Network Science **7** (1): 88–122, 2019
doi:10.1017/nws.2018.29

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Randomized optimal transport on a graph: framework and new distance measures

Guillaume Guex[1*], Ilkka Kivimäki[2], and Marco Saerens[3]

[1]ICTEAM, Université catholique de Louvain, Louvain-la-Neuve, Belgium (email: guillaume.guex@gmail.com), [2]Department of Computer Science, Aalto University, Finland & ICTEAM, Université catholique de Louvain, Belgium (email: ilkka.kivimaki@aalto.fi) and [3]ICTEAM, Université catholique de Louvain & Université Libre de Bruxelles, Belgium (email: macro.saerens@uclouvain.be)

Action Editor: Ann McCranie and Ulrik Brandes

Abstract

The recently developed bag-of-paths (BoP) framework consists in setting a Gibbs–Boltzmann distribution on all feasible paths of a graph. This probability distribution favors short paths over long ones, with a free parameter (the temperature T) controlling the entropic level of the distribution. This formalism enables the computation of new distances or dissimilarities, interpolating between the shortest-path and the resistance distance, which have been shown to perform well in clustering and classification tasks. In this work, the bag-of-paths formalism is extended by adding two independent equality constraints fixing starting and ending nodes distributions of paths (margins). When the temperature is low, this formalism is shown to be equivalent to a relaxation of the optimal transport problem on a network where paths carry a flow between two discrete distributions on nodes. The randomization is achieved by considering free energy minimization instead of traditional cost minimization. Algorithms computing the optimal free energy solution are developed for two types of paths: hitting (or absorbing) paths and non-hitting, regular, paths and require the inversion of an $n \times n$ matrix with n being the number of nodes. Interestingly, for regular paths on an undirected graph, the resulting optimal policy interpolates between the deterministic optimal transport policy ($T \to 0^+$) and the solution to the corresponding electrical circuit ($T \to \infty$). Two distance measures between nodes and a dissimilarity between groups of nodes, both integrating weights on nodes, are derived from this framework.

Keywords: network science, optimal transportation, bag-of-paths, randomized shortest path, distances between nodes, link analysis

Author ORCID. Guillaume Guex, https://orcid.org/0000-0003-1001-9525

1. Introduction

1.1 General introduction and motivation

Today, network data are studied in many different areas of science, including applied mathematics, computer science, social science, physics, chemistry, pattern recognition, applied statistics, data mining, and machine learning, to name a few (see, e.g., Barabási, 2016; Chung & Lu, 2006; Estrada, 2012; Fouss et al., 2016; Kolaczyk, 2009; Lewis, 2009; Newman, 2010; Silva & Zhao, 2016; Thelwall, 2004; Wasserman & Faust, 1994). In this context, one key problem is the definition of distances between nodes taking both direct and indirect connections into account (Chebotarev, 2011, 2012, 2013; Fouss et al., 2016; Herbster & Lever, 2009; Françoisse et al., 2017; Lü & Zhou, 2011; Alamgir & von Luxburg, 2011; Yen et al., 2008). This problem is faced in many applications such as link prediction, community detection, node classification, and network visualization among others.

© Cambridge University Press 2019

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Now, it has been shown that the standard shortest path distance and the resistance distance (Klein & Randić, 1993) suffer from important drawbacks in some situations, which sometimes hinders their use as distance measures between nodes. More precisely, the shortest path distance does not integrate the concept of high connectivity between the two nodes (it only considers the shortest paths, see, e.g., Fouss et al., 2016), while the resistance distance provides useless results when dealing with large graphs (the so-called lost-in-space effect; von Luxburg et al., 2010, 2014). Another drawback of the shortest path distance is that it provides lots of ties when comparing distances, especially on unweighted undirected graphs.

In order to avoid the drawbacks of the shortest path and resistance distances, new families of distance measures, interpolating between these two extremes, were recently suggested based on a *bag-of-paths* (BoP) *framework* (Françoisse et al., 2017; Kivimäki et al., 2014; Lebichot et al., 2014; Mantrach et al., 2010). This framework defines a Gibbs–Boltzmann probability distribution over paths on a graph, which focuses on the shortest paths, but spreads also on longer paths and random walks. The spread of the distribution is controlled by a temperature parameter monitoring the balance between choosing low-cost paths and a pure random behavior. Different distance measures between nodes are then derived based on this distribution; other ones are described in the next, related work, subsection.

Following this previous work, the effort is pursued in this paper with the introduction of *weighted* distance measures derived from new bag-of-paths (BoP) models. The weights of the distances are determined by introducing equality constraints on the path distribution margins, i.e., the a priori probabilities over starting nodes and ending nodes of paths. In other words, the new model assumes that the user knows only where paths on average start and where they on average end but not how they are distributed otherwise. In the original BoP model, which was developed for the unweighted BoP-based distances, the starting and ending node distributions are instead unconstrained and can be inferred directly from the probabilities of the paths (Françoisse et al., 2017). The model proposed in the current work will be called the *margin-constrained BoP framework*. More precisely, the work defines two models by considering two different types of paths—the first one is based on regular, non-hitting paths, and the second on hitting paths, i.e., paths where the ending node cannot appear as an intermediate node.

Weighting the nodes of a network for determining distances can be important in applications where each node represents a whole collection of items (like cities where nodes could be weighted by population). Moreover, in some situations, it could be beneficial to weigh nodes by the reciprocal of their degree in order to avoid the *hubness effect* (Radovanović et al., 2010a). This will be investigated experimentally in further work. In addition to defining weighted distances between nodes on a graph, a dissimilarity measure between groups of nodes on the graph is derived from the margin-constrained BoP framework.

The margin-constrained BoP model can also be understood as defining a *randomized policy* for the *optimal transport problem on a graph* (Ahuja et al., 1993; Kantorovich, 1942; Villani, 2003, 2008) because the starting and ending node distributions can be considered as *supply* and *demand* distributions for goods meant to be transported over the graph. The randomization is achieved by finding the probability mass on the set of paths connecting starting and ending nodes that minimize free energy (a balance between expected cost and entropy), subject to margin constraints corresponding to the predefined supply and demand distributions. As is common for such formulations, minimizing this objective function results in a Gibbs–Boltzmann probability mass on paths.

As discussed in more detail by Saerens et al. (2009), randomization from optimality can prove useful for several reasons, both in the context of transportation, as well as when measuring distance:

- If the environment is changing over time (nonstationary), the system could benefit from randomization by performing continual exploration.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

- A deterministic policy makes behavior totally predictable; on the contrary, randomness introduces unpredictability and therefore renders interception more difficult. Randomization has proved useful for this reason in game theory (see, e.g., Osborne, 2004).
- A randomized policy spreads the traffic on multiple paths, therefore reducing the danger of congestion.
- A distance measure accounting for all paths—and thus integrating the concept of high connectivity—can be more useful, e.g., in social network analysis, than relying on the best paths only.

1.2 Related work

The model proposed in this work builds on and extends previous work dedicated to the bag-of-paths (BoP) framework (Françoisse et al., 2017; Mantrach et al., 2010), as well as the *randomized-shortest-path* (RSP) framework (Akamatsu, 1996; Kivimäki et al., 2014; Kivimäki et al., 2016; Saerens et al., 2009; Yen et al., 2008) and their variants (Bavaud & Guex, 2012; Guex & Bavaud, 2015; Guex, 2016); see also Zhang et al. (2013) for a related proposition, called path integral.

The main motivation for using such models can be understood as follows (Lebichot & Saerens, 2018). Most of the traditional network measures are essentially based on two different paradigms about movement or communication occurring in the network: optimal communication based on shortest paths and random communication based on a random walk on the graph. For instance, shortest path distance as well as the standard betweenness centrality (Freeman, 1977) are defined from shortest paths, while resistance distance and random walk centrality (Brandes & Fleischer, 2005; Newman, 2005) are based on random walks (which have a strong analogy with electrical flow on the network; Doyle & Snell, 1984).

However, in reality, communication or movements over a network seldom occur either optimally or purely randomly. The BoP and RSP frameworks both relax these assumptions by interpolating between shortest paths and a pure random walk based on a temperature parameter. This enables the definition of measures with increased adaptability given by the temperature parameter. In addition to defining distances interpolating between the shortest path and resistance distances (Françoisse et al., 2017; Kivimäki et al., 2014; Yen et al., 2008), the models can also be used to define a centrality measure interpolating between a shortest path-based betweenness and the random walk betweenness (Kivimäki et al., 2016).

Besides the works mentioned above, other new families of distances have recently been developed integrating information on both the proximity (shortest-path distance) and amount of connectivity between nodes (captured, e.g., by the resistance distance) (Chebotarev, 2011, 2012, 2013; Fouss et al., 2016; Hashimoto et al., 2015; Herbster & Lever, 2009; Li et al., 2013; Lü & Zhou, 2011; Nguyen & Mamitsuka, 2016; Alamgir & von Luxburg, 2011). Many of these measures indeed interpolate (up to a constant scaling factor) between the shortest path distance (or length) and the resistance distance, therefore alleviating the previously mentioned lost-in-space effect.

A short discussion of the standard, deterministic, optimal transport on a graph problem appears in Section 4. Methods based on the optimal transport problem using entropic regularization have recently been investigated in a number of pattern recognition and machine learning tasks (e.g., Courty et al., 2017; Solomon et al., 2014). For instance, Cuturi (2013), Ferradans et al. (2014), and Guex et al. (2017) propose to regularize the standard objective function of the classical discrete optimal transport problem with an entropy term. They show on various problems, including image processing problems, that the resulting algorithm is much faster than the original one. Note that discrete entropy-regularized optimal transport problems were previously studied in economics, transportation science, and operations research (see, e.g., Wilson, 1970; Erlander & Stewart, 1990; Kapur, 1989; Kapur & Kesavan, 1992; Fang et al., 1997). The main difference with these previous contributions is that the present work defines the different quantities, such as entropy or cost, over *full paths* on the network by adopting a sum-over-paths formalism.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Finally, the hubness effect, mentioned earlier, has been studied recently in various works (Radovanović et al., 2010a,b; Suzuki et al., 2012, 2013; Tomasev et al., 2014; Hara et al., 2015). Hubness is a problem faced with high-dimensional data, e.g., when building nearest-neighbor graphs, as some nodes may become over-represented as hubs in such graphs due to concentration of distances in high-dimensional spaces. The weighting of distances provided by the margin-constrained BoP framework could help alleviate this effect in graph-based data analysis applications.

1.3 Main contributions

This work defines weighted distance measures between graph nodes by developing a *margin*-constrained BoP model. This model can be interpreted as a solution to the optimal transport problem on a graph involving a regularization term. The problem is tackled by using Kullback–Leibler divergence (also called relative entropy; Cover & Thomas, 2006) as regularization term. Furthermore, two types of paths are considered: regular paths and hitting, absorbing, paths.

The optimal randomized policy consists in the assignment of a probability distribution on the set of choices (deciding to follow an available edge) for each node of the network. It therefore defines (optimal) biased transition probabilities "attracting" the agents to the destination nodes. Furthermore, the model depends on a temperature parameter monitoring the balance between *exploitation* (expected cost) and *exploration* (entropy of paths) so that the solutions interpolate between the classical deterministic optimal transport solution (pure exploitation) and the random walk on the graph provided a priori by the user (pure exploration). Low temperatures correspond to (randomized) near-optimal solutions while high temperatures simply provide the predefined random walk behavior. Note that, when considering hitting paths, the model reduces to the standard RSP model when there is only one unique initial node and one unique destination node.

The first contribution consists in deriving the probability distribution over paths minimizing expected cost under relative entropy regularization and margin constraints, for both regular and hitting paths. Once the probability distribution over paths is derived, all the quantities of interest, such as

- the policy (optimal routing transition probabilities),
- the flow over the network based on the a priori starting and ending node distributions of paths,
- a weighted distance measure between nodes, and
- a dissimilarity between groups of nodes

can be defined and computed by simple matrix expressions.

Note that the present work is partly a re-interpretation of (Guex, 2016) in which the author already studied a similar optimal transport on a graph problem regularized by an entropic term. There, the entropic term at the node level was defined by considering, on each node, the relative entropy between the desired transition probabilities (the policy) and the reference transition probabilities corresponding to a natural random walk on the graph. Then, the global entropic regularization term was defined as a weighted sum of the relative entropies over all nodes. As by Saerens et al. (2009), the weighting factor is set to the expected number of visits to the node, therefore putting more emphasis on frequently visited nodes. In the current work, we adopt a path-based formalism and the entropic term is instead defined according to the relative entropy over *path distributions*.

Interestingly, it was found that the model derived by Guex (2016) is exactly equivalent to one of the two models introduced in this work, the one dealing with regular, non-hitting, paths, in the sense that they provide the same routing policy. Therefore, in comparison with (Guex, 2016), the present work reformulates the problem in terms of probabilities and relative entropy over paths

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

in the network, instead of transition probabilities on nodes. It also introduces another algorithm for solving the problem and it derives a new algorithm for dealing with hitting paths.

In short, the main contributions of this paper are

- the development of a new *margin-constrained bag-of-paths* (cBoP) framework, considering fixed probability distributions on starting and ending nodes,
- the introduction of a randomized solution to the optimal transport on a graph problem for both regular and hitting paths,
- the definition of a new distance measure between nodes and a dissimilarity between groups of nodes derived from this framework, and
- some illustrative simulations to explore the potential of the framework.

The remaining of the paper is as follows. Section 2 develops the formalism and derives the solution to the margin-constrained bag-of-paths problem on a graph for regular paths, while Section 3 extends the model to hitting paths. Then, Section 4 discusses the connections with the standard optimal transport on a graph problem. The derived distances are introduced in Section 5. Section 6 provides some illustrative simulations. Finally, Section 7 is the conclusion.

2. The margin-constrained bag-of-paths formalism

2.1 Background and notation

This subsection first sets the notation and terminology of the paper, after which the standard bag-of-paths (BoP) and randomized shortest-path (RSP) frameworks are briefly reviewed (note that a discussion of the standard optimal transport problem is deferred to Section 4). Then the margin-constrained BoP (cBoP) framework and the relevant related results are presented. Note that in this section, regular, non-hitting, paths are considered, whereas Section 3 restricts the set of paths to hitting, or absorbing, paths where the ending node may appear only once as final node.

Notation. In this paper, we always assume a weighted, strongly connected, directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with set of nodes $\mathcal{V} = \{1, 2, \ldots, n\}$ and set of edges $\mathcal{E} = \{(i, j)\}$ containing m edges in total. The nonnegative weights on edges, noted a_{ij}, represent local affinities between nodes and are contained in the *weighted adjacency matrix* $\mathbf{A}$. Edge weights define a natural *reference transition probabilities matrix* $\mathbf{P}^{\text{ref}}$ of a standard random walk on $\mathcal{G}$, with $\mathbf{P}^{\text{ref}} = \mathbf{D}^{-1}\mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix containing row sums of $\mathbf{A}$ (outdegrees). The Markov chain defined by these transition probabilities is assumed to be regular. Elementwise, we have

$$p_{ij}^{\text{ref}} = \frac{a_{ij}}{\sum_{k \in \mathcal{V}} a_{ik}}. \tag{1}$$

Along with weights, nonnegative *edges costs*, noted c_{ij}, are also provided. These costs are contained in the *cost matrix* $\mathbf{C}$ and can be defined either independently from weights a_{ij}, or, e.g., thanks to $c_{ij} = 1/a_{ij}$. We define a t-length *path* on the graph $\mathcal{G}$, denoted by $\wp$, as a sequence of nodes $\wp = (i_0, \ldots, i_t)$, where $t > 0$ and $(i_{\tau-1}, i_\tau) \in \mathcal{E}$ for all $\tau = 1, \ldots, t$. Note that a node can appear several times on the path (including the ending node). We denote a path starting in node i and ending in node j by $\wp_{ij}$. The *likelihood* $\tilde{\pi}^{\text{ref}}(\wp)$ of a t-length path $\wp = (i_0, \ldots, i_t)$ starting in i_0 and ending in i_t is defined by $\tilde{\pi}^{\text{ref}}(\wp) \triangleq \Pi_{\tau=1}^{t} p_{i_{\tau-1}, i_\tau}^{\text{ref}}$ and its *cost* $\tilde{c}(\wp)$ by $\tilde{c}(\wp) \triangleq \sum_{\tau=1}^{t} c_{i_{\tau-1}, i_\tau}$. We further denote respectively by $\mathcal{P}_{ij}$ and $\mathcal{P}$, the *set of paths starting in i and ending in j* and the *set of all paths* in $\mathcal{G}$, also named the *bag-of-paths*, with $\mathcal{P} = \cup_{i,j \in \mathcal{V}} \mathcal{P}_{ij}$. By convention, zero-length paths starting and ending in the same node with a zero cost are also included in the set of paths (see Françoisse et al., 2017 for details).

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

All vectors will be column vectors and denoted in lowercase bold while matrices are in uppercase bold.

The bag-of-paths and the randomized-shortest-path frameworks. The context defined above states the usual background of the *bag-of-paths* framework developed in (Françoisse et al., 2017; Mantrach et al., 2010). In these works, a probability distribution over the set of all paths, $\mathrm{P}(\wp)$ with $\wp \in \mathcal{P}$, was constructed in order to favor paths of low cost subject to a constant relative entropy constraint. It provides the probability of drawing a particular path $\wp$ from a bag, with replacement. The distribution can equivalently be obtained by solving the following problem (see Kivimäki et al., 2014)[1]:

$$\left|\begin{array}{ll} \underset{\{\mathrm{P}(\wp)\}_{\wp\in\mathcal{P}}}{\text{minimize}} & \mathrm{FE}(\mathrm{P}) = \sum_{\wp\in\mathcal{P}} \mathrm{P}(\wp)\tilde{c}(\wp) + T\sum_{\wp\in\mathcal{P}} \mathrm{P}(\wp)\log\left(\frac{\mathrm{P}(\wp)}{\mathrm{P}^{\mathrm{ref}}(\wp)}\right) \\ \text{subject to} & \sum_{\wp\in\mathcal{P}} \mathrm{P}(\wp) = 1, \end{array}\right. \tag{2}$$

where $T > 0$, the *temperature*, is a free parameter defining the relative entropy level (Cover & Thomas, 2006), and $\mathrm{P}^{\mathrm{ref}}(\wp)$ is the natural, reference, probability of a path $\wp$ depending on $\tilde{\pi}^{\mathrm{ref}}(\wp)$ (see Equation (1)) and to be discussed later. FE(P) is called the (relative) *free energy*, due to its similarity with the statistical physics quantity. It corresponds to the expected cost, or energy, to which the relative entropy weighted by temperature is added. Strictly speaking, a nonnegativity constraint should also be added to Equation (2), but this is not necessary since the resulting probability distribution will automatically be nonnegative. Indeed, following Jaynes (1957), Cover & Thomas (2006), Kapur & Kesavan (1992) and Françoisse et al. (2017), Kivimäki et al. (2014), Mantrach et al. (2010) for the paths formalism, the solution is a standard Gibbs–Boltzmann distribution of the form

$$\mathrm{P}(\wp) = \frac{\mathrm{P}^{\mathrm{ref}}(\wp)\exp\left[-\theta\tilde{c}(\wp)\right]}{\sum_{\wp'\in\mathcal{P}} \mathrm{P}^{\mathrm{ref}}(\wp')\exp\left[-\theta\tilde{c}(\wp')\right]},$$

where $\tilde{c}(\wp)$ is the total cumulated cost along path $\wp$. It provides the probability of choosing any path $\wp \in \mathcal{P}$ in the network.

In addition, the RSP framework (Saerens et al., 2009; Kivimäki et al., 2014; Yen et al., 2008) inspired by Akamatsu (1996), restricts the set of paths to hitting paths (see next section for details) connecting only two predefined nodes i and j. This defines optimal randomized policies for reaching node j from i, ranging from shortest paths to a random walk. A method for computing the RSP on large sparse graphs by restricting the set to paths with a finite predefined length was developed by Mantrach et al. (2011, Section 4).

2.2 Problem definition

In this Section, the BoP framework developed by Françoisse et al. (2017) and Mantrach et al. (2010) is extended into a *margin-constrained bag-of-paths* (cBoP) *framework* by introducing two additional density vectors on nodes, provided by the user: $\boldsymbol{\sigma}_{\mathrm{in}}$ and $\boldsymbol{\sigma}_{\mathrm{out}}$, with $\sum_{i\in\mathcal{V}} \sigma_i^{\mathrm{in}} = \sum_{i\in\mathcal{V}} \sigma_i^{\mathrm{out}} = 1$ and $\sigma_i^{\mathrm{in}}, \sigma_i^{\mathrm{out}} \geq 0, \forall i \in \mathcal{V}$. These vectors define desired *constraints on the distribution margins* of our bag-of-paths probabilities, i.e.,

$$\mathrm{P}(S = i) \triangleq \sum_{j\in\mathcal{V}} \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}(\wp_{ij}) = \sigma_i^{\mathrm{in}} \quad \forall i \in \mathcal{V}, \tag{3}$$

$$\mathrm{P}(E = j) \triangleq \sum_{i\in\mathcal{V}} \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}(\wp_{ij}) = \sigma_j^{\mathrm{out}} \quad \forall j \in \mathcal{V}, \tag{4}$$

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

where S and E denote random variables containing respectively the starting and the ending nodes of the drawn path. In turn, it means that we want to constrain the probability of picking a path in the BoP starting from i to value σ_i^{in} and the probability of picking a path ending in j to value σ_j^{out}. The intuition is as follows: the model assumes that we are carrying a unit of goods in the network from the set of nodes $\mathcal{I}n = \{i \in \mathcal{V} : \sigma_i^{\text{in}} > 0\}$ (supply nodes) to the set $\mathcal{O}ut = \{j \in \mathcal{V} : \sigma_j^{\text{out}} > 0\}$ (demand nodes) in an optimal way by minimizing a balance between expected cost and relative entropy of paths. A discussion of this model in the light of optimal transport on a graph appears later in Section 4.

Altogether, we extend problem (2) and seek the *optimal path probability distribution*, $\{\mathrm{P}(\wp)\}$, $\wp \in \mathcal{P}$, solving

$$\left|\begin{array}{ll} \underset{\{\mathrm{P}(\wp)\}_{\wp\in\mathcal{P}}}{\text{minimize}} & \mathrm{FE}(\mathrm{P}) = \sum_{\wp\in\mathcal{P}} \mathrm{P}(\wp)\tilde{c}(\wp) + T\sum_{\wp\in\mathcal{P}} \mathrm{P}(\wp)\log\left(\dfrac{\mathrm{P}(\wp)}{\mathrm{P}^{\text{ref}}(\wp)}\right) \\ \text{subject to} & \sum_{j\in\mathcal{V}}\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}(\wp_{ij}) = \sigma_i^{\text{in}} \qquad \forall i \in \mathcal{V} \\ & \sum_{i\in\mathcal{V}}\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}(\wp_{ij}) = \sigma_j^{\text{out}} \qquad \forall j \in \mathcal{V}. \end{array}\right. \tag{5}$$

Note that as we have $\sum_{i\in\mathcal{V}} \sigma_i^{\text{in}} = \sum_{i\in\mathcal{V}} \sigma_i^{\text{out}} = 1$, the constraint $\sum_{\wp\in\mathcal{P}} \mathrm{P}(\wp) = 1$ in Equation (2) can be dropped. The goal of this problem is to find a probability distribution with fixed margins, such that it favors the paths of least cost when $T \to 0$ and the paths with high likelihood $\tilde{\pi}^{\text{ref}}(\wp)$ when $T \to \infty$.

In order for the high temperature bounds to be consistent, $\mathrm{P}^{\text{ref}}(\wp)$ must be defined according to $\tilde{\pi}^{\text{ref}}(\wp)$. In the usual, unconstrained, BoP formalism with a uniform a priori probability of choosing the starting and ending node, the definition is simply $\mathrm{P}^{\text{ref}}(\wp) = \tilde{\pi}^{\text{ref}}(\wp)/\sum_{\wp'\in\mathcal{P}} \tilde{\pi}^{\text{ref}}(\wp')$ (Françoisse et al., 2017).[2] However, defining reference probabilities is not as trivial in the margin-constrained setting studied in this work because of the constraints. This definition is the goal of the next section (note that the reader mainly interested in the randomized optimal transport problem can simply assume that the reference transition probabilities as given, skip the Section 2.3 and go directly to Section 2.4). The solution of problem (5) is then stated and proved in Section 2.4.

2.3 Reference probabilities with fixed margins

The reference probability of a path, $\mathrm{P}^{\text{ref}}(\wp)$, should have appropriate margins in order to ensure the convergence $\mathrm{P}^{\star}(\wp) \to \mathrm{P}^{\text{ref}}(\wp)$ when $T \to \infty$ (pure random walk). In other words, for consistency, the reference probabilities of paths should also satisfy the constraints,

$$\mathrm{P}^{\text{ref}}(S = i) = \sum_{j\in\mathcal{V}}\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}^{\text{ref}}(\wp_{ij}) = \sigma_i^{\text{in}} \quad \forall i \in \mathcal{V}, \tag{6}$$

$$\mathrm{P}^{\text{ref}}(E = j) = \sum_{i\in\mathcal{V}}\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}^{\text{ref}}(\wp_{ij}) = \sigma_j^{\text{out}} \quad \forall j \in \mathcal{V}, \tag{7}$$

which further implies that the path probabilities $\mathrm{P}^{\text{ref}}(\wp_{ij})$ sum to 1.

In (Françoisse et al., 2017), the reference distribution of a path is simply set proportional to path likelihood $\tilde{\pi}^{\text{ref}}(\wp)$, which can be interpreted as follows: the starting distribution $\mathrm{P}^{\text{ref}}(S = i)$ is defined as uniform and the ending distribution $\mathrm{P}^{\text{ref}}(E = j)$ is equal to the proportion of time spent in each node for the Markov chain S_t, defined by transition probabilities matrix $\mathbf{P}^{\text{ref}}$, when $t \to \infty$ (stationarity). Obviously, defining the reference probability in a similar way here would lead to a problem: the ending distribution $\mathrm{P}^{\text{ref}}(E = j)$ depends entirely on the transition matrix of the Markov chain defined by $\mathbf{P}^{\text{ref}}$ and will generally not yield the desired distribution $\boldsymbol{\sigma}_{\text{out}}$.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

To address this problem, we introduce a new *killed* Markov process that will lead to the desired ending distribution, while being as similar as possible, in a certain sense, to the original chain. More precisely, we design this killed Markov process in such a way that the random walker encounters exactly the same probabilities of jumping to any adjacent node as the original transition probabilities ($\mathbf{P}^{\text{ref}}$) as long as he survives.

2.3.1 A particular killed Markov process

Let us first define a killed random process which will be helpful later.

Definition 2.1. *From the reference, regular, Markov process S_t defined on the state space $\mathcal{V}$, with initial distribution $\boldsymbol{\pi}_{\text{in}}$ and transition matrix $\mathbf{P}^{\text{ref}}$, a **killed Markov process**, denoted by K_t, is defined as a new process with the same initial distribution and following a **substochastic** transition matrix $\widehat{\mathbf{P}}^{\text{ref}}$, given by*

$$\widehat{\mathbf{P}}^{\text{ref}} \triangleq (\mathbf{I} - \mathbf{Diag}(\boldsymbol{\alpha}))\mathbf{P}^{\text{ref}}, \tag{8}$$

*where the **probability to be killed** after visiting node i is α_i, with $0 \leq \alpha_i \leq 1\ \forall i \in \mathcal{V}$. $\mathbf{Diag}(\boldsymbol{\alpha})$ is the diagonal matrix containing vector $\boldsymbol{\alpha}$ on its diagonal. In other words, the vector $\boldsymbol{\alpha} = (\alpha_i)$ contains the **killing rate** of each node. This killed Markov process can be seen as adding a new (virtual) absorbing state, the **cemetery** ω, and following the rules*

$$\mathrm{P}(K_0 = i) = \pi_i^{\text{in}}, \tag{9}$$

$$\mathrm{P}(K_{t+1} = j | K_t = i) = \widehat{p}_{ij}^{\text{ref}} = (1 - \alpha_i) p_{ij}^{\text{ref}}, \tag{10}$$

$$\mathrm{P}(K_{t+1} = \omega | K_t = i) = 1 - \widehat{p}_{i\bullet}^{\text{ref}} \triangleq \alpha_i, \tag{11}$$

$$\mathrm{P}(K_{t+1} = \omega | K_t = \omega) = 1, \tag{12}$$

where the dot in $\widehat{p}_{i\bullet}^{\text{ref}}$ means summation over the second index (over the set of nodes $\mathcal{V}$).

We observe that $\widehat{p}_{ij}^{\text{ref}} / \widehat{p}_{i\bullet}^{\text{ref}} = p_{ij}^{\text{ref}}$ for any $\boldsymbol{\alpha}$. This means that this killing process will behave similarly to the original process as long as it survives, thus arguing in favor of the similarity requirement between the two chains. However, unlike the reference Markov chain, this killed Markov process possesses an "*ending distribution*".

Definition 2.2. *A killed Markov process, as defined in Definition 2.1, possesses an **ending distribution**, denoted by $\boldsymbol{\pi}_{\text{out}} = (\pi_i^{\text{out}})$ and given by*

$$\pi_i^{\text{out}} \triangleq \mathrm{P}(K_{M-1} = i) = \sum_{t=0}^{\infty} \mathrm{P}(K_t = i, K_{t+1} = \omega), \tag{13}$$

where M is the random variable containing the time where the process is killed (it reaches the cemetery state ω).

This quantity denotes the probability of being killed in state i when starting from initial nodes with probabilities π_i^{in}: it sums up the probability of being killed after $0, 1, 2, \ldots$ steps.

Interestingly, it is possible to find the vector of killing rates $\boldsymbol{\alpha}$ corresponding to a desired ending distribution $\boldsymbol{\pi}_{\text{out}} = \boldsymbol{\sigma}_{\text{out}}$. This is important as it will allow us to design a proper reference probability distribution satisfying the predefined margins. But, we first need the following preliminary lemma.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Lemma 2.1. *The expected number of visits to j before being killed, given that the process started from state i, that is, the quantity* $\bar{n}_{ij}^{\text{ref}} = \sum_{t=0}^{\infty} \mathrm{P}(K_t = j | K_0 = i)$*, is given by element i, j of matrix* $\widehat{\mathbf{Z}}^{\text{ref}} \triangleq \sum_{t=0}^{\infty} (\widehat{\mathbf{P}}^{\text{ref}})^t = (\mathbf{I} - \widehat{\mathbf{P}}^{\text{ref}})^{-1}$ *which is well-defined for a substochastic matrix and a strongly connected graph. In other words,* $\bar{n}_{ij}^{\text{ref}} = \widehat{z}_{ij}^{\text{ref}}$.

Proof. See (Fouss et al., 2016, Section 1.5.7). □

Then, the killing rates and the ending distribution are related by the following proposition. This will allow us to determine the killing rates in order to satisfy a predefined ending distribution.

Proposition 2.1. *For a killed Markov process on a strongly connected graph (see Definition 2.1), the following equality is satisfied*

$$\boldsymbol{\alpha} = \boldsymbol{\pi}_{\text{out}} \div \bar{\mathbf{n}}^{\text{ref}}, \tag{14}$$

where $\div$ *is the elementwise division and column vector* $\bar{\mathbf{n}}^{\text{ref}} = (\bar{n}_j^{\text{ref}})$ *with* $\bar{n}_j^{\text{ref}} = \sum_{i \in \mathcal{V}} \pi_i^{\text{in}} \bar{n}_{ij}^{\text{ref}}$*, holds the expected number of visits to state j before being killed which can be computed thanks to*

$$\left(\mathbf{I} - \left(\mathbf{P}^{\text{ref}}\right)^{\top}\right)\bar{\mathbf{n}}^{\text{ref}} = \boldsymbol{\pi}_{\text{in}} - \left(\mathbf{P}^{\text{ref}}\right)^{\top}\boldsymbol{\pi}_{\text{out}}, \tag{15}$$

where $\mathbf{P}^{\text{ref}}$ *is the transition probability matrix defined in Equation (1).*

Proof. First, let us observe that the joint distribution for the starting node and the ending node of the killing process is

$$\begin{aligned}
\mathrm{P}(K_0 = i, K_{M-1} = j) &= \sum_{t=0}^{\infty} \mathrm{P}(K_0 = i, K_t = j, K_{t+1} = \omega) \\
&= \sum_{t=0}^{\infty} \mathrm{P}(K_{t+1} = \omega | K_t = j)\mathrm{P}(K_t = j | K_0 = i)\mathrm{P}(K_0 = i) \\
&= \pi_i^{\text{in}} \left[\sum_{t=0}^{\infty} \left(\widehat{\mathbf{P}}^{\text{ref}}\right)^t\right]_{ij} \alpha_j = \pi_i^{\text{in}} \left[\widehat{\mathbf{Z}}^{\text{ref}}\right]_{ij} \alpha_j \\
&= \pi_i^{\text{in}} \widehat{z}_{ij}^{\text{ref}} \alpha_j,
\end{aligned} \tag{16}$$

where $\widehat{\mathbf{Z}}^{\text{ref}}$ is defined by Lemma 2.1. Thus, the ending distribution (see Definition 2.2) reads

$$\pi_j^{\text{out}} = \sum_{i \in \mathcal{V}} \mathrm{P}(K_0 = i, K_{M-1} = j) = \sum_{i \in \mathcal{V}} \pi_i^{\text{in}} \widehat{z}_{ij}^{\text{ref}} \alpha_j. \tag{17}$$

We know from Lemma 2.1 that $\widehat{z}_{ij}^{\text{ref}}$ can be interpreted as the expected number of times the killed Markov process visits node j when starting from i. So, if we define

$$\bar{n}_j^{\text{ref}} \triangleq \sum_{i \in \mathcal{V}} \pi_i^{\text{in}} \widehat{z}_{ij}^{\text{ref}} = \sum_{i \in \mathcal{V}} \pi_i^{\text{in}} \bar{n}_{ij}^{\text{ref}}, \tag{18}$$

the column vector $\bar{\mathbf{n}}^{\text{ref}} = (\bar{n}_j^{\text{ref}})$ holds the expected number of times the process is in j before being killed. Then, Equation (17) directly provides

$$\boldsymbol{\alpha} = \boldsymbol{\pi}_{\text{out}} \div \bar{\mathbf{n}}^{\text{ref}}. \tag{19}$$

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

The second part of the proposition is obtained by starting from the definition of $\bar{\mathbf{n}}^{\text{ref}}$ and following Equation (8),

$$\begin{aligned} & \bar{\mathbf{n}}^{\text{ref}} = \left(\widehat{\mathbf{Z}}^{\text{ref}}\right)^{\top} \boldsymbol{\pi}_{\text{in}} = \left(\mathbf{I} - \left(\widehat{\mathbf{P}}^{\text{ref}}\right)^{\top}\right)^{-1} \boldsymbol{\pi}_{\text{in}}, \\ \Rightarrow \quad & \bar{\mathbf{n}}^{\text{ref}} - \left(\widehat{\mathbf{P}}^{\text{ref}}\right)^{\top} \bar{\mathbf{n}}^{\text{ref}} = \boldsymbol{\pi}_{\text{in}}, \\ \Rightarrow \quad & \bar{\mathbf{n}}^{\text{ref}} - \left(\left(\mathbf{P}^{\text{ref}}\right)^{\top} - \left(\mathbf{P}^{\text{ref}}\right)^{\top} \mathbf{Diag}(\boldsymbol{\alpha})\right) \bar{\mathbf{n}}^{\text{ref}} = \boldsymbol{\pi}_{\text{in}}. \end{aligned}$$

Then, using $\mathbf{Diag}(\boldsymbol{\alpha})\, \bar{\mathbf{n}}^{\text{ref}} = \boldsymbol{\pi}_{\text{out}}$ (see Equation (14)),

$$\left(\mathbf{I} - \left(\mathbf{P}^{\text{ref}}\right)^{\top}\right) \bar{\mathbf{n}}^{\text{ref}} = \boldsymbol{\pi}_{\text{in}} - \left(\mathbf{P}^{\text{ref}}\right)^{\top} \boldsymbol{\pi}_{\text{out}},$$

which provides the same results as the expression derived by Guex (2016, proposition 1) from another perspective. □

Equation (14) simply states that the probability of being killed in state i is equal to the expected number of visits to i times the probability of jumping to the cemetery state from i, $\pi_i^{\text{out}} = \bar{n}_i^{\text{ref}}\, \alpha_i$. This is similar to the computation of the absorption probabilities when starting from a transient state in an absorbing Markov chain (Doyle & Snell, 1984; Grinstead & Snell, 1997).

In conclusion, it is possible to determine $\bar{\mathbf{n}}^{\text{ref}}$ and then $\boldsymbol{\alpha}$ through Proposition 2.1 by considering an additional free parameter. Indeed, as $(\mathbf{I} - (\mathbf{P}^{\text{ref}})^{\top})$ is rank-deficient (its rank is $n-1$, as the initial reference chain is regular), we have (see, e.g., Graybill, 1983)

$$\bar{\mathbf{n}}^{\text{ref}} = \left(\mathbf{I} - \left(\mathbf{P}^{\text{ref}}\right)^{\top}\right)^{+} \left(\boldsymbol{\pi}_{\text{in}} - \left(\mathbf{P}^{\text{ref}}\right)^{\top} \boldsymbol{\pi}_{\text{out}}\right) + \epsilon\, \boldsymbol{\pi}_{\text{ref}}, \tag{20}$$

where $\left(\mathbf{I} - (\mathbf{P}^{\text{ref}})^{\top}\right)^{+}$ denotes the Moore–Penrose pseudoinverse, $\boldsymbol{\pi}_{\text{ref}}$ is the stationary distribution of the regular, reference, Markov chain defined by $\mathbf{P}^{\text{ref}}$ (i.e., a vector summing to 1 and generating the null-space of $(\mathbf{I} - (\mathbf{P}^{\text{ref}})^{\top})$), and ϵ is an additional free parameter, named *persistence*, such that

$$\epsilon \geq \max_i \left(\frac{\pi_i^{\text{out}} - \mathbf{e}_i^{\top} \left(\mathbf{I} - \left(\mathbf{P}^{\text{ref}}\right)^{\top}\right)^{+} \left(\boldsymbol{\pi}_{\text{in}} - \left(\mathbf{P}^{\text{ref}}\right)^{\top} \boldsymbol{\pi}_{\text{out}}\right)}{\pi_i^{\text{ref}}} \right).$$

This last inequality ensures that $\bar{n}_i^{\text{ref}} \geq \pi_i^{\text{out}}$ and thus $\alpha_i \leq 1$, $\forall i \in \mathcal{V}$. Intuitively, the persistence parameter reflects the difficulty for a process to be killed in nodes where $\alpha_i > 0$ (called killing nodes) and thus affects the expected length of the paths (see Guex, 2016, for a discussion). In (Guex, 2016), the persistence is shown to have an electrical interpretation within the well-known analogy between random-walk models and electrical models (Doyle & Snell, 1984): it corresponds to the lowest electrical potential that can be defined on nodes. However, the effect of the persistence on the behavior of the model is beyond the scope of this work, and we will set it to its lower bound in our simulations.

2.3.2 Defining the reference probabilities over paths from the killed process

We will now define the reference probabilities $\mathrm{P}^{\text{ref}}(\wp)$ over the set of paths, with the help of a killed Markov process, in order to obtain the desired starting and ending distributions $\boldsymbol{\sigma}_{\text{in}}$ and $\boldsymbol{\sigma}_{\text{out}}$, i.e., $\sum_{j \in \mathcal{V}} \sum_{\wp_{ij} \in \mathcal{P}_{ij}} \mathrm{P}^{\text{ref}}(\wp_{ij}) = \sigma_i^{\text{in}}$ and $\sum_{i \in \mathcal{V}} \sum_{\wp_{ij} \in \mathcal{P}_{ij}} \mathrm{P}^{\text{ref}}(\wp_{ij}) = \sigma_j^{\text{out}}$. We show that it suffices to set $\boldsymbol{\pi}_{\text{in}} = \boldsymbol{\sigma}_{\text{in}}$ and $\boldsymbol{\pi}_{\text{out}} = \boldsymbol{\sigma}_{\text{out}}$ in the previously defined killed Markov process and deduce the reference probabilities from it.

Proposition 2.2. *If the reference probabilities* $\mathrm{P}^{\text{ref}}(\wp)$, $\wp \in \mathcal{P}$, *are set to*

$$\mathrm{P}^{\text{ref}}(\wp_{ij}) \triangleq \sigma_i^{\text{in}} \widehat{\pi}^{\text{ref}}(\wp_{ij})\, \alpha_j \text{ for each } i, j \in \mathcal{V}, \tag{21}$$

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

where $\widehat{\pi}^{\mathrm{ref}}(\wp_{ij})$ *is the modified likelihood of the path* $\wp_{ij}$*, defined by*

$$\widehat{\pi}^{\mathrm{ref}}(\wp) \triangleq \prod_{\tau=1}^{t} \widehat{p}^{\mathrm{ref}}_{k_{\tau-1}k_\tau} \text{ for any path } \wp = (k_0, k_1, \ldots, k_t), \tag{22}$$

with $\widehat{\mathbf{P}}^{\mathrm{ref}} = (\mathbf{I} - \mathbf{Diag}(\boldsymbol{\alpha}))\mathbf{P}^{\mathrm{ref}}$*, and where* $\boldsymbol{\alpha}$ *is computed from Proposition 2.1 (with* $\boldsymbol{\pi}_{\mathrm{in}} = \boldsymbol{\sigma}_{\mathrm{in}}$ *and* $\boldsymbol{\pi}_{\mathrm{out}} = \boldsymbol{\sigma}_{\mathrm{out}}$*), then we have*

$$\mathrm{P}^{\mathrm{ref}}(S = i) = \sum_{j\in\mathcal{V}} \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}^{\mathrm{ref}}(\wp_{ij}) = \sigma_i^{\mathrm{in}} \quad \forall i \in \mathcal{V},$$

$$\mathrm{P}^{\mathrm{ref}}(E = j) = \sum_{i\in\mathcal{V}} \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}^{\mathrm{ref}}(\wp_{ij}) = \sigma_j^{\mathrm{out}} \quad \forall j \in \mathcal{V}.$$

Moreover, these reference probabilities are properly scaled as they sum to 1.

Proof. Observe that if we set the reference probabilities to $\mathrm{P}^{\mathrm{ref}}(\wp_{ij}) \triangleq \pi_i^{\mathrm{in}} \widehat{\pi}^{\mathrm{ref}}(\wp_{ij})\, \alpha_j$, then, from Equation (16), we have for the probability of picking paths $\wp_{ij} \in \mathcal{P}_{ij}$ in the bag of all paths starting in i $(S = i)$ and ending in j $(E = j)$ (see Françoisse et al., 2017)

$$\begin{aligned}\mathrm{P}(S = i, E = j) &= \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}^{\mathrm{ref}}(\wp_{ij}) = \pi_i^{\mathrm{in}} \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \widehat{\pi}^{\mathrm{ref}}(\wp_{ij})\, \alpha_j \\ &= \pi_i^{\mathrm{in}} \left[\sum_{t=0}^{\infty} (\widehat{\mathbf{P}}^{\mathrm{ref}})^t\right]_{ij} \alpha_j = \mathrm{P}(K_0 = i, K_{M-1} = j),\end{aligned}$$

and therefore,

$$\mathrm{P}(S = i) = \mathrm{P}(K_0 = i) = \pi_i^{\mathrm{in}},$$
$$\mathrm{P}(E = j) = \mathrm{P}(K_{M-1} = j) = \pi_j^{\mathrm{out}}.$$

Moreover, it also follows that $\sum_{\wp\in\mathcal{P}} \mathrm{P}^{\mathrm{ref}}(\wp) = \sum_{i,j=1}^{n} \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}^{\mathrm{ref}}(\wp_{ij}) = \sum_{i,j=1}^{n} \mathrm{P}(K_0 = i, K_{M-1} = j) = 1$ so that the reference probability distribution over paths is properly scaled. □

Note that the last quantity, $\mathrm{P}(S = i, E = j)$, was called the bag-of-paths probability matrix and played a key role in the bag-of-paths framework (see Françoisse et al., 2017, for details). It is also called the coupling matrix in optimal transportation (see later). The procedure for computing the vectors $\boldsymbol{\alpha}$ and $\bar{\mathbf{n}}^{\mathrm{ref}}$ allowing to obtain a desired ending distribution is summarized in Algorithm 1. In addition to the predefined margins, it takes as input a persistence gap parameter ϵ_{g} indicating to which extent persistence of flow is present in the network (see the discussion following Equation (20)).

2.4 Computation of the optimal probability distribution over paths

In this section, now that we have found a proper reference distribution, we focus on the computation of the optimal probability distribution $\mathrm{P}^{\star}(\wp)$ solving problem (5). This solution is obtained through its Lagrange parameter vectors, which can be obtained from the constraints.

2.4.1 The optimal path probabilities

The optimal probability distribution is obtained by the following proposition.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Algorithm 1: Killing rates for non-hitting paths

Input:
- The $n \times n$ nonnegative adjacency matrix $\mathbf{A}$ of a strongly connected directed graph, containing edge affinities.
- An input distribution vector $\boldsymbol{\sigma}_{\text{in}}$ of size n.
- An output distribution vector $\boldsymbol{\sigma}_{\text{out}}$ of size n.
- A scalar persistence gap parameter $\epsilon_{\text{g}} > 0$.

Output:
- The killing rates vector $\boldsymbol{\alpha}$ of size n.
- The expected number of visits to each node for the reference random walk, $\bar{\mathbf{n}}^{\text{ref}}$.

1. $\mathbf{D} \leftarrow \mathbf{Diag}(\mathbf{Ae})$ ▷ the out-degree matrix
2. $\mathbf{P}^{\text{ref}} \leftarrow \mathbf{D}^{-1}\mathbf{A}$ ▷ the reference transition probabilities matrix
3. $\mathbf{Q} \leftarrow \mathbf{I} - (\mathbf{P}^{\text{ref}})^{\top}$ ▷ the auxiliary matrix $\mathbf{Q}$
4. $\boldsymbol{\pi} \leftarrow \mathbf{null}(\mathbf{Q})$ ▷ computes the stationary distribution of the Markov chain whose transition matrix is $\mathbf{P}^{\text{ref}}$ ($\boldsymbol{\pi}$ spans the null space of $\mathbf{Q}$)
5. $\mathbf{Q}^{+} \leftarrow \text{pinv}(\mathbf{Q})$ ▷ the Moore–Penrose pseudoinverse matrix of $\mathbf{Q}$
6. $\bar{\mathbf{n}}_0^{\text{ref}} \leftarrow \mathbf{Q}^{+}(\boldsymbol{\sigma}_{\text{in}} - (\mathbf{P}^{\text{ref}})^{\top}\boldsymbol{\sigma}_{\text{out}})$ ▷ the initial expected number of visits to each node for the reference random walk
7. $\epsilon \leftarrow \max\left((\boldsymbol{\sigma}_{\text{out}} - \bar{\mathbf{n}}_0^{\text{ref}}) \div \boldsymbol{\pi}\right) + \epsilon_{\text{g}}$ ▷ the value of the persistence parameter
8. $\bar{\mathbf{n}}^{\text{ref}} \leftarrow \bar{\mathbf{n}}_0^{\text{ref}} + \epsilon\boldsymbol{\pi}$ ▷ the expected number of visits to each node for the reference random walk
9. $\boldsymbol{\alpha} \leftarrow \boldsymbol{\sigma}_{\text{out}} \div \bar{\mathbf{n}}^{\text{ref}}$ ▷ the killing rates vector; $\div$ is the elementwise division
10. **return** $\boldsymbol{\alpha}, \bar{\mathbf{n}}^{\text{ref}}$

Proposition 2.3. *For reference path probabilities* $\mathrm{P}^{\text{ref}}(\wp)$ *as defined in Proposition 2.2, the minimization problem* (5)*, as stated in Section 2.2, is solved by*

$$\mathrm{P}^{\star}(\wp_{ij}) = \mu_i^{\text{in}}\sigma_i^{\text{in}}\mu_j^{\text{out}}\alpha_j\widehat{\pi}^{\text{ref}}(\wp_{ij})\exp\left(-\beta\tilde{c}(\wp_{ij})\right) \text{ for each } i, j \in \mathcal{V}, \tag{23}$$

where $\beta \triangleq 1/T$ *is the inverse temperature parameter, and* $\mu_i^{\text{in}} \triangleq \exp\left(-\beta\lambda_i^{\text{in}}\right)$, $\mu_j^{\text{out}} \triangleq \exp\left(-\beta\lambda_j^{\text{out}}\right)$ *are two vectors derived from the Lagrange parameter vectors* $\boldsymbol{\lambda}_{\text{in}}$, $\boldsymbol{\lambda}_{\text{out}}$, *associated with the constraints.*

Proof. We derive the solution for the optimal probability distribution $\mathrm{P}^{\star}(\wp)$ solving problem (5). By introducing Lagrange parameter vectors $\boldsymbol{\lambda}_{\text{in}}$ and $\boldsymbol{\lambda}_{\text{out}}$, the Lagrange function associated with Equation (5) is

$$\begin{aligned}\mathcal{L}(\mathrm{P}, \boldsymbol{\lambda}_{\text{in}}, \boldsymbol{\lambda}_{\text{out}}) \triangleq & \sum_{\wp\in\mathcal{P}} \mathrm{P}(\wp)\tilde{c}(\wp) + T\sum_{\wp\in\mathcal{P}} \mathrm{P}(\wp)\log\left(\frac{\mathrm{P}(\wp)}{\mathrm{P}^{\text{ref}}(\wp)}\right) \\ & + \sum_{i\in\mathcal{V}} (\lambda_i^{\text{in}} - \tfrac{T}{2})\left[\sum_{j\in\mathcal{V}}\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}(\wp_{ij}) - \sigma_i^{\text{in}}\right] \\ & + \sum_{j\in\mathcal{V}} (\lambda_j^{\text{out}} - \tfrac{T}{2})\left[\sum_{i\in\mathcal{V}}\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}(\wp_{ij}) - \sigma_j^{\text{out}}\right],\end{aligned} \tag{24}$$

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

where the Lagrange parameters are shifted by $T/2$ to simplify the notation. Taking its partial derivative with respect to $\mathrm{P}(\wp_{ij})$, setting the result to zero, and defining the *inverse temperature* $\beta \triangleq 1/T$, provides

$$\begin{aligned}\mathrm{P}^\star(\wp_{ij}) &= \mathrm{P}^{\mathrm{ref}}(\wp_{ij}) \exp\left[-\beta\left(\tilde{c}(\wp_{ij}) + \lambda_i^{\mathrm{in}} + \lambda_j^{\mathrm{out}}\right)\right] \\ &= \mu_i^{\mathrm{in}}\mu_j^{\mathrm{out}}\mathrm{P}^{\mathrm{ref}}(\wp_{ij}) \exp\left(-\beta\tilde{c}(\wp_{ij})\right),\end{aligned}$$

where we defined

$$\mu_i^{\mathrm{in}} \triangleq \exp\left(-\beta\lambda_i^{\mathrm{in}}\right) \text{ and } \mu_i^{\mathrm{out}} \triangleq \exp\left(-\beta\lambda_i^{\mathrm{out}}\right), \tag{25}$$

which corresponds to a re-parametrization of the the Lagrange parameters that will be used instead of the original parameters. By inserting the reference probability found in Equation (21) in this last equation, we get the following form for the path probabilities

$$\mathrm{P}^\star(\wp_{ij}) = \mu_i^{\mathrm{in}}\sigma_i^{\mathrm{in}}\mu_j^{\mathrm{out}}\alpha_j\widehat{\pi}^{\mathrm{ref}}(\wp_{ij}) \exp\left(-\beta\tilde{c}(\wp_{ij})\right),$$

where the $\widehat{\pi}^{\mathrm{ref}}(\wp)$ are provided by Proposition 2.2. □

2.4.2 Computing the Lagrange parameters

The solution (23) requires the values of the Lagrange multipliers $\boldsymbol{\lambda}_{\mathrm{in}}$ and $\boldsymbol{\lambda}_{\mathrm{out}}$, or alternatively the vectors $\boldsymbol{\mu}_{\mathrm{in}}$ and $\boldsymbol{\mu}_{\mathrm{out}}$, which can be obtained from the equality constraints. Proposition 2.4 shows how to compute these vectors.

Proposition 2.4. *The two vectors, defined as* $\mu_i^{\mathrm{in}} = \exp\left(-\beta\lambda_i^{\mathrm{in}}\right)$ *and* $\mu_i^{\mathrm{out}} = \exp\left(-\beta\lambda_i^{\mathrm{out}}\right)$ *(see Proposition 2.2), verify the following equations*

$$\boldsymbol{\mu}_{\mathrm{in}} = \mathbf{e} \div \left(\widehat{\mathbf{Z}}\left(\boldsymbol{\mu}_{\mathrm{out}} \circ \boldsymbol{\alpha}\right)\right), \tag{26}$$

$$\boldsymbol{\mu}_{\mathrm{out}} = \bar{\mathbf{n}}^{\mathrm{ref}} \div \left(\widehat{\mathbf{Z}}^\top\left(\boldsymbol{\mu}_{\mathrm{in}} \circ \boldsymbol{\sigma}_{\mathrm{in}}\right)\right), \tag{27}$$

where $\widehat{\mathbf{Z}} \triangleq (\mathbf{I} - \widehat{\mathbf{W}})^{-1}$ *is the* ***fundamental matrix****, obtained from* $\widehat{\mathbf{W}} \triangleq \widehat{\mathbf{P}}^{\mathrm{ref}} \circ \exp\left[-\beta\mathbf{C}\right]$*, and* $\circ$ *is the elementwise product.*

Proof. The Lagrange parameters can be found by enforcing the constraints (3) and (4) on $\mathrm{P}^\star(\wp)$. By injecting Equation (23) for $\mathrm{P}^\star(\wp)$ in Equations (3) and (4) provides

$$\sigma_i^{\mathrm{in}}\mu_i^{\mathrm{in}} \sum_{j\in\mathcal{V}} \mu_j^{\mathrm{out}}\alpha_j \underbrace{\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \widehat{\pi}^{\mathrm{ref}}(\wp_{ij}) \exp\left(-\beta\tilde{c}(\wp_{ij})\right)}_{\widehat{z}_{ij}} = \sigma_i^{\mathrm{in}} \quad \forall i \in \mathcal{V}, \tag{28}$$

$$\alpha_j\mu_j^{\mathrm{out}} \sum_{i\in\mathcal{V}} \mu_i^{\mathrm{in}}\sigma_i^{\mathrm{in}} \underbrace{\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \widehat{\pi}^{\mathrm{ref}}(\wp_{ij}) \exp\left(-\beta\tilde{c}(\wp_{ij})\right)}_{\widehat{z}_{ij}} = \sigma_j^{\mathrm{out}} \quad \forall j \in \mathcal{V}. \tag{29}$$

By further defining the *fundamental matrix* as

$$\widehat{\mathbf{Z}} \triangleq (\mathbf{I} - \widehat{\mathbf{W}})^{-1} \tag{30}$$

with $\widehat{\mathbf{W}} \triangleq \widehat{\mathbf{P}}^{\mathrm{ref}} \circ \exp\left[-\beta\mathbf{C}\right]$, where $\circ$ is the elementwise product, it can easily be shown by using a development similar to (Françoisse et al., 2017; Mantrach et al., 2010) that

$$\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \widehat{\pi}^{\mathrm{ref}}(\wp_{ij}) \exp\left(-\beta\tilde{c}(\wp_{ij})\right) = \widehat{z}_{ij} = [\widehat{\mathbf{Z}}]_{ij}. \tag{31}$$

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

For computing the parameters, we use a variant (Kapur & Kesavan, 1992) of iterative proportional fitting procedure (discussed below) based on Equations (28) and (29). Isolating σ_i^{in} and σ_i^{out} in Equations (28) and (29) after replacing α_j by the result found with Proposition 2.1 (Equations (14) and (20)), i.e., $\alpha_j = \sigma_j^{\text{out}} / \bar{n}_j^{\text{ref}}$, in the second equation, we obtain

$$\boldsymbol{\mu}_{\text{in}} = \mathbf{e} \div \left(\widehat{\mathbf{Z}} \left(\boldsymbol{\mu}_{\text{out}} \circ \boldsymbol{\alpha} \right) \right),$$
$$\boldsymbol{\mu}_{\text{out}} = \bar{\mathbf{n}}^{\text{ref}} \div \left(\widehat{\mathbf{Z}}^{\top} \left(\boldsymbol{\mu}_{\text{in}} \circ \boldsymbol{\sigma}_{\text{in}} \right) \right).$$

Note that the above derivation is only valid for nodes i and j for which $\sigma_i^{\text{in}} > 0$ and $\sigma_j^{\text{out}} > 0$, respectively, and that μ_i^{in} and μ_j^{out} are not needed when these quantities are equal to zero because the path probabilities also vanish in this case (see Equations (23)). However, defining and computing the quantities μ_i^{in} and μ_j^{out} for all nodes i and j according to Equations (26) and (27) prove to be convenient for what follows, as these quantities appear in other meaningful expressions. For unconstrained nodes with $\sigma_i^{\text{in}} = 0$ or $\sigma_j^{\text{out}} = 0$, the Lagrange parameters are equal to zero, meaning that the corresponding $\mu_i^{\text{in}} = 1$ and $\mu_j^{\text{out}} = 1$ □

Note that by further defining the matrix $\widehat{t}_{ij} \triangleq \sigma_i^{\text{in}} \widehat{z}_{ij} \alpha_j$, Equations (28) and (29) can be rewritten in matrix form as $\mathbf{Diag}(\boldsymbol{\mu}_{\text{in}})(\mathbf{Diag}(\boldsymbol{\sigma}_{\text{in}})\widehat{\mathbf{Z}}\,\mathbf{Diag}(\boldsymbol{\alpha}))\mathbf{Diag}(\boldsymbol{\mu}_{\text{out}})\,\mathbf{e} = \mathbf{Diag}(\boldsymbol{\mu}_{\text{in}})\widehat{\mathbf{T}}\,\mathbf{Diag}(\boldsymbol{\mu}_{\text{out}})\,\mathbf{e} = \boldsymbol{\sigma}_{\text{in}}$ and $\mathbf{Diag}(\boldsymbol{\mu}_{\text{out}})\widehat{\mathbf{T}}^{\top}\mathbf{Diag}(\boldsymbol{\mu}_{\text{in}})\,\mathbf{e} = \boldsymbol{\sigma}_{\text{out}}$, where $\mathbf{Diag}(\mathbf{x})$ is a diagonal matrix with $\mathbf{x}$ on its main diagonal and $\mathbf{e}$ is a column vector full of 1's. Therefore, we have in matrix form

$$\widehat{\mathbf{T}} = \mathbf{Diag}(\boldsymbol{\sigma}_{\text{in}})\widehat{\mathbf{Z}}\,\mathbf{Diag}(\boldsymbol{\alpha}). \tag{32}$$

Thus, the computation of the Lagrange parameters reduces to the problem of finding two nonnegative row and column scaling vectors ($\boldsymbol{\mu}_{\text{in}}$ and $\boldsymbol{\mu}_{\text{out}}$), reweighting the rows and the columns of $\widehat{\mathbf{T}}$ such that the row marginals and the column marginals of the new rescaled matrix are equal to $\boldsymbol{\sigma}_{\text{in}}$ and $\boldsymbol{\sigma}_{\text{out}}$, just like in the context of standard optimal transport with entropy regularization (Cuturi, 2013).

Indeed, this procedure is closely related to the solution of the standard, relaxed, optimal transport problem with entropy regularization when using the matrix $\widehat{\mathbf{T}}$ as the matrix containing the costs or rewards of transportation (Wilson, 1970; Kapur, 1989; Erlander & Stewart, 1990; Kapur & Kesavan, 1992; Cuturi, 2013). It is usually solved by iterative proportional fitting, matrix balancing, or biproportional scaling (Bacharach, 1965; Sinkhorn, 1967; see also Kurras, 2015, and Pukelsheim, 2014, and references therein for a more recent discussion). The iterative proportional fitting algorithm has guaranteed convergence to a unique solution under some mild conditions (see, e.g., Kurras, 2015; Pukelsheim, 2014).

Our iterative procedure for solving Equations (26) and (27) consists of first fixing an arbitrary $\boldsymbol{\mu}_{\text{out}}^{(0)}$ and then computing $\boldsymbol{\mu}_{\text{in}}^{(\tau)}$ from Equation (26) at iteration τ. Thereafter, $\boldsymbol{\mu}_{\text{in}}^{(\tau)}$ is kept fixed and $\boldsymbol{\mu}_{\text{out}}^{(\tau+1)}$ is computed from $\boldsymbol{\mu}_{\text{in}}^{(\tau)}$ in Equation (27).

Finally, note that the optimal probability distribution $\mathrm{P}^{\star}(\wp)$ given by Proposition 2.3 is useless in practice, as there is an infinite number of paths. However, the different interesting and useful quantities can also be computed from the fundamental matrix and the Lagrange multipliers, as shown in the next section. Moreover, it will be shown in Section 4.3 that the optimization problem can be reduced to the estimation of the joint probabilities $\mathrm{P}^{\star}(S = i, E = j) = \sum_{\wp_{ij} \in \mathcal{P}_{ij}} \mathrm{P}^{\star}(\wp_{ij})$ (the coupling), therefore completely avoiding the introduction of the probability distribution over paths.

2.5 Computation of other important quantities

Similarly to Françoisse et al. (2017), other meaningful quantities can be computed in closed form after the convergence of $\boldsymbol{\mu}_{\text{in}}$ and $\boldsymbol{\mu}_{\text{out}}$. This section provides expressions for computing them.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Algorithm 2: Regular, non-hitting, margin-constrained bag-of-paths algorithm

Input:
- The $n \times n$ nonnegative adjacency matrix $\mathbf{A}$ of a strongly connected directed graph, containing edge affinities.
- The $n \times n$ nonnegative cost matrix $\mathbf{C}$ containing edge costs.
- An input distribution vector $\boldsymbol{\sigma}_{\text{in}}$ of size n.
- An output distribution vector $\boldsymbol{\sigma}_{\text{out}}$ of size n.
- A killing rate vector $\boldsymbol{\alpha}$ of size n.
- The expected number of visits to each node for the reference random walk, $\bar{\mathbf{n}}^{\text{ref}}$, of size n.
- A scalar inverse temperature parameter $\beta > 0$.

Output:
- The coupling matrix $\boldsymbol{\Gamma}$ of size $n \times n$.
- The minimum free energy $\text{FE}^{\min}$ (a scalar).
- The matrix $\bar{\mathbf{N}}$ of size $n \times n$, containing the expected number of visits to each edge (i, j).
- The vector $\bar{\mathbf{n}}$ of size n, containing the expected number of visits to each node.
- The resulting biased transition probabilities matrix $\mathbf{P}$ of size $n \times n$ (the policy).

$\mathbf{D} \leftarrow \mathbf{Diag}(\mathbf{Ae})$ ▷ the out-degree matrix
$\mathbf{P}^{\text{ref}} \leftarrow \mathbf{D}^{-1}\mathbf{A}$ ▷ the reference transition probabilities matrix
$\widehat{\mathbf{P}}^{\text{ref}} \leftarrow (\mathbf{I} - \mathbf{Diag}(\boldsymbol{\alpha}))\mathbf{P}^{\text{ref}}$, ▷ the killed reference transition probabilities matrix
$\widehat{\mathbf{W}} \leftarrow \widehat{\mathbf{P}}^{\text{ref}} \circ \exp[-\beta\mathbf{C}]$ ▷ the auxiliary matrix $\widehat{\mathbf{W}}$; $\circ$ is the elementwise product
$\widehat{\mathbf{Z}} \leftarrow (\mathbf{I} - \widehat{\mathbf{W}})^{-1}$ ▷ the fundamental matrix
$\boldsymbol{\mu}_{\text{out}} \leftarrow \mathbf{e}$ ▷ the initial Lagrange parameters vector for ouputs
repeat ▷ main iteration loop
  $\boldsymbol{\mu}_{\text{in}} \leftarrow \mathbf{e} \div \left(\widehat{\mathbf{Z}}\left(\boldsymbol{\mu}_{\text{out}} \circ \boldsymbol{\alpha}\right)\right)$ ▷ the update of input vector; $\div$ is the elementwise division
  $\boldsymbol{\mu}_{\text{out}} \leftarrow \bar{\mathbf{n}}^{\text{ref}} \div \left(\widehat{\mathbf{Z}}^{\top}\left(\boldsymbol{\mu}_{\text{in}} \circ \boldsymbol{\sigma}_{\text{in}}\right)\right)$ ▷ the update of output vector
until convergence
$\boldsymbol{\Gamma} \leftarrow \mathbf{Diag}(\boldsymbol{\mu}_{\text{in}} \circ \boldsymbol{\sigma}_{\text{in}})\,\widehat{\mathbf{Z}}\,\mathbf{Diag}(\boldsymbol{\mu}_{\text{out}} \circ \boldsymbol{\alpha})$ ▷ the coupling matrix
$\text{FE}^{\min} \leftarrow \frac{1}{\beta}((\log[\boldsymbol{\mu}_{\text{in}}])^{\top}\boldsymbol{\sigma}_{\text{in}} + (\log[\boldsymbol{\mu}_{\text{out}}])^{\top}\boldsymbol{\sigma}_{\text{out}})$ ▷ the minimum free energy value
$\bar{\mathbf{N}} \leftarrow \mathbf{Diag}(\bar{\mathbf{n}}^{\text{ref}} \div \boldsymbol{\mu}_{\text{out}})\,\widehat{\mathbf{W}}\,(\mathbf{Diag}(\boldsymbol{\mu}_{\text{in}}))^{-1}$ ▷ the matrix containing the expected number of visits to each edge
$\bar{\mathbf{n}} \leftarrow \bar{\mathbf{n}}^{\text{ref}} \div (\boldsymbol{\mu}_{\text{in}} \circ \boldsymbol{\mu}_{\text{out}})$ ▷ the vector containing the expected number of visits to each node
$\mathbf{P} \leftarrow \widehat{\mathbf{W}} \div \left(\left((\mathbf{e} \div \boldsymbol{\mu}_{\text{in}}) - (\boldsymbol{\mu}_{\text{out}} \circ \boldsymbol{\alpha})\right)\boldsymbol{\mu}_{\text{in}}^{\top}\right)$ ▷ the biased transition matrix (the policy)
return $\boldsymbol{\Gamma}$, $\text{FE}^{\min}$, $\bar{\mathbf{N}}$, $\bar{\mathbf{n}}$, $\mathbf{P}$

The computation of the Lagrange multipliers together with the derived quantities, for the regular, non-hitting, BoP model, is summarized in Algorithm 2.

2.5.1 The coupling matrix

First, the definition and computation of the *coupling matrix* are presented. Its name derives from the transportation science literature (Villani, 2003, 2008).

Proposition 2.5. *The coupling matrix, denoted by* $\Gamma = (\gamma_{ij})$ *and defined by*

$$\gamma_{ij} \triangleq \mathrm{P}^{\star}(S = i, E = j) = \sum_{\wp_{ij} \in \mathcal{P}_{ij}} \mathrm{P}^{\star}(\wp_{ij}), \tag{33}$$

where $\mathcal{P}_{ij}$ *is the set of paths starting in node i and ending in node j, can be computed by*

$$\Gamma = \mathbf{Diag}(\boldsymbol{\mu}_{\text{in}} \circ \boldsymbol{\sigma}_{\text{in}})\,\widehat{\mathbf{Z}}\,\mathbf{Diag}(\boldsymbol{\mu}_{\text{out}} \circ \boldsymbol{\alpha}). \tag{34}$$

Proof. From the optimal probabilities Equation (23) and the definition of the fundamental matrix Equation (30), the probability of drawing a path starting in i and ending j is

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

$$\begin{aligned}\mathrm{P}^\star(S=i,E=j) &= \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \mathrm{P}^\star(\wp_{ij}) \\ &= \mu_i^{\mathrm{in}}\sigma_i^{\mathrm{in}}\mu_j^{\mathrm{out}}\alpha_j \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \widehat{\pi}^{\mathrm{ref}}(\wp_{ij})\exp\left(-\beta\tilde{c}(\wp_{ij})\right) \qquad (35)\\ &= \mu_i^{\mathrm{in}}\sigma_i^{\mathrm{in}}\widehat{z}_{ij}\mu_j^{\mathrm{out}}\alpha_j, \qquad (36)\end{aligned}$$

which in matrix form writes out as Equation (34). □

2.5.2 The optimal free energy

Proposition 2.6. *The value of the free energy at the optimal paths probability distribution* $\mathrm{P}^\star$ *is*

$$\mathrm{FE}(\mathrm{P}^\star) = -\left(\boldsymbol{\lambda}_{\mathrm{in}}^\top\boldsymbol{\sigma}_{\mathrm{in}} + \boldsymbol{\lambda}_{\mathrm{out}}^\top\boldsymbol{\sigma}_{\mathrm{out}}\right), \qquad (37)$$

where the Lagrange parameter vectors $\boldsymbol{\lambda}_{\mathrm{in}}$, $\boldsymbol{\lambda}_{\mathrm{in}}$ *are obtained from the vectors* $\boldsymbol{\mu}_{\mathrm{in}}$, $\boldsymbol{\mu}_{\mathrm{in}}$ *as stated in Proposition 2.3.*

Proof. By replacing $\mathrm{P}^\star(\wp_{ij}) = \mu_i^{\mathrm{in}}\mu_j^{\mathrm{out}}\mathrm{P}^{\mathrm{ref}}(\wp_{ij})\exp\left(-\beta\tilde{c}(\wp_{ij})\right)$ into the free energy expression (5), we get

$$\mathrm{FE}(\mathrm{P}^\star) = -\sum_{i\in\mathcal{V}}\lambda_i^{\mathrm{in}}\sum_{j\in\mathcal{V}}\sum_{\wp_{ij}\in\mathcal{P}_{ij}}\mathrm{P}^\star(\wp_{ij}) - \sum_{j\in\mathcal{V}}\lambda_j^{\mathrm{out}}\sum_{i\in\mathcal{V}}\sum_{\wp_{ij}\in\mathcal{P}_{ij}}\mathrm{P}^\star(\wp_{ij})$$

and as the margins are fixed, we get the result. □

An interesting interpretation of this proposition based on an optimal transportation analogy is discussed in Section 4.

2.5.3 The expected number of passages through an edge

Let us define the matrix $\bar{\mathbf{N}} = (\bar{n}_{ij})$ as the matrix containing the expected number of times an edge (i,j) appears on a path drawn from the optimal distribution $\mathrm{P}^\star(\wp)$. Formally,

$$\bar{n}_{ij} \triangleq \sum_{k,l\in\mathcal{V}}\sum_{\wp_{kl}\in\mathcal{P}_{kl}}\mathrm{P}^\star(\wp_{kl})\eta(i\to j\in\wp_{kl}),$$

where $\eta(i\to j\in\wp_{kl})$ denotes the number of times the edge (i,j) appears on path $\wp_{kl}$.

In Guex (2016), $\bar{n}_{ij}$ is interpreted as the *flow on edges*, creating a "stream of matter" going from supply nodes in $\mathcal{I}n$ to destination nodes in $\mathcal{O}ut$. This interpretation will also be discussed further in Section 4.

Proposition 2.7. *The matrix* $\bar{\mathbf{N}}$*, containing the expected number of times an edge* (i,j) *appears on a drawn path, is given by*

$$\bar{\mathbf{N}} = \mathbf{Diag}(\bar{\mathbf{n}}^{\mathrm{ref}} \div \boldsymbol{\mu}_{\mathrm{out}})\,\widehat{\mathbf{W}}\,(\mathbf{Diag}(\boldsymbol{\mu}_{\mathrm{in}}))^{-1}, \qquad (38)$$

where $\widehat{\mathbf{W}}$ *is defined after Equation (30).*

Proof. Because costs are additive along paths, $\eta(i\to j\in\wp_{kl}) = \partial\tilde{c}(\wp_{kl})/\partial c_{ij}$. Then, by using Equation (23), we obtain

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

$$\bar{n}_{ij} = \sum_{k,l\in\mathcal{V}} \mu_k^{\text{in}}\sigma_k^{\text{in}}\mu_l^{\text{out}}\alpha_l \sum_{\wp_{kl}\in\mathcal{P}_{kl}} \widehat{\pi}^{\text{ref}}(\wp_{kl}) \exp\left(-\beta\tilde{c}(\wp_{kl})\right)\frac{\partial\tilde{c}(\wp_{kl})}{\partial c_{ij}}$$
$$= -\frac{1}{\beta}\sum_{k,l\in\mathcal{V}} \mu_k^{\text{in}}\sigma_k^{\text{in}}\mu_l^{\text{out}}\alpha_l\frac{\partial\widehat{z}_{kl}}{\partial c_{ij}}. \tag{39}$$

We know from direct calculus (Kivimäki et al., 2016, Equation (11)) that $\partial\widehat{z}_{kl}/\partial c_{ij} = -\beta\widehat{w}_{ij}\widehat{z}_{ki}\widehat{z}_{jl}$. Thus,

$$\bar{n}_{ij} = \widehat{w}_{ij}\left(\sum_{k\in\mathcal{V}}\mu_k^{\text{in}}\sigma_k^{\text{in}}\widehat{z}_{ki}\right)\left(\sum_{l\in\mathcal{V}}\mu_l^{\text{out}}\alpha_l\widehat{z}_{jl}\right), \tag{40}$$

and by using Equations (28) and (29) with $\alpha_l = \sigma_l^{\text{out}}/\bar{n}_l^{\text{ref}}$ (see Proposition 2.1),

$$\bar{n}_{ij} = \frac{\bar{n}_i^{\text{ref}}\widehat{w}_{ij}}{\mu_i^{\text{out}}\mu_j^{\text{in}}}, \tag{41}$$

which provides the desired result. □

2.5.4 The expected number of visits to a node

Let us further define the vector $\bar{\mathbf{n}} = (\bar{n}_i)$, containing the expected number of times node i is drawn under $\mathrm{P}^\star(\wp)$, by

$$\bar{n}_i \triangleq \sum_{\wp\in\mathcal{P}} \mathrm{P}^\star(\wp)\,\eta(i\in\wp),$$

where $\eta(i\in\wp)$ denotes the number of times node i appears on path $\wp$.

Proposition 2.8. *The vector* $\bar{\mathbf{n}}$*, containing the expected number of times node* i *is drawn, is*

$$\bar{\mathbf{n}} = \bar{\mathbf{n}}^{\text{ref}} \div (\boldsymbol{\mu}_{\text{in}} \circ \boldsymbol{\mu}_{\text{out}}), \tag{42}$$

where $\div$ *and* $\circ$ *are the elementwise division and product.*

Proof. In fact, we can decompose this quantity as

$$\eta(i\in\wp) = \sum_{j\in\mathcal{V}}\eta(i\to j\in\wp) + \delta(E(\wp)=i), \tag{43}$$

where $\delta(E(\wp)=i)$ is one if i is the ending node of the path $\wp$. This gives

$$\bar{n}_i = \sum_{j\in\mathcal{V}}\sum_{\wp\in\mathcal{P}}\mathrm{P}^\star(\wp)\eta(i\to j\in\wp) + \sum_{\wp\in\mathcal{P}}\mathrm{P}^\star(\wp)\delta(E(\wp)=i) = \sum_{j\in\mathcal{V}}\bar{n}_{ij} + \sigma_i^{\text{out}}. \tag{44}$$

By using Equation (40), we find

$$\bar{n}_i = \left(\sum_{k\in\mathcal{V}}\mu_k^{\text{in}}\sigma_k^{\text{in}}\widehat{z}_{ki}\right)\left(\sum_{l\in\mathcal{V}}\mu_l^{\text{out}}\alpha_l\sum_{j\in\mathcal{V}}\widehat{w}_{ij}\widehat{z}_{jl}\right) + \sigma_i^{\text{out}}.$$

However, because $(\mathbf{I}-\widehat{\mathbf{W}})\widehat{\mathbf{Z}} = \mathbf{I}$, $\sum_{j\in\mathcal{V}}\widehat{w}_{ij}\widehat{z}_{jl} = \widehat{z}_{il} - \delta_{il}$ as shown by Kivimäki et al. (2016, Equation (13)); therefore

$$\bar{n}_i = \sum_{k\in\mathcal{V}}\mu_k^{\text{in}}\sigma_k^{\text{in}}\widehat{z}_{ki}\left(\sum_{l\in\mathcal{V}}\mu_l^{\text{out}}\alpha_l\widehat{z}_{il} - \mu_i^{\text{out}}\alpha_i\right) + \sigma_i^{\text{out}},$$

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

which, by using Equations (28) and (29) and $\alpha_i = \sigma_i^{\text{out}}/\bar{n}_i^{\text{ref}}$ (see Proposition 2.1), provides the result

$$\bar{n}_i = \frac{\bar{n}_i^{\text{ref}}}{\mu_i^{\text{in}}\mu_i^{\text{out}}}. \tag{45}$$

□

2.5.5 The optimal randomized policy

The expected number of times edges are visited induces a *biased* random walk on the network with transition matrix $\mathbf{P} \triangleq (p_{ij})$ provided by

$$p_{ij} \triangleq \frac{\bar{n}_{ij}}{\sum_{k\in\mathcal{V}} \bar{n}_{ik}}. \tag{46}$$

Proposition 2.9. *The biased random walk transition matrix,* $\mathbf{P}$*, called the randomized routing policy, is provided by*

$$\mathbf{P} = \widehat{\mathbf{W}} \div \left(\left((\mathbf{e} \div \boldsymbol{\mu}_{\text{in}}) - (\boldsymbol{\mu}_{\text{out}} \circ \boldsymbol{\alpha}) \right) \boldsymbol{\mu}_{\text{in}}^{\top} \right). \tag{47}$$

Proof. The result is obtained in elementwise form by combining expression (46) with (41), (44), and (45) and using $\alpha_i = \sigma_i^{\text{out}}/\bar{n}_i^{\text{ref}}$ (see Proposition 2.1):

$$p_{ij} = \frac{\mu_i^{\text{in}}\widehat{w}_{ij}}{\mu_j^{\text{in}}(1 - \mu_i^{\text{in}}\mu_i^{\text{out}}\alpha_i)}.$$

□

We can observe that when $\beta \to 0$, then $\boldsymbol{\mu}_{\text{in}} \to \mathbf{e}$, $\boldsymbol{\mu}_{\text{out}} \to \mathbf{e}$, $\widehat{\mathbf{W}} \to \widehat{\mathbf{P}}^{\text{ref}}$ and we obtain

$$p_{ij} \xrightarrow[T\to\infty]{} \frac{\widehat{p}_{ij}^{\text{ref}}}{(1-\alpha_i)} = p_{ij}^{\text{ref}},$$

as it should be.

This biased random walk is the *optimal policy* that has to be followed for reaching nodes in $\mathcal{O}ut$ from nodes in $\mathcal{I}n$ and can be interpreted as follows. When $\beta \to 0$, the behavior becomes similar to the random walk defined by $\mathbf{P}^{\text{ref}}$, but as β increases, random walkers are more and more "attracted" by high σ_i^{out} nodes. These "pools of attraction," whose sizes are related to the components of $\boldsymbol{\sigma}_{\text{out}}$, get less and less "fuzzy" as β increases, eventually forcing walkers to adopt quasideterministic, optimal, paths following the solution of the optimal transport on a graph problem (see Section 6.2). Therefore, this framework can be viewed as an extension of standard electrical networks, interpolating between an optimal behavior based on least cost paths and a random behavior based on the reference probabilities (Guex, 2016).

3. The margin-constrained bag-of-hitting-paths formalism

In (Françoisse et al., 2017), the BoP formalism was defined for regular paths (as in previous section) as well as for hitting paths, i.e., paths where the final node j appears only once as last node of the path. In this section, we will now consider the margin-constrained problem for *hitting paths* and define, accordingly, the *margin-constrained bag-of-hitting-paths* (cBoHP) *framework*. This new model will yield interesting properties and analogies with other models and will require less computation time in comparison to the non-hitting, regular, bag-of-paths model considered so far in this work. We will see that while both models are similar when $T \to 0$, they are quite

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

different when $T \to \infty$ and that the hitting formalism has a somewhat more straightforward solution. Nevertheless, the hitting paths assumption can prove more practical and appropriate in practice. But, of course, the choice of whether to consider non-hitting or hitting paths depends on the application.

3.1 Problem definition

Let $\mathcal{P}_{ij}^{\mathrm{h}}$ be the set of all *hitting paths* starting in i and ending in j, i.e., all paths $\wp_{ij} = (i_0, \ldots, i_t)$ where $i_0 = i$, $i_t = j$ and $i_\tau \neq j, \forall \tau \neq t$ where t is the length of the path. This means that, technically, the ending node is turned into a killing, absorbing, node from which we cannot escape (Françoisse et al., 2017; Kivimäki et al., 2014; Fouss et al., 2016). We define the set of all hitting paths, also named *bag-of-hitting-paths*, by $\mathcal{P}_{\mathrm{h}} = \cup_{i,j \in \mathcal{V}} \mathcal{P}_{ij}^{\mathrm{h}}$. By analogy with Equation (5), the problem here is to find the *optimal hitting paths probability distribution*, $\mathrm{P}_{\mathrm{h}}^{\star}(\wp)$, solving

$$\left| \begin{array}{ll} \underset{\{\mathrm{P_h}(\wp)\}_{\wp \in \mathcal{P}_{\mathrm{h}}}}{\text{minimize}} & \mathrm{FE_h}(\mathrm{P_h}) = \sum_{\wp \in \mathcal{P}_{\mathrm{h}}} \mathrm{P_h}(\wp)\tilde{c}(\wp) + T \sum_{\wp \in \mathcal{P}_{\mathrm{h}}} \mathrm{P_h}(\wp) \log\left(\dfrac{\mathrm{P_h}(\wp)}{\mathrm{P_h^{ref}}(\wp)}\right) \\ \text{subject to} & \sum_{j \in \mathcal{V}} \sum_{\wp_{ij} \in \mathcal{P}_{ij}^{\mathrm{h}}} \mathrm{P_h}(\wp_{ij}) = \sigma_i^{\mathrm{in}} \qquad \forall i \in \mathcal{V}, \\ & \sum_{i \in \mathcal{V}} \sum_{\wp_{ij} \in \mathcal{P}_{ij}^{\mathrm{h}}} \mathrm{P_h}(\wp_{ij}) = \sigma_j^{\mathrm{out}} \qquad \forall j \in \mathcal{V}. \end{array} \right. \tag{48}$$

As probabilities for regular paths containing the final node more than once converge to zero in the non-hitting formalism when $T \to 0$ (it is suboptimal to visit several times the same node), we easily see that both problems are equivalent at this limit. However, this is not the case when $T \to \infty$, due to the difference in reference probabilities between the two models and the structure of the paths, as shown in the next section.

Yet another important difference between the hitting and the non-hitting formulations is that the former model *is equivalent to the standard entropy regularized optimal transport problem* (Wilson, 1970; Erlander & Stewart, 1990; Cuturi, 2013) based on the directed free energy distance (or potential) between nodes (Kivimäki et al., 2014; Françoisse et al., 2017; Fouss et al., 2016), which can easily be precomputed for the whole graph. This is detailed in Section 4 (see Equation (71)). A last difference is that the hitting paths formulation does not need the preprocessing step computing the reference probabilities with fixed margins described in Section 2.3 and Algorithm 1.

3.2 Reference probabilities with fixed margins

Finding reference probabilities $\mathrm{P_h^{ref}}(\wp)$ is at the heart of the difference between both formalisms and is greatly simplified in the hitting case. In fact, it was shown by Françoisse et al. (2017) that the sum of likelihoods over all hitting paths between two nodes i and j is always equal to 1. In other words,

$$\sum_{\wp \in \mathcal{P}_{ij}^{\mathrm{h}}} \tilde{\pi}^{\mathrm{ref}}(\wp) = 1,$$

where $\tilde{\pi}^{\mathrm{ref}}(\wp) \triangleq \Pi_{\tau=1}^{t} p_{i_{\tau-1}, i_\tau}^{\mathrm{ref}}$ and with the p_{ij}^{ref} being the reference transition probabilities defined in Equation (1). From this, it is easy to observe that the reference probability defined by

$$\mathrm{P_h^{ref}}(\wp_{ij}) \triangleq \sigma_i^{\mathrm{in}} \sigma_j^{\mathrm{out}} \tilde{\pi}^{\mathrm{ref}}(\wp_{ij}), \tag{49}$$

yields the correct margins as expressed in the constraints of problem (48).

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

3.3 Computation of the optimal probability distribution over paths

The reasoning for finding the solution follows the same rationale as before for the non-hitting case; therefore the details of the proofs are not repeated in this section. The main difference lies in the replacement of the reference probabilities with Equation (49) and the following new expression for hitting paths (the equivalent of Equation (31))

$$\sum_{\wp_{ij} \in \mathcal{P}_{ij}^{\mathrm{h}}} \tilde{\pi}^{\mathrm{ref}}(\wp_{ij}) \exp\left(-\beta \tilde{c}(\wp_{ij})\right) = z_{ij}^{\mathrm{h}}, \tag{50}$$

with $\mathbf{Z}_{\mathrm{h}} = (z_{ij}^{\mathrm{h}})$ being the *fundamental matrix for hitting paths* as introduced in (Kivimäki et al., 2014, Equation (12)), obtained through

$$\mathbf{Z}_{\mathrm{h}} \triangleq \mathbf{Z}\mathbf{D}_{\mathrm{h}}^{-1}, \tag{51}$$

where $\mathbf{Z} \triangleq (\mathbf{I} - \mathbf{W})^{-1}$, $\mathbf{W} \triangleq \mathbf{P}^{\mathrm{ref}} \circ \exp\left[-\beta \mathbf{C}\right]$, and $\mathbf{D}_{\mathrm{h}} \triangleq \mathbf{Diag}(\mathbf{Z})$, the diagonal matrix containing the main diagonal of $\mathbf{Z}$. Elementwise, we have $z_{ij}^{\mathrm{h}} = z_{ij}/z_{jj}$.

3.3.1 The optimal hitting-paths probabilities

The optimal hitting-paths probabilities are obtained with the following proposition.

Proposition 3.1. *When the set of paths $\mathcal{P}$ is restricted to the set of hitting paths $\mathcal{P}_{\mathrm{h}}$ and the reference path probabilities $\mathrm{P}^{\mathrm{ref}}(\wp)$ are defined according to Equation (49), then the minimization problem (48) is solved by*

$$\mathrm{P}_{\mathrm{h}}^{\star}(\wp_{ij}) = \mu_i^{\mathrm{h,in}} \sigma_i^{\mathrm{in}} \mu_j^{\mathrm{h,out}} \sigma_j^{\mathrm{out}} \tilde{\pi}^{\mathrm{ref}}(\wp_{ij}) \exp\left(-\beta \tilde{c}(\wp_{ij})\right), \tag{52}$$

where $\beta \triangleq 1/T$ is the inverse temperature parameter, and $\mu_i^{\mathrm{h,in}} \triangleq \exp\left(-\beta \lambda_i^{\mathrm{h,in}}\right)$, $\mu_j^{\mathrm{h,out}} \triangleq \exp\left(-\beta \lambda_j^{\mathrm{h,out}}\right)$ are two vectors derived from the Lagrange parameter vectors $\boldsymbol{\lambda}_{\mathrm{h,in}}$, $\boldsymbol{\lambda}_{\mathrm{h,out}}$, associated with the constraints.

Proof. The proof is similar to the proof of Proposition 2.3. □

3.3.2 Computing the Lagrange parameters

Proposition 3.2. *The two vectors $\boldsymbol{\mu}_{\mathrm{h,in}}$ and $\boldsymbol{\mu}_{\mathrm{h,out}}$, defined by $\mu_i^{\mathrm{h,in}} = \exp\left(-\beta \lambda_i^{\mathrm{h,in}}\right)$ and $\mu_i^{\mathrm{h,out}} = \exp\left(-\beta \lambda_i^{\mathrm{h,out}}\right)$, verify the following expressions*

$$\boldsymbol{\mu}_{\mathrm{in}}^{\mathrm{h}} = \mathbf{e} \div \left(\mathbf{Z}_{\mathrm{h}}\left(\boldsymbol{\mu}_{\mathrm{out}}^{\mathrm{h}} \circ \boldsymbol{\sigma}_{\mathrm{out}}\right)\right), \tag{53}$$

$$\boldsymbol{\mu}_{\mathrm{out}}^{\mathrm{h}} = \mathbf{e} \div \left(\mathbf{Z}_{\mathrm{h}}^{\top}\left(\boldsymbol{\mu}_{\mathrm{in}}^{\mathrm{h}} \circ \boldsymbol{\sigma}_{\mathrm{in}}\right)\right). \tag{54}$$

Proof. The proof is similar to the proof of Proposition 2.4, but by replacing the matrix $\mathbf{Z}$ with $\mathbf{Z}_{\mathrm{h}}$. □

As for regular paths, these two expressions are recomputed iteratively until convergence. Note that in the hitting paths case, the matrix that needs to be rescaled in order to satisfy the margin constraints is

$$\mathbf{T}_{\mathrm{h}} = \mathbf{Diag}(\boldsymbol{\sigma}_{\mathrm{in}}) \mathbf{Z}_{\mathrm{h}} \mathbf{Diag}(\boldsymbol{\sigma}_{\mathrm{out}}). \tag{55}$$

The computation of the Lagrange multipliers and the derived quantities for the hitting bag-of-paths model is summarized in Algorithm 3.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Algorithm 3: Margin-constrained bag-of-hitting-paths algorithm

Input:
- The $n \times n$ nonnegative adjacency matrix $\mathbf{A}$ of a strongly connected directed graph, containing edge affinities.
- The $n \times n$ nonnegative cost matrix $\mathbf{C}$, containing edge costs.
- An input distribution vector $\boldsymbol{\sigma}_{\text{in}}$ of size n.
- An output distribution vector $\boldsymbol{\sigma}_{\text{out}}$ of size n.
- A scalar inverse temperature parameter $\beta > 0$.

Output:
- The coupling matrix $\boldsymbol{\Gamma}_{\text{h}}$ of size $n \times n$.
- The minimum free energy scalar $\text{FE}_{\text{h}}^{\min}$.
- The matrix $\bar{\mathbf{N}}_{\text{h}}$ of size $n \times n$, containing the expected number of visits to each edge (i, j).
- The vector $\bar{\mathbf{n}}_{\text{h}}$ of size n, containing the expected number of visits to each node.
- The resulting biased transition probabilities matrix $\mathbf{P}_{\text{h}}$ of size $n \times n$ (the policy).

$\mathbf{D} \leftarrow \mathbf{Diag}(\mathbf{Ae})$ ▷ the out-degree matrix
$\mathbf{P}^{\text{ref}} \leftarrow \mathbf{D}^{-1}\mathbf{A}$ ▷ the reference transition probabilities matrix
$\mathbf{W} \leftarrow \mathbf{P}^{\text{ref}} \circ \exp[-\beta \mathbf{C}]$ ▷ the auxiliary matrix $\mathbf{W}$; $\circ$ is the elementwise product
$\mathbf{Z} \leftarrow (\mathbf{I} - \mathbf{W})^{-1}$ ▷ the fundamental matrix
$\mathbf{D}_{\text{h}} \leftarrow \mathbf{Diag}(\mathbf{Z})$ ▷ the column-normalization matrix for hitting paths probabilities
$\mathbf{Z}_{\text{h}} \leftarrow \mathbf{Z}\mathbf{D}_{\text{h}}^{-1}$ ▷ the column-normalized fundamental matrix
$\boldsymbol{\mu}_{\text{out}}^{\text{h}} \leftarrow \mathbf{e}$ ▷ the initial Lagrange parameters vector for ouputs
repeat ▷ main iteration loop
  $\boldsymbol{\mu}_{\text{in}}^{\text{h}} \leftarrow \mathbf{e} \div \left(\mathbf{Z}_{\text{h}}\left(\boldsymbol{\mu}_{\text{out}}^{\text{h}} \circ \boldsymbol{\sigma}_{\text{out}}\right)\right)$ ▷ the update of input vector
  $\boldsymbol{\mu}_{\text{out}}^{\text{h}} \leftarrow \mathbf{e} \div \left(\mathbf{Z}_{\text{h}}^{\top}\left(\boldsymbol{\mu}_{\text{in}}^{\text{h}} \circ \boldsymbol{\sigma}_{\text{in}}\right)\right)$ ▷ the update of output vector
until convergence
$\Gamma_{\text{h}} \leftarrow \mathbf{Diag}(\boldsymbol{\mu}_{\text{in}}^{\text{h}} \circ \boldsymbol{\sigma}_{\text{in}})\, \mathbf{Z}_{\text{h}}\, \mathbf{Diag}(\boldsymbol{\mu}_{\text{out}}^{\text{h}} \circ \boldsymbol{\sigma}_{\text{out}})$ ▷ the coupling matrix
$\text{FE}_{\text{h}}^{\min} \leftarrow \frac{1}{\beta}((\log[\boldsymbol{\mu}_{\text{in}}^{\text{h}}])^{\top}\boldsymbol{\sigma}_{\text{in}} + (\log[\boldsymbol{\mu}_{\text{out}}^{\text{h}}])^{\top}\boldsymbol{\sigma}_{\text{out}})$ ▷ the minimum free energy value
$\bar{\mathbf{N}}_{\text{h}} \leftarrow \left((\mathbf{e} \div \boldsymbol{\mu}_{\text{out}}^{\text{h}})(\mathbf{e} \div \boldsymbol{\mu}_{\text{in}}^{\text{h}})^{\top} - (\mathbf{Z}_{\text{h}}\mathbf{Diag}(\boldsymbol{\sigma}_{\text{out}})\mathbf{Z}_{\text{h}})^{\top}\right) \circ (\mathbf{Diag}(\mathbf{Z})\mathbf{W})$ ▷ the matrix containing the expected number of visits to each edge
$\bar{\mathbf{n}}_{\text{h}} \leftarrow \mathbf{Diag}(\mathbf{Z})\left((\mathbf{e} \div (\boldsymbol{\mu}_{\text{in}}^{\text{h}} \circ \boldsymbol{\mu}_{\text{out}}^{\text{h}})) - \mathbf{diag}(\mathbf{Z}_{\text{h}}\mathbf{Diag}(\boldsymbol{\sigma}_{\text{out}})\mathbf{Z}_{\text{h}})\right) + \boldsymbol{\sigma}_{\text{out}}$ ▷ the vector containing the expected number of visits to each node
$\mathbf{P}_{\text{h}} \leftarrow (\mathbf{Diag}(\bar{\mathbf{N}}_{\text{h}}\mathbf{e}))^{-1}\bar{\mathbf{N}}_{\text{h}}$ ▷ the biased transition matrix (the policy)
return Γ_{h}, $\text{FE}_{\text{h}}^{\min}$, $\bar{\mathbf{N}}_{\text{h}}$, $\bar{\mathbf{n}}_{\text{h}}$, $\mathbf{P}_{\text{h}}$

3.4 Computation of other important quantities

The computation of the other interesting quantities is slightly different in the hitting formalism and shows some interesting new properties. They are reviewed in this section.

3.4.1 The coupling matrix

Proposition 3.3. *The coupling matrix for hitting paths* $\Gamma_{\text{h}} = (\gamma_{ij}^{\text{h}})$ *is given by*

$$\Gamma_{\text{h}} = \mathbf{Diag}\left(\boldsymbol{\mu}_{\text{in}}^{\text{h}} \circ \boldsymbol{\sigma}_{\text{in}}\right) \mathbf{Z}_{\text{h}}\, \mathbf{Diag}\left(\boldsymbol{\mu}_{\text{out}}^{\text{h}} \circ \boldsymbol{\sigma}_{\text{out}}\right). \tag{56}$$

Proof. The proof is similar to the proof of Proposition 2.5. □

Notice that, for the hitting paths case, an alternative way of obtaining the coupling matrix by precomputing the directed free energy distances is discussed in Section 4.3 (see Equation (71)).

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

3.4.2 The optimal free energy

Proposition 3.4. *The value of the free energy at the optimal distribution* $\mathrm{P}_{\mathrm{h}}^{\star}$ *is*

$$\mathrm{FE}_{\mathrm{h}}\left(\mathrm{P}_{\mathrm{h}}^{\star}\right) = -\left(\left(\boldsymbol{\lambda}_{\mathrm{in}}^{\mathrm{h}}\right)^{\top}\boldsymbol{\sigma}_{\mathrm{in}} + \left(\boldsymbol{\lambda}_{\mathrm{out}}^{\mathrm{h}}\right)^{\top}\boldsymbol{\sigma}_{\mathrm{out}}\right). \tag{57}$$

Proof. The proof is similar to the proof of Proposition 2.6. □

3.4.3 The expected number of visits to an edge

Proposition 3.5. *The matrix* $\bar{\mathbf{N}}_{\mathrm{h}}$*, containing the expected number of times an edge* (i, j) *appears on a hitting path, is given by*

$$\bar{\mathbf{N}}_{\mathrm{h}} = \left(\left(\mathbf{e} \div \boldsymbol{\mu}_{\mathrm{out}}^{\mathrm{h}}\right)\left(\mathbf{e} \div \boldsymbol{\mu}_{\mathrm{in}}^{\mathrm{h}}\right)^{\top} - \left(\mathbf{Z}_{\mathrm{h}}\mathbf{Diag}(\boldsymbol{\sigma}_{\mathrm{out}})\mathbf{Z}_{\mathrm{h}}\right)^{\top}\right) \circ (\mathbf{Diag}(\mathbf{Z})\mathbf{W}), \tag{58}$$

where $\mathbf{Diag}(\mathbf{Z})$ *is a diagonal matrix containing the main diagonal of* $\mathbf{Z}$ *and* $\circ$ *is the elementwise matrix product.*

Proof. Let $\bar{n}_{ij}^{\mathrm{h}}$ be the expected number of times edge (i, j) appears on a path drawn according to $\mathrm{P}_{\mathrm{h}}(\wp)$, i.e.,

$$\bar{n}_{ij}^{\mathrm{h}} \triangleq \sum_{k,l\in\mathcal{V}} \sum_{\wp_{kl}\in\mathcal{P}_{kl}^{\mathrm{h}}} \mathrm{P}_{\mathrm{h}}(\wp_{kl})\, \eta(i \to j \in \wp_{kl}),$$

where $\eta(i \to j \in \wp_{kl})$ denotes the number of times the edge (i, j) is visited along hitting path $\wp_{kl}$. By a reasoning similar to Equation (39), we get

$$\bar{n}_{ij}^{\mathrm{h}} = -\tfrac{1}{\beta} \sum_{k,l\in\mathcal{V}} \mu_k^{\mathrm{h,in}} \sigma_k^{\mathrm{in}} \mu_l^{\mathrm{h,out}} \sigma_l^{\mathrm{out}} \frac{\partial z_{kl}^{\mathrm{h}}}{\partial c_{ij}}.$$

This time, we have (Kivimäki et al., 2016, Equation (11)),

$$\frac{\partial z_{kl}^{\mathrm{h}}}{\partial c_{ij}} = \frac{\partial\,(z_{kl}/z_{ll})}{\partial c_{ij}} = -\beta w_{ij}\left(\frac{z_{ki}z_{jl}}{z_{ll}} - \frac{z_{kl}z_{li}z_{jl}}{z_{ll}^2}\right),$$

leading to

$$\bar{n}_{ij}^{\mathrm{h}} = \sum_{k,l\in\mathcal{V}} \frac{\mu_k^{\mathrm{h,in}} \sigma_k^{\mathrm{in}} \mu_l^{\mathrm{h,out}} \sigma_l^{\mathrm{out}}}{z_{ll}} \left(z_{ki} - \frac{z_{kl}z_{li}}{z_{ll}}\right) w_{ij}z_{jl}. \tag{59}$$

From Equation (56), and recalling that $z_{kl}^{\mathrm{h}} = z_{kl}/z_{ll}$, we get $(\mu_k^{\mathrm{h,in}} \sigma_k^{\mathrm{in}} \mu_l^{\mathrm{h,out}} \sigma_l^{\mathrm{out}})/z_{ll} = \gamma_{kl}^{\mathrm{h}}/z_{kl}$, where γ_{kl}^{h} is element k, l of the coupling matrix (see Equation (56)), thus

$$\bar{n}_{ij}^{\mathrm{h}} = \sum_{k,l\in\mathcal{V}} \gamma_{kl}^{\mathrm{h}} \left(\frac{z_{ki}}{z_{kl}} - \frac{z_{li}}{z_{ll}}\right) w_{ij}z_{jl} = \sum_{k,l\in\mathcal{V}} \gamma_{kl}^{\mathrm{h}} \bar{n}_{ij}^{(kl)}, \tag{60}$$

and $\bar{n}_{ij}^{(kl)} \triangleq \left(\frac{z_{ki}}{z_{kl}} - \frac{z_{li}}{z_{ll}}\right) w_{ij}z_{jl}$ is the expected number of times (i, j) is visited when the starting node is fixed to k and the ending node to l, as shown by Kivimäki et al. (2016, Equation (12)). Within the constrained bag-of-hitting-paths formalism, this quantity is simply the average over all starting and ending nodes, weighted by the coupling probabilities, $\boldsymbol{\Gamma}_{\mathrm{h}}$. In fact, $\bar{n}_{ij}^{h}$ can be seen as a *weighted RSP betweenness centrality* for edges, compared to the unweighted RSP betweenness centrality defined by Kivimäki et al. (2016). More precisely, this quantity provides a weighted *group betweenness* between the two sets of nodes, $\mathcal{I}n$ and $\mathcal{O}ut$.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

While the decomposition (60) of $\bar{n}^{\mathrm{h}}_{ij}$ provides an interesting analogy, we will use another expression to actually compute this quantity. From Equation (59) with the help of Equations (53) and (54), we get

$$\bar{n}^{\mathrm{h}}_{ij} = \left(\frac{1}{\mu_i^{\mathrm{h,out}} \mu_j^{\mathrm{h,in}}} - \sum_{l \in \mathcal{V}} \sigma_l^{\mathrm{out}} z^{\mathrm{h}}_{li} z^{\mathrm{h}}_{jl} \right) z_{ii} w_{ij}, \tag{61}$$

which provides the expression for computing $\bar{\mathbf{N}}_{\mathrm{h}} \triangleq (\bar{n}^{\mathrm{h}}_{ij})$. □

3.4.4 The expected number of visits to a node

Proposition 3.6. *The vector* $\bar{\mathbf{n}}_{\mathrm{h}}$*, containing the expected number of times node i appears on a hitting path drawn from a bag of hitting paths, is provided by*

$$\bar{\mathbf{n}}_{\mathrm{h}} = \mathbf{Diag}(\mathbf{Z}) \left[\left(\mathbf{e} \div \left(\boldsymbol{\mu}^{\mathrm{h}}_{\mathrm{in}} \circ \boldsymbol{\mu}^{\mathrm{h}}_{\mathrm{out}} \right) \right) - \mathbf{diag}(\mathbf{Z}_{\mathrm{h}} \mathbf{Diag}(\boldsymbol{\sigma}_{\mathrm{out}}) \mathbf{Z}_{\mathrm{h}}) \right] + \boldsymbol{\sigma}_{\mathrm{out}}, \tag{62}$$

where $\mathbf{diag}(\mathbf{X})$ *is a column vector containing the main diagonal of matrix* $\mathbf{X}$*.*

Proof. Let us define $\bar{n}^{\mathrm{h}}_i$, the expected number of times node i appears on a path under $\mathrm{P_h}(\wp)$, by

$$\bar{n}^{\mathrm{h}}_i \triangleq \sum_{\wp \in \mathcal{P}_{\mathrm{h}}} \mathrm{P_h}(\wp)\, \eta(i \in \wp),$$

where $\eta(i \in \wp)$ denotes the number of times node i is visited along hitting path $\wp$. Using Equation (43) and (44) again, we get

$$\bar{n}^{\mathrm{h}}_i = \sum_{j \in \mathcal{V}} \bar{n}^{\mathrm{h}}_{ij} + \sigma_i^{\mathrm{out}}, \tag{63}$$

which, after using Equation (60) and $\sum_{j \in \mathcal{V}} w_{ij} z_{jl} = z_{il} - \delta_{il}$ (Kivimäki et al., 2016, Equation (13)), results in

$$\bar{n}^{\mathrm{h}}_i = \sum_{k,l \in \mathcal{V}} \gamma^{\mathrm{h}}_{kl} \bar{n}_i^{(kl)} + \sigma_i^{\mathrm{out}},$$

where $\bar{n}_i^{(kl)} \triangleq \left(\frac{z_{ki}}{z_{kl}} - \frac{z_{li}}{z_{ll}} \right) z_{il}$ is the number of times i is visited when starting in k and ending in l, as defined in (Kivimäki et al., 2016). Notice that $\bar{n}_i^{(kl)} = 0$ when $i = l$ and when $k = l$. Again, this quantity can be seen as a *weighted RSP betweenness centrality* for sets of nodes, by analogy with the unweighted RSP betweenness centrality defined in (Kivimäki et al., 2016, Equation (15)).

However, for computation, it is more convenient to apply the equivalent of Equation (61). Following the same path as for the derivation of Equation (61), we first use $\sum_{j \in \mathcal{V}} w_{ij} z_{jl} = z_{il} - \delta_{il}$ with Equation (63) on Equation (59) and then Equations (53) and (54) in order to get

$$\bar{n}^{\mathrm{h}}_i = \left(\frac{1}{\mu_i^{\mathrm{h,out}} \mu_i^{\mathrm{h,in}}} - \sum_{l \in \mathcal{V}} \sigma_l^{\mathrm{out}} z^{\mathrm{h}}_{li} z^{\mathrm{h}}_{il} \right) z_{ii} + \sigma_i^{\mathrm{out}}, \tag{64}$$

which provides the expression for computing the vector $\bar{\mathbf{n}}_{\mathrm{h}} \triangleq (\bar{n}^{\mathrm{h}}_i)$. □

3.4.5 The optimal randomized policy

Proposition 3.7. *The biased random walk transition matrix,* $\mathbf{P}_{\mathrm{h}} = (p^{\mathrm{h}}_{ij})$*, that is, the randomized routing policy, is given by*

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

$$p^{\mathrm{h}}_{ij} \triangleq \frac{\bar{n}^{\mathrm{h}}_{ij}}{\sum_{k\in\mathcal{V}} \bar{n}^{\mathrm{h}}_{ik}} = \frac{\left(\frac{1}{\mu^{\mathrm{h,out}}_i \mu^{\mathrm{h,in}}_j} - \sum_{l\in\mathcal{V}} \sigma^{\mathrm{out}}_l z^{\mathrm{h}}_{li}\, z^{\mathrm{h}}_{jl}\right)}{\left(\frac{1}{\mu^{\mathrm{h,out}}_i \mu^{\mathrm{h,in}}_i} - \sum_{l\in\mathcal{V}} \sigma^{\mathrm{out}}_l z^{\mathrm{h}}_{li}\, z^{\mathrm{h}}_{il}\right)} w_{ij}. \tag{65}$$

Proof. We get the result by combining Equations (61), (63), and (64). □

Again, we observed experimentally that this quantity converges to the reference transition matrix of the reference random walk when $\beta \to 0$, i.e.,

$$p^{\mathrm{h}}_{ij} \xrightarrow[\beta\to 0]{} p^{\mathrm{ref}}_{ij}.$$

Conversely, when $\beta \to \infty$, the problem becomes an optimal transport on a graph problem (see Section 6.2). Note that, for convenience, in Algorithm 3, matrix $\mathbf{P}_{\mathrm{h}} \triangleq (p^{\mathrm{h}}_{ij})$ is computed thanks to

$$\mathbf{P}_{\mathrm{h}} = (\mathbf{Diag}(\bar{\mathbf{N}}_{\mathrm{h}}\mathbf{e}))^{-1}\bar{\mathbf{N}}_{\mathrm{h}}.$$

Let us now turn to a discussion of the relations between the proposed models and the regularized optimal transport problem.

4. The regularized optimal transport problem analogy

In this section, we will show that both non-hitting (see Equation (5)) and hitting (see Equation (48)) problems correspond to two different kinds of *regularization* for the *optimal transport problem* (Ahuja et al., 1993; Guex et al., 2017; Kantorovich 1942; Villani 2003, 2008). It therefore generalizes discrete entropy-regularized optimal transport problems (Wilson, 1970; Erlander & Stewart, 1990; Kapur & Kesavan, 1992) to a graph structure.

4.1 The optimal transport problem

General optimal transport is a well-known problem defined, for example, in (Kantorovich, 1942; Villani, 2003, 2008), and the special case where the space is a graph is easily derived from it (Guex et al., 2017). Assume we have a subset of nodes $In \in \mathcal{V}$, called *sources*, with a supply of a certain quantity of matter, while we observe a demand of the same matter in another subset of nodes, $\mathcal{O}ut \in \mathcal{V}$, called *targets*. We suppose that the overall supply is equal to the overall demand, thus these quantities on nodes can be represented, without loss of generality, by their proportion of the total. In other words, supply and demand are represented respectively by two discrete distribution vector $\boldsymbol{\sigma}_{\mathrm{in}}$ and $\boldsymbol{\sigma}_{\mathrm{out}}$, with $\sum_{i\in\mathcal{V}} \sigma^{\mathrm{in}}_i = \sum_{i\in\mathcal{V}} \sigma^{\mathrm{out}}_i = 1$ and $\sigma^{\mathrm{in}}_i, \sigma^{\mathrm{out}}_i \geq 0,\ \forall i \in \mathcal{V}$.

The goal of the optimal transport problem is to find an optimal *attribution plan* or optimal *coupling* (Villani, 2003, 2008), i.e., a $n \times n$ matrix $\boldsymbol{\Gamma}$, where γ_{ij} represents the proportion of matter going from i to j, in order to fulfill supply and demand. Optimality here means that the cost of transportation of this attribution plan, i.e., $\sum_{i,j\in\mathcal{V}} \gamma_{ij}\, c_{ij}$ where c_{ij} is the cost of transportation from i to j, must be minimal. Altogether, we have

$$\left|\begin{array}{lll} \underset{\boldsymbol{\Gamma}}{\text{minimize}} & \sum_{i,j\in\mathcal{V}} \gamma_{ij}\, c_{ij} & \\ \text{subject to} & \sum_{j\in\mathcal{V}} \gamma_{ij} = \sigma^{\mathrm{in}}_i & \forall i \in \mathcal{V}, \\ & \sum_{i\in\mathcal{V}} \gamma_{ij} = \sigma^{\mathrm{out}}_j & \forall j \in \mathcal{V}, \\ & \gamma_{ij} \geq 0 & \forall i,j \in \mathcal{V}. \end{array}\right. \tag{66}$$

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Another interesting interpretation can be found in the *dual optimal transport problem* (Guex et al., 2017; Villani 2003, 2008), expressed by

$$\begin{vmatrix} \underset{\boldsymbol{\lambda}_{\text{in}},\boldsymbol{\lambda}_{\text{out}}}{\text{maximize}} & \sum_{i\in\mathcal{V}} \lambda_i^{\text{in}}\sigma_i^{\text{in}} + \sum_{j\in\mathcal{V}} \lambda_j^{\text{out}}\sigma_j^{\text{out}} \\ \text{subject to} & \lambda_i^{\text{in}} + \lambda_j^{\text{out}} \leq c_{ij} \qquad \forall i,j \in \mathcal{V}. \end{vmatrix} \tag{67}$$

Here, the dual vectors $\boldsymbol{\lambda}_{\text{in}} = (\lambda_i^{\text{in}})$ and $\boldsymbol{\lambda}_{\text{out}} = (\lambda_i^{\text{out}})$ can be interpreted respectively as the dual *embarkment prices* on *sources* and *disembarkment prices* on *targets*, as shown in (Villani, 2003, 2008). This is a common property of the dual problem in linear programming (Griva et al., 2009).

4.2 The standard optimal transport flow on a graph problem

For completeness, let us recall the standard (exact) transport flow problem on a graph. The linear programming optimal transport flow problem is defined as (Ahuja et al., 1993)

$$\begin{vmatrix} \underset{\mathbf{X}}{\text{minimize}} & \mathbf{e}^\top(\mathbf{X}\circ\mathbf{C}')\mathbf{e} \\ \text{subject to} & \left(\mathbf{X}-\mathbf{X}^\top\right)\mathbf{e} = \boldsymbol{\sigma}_{\text{in}} - \boldsymbol{\sigma}_{\text{out}}, \\ & x_{ij} \geq 0 \qquad \forall i,j \in \mathcal{V}, \end{vmatrix} \tag{68}$$

where $\mathbf{C}'$ is the cost matrix containing nonnegative costs on the edges of the network with infinite components set to 0. As before, it is assumed that input and output flows are nonnegative as well as $(\boldsymbol{\sigma}_{\text{in}})^\top\mathbf{e} = (\boldsymbol{\sigma}_{\text{out}})^\top\mathbf{e} = 1$. The idea is therefore to minimize the total cost of flows while satisfying the input and output constraints. The solution $\mathbf{X} = (x_{ij})$ of this problem corresponds to the matrix containing directed flows on the edges.

4.3 The regularized optimal transport problem

In order to show that the bag-of-paths formalism is closely related to the optimal transport problem, we need to compute again the minimum free energy value in terms of the elements of the coupling matrix γ_{ij}, instead of the Lagrange parameters $\lambda_i^{\text{in}}, \lambda_j^{\text{out}}$ as in Equation (37).

Let us start with the regular, non-hitting paths model. By inserting the form taken by the optimal path probability distribution $\text{P}^\star(\wp)$, given by Equation (23), in the free energy functional of Equation (5) (similarly to the proof following Proposition 2.6), we obtain

$$\text{FE}(\text{P}^\star) = T\sum_{i,j\in\mathcal{V}} \log\left(\mu_i^{\text{in}}\mu_j^{\text{out}}\right) \underbrace{\sum_{\wp_{ij}\in\mathcal{P}_{ij}} \text{P}^\star(\wp_{ij})}_{\gamma_{ij}},$$

which directly provides from $\mu_i^{\text{in}}\mu_j^{\text{out}} = \gamma_{ij}/(\sigma_i^{\text{in}}\alpha_j\widehat{z}_{ij})$ (see Equation (36)) and $\gamma_{ij} = \sum_{\wp_{ij}\in\mathcal{P}_{ij}} \text{P}^\star(\wp_{ij})$ (Equation (33))

$$\text{FE}(\text{P}^\star) = -T\sum_{i,j\in\mathcal{V}} \log\left(\widehat{z}_{ij}\right)\gamma_{ij} + T\sum_{i,j\in\mathcal{V}} \gamma_{ij}\log\left(\frac{\gamma_{ij}}{\sigma_i^{\text{in}}\alpha_j}\right). \tag{69}$$

Following the same reasoning for the hitting paths formalism, we obtain

$$\text{FE}_{\text{h}}(\text{P}_{\text{h}}^\star) = \sum_{i,j\in\mathcal{V}} \underbrace{-T\log\left(z_{ij}^{\text{h}}\right)}_{\phi_T(i,j)}\gamma_{ij}^{\text{h}} + T\sum_{i,j\in\mathcal{V}} \gamma_{ij}^{\text{h}}\log\left(\frac{\gamma_{ij}^{\text{h}}}{\sigma_i^{\text{in}}\sigma_j^{\text{out}}}\right). \tag{70}$$

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

This shows that, instead of working with the whole probability distribution as required by Equation (5), it is sufficient to directly compute the elements of the coupling matrix γ_{ij} if the quantities $\phi_T(i,j)$ are precomputed.

Indeed, the resulting expression in the hitting formalism has a nice interpretation. In (Kivimäki et al., 2014), it was shown that in the simple RSP framework, $\phi_T(i,j) \triangleq \mathrm{FE}_{\mathrm{h}}(\mathrm{P}_{\mathrm{h}}^{\star}) = -T \log(z_{ij}^{\mathrm{h}})$ is known to be the minimum free energy when problem (2) is restricted to hitting paths connecting a single source i to a single destination j. This quantity corresponds to the pairwise directed *free energy distance* between nodes of a graph introduced in (Kivimäki et al., 2014) and (Françoisse et al., 2017), where it is proved that it is a distance metric. This distance provided competitive results in pattern recognition tasks (Françoisse et al., 2017; Sommer et al., 2016, 2017).

Moreover, it is shown in (Françoisse et al., 2017) that $\phi_T(i,j)$ converges to the directed *shortest path distance* between i and j when $T \to 0$ and to the *average first passage time* (up to a scaling factor) between i and j when $T \to \infty$. The free energy distances between all pairs of nodes can easily be computed in matrix form (Kivimäki et al., 2014; Françoisse et al., 2017; Fouss et al., 2016). It has further been shown that, when computing the continuous time—continuous state equivalent to the RSP model by densifying the graph, the minimum free energy becomes a *potential* attracting the agents to the goal state (García-Díez et al., 2011).

From Equation (70), once the directed free energy distances have been computed, we observe that problem (48) can therefore be restated as

$$\begin{array}{ll} \underset{\mathbf{\Gamma_h}}{\text{minimize}} & \sum_{i,j\in\mathcal{V}} \phi_T(i,j)\gamma_{ij}^{\mathrm{h}} + T\sum_{i,j\in\mathcal{V}} \gamma_{ij}^{\mathrm{h}} \log\left(\dfrac{\gamma_{ij}^{\mathrm{h}}}{\sigma_i^{\mathrm{in}}\sigma_j^{\mathrm{out}}}\right) \\ \text{subject to} & \sum_{j\in\mathcal{V}} \gamma_{ij}^{\mathrm{h}} = \sigma_i^{\mathrm{in}} \qquad \forall i \in \mathcal{V}, \\ & \sum_{i\in\mathcal{V}} \gamma_{ij}^{\mathrm{h}} = \sigma_j^{\mathrm{out}} \qquad \forall j \in \mathcal{V}, \\ & \gamma_{ij}^{\mathrm{h}} \geq 0 \qquad\qquad \forall i,j \in \mathcal{V}. \end{array} \tag{71}$$

Knowing that the directed free energy distance $\phi_T(i,j)$ converges to the directed shortest path distance when $T \to 0$, we conclude that problem (71) reduces to the optimal transport problem at this limit. Thus, problem (71) is actually a "soft" (entropy regularized) optimal transport problem, similar to the one studied in, e.g., (Wilson, 1970; Erlander & Stewart, 1990; Cuturi, 2013) and based on the directed free energy distance (Kivimäki et al., 2014; Françoisse et al., 2017; Fouss et al., 2016), monitored by temperature.

Therefore, in the case of hitting paths, an alternative way of solving the entropy-regularized optimal transport on a graph problem (48) is to *pre-compute the free energy distances* $\phi_T(i,j)$ and then solve problem (71) (see Cuturi, 2013 for a recent discussion).

The non-hitting problem is another regularization of the optimal transport problem as both Equations (69) and (70) converge to the same limit when $T \to 0$. As a matter of fact, when $T \to 0$, $\widehat{z}_{ij}$ and z_{ij}^{h} have the same limit, as the effect of α_i vanishes in the first case and $z_{jj} \to 1$ in the second case. In contrast, these two formalisms diverge when $T \to \infty$. The hitting paths formalism converges to the problem described in (Guex et al., 2017) and thus results in the trivial, independent, coupling $\gamma_{ij}^{\mathrm{h}} = \sigma_i^{\mathrm{in}}\sigma_j^{\mathrm{out}}$ when $T \to \infty$, while the non-hitting paths formalism provides a more interesting solution, though harder to interpret (see Equation (69)). Actually, following the derivations appearing so far in this paper and results discussed in (Guex, 2016), it appears that the high temperature limit of the non-hitting formalism corresponds, in the case of an undirected graph, to the *electrical circuit formalism*, where sources and targets correspond to nodes with potentials fixed by the user (a high potential on sources and a low potential on targets) (see Guex, 2016, for details). A study of this interesting question is left for further work.

Interestingly, Equations (37) and (57) give two different, alternative, expressions for the minimal free energy, which are equivalent to the objective function of the dual optimal transport

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

problem (with Lagrange multipliers corresponding to the dual variables multiplied by -1). It implies that, when $T \to 0$, $-\boldsymbol{\lambda}_{\text{in}}$ and $-\boldsymbol{\lambda}_{\text{out}}$ (and equivalently, $-\boldsymbol{\lambda}^{\text{h}}_{\text{in}}$ and $-\boldsymbol{\lambda}^{\text{h}}_{\text{out}}$) converge respectively to the dual embarkment prices on sources and disembarkment prices on targets. Therefore, a mapping of these variables on nodes can highlight problematic sources and targets, in terms of optimal transport (see, e.g., Guex et al., 2017).

4.4 The optimal transportation flow

Within the context of the optimal transport problem, it is interesting to discuss the interpretation of the matrix containing the expected number of visits to edges, i.e., $\bar{\mathbf{N}} = (\bar{n}_{ij})$. Note that the discussion is developed within the non-hitting formalism but remains valid in the hitting case.

In Section 2.5.4, the number of times node i appears on a path $\wp$ was decomposed in $\eta(i \in \wp) = \sum_{j \in \mathcal{V}} \eta(i \to j \in \wp) + \delta(E(\wp) = i)$ (Equation (43)), leading to $\bar{n}_i = \sum_{j \in \mathcal{V}} \bar{n}_{ij} + \sigma_i^{\text{out}}$ (Equation (44)). However, it is also possible to write

$$\eta(i \in \wp) = \sum_{j \in \mathcal{V}} \eta(j \to i \in \wp) + \delta(S(\wp) = i),$$

where $\delta(S(\wp) = i)$ is 1 iff $\wp$ begins with i and 0 otherwise. This second version results in having

$$\bar{n}_i = \sum_{j \in \mathcal{V}} \bar{n}_{ji} + \sigma_i^{\text{in}},$$

and combining this result with Equation (44), we obtain

$$\sum_{j \in \mathcal{V}} \bar{n}_{ij} - \sum_{j \in \mathcal{V}} \bar{n}_{ji} = \sigma_i^{\text{in}} - \sigma_i^{\text{out}}, \ \forall i \in \mathcal{V}.$$

This last equation shows that $\bar{n}_{ij}$ can be interpreted as a *directed flow on edges*, as shown in Guex (2016). This flow is emitted by sources, absorbed by targets, and conserved everywhere else. Mapping this flow allows us to analyse the transportation of matter along the edges of the graph and is illustrated in Section 6. In (Guex, 2016), it is shown that the *net flow* on an undirected graph, i.e., $\bar{n}_{ij} - \bar{n}_{ji}$, converges to the *electrical flow* for the non-hitting formalism when $T \to \infty$. Furthermore, in the hitting formalism, summing the absolute values of the net flows over the edges results in a RSP *net betweenness centrality* (Kivimäki et al., 2016). For an unweighted graph, the standard, unweighted RSP net betweenness converges to the *current flow betweenness* in the limit $T \to \infty$ (Newman, 2005; Brandes & Fleischer, 2005).

5. Derived distances and dissimilarities

Two general families of distances are derived from our framework: distances between nodes and dissimilarities between groups of nodes (both for the non-hitting and the hitting case).

5.1 Distances between nodes

Let us first discuss distances between nodes.

5.1.1 Definitions

For both the hitting and non-hitting formalism, we can now define a distance named the *surprisal distance*, generalizing the one introduced in (Françoisse et al., 2017; Kivimäki et al., 2014). The particularity here is that we can attach positive *weights* $\mathbf{w} = (w_i)$ (with $w_i > 0$ and summing to 1) to nodes, which affect the distances through $\boldsymbol{\sigma}_{\text{in}}$ and $\boldsymbol{\sigma}_{\text{out}}$. More precisely, for a strongly connected

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

graph, we define the *margin-constrained bag-of-paths surprisal distance* $\mathbf{\Delta}_{\text{sur}} = (\Delta_{ij}^{\text{sur}})$, and *margin-constrained bag-of-hitting-paths surprisal distance* $\mathbf{\Delta}_{\text{sur}}^{\text{h}} = (\Delta_{ij}^{\text{h,sur}})$ by, respectively,

$$\Delta_{ij}^{\text{sur}} \triangleq \begin{cases} -\frac{1}{2}(\log(\gamma_{ij}) + \log(\gamma_{ji})) & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases} \tag{72}$$

$$\Delta_{ij}^{\text{h,sur}} \triangleq \begin{cases} -\frac{1}{2}\left(\log\left(\gamma_{ij}^{\text{h}}\right) + \log\left(\gamma_{ji}^{\text{h}}\right)\right) & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases} \tag{73}$$

where $\mathbf{\Gamma} = (\gamma_{ij})$ and $\mathbf{\Gamma}_{\text{h}} = (\gamma_{ij}^{\text{h}})$ (the elements of the coupling matrix, see Sections 2.5.1 and 3.4.1) are obtained from, respectively, the non-hitting and hitting path formalisms with $\boldsymbol{\sigma}_{\text{in}} = \boldsymbol{\sigma}_{\text{out}} = \mathbf{w}$. From this definition, each node acts as a source and a target and larger weights induce a stronger influence over the graph, as the flows along the paths starting and ending in a particular node will scale accordingly.

Proposition 5.1. *The margin-constrained bag-of-paths surprisal distance,* $\mathbf{\Delta}_{\text{sur}}$*, defined in Equation (72), and the margin-constrained bag-of-hitting-paths surprisal distance,* $\mathbf{\Delta}_{\text{sur}}^{\text{h}}$ *defined in Equation (73), are both metrics.*

Proof. The triangle inequality for both surprisal distances, i.e., $\Delta_{ij}^{\text{sur}} \leq \Delta_{ik}^{\text{sur}} + \Delta_{kj}^{\text{sur}}$ is trivially proven if $i = k$, $i = k$ or $j = k$. So, we will assume here that $i \neq j \neq k \neq i$.

Non-hitting formalism. First, notice that the reasoning found in (Françoisse et al., 2017, Appendix B) is still valid with the non-hitting reference probabilities derived in Section 2.3, namely, we have $\widehat{\pi}^{\text{ref}}(\wp_{ij}) = \widehat{\pi}^{\text{ref}}(\wp_{ij}^{\text{h}})\widehat{\pi}^{\text{ref}}(\wp_{jj})$ and $\tilde{c}(\wp_{ij}) = \tilde{c}(\wp_{ij}^{\text{h}}) + \tilde{c}(\wp_{jj})$, where $\wp_{ij}^{\text{h}}$ is the hitting path consisting of the first part of $\wp_{ij}$, until it reaches j for the first time, and $\wp_{jj}$ is the remaining part of $\wp_{ij}$. Thus, we also have

$$\widehat{z}_{ij} = \widehat{z}_{ij}^{\text{h}}\widehat{z}_{jj}, \tag{74}$$

where $\widehat{z}_{ij}^{\text{h}} \triangleq \sum_{\wp_{ij}^{\text{h}} \in \mathcal{P}_{ij}^{\text{h}}} \widehat{\pi}^{\text{ref}}(\wp_{ij}^{\text{h}}) \exp(-\beta\tilde{c}(\wp_{ij}^{\text{h}}))$. Now, it is easy to see that, for the optimal path probabilities obtained in Equation (23),

$$\text{P}(S = i, E = j) \geq \text{P}(S = i, k \in \wp, E = j) \triangleq \sum_{\wp_{ij} \in \mathcal{P}_{ij}} \delta(k \in \wp_{ij})\text{P}^{\star}(\wp_{ij}), \tag{75}$$

where $\delta(k \in \wp_{ij})$ is equal to 1 if node k lies on path $\wp_{ij}$ and 0 otherwise. By developing with Equation (23), we obtain

$$\begin{aligned} \text{P}(S = i, k \in \wp, E = j) &= \mu_i^{\text{in}}\sigma_i^{\text{in}}\mu_j^{\text{out}}\alpha_j \sum_{\wp_{ij} \in \mathcal{P}_{ij}} \delta(k \in \wp_{ij})\widehat{\pi}^{\text{ref}}(\wp_{ij}) \exp(-\beta\tilde{c}(\wp_{ij})) \\ &= \mu_i^{\text{in}}\sigma_i^{\text{in}}\mu_j^{\text{out}}\alpha_j \sum_{\wp_{ik}^{\text{h}} \in \mathcal{P}_{ik}^{\text{h}}} \tilde{\pi}^{\text{ref}}(\wp_{ik}^{\text{h}}) \exp(-\beta\tilde{c}(\wp_{ik}^{\text{h}})) \sum_{\wp_{kj} \in \mathcal{P}_{kj}} \widehat{\pi}^{\text{ref}}(\wp_{kj}) \exp(-\beta\tilde{c}(\wp_{kj})) \\ &= \mu_i^{\text{in}}\sigma_i^{\text{in}}\mu_j^{\text{out}}\alpha_j \widehat{z}_{ik}^{\text{h}}\widehat{z}_{kj}, \end{aligned}$$

where each path from i to j is again cut in two sequential subpaths.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Then, using Equations (74) and (36) provides

$$\begin{aligned}\mathrm{P}(S=i,k\in\wp,E=j)&=\frac{\mu_i^{\mathrm{in}}\sigma_i^{\mathrm{in}}\mu_k^{\mathrm{out}}\alpha_k\widehat{z}_{ik}\mu_k^{\mathrm{in}}\sigma_k^{\mathrm{in}}\mu_j^{\mathrm{out}}\alpha_j\widehat{z}_{kj}}{\mu_k^{\mathrm{in}}\sigma_k^{\mathrm{in}}\mu_k^{\mathrm{out}}\alpha_k\widehat{z}_{kk}}\\&=\frac{\mathrm{P}(S=i,E=k)\mathrm{P}(S=k,E=j)}{\mathrm{P}(S=k,E=k)}.\end{aligned}$$

Using the inequality in Equation (75) and $\mathrm{P}(S=k,E=k)\leq 1$, we get

$$-\log\left(\mathrm{P}(S=i,E=j)\right)\leq-\log\left(\mathrm{P}(S=i,E=k)\right)-\log\left(\mathrm{P}(S=k,E=j)\right),$$

which proves the triangle inequality for the non-hitting surprisal distance for $i\neq j\neq k\neq i$.

Hitting formalism. The reasoning is similar to the previous case. First let us consider

$$\begin{aligned}\mathrm{P_h}(S=i,E=j)&=\mu_i^{\mathrm{h,in}}\sigma_i^{\mathrm{in}}\mu_j^{\mathrm{h,out}}\sigma_j^{\mathrm{out}}z_{ij}^{\mathrm{h}}=\mu_i^{\mathrm{h,in}}\sigma_i^{\mathrm{in}}\mu_j^{\mathrm{h,out}}\sigma_j^{\mathrm{out}}\frac{z_{ij}}{z_{jj}}\\&\geq\mu_i^{\mathrm{h,in}}\sigma_i^{\mathrm{in}}\mu_j^{\mathrm{h,out}}\sigma_j^{\mathrm{out}}\frac{\sum_{\wp_{ij}\in\mathcal{P}_{ij}}\delta(k\in\wp_{ij})\tilde{\pi}^{\mathrm{ref}}(\wp_{ij})\exp\left(-\beta\tilde{c}(\wp_{ij})\right)}{z_{jj}}.\end{aligned}\tag{76}$$

And again,

$$\begin{aligned}&\mu_i^{\mathrm{h,in}}\sigma_i^{\mathrm{in}}\mu_j^{\mathrm{h,out}}\sigma_j^{\mathrm{out}}\frac{\sum_{\wp_{ij}\in\mathcal{P}_{ij}}\delta(k\in\wp_{ij})\tilde{\pi}^{\mathrm{ref}}(\wp_{ij})\exp\left(-\beta\tilde{c}(\wp_{ij})\right)}{z_{jj}}\\&=\frac{\mu_i^{\mathrm{h,in}}\sigma_i^{\mathrm{in}}\mu_j^{\mathrm{h,out}}\sigma_j^{\mathrm{out}}}{z_{jj}}\sum_{\wp_{ik}^{\mathrm{h}}\in\mathcal{P}_{ik}^{\mathrm{h}}}\tilde{\pi}^{\mathrm{ref}}(\wp_{ik}^{\mathrm{h}})\exp\left(-\beta\tilde{c}(\wp_{ik}^{\mathrm{h}})\right)\sum_{\wp_{kj}\in\mathcal{P}_{kj}}\tilde{\pi}^{\mathrm{ref}}(\wp_{kj})\exp\left(-\beta\tilde{c}(\wp_{kj})\right)\\&=\frac{\mu_i^{\mathrm{h,in}}\sigma_i^{\mathrm{in}}\mu_k^{\mathrm{h,out}}\sigma_k^{\mathrm{out}}z_{ik}^{\mathrm{h}}\,\mu_k^{\mathrm{h,in}}\sigma_k^{\mathrm{in}}\mu_j^{\mathrm{h,out}}\sigma_j^{\mathrm{out}}z_{kj}^{\mathrm{h}}}{\mu_k^{\mathrm{h,in}}\sigma_k^{\mathrm{in}}\mu_k^{\mathrm{h,out}}\sigma_k^{\mathrm{out}}}=\frac{\mathrm{P_h}(S=i,E=k)\mathrm{P_h}(S=k,E=j)}{\mathrm{P_h}(S=k,E=k)},\end{aligned}$$

where we used $z_{kk}^{\mathrm{h}}=1$. With Equation (76) and $\mathrm{P_h}(S=k,E=k)\leq 1$, we finally obtain $-\log\left(\mathrm{P_h}(S=i,E=j)\right)\leq-\log\left(\mathrm{P_h}(S=i,E=k)\right)-\log\left(\mathrm{P_h}(S=k,E=j)\right)$, which shows the triangle inequality for the hitting surprisal distance for $i\neq j\neq k\neq i$. □

5.2 Distances between groups of nodes

A different family of dissimilarities naturally arises from the optimal transport interpretation of the margin-constrained BoP formalism, namely, dissimilarities between *groups of nodes*. These dissimilarities can be viewed as an extension of the *Wasserstein distance* between node distributions on a graph, also known under the name of the *Monge–Kantorovich distance* or the *earth mover distance* in the literature (see e.g. Dobrushin, 1970; Villani, 2003, 2008; Zolotarev, 1983).

This dissimilarity is defined as the total cost of transportation in order to move from the distribution on source nodes, $\boldsymbol{\sigma}_{\mathrm{in}}$, to the distribution on target nodes, $\boldsymbol{\sigma}_{\mathrm{out}}$. Similarly to the usual free energy distance described in (Kivimäki et al., 2014; Françoisse et al., 2017), interpolating between the shortest path distance and the commute cost distance (which is proportional to the resistance distance for undirected graphs), the margin-constrained BoP formalism uses the value of the free energy functional in order to derive a dissimilarity which interpolates between the Wasserstein distance and an electrical circuit-based dissimilarity between groups of node.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

5.2.1 Definitions

Let $\mathcal{G}$ be a directed, strongly connected, graph with n nodes, weighted by vector $\mathbf{w}$. Suppose we have p groups of nodes, and the *membership matrix* $\mathbf{M} = (m_{ig})$ with $m_{ig} \geq 0$ and $\sum_{g=1}^{p} m_{ig} = 1$ represents the membership degree of node i to group g (fuzzy memberships are allowed). From that, we can compute the *node distribution in group g*, $\boldsymbol{\sigma}_g = (\sigma_i^g)$, as , e.g.,

$$\sigma_i^g = \frac{w_i m_{ig}}{\sum_{j \in \mathcal{V}} w_j m_{jg}}.$$

Then, as for the standard free energy distance between two nodes (Kivimäki et al., 2014, see this paper for details), the *bag-of-paths free energy dissimilarity* between groups g and h is defined as the symmetrized minimum free energy between these two groups of nodes

$$\Delta_{gh}^{\mathrm{FE}} \triangleq \begin{cases} \frac{1}{2}(\mathrm{FE}^{\min}(\boldsymbol{\sigma}_g, \boldsymbol{\sigma}_h) + \mathrm{FE}^{\min}(\boldsymbol{\sigma}_h, \boldsymbol{\sigma}_g)) & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases} \tag{77}$$

and the *bag-of-hitting-paths free energy dissimilarity* between groups g and h by

$$\Delta_{gh}^{\mathrm{h,FE}} \triangleq \begin{cases} \frac{1}{2}(\mathrm{FE}_{\mathrm{h}}^{\min}(\boldsymbol{\sigma}_g, \boldsymbol{\sigma}_h) + \mathrm{FE}_{\mathrm{h}}^{\min}(\boldsymbol{\sigma}_h, \boldsymbol{\sigma}_g)) & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases} \tag{78}$$

where $\mathrm{FE}^{\min}(\boldsymbol{\sigma}_{\mathrm{in}}, \boldsymbol{\sigma}_{\mathrm{out}})$ is the non-hitting free energy (see Equation (37)), and $\mathrm{FE}_{\mathrm{h}}^{\min}(\boldsymbol{\sigma}_{\mathrm{in}}, \boldsymbol{\sigma}_{\mathrm{out}})$ the hitting free energy (see Equation (57)), with starting and ending node flows respectively equal to $\boldsymbol{\sigma}_{\mathrm{in}}$ and $\boldsymbol{\sigma}_{\mathrm{out}}$. By definition of the free energy, we are sure that this quantity is always positive. When $T \to 0$, the dissimilarity between groups g and h will yield the *optimal cost of transportation* from group g to h and from group h to g, which is obviously a metric (Dobrushin, 1970; Villani, 2003, 2008; Zolotarev, 1983). It is, however, possible that this dissimilarity is not a metric anymore for other values of T.

From a computational point of view, there exists an important difference between the bag-of-hitting-paths and the bag-of-paths algorithms computing their respective free energy dissimilarities. As a matter of fact, in the hitting formalism, the matrix $\mathbf{Z}_{\mathrm{h}}$ is only computed once, and dissimilarities between each pair of groups can be obtained afterward by solely changing the values of $\boldsymbol{\sigma}_{\mathrm{in}}$ and $\boldsymbol{\sigma}_{\mathrm{out}}$ in the iterative procedure defined by Equations (53) and (54). It is however impossible to proceed that way for the non-hitting formalism, as the computation of $\widehat{\mathbf{Z}}$ requires the values of $\boldsymbol{\sigma}_{\mathrm{in}}$ and $\boldsymbol{\sigma}_{\mathrm{out}}$. Therefore, the bag-of-hitting-paths distances are obtained in a more efficient way than the bag-of-paths distances.

6. Some illustrations

Although the main contribution of this work lies in the theoretical development of the margin-constrained bag-of-paths models, we provide here an illustration of the algorithms on a toy example. The bag-of-paths formalism (as well as the bag-of-hitting-paths, as they converge to the same solution when $T \to 0$) defines an efficient way to find an approximate solution of the transportation problem on a graph. Moreover, by varying the temperature of the model, we can add uncertainty to the optimal paths and offer a flexible, stochastic, alternative to the optimal solution. With this feature, we can, e.g., pinpoint target nodes for which the optimal source coupling is most unclear. In a practical setting, when focusing on a deterministic transportation policy, this can help in evaluating the importance of each source-target coupling decision. Unlike other efficient optimal transport solvers, the algorithm here provides not only the coupling between pairs of nodes but also the flows on edges. Knowing the most frequented edges could be a major asset for real-life applications, for example, in order to forecast network traffic. This section illustrates this idea on a toy graph (a 10×10 lattice). In addition, we evaluate the computational efficiency of Algorithms 2 and 3 by comparing the computation times with a baseline linear solver on lattices of different sizes.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

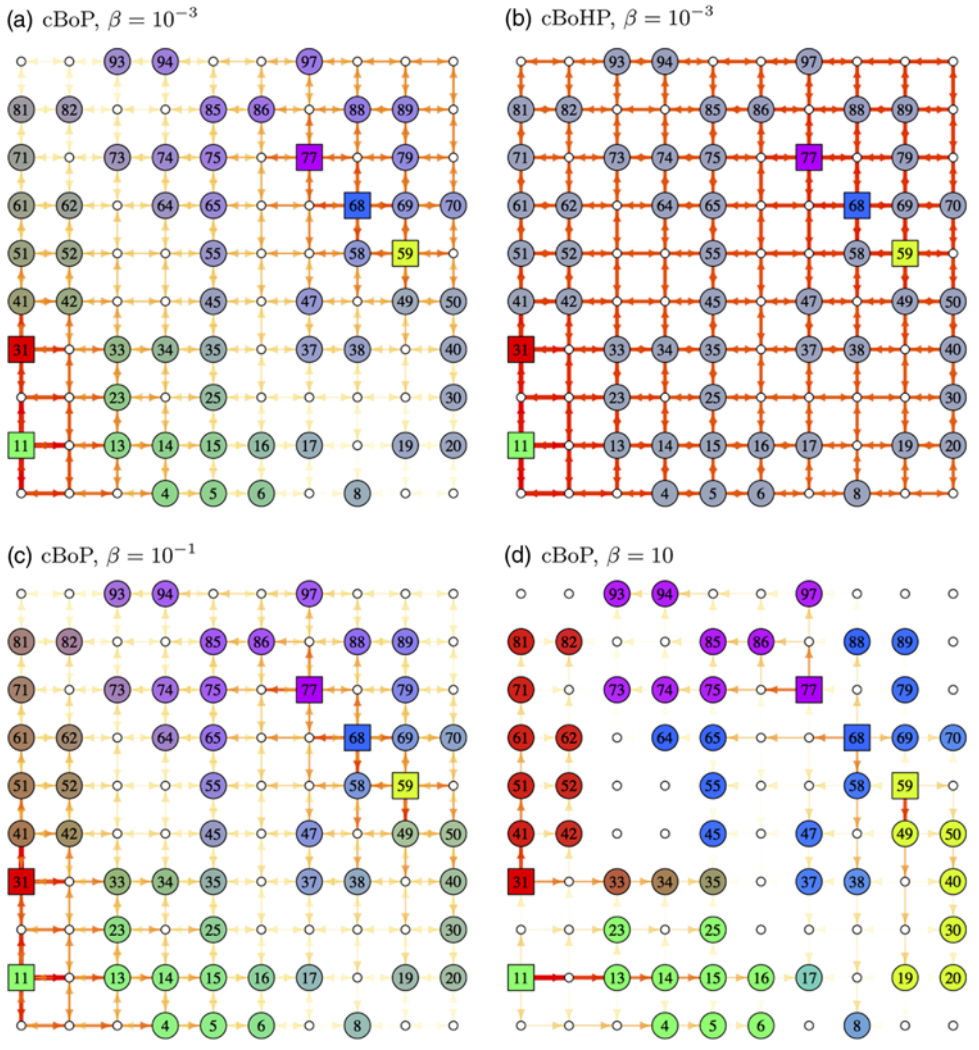


Figure 1. Resulting coupling $\boldsymbol{\Gamma}$ and edge flows $\bar{\mathbf{N}}$ for the constrained bag-of-paths (cBoP) and the constrained bag-of-hitting–paths (cBoHP) models, for different values of β. Source nodes are represented by a square and targets by a circle. $\boldsymbol{\Gamma}$ is shown on targets by a color interpolation representing their membership to corresponding sources (the membership of target j to source i is $\mathrm{P}(S = i|E = j) = \gamma_{ij}/\sigma_j^{\mathrm{out}}$). $\bar{\mathbf{N}}$ is represented on edges.

6.1 Illustrations on a 10 × 10 *lattice*

In this illustrative example, 5 random nodes were picked with $\sigma_i^{\mathrm{in}} = 0.2$ and 50 others with $\sigma_i^{\mathrm{out}} = 0.02$. The resulting coupling Γ and flows on edges $\bar{\mathbf{N}}$ are represented in Figure 1 for the constrained bag-of-paths (cBoP) and the constrained bag-of-hitting-paths (cBoHP), with different values of β. In this figure, target nodes j are colored to represent their distribution of membership over sources, i.e., $\mathrm{P}(S = i|T = j) = \gamma_{ij}/\sigma_j^{\mathrm{out}}$ and edge colors display the flows, $\bar{\mathbf{N}}$.

We observe that results obtained by the bag-of-paths model and the bag-of-hitting-paths model are quite different for $\beta = 10^{-3}$ (Figure 1, top row): the bag-of-paths model displays a behavior similar to a diffusive process, with edges near sources drawn more frequently, which is known to be similar to the electrical current (Ahuja et al., 1993; Doyle & Snell, 1984; Guex, 2016). On the other hand, the bag-of-hitting-paths solution for $\beta = 10^{-3}$ is quite trivial, with a uniform distribution of memberships of every target j to sources and the flow almost similar on every edge. In contrast, when the temperature is low, both models converge to the same solution and,

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

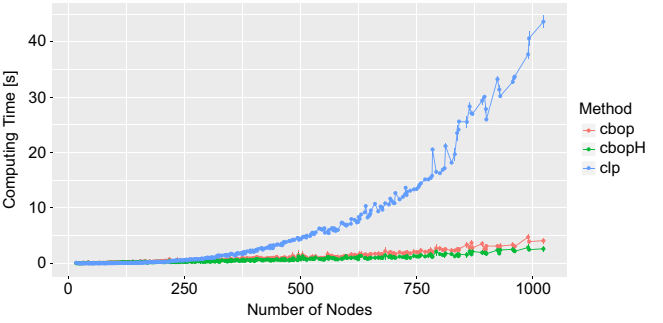


Figure 2. Computation times for, respectively, the open-source "Coin-or" linear programming solver (`clp`, blue), the margin–constrained bag-of-paths algorithm (`cbop`, red), and the margin-constrained bag-of-hitting-paths algorithm (`cbopH`, green) on lattices containing n nodes (x-axis), $\lfloor n/3 \rfloor$ sources, and $\lfloor n/3 \rfloor$ targets with uniform weights.

to avoid redundancy, only the bag-of-paths model is shown here (Figure 1, bottom row). With $\beta = 10$, this model displays an optimal transport solution, with only shortest paths followed and almost deterministic distributions of targets-to-sources memberships.

Therefore, in the present problem, the constrained bag-of-hitting-paths is perhaps less useful than the constrained bag-of-paths when the parameter β is close to zero. However, this depends on the application at hand and, essentially, on the desired behavior of the system when $\beta \to 0$, either the solution of an electrical circuit or the independence between sources and destinations.

6.2 Comparison of computation time

To assess the computation time of the constrained bag-of-paths and bag-of-hitting-paths algorithms, we compare them to the open-source *Computational Infrastructure for Operations Research* (Coin-or) linear programming solver (`clp`) written in C++ (Lougee-Heimer, 2003), which is considered as an efficient baseline algorithm for finding the coupling and the flow of the exact optimal transport problem (see Equation (68)). We run the algorithms on lattices of various dimensions in order to increase the number of nodes n. The number of source nodes and target nodes are set to be both $\lfloor n/3 \rfloor$ (rounded down) with similar weights, and their locations are randomly selected.

The results are presented in Figure 2. We can observe that on large graphs, both hitting and non-hitting bag-of-paths algorithms perform much faster than the linear programming baseline, with a slight advantage for the hitting algorithm. This was already observed by Cuturi (2013) for entropy-regularized optimal transport problems. All results were obtained with Julia (version 0.5.0) running on an Intel Xeon with 2×8 3.6 GHz processors and 128 GB of RAM.

7. Conclusion

This work extends the bag-of-paths framework introduced in (Mantrach et al., 2010; Françoisse et al., 2017) by allowing the user to set constraints on the starting and ending nodes of paths, $\boldsymbol{\sigma}_{\text{in}}$ and $\boldsymbol{\sigma}_{\text{out}}$. Like its predecessor, this formalism is derived for two types of paths, non-hitting paths and hitting paths. It also depends on a user-defined parameter, the temperature $T > 0$, according to which the model interpolates between a deterministic optimal policy and a completely random behavior. Both the non-hitting and hitting paths formalisms allow the computation of various quantities: the coupling, $\gamma_{ij} = \mathrm{P}(S = i, E = j)$; the expected number of times a node appears on the paths (a betweenness value), $\bar{n}_{ij}$; and the optimal policy defining a biased random walk with transition probabilities p_{ij}. All these quantities are expressed in terms of three computational elements: a fundamental matrix $\mathbf{Z}$, also found in (Françoisse et al., 2017), and Lagrange multipliers $\boldsymbol{\lambda}_{\text{in}}$ and $\boldsymbol{\lambda}_{\text{out}}$.

The addition of the set of constraints over starting and ending node distributions adds flexibility to its unconstrained predecessor and yields interesting connections with other models. When $T \to 0^+$, both the non-hitting and hitting formalisms are shown to be similar and converge to a

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

solution of the optimal transport on a graph problem. Unlike most algorithms solving the transportation problem, both BoP formalisms not only give sources-to-targets attributions, i.e., the coupling γ_{ij}, but also corresponding embarkment and disembarkment prices (with $\boldsymbol{\lambda}_{\text{in}}$ and $\boldsymbol{\lambda}_{\text{out}}$) and the flow on edges ($\bar{n}_{ij}$) while running with a competitive computation time compared to a baseline linear solver. In contrast, when $T \to 0$, each formalism behaves differently, each having its own merits. The non-hitting formalism converges to the electrical solution, with starting and ending node distributions corresponding to different potentials defined on nodes, and the hitting formalism, which is faster to compute, converges to the trivial, independent coupling.

These constraints also enlarge the range of applications of the BoP formalism, and it was shown here how to derive two families of dissimilarities from it. The first family of dissimilarities is defined as the surprisal distance between nodes, and constraints on starting and ending nodes provide a way to associate weights on nodes. The second family of dissimilarities is the free energy dissimilarity between groups of nodes. For the moment, these dissimilarities are quite theoretical and their applications are not explored in this paper. However, future research will investigate the use of these new dissimilarities in semi-supervised classification, hierarchical clustering, as well as other applications.

Generally speaking, the flexibility and the richness of this model could lead to different use cases, and future investigations will aim at finding various applications of the different introduced quantities. An on-going study will also investigate the introduction of flow constraints in the bag-of-paths framework.

Acknowledgments. This work was partially supported by the Immediate and the Brufence projects funded by InnovIris (Brussels Region), as well as former projects funded by the Walloon region, Belgium. Ilkka Kivimäki was partially funded by Emil Aaltonen Foundation, Finland. We thank these institutions for giving us the opportunity to conduct both fundamental and applied research. We also thank the anonymous reviewers and the editor whose remarks allowed to improve significantly the manuscript.

Conflict of interest. G. Guex, Ilkka Kivimäki and Marco Saerens have nothing to disclose.

Note

1 Alternatively, it can also be obtained by following a maximum entropy argument (Jaynes, 1957; Cover & Thomas, 2006; Kapur & Kesavan, 1992).

2 Note that nonuniform prior probabilities on the starting and ending node are briefly discussed in (Françoisse et al., 2017).

References

Ahuja, R. K., Magnanti, T. L., & Orlin, J. B. (1993). *Network flows: Theory, algorithms, and applications*. Prentice Hall.

Akamatsu, T. (1996). Cyclic flows, Markov process and stochastic traffic assignment. *Transportation Research B*, *30*(5), 369–386.

Alamgir, M., & von Luxburg, U. (2011). Phase transition in the family of p-resistances. In *Advances in neural information processing systems 24: Proceedings of the NIPS '11 conference* (pp. 379–387).

Bacharach, M. (1965) Estimating nonnegative matrices from marginal data. *International Economic Review*, *6*(3), 294–310.

Barabási, A.-L. (2016) *Network science*. Cambridge University Press.

Bavaud, F., & Guex, G. (2012). Interpolating between random walks and shortest paths: A path functional approach. In *International conference on social informatics* (pp. 68–81).

Brandes, U., & Fleischer, D. (2005). Centrality measures based on current flow. In *Proceedings of the 22nd annual symposium on theoretical aspects of computer science (STACS '05)* (pp. 533–544).

Chebotarev, P. (2011). A class of graph-geodetic distances generalizing the shortest-path and the resistance distances. *Discrete Applied Mathematics*, *159*(5), 295–302.

Chebotarev, P. (2012) The walk distances in graphs. *Discrete Applied Mathematics*, *160*(10–11), 1484–1500.

Chebotarev, P. (2013). Studying new classes of graph metrics. In F. Nielsen & F. Barbaresco (Eds.), *Proceedings of the 1st international conference on geometric science of information (GSI '13)* (vol. 8085, pp. 207–214).

Chung, F. R., & Lu, L. (2006). *Complex graphs and networks*. American Mathematical Society.

Courty, N., Flamary, R., Tuia, D., & Rakotomamonjy, A. (2017). Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(9), 1853–1865.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). John Wiley and Sons.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems 26: Proceedings of the NIPS '13 conference* (pp. 2292-2300). MIT Press.

Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, *15*(3), 458–486.

Doyle, P. G., & Snell, J. L. (1984). *Random walks and electric networks*. The Mathematical Association of America.

Erlander, S., & Stewart, N. (1990). *The gravity model in transportation analysis. Theory and extensions*. VSP International Science Publishers.

Estrada, E. (2012). *The structure of complex networks*. Oxford University Press.

Fang, S., Rajasekera, J., & Tsao, H. (1997). *Entropy optimization and mathematical programming*. Springer.

Ferradans, S., Papadakis, N., Peyré, G., & Aujol, J.-F. (2014) Regularized discrete optimal transport. *SIAM Journal on Imaging Sciences*, *7*(3), 1853–1882.

Fouss, F., Saerens, M., & Shimbo, M. (2016). *Algorithms and models for network data and link analysis*. Cambridge University Press.

Françoisse, K., Kivimäki, I., Mantrach, A., Rossi, F., & Saerens, M. (2017). A bag-of-paths framework for network data analysis. *Neural Networks*, *90*, 90–111.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41.

García-Díez, S., Vandenbussche, E., & Saerens, M. (2011). A continuous-state version of discrete randomized shortest-paths. In *Proceedings of the 50th IEEE international conference on decision and control (IEEE CDC 2011)* (pp. 6570–6577).

Graybill, F. (1983). *Matrices with applications in statistics*. Wadsworth International Group.

Grinstead, C., & Snell, J. L. (1997). *Introduction to probability* (2nd ed.). The Mathematical Association of America.

Griva, I., Nash, S. G., & Sofer, A. (2009). *Linear and nonlinear optimization: Second edition*. Society for Industrial and Applied Mathematics (SIAM).

Guex, G. (2016). Interpolating between random walks and optimal transportation routes: Flow with multiple sources and targets. *Physica A: Statistical Mechanics and its Applications*, *450*, 264–277.

Guex, G., & Bavaud, F. (2015). Flow-based dissimilarities: shortest path, commute time, max-flow and free energy. In B. Lausen, S. Krolak-Schwerdt, & M. Bohmer (Eds.), *Data science, learning by latent structures, and knowledge discovery* (vol. 1564, pp. 101–111). Springer.

Guex, G., Emmanouilidis, T., & Bavaud, F. (2017). *Transportation clustering: A regularized version of the optimal transportation problem.* (Submitted for publication)

Hara, K., Suzuki, I., Shimbo, M., Kobayashi, K., Fukumizu, K., & Radovanovic, M. (2015). Localized centering: Reducing hubness in large-sample data. In *Proceedings of the 29th AAAI conference on artificial intelligence (AAAI '15)* (pp. 2645–2651).

Hashimoto, T., Sun, Y., & Jaakkola, T. (2015). From random walks to distances on unweighted graphs. In *Advances in neural information processing systems 28: Proceedings of the NIPS '15 conference* (pp. 3429–3437). MIT Press.

Herbster, M., & Lever, G. (2009). Predicting the labelling of a graph via minimum p-seminorm interpolation. In *Proceedings of the 22nd conference on learning theory (COLT '09)* (pp. 18–21).

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*, 620–630.

Kantorovich, L. V. (1942). On the translocation of masses. *Doklady Akademii Nauk SSSR*, *37*(7–8), 227–229.

Kapur, J. N. (1989). *Maximum-entropy models in science and engineering*. Wiley.

Kapur, J. N., & Kesavan, H. K. (1992). *Entropy optimization principles with applications*. Academic Press.

Kivimäki, I., Lebichot, B., Saramäki, J., & Saerens, M. (2016). Two betweenness centrality measures based on randomized shortest paths (Scientific Reports, 6, srep19668).

Kivimäki, I., Shimbo, M., & Saerens, M. (2014). Developments in the theory of randomized shortest paths with a comparison of graph node distances. *Physica A: Statistical Mechanics and its Applications*, *393*, 600–616.

Klein, D. J., & Randić, M. (1993). Resistance distance. *Journal of Mathematical Chemistry*, *12*(1), 81–95.

Kolaczyk, E. D. (2009). *Statistical analysis of network data: Methods and models*. Springer.

Kurras, S. (2015). Symmetric iterative proportional fitting. In *Proceedings of the 18th international conference on artificial intelligence and statistics (AISTATS)* (vol. 38, pp. 526–534).

Lebichot, B., Kivimäki, I., Françoisse, K., & Saerens, M. (2014). Semi-supervised classification through the bag-of-paths group betweenness. *IEEE Transactions on Neural Networks and Learning Systems*, *25*, 1173–1186.

Lebichot, B., & Saerens, M. (2018). A bag-of-paths node criticality measure. *Neurocomputing*, *275*, 224–236.

Lewis, T. G. (2009). *Network science*. Wiley.

Li, Y., Zhang, Z.-L., & Boley, D. (2013). From shortest-path to all-path: The routing continuum theory and its applications. *IEEE Transactions on Parallel and Distributed Systems*, *25*(7), 1745–1755.

Lougee-Heimer, R. (2003). The Common Optimization INterface for Operations Research: Promoting open-source software in the operations research community. *IBM Journal of Research and Development*, *47*(1), 57–66. doi: 10.1147/rd.471.0057.

Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, *390*, 1150–1170.

Mantrach, A., Yen, L., Callut, J., Françoisse, K., Shimbo, M., & Saerens, M. (2010). The sum-over-paths covariance kernel: A novel covariance between nodes of a directed graph. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(6), 1112–1126.

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press

Mantrach, A., Zeebroeck, N. V., Francq, P., Shimbo, M., Bersini, H., & Saerens, M. (2011). Semi-supervised classification and betweenness computation on large, sparse, directed graphs. *Pattern Recognition*, *44*(6), 1212–1224.
Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, *27*(1), 39–54.
Newman, M. E. (2010). *Networks: An introduction*. Oxford University Press.
Nguyen, C. H., & Mamitsuka, H. (2016). New resistance distances with global information on large graphs. In *Proceedings of the 19th international conference on artificial intelligence and statistics (AISTATS)* (pp. 639–647).
Osborne, M. J. (2004). *An introduction to game theory*. Oxford University Press.
Pukelsheim, F. (2014). Biproportional scaling of matrices and the iterative proportional fitting procedure. *Annals of Operations Research*, *215*(1), 269–283.
Radovanović, M., Nanopoulos, A., & Ivanović, M. (2010a). Hubs in space: Popular nearest neighbors in high-dimensional data. *Journal of Machine Learning Research*, *11*, 2487–2531.
Radovanović, M., Nanopoulos, A., & Ivanović, M. (2010b). On the existence of obstinate results in vector space models. In *Proceedings of the 33rd annual international ACM SIGIR conference on research and development in information retrieval (SIGIR '10)* (pp. 186–193).
Saerens, M., Achbany, Y., Fouss, F., & Yen, L. (2009). Randomized shortest-path problems: Two related models. *Neural Computation*, *21*(8), 2363–2404.
Silva, T., & Zhao, L. (2016). *Machine learning in complex networks*. Springer.
Sinkhorn, R. (1967). Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, *74*(4), 402–405.
Solomon, J., Rustamov, R., Guibas, L., & Butscher, A. (2014). Wasserstein propagation for semi-supervised learning. In *Proceedings of the 31 international conference on machine learning (ICML '14)* (pp. 306–314).
Sommer, F., Fouss, F., & Saerens, M. (2016). Comparison of graph node distances on clustering tasks. In *Proceedings of the 25th international conference on artificial neural networks (ICANN 2016)* (vol. 9886, pp. 192–201).
Sommer, F., Fouss, F., & Saerens, M. (2017). Modularity-driven kernel k-means for community detection. In *Proceedings of the 26th international conference on artificial neural networks (ICANN 2017)* (vol. 10614, pp. 423–433).
Suzuki, I., Hara, K., Shimbo, M., Matsumoto, Y., & Saerens, M. (2012). Investigating the effectiveness of Laplacian-based kernels in hub reduction. In *Proceedings of the 26th AAAI conference on artificial intelligence (AAAI '12)* (pp. 1112–1118).
Suzuki, I., Hara, K., Shimbo, M., Saerens, M., & Fukumizu, K. (2013). Centering similarity measures to reduce hubs. In *Proceedings of the international conference on empirical methods in natural language processing (EMNLP 2013)* (pp. 613–623).
Thelwall, M. (2004). *Link analysis: An information science approach*. Elsevier.
Tomasev, N., Radovanovic, M., Mladenic, D., & Ivanovic, M. (2014). The role of hubness in clustering high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(3), 739–751.
Villani, C. (2003). *Topics in optimal transportation*. American Mathematical Society.
Villani, C. (2008). *Optimal transport: Old and new*. Springer.
von Luxburg, U., Radl, A., & Hein, M. (2010). Getting lost in space: Large sample analysis of the commute distance. In *Advances in neural information processing systems 23: Proceedings of the NIPS '10 conference* (pp. 2622–2630). MIT Press.
von Luxburg, U., Radl, A., & Hein, M. (2014). Hitting and commute times in large random neighborhood graphs. *Journal of Machine Learning Research*, *15*(1), 1751–1798.
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.
Wilson, A. (1970). *Entropy in urban and regional modelling*. Routledge.
Yen, L., Mantrach, A., Shimbo, M., & Saerens, M. (2008). A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '08)* (pp. 785–793).
Zhang, W., Zhao, D., & Wang, X. (2013). Agglomerative clustering via maximum incremental path integral. *Pattern Recognition*, *46*(11), 3056–3065.
Zolotarev, V. M. (1983). Probability metrics. *Teoriya Veroyatnostei i ee Primeneniya*, *28*(2), 264–287.

Cite this article: Guex G., Kivimäki I., and Saerens M. Randomized optimal transport on a graph: Framework and new distance measures. *Network Science* **7**, 88–122. https://doi.org/10.1017/nws.2018.29

https://doi.org/10.1017/nws.2018.29 Published online by Cambridge University Press