^{\top})\,. \tag{197}$$

The equilibrium of the dynamics given by Eq. (197) gives the consensus probability F_i^{VM} for opinion 0 when only node v_i initially has opinion 0. By setting the left-hand side of Eq. (197) to 0, we obtain

$$(F_1^{\text{VM}}, \ \dots \ , F_N^{\text{VM}}) = (F_1^{\text{VM}}, \ \dots \ , F_N^{\text{VM}})(D^{\text{rev}})^{-1}A^{\top}\,, \tag{198}$$

which is equal to the stationary density of a DTRW on the edge-reversed network. Because Eqs. (192) and (197), respectively, represent a Poissonian edge-centric CTRW and a DTRW on the same network, we obtain

$$F_i^{\text{VM}} = s_i^{\text{in}} F_i^{\text{ED}} \tag{199}$$

for arbitrary networks (Section 3.3.1). When a network is undirected, the edge-reversed network is the same as the original network, and we thereby see that

$$F_i^{\text{VM}} = \frac{s_i}{\sum_{s_\ell=1}^{N} s_\ell}\,. \tag{200}$$

When a network is directed, the first-order approximation is given by

$$F_i^{\text{VM}} \propto s_i^{\text{out}}\,. \tag{201}$$

In the so-called "invasion process" (IP) updating rule, one first selects a node v_i uniformly at random (i.e., with probability $1/N$) at each time step to propagate its opinion to one of its out-neighbors. One then selects an out-neighbor v_j of v_i with probability A_{ij}/s_i^{out} (i.e., uniformly at random), and then node v_j copies the opinion of v_i with probability 1. One then advances time by $1/N$. One can map IP dynamics to ED dynamics with the modified weighted adjacency matrix $D^{-1}A$, whose (i,j)th element is equal to A_{ij}/s_i^{out}. The master equation for a single walker in the edge-reversed network is

$$\frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = \boldsymbol{p}(t)(-D^{\text{IP}} + A^{\top}D^{-1})\,, \tag{202}$$

where D^{IP} is the diagonal matrix whose (i,i)th element is given by $\sum_{j=1}^{N}\left(A_{ji}/s_j^{\text{out}}\right)$. The consensus probability F_i^{IP} satisfies

$$(F_1^{\text{IP}}, \ \dots \ , F_N^{\text{IP}}) = (F_1^{\text{IP}}, \ \dots \ , F_N^{\text{IP}})A^{\top}D^{-1}(D^{\text{IP}})^{-1}\,. \tag{203}$$

For an undirected network, $p_i^* \propto 1/s_i$ solves Eq. (203), so nodes with small strengths are good at disseminating their opinions. For a directed network, the first-order approximation to Eq. (203) is

$$\begin{aligned} F_i^{\mathrm{IP}} &= \sum_{j=1}^{N} \frac{F_j^{\mathrm{IP}} A_{ij}/s_i^{\mathrm{out}}}{\sum_{\ell=1}^{N} A_{\ell i}/s_\ell^{\mathrm{out}}} \\ &\approx \sum_{j=1}^{N} \frac{(\mathrm{const}) \times A_{ij}/s_i^{\mathrm{out}}}{\sum_{\ell=1}^{N} A_{\ell i}/(\mathrm{const})} \\ &\propto \frac{1}{s_i^{\mathrm{in}}} \,. \end{aligned} \tag{204}$$

5.8. DeGroot model

The "DeGroot model" is a linear deterministic model that describes opinion-formation dynamics towards consensus [400–402]. Control theorists have studied it as an example of a decentralized consensus algorithm (or protocol) [403]. Although the DeGroot model is not usually discussed as an application of RWs, there are relationships between the extent of a node's influence on the final collective opinion in the DeGroot model and the stationary density of RWs. Before proceeding with our discussion, note that a recent generalization of the DeGroot model combines the averaging rule of the former with an appraisal mechanism (See Ref. [404] and references therein.) to describe the dynamics of individuals' self-appraisal and social power in a network [405]. For nonlinear opinion-formation dynamics that allow non-consensus steady states, see Refs. [392, 406–409].

In the DeGroot model, the opinion of node v_i at discrete time n is given by a continuous variable $x_i(n)$. One assumes that node v_j weighs the opinion $x_i(n)$ of node v_i with weight A_{ij} to determine its opinion in the next time step (i.e., $x_j(n+1)$). The normalization is $\sum_{i=1}^{N} A_{ij} = 1$, and the dynamics are given by

$$x_i(n) = \sum_{j=1}^{N} A_{ji} x_j(n-1) \qquad (i \in \{1, \ldots, N\}) \,. \tag{205}$$

In the DeGroot model, the column sum of A is equal to 1 for every column, and recall that the row sum of T is equal to 1 for every row in a DTRW. To see the correspondence between the two models, it is convenient to write Eq. (205) in vector form as follows:

$$\boldsymbol{x}(n) = A^{\top} \boldsymbol{x}(n-1) \,, \tag{206}$$

where $\boldsymbol{x}(n) = (x_1(n), \ \ldots \ , x_N(n))^{\top}$. Because the row sum of $A^{\top}$ equals 1, we can identify $A^{\top}$ with T. The DeGroot model and DTRWs are thus driven by the same matrix, so their dynamics are essentially the same. The only difference is that the state vector is multiplied on the left in the RW, but it is multiplied on the right in the DeGroot model. Up to rescaling, the models are characterized by the same eigenvalues and eigenvectors.

As long as the spectral gap of T (i.e., $A^\top$) is positive, the stationary density of a DTRW is given uniquely by the left eigenvector of T whose corresponding eigenvalue is 1. Under the same condition, the asymptotic state of the DeGroot model is given by the corresponding right eigenvector of $A^\top$. This eigenvector is $\boldsymbol{x}^* = (x_1^*, \dots, x_N^*)^\top \propto (1, \dots, 1)^\top$, and it corresponds to a state with full consensus.

The initial opinion $x_i(0)$ of node v_i affects the value of the final opinion $x_1^* = \cdots = x_N^*$ in consensus. If $x_1^* = \cdots = x_N^*$ is close to $x_i(0)$ (for a general set of initial conditions that we will specify below) one interprets node v_i as being influential. To quantify this idea, we postulate that $\sum_{i=1}^N F_i^{\mathrm{DG,disc}} x_i(n)$ is conserved over time for positive constants $F_i^{\mathrm{DG,disc}}$ (with $i \in \{1, \dots, N\}$), where the superscript "disc" stands for discrete time and $\sum_{i=1}^N F_i^{\mathrm{DG,disc}} = 1$ gives the normalization. If such a conserved quantity exists, one obtains

$$\sum_{i=1}^N F_i^{\mathrm{DG,disc}} x_i(0) = \sum_{i=1}^N F_i^{\mathrm{DG,disc}} x_i^* = x_1 = \cdots = x_N^* \,. \tag{207}$$

Equation (207) implies that $F_i^{\mathrm{DG,disc}}$ quantifies the influence of v_i on the final opinion in consensus. By imposing this conservation law, one obtains

$$\begin{aligned} \sum_{i=1}^N F_i^{\mathrm{DG,disc}} x_i(n-1) &= \sum_{i=1}^N F_i^{\mathrm{DG,disc}} x_i(n) \\ &= \sum_{i=1}^N F_j^{\mathrm{DG,disc}} \left(\sum_{j=1}^N A_{ji} x_j(n-1) \right) . \end{aligned} \tag{208}$$

By requiring that Eq. (208) holds for arbitrary $x_i(n-1)$ (with $i \in \{1, \dots, N\}$), we obtain

$$F_i^{\mathrm{DG,disc}} = \sum_{j=1}^N A_{ij} F_j^{\mathrm{DG,disc}} \,. \tag{209}$$

Equation (209) indicates that $F_i^{\mathrm{DG,disc}}$ is the stationary density of the DTRW whose transition-probability matrix is $A^\top$.

A continuous-time variant of the DeGroot model has similar relationships [222]. Consider the continuous-time DeGroot model [403]

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = \sum_{j=1}^N A_{ji} \left[x_j(t) - x_i(t) \right] , \tag{210}$$

and note that we do not impose $\sum_{j=1}^N A_{ji} = 1$. The asymptotic state of Eq. (210) is given by $x_1^* = \cdots = x_N^*$. Similar to the discrete-time DeGroot model above, we rewrite Eq. (210) as

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \left(A^\top - D^{\mathrm{rev}} \right) \boldsymbol{x}(t) \equiv -L^{\mathrm{rev}} \boldsymbol{x}(t) \,. \tag{211}$$

Recall that D^{rev} is the diagonal matrix whose $(i,\ i)$th element equals s_i^{in}, and L^{rev} is the combinatorial Laplacian matrix for the edge-reversed network. The left eigenvector of L^{rev} corresponding to eigenvalue 0 gives the stationary density of the Poissonian edge-centric CTRW on the edge-reversed network. The corresponding right eigenvector gives the asymptotic state of the continuous-time DeGroot model. Moreover, this eigenvector is the consensus state $\boldsymbol{x}^* \propto (1,\ \dots\ ,1)^\top$. Equation (211) also has a fascinating interpretation as linear synchronization dynamics that results from linearizing nonlinear systems such as coupled Kuramoto oscillators [54, 410]. See, for example, the discussion in [376].

Equation (211) yields

$$\boldsymbol{p}^* \frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = (\boldsymbol{p}^* L^{\mathrm{rev}})\, \boldsymbol{x}(t) = 0\,, \tag{212}$$

where $\boldsymbol{p}^* = (p_1^*,\ \dots\ ,p_N^*)$, and p_i^* is the stationary density of the Poissonian edge-centric CTRW at node v_i in the edge-reversed network. Therefore, $\boldsymbol{p}^*\boldsymbol{x}(t)$ is conserved, implying that $\sum_{i=1}^N p_i^* x_i(0) = \sum_{i=1}^N p_i^* x_i^* = x_1^* = \cdots = x_N^*$. We thereby see that p_i^* quantifies the influence of node v_i on the final opinion, similar to the case of the discrete-time DeGroot model.

6. Conclusions and outlook

Random walks play a central role in network science. As we have seen in this review, RWs are at the core of numerous methods to extract information from networked systems, and they serve as a leading-order model for (conservative) diffusion processes on networks. Because conventional RWs are linear processes, they are amenable to analysis. For example, one can exploit methods from linear algebra to characterize dynamics in terms of modes relaxing on different time scales, and one can even derive analytical solutions (e.g., via recursive equations) for quantities such as mean first-passage time (MFPT). The simplicity of RWs is crucial, because associated dynamical properties on networks can be analyzed exactly, allowing one to uncover mechanisms by which network structure affects dynamical processes, which is perhaps the primary goal of studying dynamical processes on networks [49]. Many nonlinear processes (e.g., reaction–diffusion systems) include terms related to linear diffusion, so studying RWs on networks also yields important insights into the linear stability (and weakly nonlinear regimes) of numerous nonlinear processes.

RWs have been studied thoroughly (especially on networks) for many decades, but there remains much exciting work to be done. In the following paragraphs, we discuss a few important directions in the study of RWs on networks. As with the rest of our paper, these suggestions are far from exhaustive, and we look forward to seeing new theory and applications of RWs. As we have discussed at length, RWs have connections both to many other processes and to a diverse variety of applications, and we look forward especially to new, unexpected connections that will come to light in the coming years.

One prominent research direction is "non-backtracking RWs", which have opened new perspectives in recent years in topics such as community detection [262–264], because of the convenient properties of their spectrum for sparse networks. Non-backtracking spreading processes have also been used in the examination of network centralities [261], percolation theory [259, 260], and the design of efficient immunization algorithms [265]. Non-backtracking RWs are a type of second-order Markov chain (see Section 4.2.2), and their further study may provide algorithms for clustering and other applications that are more efficient and/or realistic than current ones. As we have illustrated in this review, one can define different types of RWs on the same network, and different RWs lead to different processes, algorithms, and insights.

Intrinsically, community detection and other forms of clustering are a type of model reduction, as one seeks to represent a given network (or dynamical process on a network) using a smaller amount of information. InfoMap (see Section 5.3.3) is a community-detection algorithm that is constructed explicitly on this principle. Related techniques include coarse-graining RWs in a way that preserves the spectral properties of relevant matrices [411, 412], external equitable partitions [413], and using computational group theory to find "hidden" symmetries in networks [414]. More generally, RWs are at the heart of flow-based algorithms, and they have been exploited to examine node centralities (see Section 5.2), community structure (see Section 5.3), core–periphery structure (see Section 5.4), and the mapping of networks into a Euclidean feature space [415]. It may also be fruitful to exploit similar ideas to examine other types of network properties (e.g., "role similarity" [349, 416], "rich clubs" [417, 418], and approximately multipartite structure [419]). RWs have also been used for some studies of community structure in temporal and multilayer networks [68, 234–236] as well as for examining diffusion processes and centralities in such networks [62, 237, 238, 240, 281, 283, 284], and much more remains to be discovered in such applications. In temporal networks, for example, it is important to consider the relative timescales of the network dynamics and the RW dynamics. Novel types of RWs also play an important role in examining higher-order network structure. Examples include the spacey RW [269, 420], RWs on hypergraphs [421], and RWs on simplicial complexes [422].

One can also combine RWs with other dynamical processes to model real-world phenomena in fascinating and insightful ways. For example, one can couple RWs to other processes in multilayer networks [62, 423], where it is important to study scenarios such as infection spreading coupled to human/animal mobility (and more generally to study diffusion dynamics coupled to other types of dynamics). One very successful family of models that combines multiple types of dynamics is metapopulation models of biological contagions, in which individuals move from one subpopulation to another in some way (e.g., according to an RW) and infection events occur within each subpopulation [216, 217]. Metapopulation models, reaction–diffusion models [95, 216], and many other dynamical processes on networks often feature diffusion in the form of a simple, memoryless Poisson process. The use of more complicated and realistic RW processes such as higher-order Markov chains (see Section 4.2.2) and CTRWs

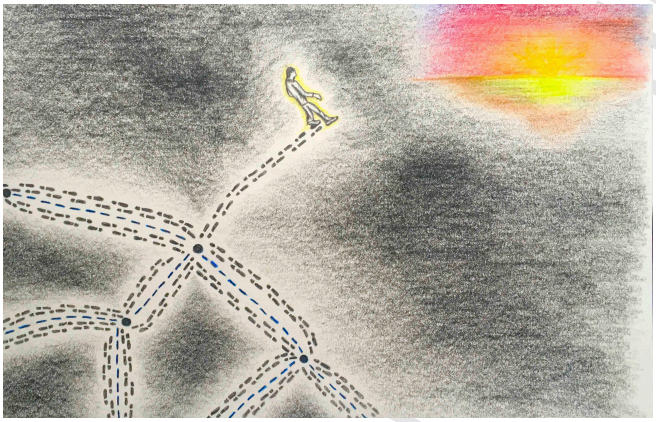

Figure 12: The weary random walker retires from the network and heads off into the distant sunset. [This picture was drawn by Yulian Ng.]

driven by non-Poissonian renewal processes (see Sections 2.2 and 3.3) may yield interesting results.

Various types of RWs continue to be employed actively for a diverse array of applications. We mentioned several examples in Section 1, and we now indicate a few more applications of different types of RWs. For example, a "hungry RW" (taking some inspiration from the arcade game Pac-Man) has yielded insights into anomalous diffusion in bacteria [86], a "waddling RW" allows one to devise an efficient sampler for estimating the frequency of small subgraphs in a network [424], Lévy flights can help capture features of animal foraging [9, 11], multiplicative RWs are a useful approach for examining the dynamics of financial markets [20, 21], self-avoiding RWs have helped improve understanding of polymer chains [18, 19], the stochastic dynamics of neuronal firing have been studied using Ornstein–Uhlenbeck processes (a type of CTRW with a leak term) [13, 14], and the dynamics of correlated novelties (and Kauffman's so-called "adjacent possible") have been modeled using an RW on a growing network (representing the growing space of possible innovations) [425].

In the coming years, we expect that RWs will continue to play a crucial role in physics, computer science, biology, sociology, and numerous other fields. The study of RWs continues to yield fascinating, important, and inspiring insights. Given how much random walkers have contributed to our scientific knowledge, they must be exhausted by now (see Fig. 12).

Acknowledgments

We thank Mariano Beguerisse-Díaz, Michael T. Schaub, Taro Takaguchi, Johan Ugander, and Fabian Ying for valuable feedback on the manuscript. N.M. acknowledges the support provided through JST, CREST; and the JST, ERATO, Kawarabayashi Large Graph Project. R.L. acknowledges the support pro-

vided through Actions de Recherche Concertée (ARC) Mining and Optimization of Big Data Models and the Belgian Network Dynamical Systems, Control, and Optimization (funded by the Interuniversity Attraction Poles Programme). We thank Yulian Ng for drawing Fig. 12.

References

[1] D. Aldous, J. A. Fill, Reversible markov chains and random walks on graphs, unfinished monograph, recompiled 2014, available at `http://www.stat.berkeley.edu/~aldous/RWG/book.html` (2002).

[2] W. Feller, An Introduction to Probability Theory and its Applications, Volume I, 3rd Edition, John Wiley & Sons, Hoboken, NJ, USA, 1968.

[3] W. Feller, An Introduction to Probability Theory and its Applications, Volume II, 2nd Edition, John Wiley & Sons, Hoboken, NJ, USA, 1971.

[4] B. D. Hughes, Random Walks and Randon Environments, Volume 1: Random Walks, Oxford University Press, Oxford, UK, 1995.

[5] R. Kutner, J. Masoliver, The continuous time random walk, still trendy: Fifty-year history, state of art, and outlook, Eur. Phys. J. B 90 (2017) 50.

[6] K. Pearson, The problem of the random walk, Nature 72 (1905) 294.

[7] O. Ore, Pascal and the invention of probability theory, Am. Math. Monthly 67 (1960) 409–419.

[8] A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Ann. Phys. 322 (1905) 549–560.

[9] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. G. E. da Luz, E. P. Raposo, H. E. Stanley, Optimizing the success of random searches, Nature 401 (1999) 911–914.

[10] E. A. Codling, M. J. Plank, S. Benhamou, Random walk models in biology, J. R. Soc. Interface 5 (2008) 813–834.

[11] N. E. Humphries et al., Environmental context explains Lévy and Brownian movement patterns of marine predators, Nature 465 (2010) 1066–1069.

[12] A. Okubo, S. A. Levin, Diffusion and Ecological Problems: Modern Perspectives, 2nd Edition, Springer, New York, NY, USA, 2001.

[13] H. C. Tuckwell, Introduction to Theoretical Neurobiology: Volume 2, Nonlinear and Stochastic Theories, Cambridge University Press, Cambridge, UK, 1988.

[14] F. Gabbiani, S. J. Cox, Mathematics for Neuroscientists, Academic Press, Amsterdam, The Netherlands, 2010.

[15] M. Usher, J. L. McClelland, The time course of perceptual choice: The leaky, competing accumulator model, Psychol. Rev. 108 (2001) 550–592.

[16] J. I. Gold, M. N. Shadlen, The neural basis of decision making, Annu. Rev. Neurosci. 30 (2007) 535–574.

[17] W. J. Ewens, Mathematical Population Genetics I. Theoretical Introduction, Springer, New York, NY, USA, 2010.

[18] M. E. Fisher, Shape of a self-avoiding walk or polymer chain, J. Chem. Phys. 44 (1966) 616–622.

[19] M. B. Isichenko, Percolation, statistical topography, and transport in random media, Rev. Mod. Phys. 64 (1992) 961–1043.

[20] J. Y. Campbell, A. W. Lo, A. C. MacKinlay, The Econometrics of Financial Markets, Princeton University Press, Princeton, NJ, USA, 1996.

[21] R. N. Mantegna, H. E. Stanley, An Introduction to Econophysics, Cambridge University Press, Cambridge, UK, 1999.

[22] T. Jia, D. Wang, B. K. Szymanski, Quantifying patterns of research-interest evolution, Nat. Human Behav. 1 (2017) 0078.

[23] D. F. Gleich, PageRank beyond the Web, SIAM Rev. 57 (2015) 321–363.

[24] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, S. W. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, Proc. Natl. Acad. Sci. USA 102 (2005) 7426–7431.

[25] R. R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (2006) 5–30.

[26] A. Clauset, M. Kogan, S. Redner, Safe leads and lead changes in competitive team sports, Phys. Rev. E 91 (2015) 062815.

[27] C. Godrèche, S. N. Majumdar, G. Schehr, Record statistics of a strongly correlated time series: Random walks and Lévy flights, arXiv:1702.00586 (to appear in J. Phys. A).

[28] F. Iannelli, A. Koher, D. Brockmann, P. Hövel, I. M. Sokolov, Effective distances for epidemics spreading on complex networks, Phys. Rev. E 95 (2017) 012313.

[29] F. Spitzer, Principles of Random Walk, 2nd Edition, Springer, New York, NY, USA, 1976.

[30] G. H. Weiss, Aspects and Applications of the Random Walk, North-Holland, Amsterdam, The Netherlands, 1994.

[31] S. Redner, A Guide to First-Passage Processes, Cambridge University Press, Cambridge, UK, 2001.

[32] R. Burioni, D. Cassi, Random walks on graphs: Ideas, techniques and results, J. Phys. A 38 (2005) R45–R78.

[33] P. L. Krapivsky, S. Redner, E. Ben-Naim, A Kinetic View of Statistical Physics, Cambridge University Press, Cambridge, UK, 2010.

[34] J. Klafter, I. M. Sokolov, First Steps in Random Walks, Oxford University Press, Oxford, UK, 2011.

[35] O. C. Ibe, Elements of Random Walk and Diffusion Processes, Wiley, Hoboken, NJ, USA, 2013.

[36] R. Rammal, Random-walk statistics on fractal structures, J. Stat. Phys. 36 (1984) 547–560.

[37] S. Havlin, D. ben-Avraham, Diffusion in disordered media, Adv. Phys. 36 (1987) 695–798.

[38] J. P. Bouchaud, A. Georges, Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications, Phys. Rep. 195 (1990) 127–293.

[39] D. ben-Avraham, S. Havlin, Diffusion and Reactions in Fractals and Disordered Systems, Cambridge University Press, Cambridge, UK, 2000.

[40] O. Bénichou, R. Voituriez, From first-passage times of random walks in confinement to geometry-controlled kinetics, Phys. Rep. 539 (2014) 225–284.

[41] P. G. Doyle, J. L. Snell, Random Walks and Electric Networks, Mathematical Association of America, Washington, DC, USA, 1984.

[42] L. Lovász, Random walks on graphs: A survey, in: D. Miklós, V. T. Sós, T. Szőnyi (Eds.), Combinatorics, Paul Erdős is Eighty, Volume 1, János Bolyai Math. Soc., Budapest, Hungary, 1993, pp. 353–398.

[43] M. A. Porter, G. Bianconi, Editorial: Network analysis and modelling: Special issue of *European Journal of Applied Mathematics*, Eur. J. Appl. Math. 27 (2016) 807–811.

[44] M. E. J. Newman, Networks: An Introduction, Oxford University Press, Oxford, UK, 2010.

[45] A. L. Barabási, Network Science, Cambridge University Press, Cambridge, UK, 2016.

[46] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. 424 (2006) 175–308.

[47] S. H. Strogatz, Exploring complex networks, Nature 410 (2001) 268–276.

[48] A. Barrat, M. Barthélemy, A. Vespignani, Dynamical Processes on Complex Networks, Cambridge University Press, Cambridge, UK, 2008.

[49] M. A. Porter, J. P. Gleeson, Dynamical Systems on Networks: A Tutorial, Springer, Heidelberg, Germany, 2016.

[50] S. N. Dorogovtsev, J. F. F. Mendes, Evolution of Networks: From Biological Nets to the Internet and WWW, Oxford University Press, Oxford, UK, 2003.

[51] R. Cohen, S. Havlin, Complex Networks: Structure, Robustness and Function, Cambridge University Press, Cambridge, UK, 2010.

[52] E. Estrada, The Structure of Complex Networks: Theory and Applications, Oxford University Press, Oxford, UK, 2012.

[53] M. E. J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256.

[54] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, C. Zhou, Synchronization in complex networks, Phys. Rep. 469 (2008) 93–153.

[55] R. Albert, A. L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. 74 (2002) 47–97.

[56] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, Critical phenomena in complex networks, Rev. Mod. Phys. 80 (2008) 1275–1335.

[57] P. Holme, J. Saramäki, Temporal networks, Phys. Rep. 519 (2012) 97–125.

[58] P. Holme, Modern temporal network theory: A colloquium, Eur. Phys. J. B 88 (2015) 234.

[59] N. Masuda, R. Lambiotte, A Guide to Temporal Networks, World Scientific, London, UK, 2016.

[60] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, M. Zanin, The structure and dynamics of multilayer networks, Phys. Rep. 544 (2014) 1–122.

[61] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, M. A. Porter, Multilayer Networks, J. Comp. Netw. 2 (2014) 203–271.

[62] M. De Domenico, C. Granell, M. A. Porter, A. Arenas, The physics of spreading processes in multilayer networks, Nat. Phys. 12 (2016) 901–906.

[63] M. A. Porter, J. P. Onnela, P. J. Mucha, Communities in networks, Notices of the AMS 56 (2009) 1082–1097, 1164–1166.

[64] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (2010) 75–174.

[65] S. Fortunato, D. Hric, Community detection in networks: A user guide, Phys. Rep. 659 (2016) 1–44.

[66] S. Brin, L. Page, Anatomy of a large-scale hypertextual web search engine, Proceedings of the Seventh International World Wide Web Conference (1998) 107–117.

[67] P. Bonacich, Factoring and weighting approaches to status scores and clique identification, J. Math. Sociol. 2 (1972) 113–120.

[68] L. G. S. Jeub, P. Balachandran, M. A. Porter, P. J. Mucha, M. W. Mahoney, Think locally, act locally: Detection of small, medium-sized, and large communities in large networks, Phys. Rev. E 91 (2015) 012821.

[69] J. Gillis, Correlated random walk, Math. Proc. Camb. Philos. Soc. 51 (1955) 639–651.

[70] C. Domb, From random to self-avoiding walks, J. Stat. Phys. 30 (1983) 425–436.

[71] N. Madras, G. Slade, The Self-Avoiding Walk, Birkhäuser, Boston, MA, USA, 1993.

[72] M. R. Evans, T. Hanney, Nonequilibrium statistical mechanics of the zero-range process and related models, J. Phys. A 38 (2005) R195–R240.

[73] A. Schenzle, H. Brand, Multiplicative stochastic processes in statistical physics, Phys. Rev. A 20 (1979) 1628–1647.

[74] S. Havlin, R. B. Selinger, M. Schwartz, H. E. Stanley, A. Bunde, Random multiplicative processes and transport in structures with correlated spatial disorder, Phys. Rev. Lett. 61 (1988) 1438–1441.

[75] R. Pemantle, A survey of random processes with reinforcement, Prob. Surveys 4 (2007) 1–79.

[76] R. B. Schinazi, Classical and Spatial Stochastic Processes, Birkhäuser, Boston, MA, USA, 1999.

[77] G. M. Schütz, S. Trimper, Elephants can always remember: Exact long-range memory effects in a non-Markovian random walk, Phys. Rev. E 70 (2004) 045101(R).

[78] J. Kempe, Quantum random walks: An introductory overview, Contemp. Phys. 44 (2003) 307–327.

[79] O. Mülken, A. Blumen, Continuous-time quantum walks: Models for coherent transport on complex networks, Phys. Rep. 502 (2011) 37–87.

[80] O. Bénichou, C. Loverdo, M. Moreau, R. Voituriez, Intermittent search strategies, Rev. Mod. Phys. 83 (2011) 81–130.

[81] V. Tejedor, R. Voituriez, O. Bénichou, Optimizing persistent random searches, Phys. Rev. Lett. 108 (2012) 088103.

[82] O. Bénichou, S. Redner, Depletion-controlled starvation of a diffusing forager, Phys. Rev. Lett. 113 (2014) 238101.

[83] O. Bénichou, M. Chupeau, S. Redner, Role of depletion on the dynamics of a diffusing forager, J. Phys. A 49 (2016) 394003.

[84] U. Bhat, S. Redner, O. Bénichou, Starvation dynamics of a greedy forager, arXiv:1704.05861.

[85] D. S. Grebenkov, J. F. Rupprecht, The escape problem for mortal walkers, J. Chem. Phys. 146 (2017) 084106.

[86] T. Schilling, T. Voigtmann, Clearing out a maze: The hungry random walker and its anomalous diffusion, arXiv:1607.01123.

[87] R. Ghosh, K. Lerman, Rethinking centrality: The role of dynamical processes in social network analysis, Disc. Cont. Dyn. Syst. Ser. B 19 (2014) 1355–1372.

[88] X. Yan, S. h. Teng, K. Lerman, R. Ghosh, Capturing the interplay of dynamics and networks through parameterizations of Laplacian operators, PeerJ Comput. Sci. 2 (2016) e57.

[89] R. M. Anderson, R. M. May, Infectious Diseases of Humans, Oxford University Press, Oxford, UK, 1991.

[90] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. 87 (2015) 925–979.

[91] Y. Wang, D. Chakrabarti, C. Wang, C. Faloutsos, Epidemic spreading in real networks: an eigenvalue viewpoint, Proc. 22nd International Symposium on Reliable Distributed Systems (SRDS'03) (2003) 25–34.

[92] K. Klemm, M. A. Serrano, V. M. Eguíluz, M. San Miguel, A measure of individual role in collective dynamics, Sci. Rep. 2 (2012) 292.

[93] C. Godsil, G. Royle, Algebraic Graph Theory, Springer, New York, NY, USA, 2001.

[94] T. W. Valente, Network Models of the Diffusion of Innovations, Hampton Press, Cresskill, NJ, USA, 1995.

[95] M. Asllani, J. D. Challenger, F. S. Pavone, L. Sacconi, D. Fanelli, The theory of pattern formation on directed networks, Nat. Comm. 5 (2014) 4517.

[96] E. W. Montroll, G. H. Weiss, Random walks on lattices. II, J. Math. Phys. 6 (1965) 167–181.

[97] D. R. Cox, Renewal Theory, Methuen & Co. Ltd, Frome, UK, 1962.

[98] P. Grigolini, L. Palatella, G. Raffaelli, Asymmetric anomalous diffusion: An efficient way to detect memory in time series, Fractals 9 (2001) 439–449.

[99] S. Hoory, N. Linial, A. Wigderson, Expander graphs and their applications, Bull. Amer. Math. Soc. 43 (2006) 439–561.

[100] A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (2009) 661–703.

[101] Z. Eisler, J. Kertész, Random walks on complex networks with inhomogeneous impact, Phys. Rev. E 71 (2005) 057104.

[102] W. X. Wang, B. H. Wang, C. Y. Yin, Y. B. Xie, T. Zhou, Traffic dynamics based on local routing protocol on a scale-free network, Phys. Rev. E 73 (2006) 026111.

[103] A. Fronczak, P. Fronczak, Biased random walks in complex networks: The role of local navigation rules, Phys. Rev. E 80 (2009) 016107.

[104] S. Lee, S. H. Yook, Y. Kim, Centrality measure of complex networks using biased random walks, Eur. Phys. J. B 68 (2009) 277–281.

[105] A. Baronchelli, R. Pastor-Satorras, Mean-field diffusive dynamics on weighted networks, Phys. Rev. E 82 (2010) 011111.

[106] M. Bonaventura, V. Nicosia, V. Latora, Characteristic times of biased random walks on complex networks, Phys. Rev. E 89 (2014) 012803.

[107] L. Demetrius, T. Manke, Robustness and network evolution—an entropic principle, Physica A 346 (2005) 682–696.

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

[109] Z. Burda, J. Duda, J. M. Luck, B. Waclaw, Localization of the maximal entropy random walk, Phys. Rev. Lett. 102 (2009) 160602.

[110] J. C. Delvenne, A. S. Libert, Centrality measures and thermodynamic formalism for complex networks, Phys. Rev. E 83 (2011) 046117.

[111] R. Sinatra, J. Gómez-Gardeñes, R. Lambiotte, V. Nicosia, V. Latora, Maximal-entropy random walks in complex networks with limited information, Phys. Rev. E 83 (2011) 030103(R).

[112] J. G. Kemeny, J. L. Snell, Finite Markov Chains (Second Printing), Springer-Verlag, New York, NY, USA, 1976.

[113] A. Papoulis, S. U. Pillai, Probability, Random Variables, and Stochastic Processes, 4th Edition, McGraw-Hill, New York, NY, USA, 2002.

[114] M. Iosifescu, Finite Markov Processes and Their Applications, Dover Publications, Mineola, NY, USA, 1980.

[115] W. J. Stewart, Introduction to the Numerical Solution of Markov Chains, Princeton University Press, Princeton, NJ, USA, 1994.

[116] J. R. Norris, Markov Chains, Cambridge University Press, Cambridge, UK, 1997.

[117] H. M. Taylor, S. Karlin, An Introduction to Stochastic Modeling, 3rd Edition, Academic Press, San Diego, CA, USA, 1998.

[118] D. A. Levin, Y. Peres, E. L. Wilmer, Markov Chains and Mixing Times, American Mathematical Society, Providence, RI, USA, 2009.

[119] P. Blanchard, D. Volchenkov, Random Walks and Diffusions on Graphs and Databases: An Introduction, Springer, Heidelberg, Germany, 2011.

[120] N. Privault, Understanding Markov Chains: Examples and Applications, Springer, Singapore, 2013.

[121] J. P. Sethna, Statistical Mechanics: Entropy, Order Parameters and Complexity, Oxford University Press, Oxford, UK, 2006.

[122] B. K. Fosdick, D. B. Larremore, J. Nishimura, J. Ugander, Configuring random graph models with fixed degree sequences, arXiv:1608.00607.

[123] V. Colizza, A. Vespignani, Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations, J. Theor. Biol. 251 (2008) 450–467.

[124] Y. Lin, Z. Zhang, Random walks in weighted networks with a perfect trap: An application of Laplacian spectra, Phys. Rev. E 87 (2013) 062140.

[125] Z. Zhang, T. Shan, G. Chen, Random walks on weighted networks, Phys. Rev. E 87 (2013) 012112.

[126] D. Donato, L. Laura, S. Leonardi, S. Millozzi, Large scale properties of the Webgraph, Eur. Phys. J. B 38 (2004) 239–243.

[127] S. Fortunato, A. Flammini, F. Menczer, A. Vespignani, Topical interests and the mitigation of search engine bias, Proc. Natl. Acad. Sci. USA 103 (2006) 12684–12689.

[128] J. G. Restrepo, E. Ott, B. R. Hunt, Characterizing the dynamical importance of network nodes and links, Phys. Rev. Lett. 97 (2006) 094102.

[129] P. M. Davis, Eigenfactor: Does the principle of repeated improvement result in better estimates than raw citation counts?, J. Amer. Soc. Info. Sci. Tech. 59 (2008) 2186–2188.

[130] S. Fortunato, M. Boguñá, A. Flammini, F. Menczer, Approximating PageRank from in-degree, LNCS 4936 (2008) 59–71.

[131] A. Fersht, The most influential journals: Impact factor and eigenfactor, Proc. Natl. Acad. Sci. USA 106 (2009) 6883–6884.

[132] N. Masuda, H. Ohtsuki, Evolutionary dynamics and fixation probabilities in directed networks, New J. Phys. 11 (2009) 033012.

[133] G. Ghoshal, A. L. Barabási, Ranking stability and super-stable nodes in complex networks, Nat. Commun. 2 (2011) 394.

[134] A. Sandryhaila, J. M. F. Moura, Discrete signal processing on graphs, IEEE Trans. Signal Proc. 61 (2013) 1644–1656.

[135] N. Tremblay, P. Borgnat, Graph wavelets for multiscale community mining, IEEE Trans. Signal Proc. 62 (2014) 5227–5239.

[136] A. N. Samukhin, S. N. Dorogovtsev, J. F. F. Mendes, Laplacian spectra of, and random walks on, complex networks: Are scale-free architectures really important?, Phys. Rev. E 77 (2008) 036115.

[137] F. R. K. Chung, Spectral Graph Theory, American Mathematical Society, Providence, RI, USA, 1997.

[138] L. Donetti, F. Neri, M. A. Muñoz, Optimal network topologies: expanders, cages, Ramanujan graphs, entangled networks and all that, J. Stat. Mech. 2006 (2006) P08007.

[139] D. Cvetković, P. Rowlinson, S. Simić, An Introduction to the Theory of Graph Spectra, Cambridge University Press, Cambridge, UK, 2010.

[140] P. Van Mieghem, Graph Spectra for Complex Networks, Cambridge University Press, 2011.

[141] G. H. Golub, C. F. Van Loan, Matrix Computations, 3rd Edition, The Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[142] E. Estrada, N. Hatano, M. Benzi, The physics of communicability in complex networks, Physics Reports 514 (2012) 89–119.

[143] U. Miekkala, Graph properties for splitting with grounded Laplacian matrices, BIT 33 (1993) 485–495.

[144] J. D. Noh, H. Rieger, Random walks on complex networks, Phys. Rev. Lett. 92 (2004) 118701.

[145] H. S. Wilf, generatingfunctionology, 3rd Edition, A K Peters, Ltd., Wellesley, MA, USA, 2005.

[146] A. Kittas, S. Carmi, S. Havlin, P. Argyrakis, Trapping in complex networks, EPL 84 (2008) 40008.

[147] N. Perra, A. Baronchelli, D. Mocanu, B. Gonçalves, R. Pastor-Satorras, A. Vespignani, Random walks and search in time-varying networks, Phys. Rev. Lett. 109 (2012) 238701.

[148] M. Starnini, A. Baronchelli, A. Barrat, R. Pastor-Satorras, Random walks on temporal networks, Phys. Rev. E 85 (2012) 056115.

[149] E. Almaas, R. V. Kulkarni, D. Stroud, Scaling properties of random walks on small-world networks, Phys. Rev. E 68 (2003) 056105.

[150] N. Masuda, N. Konno, Return times of random walk on generalized random graphs, Phys. Rev. E 69 (2004) 066113.

[151] S. Hwang, D. S. Lee, B. Kahng, Effective trapping of random walkers in complex networks, Phys. Rev. E 85 (2012) 046110.

[152] S. Hwang, D. S. Lee, B. Kahng, Origin of the hub spectral dimension in scale-free networks, Phys. Rev. E 87 (2013) 022816.

[153] J. Peng, E. Agliari, Z. Zhang, Exact calculations of first-passage properties on the pseudofractal scale-free web, Chaos 25 (2015) 073118.

[154] E. W. Montroll, Random walks on lattices. III. Calculation of first-passage times with application to exciton trapping on photosynthetic units, J. Math. Phys. 10 (1969) 753–765.

[155] E. M. Bollt, D. ben-Avraham, What is special about diffusion on scale-free nets?, New J. Phys. 7 (2005) 26.

[156] V. Sood, S. Redner, D. ben-Avraham, First-passage properties of the Erdős-Renyi random graph, J. Phys. A 38 (2005) 109–123.

[157] M. Löwe, F. Torres, On hitting times for a simple random walk on dense Erdös–Rényi random graphs, Stat. Prob. Lett. 89 (2014) 81–88.

[158] I. Tishby, O. Biham, E. Katzav, The distribution of first hitting times of random walks on directed Erdős-Rényi networks, arXiv:1703.10269.

[159] S. Jespersen, I. M. Sokolov, A. Blumen, Relaxation properties of small-world networks, Phys. Rev. E 62 (2000) 4405–4408.

[160] S. A. Pandit, R. E. Amritkar, Random spread on the family of small-world networks, Phys. Rev. E 63 (2001) 041104.

[161] J. Lahtinen, J. Kertész, K. Kaski, Scaling of random spreading in small world networks, Phys. Rev. E 64 (2001) 057105.

[162] F. Jasch, A. Blumen, Trapping of random walks on small-world networks, Phys. Rev. E 64 (2001) 066104.

[163] P. E. Parris, V. M. Kenkre, Traversal times for random walks on small-world networks, Phys. Rev. E 72 (2005) 056119.

[164] S. J. Yang, Exploring complex networks by walking on them, Phys. Rev. E 71 (2005) 016107.

[165] M. A. Porter, Small-world network, Scholarpedia 7 (2) (2012) 1739.

[166] S. Hwang, D. S. Lee, B. Kahng, First passage time for random walks in heterogeneous networks, Phys. Rev. Lett. 109 (2012) 088701.

[167] S. Condamin, O. Bénichou, V. Tejedor, R. Voituriez, J. Klafter, First-passage times in complex scale-invariant media, Nature 450 (2007) 77–80.

[168] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, Pseudofractal scale-free web, Phys. Rev. E 65 (2002) 066122.

[169] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A. L. Barabási, Hierarchical organization of modularity in metabolic networks, Science 297 (2002) 1551–1555.

[170] E. Ravasz, A. L. Barabási, Hierarchical organization in complex networks, Phys. Rev. E 67 (2003) 026112.

[171] H. D. Rozenfeld, S. Havlin, D. ben Avraham, Fractal and transfractal recursive scale-free nets, New J. Phys. 9 (2007) 175.

[172] C. Song, S. Havlin, H. A. Makse, Self-similarity of complex networks, Nature 433 (2005) 392–395.

[173] C. M. Song, S. Havlin, H. A. Makse, Origins of fractality in the growth of complex networks, Nat. Phys. 2 (2006) 275–281.

[174] H. A. Simon, The architecture of complexity, Proc. Amer. Philos. Soc. 106 (1962) 467–482.

[175] J. Leskovec, D. Huttenlocher, J. Kleinberg, Signed networks in social media, in: Proc. 28th International Conference on Human Factors in Computing Systems - CHI '10, 2010, pp. 1361–1370.

[176] B. Karrer, M. E. J. Newman, Random graph models for directed acyclic networks, Phys. Rev. E 80 (2009) 046110.

[177] K. Falconer, Fractals: A Very Short Introduction, Oxford University Press, Oxford, UK, 2013.

[178] J. J. Kozak, V. Balakrishnan, Analytic expression for the mean time to absorption for a random walker on the Sierpinski gasket, Phys. Rev. E 65 (2002) 021105.

[179] S. Havlin, H. Weissman, Mapping between hopping on hierarchical structures and diffusion on a family of fractals, J. Phys. A 19 (1986) L1021–L1026.

[180] B. Kahng, S. Redner, Scaling of the first-passage time and the survival probability on exact and quasi-exact self-similar structures, J. Phys. A 22 (1989) 887–902.

[181] E. Agliari, Exact mean first-passage time on the T-graph, Phys. Rev. E 77 (2008) 011128.

[182] T. Vicsek, Fractal models for diffusion controlled aggregation, J. Phys. A 16 (1983) L647–L652.

[183] A. Blumen, A. Jurjiu, T. Koslowski, C. von Ferber, Dynamics of Vicsek fractals, models for hyperbranched polymers, Phys. Rev. E 67 (2003) 061103.

[184] Z. Zhang, B. Wu, H. Zhang, S. Zhou, J. Guan, Z. Wang, Determining global mean-first-passage time of random walks on Vicsek fractals using eigenvalues of Laplacian matrices, Phys. Rev. E 81 (2010) 031118.

[185] O. Matan, S. Havlin, Mean first-passage time on loopless aggregates, Phys. Rev. A 40 (1989) 6573–6579.

[186] A. N. Berker, S. Ostlund, Renormalization-group calculations of finite systems: Order parameter and specific heat for epitaxial ordering, J. Phys. C 12 (1979) 4961–4975.

[187] V. Tejedor, O. Bénichou, R. Voituriez, Global mean first-passage times of random walks on complex networks, Phys. Rev. E 80 (2009) 065104(R).

[188] Z. Zhang, Y. Lin, Y. Ma, Effect of trap position on the efficiency of trapping in treelike scale-free networks, J. Phys. A 44 (2011) 075102.

[189] C. K. Yun, B. Kahng, D. Kim, Annihilation of two-species reaction–diffusion processes on fractal scale-free networks, New J. Phys. 11 (2009) 063025.

[190] A. L. Barabási, E. Ravasz, T. Vicsek, Deterministic scale-free networks, Physica A 299 (2001) 559–564.

[191] S. Jung, S. Kim, B. Kahng, Geometric fractal growth model for scale-free networks, Phys. Rev. E 65 (2002) 056101.

[192] J. S. Andrade Jr., H. J. Herrmann, R. F. S. Andrade, L. R. da Silva, Apollonian networks: Simultaneously scale-free, small world, Euclidean, space filling, and with matching graphs, Phys. Rev. Lett. 94 (2005) 018702.

[193] J. P. K. Doye, C. P. Massen, Self-similar disk packings as model spatial scale-free networks, Phys. Rev. E 71 (2005) 016128.

[194] Z. Zhang, Y. Qi, S. Zhou, W. Xie, J. Guan, Exact solution for mean first-passage time on a pseudofractal scale-free web, Phys. Rev. E 79 (2009) 021127.

[195] Z. Zhang, Y. Qi, S. Zhou, S. Gao, J. Guan, Explicit determination of mean first-passage time for random walks on deterministic uniform recursive trees, Phys. Rev. E 81 (2010) 016114.

[196] Z. Zhang, X. Guo, Y. Lin, Full eigenvalues of the Markov matrix for scale-free polymer networks, Phys. Rev. E 90 (2014) 022816.

[197] E. Agliari, R. Burioni, Random walks on deterministic scale-free networks: Exact results, Phys. Rev. E 80 (2009) 031125.

[198] Z. Zhang, Y. Lin, S. Gao, S. Zhou, J. Guan, M. Li, Trapping in scale-free networks with hierarchical organization of modularity, Phys. Rev. E 80 (2009) 051120.

[199] Z. Zhang, J. Guan, W. Xie, Y. Qi, S. Zhou, Random walks on the Apollonian network with a single trap, EPL 86 (2009) 10006.

[200] S. Reuveni, R. Granek, J. Klafter, Vibrational shortcut to the mean-first-passage-time problem, Phys. Rev. E 81 (2010) 040103(R).

[201] L. K. Gallos, C. Song, S. Havlin, H. A. Makse, Scaling theory of transport in complex biological networks, Proc. Natl. Acad. Sci. USA 104 (2007) 7746–7751.

[202] A. Baronchelli, V. Loreto, Ring structures and mean first passage time in networks, Phys. Rev. E 73 (2006) 026103.

[203] H. W. Lau, K. Y. Szeto, Asymptotic analysis of first passage time in complex networks, EPL 90 (2010) 40005.

[204] C. Cooper, A. Frieze, The cover time of sparse random graphs, Random Struct. Alg. 30 (2007) 1–16.

[205] C. Cooper, A. Frieze, The cover time of the preferential attachment graph, J. Comb. Th. Ser. B 97 (2007) 269–290.

[206] M. Chupeau, O. Bénichou, R. Voituriez, Cover times of random searches, Nat. Phys. 11 (2015) 844–847.

[207] B. F. Maier, D. Brockmann, Cover time for random walks on arbitrary complex networks, arXiv:1706.02356.

[208] G. Barnes, U. Feige, Short random walks on graphs, in: Proc. Twenty-fifth Annual ACM Symposium on Theory of Computing (STOC '93), 1993, pp. 728–737.

[209] L. K. Gallos, Random walk and trapping processes on scale-free networks, Phys. Rev. E 70 (2004) 046116.

[210] D. Stauffer, M. Sahimi, Diffusion in scale-free networks with annealed disorder, Phys. Rev. E 72 (2005) 046128.

[211] L. da Fontoura Costa, G. Travieso, Exploring complex networks through random walks, Phys. Rev. E 75 (2007) 016102.

[212] A. Baronchelli, M. Catanzaro, R. Pastor-Satorras, Random walks on complex trees, Phys. Rev. E 78 (2008) 011114.

[213] A. Asztalos, Z. Toroczkai, Network discovery by generalized random walks, EPL 92 (2010) 50008.

[214] T. Hoffmann, M. A. Porter, R. Lambiotte, Generalized master equations for non-Poisson dynamics on networks., Phys. Rev. E 86 (2012) 046102.

[215] L. Speidel, R. Lambiotte, K. Aihara, N. Masuda, Steady state and mean recurrence time for random walks on stochastic temporal networks., Phys. Rev. E 91 (2015) 012806.

[216] V. Colizza, R. Pastor-Satorras, A. Vespignani, Reaction-diffusion processes and metapopulation models in heterogeneous networks, Nat. Phys. 3 (2007) 276–282.

[217] A. Vespignani, Modelling dynamical processes in complex socio-technical systems, Nat. Phys. 8 (2011) 32–39.

[218] P. S. Scardal, D. Taylor, J. Sun, A. Arenas, Collective frequency variation in network synchronization and reverse PageRank, Phys. Rev. E 93 (2016) 042314.

[219] N. Biggs, Algebraic potential theory on graphs, Bull. London Math. Soc. 29 (1997) 641–682.

[220] R. P. Agaev, P. Y. Chebotarev, The matrix of maximum out forests of a digraph and its ppplications, Autom. Rem. Cont. 61 (2000) 1424–1450.

[221] N. Masuda, Y. Kawamura, H. Kori, Analysis of relative influence of nodes in directed networks, Phys. Rev. E 80 (2009) 046114.

[222] N. Masuda, Y. Kawamura, H. Kori, Impact of hierarchical modular structure on ranking of individual nodes in directed networks, New J. Phys. 11 (2009) 113002.

[223] G. B. Ermentrout, Stable periodic solutions to discrete and continuum arrays of weakly coupled nonlinear oscillators, SIAM J. Appl. Math. 52 (1992) 1665–1687.

[224] R. Merris, Laplacian matrices of graphs: A survey, Lin. Alge. Its Appl. 197-198 (1994) 143–176.

[225] B. Mohar, Some applications of Laplace eigenvalues of graphs, in: G. Hahn, G. Sabidussi (Eds.), Graph Symmetry: Algebraic Methods and Applications, Kluwer, Dordrecht, The Netherlands, 1997, pp. 225–275.

[226] L. M. Pecora, T. L. Carroll, Synchronization of chaotic systems, Chaos 25 (2015) 097611.

[227] A. E. Motter, Bounding network spectra for network design, New J. Phys. 9 (2007) 182.

[228] A. Vázquez, J. G. Oliveira, Z. Dezsö, K. I. Goh, I. Kondor, A. L. Barabási, Modeling bursts and heavy tails in human dynamics, Phys. Rev. E 73 (2006) 036127.

[229] M. Kivelä, M. A. Porter, Estimating interevent time distributions from finite observation periods in communication networks, Phys. Rev. E 92 (2015) 052813.

[230] J. C. Delvenne, R. Lambiotte, L. E. C. Rocha, Diffusion on networked systems is a question of time or structure, Nat. Comm. 6 (2015) 7366.

[231] S. De Nigris, A. Hastir, R. Lambiotte, Burstiness and fractional diffusion on complex networks, Eur. Phys. J. B 89 (2016) 114.

[232] A. O. Allen, Probability, Statistics, and Queueing Theory: With Computer Science Applications, 2nd Edition, Academic Press, Boston, MA, USA, 1990.

[233] D. R. DeFord, S. D. Pauls, A new framework for dynamical models on multiplex networks, arXiv:1507.00695v2.

[234] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, J. P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, Science 328 (2010) 876–878.

[235] M. De Domenico, A. Lancichinetti, A. Arenas, M. Rosvall, Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems, Phys. Rev. X 5 (2015) 011027.

[236] L. G. S. Jeub, M. W. Mahoney, P. J. Mucha, M. A. Porter, A local perspective on community structure in multilayer networks, Netw. Sci. 5 (2017) 144–163.

[237] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, A. Arenas, Diffusion dynamics on multiplex networks, Phys. Rev. Lett. 110 (2013) 028701.

[238] A. Solé-Ribalta, M. De Domenico, N. E. Kouvaris, A. Díaz-Guilera, S. Gómez, A. Arenas, Spectral properties of the Laplacian of multiplex networks, Phys. Rev. E 88 (2013) 032807.

[239] F. Radicchi, A. Arenas, Abrupt transition in the structural formation of interconnected networks, Nat. Phys. 9 (2013) 717–720.

[240] M. De Domenico, A. Solé-Ribalta, S. Gómez, A. Arenas, Navigability of interconnected networks under random failures, Proc. Natl. Acad. Sci. USA 111 (2014) 8351–8356.

[241] L. Lacasa, I. P. Mariño, J. Miguez, V. Nicosia, J. Gómez-Gardeñes, Identifying the hidden multiplex architecture of complex systems, arXiv:1705.04661.

[242] N. Perra, B. Gonçalves, R. Pastor-Satorras, A. Vespignani, Activity driven modeling of time varying networks., Sci. Rep. 2 (2012) 469.

[243] B. Bollobás, Random Graphs, 2nd Edition, Cambridge University Press, Cambridge, UK, 2001.

[244] B. Ribeiro, N. Perra, A. Baronchelli, Quantifying the effect of temporal resolution on time-varying networks., Sci. Rep. 3 (2013) 3006.

[245] L. Alessandretti, K. Sun, A. Baronchelli, N. Perra, Random walks on activity-driven networks with attractiveness, Phys. Rev. E 95 (2017) 052318.

[246] N. Masuda, K. Klemm, V. M. Eguíluz, Temporal networks: Slowing down diffusion by long lasting interactions, Phys. Rev. Lett. 111 (2013) 188701.

[247] R. Lambiotte, V. Salnikov, M. Rosvall, Effect of memory on the dynamics of random walks on networks, J. Comp. Netw. 3 (2015) 177–188.

[248] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone, F. Schweitzer, Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks, Nat. Comm. 5 (2014) 5024.

[249] A. Sousa da Mata, R. Pastor-Satorras, Slow relaxation dynamics and aging in random walks on activity driven temporal networks, Eur. Phys. J. B 88 (2015) 38.

[250] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, R. Lambiotte, Memory in network flows and its effects on spreading dynamics and community detection, Nat. Comm. 5 (2014) 4630.

[251] L. E. C. Rocha, N. Masuda, Random walk centrality for temporal networks, New J. Phys. 16 (2014) 063023.

[252] J. Saramäki, P. Holme, Exploring temporal networks with greedy walks, Eur. Phys. J. B 88 (2015) 334.

[253] F. Harary, R. Z. Norman, Some properties of line digraphs, Rend. Circ. Mat. Palermo 9 (1960) 161–168.

[254] T. S. Evans, R. Lambiotte, Line graphs, link partitions, and overlapping communities, Phys. Rev. E 80 (2009) 016105.

[255] V. Salnikov, M. T. Schaub, R. Lambiotte, Using higher-order Markov models to reveal flow-based communities in networks, Sci. Rep. 6 (2016) 23194.

[256] J. Xu, T. L. Wickramarathne, N. V. Chawla, Representing higher-order dependencies in networks, Sci. Adv. 2 (2016) e1600028.

[257] N. Alon, I. Benjamini, E. Lubetzky, S. Sodin, Non-backtracking random walks mix faster, Comm. Contemp. Math. 9 (2007) 585–603.

[258] R. Fitzner, R. van der Hofstad, Non-backtracking random walk, J. Stat. Phys. 150 (2013) 264–284.

[259] B. Karrer, M. E. J. Newman, L. Zdeborová, Percolation on sparse networks, Phys. Rev. Lett. 113 (2014) 208702.

[260] K. E. Hamilton, L. P. Pryadko, Tight lower bound for percolation threshold on an infinite graph, Phys. Rev. Lett. 113 (2014) 208701.

[261] T. Martin, X. Zhang, M. E. J. Newman, Localization and centrality in networks, Phys. Rev. E 90 (2014) 052808.

[262] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, P. Zhang, Spectral redemption in clustering sparse networks, Proc. Natl. Acad. Sci. USA 110 (2013) 20935–20940.

[263] M. E. J. Newman, Spectral community detection in sparse networks, arXiv:1308.6494v1.

[264] C. Bordenave, M. Lelarge, L. Massoulié, Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs, in: Proc. IEEE 56th Annual Symp. Found. Comput. Sci. (FOCS '15), 2015, pp. 1347–1357.

[265] F. Morone, H. A. Makse, Influence maximization in complex networks through optimal percolation, Nature 524 (2015) 65–68.

[266] T. Takaguchi, M. Nakamura, N. Sato, K. Yano, N. Masuda, Predictability of conversation partners, Phys. Rev. X 1 (2011) 011008.

[267] R. Pfitzner, I. Scholtes, A. Garas, C. J. Tessone, F. Schweitzer, Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks, Phys. Rev. Lett. 110 (2013) 198701.

[268] W. Li, M. K. Ng, On the limiting probability distribution of a transition probability tensor, Lin. Multilin. Algebra 62 (2014) 362–385.

[269] A. R. Benson, D. F. Gleich, L. H. Lim, The spacey random walk: A stochastic process for higher-order data, SIAM Rev. 59 (2017) 321–345.

[270] P. Kareiva, N. Shigesada, Analyzing insect movement as a correlated random walk, Oecologia 56 (1983) 234–238.

[271] F. Chierichetti, R. Kumar, P. Raghavan, T. Sarlós, Are web users really Markovian?, in: Proc. 21st Intl. Conf. World Wide Web (WWW '12), 2012, pp. 609–618.

[272] J. Kleinberg, Complex networks and decentralized search algorithms, in: Proc. Internat. Congress Math., 2006, pp. 1019–1044.

[273] Q. Lv, P. Cao, E. Cohen, K. Li, S. Shenker, Search and replication in unstructured peer-to-peer networks, in: Proc. 16th Internat. Conf. Supercomputing (ICS '02), 2002, pp. 84–95.

[274] N. Bisnik, A. Abouzeid, Modeling and analysis of random walk search algorithms in P2P networks, in: Proc. Second International Workshop on Hot Topics in Peer-to-Peer Systems, 2005, pp. 95–103.

[275] A. S. Tanenbaum, D. J. Wetherall, Computer Networks, 5th Edition, Prentice Hall, Boston, MA, USA, 2011.

[276] L. A. Adamic, R. M. Lukose, A. R. Puniyani, B. A. Huberman, Search in power-law networks, Phys. Rev. E 64 (2001) 046135.

[277] L. Lü, D. Chen, X. L. Ren, Q. M. Zhang, Y. C. Zhang, T. Zhou, Vital nodes identification in complex networks, Phys. Rep. 650 (2016) 1–63.

[278] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1953) 39–43.

[279] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM 46 (1999) 604–632.

[280] P. Grindrod, M. C. Parsons, D. J. Higham, E. Estrada, Communicability across evolving networks, Phys. Rev. E 83 (2011) 046120.

[281] A. Halu, R. J. Mondragón, P. Panzarasa, G. Bianconi, Multiplex PageRank, PLOS ONE 8 (2013) e78293.

[282] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, A. Arenas, Ranking in interconnected multilayer networks reveals versatile nodes, Nat. Comm. 6 (2015) 6868.

[283] A. Solé-Ribalta, M. De Domenico, S. Gómez, A. Arenas, Random walk centrality in interconnected multilayer networks, Physica D 323–324 (2016) 73–79.

[284] D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, P. J. Mucha, Eigenvector-based centrality measures for temporal networks, Multiscale Model. Simul. 15 (2017) 537–574.

[285] H. Liao, M. S. Mariani, M. Medo, Y. C. Zhang, M. Y. Zhou, Ranking in evolving complex networks, arXiv:1704.08027.

[286] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank citation ranking: Bringing order to the web, Technical Report 1999-66, Stanford Infolab (1999).

[287] G. Pinski, F. Narin, Citation influence for journal aggregates of scientific publications: Theory, with application to literature of physics, Info. Proc. Manag. 12 (1976) 297–312.

[288] A. N. Langville, C. D. Meyer, Deeper inside PageRank, Internet Math. 1 (2004) 335–380.

[289] P. Berkhin, A survey on PageRank computing, Internet Math. 2 (2005) 73–120.

[290] A. N. Langville, C. D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Rankings, Princeton University Press, Princeton, NJ, USA, 2006.

[291] S. Kamvar, Numerical Algorithms for Personalized Search in Self-organizing Information Networks, Princeton University Press, Princeton, NJ, USA, 2010.

[292] L. Ermann, K. M. Frahm, D. L. Shepelyansky, Google matrix analysis of directed networks, Rev. Mod. Phys. 88 (2016) 039905.

[293] C. Winter et al., Google goes cancer: Improving outcome prediction for cancer patients by network-based ranking of marker genes, PLOS Comput. Biol. 8 (2012) e1002511.

[294] F. Radicchi, Who is the best player ever? A complex network analysis of the history of professional tennis, PLOS ONE 6 (2011) e17249.

[295] T. Callaghan, P. J. Mucha, M. A. Porter, Random walker ranking for NCAA division I-A football, Amer. Math. Monthly 114 (2007) 761–777.

[296] J. Park, M. E. J. Newman, A network-based ranking system for US college football, J. Stat. Mech. (2005) P10014.

[297] S. Saavedra, S. Powers, T. McCotter, M. A. Porter, P. J. Mucha, Mutually-antagonistic interactions in baseball networks, Physica A 389 (2010) 1131–1141.

[298] J. Lages, A. Patt, D. L. Shepelyansky, Wikipedia ranking of world universities, Eur. Phys. J. B 89 (2016) 69.

[299] S. A. Myers, P. J. Mucha, M. A. Porter, Mathematical genealogy and department prestige, Chaos 21 (2011) 041104.

[300] F. Mitzlaff, G. Stumme, Recommending given names, arXiv:1302.4412.

[301] S. Brin, R. Motwani, L. Page, T. Winograd, What can you do with a Web in your Pocket?, Bull. IEEE Comput. Soc. Tech. Comm. Data Eng. 21 (1998) 37–47.

[302] T. H. Haveliwala, Topic-sensitive PageRank, in: Proc. Eleventh Intl. Conf. World Wide Web (WWW '02), 2002, pp. 517–526.

[303] G. Jeh, J. Widom, Scaling personalized web search, in: Proc. Twelfth Intl. Conf. World Wide Web (WWW '03), 2003, pp. 271–279.

[304] W. Xie, D. Bindel, A. Demers, J. Gehrke, Edge-weighted personalized PageRank: Breaking a decade-old performance barrier, in: Proc. 21st ACM SIGKDD Intl. Conf. Knowledge Disc. Data Mining (KDD '15), 2015, pp. 1325–1334.

[305] R. Lambiotte, M. Rosvall, Ranking and clustering of nodes in networks with smart teleportation, Phys. Rev. E 85 (2012) 056107.

[306] P. Boldi, M. Santini, S. Vigna, PageRank as a function of the damping factor, in: Proc. 14th Internat. Conf. on World Wide Web, 2005, pp. 557–566.

[307] M. Brinkmeier, PageRank revisited, ACM Trans. Internet Tech. 6 (2006) 282–301.

[308] F. Chung, The heat kernel as the pagerank of a graph, Proc. Natl. Acad. Sci. USA 104 (2007) 19735–19740.

[309] F. Chung, A local graph partitioning algorithm using heat kernel pagerank, Internet Math. 6 (2009) 315–330.

[310] D. Zhou, S. A. Orshanskiy, H. Zha, C. L. Giles, Co-ranking authors and documents in a heterogeneous network, in: Seventh IEEE Int. Conf. Data Mining (ICDM 2007), 2007, pp. 739–744.

[311] M. K. Ng, X. Li, Y. Ye, MultiRank: Co-ranking for objects and relations in multi-relational data, in: Proc. 17th ACM SIGKDD Int. Conf. Knowledge Disc. Data Mining (KDD '11), 2011, pp. 1217–1225.

[312] F. Pedroche, M. Romance, R. Criado, A biplex approach to PageRank centrality: From classic to multiplex networks, Chaos 26 (2016) 065301.

[313] H. E. Daniels, Round-robin tournament scores, Biometrika 56 (1969) 295–299.

[314] J. W. Moon, N. J. Pullman, On generalized tournament matrices, SIAM Rev. 12 (1970) 384–399.

[315] K. A. Berman, A graph theoretical approach to handicap ranking of tournaments and paired comparisons, SIAM Alg. Disc. Methods 1 (1980) 359–361.

[316] N. E. Borm, R. V. D. Brink, M. Slikker, An iterative procedure for evaluating digraph competitions, Ann. Operat. Res. 109 (2002) 61–75.

[317] P. D. Taylor, Allele-frequency change in a class-structured population, Am. Nat. 135 (1990) 95–106.

[318] P. D. Taylor, Inclusive fitness arguments in genetic models of behaviour, J. Math. Biol. 34 (1996) 654–674.

[319] P. Grindrod, D. J. Higham, A matrix iteration for dynamic network summaries, SIAM Rev. 55 (2013) 118–128.

[320] L. C. Freeman, A set of measures of centrality based on betweenness, Sociometry 40 (1977) 35–41.

[321] M. E. J. Newman, A measure of betweenness centrality based on random walks, Soc. Netw. 27 (2005) 39–54.

[322] M. Ben-Akiva, S. R. Lerman, Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press, Cambridge, MA, 1985.

[323] K. E. Train, Discrete Choice Methods with Simulation, 2nd Edition, Cambridge University Press, Cambridge, UK, 2009.

[324] C. Dwork, R. Kumar, M. Naor, D. Sivakumar, Rank aggregation methods for the Web, in: Proc. Tenth Intl. Conf. World Wide Web (WWW '01), 2001, pp. 613–622.

[325] R. A. Bradley, M. E. Terry, Rank analysis of incomplete block designs: I. The method of paired comparisons, Biometrika 39 (1952) 324–345.

[326] R. D. Luce, Individual Choice Behavior: A Theoretical Analysis, Wiley, New York, NY, 1959.

[327] S. Ragain, J. Ugander, Pairwise choice Markov chains, in: D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems 29, Curran Associates, Inc., 2016, pp. 3198–3206.

[328] H. D. Block, J. Marschak, Random orderings and stochastic theories of responses, in: I. Olkin, S. G. Ghuyre, W. Hoeffding, W. G. Madow, H. B. Mann (Eds.), Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, Stanford University Press, Redwood City, CA, 1960, pp. 97–132.

[329] R. L. Plackett, The Analysis of permutations, J. R. Stat. Soc. Ser. C 24 (1975) 193–202.

[330] L. Maystre, M. Grossglauser, Fast and accurate inference of Plackett–Luce models, in: C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems 28, Curran Associates, Inc., 2015, pp. 172–180.

[331] S. Negahban, S. Oh, D. Shah, Iterative ranking from pair-wise comparisons, in: F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25, Curran Associates, Inc., 2012, pp. 2474–2482.

[332] S. Negahban, S. Oh, D. Shah, Rank centrality: Ranking from pairwise comparisons, Oper. Res. 65 (2017) 266–287.

[333] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 888–905.

[334] S. M. van Dongen, Graph clustering by flow simulation, PhD thesis, University of Utrecht (2001).

[335] K. A. Eriksen, I. Simonsen, S. Maslov, K. Sneppen, Modularity and extreme edges of the Internet, Phys. Rev. Lett. 90 (2003) 148701.

[336] H. Zhou, Network landscape from a Brownian particle's perspective, Phys. Rev. E 67 (2003) 041908.

[337] A. Arenas, A. Díaz-Guilera, C. J. Pérez-Vicente, Synchronization reveals topological scales in complex networks, Phys. Rev. Lett. 96 (2006) 114102.

[338] M. E. J. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (2006) 036104.

[339] L. Danon, A. Arenas, A. Díaz-Guilera, Impact of community structure on information transfer, Phys. Rev. E 77 (2008) 036103.

[340] X. Q. Cheng, H. W. Shen, Uncovering the community structure associated with the diffusion dynamics on networks, J. Stat. Mech 2010 (2010) P04024.

[341] R. Lambiotte, R. Sinatra, J. C. Delvenne, T. S. Evans, M. Barahona, V. Latora, Flow graphs: Interweaving dynamics and structure, Phys. Rev. E 84 (2011) 017102.

[342] C. Piccardi, Finding and testing network communities by lumped Markov chains, PLOS ONE 6 (2011) e27028.

[343] M. Sarzynska, E. A. Leicht, G. Chowell, M. A. Porter, Null models for community detection in spatially embedded, temporal networks, J. Comp. Netw. 4 (2016) 363–406.

[344] M. Bazzi, M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, S. D. Howison, Community detection in temporal multilayer networks, with an application to correlation networks, Mult. Model. Simul. 14 (2016) 1–41.

[345] S. Fortunato, M. Barthélemy, Resolution limit in community detection, Proc. Natl. Acad. Sci. USA 104 (2007) 36–41.

[346] B. H. Good, Y. A. de Montjoye, A. Clauset, Performance of modularity maximization in practical contexts, Phys. Rev. E 81 (2010) 046106.

[347] M. Lambiotte, J. C. Delvenne, M. Barahona, Laplacian dynamics and multiscale modular structure in networks, arXiv:0812.1770v2.

[348] M. T. Schaub, J. C. Delvenne, S. N. Yaliraki, M. Barahona, Markov dynamics as a zooming lens for multiscale community detection: Non clique-like communities and the field-of-view limit, PLOS ONE 7 (2012) e32210.

[349] M. Beguerisse-Díaz, G. Garduño-Hernández, B. Vangelov, S. N. Yaliraki, M. Barahona, Interest communities and flow roles in directed networks: The Twitter network of the UK riots, J. R. Soc. Interface 11 (2014) 20140940.

[350] R. Lambiotte, J. C. Delvenne, M. Barahona, Random walks, Markov processes and the multiscale modular organization of complex networks, IEEE Trans. Netw. Sci. Eng. 1 (2014) 76–90.

[351] J. C. Delvenne, S. N. Yaliraki, M. Barahona, Stability of graph communities across time scales, Proc. Natl. Acad. Sci. USA 107 (2010) 12755–12760.

[352] J. C. Delvenne, M. T. Schaub, S. N. Yaliraki, M. Barahona, The stability of a graph partition: A dynamics-based framework for community detection, in: A. Mukherjee, M. Choudhury, F. Peruani, N. Ganguly, B. Mitra (Eds.), Dynamics on and of Complex Networks, Volume 2: Applications to Time-Varying Dynamical Systems, Springer, New York, NY, USA, 2013, pp. 221–242.

[353] J. Reichardt, S. Bornholdt, Statistical mechanics of community detection, Phys. Rev. E 74 (2006) 016110.

[354] M. E. J. Newman, Equivalence between modularity optimization and maximum likelihood methods for community detection, Phys. Rev. E 94 (2016) 052315.

[355] D. S. Bassett, M. A. Porter, N. F. Wymbs, S. T. Grafton, J. M. Carlson, P. J. Mucha, Robust detection of dynamic community structure in networks, Chaos 23 (2013) 013142.

[356] P. Pons, M. Latapy, Computing communities in large networks using random walks, J. Graph Algo. Appl. 10 (2006) 191–218.

[357] H. Zhou, R. Lipowsky, Network Brownian motion: A new method to measure vertex-vertex proximity and to identify communities and subcommunities, LNCS 3038 (2004) 1062–1069.

[358] M. E. J. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69 (2004) 066133.

[359] F. Fouss, A. Pirotte, J. M. Renders, M. Saerens, Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation, IEEE Trans. Knowl. Data Eng. 19 (2007) 355–369.

[360] S. Barnett, Matrices: Methods and Applications, Oxford University Press, Oxford, UK, 1990.

[361] M. Rosvall, C. T. Bergstrom, Maps of random walks on complex networks reveal community structure, Proc. Natl. Acad. Sci. USA 105 (2008) 1118–1123.

[362] M. Rosvall, C. T. Bergstrom, Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems, PLOS ONE 6 (2011) e18209.

[363] D. A. Huffman, A method for the construction of minimum-redundancy codes, Proc. IRE 40 (1952) 1098–1101.

[364] M. Rosvall, Mapdemo (demonstration of map equation), `http://www.mapequation.org/apps/MapDemo.html` (3 September 2016) (2016).

[365] D. A. Spielman, S. H. Teng, Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems, in: Proc. Thirty-sixth Annual ACM Symp. Theory of Comput. (STOC '04), 2004, pp. 81–90.

[366] R. Andersen, F. Chung, K. Lang, Local graph partitioning using PageRank vectors, in: Proc. IEEE 47th Annual Symp. Found. Comput. Sci. (FOCS '06), 2006, pp. 475–486.

[367] D. A. Spielman, S. H. Teng, A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning, SIAM J. Comput. 42 (2013) 1–26.

[368] K. Kloster, D. F. Gleich, Heat kernel based community detection, in: Proc. 20th ACM SIGKDD Intl. Conf. Knowledge Disc. Data Mining (KDD '14), 2014, pp. 1386–1395.

[369] I. M. Kloumann, J. M. Kleinberg, Community membership identification from small seed sets, in: Proc. 20th ACM SIGKDD Intl. Conf. Knowledge Disc. Data Mining (KDD '14), 2014, pp. 1366–1375.

[370] I. M. Kloumann, J. Ugander, J. Kleinberg, Block models and personalized PageRank, Proc. Natl. Acad. Sci. USA 114 (2017) 33–38.

[371] P. J. Mucha, M. A. Porter, Communities in multislice voting networks, Chaos 20 (2010) 041108.

[372] P. Csermely, A. London, L. Y. Wu, B. Uzzi, Structure and dynamics of core/periphery networks, J. Compl. Netw. 1 (2013) 93–123.

[373] M. P. Rombach, M. A. Porter, J. H. Fowler, P. J. Mucha, Core-periphery structure in networks, SIAM J. Appl. Math. 74 (2014) 167–190.

[374] F. Della Rossa, F. Dercole, C. Piccardi, Profiling core-periphery network structure by random walkers, Sci. Rep. 3 (2013) 1467.

[375] J. A. Lee, M. Verleysen, Nonlinear Dimensionality Reduction, Springer, New York, NY, 2007.

[376] M. De Domenico, Diffusion geometry unravels the emergence of functional clusters in collective phenomena, Phys. Rev. Lett. 118 (2017) 168301.

[377] D. D. Heckathorn, Respondent-driven sampling: A new approach to the study of hidden populations, Soc. Prob. 44 (1997) 174–199.

[378] M. J. Salganik, D. D. Heckathorn, Sampling and estimation in hidden populations using respondent-driven sampling, Sociol. Methodol. 34 (2004) 193–240.

[379] E. Volz, D. D. Heckathorn, Probability based estimation theory for respondent driven sampling, J. Official Stat. 24 (2008) 79–97.

[380] C. McCarty, P. D. Killworth, H. R. Bernard, E. C. Johnsen, G. A. Shelley, Comparing two methods for estimating network size, Human Organ. 60 (2001) 28–39.

[381] P. V. Marsden, Recent developments in network measurement, in: P. J. Carrington, J. Scott, S. Wasserman (Eds.), Models and Methods in Social Network Analysis, Cambridge University Press, Cambridge, UK, 2005, pp. 8–30.

[382] L. E. C. Rocha, A. E. Thorson, R. Lambiotte, F. Liljeros, Respondent-driven sampling bias induced by community structure and response rates in social networks, J. R. Statist. Soc. A 180 (2017) 99–118.

[383] J. Malmros, L. E. C. Rocha, Multiple seed structure and disconnected networks in respondent-driven sampling, arXiv.orgarXiv:1603.04222.

[384] X. Lu, L. Bengtsson, T. Britton, M. Camitz, B. J. Kim, A. Thorson, F. Liljeros, The sensitivity of respondent-driven sampling, J. R. Statist. Soc. A 175 (2012) 191–216.

[385] J. Malmros, N. Masuda, T. Britton, Random walks on directed networks: Inference and respondent-driven sampling, J. Off. Stat. 32 (2016) 433–459.

[386] K. J. Gile, Improved inference for respondent-driven sampling data with application to HIV prevalence estimation, J. Am. Stat. Assoc. 106 (2011) 135–146.

[387] W. R. Gilks, S. Richardson, D. J. Spiegelhalter (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, London, UK, 1996.

[388] D. Stutzbach, R. Rejaie, N. Duffield, S. Sen, W. Willinger, On unbiased sampling for unstructured peer-to-peer networks, IEEE/ACM Trans. Netw. 17 (2009) 377–390.

[389] M. Gjoka, M. Kurant, C. T. Butts, A. Markopoulou, Walking in facebook: A case study of unbiased sampling of OSNs, in: Proc. 29th Conf. Info. Comm. (INFOCOM '10), 2010, pp. 2498–2506.

[390] T. M. Liggett, Interacting Particle Systems, Springer, New York, NY, USA, 1985.

[391] D. Durrett, Lecture Notes on Particle Systems and Percolation, Belmont, Wadsworth, CA, USA, 1988.

[392] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. 81 (2009) 591–646.

[393] T. Antal, S. Redner, V. Sood, Evolutionary dynamics on degree-heterogeneous graphs, Phys. Rev. Lett. 96 (2006) 188104.

[394] V. Sood, T. Antal, S. Redner, Voter models on heterogeneous networks, Phys. Rev. E 77 (2008) 041121.

[395] P. Donnelly, D. Welsh, Finite particle systems and infection models, Math. Proc. Camb. Phil. Soc. 94 (1983) 167–182.

[396] D. Griffeath, Annihilating and coalescing random walks on $\mathbb{Z}^d$, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 46 (1978) 55–65.

[397] C. Cooper, R. Elsässer, H. Ono, T. Radzik, Coalescing random walks and voting on connected graphs, SIAM J. Disc. Math. 27 (2013) 1748–1758.

[398] N. Masuda, Voter model on the two-clique graph, Phys. Rev. E 90 (2014) 012802.

[399] Y. Iwamasa, N. Masuda, Networks maximizing the consensus time of voter models, Phys. Rev. E 90 (2014) 012816.

[400] R. P. Abelson, Mathematical models of the distribution of attitudes under controversy, in: N. Frederiksen, H. Gulliksen (Eds.), Contributions to Mathematical Psychology, Holt, Rinehart and Winston, New York, NY, USA, 1964, pp. 141–160.

[401] M. H. DeGroot, Reaching a consensus, J. Am. Stat. Assoc. 69 (1974) 118–121.

[402] M. O. Jackson, Social and Economic Networks, Princeton University Press, Princeton, NJ, USA, 2008.

[403] R. Olfati-Saber, J. A. Fax, R. M. Murray, Consensus and cooperation in networked multi-agent systems, Proc. IEEE 95 (2007) 215–233.

[404] N. E. Friedkin, E. C. Johnsen, Social Influence Network Theory: A Sociological Examination of Small Group Dynamics, Cambridge University Press, Cambridge, UK, 2011.

[405] P. Jia, A. MirTabatabaei, N. E. Friedkin, F. Bullo, Opinion dynamics and the evolution of social power in influence networks, SIAM Rev. 57 (2015) 367–397.

[406] K. Sznajd-Weron, J. Sznajd, Opinion evolution in closed community, Int. J. Mod. Phys. C 11 (2000) 1157–1165.

[407] P. L. Krapivsky, S. Redner, Dynamics of majority rule in two-state interacting spin systems, Phys. Rev. Lett. 90 (23) (2003) 238701.

[408] F. Schweitzer, L. Behera, Nonlinear voter models: The transition from invasion to coexistence, Eur. Phys. J. B 67 (2009) 301–318.

[409] J. Shao, S. Havlin, H. E. Stanley, Dynamic opinion model and invasion percolation, Phys. Rev. Lett. 103 (2009) 018701.

[410] F. A. Rodrigues, T. K. D. Peron, P. Ji, J. Kurths, The Kuramoto model in complex networks, Phys. Rep. 610 (2016) 1–98.

[411] D. Gfeller, P. De Los Rios, Spectral coarse graining of complex networks, Phys. Rev. Lett. 99 (2007) 038701.

[412] D. Gfeller, P. De Los Rios, Spectral coarse graining and synchronization in oscillator networks, Phys. Rev. Lett. 100 (2008) 174104.

[413] N. O'Clery, Y. Yuan, G. B. Stan, M. Barahona, Observability and coarse graining of consensus dynamics through the external equitable partition, Phys. Rev. E 88 (2013) 042805.

[414] L. M. Pecora, F. Sorrentino, A. M. Hagerstrom, T. E. Murphy, R. Roy, Cluster synchronization and isolated desynchronization in complex networks with symmetries, Nat. Comm. 5 (2014) 4079.

[415] L. Li, W. M. Campbell, R. S. Caceres, Graph model selection via random walks , arXiv:1704.05516.

[416] K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, C. Faloutsos, L. Li, RolX: Structural role extraction & mining in large graphs, in: Proc. 18th ACM SIGKDD Intl. Conf. Knowledge Disc. Data Mining (KDD '12), 2012, pp. 1231–1239.

[417] S. Zhou, R. J. Mondragón, The rich-club phenomenon in the Internet topology, IEEE Comm. Lett. 8 (2004) 180–182.

[418] V. Colizza, M. A. S. A. Flammini, A. Vespignani, Detecting rich-club ordering in complex networks, Nat. Phys. 2 (2006) 110–115.

[419] M. E. J. Newman, E. A. Leicht, Mixture models and exploratory analysis in networks, Proc. Natl. Acad. Sci. USA 104 (2007) 9564–9569.

[420] T. Wu, A. R. Benson, D. F. Gleich, General tensor spectral co-clustering for higher-order data, in: D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems 29, Curran Associates, Inc., 2016, pp. 2559–2567.

[421] L. Lu, X. Peng, High-ordered random walks and generalized Laplacians on hypergraphs, LNCS 6732 (2011) 14–25.

[422] S. Mukherjee, J. Steenbergen, Random walks on simplicial complexes and harmonics, Rand. Struct. Algor. 49 (2016) 379–405.

[423] V. Nicosia, P. S. Skardal, V. Latora, A. Arenas, Spontaneous synchronization driven by energy transport in interconnected networks, arXiv:1405.5855 (to appear in Phys. Rev. Lett.).

[424] G. Han, H. Sethu, Waddling random walk: Fast and accurate mining of motif statistics in large graphs, in: Sixteenth IEEE Intl. Conf. Data Mining, (ICDM 2016), 2016, pp. 181–190.

[425] F. Tria, V. Loreto, V. D. P. Servedio, S. H. Strogatz, The dynamics of correlated novelties, Sci. Rep. 4 (2014) 5890.