Physica A 450 (2016) 264–277

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Interpolating between random walks and optimal transportation routes: Flow with multiple sources and targets

Guillaume Guex

Department of Geography, University of Lausanne, Switzerland

HIGHLIGHTS

- New efficient algorithm to compute the optimal transportation flow.
- Source–target coupling corresponding to the flow.
- Strong mathematical formalism, linking shortest-path flow, random-walk flow and electrical current flow.
- Generalization of the randomized-shortest path with multiple sources and targets.
- Applications of the algorithm on case studies.

ARTICLE INFO

Article history:
Received 23 June 2015
Received in revised form 8 October 2015
Available online 12 January 2016

Keywords:
Functional minimization
Random walks
Optimal transportation problem
Multiple sources and targets

ABSTRACT

In recent articles about graphs, different models proposed a formalism to find a type of path between two nodes, the source and the target, at crossroads between the shortest-path and the random-walk path. These models include a freely adjustable parameter, allowing to tune the behavior of the path toward randomized movements or direct routes. This article presents a natural generalization of these models, namely a model with *multiple* sources and targets. In this context, source nodes can be viewed as locations with a supply of a certain good (e.g. people, money, information) and target nodes as locations with a demand of the same good. An algorithm is constructed to display the flow of goods in the network between sources and targets. With again a freely adjustable parameter, this flow can be tuned to follow routes of minimum cost, thus displaying the flow in the context of the *optimal transportation problem* or, by contrast, a random flow, known to be similar to the *electrical current flow* if the random-walk is reversible. Moreover, a *source–target* coupling can be retrieved from this flow, offering an optimal assignment to the transportation problem. This algorithm is described in the first part of this article and then illustrated with case studies.

1. Introduction

Over the years, the old *Monge–Kantorovich optimal transportation problem* has continued to show its usefulness and richness for numerous applications [1–5] (and many others). When presented in a graph setting [4], this problem consists in finding the optimal assignment of a resource supplied by a countable set of nodes, the *sources*, to another countable set of nodes, the *targets*, while minimizing the *cost* $U(X)$ of the transportation *flow* X. Different algorithms are capable of finding optimal solutions while respecting constraints of source supplies and target demands [6–8]. However, none are convenient to display routes of transportation, as it requires to solve the allocation problem first and, subsequently, run a shortest-path algorithm for every source–target pair. In this article, a new algorithm allowing the visualization of optimal transportation

E-mail address: guillaume.guex@unil.ch.

http://dx.doi.org/10.1016/j.physa.2015.12.117

routes is presented. This algorithm works with the well-known principle of *regularization*. By adding a suitable nonlinear functional to the linear functional $U(X)$, the resulting functional becomes derivable and an approximation of the optimal solution can be found with considerably less computing effort [9,10]. Here, this functional is named the *entropy* $G(X)$, as flows X minimizing it will display a random behavior, and the new functional $F(X) := U(X)+TG(X)$ is named the *free energy*, with a freely adjustable parameter $T > 0$, the *temperature*. Defining this new objective functional $F(X)$ does not only serve the purpose of reducing computation time, but also enables modeling uncertainty in transportation, which can be more realistic than using a deterministic model in real-life situations. When $T \to 0$, optimal transportation routes are displayed, but when $T \to \infty$, the cost minimization is negligible and the flow follows a random-walk pattern. If this random-walk is reversible, it has been shown to be similar to the *electrical current flow* with sources and targets being respectively current inputs and outputs [11,12]. A *source–target coupling*, i.e. an assignment of resources between sources and targets, can be retrieved from this flow. When $T \to 0$, this coupling gives an optimal solution to the transportation problem. To summarize, this algorithm enables to model an array of flows, ranging from optimal transportation routes to random-walk routes, when supply and demand on nodes are fixed, and to find a source–target coupling corresponding to these flows.

This algorithm is, in fact, a generalization of the one exposed in Refs. [13,14] and the notation is similar, even though not entirely compatible. In the latter, and in other articles, e.g. Refs. [15–18], the *randomized shortest-path* computation is limited to two nodes, whereas in the present case, multiples sources and targets are allowed.

The present article is divided in two parts. The formalism needed to construct the algorithm is exposed first, followed by illustrations of the algorithm running on a toy graph and a real graph.

2. Formalism

2.1. Admissible flows

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a simple oriented connected graph with $|\mathcal{V}| = n$, and $\mathcal{S}, \mathcal{T} \subset \mathcal{V}$, $\mathcal{S} \cap \mathcal{T} = \emptyset$, respectively, the set of *sources* and the set of *targets*. Each source generates a defined flow, which is transported on the graph before being absorbed by targets. Let us define $f = (f_i)$ with $f_i > 0\ \forall i \in \mathcal{S}$, $f_i = 0\ \forall i \notin \mathcal{S}$ and $\sum_i f_i = 1$, the *in-flow vector*, representing the proportion of the flow created by source nodes. Similarly, let $\rho = (\rho_i)$ with $\rho_i > 0\ \forall i \in \mathcal{T}$, $\rho_i = 0\ \forall i \notin \mathcal{T}$ and $\sum_i \rho_i = 1$ be the *out-flow vector*, representing the proportion of the flow absorbed by target nodes. All these quantities are provided initially.

The unknown *flow matrix*, noted $X = (x_{ij})$, represents the quantity of flow on arcs (i, j), $\forall i, j \in \mathcal{V}$. Components x_{ij} must verify:

$$x_{ij} \geq 0 \qquad \text{positivity} \tag{1}$$

$$x_{i\bullet} - x_{\bullet i} = f_i - \rho_i \qquad \text{unit flow conservation.} \tag{2}$$

An alternative way to consider the out-flow vector is as a set of constraints on flows coming from each node to a "virtual" node ω, the *ground node*. Similarly, the in-flow vector can be viewed as constraints set on flows coming from another "virtual" node ϕ, the *generator node*, to $\mathcal{V}$. Formally, $\forall i \in \mathcal{V}$:

$$x_{\phi i} = f_i \qquad x_{i\phi} = 0 \tag{3}$$

$$x_{i\omega} = \rho_i \qquad x_{\omega i} = 0. \tag{4}$$

Further in this article, the ground node must be added to the original graph in order to solve the problem.

Note that the *set of admissible flows*, $\mathcal{X}$, that is the set of flows verifying constraints (1) and (2), is a *convex set*, i.e. if X and Y are in $\mathcal{X}$, so is their *mixture* $\alpha X + (1-\alpha)Y,\ \forall \alpha \in [0, 1]$.

2.2. Flow entropy and energy

Let $W = (w_{ij})$ be the $(n \times n)$ transition matrix of some irreducible Markov chain defined on $\mathcal{G}$. A flow matrix X will follow the random-walk defined by W iff $x_{ij}/x_{i\bullet} = w_{ij}$ for all nodes i with $x_{i\bullet} > 0$. Therefore a "random-walk" flow will minimize the *entropy functional*:

$$G(X) = G(X \parallel W) := \sum_{i,j \in \mathcal{V}} x_{ij} \ln \frac{x_{ij}}{x_{i\bullet} w_{ij}} = x_{\bullet\bullet} \sum_{i \in \mathcal{V}} \frac{x_{i\bullet}}{x_{\bullet\bullet}} K_i(X \parallel W) \tag{5}$$

where $K_i(X \parallel W) := \sum_{j \in \mathcal{V}} \frac{x_{ij}}{x_{i\bullet}} \ln \frac{x_{ij}}{x_{i\bullet} w_{ij}} \geq 0$ is the Kullback–Leibler divergence between the transition distributions X and W. This divergence is weighted by $x_{i\bullet}/x_{\bullet\bullet}$ to take into account the visit frequency of nodes, and the sum of flows on all arcs, $x_{\bullet\bullet}$, is included in $G(X)$ to transform the entropy in a *homogeneous* functional, that is $G(\nu X) = \nu G(X)$ for $\nu > 0$, reflecting the *extensivity* of $G(X)$ in the thermodynamic sense.

By contrast, flows minimizing the following *energy functional*:

$$U(X) = U(X \parallel R) := \sum_{i,j \in \mathcal{V}} r_{ij} x_{ij} \tag{6}$$