# An adaptation of InfoMap to absorbing random walks using absorption-scaled graphs

Esteban Vargas, Mason A. Porter, Joseph H. Tien

April 25, 2024

## Abstract

InfoMap is a popular approach to detect densely connected "communities" of nodes in networks. To detect such communities, InfoMap uses random walks and ideas from information theory. Motivated by the dynamics of disease spread on networks, whose nodes can have heterogeneous disease-removal rates, we adapt InfoMap to absorbing random walks. To do this, we use *absorption-scaled graphs* (in which edge weights are scaled according to absorption rates) and Markov time sweeping. One of our adaptations of InfoMap converges to the standard version of InfoMap in the limit in which the node-absorption rates approach 0. We demonstrate that the community structure that one obtains using our adaptations of InfoMap can differ markedly from the community structure that one detects using methods that do not account for node-absorption rates. We also illustrate that the community structure that is induced by heterogeneous absorption rates can have important implications for susceptible–infected–recovered (SIR) dynamics on ring-lattice networks. For example, in some situations, the outbreak duration is maximized when a moderate number of nodes have large node-absorption rates.

## 1 Introduction

Random walks are one of the most fundamental dynamical processes, and many studies have used random walks on networks (i.e., graphs) to gain insights into network structure and how such structure affects dynamical processes [25]. Much research has focused on standard random walks, in which the distribution of the occupation probabilities of a network's nodes converges to a stationary distribution with all positive entries in the limit of infinitely many walker steps. It is important to understand the relationship between network structure and different types of random walks. In the present paper, we consider absorbing random walks, in which the probability to reach an "absorbing state" converges to 1 as the number of walker steps becomes infinite. We examine dynamical processes that involve absorbing random walks, for which there is a nonzero rate (the so-called "absorption rate") of transitioning to an absorbing state from each node of a graph.

Absorbing random walks have been used to develop centrality measures [14], other methods to rank the nodes of a network [46], transduction algorithms (which one can use to infer the labels of the nodes of a graph from the labels of a subset of the nodes) [7], and more. For example, Jaydeep et al. [7] proposed a transduction algorithm that uses the number of visits before absorption of an absorbing random walk as a measure of affinity between the nodes of a graph. Absorbing random walks also arise naturally in many modeling contexts, including population dynamics [11], the spread of infectious diseases on networks [21], and the propagation of content in online social networks [3]. In the setting of population dynamics, consider a collection of habitat patches that are connected through some mobility network. In this context, a random walk corresponds to an individual moving between patches and absorption corresponds to death [1, 10, 11]. Similar situations occur in contagion processes, where one can model the transfer of a pathogen between sites as a random walk and pathogen clearance (e.g., recovery from infection) as absorption. This interpretation of pathogen transfer underlies the next-generation-matrix approach [44] to calculate the basic reproduction number $\mathcal{R}_0$, which is a staple quantity of mathematical epidemiology that indicates the mean number of secondary infections that arise when a single infected individual enters a population of susceptible individuals [5].

In the present paper, we examine absorbing random walks on graphs in which different nodes can have different absorption rates, inducing an "effective" network structure that is reflected only partially by the edge weights of a network. Many notions of network community structure arise from the analysis of random walks [12, 25], and we expect different types of random walks to yield different community structures [17, 18]. A "community" in a network is a tightly knit set of nodes that is connected sparsely to other tightly knit sets of nodes [12, 33]. Communities are a common feature of many real-world networks, and community structure influences dynamical processes such as the spread of infectious diseases [42] and online content [15, 45]. For

example, community structure can affect the size and duration of a disease outbreak [38]. There is intense interest in understanding how community structure and node characteristics combine to influence contagions on networks [24, 34, 37].

We develop community-detection algorithms that account for node-absorption rates. We adapt the widely-used community-detection algorithm InfoMap [35, 36, 41] to absorbing random walks and thereby account for heterogeneous node-absorption rates in the detected communities. In our adaptation, we apply InfoMap to *absorption-scaled graphs*, which account for absorption by scaling the edge weights of a network [16]. These absorption-scaled graphs are related to their associated absorbing random walks by a generalized inverse (the so-called "absorption inverse") and a fundamental matrix [16]. We use absorption inverses and results from [16] to study the absorption-scaled graphs that are associated with our adaptations of InfoMap.

Community structure can greatly influence disease dynamics on networks [26, 30]. Salathé and Jones [38] illustrated that changes in community structure are correlated with changes in disease quantities for susceptible–infected–recovered (SIR) dynamics on a network. One of their findings is that outbreak duration can achieve a maximum at intermediate modularity values. Inspired by Salathé and Jones [38], we use our adaptation of InfoMap to study an example contagion process that illustrates how absorption in disease dynamics affects community structure, which in turn affects disease spread. We investigate the association between changes in the effective community structure that is induced by the node-absorption rates with disease quantities such as outbreak size and duration.

Our paper proceeds as follows. In Section 2, we present the original InfoMap algorithm and the Markov time-sweeping technique that we use in our adaptations of InfoMap. In Section 3, we apply InfoMap to absorption-scaled graphs. In Section 4, we introduce a definition of a map function $L^{(a)}$ for absorbing random walks. We relate this new map function to our adaptations of InfoMap. In Section 5, we discuss toy examples to illustrate that our adaptations of InfoMap yield effective community structures, which arise from the heterogeneous node-absorption rates. In Section 6, we examine effective community structure and SIR disease dynamics on a synthetic network. In Section 7, we summarize and discuss our key conclusions. In Appendix A, we present the proofs of three key propositions from Section 4. Our code is available at https://gitlab.com/esteban_vargas_bernal/extending-infomap-to-absorbing-random-walks.

## 2   Background on InfoMap

We now present background material on InfoMap and an adaptation of it to continuous-time Markov chains using Markov time sweeping [39]. We summarize our notation in Table 1.

| Symbol | Expression | Description |
|---|---|---|
| $G$ | | Directed, weighted graph |
| $A$ | | Adjacency matrix of a directed, weighted graph |
| $W$ | | Diagonal matrix of out-degrees |
| $\mathcal{L}$ | $W - A$ | Unnormalized graph Laplacian matrix |
| $P$ | $AW^{-1}$ | Transition-probability matrix of the Markov chain that is associated with $A$ |
| $M$ | $\{M_1, \ldots, M_m\}$ | Partition of a graph into communities |
| $M_i$ | $\{k_1, \ldots, k_{n_i}\}$ | Community that consists of nodes $k_1, \ldots, k_{n_i}$ |
| $\pi$ | $(\pi_1, \ldots, \pi_n)^{\mathrm{T}}$ | Probability distribution on the set of nodes for the map function |
| $q_{i\curvearrowright}$ | $\sum_{j \in M_i, k \notin M_i} \pi_j p_{kj}$ | Probability of a transition out of community $M_i$ |
| $q_{i\curvearrowleft}$ | $\sum_{k \in M_i, j \notin M_i} \pi_j p_{kj}$ | Probability of a transition into community $M_i$ |
| $q_{\curvearrowleft}$ | $\sum_{i \in \{1, \ldots, m\}} q_{i\curvearrowleft}$ | Probability of a transition into a different community |
| $p_{\circlearrowleft}^i$ | $q_{i\curvearrowright} + \sum_{j \in M_i} \pi_j$ | Normalization factor for the probability distribution $\mathcal{P}^i$ |
| $\mathcal{Q}$ | $\left(q_{1\curvearrowleft}/q_{\curvearrowleft}, \ldots, q_{m\curvearrowleft}/q_{\curvearrowleft}\right)^{\mathrm{T}}$ | Probability distribution for the map function that is associated with transitions into a different community |
| $\mathcal{P}^i$ | $\left(q_{i\curvearrowright}/p_{\circlearrowleft}^i, \pi_{k_1}/p_{\circlearrowleft}^i, \ldots, \pi_{k_{n_i}}/p_{\circlearrowleft}^i\right)^{\mathrm{T}}$ | Probability distribution for the map function that is associated with transitions out of community $M_i$ |
| $\mathcal{H}$ | $\mathcal{H}(\mathcal{P}), \mathcal{H}(\mathcal{Q})$ | Entropy of a distribution |
| $L(M, P, \pi)$ | $q_{\curvearrowleft}\mathcal{H}(\mathcal{Q}) + \sum_{i=1}^m p_{\circlearrowleft}^i \mathcal{H}(\mathcal{P}^i)$ | Map function |
| $H$ | | Diagonal matrix with non-negative entries on its diagonal |
| $\vec{d}$ | | Generic absorption-rate vector |
| $\vec{\delta}$ | | Node-absorption-rate vector |
| $D_\delta$ | $\mathrm{diag}\{\vec{\delta}\}$ | Diagonal matrix of node-absorption rates |
| $\vec{d}_{\mathrm{s}}(D_\delta, H)$ | $(d_{\mathrm{s}})_i = h_i w_i + \delta_i$ | Scaled rate vector, which is given by the diagonal of the matrix $HW + D_\delta$ |
| $\tilde{G}(D_\delta, H)$ | | Absorption-scaled graph that is associated with the pair $(G, \vec{d}_{\mathrm{s}}(D_\delta, H))$ |
| $\tilde{\mathcal{L}}(D_\delta, H)$ | $(W - A)(HW + D_\delta)^{-1}$ | Unnormalized graph Laplacian matrix of an absorption-scaled graph |
| $\vec{u}$ | | Vector in $\mathrm{Ker}\,\mathcal{L}$ with non-negative entries that satisfy $\sum_{i=1}^n u_i = 1$ |
| $P_e(D_\delta, H)$ | $e^{-t\tilde{\mathcal{L}}(D_\delta, H)}$ | Transition-probability matrix for a Markov chain on an absorption-scaled graph |
| $P_l(D_\delta, H)$ | $I - t\tilde{\mathcal{L}}(D_\delta, H)$ | Linearized transition-probability matrix for a Markov chain on an absorption-scaled graph |
| $\tilde{A}$ | $\begin{pmatrix} A & \vec{0} \\ \vec{\delta}^{\mathrm{T}} & 0 \end{pmatrix}$ | Adjacency matrix of a graph with an absorbing state |
| $\tilde{P}$ | | Transition-probability matrix that is induced by $\tilde{A}$ |
| $Q$ | $A(W + D_\delta)^{-1}$ | Submatrix of the transition-probability matrix $\tilde{P}$ for transitions between non-absorbing states |
| $N$ | $(I - Q)^{-1}$ | Fundamental matrix of a discrete-time absorbing Markov chain |
| $\vec{t}$ | $N^T \vec{1}$ | The entry $t_i$ of $\vec{t}$ is the expected number of transitions before absorption when the initial state is $i$ |
| $\hat{N}$ | $ND_{\vec{t}}^{-1}$ | Normalized fundamental matrix of a discrete-time absorbing Markov chain |
| $P_\delta$ | $P_l(D_\delta, I, 1)$ | Special case of $P_l(D_\delta, H)$ with $H = I$ and $t = 1$ |
| $\pi_\delta^{(a)}$ | $\hat{N}\pi_0$ | The quantity $(\pi_\delta^{(a)})_i$ is the probability that node $i$ is the last node before absorption if the initial distribution is $\pi_0$ |
| $L^{(a)}(M, A, \vec{\delta}, \pi_0)$ | $L(M, P_\delta, \pi_\delta^{(a)})$ | Map function for a Markov chain with an absorbing state |

Table 1: Summary of our notation.

## 2.1 The map function and the standard InfoMap algorithm

In this subsection, we define an objective function, which is called the "map function", that the InfoMap algorithm seeks to minimize [4, 35, 36, 41].

**Definition 1. (Map function)** *Let $P = (p_{ij})_{i,j \in \{1,\dots,n\}}$ be a stochastic matrix, and let $\pi = (\pi_1, \dots, \pi_n)^{\mathrm{T}}$ be a probability distribution. For a partition $M = \{M_1, \dots, M_m\}$ of the set of nodes of a network, we define the probability of a transition out of community $M_i$ with an initial state from the distribution $\pi$ by $q_{i\frown} := \sum_{j \in M_i, k \notin M_i} \pi_j p_{kj}$ and the probability of a transition into community $M_i$ with an initial state from the distribution $\pi$ by $q_{i\frown} := \sum_{k \in M_i, j \notin M_i} \pi_j p_{kj}$. The probability of a transition into a community that is different from the community of an initial state from the distribution $\pi$ is $q_\frown := \sum_{i \in \{1,\dots,m\}} q_{i\frown}$. The* optimal mean encoding length that is associated with the probabilities of transitions into a different community *is the entropy[1] $\mathcal{H}(\mathcal{Q})$, where the distribution $\mathcal{Q}$ is*

$$\mathcal{Q} := \left( q_{1\frown}/q_\frown, \dots, q_{m\frown}/q_\frown \right)^{\mathrm{T}}.$$

*The* optimal mean encoding length that is associated with the probabilities of transitions out of community $M_i$ *is the entropy $\mathcal{H}(\mathcal{P}^i)$, where the probability distribution $\mathcal{P}^i$ is*

$$\mathcal{P}^i := \left( q_{i\frown}/p_\circlearrowleft^i, \pi_{k_1}/p_\circlearrowleft^i, \dots, \pi_{k_{n_i}}/p_\circlearrowleft^i \right)^{\mathrm{T}},$$

*where $M_i = \{k_1, \dots, k_{n_i}\}$ (with $k_1 < \cdots < k_{n_i}$) and $p_\circlearrowleft^i := q_{i\frown} + \sum_{j \in M_i} \pi_j$ is a normalization factor.[2] For the partition $M$, the* map function that is associated with $P$ and $\pi$ *is*

$$L(M, P, \pi) := q_\frown \mathcal{H}(\mathcal{Q}) + \sum_{i=1}^{m} p_\circlearrowleft^i \mathcal{H}(\mathcal{P}^i). \tag{1}$$

**Definition 2. (Standard map function)** *Let $P$ be a regular matrix (i.e., some power of $P$ has only positive entries), and let $\pi$ be its corresponding stationary distribution. We define $L(M, P)$ by $L(M, P, \pi)$, and we refer to $L(M, P)$ as a* standard map function.

The standard map function (see [36]) at a partition of a node set is based on entropies that are associated with codes that describe three types of transitions: (1) transitions into a different community, (2) transitions that end in a given community, and (3) transitions that start in a given community and then leave it. The quantity $\mathcal{H}(\mathcal{Q})$ in (1) is the entropy of a distribution $\mathcal{Q}$ that is associated with transitions into a different community, and the $\mathcal{H}(\mathcal{P}^i)$ terms in (1) are entropies of the distributions $\mathcal{P}^i$ that are associated with transitions that either end in community $M_i$ or start in $M_i$ and then leave it. One can interpret these entropies as optimal mean encoding lengths (see Definition 1) in the sense of Shannon's source-coding theorem [6]. Shannon's source-coding theorem states that if $X$ is a random variable with finitely many states and $p$ is a probability mass function with entropy $\mathcal{H}(p)$, then the mean length of a code that describes the states of $X$ cannot be smaller than $\mathcal{H}(p)$. Additionally, as the size of the set of states of $X$ becomes infinite, one can approach the lower bound arbitrarily closely. Therefore, we refer to $\mathcal{H}(p)$ as the "optimal mean encoding length" (see Theorem 6 in [40]), and we regard the map function as a weighted sum of optimal mean encoding lengths for one-step transitions between and within communities.

Let $A = (a_{ij})_{i,j \in \{1,\dots,n\}}$ be the adjacency matrix of a directed and weighted graph with node set $\{1, \dots, n\}$. The adjacency-matrix element $a_{ij}$ encodes the weight of the edge from node $j$ to node $i$, and $a_{ij} = 0$ implies that there is no edge from $j$ to $i$. The map function $L(M)$ in the following definition corresponds to the case in which the Markov chain that is induced by $A$ is regular.[3]

**Definition 3. (Map function associated with an adjacency matrix $A$)** *Let $A$ be an adjacency matrix such that $AW^{-1}$ is regular, where $W := \mathrm{diag}\{\omega_1, \dots, \omega_n\}$ and $\omega_j := \sum_i a_{ij} \neq 0$ for $j \in \{1, \dots, n\}$ is the out-degree of node $j$. The* map function $L(M)$ that is associated with $A$ *is*

$$L(M) := L(M, AW^{-1}).$$

---

[1]The entropy of a distribution with strictly positive probabilities $p_1, \dots, p_r$ is $\mathcal{H}(\{p_1, \dots, p_r\}) := -\sum_i p_i \log_2(p_i)$.

[2]If $\pi$ is the stationary distribution corresponding to $P$, then $\pi_{k_1}, \dots, \pi_{k_{n_i}}$ are the associated probabilities that a transition ends at a node in community $M_i$ and we can interpret the probabilities in $\mathcal{P}^i$ as the probabilities of transitions that start in community $M_i$ and either leave it or remain in community $M_i$.

[3]A Markov chain is *regular* if its associated transition-probability matrix is regular.

The map function $L(M)$ measures the strength of the community structure of a partition $M = \{M_1, \ldots, M_m\}$ of the set of nodes. Intuitively, if the term $q_\curvearrowright \mathcal{H}(\mathcal{Q})$ in (1) is small, then the connections between communities are sparse. If the terms $p_\circlearrowright^i \mathcal{H}(\mathcal{P}^i)$ are small, then there are dense intra-community connections [36, 39]. Therefore, we expect that minimizing $L(M)$ yields a partition with dense connections within communities and sparse connections between communities. The standard InfoMap algorithm attempts to minimize $L(M)$ using a greedy approach [9] that is reminiscent of the Leiden algorithm [43].

## 2.2 Markov time sweeping

In this subsection, we present Markov time sweeping for a version of the map function that includes a resolution parameter for community detection [39]. This resolution parameter, which amounts to a "Markov time" (which we will explain shortly), allows us to tune the sizes of the communities that we obtain using InfoMap.

The map function (1) is associated with one-step transitions of a random walk. Schaub et al. [39] incorporated Markov time sweeping[4] into InfoMap to tune the time scales of transitions by encoding transitions with steps of any length $t > 0$. We think of the Markov chain that is determined by $T_J := AW^{-1}$ (with $AW^{-1}$ specified as in Definition 3) as having time steps of size 1. Markov time sweeping uses the transition-probability matrix of a continuous-time Markov chain in which the time step is $t$ instead of 1. We refer to the time $t$ as a "Markov time". Markov times $t < 1$ yield an encoding by the map function at a smaller transition time than $t = 1$. This, in turn, yields small communities. By contrast, Markov times $t > 1$ yield an encoding at a larger transition time than $t = 1$, so the encoding is able to capture transitions of a random walker that take more than one step. We thereby obtain large communities. See [20] for further discussion of encodings and Markov times.

As discussed in [20], for $t < 1$, one can also consider the linearization

$$e^{-t(I-T_J)} \approx I - t(I - T_J) = (1 - t)I + tT_J \tag{2}$$

as an input of InfoMap. The matrix $(1-t)I + tT_J$ has diagonal elements that are all equal to $1 - t$ if we assume that $a_{ii} = 0$ for $i \in \{1, \ldots, n\}$. These diagonal elements correspond to self-edges with weight $1 - t$. If we use the standard map function with the transition-probability matrix $(1-t)I + tT_J$ instead of $T_J$, we recover the map function in Definition 3 by setting $t = 1$.

# 3 Markov time sweeping and adaptations of InfoMap to absorbing random walks

In this section, we present the adaptations of InfoMap that we use to explore how absorption in a dynamical process can affect community structure. To illustrate the implicit impact of absorption on community structure, consider an absorbing random walk on an undirected and unweighted line network with four nodes (see Figure 1). We represent an absorbing state of an absorbing random walk on a graph as a node[5] with out-degree 0. For a given node, we refer to the sum of the weights of the edges between the node and absorbing states as the "absorption rate" of the node. Suppose that the absorption rate of node 2 in Figure 1 is much larger than the absorption rates of the other nodes. The large absorption rate of node 2 is a barrier to the absorbing random walk, as transitions from nodes in the set $\{3, 4\}$ to nodes in the set $\{1\}$ (and vice versa) are unlikely. The local dynamics (specifically, the absorption at the nodes) induces a partition of the node set that we can interpret as an *effective community structure* of the network. Such effective community structure can be rather different from the community structure that one detects based on network structure alone.

---

[4]Markov sweeping time had been used previously in contexts other than InfoMap [8, 23].

[5]Throughout our paper, our figures do not show the nodes for the absorbing states that are associated with absorbing graphs. Our figures only show nodes that are associated with transient (i.e., non-absorbing) states of the corresponding absorbing random walks.
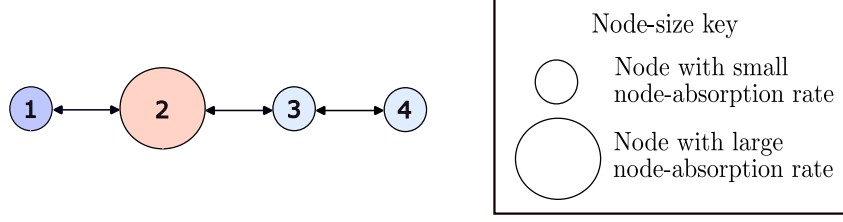
Figure 1: Consider an absorbing random walk on the depicted four-node network, and suppose that the absorption rate of node 2 is much larger than the absorption rates of the other nodes. Detecting communities via modularity maximization or the standard InfoMap algorithm produces a partition of the network into a single community that includes all nodes. However, the flow of an absorbing random walk is trapped in either the set $\{1\}$ (in dark blue) or the set $\{3, 4\}$ (in light blue). Consequently, a partition that separates node 1 from nodes 3 and 4 better captures the dynamics of an absorbing random walk than a partition of the network into a single community.

We now introduce adaptations of InfoMap that account for the absorption rates of the nodes of a network. Our approach uses *absorption-scaled graphs*, which arise naturally in the context of absorbing random walks [16].

**Definition 4. (Absorption-scaled graph)** *Let $G$ be a directed and weighted graph with adjacency matrix $A = (a_{ij})_{i,j \in \{1,\ldots,n\}}$, where $a_{ij}$ encodes the weight of the edge from node $j$ to node $i$. Let $\vec{d} = (d_1, \ldots, d_n)^{\mathrm{T}}$ be a vector (which we call an "absorption-rate vector") with positive entries that we call the "absorption rates". We define the* absorption-scaled graph *that is associated with the pair $(G, \vec{d})$ as the graph $\tilde{G}$ with adjacency matrix $\tilde{A} := AD^{-1}$, where $D := \mathrm{diag}\{d_1, \ldots, d_n\}$.*


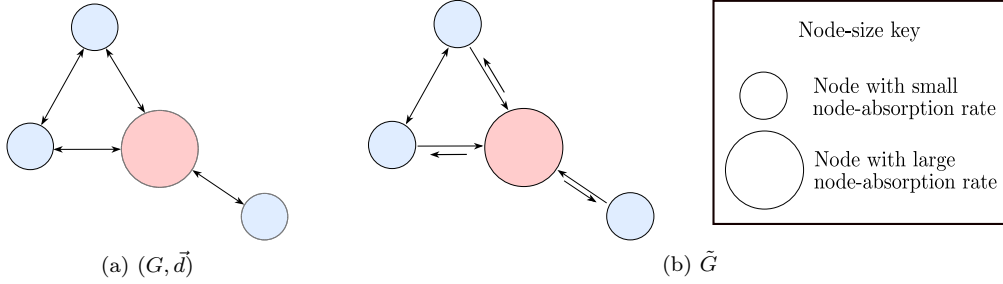
(a) $(G, \vec{d})$        (b) $\tilde{G}$

Figure 2: An absorption-scaled graph. (a) A graph $G$ with absorption-rate vector $\vec{d}$. The pink node has a large absorption rate. (b) The associated absorption-scaled graph $\tilde{G}$, where the arrow length is proportional to the corresponding edge weight.

We now define some important mathematical objects that we use in our adaptations of InfoMap.

**Definition 5.** *Let $A = (a_{ij})_{i,j \in \{1,\ldots,n\}}$ be the adjacency matrix of a graph $G$. Let $\vec{\delta} = (\delta_1, \ldots, \delta_n)^{\mathrm{T}}$ be a vector with positive entries, where $\delta_i$ is the* node-absorption rate *of node $i$ and is specified independently from the matrix $A$. We refer to the vector $\vec{\delta}$ as the* node-absorption-rate vector. *Let $D_\delta := \mathrm{diag}\{\vec{\delta}\}$ and consider a diagonal matrix $H = \mathrm{diag}\{h_1, \ldots, h_n\}$, where $h_i \geq 0$. The* scaled rate vector *$\vec{d}_{\mathrm{s}}(D_\delta, H)$ is given by the diagonal of $HW + D_\delta$.*

The node-absorption-rate vector $\vec{\delta}$ in Definition 5 does not depend on the adjacency matrix $A$, whereas the scaled rate vector $\vec{d}_{\mathrm{s}}(D_\delta, H)$ depends on the out-degrees $\omega_i$ through $H$. In particular, $\vec{\delta} = \vec{d}_{\mathrm{s}}(D_\delta, \mathbf{0})$ when $H = \mathbf{0}$. We use the notation $\vec{d}$ for a generic absorption-rate vector; it is not necessarily a node-absorption-rate vector or a scaled rate vector.

Our adaptations of InfoMap use Markov time sweeping (see Section 2.2) on absorption-scaled graphs. Consider a directed and weighted graph $G$ with adjacency matrix $A$ and node-absorption-rate vector $\vec{\delta} = (\delta_1, \ldots, \delta_n)^{\mathrm{T}}$. We consider a family of absorption-scaled graphs with absorption-rate vectors that are equal to scaled rate vectors $\vec{d}_{\mathrm{s}}(D_\delta, H)$, where $H = \mathrm{diag}\{h_1, \ldots, h_n\}$ and $h_i \geq 0$.

Let $\tilde{G}(D_\delta, H)$ denote the absorption-scaled graph that is associated with the pair $(G, \vec{d}_{\mathrm{s}}(D_\delta, H))$. The unnormalized graph Laplacian matrix of $\tilde{G}(D_\delta, H)$ is

$$\tilde{\mathcal{L}}(D_\delta, H) := (W - A)(HW + D_\delta)^{-1}. \tag{3}$$

For any Markov time $t > 0$, the transition-probability matrix that is associated with the infinitesimal generator $-\tilde{\mathcal{L}}(D_\delta, H)$ is

$$P_e(D_\delta, H, t) := e^{-t\tilde{\mathcal{L}}(D_\delta, H)} \,. \tag{4}$$

For a Markov time $t$, the linearization of $P_e(D_\delta, H, t)$ is

$$P_l(D_\delta, H, t) := I - t\tilde{\mathcal{L}}(D_\delta, H) = \left( I - tW(HW + D_\delta)^{-1} \right) + tA(HW + D_\delta)^{-1} \,. \tag{5}$$

For this linearization, we require that $0 < t < 1/\max_i\{\omega_i/(h_i\omega_i + \delta_i)\}$ to ensure that $P_l(D_\delta, H, t)$ is a transition-probability matrix.

We use the matrix $H$ in Definition 5 because this matrix allows us to tune the relative effects of the edge weights and the node-absorption rates on the communities that we detect using our adaptations of InfoMap. In Algorithms 1a and 1b, we summarize our adaptations of InfoMap.

---

**Algorithm 1a** InfoMap for absorbing random walks with a linear input.

---

Input: An adjacency matrix $A = (a_{ij})_{i,j \in \{1,\dots,n\}}$ of a directed and weighted graph, a node-absorption-rate vector $\vec{\delta} = (\delta_1, \dots, \delta_n)^{\mathrm{T}}$ with strictly positive entries, and a diagonal matrix $H$ with non-negative entries.

Output: A partition $M$ of the set of nodes that minimizes $L(M, P_l(D_\delta, H, t))$ for a Markov time $t$.

  1: Construct the unnormalized graph Laplacian matrix $\tilde{\mathcal{L}}(D_\delta, H) = (W - A)(HW + D_\delta)^{-1}$ for the absorption-scaled graph $\tilde{G}(D_\delta, H)$.

  2: Choose a Markov time such that

$$0 < t < 1/\max_i\{\omega_i/(h_i\omega_i + \delta_i)\} \,. \tag{6}$$

  3: Apply the standard InfoMap algorithm to minimize $L\left(M, P_l(D_\delta, H, t)\right)$.

---

**Algorithm 1b** InfoMap for absorbing random walks with an exponential input.

---

Input: An adjacency matrix $A = (a_{ij})_{i,j \in \{1,\dots,n\}}$ of a directed and weighted graph, a node-absorption-rate vector $\vec{\delta} = (\delta_1, \dots, \delta_n)^{\mathrm{T}}$ with positive entries, and a diagonal matrix $H$ with non-negative entries.

Output: A partition $M$ of the set of nodes that minimizes $L(M, P_e(D_\delta, H, t))$ for a Markov time $t$.

  1: Construct the unnormalized graph Laplacian matrix $\tilde{\mathcal{L}}(D_\delta, H) = (W - A)(HW + D_\delta)^{-1}$ for the absorption-scaled graph $\tilde{G}(D_\delta, H)$.

  2: Choose any Markov time $t > 0$.

  3: Apply the standard InfoMap algorithm to minimize $L\left(M, P_e(D_\delta, H, t)\right)$.

---

In our adaptations of InfoMap to absorbing random walks, we introduce a family of associated absorption-scaled graphs and then apply Markov time sweeping to these absorption-scaled graphs. To illustrate how the node-absorption rates impact the communities that we detect, consider the matrix $P_l$ in Algorithm 1a. In the expression for $P_l(D_\delta, H, t)$ in (5), the term

$$I - tW(HW + D_\delta)^{-1} = \mathrm{diag}\left\{ 1 - \frac{t\omega_i}{h_i\omega_i + \delta_i} \right\}$$

creates self-edges that are positively correlated with the node-absorption rates $\delta_i$. This correlation reflects the idea that a random walk gets stuck longer in nodes with larger node-absorption rates.

If we let $D_\delta = \delta_* I$ and $\omega_i \geq 1$ (with $i \in \{1, \dots, n\}$) and set $h_i = \delta_*(\omega_i - 1)/\omega_i$ and $t = \delta_*$, then $P_l(\delta_* I, H, \delta_*) = AW^{-1}$. We thereby recover the input of the standard InfoMap algorithm when all of the absorption rates are the same. By letting $H = hI$ and $t = h$, we obtain

$$\lim_{\|\vec{\delta}\| \to 0} P_l(D_\delta, hI, h) = \lim_{\|\vec{\delta}\| \to 0} (I - hW(hW + D_\delta)^{-1}) + hA(hW + D_\delta)^{-1} = AW^{-1} \,, \tag{7}$$

which again recovers the input of the standard InfoMap algorithm.

# 4 A map function $L^{(a)}$ for Markov chains with an absorbing state

## 4.1 A map function for an absorbing random walk

It is natural to ask if one can interpret map functions that are associated with our adaptations of InfoMap in terms of corresponding Markov chains with an absorbing state. We take a step towards answering this question by defining a map function $L^{(a)}$ for Markov chains with an absorbing state. The main results of this subsection are that (1) the map function $L^{(a)}$ is the map function that is associated with our adaptations of InfoMap in Algorithm 1a with $H = I$ and that (2) the map function $L^{(a)}$ converges to the standard map function as the absorption rates approach 0.

### 4.1.1 Construction of a map function for an absorbing random walk

We assume that the Markov chain that is associated with the adjacency matrix $A$ is regular, and we add an absorbing state (i.e., a new node with out-degree 0) and node-absorption rates $\delta_1, \ldots, \delta_n$, which are the transition rates from states that are associated with $A$ to the absorbing state. The adjacency matrix of the absorbing Markov chain is

$$\tilde{A} = \begin{pmatrix} A & \vec{0} \\ \vec{\delta}^{\mathrm{T}} & 0 \end{pmatrix},$$

where $\vec{\delta} = (\delta_1, \ldots, \delta_n)^{\mathrm{T}}$. From $\tilde{A}$, we obtain the transition-probability matrix

$$\tilde{P} = \begin{pmatrix} Q & \vec{0} \\ \vec{r}^{\mathrm{T}} & 1 \end{pmatrix}, \tag{8}$$

where the transition probabilities to absorption are the entries of $\vec{r} = (\delta_1/(\omega_1 + \delta_1), \ldots, \delta_n/(\omega_n + \delta_n))^{\mathrm{T}}$, with $\omega_j = \sum_i a_{ij}$ (for $j \in \{1, \ldots, n\}$), and $Q = A(W + D_\delta)^{-1}$, with $D_\delta = \mathrm{diag}\{\vec{\delta}\}$ and $W = \mathrm{diag}\{\omega_1, \ldots, \omega_n\}$. The Markov chain with the transition-probability matrix $\tilde{P}$ is the absorbing Markov chain that is associated with $A$ and $\delta_1, \ldots, \delta_n$.

By Definition 1, we know that if $P$ is a regular transition-probability matrix and $\pi$ is a probability mass function, then $L(M, P, \pi)$ depends only on $M$, $P$, and $\pi$. We write $L(M, P, \pi)$ as $L(M, P)$ if $\pi$ is the unique stationary distribution of $P$. The quantity $L(M, P, \pi)$ is a weighted sum of optimal mean encoding lengths for one-step transitions (with probabilities in $P$ and starting from the distribution $\pi$) between and within communities. We define a map function $L^{(a)}$ for the absorbing random walk that is associated with (8) by $L(M, P_\delta, \pi_\delta^{(a)})$ for an appropriate distribution $\pi_\delta^{(a)}$ (see Section 4.1.1.1) and an appropriate transition-probability matrix $P_\delta$ (see Section 4.1.1.2).

#### 4.1.1.1 The distribution $\pi_\delta^{(a)}$

We define a distribution $\pi_\delta^{(a)}$ with the desirable property of recovering the stationary distribution $\pi$ of a Markov chain without absorption when the absorption rates approach 0. (This Markov chain has the transition-probability matrix $AW^{-1}$.) That is,

$$\lim_{\|\vec{\delta}\| \to 0} \pi_\delta^{(a)} = \pi. \tag{9}$$

Let $N = (I - Q)^{-1} = \sum_{k=0}^{\infty} Q^k$. The entry $n_{ij}$ of $N$ gives the expected number of visits to node $i$ that start from node $j$. The matrix $N$ is the *fundamental matrix* of the corresponding absorbing Markov chain [19]. In particular, $N$ is the fundamental matrix of the absorbing Markov chain with transition-probability matrix $\tilde{P}$. Consider the normalized fundamental matrix

$$\hat{N} = N D_{\vec{t}}^{-1}, \tag{10}$$

where $\vec{t} = N^{\mathrm{T}} \vec{1} = (t_1, \ldots, t_n)^{\mathrm{T}}$, with $\vec{1} := (1, \ldots, 1)^{\mathrm{T}}$, is the vector whose entries give the expected numbers of steps before absorption starting from each non-absorbing state. For each node pair $(i, j)$, the entry $n_{ij}/t_j$ of $\hat{N}$ gives the probability that node $i$ is the last node before absorption if we start at node $j$. This probability depends on the node-absorption rates. Given an initial distribution $\pi_0$, we obtain the distribution

$$\pi_\delta^{(a)} = \hat{N} \pi_0. \tag{11}$$

The following proposition states that equation (9) holds.

**Proposition 1.** *Suppose that the Markov chain with transition-probability matrix $AW^{-1}$ is regular. Let $\vec{\delta}$ be a node-absorption-rate vector in which all entries are strictly positive, and let $D_\delta := \mathrm{diag}\{\vec{\delta}\}$. Let $N = (I - A(W + D_\delta)^{-1})^{-1}$ be the fundamental matrix of the absorbing Markov chain that is associated with $A$ and $\delta_1, \ldots, \delta_n$. Additionally, let $D_{\vec{t}}$ be the diagonal matrix with the column sums of $N$ in its diagonal, and let $\pi$ be the stationary-distribution vector that is associated with $AW^{-1}$.*

*It then follows that*

$$\lim_{\|\vec{\delta}\| \to 0} N D_{\vec{t}^{-1}} = \pi \vec{1}^{\mathrm{T}}.$$

*Proof.* Consider the vector 1-norm $\|\vec{b}\|_1 = \sum_i |b_i|$ and its induced matrix norm $\|B\|_1 = \max_j \sum_i |b_{ij}|$. Fix $\epsilon > 0$. Because $\lim_{n \to \infty} (AW^{-1})^n = \pi \vec{1}^{\mathrm{T}}$, it follows that there is a positive integer $N_1$ such that

$$\|(AW^{-1})^n - \pi \vec{1}^{\mathrm{T}}\|_1 < \epsilon \quad \text{for} \quad n \geq N_1. \tag{12}$$

In particular,

$$(AW^{-1})^{N_1} = \pi \vec{1}^{\mathrm{T}} + \Lambda, \tag{13}$$

where $\|\Lambda\|_1 < \epsilon$. Let $Q := A(W + D_\delta)^{-1}$. Because $\|Q^j\|_1 \leq \|Q\|_1^j \leq \|AW^{-1}\|_1^j = 1$ and $t_i \to \infty$ as $\|\vec{\delta}\|_1 \to 0$, it follows that

$$\lim_{\|\vec{\delta}\|_1 \to 0} \|\sum_{j < N_1} Q^j D_{\vec{t}}^{-1}\|_1 = 0. \tag{14}$$

Additionally, $\lim_{\|\vec{\delta}\|_1 \to 0} Q^{N_1} = (AW^{-1})^{N_1}$. Therefore, there is a $\eta_0 > 0$ such that

$$\left\| \sum_{j < N_1} Q^j D_{\vec{t}}^{-1} \right\|_1 < \epsilon \quad \text{and} \quad Q^{N_1} = (AW^{-1})^{N_1} + \Delta \tag{15}$$

whenever $0 < \|\vec{\delta}\|_1 < \eta_0$, where $\|\Delta\|_1 < \epsilon$. Let $\Delta' := \Lambda + \Delta$. Using (13) and (15), we obtain

$$\begin{aligned}
\sum_{j \geq N_1} Q^j D_{\vec{t}}^{-1} &= Q^{N_1} \sum_{j \geq 0} Q^j D_{\vec{t}}^{-1} \\
&= Q^{N_1} N D_{\vec{t}}^{-1} \\
&= (\pi \vec{1}^{\mathrm{T}} + \Delta') N D_{\vec{t}}^{-1} \\
&= \pi \vec{1}^{\mathrm{T}} N D_{\vec{t}}^{-1} + \Delta' N D_{\vec{t}}^{-1} \\
&= \pi \vec{t}^{\mathrm{T}} D_{\vec{t}}^{-1} + \Delta' N D_{\vec{t}}^{-1} \\
&= \pi \vec{1}^{\mathrm{T}} + \Delta' N D_{\vec{t}}^{-1}, 
\end{aligned} \tag{16}$$

where $\|\Delta' N D_{\vec{t}}^{-1}\|_1 \leq \|\Delta'\|_1 \|N D_{\vec{t}}^{-1}\|_1 < 2\epsilon$. From (15) and (16), it follows for $0 < \|\vec{\delta}\|_1 < \eta_0$ that

$$\|N D_{\vec{t}}^{-1} - \pi \vec{1}^{\mathrm{T}}\|_1 \leq \|\sum_{j < N_1} Q^j D_{\vec{t}}^{-1}\|_1 + \|\Delta' N D_{\vec{t}}^{-1}\|_1 < 3\epsilon.$$

$\square$

#### 4.1.1.2 The transition-probability matrix $P_\delta$

Let $P_\delta$ denote the linearization (5) with $H = I$ and $t = 1$. That is,

$$P_\delta := P_l(D_\delta, I, 1) = (I - W(W + D_\delta)^{-1}) + A(W + D_\delta)^{-1} = D_{\vec{r}} + Q, \tag{17}$$

where $D_{\vec{r}} = \mathrm{diag}\{\vec{r}\}$ with $\vec{r}$ as in (8). Our choice of $P_\delta$ is motivated by the property

$$\lim_{\|\vec{\delta}\| \to 0} P_\delta = P = AW^{-1}. \tag{18}$$

From (18), we recover the transition-probability matrix of the regular Markov chain that is induced by $A$ in the limit in which there is no absorption. The diagonal entries of $P_\delta$ correspond to the one-step absorption probabilities $\delta_i/(\omega_i + \delta_i)$ in the absorbing Markov chain that is associated with the transition-probability matrix $\tilde{P}$ (see (8)). The following proposition states that the time to self-transitions (i.e., a transition from a node to itself) of the Markov chain that is associated with $P_\delta$ is equal to the time to absorption in the absorbing Markov chain that is associated with $\tilde{P}$.

**Proposition 2.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be the Markov chain with transition-probability matrix $P_\delta = P_l(D_\delta, I, 1) = (p_{ij}^{(1)})_{i,j\in\{1,\dots,n\}}$. Define the random variable*

$$T_j := \min\{n : X_n = X_{n-1} \text{ and } X_0 = j\}. \tag{19}$$

*Let $\theta_j = \mathbb{E}(T_j)$ be the expected time to the first self-transition, and let $\vec{\theta} = (\theta_1, \dots, \theta_n)^{\mathrm{T}}$. If $N = (I - Q)^{-1}$ is the fundamental matrix of the absorbing Markov chain that is associated with $A$ and $\delta_1, \dots, \delta_n$, it follows that*

$$\vec{\theta}^{\mathrm{T}} = \vec{1}^{\mathrm{T}} N. \tag{20}$$

*Proof.* From the law of total expectation,

$$\theta_j = \sum_{i\neq j}(\mathbb{E}(T_i) + 1)p_{ij}^{(1)} + p_{jj}^{(1)} = \sum_{i\neq j}\theta_i p_{ij}^{(1)} + 1. \tag{21}$$

Let $(P_\delta)_{\mathrm{diag}}$ denote the diagonal matrix with the same diagonal as $P_\delta$. We write (21) as

$$\vec{\theta}^{\mathrm{T}}[I - (P_\delta - (P_\delta)_{\mathrm{diag}})] = \vec{1}^{\mathrm{T}}. \tag{22}$$

Because $P_\delta = D_{\vec{r}} + Q$, it follows from (22) that $Q = P_\delta - (P_\delta)_{\mathrm{diag}}$ and $\vec{\theta}^{\mathrm{T}} = \vec{1}^{\mathrm{T}} N$. $\qquad\square$

We now define a map function for the absorbing random walk that is associated with $A$ and $\delta_1, \dots, \delta_n$.

**Definition 6. (Map function for an absorbing random walk)** *Let $A$ be the adjacency matrix of an absorbing random walk, and let $\vec{\delta}$ be the walk's node-absorption-rate vector. Let $M$ be a partition of the node set that is associated with $A$, and let $\pi_0$ be an initial probability distribution. We define the* map function *$L^{(a)}(M, A, \vec{\delta}, \pi_0) := L(M, P_\delta, \hat{N}\pi_0)$ with $P_\delta$ in (17) and $\hat{N}$ by (10).*

As an instructive example, we calculate the map function $L^{(a)}$ for all possible partitions $M$ of the three-node network with node set $\{1, 2, 3\}$ and adjacency matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}. \tag{23}$$

Intuitively, if the node-absorption rate of node 2 is larger than the node-absorption rates of nodes 1 and 3, then node 2 is in a different community than nodes 1 and 3 in the effective community structure. We want to check whether or not the partition with the minimum value of $L^{(a)}$ captures this intuition.

Let $\vec{\delta} = (\delta_1, \delta_2, \delta_3)^{\mathrm{T}}$ be a node-absorption-rate vector. We fix $\delta_1 = \delta_3 = 0.1$ and vary $\delta_2$ in the interval $[0.1, 10]$. In Figure 3, we show the values of $L^{(a)}(M, A, \vec{\delta}, \pi_0)$ for all five possible partitions $M$ of the set $\{1, 2, 3\}$ of nodes, where $A$ is specified in (23) and $\pi_0$ is the uniform distribution on $\{1, 2, 3\}$. We always attain the minimum value of $L^{(a)}(M, A, \vec{\delta}, \pi_0)$ for the partition $\{\{2\}, \{1, 3\}\}$, so we obtain this partition if we select the optimal encoding $L^{(a)}(M, A, \vec{\delta}, \pi_0)$. Because $\delta_2 > \delta_1$ and $\delta_2 > \delta_3$, this result is consistent with the intuition that $\{\{2\}, \{1, 3\}\}$ is the effective community structure.
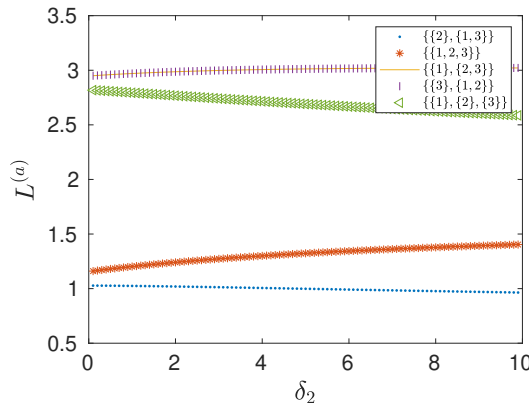


Figure 3: The map function $L^{(a)}(M, A, \vec{\delta}, \pi_0)$ for all five possible partitions $M$ of the three-node network with adjacency matrix (23). The node-absorption-rate vector is $\vec{\delta} = (\delta_1, \delta_2, \delta_3)^{\mathrm{T}}$, with $\delta_1 = \delta_3 = 0.1$ and $0.1 \leq \delta_2 \leq 10$. The initial distribution is $\pi_0 = (1/3, 1/3, 1/3)^{\mathrm{T}}$.

### 4.1.2 Main results for the map function $L^{(a)}$

Our first key result in the present subsubsection is that the standard map function $L(M, P_\delta)$, with $P_\delta$ as in (17), equals the map function $L^{(a)}(M, A, \vec{\delta}, \pi_0)$ for an appropriate distribution $\pi_0$.

**Proposition 3.** *Suppose that $P_\delta = P_l(D_\delta, I, 1) = (I - W(W + D_\delta)^{-1}) + A(W + D_\delta)^{-1} = D_{\vec{r}} + Q$. Let $\pi_\delta^{(na)}$ be the stationary distribution of $P_\delta$, and let $\pi_0 := D_{\vec{t}} D_{\vec{r}} \pi_\delta^{(na)} = (\pi_{\delta,i}^{(na)} t_i \delta_i / (\omega_i + \delta_i))_i$. We have that*

$$L(M, P_\delta) = L^{(a)}(M, A, \vec{\delta}, \pi_0). \tag{24}$$

*Proof.* Because $\pi_\delta^{(na)}$ is the stationary distribution of $P_\delta$, it follows that $(D_{\vec{r}} + Q)\pi_\delta^{(na)} = \pi_\delta^{(na)}$, which implies that

$$\pi_\delta^{(na)} = \hat{N} D_{\vec{t}} D_{\vec{r}} \pi_\delta^{(na)} = \hat{N} \pi_0. \tag{25}$$

Because the entries of $\pi_\delta^{(na)}$ sum to 1 and the columns of $\hat{N}$ each sum to 1, we know that $\pi_0$ is a probability distribution. Therefore, for this choice of $\pi_0$, it follows that

$$L(M, P_\delta) = L(M, P_\delta, \pi_\delta^{(na)}) = L(M, P_\delta, \hat{N}\pi_0) = L^{(a)}(M, A, \vec{\delta}, \pi_0), \tag{26}$$

where $L(M, P_\delta)$ is the standard map function with input $P_\delta$ and $L^{(a)}(M, A, \vec{\delta}, \pi_0)$ is the map function in Definition 6. $\square$

**Remark.** *The "na" in the superscript of $\pi_\delta^{(na)}$ stands for "non-absorbing". (The probability distribution $\pi_\delta^{(na)}$ is the stationary distribution of the non-absorbing Markov chain with transition-probability matrix $P_\delta$.)*

Our second key result in the present subsubsection is that

$$L^{(a)}(M, A, \vec{\delta}, \pi_0) \to L(M, P) \quad \text{as} \quad \|\vec{\delta}\| \to 0 \tag{27}$$

for any $\pi_0$. This result follows from $\lim_{\|\vec{\delta}\| \to 0} \pi_\delta^{(a)} = \pi$ (see Proposition 1) and $\lim_{\|\delta\| \to 0} P_\delta = P = AW^{-1}$. Therefore, the map function $L^{(a)}(M, A, \vec{\delta}, \pi_0)$ that is associated with the absorbing random walk converges to the map function $L(M, P)$ that is associated with the (non-absorbing) Markov chain in the limit $\|\vec{\delta}\| \to 0$.

## 4.2 Relating $\tilde{G}(D_\delta, I)$ to $\tilde{G}(D_\delta, \mathbf{0})$

Our adaptation of InfoMap to absorbing random walks involves a family of absorption-scaled graphs $\tilde{G}(D_\delta, H)$, where $H$ is a scaling matrix that controls the relative importances of edge weights and node-absorption rates to community detection. The choice $H = \mathbf{0}$ corresponds to an absorption-scaled graph with an absorption-rate vector whose entries are the node-absorption rates. As discussed in [16], the absorption-scaled graph $\tilde{G}(D_\delta, H)$ with $H = \mathbf{0}$ is related to the fundamental matrix of the continuous-time absorbing random walk that is associated with $A$ and $\vec{\delta}$. This relationship arises through the absorption inverse, which is a particular generalized inverse of the unnormalized graph Laplacian matrix. Our results in Section 4.1 indicate that $H = I$ is also a distinguished choice. It is natural to ask how the absorption-scaled graphs $\tilde{G}(D_\delta, \mathbf{0})$ and $\tilde{G}(D_\delta, I)$ are related. In this subsection, we find relationships between the Markov chains that are associated with the graph Laplacian matrices $\tilde{\mathcal{L}}(D_\delta, \mathbf{0})$ and $\tilde{\mathcal{L}}(D_\delta, I)$ through their associated fundamental matrices and absorption inverses. We describe these relationships, which are the main results of this subsection, in Propositions 4, 8, and 9. These results also yield connections between $\tilde{G}(D_\delta, \mathbf{0})$, $\tilde{G}(D_\delta, I)$, and the fundamental matrix $(\mathcal{L} + D_\delta)^{-1}$ through Propositions 5 and 6.

We first look at the fundamental matrices of the discrete-time Markov chains that are associated with $\tilde{\mathcal{L}}(D_\delta, \mathbf{0})$, and $\tilde{\mathcal{L}}(D_\delta, I)$. Definition 7 describes the fundamental matrix of a regular Markov chain.

**Definition 7. (Fundamental matrix)** *Let $P$ be a regular transition-probability matrix, and let $\vec{p}$ be its corresponding stationary distribution. The* fundamental matrix $Z$ *of the Markov chain that is associated with $P$ is*

$$Z = (I - P + \vec{p}\vec{1}^{\mathrm{T}})^{-1}. \tag{28}$$

The entries of the matrix $Z - \vec{p}\vec{1}^T$ approximate the differences between the expected numbers of visits of the Markov chain that is associated with $P$ and the expected numbers of visits of the Markov chain that is associated with $\vec{p}\vec{1}^{\mathrm{T}}$. (See Section 4.3 of [19].)

The following proposition gives a relationship between the fundamental matrices of the discrete-time Markov chains that are associated with the graph Laplacian matrices $\tilde{\mathcal{L}}(D_\delta, \mathbf{0})$ and $\tilde{\mathcal{L}}(D_\delta, I)$. We prove this proposition in Appendix A.

**Proposition 4.** *Let $P_0 = AW^{-1}$ be the transition-probability matrix of the discrete-time Markov chain that is associated with $\tilde{\mathcal{L}}(D_\delta, \mathbf{0})$, and let $P_1 = (A + D_\delta)(W + D_\delta)^{-1}$ be the transition-probability matrix of the discrete-time Markov chain that is associated with $\tilde{\mathcal{L}}(D_\delta, I)$. Let $\pi$ and $\pi'$ be the stationary distributions that are associated with $P_0$ and $P_1$, respectively. Let $Z_i$ be the fundamental matrix that is associated with $P_i$ (with $i \in \{1, 2\}$). Let $U := (1/(\vec{\delta}^{\mathrm{T}}\vec{u}))\vec{u}\vec{1}^{\mathrm{T}}$ and $\alpha := \vec{\delta}^{\mathrm{T}}\vec{u}/(\vec{w}^{\mathrm{T}}\vec{u} + \vec{\delta}^{\mathrm{T}}\vec{u})$, where $\vec{u} = (u_1, \ldots, u_n)^{\mathrm{T}}$ is a vector in the kernel $\mathrm{Ker}\,(W - A)$ with positive entries $u_i$ such that $\sum_{i=1}^{n} u_i = 1$. We have that*

$$Z_1 = W^{-1}(W + D_\delta)\left[Z_0 + \alpha(1-\alpha)\pi\vec{1}^{\mathrm{T}} - \alpha\left(Z_0 D_\delta U + W\frac{\vec{u}\vec{\delta}^{\mathrm{T}}}{\vec{\delta}^{\mathrm{T}}\vec{u}}W^{-1}Z_0(I - \alpha D_\delta U)\right)\right]. \tag{29}$$

**Remark.** *The expression in square brackets in the right-hand side of (29) is a fundamental matrix that is associated with $P_0$ [19].*

We now seek a connection between the unnormalized graph Laplacian matrices $\tilde{\mathcal{L}}(D_\delta, \mathbf{0})$ and $\tilde{\mathcal{L}}(D_\delta, I)$ through absorption inverses. Definition 8 defines the absorption inverse of an unnormalized graph Laplacian matrix and absorption-rate vector $\vec{d}$.

**Definition 8. (Absorption inverse)** *Let $\tilde{\mathcal{L}}$ be the unnormalized graph Laplacian matrix of a strongly connected graph, and let $\vec{d}$ be an absorption-rate vector. Let $D := \mathrm{diag}\{\vec{d}\}$, $N_{1,0} := \{\vec{x} \in \mathbb{R}^n : D\vec{x} \in \mathrm{Range}\,\tilde{\mathcal{L}}\}$, and $R_{1,0} := \{D\vec{x} : \vec{x} \in \mathrm{Ker}\,\tilde{\mathcal{L}}\}$. An absorption inverse $\tilde{\mathcal{L}}^{\vec{d}}$ of $\tilde{\mathcal{L}}$ with respect to $\vec{d}$ is defined by the following properties:*

$$\begin{aligned}
\tilde{\mathcal{L}}^{\vec{d}}\tilde{\mathcal{L}}\vec{y} &= \vec{y} \quad \text{for} \quad \vec{y} \in N_{1,0}\,, \\
\tilde{\mathcal{L}}^{\vec{d}}\vec{y} &= \vec{0} \quad \text{for} \quad \vec{y} \in R_{1,0}\,.
\end{aligned} \tag{30}$$

The absorption inverse of a graph exists and is unique if the graph is strongly connected and the absorption rates are all positive. (See Theorem 2 in [16].) As described in [16], the absorption inverse of a graph is closely related to the group inverse of an associated matrix.

**Definition 9. (Group inverse)** *Let $X$ be a square matrix such that $\mathrm{rank}(X) = \mathrm{rank}(X^2)$. The group inverse of $X$ is the unique matrix $X^\#$ that satisfies*

$$\begin{aligned}
XX^\# X &= X\,, \\
X^\# XX^\# &= X^\#\,, \\
XX^\# &= X^\# X\,.
\end{aligned} \tag{31}$$

**Proposition 5** ([16], Proposition 2). *Let $\tilde{\mathcal{L}}$ be the unnormalized graph Laplacian matrix of a strongly connected graph, let $\vec{d}$ be an absorption-rate vector, and let $D = \mathrm{diag}\{\vec{d}\}$. The following relationship holds:*

$$\left(\tilde{\mathcal{L}}D^{-1}\right)^\# = D\tilde{\mathcal{L}}^{\vec{d}}\,.$$

The absorption inverse $\tilde{\mathcal{L}}^{\vec{d}}$ is related both to the fundamental matrix $(\tilde{\mathcal{L}} + D)^{-1}$ of the continuous-time absorbing Markov chain that is associated with $\tilde{\mathcal{L}}$ and $d_1, \ldots, d_n$ and to the fundamental matrix of the associated discrete-time Markov chain. Propositions 6 and 7 give two such relationships. See [16] for the proofs of these propositions. We use these prior results to establish relationships between $\tilde{G}(D_\delta, \mathbf{0})$ and $\tilde{G}(D_\delta, I)$. We give these results in Propositions 8 and 9.

**Proposition 6** ([16], Proposition 4). *Let $\tilde{\mathcal{L}}^{\vec{d}}$ be the absorption inverse of $\tilde{\mathcal{L}}$ with respect to $\vec{d}$. Let $\vec{u}$ be a vector in $\mathrm{Ker}\,\tilde{\mathcal{L}}$ with positive entries $u_i$ such that $\sum_i u_i = 1$. Additionally, let $D := \mathrm{diag}\{\vec{d}\}$ and $U := (1/\hat{\delta})\vec{u}\vec{1}^{\mathrm{T}}$, where $\hat{d} = \sum_i(u_i d_i)$. The following relationship holds:*

$$(\tilde{\mathcal{L}} + D)^{-1} = U + (I + \tilde{\mathcal{L}}^{\vec{d}}D)^{-1}\tilde{\mathcal{L}}^{\vec{d}}\,. \tag{32}$$

If $\vec{d} = \vec{\delta}$ and $\tilde{\mathcal{L}} = \mathcal{L} = W - A$, then Propositions 5 and 6 relate the graph Laplacian matrix $\tilde{\mathcal{L}}(D_\delta, \mathbf{0}) = \mathcal{L}D_\delta^{-1}$ that has the associated absorption-scaled graph $\tilde{G}(D_\delta, \mathbf{0})$ to the continuous-time absorbing Markov chain that is

associated with $\mathcal{L} = W - A$ and $\delta_1, \ldots, \delta_n$. We suppose that the spectral radius $\rho(\mathcal{L}^{\vec{\delta}} D_\delta)$ satisfies $\rho(\mathcal{L}^{\vec{\delta}} D_\delta) < 1$ and insert the series expansion $(I + \mathcal{L}^{\vec{\delta}} D_\delta)^{-1} = \sum_{k=0} (-\mathcal{L}^{\vec{\delta}} D_\delta)^k$ into (32) to obtain

$$(\mathcal{L} + D_\delta)^{-1} = U + \mathcal{L}^{\vec{\delta}} + \sum_{k=1} (-\mathcal{L}^{\vec{\delta}} D_\delta)^k \mathcal{L}^{\vec{\delta}} . \tag{33}$$

For $\epsilon := \|\mathcal{L}^{\vec{\delta}} D_\delta\| \ll 1$, it follows from (33) that

$$\mathcal{L}^{\vec{\delta}} = (\mathcal{L} + D_\delta)^{-1} - (1/\hat{\delta}) u \vec{1}^{\mathrm{T}} + \mathcal{O}(\epsilon) . \tag{34}$$

This approximation indicates that when $\epsilon \ll 1$, the entries of $\mathcal{L}^{\vec{\delta}}$ approximate the differences in the expected times to absorption between the continuous-time absorbing Markov chain that is associated with $\mathcal{L}$ and $\delta_1, \ldots, \delta_n$ and the Markov chain without absorption that is associated with $(1/\hat{\delta}) u \vec{1}^{\mathrm{T}}$.

To compute the absorption inverse, we use the following proposition.

**Proposition 7** ([16], Lemma 3 and Theorem 3). *Let $\vec{d}$ be an absorption-rate vector, and let $\vec{u} \in \operatorname{Ker} \tilde{\mathcal{L}} = \operatorname{Ker}(W - A)$. Additionally, let $\vec{w} := W_{\mathrm{diag}}$, $\pi := W\vec{u}/(\vec{w}^{\mathrm{T}} \vec{u})$, $D := \operatorname{diag}\{\vec{\delta}\}$, $U := (1/(\vec{d}^{\mathrm{T}} \vec{u})) \vec{u} \vec{1}^{\mathrm{T}}$, and $Z := W^{-1} Z_0$, where $Z_0 = \left(I - AW^{-1} + \pi\vec{1}^{\mathrm{T}}\right)^{-1}$ is the fundamental matrix that is associated with $AW^{-1}$. We have that*

$$\tilde{\mathcal{L}}^{\vec{d}} = (I - UD) Z (I - DU) . \tag{35}$$

The following proposition relates the absorption inverse $\mathcal{L}^{\vec{\delta}}$ (which is associated with $\tilde{G}(D_\delta, \mathbf{0})$) and an absorption inverse of $\tilde{\mathcal{L}}(D_\delta, I)$ (which is associated with $\tilde{G}(D_\delta, I)$). We prove this proposition in Appendix A.

**Proposition 8.** *Let $\tilde{\mathcal{L}}_1 := \tilde{\mathcal{L}}(D_\delta, I) = (W - A)(W + D_\delta)^{-1}$, and let $\vec{d_1}$ be the diagonal of $D_\delta(W + D_\delta)^{-1}$. Additionally, let $U := \vec{u}\vec{1}^{\mathrm{T}}/(\vec{\delta}^{\mathrm{T}} \vec{u})$, $U_1 := (W + D_\delta) U$, and $D_1 := D_\delta(W + D_\delta)^{-1}$. We have that*

$$\tilde{\mathcal{L}}_1^{\vec{d_1}} = (W + D_\delta) \mathcal{L}^{\vec{\delta}} \tag{36}$$

*and*

$$(\mathcal{L} + D_\delta)^{-1} = (W + D_\delta)^{-1} \left( U_1 + (I + \tilde{\mathcal{L}}_1^{\vec{d_1}} D_1)^{-1} \tilde{\mathcal{L}}_1^{\vec{d_1}} \right) . \tag{37}$$

Let $\vec{d'} := \vec{d}_{\mathrm{s}}(D_\delta, I)$ be the absorption-rate vector that is associated with $H = I$. The following proposition relates the absorption inverse $\mathcal{L}^{\vec{\delta}}$ (which is associated with $\tilde{G}(D_\delta, \mathbf{0})$) and the absorption inverse $\mathcal{L}^{\vec{d'}}$ (which is associated with $\tilde{G}(D_\delta, I)$). We prove this proposition in Appendix A.

**Proposition 9.** *Let $\vec{d'} := \vec{d}_{\mathrm{s}}(D_\delta, I) = \vec{w} + \vec{\delta} = (\omega_1 + \delta_1, \ldots, \omega_n + \delta_n)^{\mathrm{T}}$ be the scaled rate vector that is associated with the absorption-scaled graph $\tilde{G}(D_\delta, I)$. With $\mathcal{L} = W - A$, $\alpha := \vec{\delta}^{\mathrm{T}} \vec{u}/(\vec{w}^{\mathrm{T}} \vec{u} + \vec{\delta}^{\mathrm{T}} \vec{u})$, $\pi = W\vec{u}/(\vec{w}^{\mathrm{T}} \vec{u})$, $Z_0 = \left(I - AW^{-1} + \pi\vec{1}^{\mathrm{T}}\right)^{-1}$, and $Z_* = W^{-1}\left(Z_0 - \pi\vec{1}^{\mathrm{T}}\right)$, it follows that*

$$\mathcal{L}^{\vec{d'}} = \alpha^2 \mathcal{L}^{\vec{\delta}} + \alpha(1 - \alpha)\left(\mathcal{L}^{\vec{\delta}} \mathcal{L} Z_* + Z_* \mathcal{L} \mathcal{L}^{\vec{\delta}}\right) + (1 - \alpha)^2 Z_* . \tag{38}$$

Because $Z_0$ is the fundamental matrix of the regular Markov chain with transition-probability matrix $AW^{-1}$, the entries of $Z_* = W^{-1}(Z_0 - \pi\vec{1}^{\mathrm{T}})$ measure the differences in expected visit times between the Markov chain that is associated with $AW^{-1}$ and the Markov chain that is associated with $\pi\vec{1}^{\mathrm{T}}$. Therefore, if $\|D_\delta\|/\|\mathcal{L}\| \ll 1$, then the matrix $Z_*$ is the analogue of $\mathcal{L}^{\vec{\delta}}$. From (38), we see that $\mathcal{L}^{\vec{d'}}$ is a linear combination of terms that include $\mathcal{L}^{\vec{\delta}}$ and $Z_*$, where the corresponding coefficients depend on $\alpha = \vec{\delta}^{\mathrm{T}} \vec{u}/(\vec{w}^{\mathrm{T}} \vec{u} + \vec{\delta}^{\mathrm{T}} \vec{u})$. Additionally, $\alpha \approx 1$ implies that $\mathcal{L}^{\vec{d'}} \approx \mathcal{L}^{\vec{\delta}}$ and $\alpha \approx 0$ implies that $\mathcal{L}^{\vec{d'}} \approx Z_*$.

# 5    Examples

In this section, we apply Algorithms 1a and 1b to three small networks.

## 5.1 A three-node network

We consider the three-node network (see Figure 4) with adjacency matrix $A$ in (23) and node-absorption-rate vector $\vec{\delta} = (\delta_1, \delta_2, \delta_3)^{\mathrm{T}}$.
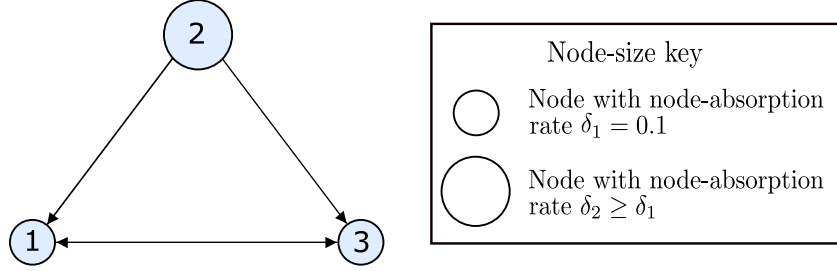


Figure 4: An example three-node network. The node-absorption rate of node 2 is greater than or equal to the node-absorption rates of nodes 1 and 3.

Consider the absorption-scaled graph $\tilde{G}(D_\delta, H)$ with $H = \mathbf{0}$ and Algorithm 1a with input $P_l(D_\delta, \mathbf{0}, t)$ for a fixed Markov time $t$. In Figure 5, we show the values of $L(M) = L(M, P_l(D_\delta, \mathbf{0}, 1/20))$ for all five possible partitions $M$ of the three nodes, where we fix $\delta_1 = \delta_3 = 0.1$ and vary $\delta_2$ in the interval $[0.1, 1]$. When all of the absorption rates are equal (i.e., when $\delta_2 = 0.1$), we see that $L(M)$ is minimized by the partitions $M = \{\{2\}, \{1, 3\}\}$ and $M = \{\{1, 2, 3\}\}$. (See the dashed orange curve.) However, when $\delta_2 > 0.1$, the partition $M = \{\{2\}, \{1, 3\}\}$ produces a smaller value of the map function (see the solid blue curve); this partition is the output of Algorithm 1a.
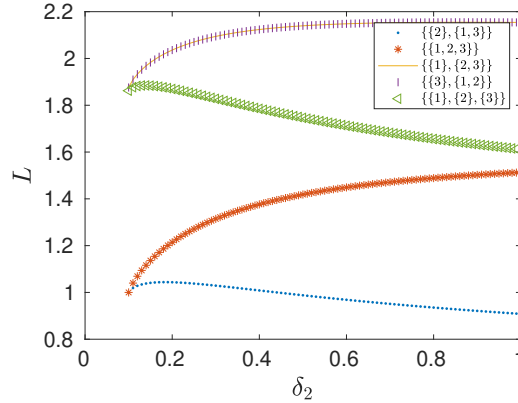


Figure 5: The values of $L(M, P_l(D_\delta, \mathbf{0}, 1/20))$ for all five possible partitions $M$ of the three-node network in Figure 4 with $\delta_1 = \delta_3 = 0.1$ and Markov time $t = 1/20$.

## 5.2 A four-clique network

Consider a network that consists of four planted cliques [see Figure 6(a)], with four nodes each. Additionally, suppose that all of the edge weights are 1. The node-absorption rates of the nodes in the cliques $C_1 := \{1, 2, 3, 4\}$ and $C_3 := \{9, 10, 11, 12\}$ are $\delta_i = 7$ (with $i \in C_1 \cup C_3$), and the node-absorption rates of the nodes in the cliques $C_2 := \{5, 6, 7, 8\}$ and $C_4 := \{13, 14, 15, 16\}$ are $\delta_i = 1$ (with $i \in C_2 \cup C_4$).

Consider the node-absorption-rate vector $\vec{\delta} = (\delta_1, \ldots, \delta_{16})^{\mathrm{T}}$ and the absorption-scaled graphs $\tilde{G}(D_\delta, \mathbf{0})$ and $\tilde{G}(D_\delta, (3/2)I)$. We use Algorithm 1a with inputs $P_l(D_\delta, \mathbf{0}, t)$ and $P_l(D_\delta, (3/2)I, t)$ and Algorithm 1b with inputs $P_e(D_\delta, \mathbf{0}, t)$ and $P_e(D_\delta, (3/2)I, t)$ for different Markov times $t$. Arguably, the network structure (i.e., the network topology and the edge weights) on its own favors the partition $M^* := \{C_1, C_2, C_3, C_4\}$ in Figure 6(b). However, the larger absorption rates in the cliques $C_1$ and $C_3$ lead to the partition $M^{**} := \{\{1\}, \{2\}, \{3\}, \{4\}, C_2, \{9\}, \{10\}, \{11\}, \{12\}, C_4\}$ in Figure 6(c).

(a) An example network with four planted cliques

(b) The partition $M^*$

(c) The partition $M^{**}$

Node-size key

○ Node with node-absorption rate $\delta_i = 1$

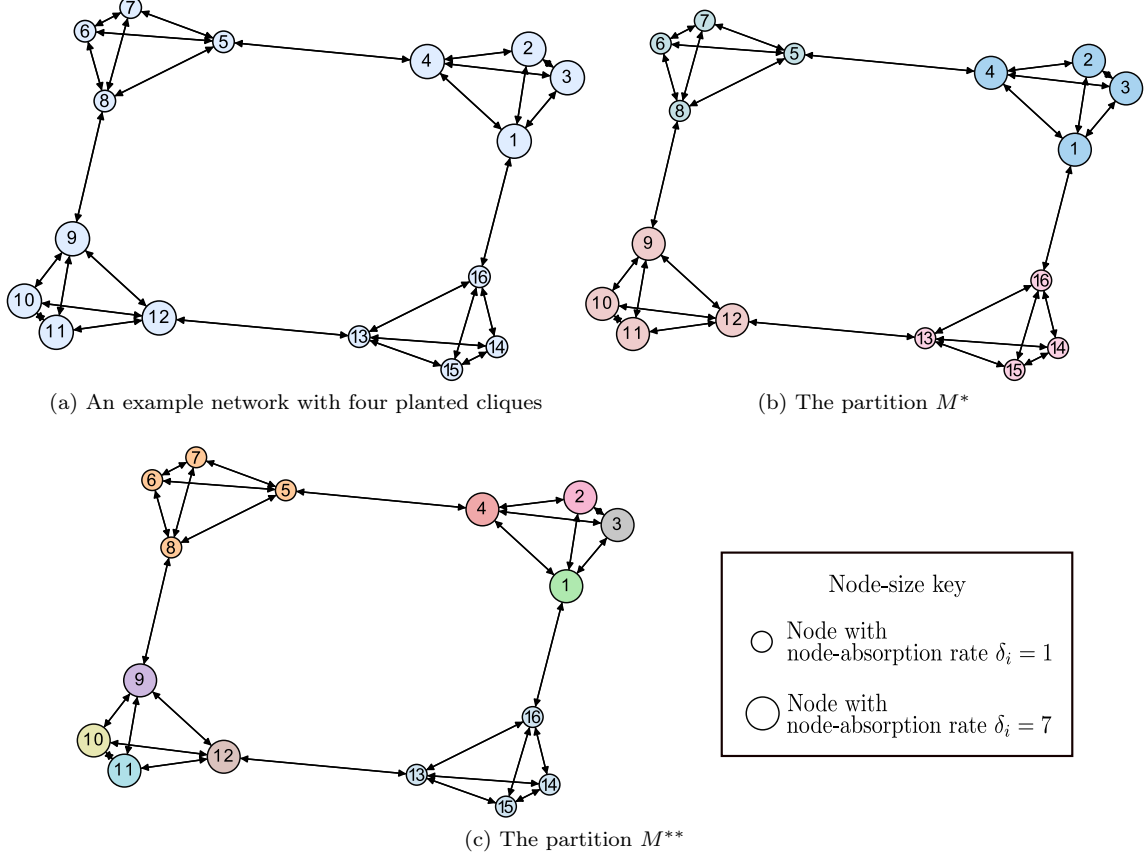◯ Node with node-absorption rate $\delta_i = 7$

Figure 6: An example network with four planted cliques and two partitions of its set of nodes. In (a), the node-absorption rates of the large nodes are $\delta_i = 7$ and the node-absorption rates of the small nodes are $\delta_i = 1$. The edge weights are all equal to 1. In (b) and (c), each color indicates a different community. The partition $M^*$ in (b) is the four-clique planted partition, which arises from the network structure (i.e., the network topology and the edge weights). The partition $M^{**}$ in (c) arises from a combination of the network structure and the node-absorption rates.

Algorithms 1a and 1b with $H = \mathbf{0}$ produce the partition $M^*$ for a smaller range of Markov times than Algorithms 1a and 1b with $H = (3/2)I$. Specifically, Algorithm 1a with the input $P_l(D_\delta, \mathbf{0}, t)$ does not produce the partition $M^*$ for any Markov time $t$ that satisfies (6) [see Figure 7(a)], whereas Algorithm 1a with the input $P_l(D_\delta, (3/2)I, t)$ produces the partition $M^*$ when $1.47 \lessapprox t \lessapprox 1.75$ [see Figure 7(c)]. Additionally, Algorithm 1b with the input $P_e(D_\delta, \mathbf{0}, t)$ produces $M^*$ when $1.28 \lessapprox t \lessapprox 2.67$ [see Figure 7(b)], whereas Algorithm 1b with the input $P_e(D_\delta, (3/2)I, t)$ produces $M^*$ when $2.01 \lessapprox t \lessapprox 13.73$ [see Figure 7(d)]. These results are consistent with the fact that the choice $H = (3/2)I$ gives more importance to the edge weights than the choice $H = \mathbf{0}$.

(a) Input $P_l(D_\delta, \mathbf{0}, t)$              (b) Input $P_e(D_\delta, \mathbf{0}, t)$

(c) Input $P_l(D_\delta, (3/2)I, t)$       (d) Input $P_e(D_\delta, (3/2)I, t)$
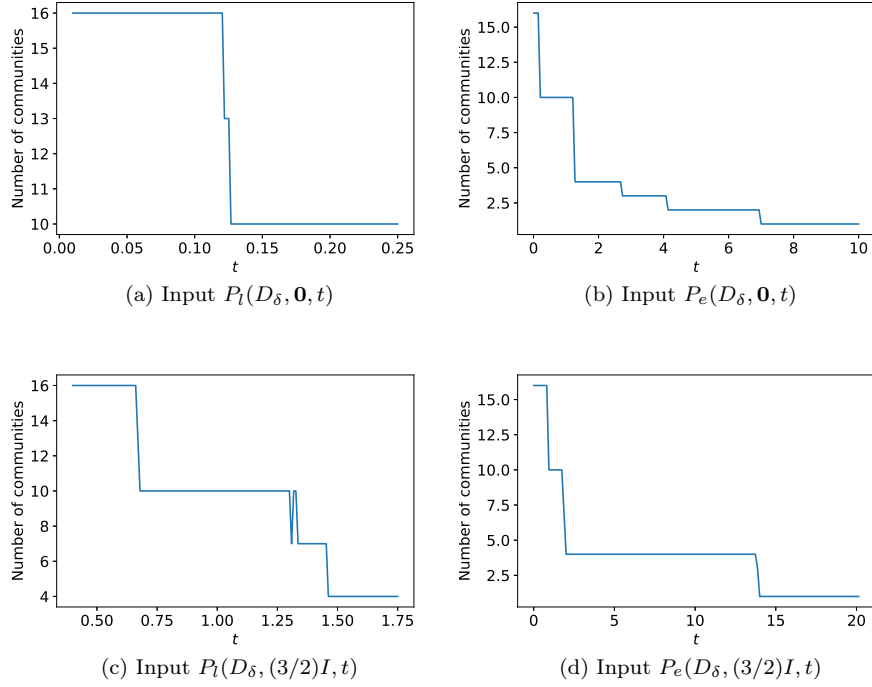
Figure 7: The numbers of communities in the partitions that we obtain using Algorithms 1a and 1b with four different inputs. The partitions that consist of four communities are the same as the partition $M^*$ in Figure 6(b), and the partitions that consist of ten communities are the same as the partition $M^{**}$ in Figure 6(c).

## 5.3    A square-lattice network with four different node-absorption rates

We consider a square-lattice network with 36 nodes (see Figure 8). We divide the set of nodes into four sublattices (with labels $B_1$, $B_2$, $B_3$, and $B_4$), which we illustrate using nodes of different sizes in Figure 8. We endow the nodes in sublattice $B_1$ with a node-absorption rate of 0.2, the nodes in sublattice $B_2$ with a node-absorption rate of 0.7, the nodes in sublattice $B_3$ with a node-absorption rate of 1.2, and the nodes in sublattice $B_4$ with a node-absorption rate of 1.7.
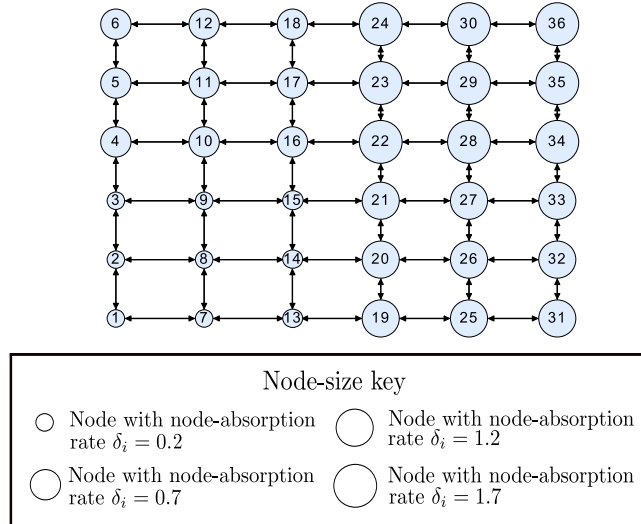


Figure 8: A square-lattice network with nodes that have node-absorption rates in the set $\{0.2, 0.7, 1.2, 1.7\}$.

We first look at the community structure that we obtain using Algorithm 1a with the input $P_l(D_\delta, \mathbf{0}, t)$ for a Markov time $t$ that satisfies (6). In Figure 9(a), we see that we obtain a partition with 28 communities

when $t \gtrsim 0.0345$. In this partition, $B_1$ is a community and all other communities consist of individual nodes [see Figures 9(a) and 9(c)]. By contrast, Algorithm 1b produces different communities than Algorithm 1a for sufficiently large Markov times $t$. For example, Algorithm 1b with the input $P_e(D_\delta, I, 5.25)$ produces the partition $\{B_1, B_2, B_3, B_4\}$ [see Figure 9(d)].



(a) Input $P_l(D_\delta, \mathbf{0}, t)$



(b) Input $P_e(D_\delta, I, t)$



(c) Resulting partition for $P_l(D_\delta, \mathbf{0}, 0.04)$

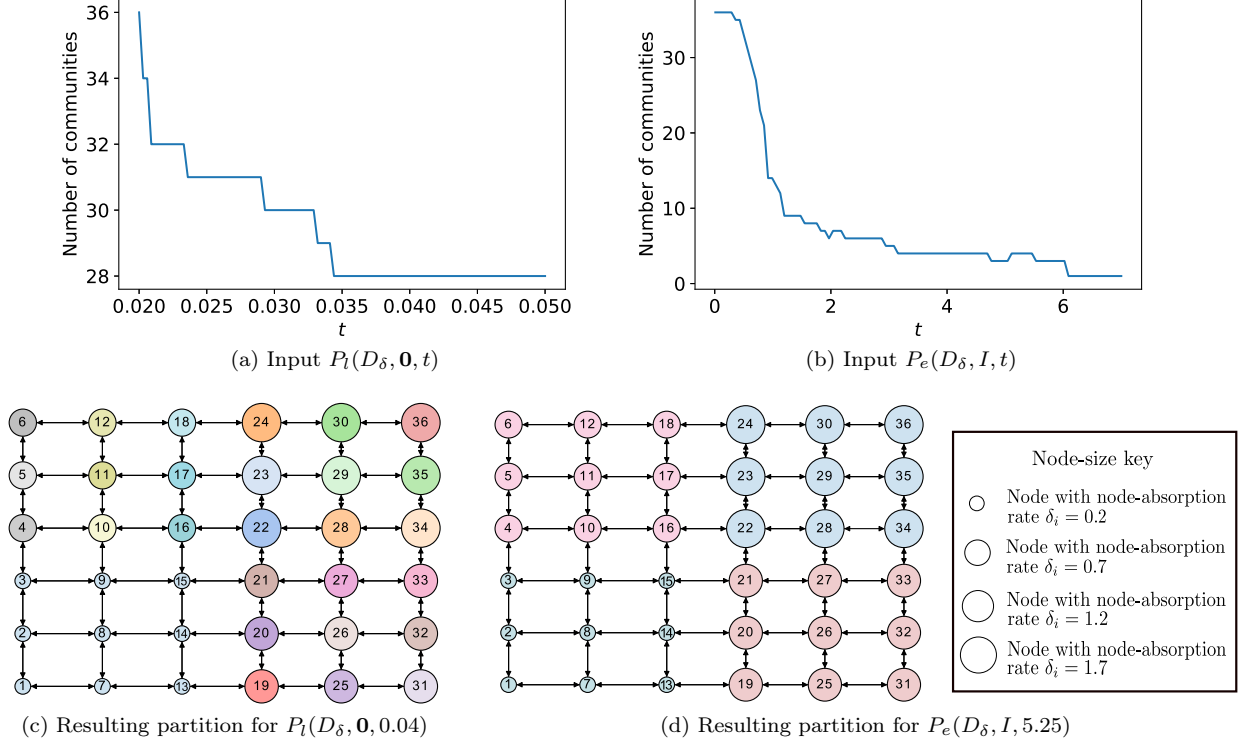(d) Resulting partition for $P_e(D_\delta, I, 5.25)$

Figure 9: (a,b) The numbers of communities in the partitions that we obtain using (a) Algorithm 1a with the input $P_l(D_\delta, \mathbf{0}, t)$ and (b) Algorithm 1b with the input $P_e(D_\delta, I, t)$. In (c) and (d), we show the partitions that we obtain from Algorithm 1a with the input $P_l(D_\delta, \mathbf{0}, 0.04)$ and Algorithm 1b with the input $P_e(D_\delta, I, 5.25)$, respectively. The colors in panels (c) and (d) indicate community assignments.

# 6 Effective community structure and susceptible–infected–recovered dynamics on networks with ring-lattice communities

Inspired by an example in Salathé and Jones [38] about the impact of community structure on disease spread, we explore the impact of the effective community structure of a contact network on the outbreak duration, final size, and outbreak peak of a disease in simulations of a susceptible–infected–recovered (SIR) model of disease spread on the network. See [22, 30] for an introduction to disease dynamics on networks.

## 6.1 An example of Salathé and Jones [38]

Salathé and Jones [38] explored the effect of differences in community structure on the outbreak duration, final size, and outbreak peak in simulations of an SIR model on networks. In our work, we consider a modified version of one of the examples in [38]. Salathé and Jones rewired edges to consider networks with different community structures, and they thereby examined how community structure impacts various aspects of disease outbreaks. Inspired by their exploration, we examine different node-absorption rates (instead of rewiring edges) to consider different community structures, and we simulate disease dynamics on the resulting networks. One of the findings of Salathé and Jones is that outbreak duration depends non-monotonically on how modular the network is. We examine whether there is a similar relationship between outbreak duration and the community structure that we detect with our adaptation of InfoMap.

We begin by describing the example of Salathé and Jones. Each stage $s$ of their example has an associated network $G_s^{\text{WS}}$ and SIR simulations on that network. In stage $s = 1$, Salathé and Jones [38] considered a network

$G_1^{\mathrm{WS}}$ that consists of Watts–Strogatz (WS) small-world graphs [32], which yield "planted communities" that are connected to each other by "community bridges" through edges that are assigned uniformly at random from all possible node pairs. The set of nodes is the same in all stages, but Salathé and Jones rewire the set of edges in each stage. In stage $s+1$, Salathé and Jones rewired one community bridge of $G_s^{\mathrm{WS}}$ (which is known from stage $s$) into one edge within a planted community in the following manner. First, they selected a community bridge $\{i_1, i_2\}$ uniformly at random from all of the community bridges of $G_s^{\mathrm{WS}}$. They then selected a node $i_k$ uniformly at random from $\{i_1, i_2\}$ and selected a node $i_3 \neq i_k$ uniformly at random from the planted community of $i_k$. Finally, they obtained the set of edges of $G_{s+1}^{\mathrm{WS}}$ by removing $\{i_1, i_2\}$ from the union of the edge $\{i_k, i_3\}$ and the set of edges of $G_s^{\mathrm{WS}}$. At each stage $s$, they simulated an SIR process on $G_s^{\mathrm{WS}}$ and recorded the means of the outbreak durations, final sizes, and outbreak peaks. In all stages, they considered the same time-independent transmission and recovery intensities in their SIR simulations.

We also describe the stochastic SIR model that Salathé and Jones employed in [38]. We simulate this SIR model on networks using a Gillespie algorithm [2, 13, 22]. Each node is either in the susceptible state (i.e., compartment) $S$, the infected state $I$, or the recovered state $R$. If a susceptible node has $k$ infected neighbors, then the time that it remains susceptible is exponentially distributed with *transmission intensity*[6] $\beta k$. With this distribution, the mean time is $1/(\beta k)$. The time that an infected node remains infected is exponentially distributed with *recovery intensity* $\gamma$, which yields a mean time of $1/\gamma$. Let |S| denote the number of susceptible nodes, |I| denote the number of infected nodes, and |R| denote the number of recovered nodes. Additionally, |SI| denotes the number of neighboring node pairs in which one node is susceptible and the others node is infected. This SIR model is a continuous-time Markov chain with two possible events: (1) one susceptible node is infected at a rate $\beta$|SI| or (2) one infected node recovers at a rate $\gamma$|I|. Two events cannot occur simultaneously. In Table 2, we show the transitions in the numbers of nodes in each state for this model.

| Transition | Rate |
|---|---|
| $\vert S\vert \to \vert S\vert - 1$, $\vert I\vert \to \vert I\vert + 1$, $\vert R\vert \to \vert R\vert$ | $\beta\vert SI\vert$ |
| $\vert S\vert \to \vert S\vert$, $\vert I\vert \to \vert I\vert - 1$, $\vert R\vert \to \vert R\vert + 1$ | $\gamma\vert I\vert$ |

Table 2: Transitions in the numbers of nodes in each compartment in the examined SIR model.

In each stage $s$, Salathé and Jones computed the value of modularity of the partition of the network $G_s^{\mathrm{WS}}$. They observed that these modularity values increase with $s$. This reflects a sparsification of the edges between the planted communities due to the edge-rewiring process.

## 6.2 Increasing node-absorption rates of bridging nodes instead of removing community bridges

### 6.2.1 Ring-lattice graphs as planted communities

We study an example that plays a similar role to the example of Salathé and Jones [38]. In our example, setting the absorption rates of bridge nodes to larger values than the absorption rates of other nodes is analogous to removing community bridges. Unlike in the example of Salathé and Jones, our example uses the same network at each stage and we change the absorption rates of specific bridging nodes instead of rewiring community bridges. Because the network is the same in all stages, maximizing modularity or using the standard InfoMap algorithm cannot reveal how effective community structure changes as we change the node-absorption rates, which correspond to the recovery intensities of a disease. However, our adaptations of InfoMap are designed for such situations. The values that we obtain for the associated map function play the role of the modularity values in [38]. They reflect an effective sparsification between planted communities due to increasing the node-absorption rates of some bridging nodes. This effective sparsification between communities, which entails a corresponding increased isolation of communities from each other, is analogous to the more literal sparsification in [38].

We use Algorithm 2 to generate the network $G^{\mathrm{RL}}$ (which is inspired by the example in [38] that we described in Section 6.1) on which we simulate an SIR process. [See Figure 10(a) for an example of a subgraph of a network that we generate using Algorithm 2.] We generate a graph $G^{\mathrm{RL}}$ using Algorithm 2 with $N_{\mathrm{WS}} = 20$ directed and unweighted ring-lattice subgraphs $G_1, \ldots, G_{20}$ (the planted communities) of size $n_{\mathrm{WS}} = 12$ (i.e., with 12 nodes each) that are connected by community bridges (which we place between node pairs that we choose uniformly at random from all possible pairs with nodes in distinct planted communities). Each node of $G^{\mathrm{RL}}$ has $k_{\mathrm{WS}} = 6$ neighbors in its planted community. We choose a small size for the ring-lattice graphs to visualize the effective

---

[6]Our use of the word "intensity" follows the terminology in [2].

community structure of one of the 20 ring-lattice subgraphs [see Figures 12(c,d)]. Each of these ring-lattice subgraphs is a one-dimensional lattice with periodic boundary conditions and additional local connections, which we specify in Algorithm 2. In the context of the analogy with the example in [38], each ring-lattice subgraph in our example is a directed WS network in which all edges are bidirectional (i.e., each edge is reciprocated) and the edge-rewiring probability is 0 [27]. We use an edge-rewiring probability of 0 because we seek to examine the effects of changing absorption rates instead of the effects of rewiring edges (which was explored in [38]). The degree of a node of a network $G^{\mathrm{RL}}$ that we generate with Algorithm 2 is the random variable $k_{\mathrm{WS}} + N_{\mathrm{bridges}}$, where $N_{\mathrm{bridges}}$ is the number of community bridges that are adjacent to the node. This random variable follows a binomial distribution with parameters $n_{\mathrm{WS}}(N_{\mathrm{WS}} - 1)$ and $2/(n_{\mathrm{WS}}(N_{\mathrm{WS}} - 1))$. The mean degree of $G^{\mathrm{RL}}$ is $k_{\mathrm{WS}} + 2$, and the variance of the degree is $2\left[1 - 2/(n_{\mathrm{WS}}(N_{\mathrm{WS}} - 1))\right]$. For the graph $G^{\mathrm{RL}}$, the standard InfoMap algorithm yields the partition that consists of the planted communities $G_1, \ldots, G_{N_{\mathrm{WS}}}$.

We use a similar process (see Algorithm 2) as in [38] to generate the network $G^{\mathrm{RL}}$ on which we simulate an SIR process. We generate a graph $G^{\mathrm{RL}}$ using Algorithm 2 with $N_{\mathrm{WS}} = 20$ directed and unweighted ring-lattice subgraphs $G_1, \ldots, G_{20}$ (the planted communities) of size $n_{\mathrm{WS}} = 12$ (i.e., with 12 nodes each) that are connected by community bridges (which we place between node pairs that we choose uniformly at random from all possible pairs with nodes in distinct planted communities). Each node of $G^{\mathrm{RL}}$ has $k_{\mathrm{WS}} = 6$ neighbors in its planted community. We choose a small size for the ring-lattice graphs to visualize the effective community structure of one of the 20 ring-lattice subgraphs [see Figures 12(c,d)]. Each of these ring-lattice subgraphs is a one-dimensional lattice with periodic boundary conditions and additional local connections, which we specify in Algorithm 2. In the context of the analogy with the example in [38], each ring-lattice subgraph in our example is a directed WS network in which all edges are bidirectional (i.e., each edge is reciprocated) and the edge-rewiring probability is 0 [27]. We use an edge-rewiring probability of 0 because we seek to examine the effects of changing absorption rates instead of the effects of rewiring edges (which was explored in [38]).

---

**Algorithm 2** Generation of the network $G^{\mathrm{RL}}$ on which we simulate an SIR process.

Input: Positive integers $n_{\mathrm{WS}}$ and $N_{\mathrm{WS}}$; an even positive integer $k_{\mathrm{WS}}$.

Output: An $N$-node directed and unweighted graph $G^{\mathrm{RL}}$, which we partition into $N_{\mathrm{WS}}$ ring-lattice subgraphs $G_1, \ldots, G_{N_{\mathrm{WS}}}$, which each have size $n_{\mathrm{WS}}$. The graph $G^{\mathrm{RL}}$ has $n_{\mathrm{WS}} \times N_{\mathrm{WS}}$ edges between nodes from distinct ring-lattice subgraphs. Each node of $G^{\mathrm{RL}}$ has $k_{\mathrm{WS}}$ neighbors in its corresponding ring-lattice subgraph.

1: Define a ring-lattice subgraph $G_1$ with the set $\{1, \ldots, n_{\mathrm{WS}}\}$ of nodes, where the neighbors of each node $i$ are the nodes $i \pm \ell \pmod{n_{\mathrm{WS}}}$, with $\ell \in \{1, \ldots, k_{\mathrm{WS}}/2\}$. The edges of this subgraph are bidirectional. (In other words, if $(i_1, i_2)$ is an edge of $G_1$, then so is $(i_2, i_1)$.)

2: For $\varphi \in \{2, \ldots, N_{\mathrm{WS}}\}$, define a ring-lattice subgraph $G_\varphi$ with the set $\{(\varphi - 1) \cdot n_{\mathrm{WS}} + 1, \ldots, \varphi \cdot n_{\mathrm{WS}}\}$ of nodes, where $G_\varphi$ is isomorphic to $G_1$. (The subgraphs $G_1, \ldots, G_{N_{\mathrm{WS}}}$ are "planted communities".)

3: Select $n_{\mathrm{WS}} \times N_{\mathrm{WS}}$ pairs $\{i_1, i_2\}$ of nodes uniformly at random from all node pairs that belong to distinct ring-lattice graphs. Add bidirectional edges between the nodes of each pair. (The pair $\{i_1, i_2\}$ is a "community bridge", and the nodes $i_1$ and $i_2$ are "bridging nodes".)

4: Set all of the edge weights to 1. (That is, we use an unweighted network.)

---

### 6.2.2 Duration, final size, and peak of SIR outbreaks

We use Algorithm 3 to run SIR simulations on the graph $G^{\mathrm{RL}}$ for different node-absorption rates (which correspond to the recovery intensities) and transmission intensities. We refer to a set of node-absorption rates $\delta_1, \ldots, \delta_N$ and transmission intensities $\beta_1, \ldots, \beta_N$ of the $N$ nodes of $G^{\mathrm{RL}}$ as a *parameter configuration*. In steps 1 and 2 of Algorithm 3, we recursively determine parameter configurations. We refer to each step of this recursive process as a *stage*. For the parameter configurations in two consecutive stages of Algorithm 3, increasing the node-absorption rates of two bridging nodes from $\delta_*$ to $\delta_{**}$ is analogous to removing a community bridge in [38] and increasing the transmission intensities of other nodes (which we call "balancing nodes") from $\beta_*$ to $\beta_{**}$ is analogous to adding edges within a community in [38]. In Figure 10, we show example parameter configurations in two stages of Algorithm 3 for a subgraph of a network that we generate with Algorithm 2.

**Algorithm 3** SIR simulations on a network $G^{\mathrm{RL}}$ (which we generate with Algorithm 2) for different node-absorption rates (i.e., recovery intensities) and transmission intensities.

Input: A graph $G^{\mathrm{RL}}$ that is the output of Algorithm 2 with inputs $n_{\mathrm{WS}}$ (the size of each planted community), $N_{\mathrm{WS}}$ (the number of planted communities), and $k_{\mathrm{WS}}$ (the number of neighbors of a node within its planted community). (Given these parameters, the graph $G^{\mathrm{RL}}$ has $n_{\mathrm{WS}} \times N_{\mathrm{WS}}$ bidirectional edges (i.e., community bridges) that connect nodes from distinct planted communities.) Node-absorption rates $\delta_*$ and $\delta_{**}$, with $\delta_{**} > \delta_*$. (They correspond to the recovery intensities.) Transmission intensities $\beta_*$ and $\beta_{**}$, with $\beta_{**} > \beta_*$. Positive integers $N_{\mathrm{S}}$ and $N_{\mathrm{sim}}$.

Output: The means of the outbreak durations, final sizes, and outbreak peaks of $N_{\mathrm{sim}}$ SIR simulations for each of the $N_{\mathrm{S}}$ parameter configurations that are defined by steps 1 and 2 of this algorithm. (A parameter configuration consists of the node-absorption rates and transmission intensities of the nodes of $G^{\mathrm{RL}}$.)

1: <u>Determine the parameter configuration in stage 1:</u> Set the node-absorption rate of each node of $G^{\mathrm{RL}}$ to $\delta_*$, and set the transmission intensity of each node of $G^{\mathrm{RL}}$ to $\beta_*$.

2: <u>Determine the parameter configuration in stage $s + 1$ for $s \in \{1, \ldots, N_{\mathrm{S}} - 1\}$:</u> Once we have determined the parameter configuration in stage $s$, select a bridging-node pair $\{i_1, i_2\}$ of $G^{\mathrm{RL}}$ uniformly at random from the set of pairs of bridging nodes that have node-absorption rate $\delta_*$. Set the node-absorption rates of nodes $i_1$ and $i_2$ to $\delta_{**}$. Additionally, select a node $j_k$ uniformly at random from the ring-lattice subgraph that is associated with $i_k$ for $k \in \{1, 2\}$. (We require that $j_k$ is distinct from $i_k$ and that it is not a neighbor of $i_k$ for $k \in \{1, 2\}$.) Set the transmission intensities of $j_1$ and $j_2$ to $\beta_{**}$.

3: <u>Run SIR simulations:</u> Run $N_{\mathrm{sim}}$ simulations of the SIR model on the graph $G^{\mathrm{RL}}$ with transmission intensities $\beta \in \{\beta_*, \beta_{**}\}$ and recovery intensities $\gamma \in \{\gamma_*, \gamma_{**}\}$ that correspond to the parameter configuration in each stage $s \in \{1, \ldots, N_{\mathrm{S}}\}$. The recovery intensities $\gamma_*$ and $\gamma_{**}$ equal the associated node-absorption rates $\delta_*$ and $\delta_{**}$. To run these simulations, we use a Gillespie algorithm [13]. Record the means of the outbreak durations, the final sizes, and the outbreak peaks of the simulations.

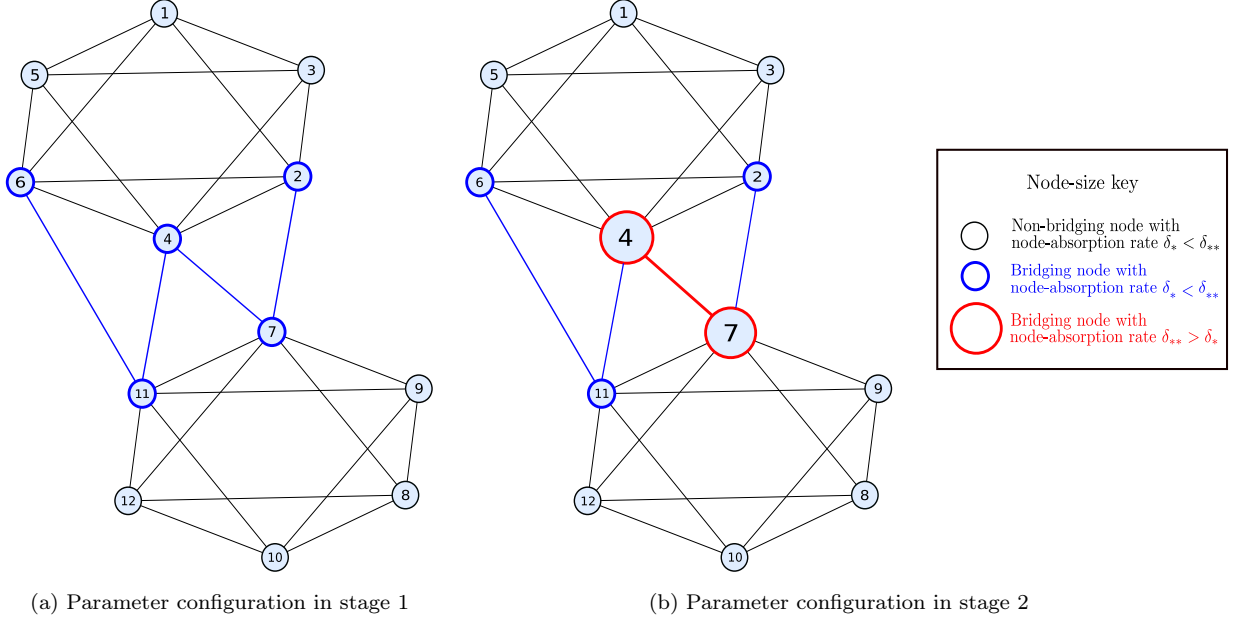(a) Parameter configuration in stage 1          (b) Parameter configuration in stage 2

Figure 10: Illustration of a subgraph of a network that we generate with Algorithm 2 and two stages of Algorithm 3 for that network. We show the parameter configurations in stages 1 and 2 of Algorithm 3 for a network $G^{\mathrm{RL}}$ from Algorithm 2 with $n_{\mathrm{WS}} = 6$, $N_{\mathrm{WS}} = 2$, and $k_{\mathrm{WS}} = 4$. In this figure, we show only four community bridges (the blue segments) of the $n_{\mathrm{WS}} \times N_{\mathrm{WS}} = 12$ community bridges of $G^{\mathrm{RL}}$ between the planted communities $G_1 = \{1, \ldots, 6\}$ and $G_2 = \{7, \ldots, 12\}$. (a) In the parameter configuration in stage 1, all 12 nodes have the same node-absorption rate $\delta_*$ and the same transmission intensity $\beta_*$. (b) In the parameter configuration in stage 2, the bridging nodes 4 and 7 have the same node-absorption rate $\delta_{**} > \delta_*$ and all other nodes have the node-absorption rate $\delta_*$. In this stage, nodes 1 and 10 (which are not in the neighborhoods of nodes 4 and 7) have the same transmission intensity $\beta_{**} > \beta_*$ and all other nodes have the transmission intensity $\beta_*$.

For each node of $G^{\mathrm{RL}}$, the recovery intensity $\gamma$ equals the associated node-absorption rate $\delta$. The recovery intensities $\gamma_*$ and $\gamma_{**}$ are $\delta_*$ and $\delta_{**}$, respectively. We choose the transmission intensity $\beta_{**}$ to compensate for the decrease in new infections that occur due to the increase of the absorption rates of the chosen bridging nodes. We estimate the decrease in the basic reproduction number $\mathcal{R}_0$ by calculating $\langle k \rangle \beta_*/\gamma_* - \langle k \rangle \beta_*/\gamma_{**}$ (using an approximation that is similar to one in [28]), where $\langle k \rangle$ is the mean degree of the bridging nodes. We compensate for the decrease in infections by choosing $\beta_{**}$ such that $\langle k \rangle \beta_{**}/\gamma_* = \langle k \rangle \beta_*/\gamma_* + \alpha \left( \langle k \rangle \beta_*/\gamma_* - \langle k \rangle \beta_*/\gamma_{**} \right)$, which estimates the new infections that arise from balancing nodes, where $\alpha$ is a tuning parameter that we use to preserve the value of the basic reproduction number. This yields

$$\beta_{**} = \beta_* + \alpha \gamma_* \beta_* \left( \frac{1}{\gamma_*} - \frac{1}{\gamma_{**}} \right). \tag{39}$$

We produce parameter configurations in $N_{\mathrm{S}} = 68$ stages [with $\delta_* = 0.2$, $\delta_{**} = 1$, $\beta_* = 0.125$, $\alpha = 0.1$, and the corresponding value of $\beta_{**}$ that we obtain using (39)]. For each parameter configuration, we run $N_{\mathrm{sim}} = 1000$ simulations of SIR dynamics on $G^{\mathrm{RL}}$ (see Algorithm 3). In Figures 11(a)–(c), we show the mean outbreak duration, final size, and outbreak peak for the parameter configuration in each of the 68 stages. The qualitative behavior of these epidemiological quantities is consistent with the observations in the example of Salathé and Jones [38]. Specifically, the mean final outbreak size and the mean outbreak peak decrease monotonically as we increase the stage number [see Figures 11(b,c)]. Additionally, the mean outbreak duration peaks at intermediate stages [see Figure 11(a)].

In Figure 11(d), we show the value of the map function $L(M_0) = L\left(M_0, P_e(D_\delta^{(s)}, \mathbf{0}, t = 0.025)\right)$ for the partition $M_0$ that consists of the planted communities of $G^{\mathrm{RL}}$, where the node-absorption rates in the diagonal entries of $D_\delta^{(s)}$ correspond to the parameter configuration in stage $s$ of Algorithm 3. (We justify the choice of Markov time $t = 0.025$ in Section 6.2.3.) This map function, which is associated with our adaptation of InfoMap in Algorithm 1b, captures the effect on community structure of changes in node-absorption rates. The standard map function cannot capture these effects. In Figure 11(d), we observe that the value of the map function $L(M_0)$ decreases with the stage number. Consequently, we conclude from Figure 11 that the map

function $L(M_0)$ decreases as one increases the number of bridging nodes with the large node-absorption rate $\delta_{**}$. (For the parameter configuration in stage $s$, there are $2s$ bridging nodes with the large node-absorption rate.) This reflects an increased isolation of the planted communities. The monotonicity of $L(M_0)$ is analogous to the increase of modularity that Salathé and Jones [38] obtained for their planted partition as they removed community bridges.
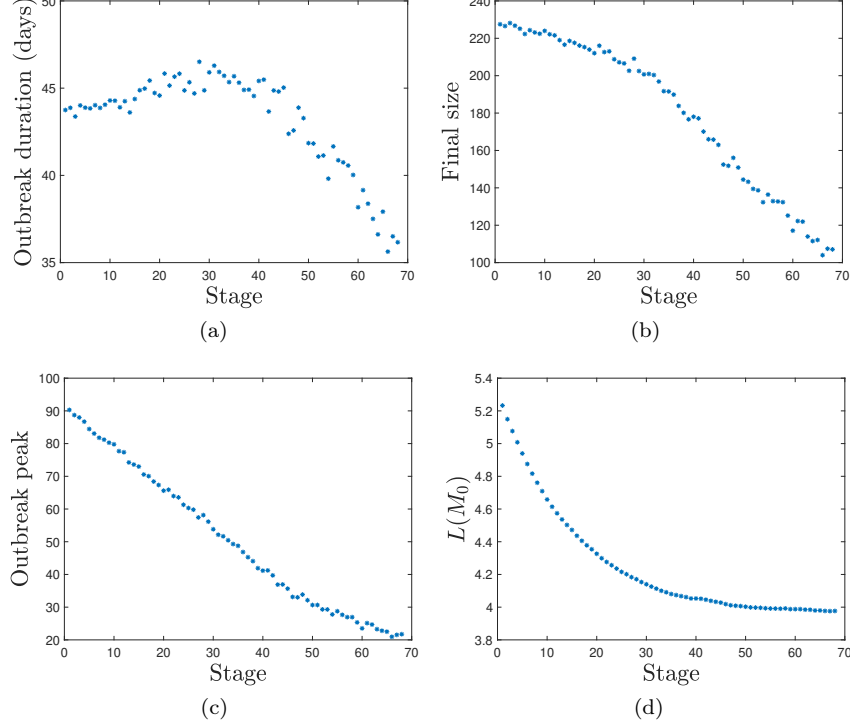


Figure 11: The (a) mean outbreak duration, (b) mean final outbreak size, and (c) mean outbreak peak as a function of the stage number of Algorithm 3 with $N_S = 68$ stages, the graph $G^{RL}$ that we generate with Algorithm 2 (for $n_{WS} = 12$, $N_{WS} = 20$, and $k_{WS} = 6$), and $N_{sim} = 1000$ simulations of SIR dynamics [for parameter configurations with $\alpha = 0.1$, $\delta_* = 0.2$, $\delta_{**} = 1$, $\beta_* = 0.125$, and the associated value of $\beta_{**}$ that we obtain using (39)]. In (d), we show the map function $L(M_0) = L(M_0, P_e(D_\delta^{(s)}, \mathbf{0}, t = 0.025))$ for the planted partition $M_0$ of $G^{RL}$ as a function the stage number of Algorithm 3 with the same inputs as in (a)–(c), where the node-absorption rates in the diagonal entries of $D_\delta^{(s)}$ correspond to the parameter configuration in stage $s$ of Algorithm 3.

### 6.2.3 Effective community structure of a single planted community

The 20 ring-lattice subgraphs $G_1, \ldots, G_{20}$ of the network $G^{RL}$ in Section 6.2.1 are isomorphic and randomly connected according to the procedure in Algorithm 2. We now examine the effective community structure of one ring-lattice subgraph of $G^{RL}$. We use Algorithm 1b with the input $P_e(D_\delta, \mathbf{0}, t)$ for the adjacency matrix of $G^{RL}$, Markov times $t \in (0.01, 0.05)$, and node-absorption rates $\delta_1, \ldots, \delta_{240}$ from the parameter configurations in three stages of Algorithm 3. We use $H = \mathbf{0}$ as an input of Algorithm 1b because the influence of the absorption on the effective community structure is particularly noticeable when $H = \mathbf{0}$ [e.g., see Figures 7(b,d)].

We use the parameter configurations from the following three stages of Algorithm 3: the initial stage (stage 1), the stage with the peak duration (stage 29), and the final stage (stage 68). We then select the planted community $G_5$ [which consists of the nodes with labels 49–60 in Figures 12(c,d)]. We choose this planted community because it has a relatively large number of bridging nodes (these are nodes 50, 53, 54, 56, and 59) and because the node-absorption rates of these nodes increases from $\delta_*$ to $\delta_{**}$ in one of the three examined stages. Given a partition $M$ of $G^{RL}$, we refer to a subset of the planted community $G_5$ as a "subcommunity" of $G_5$ if it is the intersection between $G_5$ and a community in $M$.

In Figures 12(a) and 12(b), we show the number of communities of $G^{RL}$ and number of subcommunities of $G_5$, respectively, as functions of the Markov time $t$ for partitions that we obtain using Algorithm 1b with the input $P_e(D_\delta, \mathbf{0}, t)$, where the node-absorption rates in the diagonal entries of $D_\delta$ correspond to the parameter configurations in the initial stage (see the dash-dotted green curve), the stage with the peak duration (see the

solid red curve), and the final stage (see the dashed blue curve). In Figures 12(a,b), we observe that the number of communities of the network $G^{\mathrm{RL}}$ [in Figure 12(a)] and the number of subcommunities of the subgraph $G_5$ [in Figure 12(b)] increase with the stage number.

In Figures 12(c,d), we show the subcommunities of $G_5$ for Markov time $t = 0.025$. We select $t = 0.025$ because the curves in Figures 12(a,b) are either flat near this value or change relatively little near it. In Figure 12(c), we see that the nodes with the large node-absorption rate $\delta_{**}$ (i.e., nodes 54, 56, and 59) are in different communities but that the disease can flow through the blue community and enter a different planted community in the peak-duration stage. For example, the increase in the transmission intensities of some of the nodes that belong to the same community (e.g., nodes 50, 51, and 53) implies that the disease can still spread to other planted communities (through the bridging nodes 50 and 53), so the outbreak lasts longer in this stage than in the initial stage [see Figure 11(a)].

In Figure 12(d), we see that there are more bridging nodes with the large node-absorption rate $\delta_{**}$ in the final stage than in the peak-duration stage. (Specifically, nodes 50 and 53 have the large node-absorption rate in the final stage.) Therefore, the spread of the disease to other planted communities is less likely in the final stage than in the peak-duration stage. Specifically, in stage 68 (unlike in stage 29), the disease can disappear more easily at nodes 50 and 53 (which belong to different communities in the final stage) in the final stage because nodes 50 and 53 have large node-absorption rates. Therefore, we expect that the outbreak duration, final outbreak size, and outbreak peak in Figures 11(a)–(c) in the final stage are smaller than in earlier stages.
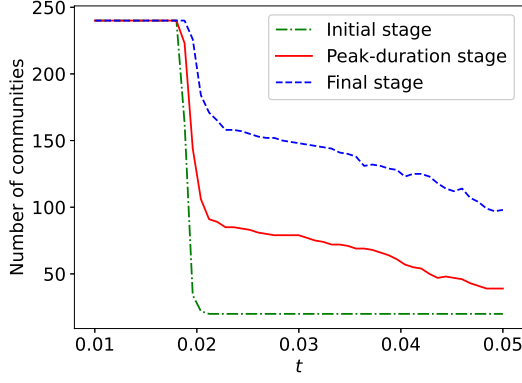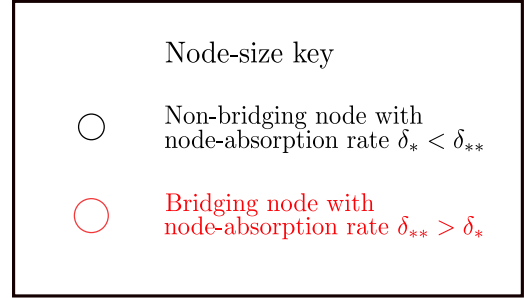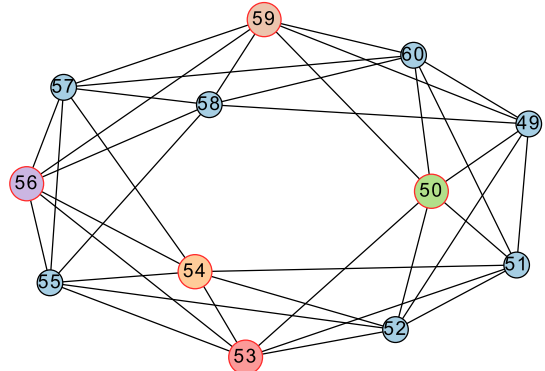
(a) Total number of communities as a function of the Markov time $t$
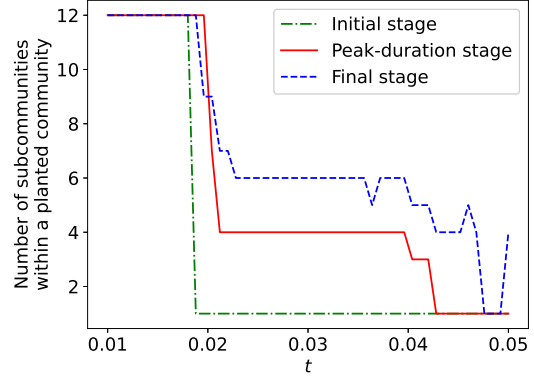
(b) Number of subcommunities of the planted community $G_5$ as a function of the Markov time $t$
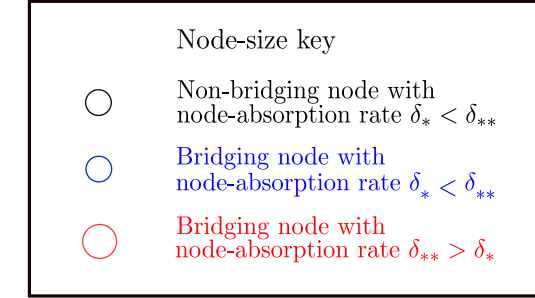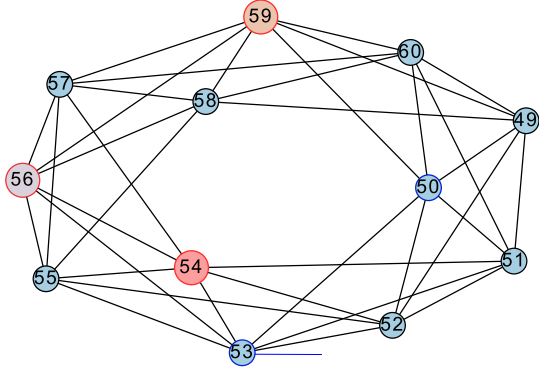
(c) Peak-duration stage ($t = 0.025$), planted community $G_5$

(d) Final stage ($t = 0.025$), planted community $G_5$

Figure 12: Comparison of community structures in the initial, peak-duration, and final stages of Algorithm 3 for SIR dynamics on $G^{\mathrm{RL}}$. (a) The number of communities in $G^{\mathrm{RL}}$ that we obtain using Algorithm 1b with the input $P_e(D_\delta, \mathbf{0}, t)$ for the node-absorption rates from the parameter configurations in three stages of Algorithm 3. (b) The number of subcommunities in the planted community $G_5$ that we obtain using Algorithm 1b for the node-absorption rates from the parameter configurations in three stages of Algorithm 3. (c,d) The subcommunities of $G_5$ that we obtain using Algorithm 1b in (c) the peak-duration stage with $t = 0.025$ and (d) the final stage with $t = 0.025$. Each color indicates a subcommunity.

## 7 Conclusions and Discussion

We adapted the community-detection algorithm InfoMap by using absorbing random walks, which are important for a variety of dynamical processes on networks. Through theoretical analysis and numerical computations of examples with small networks, we demonstrated that incorporating heterogeneous node-absorption rates leads to an effective community structure that can differ markedly from conventional notions of community structure. Through simulations of an SIR compartmental model of disease spread, we illustrated that such effective community structure can significantly influence dynamical processes on a network.

24

In our work, we considered a specific compartmental model on a specific type of random graph. It is important to examine the relationships between node-absorption rates, effective community structure, and dynamics in more general settings, including for other random-graph models and for empirical networks. Some relevant settings include mobility networks that link spatial locations, with node-absorption rates that reflect variations in habitat quality; sexual networks, with nodes corresponding to individuals and node-absorption rates that reflect heterogeneities in treatment rates in different subsets of a population; and propagation of online content, with node-absorption rates that reflect different rates of content removal.

The community-detection algorithm InfoMap is based on random walks, so it is natural to adapt it to absorbing random walks. However, there are numerous approaches to community detection [12, 33], and it is worthwhile to adapt other approaches, such as modularity maximization [29] and statistical influence using stochastic block models [31], to account for node-absorption rates. Absorption-scaled graphs provide a useful way to adapt community-detection methods (including ones that are not based on random walks) to account for heterogeneous node-absorption rates. Community structure depends not only on network structure but also on network dynamics (see, e.g., [17]), and it is important to use a variety of perspectives to examine the "effective community structure" that is associated with different dynamical processes.

# A    Appendix

We now prove Propositions 4, 8, and 9 from Section 4.2. We start with Proposition 4.

**Proposition 4.** *Let $P_0 = AW^{-1}$ be the transition-probability matrix of the discrete-time Markov chain that is associated with $\tilde{\mathcal{L}}(D_\delta, \mathbf{0})$, and let $P_1 = (A + D_\delta)(W + D_\delta)^{-1}$ be the transition-probability matrix of the discrete-time Markov chain that is associated with $\tilde{\mathcal{L}}(D_\delta, I)$. Let $\pi$ and $\pi'$ be the stationary distributions that are associated with $P_0$ and $P_1$, respectively. Let $Z_i$ be the fundamental matrix that is associated with $P_i$ (with $i \in \{1, 2\}$). Let $U := (1/(\vec{\delta}^{\mathrm{T}} \vec{u})) \vec{u} \vec{1}^{\mathrm{T}}$ and $\alpha := \vec{\delta}^{\mathrm{T}} \vec{u}/(\vec{w}^{\mathrm{T}} \vec{u} + \vec{\delta}^{\mathrm{T}} \vec{u})$, where $\vec{u} = (u_1, \ldots, u_n)^{\mathrm{T}}$ is a vector in the kernel $\mathrm{Ker}\,(W - A)$ with positive entries $u_i$ such that $\sum_{i=1}^n u_i = 1$. We have that*

$$Z_1 = W^{-1}(W + D_\delta) \left[ Z_0 + \alpha(1-\alpha)\pi\vec{1}^{\mathrm{T}} - \alpha \left( Z_0 D_\delta U + W \frac{\vec{u}\vec{\delta}^{\mathrm{T}}}{\vec{\delta}^{\mathrm{T}}\vec{u}} W^{-1} Z_0 (I - \alpha D_\delta U) \right) \right]. \qquad (29)$$

*Proof.* Let $\vec{u}$ be a column vector in $\mathrm{Ker}\,(W - A)$ whose entries $u_i$ are all positive and sum to 1. It follows that $\pi = W\vec{u}/(\vec{w}^{\mathrm{T}}\vec{u})$, $\pi' = (W + D_\delta)\vec{u}/\left(\vec{w}^{\mathrm{T}}\vec{u} + \vec{\delta}^{\mathrm{T}}\vec{u}\right)$, and

$$\pi' = (1-\alpha)\pi + \alpha\frac{D_\delta\vec{u}}{\vec{\delta}^{\mathrm{T}}\vec{u}}. \qquad (40)$$

Additionally,

$$Z_1 = \left(I - P_1 + \pi'\vec{1}^{\mathrm{T}}\right)^{-1},$$

which we can write as

$$Z_1 = W^{-1}(W + D_\delta) \left[ I - P_0 + \pi\vec{1}^{\mathrm{T}} - \alpha\pi\vec{1}^{\mathrm{T}} + \alpha\frac{D_\delta\vec{u}}{\vec{\delta}^{\mathrm{T}}\vec{u}}\vec{1}^{\mathrm{T}} + \left((1-\alpha)\pi + \alpha\frac{D_\delta\vec{u}}{\vec{\delta}^{\mathrm{T}}\vec{u}}\right)\vec{1}^{\mathrm{T}}D_\delta W^{-1} \right]^{-1}. \qquad (41)$$

We now use the Sherman–Morrison formula to compute the inverse in (41). The Sherman–Morrison formula states that if $F$ is a non-singular matrix and $\vec{v}_1$ and $\vec{v}_2$ are column vectors such that $1 + \vec{v}_2^{\mathrm{T}}F^{-1}\vec{v}_1 \neq 0$, then

$$\left(F + \vec{v}_1\vec{v}_2^{\mathrm{T}}\right)^{-1} = F^{-1} - \frac{F^{-1}\vec{v}_1\vec{v}_2^{\mathrm{T}}F^{-1}}{1 + \vec{v}_2^{\mathrm{T}}F^{-1}\vec{v}_1}. \qquad (42)$$

Define

$$\begin{aligned} F_0 &:= I - P_0 + \pi\vec{1}^{\mathrm{T}}, \\ F_1 &:= I - P_0 + \pi\vec{1}^{\mathrm{T}} - \alpha\pi\vec{1}^{\mathrm{T}}, \\ F_2 &:= I - P_0 + \pi\vec{1}^{\mathrm{T}} - \alpha\pi\vec{1}^{\mathrm{T}} + \alpha\frac{D_\delta\vec{u}}{\vec{\delta}^{\mathrm{T}}\vec{u}}\vec{1}^{\mathrm{T}} \\ F_3 &:= I - P_0 + \pi\vec{1}^{\mathrm{T}} - \alpha\pi\vec{1}^{\mathrm{T}} + \alpha\frac{D_\delta\vec{u}}{\vec{\delta}^{\mathrm{T}}\vec{u}}\vec{1}^{\mathrm{T}} + \left((1-\alpha)\pi + \alpha\frac{D_\delta\vec{u}}{\vec{\delta}^{\mathrm{T}}\vec{u}}\right)\vec{1}^{\mathrm{T}}D_\delta W^{-1}. \qquad (43) \end{aligned}$$

Additionally, recall that the fundamental matrix $Z$ of a regular Markov chain with stationary distribution $\vec{p}$ satisfies

$$Z\vec{p} = \vec{p} \quad \text{and} \quad \vec{1}^{\mathrm{T}}Z = \vec{1}^{\mathrm{T}}. \tag{44}$$

Using (42) and (44), we obtain

$$F_1^{-1} = \left(F_0 - \alpha\pi\vec{1}^{\mathrm{T}}\right)^{-1} = Z_0 + \frac{\alpha}{1-\alpha}\pi\vec{1}^{\mathrm{T}},$$

$$F_2^{-1} = \left(F_1 + \alpha\frac{D_\delta\vec{u}}{\vec{\delta}^{\mathrm{T}}\vec{u}}\vec{1}^{\mathrm{T}}\right)^{-1} = Z_0 + \alpha\pi\vec{1}^{\mathrm{T}} - \alpha Z_0 D_\delta U,$$

$$F_3^{-1} = \left(F_2 + \left((1-\alpha)\pi + \alpha\frac{D_\delta\vec{u}}{\vec{\delta}^{\mathrm{T}}\vec{u}}\right)\vec{1}^{\mathrm{T}}D_\delta W^{-1}\right)^{-1}$$

$$= Z_0 + \alpha(1-\alpha)\pi\vec{1}^{\mathrm{T}} - \alpha\left[Z_0 D_\delta U + W\frac{\vec{u}\vec{\delta}^{\mathrm{T}}}{\vec{\delta}^{\mathrm{T}}\vec{u}}W^{-1}Z_0(I - \alpha D_\delta U)\right]. \tag{45}$$

From equation (41), it follows that $Z_1 = W^{-1}(W + D_\delta)F_3^{-1}$. Combining this relation with (45) yields (29). $\square$

We now prove Proposition 8.

**Proposition 8.** Let $\tilde{\mathcal{L}}_1 := \tilde{\mathcal{L}}(D_\delta, I) = (W - A)(W + D_\delta)^{-1}$, and let $\vec{d}_1$ be the diagonal of $D_\delta(W + D_\delta)^{-1}$. Additionally, let $U := \vec{u}\vec{1}^{\mathrm{T}}/(\vec{\delta}^{\mathrm{T}}\vec{u})$, $U_1 := (W + D_\delta)U$, and $D_1 := D_\delta(W + D_\delta)^{-1}$. We have that

$$\tilde{\mathcal{L}}_1^{\vec{d}_1} = (W + D_\delta)\mathcal{L}^{\vec{\delta}} \tag{36}$$

and

$$(\mathcal{L} + D_\delta)^{-1} = (W + D_\delta)^{-1}\left(U_1 + (I + \tilde{\mathcal{L}}_1^{\vec{d}_1}D_1)^{-1}\tilde{\mathcal{L}}_1^{\vec{d}_1}\right). \tag{37}$$

*Proof.* The adjacency matrix $A_1 = P_1 = (A + D_\delta)(W + D_\delta)$ has the associated unnormalized graph Laplacian matrix $\tilde{\mathcal{L}}_1 = (W - A)(W + D_\delta)^{-1}$.

From Proposition 4, it follows that

$$Z_1 = W^{-1}(W + D_\delta)[Z_0 + \alpha R], \tag{46}$$

where

$$R = (1-\alpha)\pi\vec{1}^{\mathrm{T}} - Z_0 D_\delta U - WUD_\delta W^{-1}Z_0(I - \alpha D_\delta U).$$

Proposition 7 then implies that

$$\tilde{\mathcal{L}}_1^{\vec{d}_1} = (I - U_1 D_1)Z_1(I - D_1 U_1). \tag{47}$$

Additionally,

$$I - U_1 D_1 = (W + D_\delta)(I - U D_\delta)(W + D_\delta)^{-1},$$
$$I - D_1 U_1 = I - U D_\delta. \tag{48}$$

Substituting (48) into (47) yields

$$\tilde{\mathcal{L}}_1^{\vec{d}_1} = (W + D_\delta)\mathcal{L}^{\vec{\delta}} + \alpha(W + D_\delta)(I - U D_\delta)W^{-1}R(I - D_\delta U). \tag{49}$$

Using the relations $\pi\vec{1}^{\mathrm{T}}(I - D_\delta U) = \mathbf{0}$, $D_\delta U(I - D_\delta U) = \mathbf{0}$, and $(I - \alpha D_\delta U)(I - D_\delta U) = I - D_\delta U$ yields $R(I - D_\delta U) = -WUD_\delta W^{-1}Z_0(I - D_\delta U)$. We then use the fact that $(I - U D_\delta)U D_\delta = \mathbf{0}$ to obtain

$$(I - U D_\delta)W^{-1}R(I - D_\delta U) = \mathbf{0}. \tag{50}$$

Substituting (50) into the right-hand side of (49) yields (36).

We express the fundamental matrix $(\mathcal{L} + D_\delta)^{-1}$ as

$$(\mathcal{L} + D_\delta)^{-1} = (W + D_\delta)^{-1}\left((W - A)(W + D_\delta)^{-1} + D_\delta(W + D_\delta)^{-1}\right)^{-1}$$

$$= (W + D_\delta)(\tilde{\mathcal{L}}_1 + D_1)^{-1}. \tag{51}$$

By Proposition 6 and (36), we have

$$(\tilde{\mathcal{L}}_1 + D_1)^{-1} = U_1 + (I + \tilde{\mathcal{L}}_1^{\vec{d}_1} D_1)^{-1} \tilde{\mathcal{L}}_1^{\vec{d}_1}. \tag{52}$$

Equations (51) and (52) then yield (37).

$\square$

We now prove Proposition 9.

**Proposition 9.** *Let $\vec{d'} := \vec{d}_s(D_\delta, I) = \vec{w} + \vec{\delta} = (\omega_1 + \delta_1, \ldots, \omega_n + \delta_n)^{\mathrm{T}}$ be the scaled rate vector that is associated with the absorption-scaled graph $\tilde{G}(D_\delta, I)$. With $\mathcal{L} = W - A$, $\alpha := \vec{\delta}^{\mathrm{T}} \vec{u}/(\vec{w}^{\mathrm{T}} \vec{u} + \vec{\delta}^{\mathrm{T}} \vec{u})$, $\pi = W\vec{u}/(\vec{w}^{\mathrm{T}} \vec{u})$, $Z_0 = \left(I - AW^{-1} + \pi \vec{1}^{\mathrm{T}}\right)^{-1}$, and $Z_* = W^{-1}\left(Z_0 - \pi \vec{1}^{\mathrm{T}}\right)$, it follows that*

$$\mathcal{L}^{\vec{d'}} = \alpha^2 \mathcal{L}^{\vec{\delta}} + \alpha(1 - \alpha)\left(\mathcal{L}^{\vec{\delta}} \mathcal{L} Z_* + Z_* \mathcal{L} \mathcal{L}^{\vec{\delta}}\right) + (1 - \alpha)^2 Z_*. \tag{38}$$

*Proof.* By Proposition 7, we have

$$\mathcal{L}^{\vec{d'}} = \left(I - \frac{\vec{u}\vec{1}^{\mathrm{T}}}{\vec{w}^{\mathrm{T}}\vec{u} + \vec{\delta}^{\mathrm{T}}\vec{u}}(W + D_\delta)\right) W^{-1} Z_0 \left(I - \frac{1}{\vec{w}^{\mathrm{T}}\vec{u} + \vec{\delta}^{\mathrm{T}}\vec{u}}(W + D_\delta)\vec{u}\vec{1}^{\mathrm{T}}\right). \tag{53}$$

Furthermore,

$$I - \frac{\vec{u}\vec{1}^{\mathrm{T}}}{\vec{w}^{\mathrm{T}}\vec{u} + \vec{\delta}^{\mathrm{T}}\vec{u}}(W + D_\delta) = \alpha\left(I - \frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}D_\delta\right) + (1 - \alpha)W^{-1}(I - \pi\vec{1}^{\mathrm{T}})W,$$

$$I - \frac{1}{\vec{w}^{\mathrm{T}}\vec{u} + \vec{\delta}^{\mathrm{T}}\vec{u}}(W + D_\delta)\vec{u}\vec{1}^{\mathrm{T}} = \alpha\left(I - D_\delta\frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}\right) + (1 - \alpha)(I - \pi\vec{1}^{\mathrm{T}}). \tag{54}$$

Substituting (54) into (53) yields

$$\mathcal{L}^{\vec{d'}} = \alpha^2\left(I - \frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}D_\delta\right) W^{-1} Z_0\left(I - D_\delta\frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}\right) + \alpha(1 - \alpha)\left(I - \frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}D_\delta\right) W^{-1} Z_0(I - \pi\vec{1}^{\mathrm{T}})$$

$$+ \alpha(1 - \alpha)W^{-1}(I - \pi\vec{1}^{\mathrm{T}})WW^{-1} Z_0\left(I - D_\delta\frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}\right) + (1 - \alpha)^2 W^{-1}(I - \pi\vec{1}^{\mathrm{T}})WW^{-1} Z_0(I - \pi\vec{1}^{\mathrm{T}}). \tag{55}$$

By Proposition 6, the first term of the right-hand side of (55) is $\alpha^2 \mathcal{L}^{\vec{\delta}}$. Note that $\pi$ is the stationary distribution of the Markov chain with transition-probability matrix $P_0$. Additionally, $Z_0$ is the fundamental matrix of this Markov chain. Therefore, using (44), it follows that $W^{-1} Z_0(I - \pi\vec{1}^{\mathrm{T}}) = W^{-1}(I - \pi\vec{1}^{\mathrm{T}})Z_0 = W^{-1}(Z_0 - \pi\vec{1}^{\mathrm{T}}) = Z_*$ and $W^{-1}(I - \pi\vec{1}^{\mathrm{T}})Z_0(I - \pi\vec{1}^{\mathrm{T}}) = W^{-1}(Z_0 - \pi\vec{1}^{\mathrm{T}}) = Z_*$. Consequently, from (55), we obtain

$$\mathcal{L}^{\vec{d'}} = \alpha^2 \mathcal{L}^{\vec{\delta}} + \alpha(1 - \alpha)\left[(I - \frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}D_\delta)Z_* + Z_*\left(I - D_\delta\frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}\right)\right] + (1 - \alpha)^2 Z_*. \tag{56}$$

Additionally, $\mathcal{L}^{\vec{\delta}}\mathcal{L} = I - \frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}D_\delta$ and $\mathcal{L}\mathcal{L}^{\vec{\delta}} = I - D_\delta\frac{\vec{u}\vec{1}^{\mathrm{T}}}{\delta^{\mathrm{T}}\vec{u}}$ (see Theorem 1 and Lemma 1 in [16]), so (38) follows from (56). $\square$

# Data and code availability

The code that yields the numerical results in Figures 11 and 12 of Section 6 are available at https://gitlab.com/esteban_vargas_bernal/extending-infomap-to-absorbing-random-walks.

# Acknowledgements

# References

[1] M. A. Acevedo, J. A. Sefair, J. C. Smith, B. Reichert, and R. J. Fletcher. Conservation under uncertainty: Optimal network protection strategies for worst-case disturbance events. *Journal of Applied Ecology*, 52(6):1588–1597, 2015.

[2] H. Andersson and T. Britton. *Stochastic epidemic models and their statistical analysis*, volume 151. Springer Science & Business Media, 2012.

[3] S. Bhagat, G. Cormode, and S. Muthukrishnan. Node classification in social networks. *Social Network Data Analytics*, pages 115–148, 2011.

[4] L. Bohlin, D. Edler, A. Lancichinetti, and M. Rosvall. Community detection and visualization of networks with the Map Equation framework. In *Measuring Scholarly Impact*, pages 3–34. Springer-Verlag, Heidelberg, Germany, 2014.

[5] F. Brauer, C. Castillo-Chavez, and Z. Feng. *Mathematical Models in Epidemiology*. Springer-Verlag, Heidelberg, Germany, 2019.

[6] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Hoboken, NJ, USA, 2012.

[7] J. De, X. Zhang, F. Lin, and L. Cheng. Transduction on directed graphs via absorbing random walks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(7):1770–1784, 2017.

[8] J. C. Delvenne, S. N. Yaliraki, and M. Barahona. Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences of the United States of America*, 107(29):12755–12760, 2010.

[9] D. Edler, L. Bohlin, and M. Rosvall. Mapping higher-order network flows in memory and multilayer networks with Infomap. *Algorithms*, 10(4):112, 2017.

[10] R. J. Fletcher, M. E. Iezzi, R. Guralnick, A. J. Marx, S. J. Ryan, and D. Valle. A framework for linking dispersal biology to connectivity across landscapes. *Landscape Ecology*, 38:2487–2500, 2023.

[11] R. J. Fletcher, J. A. Sefair, C. Wang, C. L. Poli, T. Smith, E. M. Bruna, R. D. Holt, M. Barfield, A. J. Marx, and M. A. Acevedo. Towards a unified framework for connectivity that disentangles movement and mortality in space and time. *Ecology Letters*, 22(10):1680–1689, 2019.

[12] S. Fortunato and D. Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1–44, 2016.

[13] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[14] A. J. Gurfinkel and P. A. Rikvold. Absorbing random walks interpolating between centrality measures on complex networks. *Physical Review E*, 101(1):012302, 2020.

[15] L. Huang, K. Park, and Y.-C. Lai. Information propagation on modular networks. *Physical Review E*, 73(3):035103, 2006.

[16] K. A. Jacobsen and J. H. Tien. A generalized inverse for graphs with absorption. *Linear Algebra and its Applications*, 537:118–147, 2018.

[17] L. G. S. Jeub, P. Balachandran, M. A. Porter, P. J. Mucha, and M. W. Mahoney. Think locally, act locally: Detection of small, medium-sized, and large communities in large networks. *Physical Review E*, 91(1):012821, 2015.

[18] L. G. S. Jeub, M. W. Mahoney, P. J. Mucha, and M. A. Porter. A local perspective on community structure in multilayer networks. *Network Science*, 5(2):144–163, 2017.

[19] J. G. Kemeny and J. L. Snell. *Finite Markov Chains: With a New Appendix "Generalization of a Fundamental Matrix"*. Springer-Verlag, Heidelberg, Germany, 1983.

[20] M. Kheirkhahzadeh, A. Lancichinetti, and M. Rosvall. Efficient community detection of network flows for varying Markov times and bipartite networks. *Physical Review E*, 93(3):032309, 2016.

[21] S. Kim, J. Breen, E. Dudkina, F. Poloni, and E. Crisostomi. On the use of Markov chains for epidemic modeling on networks. *arXiv preprint arXiv:2207.02737*, 2022.

[22] I. Z. Kiss, J. C. Miller, and P. L. Simon. *Mathematics of Epidemics on Networks: From Exact to Approximate Models*. Springer International Publishing, Cham, Switzerland, 2017.

[23] R. Lambiotte, J. C. Delvenne, and M. Barahona. Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*, 2008.

[24] B. L. Levy, K. Vachuska, S. V. Subramanian, and R. J. Sampson. Neighborhood socioeconomic inequality based on everyday mobility predicts COVID-19 infection in San Francisco, Seattle, and Wisconsin. *Science Advances*, 8(7):eabl3825, 2022.

[25] N. Masuda, M. A Porter, and R. Lambiotte. Random walks and diffusion on networks. *Physics Reports*, 716:1–58, 2017.

[26] D. Mistry et al. Inferring high-resolution human mixing patterns for disease modeling. *Nature Communications*, 12(1):323, 2021.

[27] M. E. J. Newman. Models of the small world. *Journal of Statistical Physics*, 101(3):819–841, 2000.

[28] M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, 2002.

[29] M. E. J. Newman. *Networks*. Oxford University Press, Oxford, UK, second edition, 2018.

[30] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925, 2015.

[31] T. P. Peixoto. Bayesian stochastic blockmodeling. In Patrick Doreian, Vladimir Batagelj, and Anuska Ferligoj, editors, *Advances in Network Clustering and Blockmodeling*, pages 289–332. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2019.

[32] M. A. Porter. Small-world network. *Scholarpedia*, 7(2):1739, 2012.

[33] M. A. Porter, J.-P. Onnela, and P. J. Mucha. Communities in networks. *Notices of the American Mathematical Society*, 56(9):1082–1097, 1164–1166, 2009.

[34] O. Ratmann et al. Quantifying HIV transmission flow between high-prevalence hotspots and surrounding communities: A population-based study in Rakai, Uganda. *The Lancet HIV*, 7(3):e173–e183, 2020.

[35] M. Rosvall, D. Axelsson, and C. T. Bergstrom. The map equation. *The European Physical Journal — Special Topics*, 178(1):13–23, 2009.

[36] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, 105(4):1118–1123, 2008.

[37] N. W. Ruktanonchai et al. Identifying malaria transmission foci for elimination using human mobility data. *PLOS Computational Biology*, 12(4):1–19, 04 2016.

[38] M. Salathé and J. H. Jones. Dynamics and control of diseases in networks with community structure. *PLoS Computational Biology*, 6(4):e1000736, 2010.

[39] M. T. Schaub, R. Lambiotte, and M. Barahona. Encoding dynamics for multiscale community detection: Markov time sweeping for the map equation. *Physical Review E*, 86(2):026112, 2012.

[40] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[41] J. Smiljanić, C. Blöcker, A. Holmgren, D. Edler, M. Neuman, and M. Rosvall. Community detection with the map equation and Infomap: Theory and applications. *arXiv preprint arXiv:2311.04036*, 2023.

[42] C. Stegehuis, R. van der Hofstad, and J. S. H. Van Leeuwaarden. Epidemic spreading on complex networks with community structures. *Scientific Reports*, 6:29748, 2016.

[43] V. A. Traag, L. Waltman, and N. J. Van Eck. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9:5233, 2019.

[44] P. van den Driessche and J. Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180:29–48, 2002.

[45] L. Weng, F. Menczer, and Y.-Y. Ahn. Virality prediction and community structure in social networks. *Scientific Reports*, 3:2522, 2013.

[46] X. Zhu, A. B. Goldberg, J. Van Gael, and D. Andrzejewski. Improving diversity in ranking using absorbing random walks. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 97–104, 2007.