Efficient Robust Routing for Single Commodity Network Flows

Yongxin Chen , *Student Member, IEEE*, Tryphon T. Georgiou , *Fellow, IEEE*, Michele Pavon , and Allen Tannenbaum , *Fellow, IEEE*

Abstract—**We study single commodity network flows with suitable robustness and efficiency specs. An original use of a maximum entropy problem for distributions on the paths of the graph turns this problem into a *steering problem* for Markov chains with prescribed initial and final marginals. From a computational standpoint, viewing scheduling this way is especially attractive in light of the existence of an iterative algorithm to compute the solution. This paper builds on the work proposed by Y. Chen *et al.*, ["Robust transport over networks," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4675–4682, Sep. 2017] by introducing an index of *efficiency* of a transportation plan and points, accordingly, to efficient robust transport policies. In developing the theory, we establish two new *invariance properties* of the solution (called *bridge*)—an *iterated bridge* invariance property and the invariance of the *most probable paths*. These properties, which were tangentially mentioned in our previous work, are fully developed here. We also show that the distribution on paths of the optimal transport policy, which depends on a "*temperature*" parameter, tends to the solution of the "most economical" but possibly less robust *optimal mass transport* problem as the temperature goes to zero. The relevance of all of these properties for transport over networks is illustrated in an example.**

Index Terms—**Maximum entropy problem, most probable path, temperature parameter, transport over networks.**

I. INTRODUCTION

Consider a company owning a factory F and a warehouse W. The company wants to ship a certain quantity of goods from F so that they reach W in at most N time units. The flow must occur on the available road network connecting the two facilities. On one hand, it is desirable that the transport plan utilizes as many different routes as possible so that most of the goods arrive within the prescribed time horizon even in the presence of road congestion, roadwork, etc. On the other hand, it is also important that shorter paths are used to keep the vehicles fuel consumption within a budgetary constraint.

In this paper, continuing the research initiated in [13], we provide a precise mathematical formulation of the above single commodity network flow problem. Normalizing the mass of goods to one, we formulate a *maximum entropy problem* for Markovian distributions on the paths of the network. The optimal feedback control suitably modifies a prior transition mechanism thereby achieving robustness while limiting the cost. This is accomplished through an appropriate choice of the prior transition involving the *adjacency matrix* of the graph. The optimal scheduling, while spreading the mass over all feasible paths, assigns *maximum probability* to all *minimum cost paths*.

Differently from the standard literature on controlled Markov chains, the optimal policy (*Schrödinger bridge*) is *not computed through dynamic programming*. The constraint on the final marginal (all the goods should be in the warehouse by day N) dictates a different approach. The solution is computed by solving iteratively a Schrödinger–Bernstein linear system with nonlinear coupling at the initial and final times. This algorithm, whose convergence was established in [22], is related to recent work in entropy regularization [16] and equilibrium assignment in economics [23] as well as to classical work in statistics [26].

Our straightforward approach also avoids altogether modeling *cascading failures,* which is a complex and controversial task [42]. It is also worthwhile remarking that maximum entropy problems [14] that constitute a powerful inference method find here an alternative use as a tool to produce a desired flow in a network by exploiting the properties of the prior transition mechanism.

Our intuitive notion of robustness of the routing policy should not be confused with other notions of robustness concerning networks that have been put forward and studied, see, e.g., [1], [5]–[7], [19], [42]. In particular, in [7] and [19], robustness has been defined through a fluctuation–dissipation relation involving the entropy rate. This latter notion captures relaxation of a process back to equilibrium after a perturbation and has been used to study both financial and biological networks [40], [41]. Our study, inspired by transportation and data networks, does not concern equilibrium or near equilibrium cases.

This paper features the following novel contributions.
1) It introduces an explicit index of *efficiency* of a transportation plan.
2) The choice of the adjacency matrix as prior transition mechanism, which was justified in [13] on an intuitive basis, is here further motivated through a specific optimization problem.
3) We derive an *iterated bridge* invariance property.
4) We establish the invariance of the *most probable paths*. These two invariance properties, which were only briefly mentioned in [13]

Y. Chen is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: yongchen@iastate.edu).

T. T. Georgiou is with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697-1075 USA (e-mail: tryphon@uci.edu).

M. Pavon is with the Dipartimento di Matematica "Tullio Levi Civita", Università di Padova, Padova 35121, Italy (e-mail: pavon@math.unipd.it).

A. Tannenbaum is with the Departments of Computer Science and Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794 USA (e-mail: allen.tannenbaum@stonybrook.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2017.2763418

in some special cases, are fully investigated here. Their relevance for transport over networks is also illustrated.

5) We study the dependence of the optimal transport on a temperature parameter.

The possibility of employing the solution for near-zero temperature as an approximation of the solution to *optimal mass transport* (OMT) is also discussed and illustrated through examples.

The outline of this paper is as follows. In Section II, we introduce generalized maximum entropy problems. In Section II-A, we establish the iterated bridge property, and in Section II-B, the invariance of the most probable paths. Efficiency of a transport policy is introduced in Section III-A. In Section III-B, we introduce robust transport with fixed average path length. Section IV deals with efficient robust transportation. In Section V, the dependence of the optimal transport on the temperature parameter is investigated. The results are then illustrated through academic examples in Section VI.

II. GENERALIZED MAXIMUM ENTROPY PROBLEMS

We are given a directed, strongly connected (i.e., with a path in each direction between each pair of vertices) aperiodic graph $\mathbf{G} = (\mathcal{X}, \mathcal{E})$ with vertex set $\mathcal{X} = \{1, 2, \ldots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{X} \times \mathcal{X}$. We let time vary in $\mathcal{T} = \{0, 1, \ldots, N\}$, and let $\mathcal{FP}_0^N \subseteq \mathcal{X}^{N+1}$ denote the family of length N, feasible paths $x = (x_0, \ldots, x_N)$, namely, paths such that $x_i x_{i+1} \in \mathcal{E}$ for $i = 0, 1, \ldots, N-1$.

We seek a probability distribution P on $\mathcal{FP}_0^N$ with prescribed initial and final marginal probability distributions $\nu_0(\cdot)$ and $\nu_N(\cdot)$, respectively, and such that the resulting random evolution is closest to a "prior" measure $\mathfrak{M}$ on $\mathcal{FP}_0^N$ in a suitable sense. The prior law $\mathfrak{M}$ is induced by the Markovian evolution as

$$\mu_{t+1}(x_{t+1}) = \sum_{x_t \in \mathcal{X}} \mu_t(x_t) m_{x_t x_{t+1}}(t) \tag{1}$$

with nonnegative distributions $\mu_t(\cdot)$ over $\mathcal{X}$, $t \in \mathcal{T}$, and weights $m_{ij}(t) \geq 0$ for all indices $i, j \in \mathcal{X}$ and all times. Moreover, to respect the topology of the graph, $m_{ij}(t) = 0$ for all t whenever $ij \notin \mathcal{E}$. Often, but not always, the matrix

$$M(t) = [m_{ij}(t)]_{i,j=1}^n \tag{2}$$

does not depend on t. The rows of the transition matrix $M(t)$ do not necessarily sum up to one, so that the "total transported mass" is not necessarily preserved. It occurs, for instance, when M simply encodes the topological structure of the network with m_{ij} being zero or one, depending on whether a certain link exists. The evolution (1) together with the measure $\mu_0(\cdot)$, which we assume positive on $\mathcal{X}$, i.e.,

$$\mu_0(x) > 0 \text{ for all } x \in \mathcal{X} \tag{3}$$

induces a measure $\mathfrak{M}$ on $\mathcal{FP}_0^N$ as follows. It assigns to a path $x = (x_0, x_1, \ldots, x_N) \in \mathcal{FP}_0^N$ the value

$$\mathfrak{M}(x_0, x_1, \ldots, x_N) = \mu_0(x_0) m_{x_0 x_1} \ldots m_{x_{N-1} x_N} \tag{4}$$

and gives rise to a flow of *one-time marginals*

$$\mu_t(x_t) = \sum_{x_\ell \neq t} \mathfrak{M}(x_0, x_1, \ldots, x_N), \quad t \in \mathcal{T}.$$

Definition 1: We denote by $\mathcal{P}(\nu_0, \nu_N)$ the family of probability distributions on $\mathcal{FP}_0^N$ having the prescribed marginals $\nu_0(\cdot)$ and $\nu_N(\cdot)$.

We seek a distribution in this set which is closest to the prior $\mathfrak{M}$ in *relative entropy* where, for P and Q measures on $\mathcal{X}^{N+1}$, the relative entropy (divergence, Kullback–Leibler index) $\mathbb{D}(P\|Q)$ is

$$\mathbb{D}(P\|Q) := \begin{cases} \sum_x P(x) \log \frac{P(x)}{Q(x)}, & \mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q) \\ +\infty, & \mathrm{Supp}(P) \nsubseteq \mathrm{Supp}(Q). \end{cases}$$

Here, by definition, $0 \cdot \log 0 = 0$. Naturally, while the value of $\mathbb{D}(P\|Q)$ may turn out negative due to mismatch of scaling (in case $Q = \mathfrak{M}$ is not a probability measure), the relative entropy is always jointly convex. We consider the *Schrödinger bridge problem* (SBP) as follows.

Problem 1: Determine

$$\mathfrak{M}^*[\nu_0, \nu_N] := \arg\min\{\mathbb{D}(P\|\mathfrak{M}) \mid P \in \mathcal{P}(\nu_0, \nu_N)\}. \tag{5}$$

The following result is a slight generalization (to time inhomogeneous prior) of [13, Th. 2.3].

Theorem 1: Assume that the product $M(N-1)M(N-2)\cdots M(1)M(0)$ has all entries positive. Then there exist nonnegative functions $\varphi(\cdot)$ and $\hat{\varphi}(\cdot)$ on $[0, N] \times \mathcal{X}$ satisfying

$$\varphi(t, i) = \sum_j m_{ij}(t) \varphi(t+1, j) \tag{6a}$$

$$\hat{\varphi}(t+1, j) = \sum_i m_{ij}(t) \hat{\varphi}(t, i) \tag{6b}$$

for $t \in [0, N-1]$, along with the (nonlinear) boundary conditions

$$\varphi(0, x_0) \hat{\varphi}(0, x_0) = \nu_0(x_0) \tag{6c}$$

$$\varphi(N, x_N) \hat{\varphi}(N, x_N) = \nu_N(x_N) \tag{6d}$$

for $x_0, x_N \in \mathcal{X}$. Moreover, the solution $\mathfrak{M}^*[\nu_0, \nu_N]$ to Problem 1 is unique and obtained by

$$\mathfrak{M}^*(x_0, \ldots, x_N) = \nu_0(x_0) \pi_{x_0 x_1}(0) \cdots \pi_{x_{N-1} x_N}(N-1)$$

where the one-step transition probabilities

$$\pi_{ij}(t) := m_{ij}(t) \frac{\varphi(t+1, j)}{\varphi(t, i)} \tag{7}$$

are well defined.

The factors φ and $\hat{\varphi}$ are unique up to multiplication of φ by a positive constant and division of $\hat{\varphi}$ by the same constant. Let $\varphi(t)$ and $\hat{\varphi}(t)$ denote the column vectors with components $\varphi(t, i)$ and $\hat{\varphi}(t, i)$, respectively, with $i \in \mathcal{X}$. In matricial form, (6a), (6b), and (7) read

$$\varphi(t) = M(t) \varphi(t+1), \quad \hat{\varphi}(t+1) = M(t)^T \hat{\varphi}(t) \tag{8}$$

and

$$\Pi(t) := [\pi_{ij}(t)] = \mathrm{diag}(\varphi(t))^{-1} M(t) \mathrm{diag}(\varphi(t+1)). \tag{9}$$

Historically, the SBP was posed in 1931 by Schrödinger for Brownian particles with a large deviations of the empirical distribution motivation [43], see [30] for a survey. The problem was considered in the context of Markov chains and studied in [22] and [36], and some generalizations have been discussed in [13]. Important connections between SBP and OMT [3], [45], [46] have been discovered and developed in [10]–[12], [29], [30], [32], [33].

A. Iterated Bridges

In this section, we explain a rather interesting property of Schrödinger bridges, which is the following. If, after solving an SBP for a given set of marginals (ν_0, ν_N) and a Markovian prior $\mathfrak{M}$ to obtain $\mathfrak{M}^*[\nu_0, \nu_N]$, we decided to update the data (ν_0, ν_N) to another set of marginals (π_0, π_N) then, whether we use as prior $\mathfrak{M}$ or $\mathfrak{M}^*[\nu_0, \nu_N]$ for the SBP with the new marginals π_0 and π_N, we obtain precisely

the same solution $\mathfrak{M}^*[\pi_0, \pi_N]$. The significance of this property will be discussed later on in the context of robust transportation.

Indeed, take $\mathfrak{M}^*[\nu_0, \nu_N]$ as prior and consider the corresponding new Schrödinger system (in matrix form)

$$\psi(t) = \Pi(t)\psi(t+1), \quad \hat{\psi}(t+1) = \Pi(t)^T \hat{\psi}(t)$$

with boundary conditions

$$\psi(0, x_0)\hat{\psi}(0, x_0) = \pi_0(x_0) \tag{10a}$$

$$\psi(N, x_N)\hat{\psi}(N, x_N) = \pi_N(x_N). \tag{10b}$$

Note in the above $\Pi(t) = \mathrm{diag}(\varphi(t))^{-1} M(t) \mathrm{diag}(\varphi(t+1))$, therefore it can be written as

$$\mathrm{diag}(\varphi(t))\psi(t) = M(t) \mathrm{diag}(\varphi(t+1))\psi(t+1) \tag{11a}$$

$$\mathrm{diag}(\varphi(t+1))^{-1}\hat{\psi}(t+1) = M(t)^T \mathrm{diag}\,\varphi(t))^{-1}\hat{\psi}(t). \tag{11b}$$

The new transition matrix Q^* is given by

$$\begin{aligned} Q^*(t) &= \mathrm{diag}(\psi(t))^{-1}\Pi(t)\,\mathrm{diag}(\psi(t+1)) \\ &= \mathrm{diag}(\psi(t))^{-1}\,\mathrm{diag}(\varphi(t))^{-1} \\ &\quad \times M(t)\,\mathrm{diag}(\varphi(t+1))\,\mathrm{diag}(\psi(t+1)). \end{aligned}$$

Let $\psi_1(t) = \mathrm{diag}(\varphi(t))\psi(t)$ and $\hat{\psi}_1(t) = \mathrm{diag}(\varphi(t))^{-1}\hat{\psi}(t)$, then

$$Q^*(t) = \mathrm{diag}(\psi_1(t))^{-1} M(t)\,\mathrm{diag}(\psi_1(t+1)).$$

By (11), ψ and $\hat{\psi}$ are vectors with positive components satisfying

$$\psi_1(t) = M(t)\psi_1(t+1),\ \hat{\psi}_1(t+1) = M(t)^T \hat{\psi}_1(t).$$

Moreover, they satisfy the boundary conditions

$$\psi_1(0, x_0)\hat{\psi}_1(0, x_0) = \pi_0(x_0) \tag{12a}$$

$$\psi_1(N, x_N)\hat{\psi}_1(N, x_N) = \pi_N(x_N). \tag{12b}$$

Thus, $(\psi_1, \hat{\psi}_1)$ provide the solution to Problem 1 when $\mathfrak{M}$ is taken as prior.

Alternatively, observe the transition matrix $Q^*(t)$ resulting from the two problems is the same and so is the initial marginal. Hence, the solutions of the SBP with marginals π_0 and π_N and prior transitions $\Pi(t)$ and $M(t)$ are identical.

Thus, "the bridge over a bridge over a prior" is the same as the "bridge over the prior," i.e., iterated bridges produce the same result. It should be observed that this result for probability distributions is not surprising since the solution is in the same *reciprocal class* as the prior (namely, it has the same three times transition probability), cf., [27], [31], [49]. It could then be described as the fact that only the reciprocal class of the prior matters; this is can be seen from Schrödinger's original construction [43], and also [22, Section III-B] for the case of Markov chains. This result, however, is more general since the prior is not necessarily a probability measure.

In information theoretic terms, the bridge (i.e., probability law on path spaces) corresponding to Q^* is the I-projection in the sense of Cziszar [15] of the prior onto the set of measures that are consistent with the initial-final marginals. The above result, however, is not simply an "iterated information-projection" property, since $\mathfrak{M}^*[\nu_0, \nu_N]$ is the I-projection of $\mathfrak{M}$ onto $\mathcal{P}(\nu_0, \nu_N)$ which does not contain $\mathcal{P}(\pi_0, \pi_N)$ being in fact disjoint from it.

B. Invariance of Most Probable Paths

Building on the *logarithmic transformation* of Fleming, Holland, Mitter, and others, the connection between SBP and stochastic control was developed from the early 1990s on [8], [17], [18], [37]. More

recently, Brockett studied steering of the Louiville equation [9]. In [17, Sec. 5], Dai Pra established an interesting path-space property of the Schrödinger bridge for diffusion processes that the "most probable path" [20], [44] of the prior and the solution is the same. Loosely speaking, a most probable path is similar to a *mode* for the path-space measure P. More precisely, if both drift $b(\cdot, \cdot)$ and diffusion coefficient $\sigma(\cdot, \cdot)$ of the Markov diffusion process

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t$$

are smooth and bounded, with $\sigma(x, t)\sigma(x, t)^T > \eta I$, $\eta > 0$, and $\{x(t) \mid 0 \le t \le T\}$ is a path of class C^2, then there exists an asymptotic estimate of the probability P of a small tube around $x(t)$ of radius ϵ. It follows from this estimate that the most probable path is the minimizer in a deterministic calculus of variations problem where the Lagrangian is an *Onsager–Machlup functional*, see [25, p. 532] for the full story.[1]

The concept of the most probable path is, of course, much less delicate in our discrete setting. We define it for general positive measures on paths. Given a positive measure $\mathfrak{M}$ as in Section II on the feasible paths of our graph $\mathbf{G}$, we say that $x = (x_0, \dots, x_N) \in \mathcal{FP}_0^N$ is of *maximal mass* if for all other feasible paths $y \in \mathcal{FP}_0^N$ we have $\mathfrak{M}(y) \le \mathfrak{M}(x)$. Likewise, we consider paths of *maximal mass* connecting particular nodes. It is apparent that paths of maximal mass always exist but are, in general, not unique. If $\mathfrak{M}$ is a probability measure, then the maximal mass paths—most probable paths—are simply the modes of the distribution. We establish in the following that the maximal mass paths joining two given nodes under the solution of an SBP as in Section II are the same as for the prior measure.

Proposition 1: Consider marginals ν_0 and ν_1 in Problem 1. Assume that $\nu_0(x) > 0$ on all nodes $x \in \mathcal{X}$ and that the product $M(N-1) \cdot M(N-2) \cdots M(1) \cdot M(0)$ of transition probability matrices of the prior has all positive elements [cf., with M's as in (2)]. Let x_0 and x_N be any two nodes. Then, under the solution $\mathfrak{M}^*[\nu_0, \nu_N]$ of the SBP, the family of maximal mass paths joining x_0 and x_N in N steps is the same as under the prior measure $\mathfrak{M}$.

Proof: Suppose path $y = (y_0 = x_0, y_1, \dots, y_{N-1}, y_N = x_N)$ has maximal mass under the prior $\mathfrak{M}$. In view of (4) and (7) and assumption (3), we have

$$\begin{aligned} \mathfrak{M}^*[\nu_0, \nu_N](y) &= \nu_0(y_0)\pi_{y_0 y_1}(0) \cdots \pi_{y_{N-1} y_N}(N-1) \\ &= \frac{\nu_0(x_0)}{\mu_0(x_0)} \frac{\varphi(N, x_N)}{\varphi(0, x_0)} \mathfrak{M}(y_0, y_1, \dots, y_N). \end{aligned}$$

Since the quantity

$$\frac{\nu_0(x_0)}{\mu_0(x_0)} \frac{\varphi(N, x_N)}{\varphi(0, x_0)}$$

is positive and does not depend on the particular path joining x_0 and x_N, the conclusion follows. $\qquad\square$

The calculation in the above proof actually establishes the following stronger result.

Proposition 2: Let x_0 and x_N be any two nodes in $\mathcal{X}$. Then, under the assumptions of Proposition 1, the measures $\mathfrak{M}$ and $\mathfrak{M}^*[\nu_0, \nu_N]$, restricted on the set of paths that begin at x_0 at time 0 and end at x_N at time N, are identical.

III. ROBUST TRANSPORT

In this section, we first discuss notions of efficiency of a transportation plan and then introduce entropy as a surrogate for robustness.

---

[1] The Onsager–Machlup functional was introduced in [34] to develop a theory of fluctuations in equilibrium and nonequilibrium thermodynamics.

A. Efficiency of a Transport Plan

Inspired by the celebrated paper [48], we introduce a measure of *efficiency* of a *transportation plan over a certain finite-time horizon and a given network* in the following.

For the case of undirected and connected graphs, *small-world networks* [48] were identified as networks being highly clustered but with small characteristic path length L, where

$$L := \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}$$

and d_{ij} is the shortest path length between vertices i and j. The inverse of the characteristic path length L^{-1} is an index of efficiency of $\mathbf{G}$. There are other such indexes, most noticeably the global efficiency E_{glob} introduced in [28]. This is defined as $E_{\mathrm{glob}} = E(\mathbf{G})/E(\mathbf{G}_{\mathrm{id}})$ where

$$E(\mathbf{G}) = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$$

and $\mathbf{G}_{\mathrm{id}}$ is the complete network with all possible edges in place. Thus, $0 \leq E_{\mathrm{glob}} \leq 1$. However, as argued on [28, p. 198701–2], it is $1/L$ which "measures the efficiency of a sequential system (i.e., only one packet of information goes along the network)." E_{glob}, instead, measures the efficiency of a parallel system, namely, one in which all nodes concurrently exchange packets of information. Since we are interested in the efficiency of a specific transportation plan, we define the efficiency by a suitable adaptation of the index L as follows.

Consider a strongly connected, aperiodic, directed graph $\mathbf{G} = (\mathcal{X}, \mathcal{E})$ as in Section II. To each edge, ij is now associated with a length $l_{ij} \geq 0$. If $ij \notin \mathcal{E}$, we set $l_{ij} = +\infty$. The length may represent distance, cost of transport/communication/etc. Let $\mathcal{T} = \{0, 1, \ldots, N\}$ be the time-indexing set. For a path $x = (x_0, \ldots, x_N) \in \mathcal{X}^{N+1}$, we define the length of x to be

$$l(x) = \sum_{t=0}^{N-1} l_{x_t x_{t+1}}.$$

We consider the situation where initially at time $t = 0$ the mass is distributed on $\mathcal{X}$ according to $\nu_0(x)$ and needs to be distributed according to $\nu_N(x)$ at the final time $t = N$. These masses are normalized to sum to one, so that they are probability distributions. A transportation plan P is a probability measure on the (feasible) paths of the network having the prescribed marginals ν_0 and ν_N at the initial and final time, respectively. A natural adaptation of the characteristic path length is to consider the *average path length of the transportation plan P*, which we define as

$$L(P) = \sum_{x \in \mathcal{X}^{N+1}} l(x) P(x) \tag{13}$$

with the usual convention $+\infty \times 0 = 0$. This is entirely analogous to a thermodynamic quantity, the *internal energy*, which is defined as the expected value of the Hamiltonian observable in state P. Clearly, $L(P)$ is finite if and only if the transport takes place on actual, existing links of $\mathbf{G}$. Moreover, only the paths which are in the *support* of P enter in the computation of $L(P)$. One of the goals of a transportation plan is of course to have small average path length since, for instance, cost might simply be proportional to length. Determining the probability measure that minimizes (13) can be seen to be an OMT problem.

B. Problem Formulation

Besides efficiency, another desirable property of a transport strategy is to ensure robustness with respect to links/nodes failures, the latter being due possibly to malicious attacks. We, therefore, seek a transport plan in which the mass spreads, as much as it is allowed by the network topology, before reconvening at time $t = N$ in the sink nodes. We achieve this by selecting a transportation plan P that has a suitably high entropy $S(P)$, where

$$S(P) = - \sum_{x \in \mathcal{X}^{N+1}} P(x) \ln P(x). \tag{14}$$

Thus, in order to attain a level of robustness while guaranteeing a relatively low average path length (cost), we formulate a constrained optimization problem that weighs in both $S(P)$ as well as $L(P)$ as follows.

We begin by letting $\bar{L}$ designate a suitable bound on the average path length (cost) that we are willing to accept. Clearly, we need that

$$l_m := \min_{x \in \mathcal{X}^{N+1}} l(x) \leq \bar{L}. \tag{15a}$$

We will also assume that

$$\bar{L} \leq \frac{1}{|\mathcal{FP}_0^N|} \sum_{x \in \mathcal{FP}_0^N} l(x). \tag{15b}$$

The rationale behind the latter, i.e., requiring an upper bound as stated, will be explained in Proposition 3 as follows.

Let $\mathcal{P}$ denote the family of probability measures on $\mathcal{X}^{N+1}$. The probability measure that maximizes the entropy $S(P)$ subject to a path-length constraint $L(P) = \bar{L}$ is the Boltzmann distribution

$$P_T^*(x) = Z(T)^{-1} \exp\left[-\frac{l(x)}{T}\right], \quad Z(T) = \sum_x \exp\left[-\frac{l(x)}{T}\right] \tag{16}$$

where the parameter (temperature) T depends on $\bar{L}$. To see this, consider the Lagrangian

$$\mathcal{L}(P, \lambda) := S(P) + \lambda(\bar{L} - L(P)) \tag{17}$$

and observe that the Boltzmann distribution (16) satisfies the first-order optimality condition of $\mathcal{L}$ with $T = 1/\lambda$. Clearly, the Boltzmann distribution has support on the feasible paths $\mathcal{FP}_0^N$. Hence, we get a version of Gibbs' variational principle that the Boltzmann distribution P_T^* minimizes the *free energy* functional

$$F(P, T) := L(P) - T S(P) \tag{18}$$

over $\mathcal{P}$. An alternative way to establish the minimizing property of Boltzmann's distribution is to observe that

$$F(P, T) = T \mathbb{D}(P \| P_T^*) - T \log Z \tag{19}$$

and, therefore, minimizing the free energy over $\mathcal{P}$ is equivalent to minimizing the relative entropy $\mathbb{D}(P \| P_T^*)$ over $P \in \mathcal{P}$, which ensures that the minimum is unique. The following properties of P_T^* are noted, see, e.g., [35, Ch. 2].

Proposition 3: The following holds:
i) For $T \nearrow +\infty$, P_T^* tends to the uniform distribution on all feasible paths.
ii) For $T \searrow 0$, P_T^* tends to concentrate on the set of feasible paths having minimal length.
iii) Assuming that $l(\cdot)$ is not constant over $\mathcal{FP}_0^N$, then for each value $\bar{L}$ satisfying the bounds (15), there exists a unique nonnegative value of $T = \lambda^{-1} \in [0, \infty]$ such that P_T^* maximizes $S(P)$ subject to $L(P) = \bar{L}$.

We also observe the Markovian nature of the measure P_T^*. Indeed, recall that a positive measure $\mathfrak{M}$ on $\mathcal{X}^{N+1}$ is Markovian if it can be expressed as in (4). Since

$$P_T^*(x_0, x_1, \ldots, x_N) = Z(T)^{-1} \prod_{t=0}^{N-1} \exp\left[-\frac{l_{x_t x_{t+1}}}{T}\right] \quad (20)$$

which is exactly in the form (4), we conclude that P_T^* is (*time homogeneous*) Markovian with uniform initial measure $\mu(x_0) \equiv Z(T)^{-1}$ and time-invariant transition matrix given by

$$M_T = \left[\exp\left(-\frac{l_{ij}}{T}\right)\right]_{i,j=1}^n. \quad (21)$$

However, observe that, in general, M_T is not stochastic (rows do not sum to one). Moreover, observe that, after suitable normalization, M_T represents the transition matrix of a chain where probabilities of transition between nodes are inversely proportional to the length of the links.

Consider now ν_0 and ν_N distributions on $\mathcal{X}$. These are the "starting" and "ending" concentrations of resources for which we seek a transportation plan. We denote by $\mathcal{P}(\nu_0, \nu_N)$ the family of probability distributions on paths $x \in \mathcal{X}^{N+1}$ having ν_0 and ν_N as initial and final marginals, respectively, and we consider the problem to maximize the entropy subject to marginal and length constraints as follows.

Problem 2: Maximize $S(P)$ subject to $P \in \mathcal{P}(\nu_0, \nu_N)$ and $L(P) = \bar{L}$.

Note that the solution to Problem 2 depends on $\bar{L}$ as well as the two marginals ν_0, ν_N and that when $\bar{L}$ is too close to l_m, the problem may be infeasible.

Once again, bringing in the Lagrangian (17), which now needs to be minimized over $\mathcal{P}(\nu_0, \nu_N)$, we see that Problem 2 is equivalent to solving the following SBP for a suitable value of the parameter T.

Problem 3: minimize $\{\mathbb{D}(P\|P_T^*) \mid P \in \mathcal{P}(\nu_0, \nu_N)\}$.

Thus, *employing path-space entropy as a measure of robustness*, the solution to Problem 3, denoted by $\mathfrak{M}_T^*(\nu_0, \nu_N)$ and constructed in accordance with Theorem 1, minimizes a suitable *free energy functional* with the temperature parameter specifying the tradeoff between efficiency and robustness. Thus, Problem 3 can be viewed as an SBP as in Section II where the "prior" measure P_T^* is Markovian.

IV. STRUCTURE OF ROBUST TRANSPORT

We now address in detail Problem 3, namely, to identify a probability distribution P on $\mathcal{FP}_0^N$ that minimizes $\mathbb{D}(\cdot\|P_T^*)$ over $\mathcal{P}(\nu_0, \nu_N)$ where P_T^* is the Boltzmann distribution (20)—the minimizing law being denoted by $\mathfrak{M}_T^*[\nu_0, \nu_N]$ as before. In the following, we show that the two invariant properties discussed in the previous two sessions can be used to determine an optimal transport policy. We also show that the $\mathfrak{M}_T^*[\nu_0, \nu_N]$ inherits from the Boltzmann distribution P_T^* properties as dictated by Proposition 3.

Initially, for simplicity, we consider the situation where at time $t = 0$ the whole mass is concentrated on node 1 (*source*), and at time $t = N$ it is concentrated on node n (*sink*), i.e., $\nu_0(x) = \delta_1(x)$ and $\nu_N(x) = \delta_n(x)$. We want to allow (part of) the mass to reach the end-point "sink" node, if this is possible, in less than N steps and then remain there until $t = N$. In order to ensure that it is possible, we assume that there exists a self-loop at node n, i.e., $M_{T\,nn} > 0$. Clearly, $\mathfrak{M}_T^*(\delta_1, \delta_n)(\cdot) = P_T^*[\cdot|Y_0 = 1, Y_N = n]$. The Schrödinger bridge theory provides transition probabilities so that, for a path

$y = (y_0, y_1, \ldots, y_N)$

$$\mathfrak{M}_T^*(\delta_1, \delta_n)(y) = \delta_1(y_0) \prod_{t=1}^{N-1} \exp\left(-\frac{l_{y_t y_{t+1}}}{T}\right) \frac{\varphi_T(t+1, y_{t+1})}{\varphi_T(t, y_t)}$$

$$= \delta_1(y_0) \frac{\varphi_T(N, y_N)}{\varphi_T(0, y_0)} \left[\exp\left(-\frac{l(y)}{T}\right)\right] \quad (22)$$

cf., (4) and (7). Here, $l(y) = \sum_{t=0}^{N-1} l_{y_t y_{t+1}}$ is the length of path y and φ_T satisfies together with $\hat{\varphi}_T$ the Schrödinger system (6) with $m_{ij}(t) = \exp\left(-\frac{l_{ij}}{T}\right)$ and $\nu_0(x) = \delta_1(x), \nu_N(x) = \delta_n(x)$.

In [13, Sec. VI], Problem 3 was first studied with a prior measure $\mathfrak{M}_l$ having certain special properties. To introduce this particular measure, we first recall (part of) a fundamental result from linear algebra [24].

Theorem 2 (Perron–Frobenius): Let $A = (a_{ij})$ be an $n \times n$ matrix with nonnegative entries. Suppose there exists N such that A^N has only positive entries, and let λ_A be its spectral radius. Then,
  i) $\lambda_A > 0$ is an eigenvalue of A;
  ii) λ_A is a simple eigenvalue;
  iii) there exists an eigenvector v corresponding to λ_A with strictly positive entries.

Consider now the weighted adjacency matrix $B = M_T$ in (21) (where we dropped the subscript T as it will be fixed throughout this section). Assume that B^N has all positive elements so that we can apply the Perron–Frobenius theorem. Let u and v be the left and right eigenvectors with positive components of the matrix B corresponding to the spectral radius λ_B. We have

$$B^T u = \lambda_B u, \quad Bv = \lambda_B v. \quad (23)$$

We assume throughout that u and v are chosen so that $\sum_i u_i v_i = 1$. Then, for $y_0 = i$ and $y_t = j$, define

$$\mathfrak{M}_l(i, y_1, \ldots, x_{0\,t-1}, j) := \lambda_B^{-t} u_i v_j e^{-\sum_{k=0}^{t-1} l_{y_k y_{k+1}}}. \quad (24)$$

The corresponding transition matrix is

$$R_l = \lambda_B^{-1} \operatorname{diag}(v)^{-1} B \operatorname{diag}(v). \quad (25)$$

It admits the *invariant measure*

$$\mu_l(i) = u_i v_i. \quad (26)$$

Note that $\mathfrak{M}_l$ and the Boltzmann distribution P_T^* have the same transition matrix but different initial distributions. In [13], to which we refer for motivation and more details, the following problem was studied.

Problem 4: minimize $\{\mathbb{D}(P\|\mathfrak{M}_l) \mid P \in \mathcal{P}(\nu_0, \nu_N)\}$.

Under the assumption that B^N has all positive entries, this SBP has a unique solution $\mathfrak{M}_l^*$. In [13, Th. 3.4], it was also shown that $\mathfrak{M}_l$ is itself the solution of an SBP with equal marginals and the Boltzmann distribution (16) as prior. Thus, by the iterated bridge property of Section II-A, $\mathfrak{M}_l^*$ coincides with the solution of Problem 3 for any choice of the initial–final marginals ν_0 and ν_N.

We recall the following rather surprising result [13, Th. 6.1], which includes the invariance of the most probable paths in Problem 3 (Proposition 1).

Theorem 3: $\mathfrak{M}_l^*$ gives equal probability to paths $y \in \mathcal{X}^{N+1}$ of equal length between any two given nodes. In particular, it assigns maximum and equal probability to minimum length paths.

This result is relevant when the solution of Problem 3 for low temperature is used as an approximation to OMT, see Remark 1 Section V. Finally, an important special case occurs when $l_{ij} = 0$ for existing links and $+\infty$ for nonexisting. Then, the matrix B reduces to the unweighted adjacency matrix A and the measure $\mathfrak{M}_l$ to the so-called Ruelle–Bowen random walk $\mathfrak{M}_{RB}$. The only concern in the transport

policy is in maximizing path family entropy to achieve robustness, see [13, Sec. 4 and 5] for details.

V. Dependence of Robust Transport on T

Following we study how the solution $\mathfrak{M}_T^* [\delta_{x_0}, \delta_{x_N}]$ to Problem 3 varies with the temperature parameter T. Here, x_0, x_N are specified nodes where mass is concentrated at the start and end times, and $\delta_{x'}(x) = 1$ when $x = x'$ and zero otherwise. It should be noted that similar results hold for general marginal distributions as well, which are not necessarily Dirac.

Theorem 4: Consider the solution $\mathfrak{M}_T^* [\delta_{x_0}, \delta_{x_N}] =: \mathfrak{M}_T^*$ to *Problem 3* with $\nu_0(x) = \delta_{x_0}(x)$ and $\nu_N(x) = \delta_{x_N}(x)$. Let $l_m(x_0, x_N) = \min_{y \in \mathcal{X}^{N+1}(x_0, x_N)} l(y)$, i.e., the minimum length of N-step paths originating in x_0 and terminating in x_N. Then,

i) For $T \searrow 0$, $\mathfrak{M}_T^*$ tends to concentrate itself on the set of feasible, minimum length paths joining x_0 and x_N in N steps. Namely, if $y = (y_0 = x_0, y_1, \ldots, y_{N-1}, y_N = x_N)$ is such that $l(y) > l_m(x_0, x_N)$, then $\mathfrak{M}_T^*(y) \searrow 0$ as $T \searrow 0$.

ii) For $T \nearrow +\infty$, $\mathfrak{M}_T^*$ tends to the uniform distribution on all feasible paths joining x_0 and x_N in N steps.

iii) Suppose $\mathcal{X}^{N+1}(x_0, x_N)$ is not a singleton and that $l(\cdot)$ is not constant over it. Then, for each value $\bar{L}$ satisfying the bounds

$$l_m(x_0, x_N) \le \bar{L} \le \frac{1}{|\mathcal{X}^{N+1}(x_0, x_N)|} \sum_{y \in \mathcal{X}^{N+1}(x_0, x_N)} l(y)$$

there exists a unique value of $T \in [0, +\infty]$ such that $\mathfrak{M}_T^*$ satisfies the constraint $L(\mathfrak{M}_T^*) = \bar{L}$ and therefore solves Problem 2.

Proof: Observe first that, since $\mathfrak{M}_T^*$ is a probability measure on $\mathcal{X}^{N+1}$, it must satisfy by (22)

$$1 = \sum_{y \in \mathcal{X}^{N+1}} \mathfrak{M}_T^*(y) = \sum_{y \in \mathcal{X}^{N+1}} \delta_1(y_0) \frac{\varphi_T(N, y_N)}{\varphi_T(0, y_0)} \left[\exp\left(-\frac{l(y)}{T} \right) \right]$$

$$= \sum_{y \in \mathcal{X}^{N+1}(x_0, x_N)} \delta_1(y_0) \frac{\varphi_T(N, x_N)}{\varphi_T(0, x_0)} \left[\exp\left(-\frac{l(y)}{T} \right) \right] \quad (27)$$

where we have used the fact that the initial and final marginals of $\mathfrak{M}_T^*$ are δ_1 and δ_n, respectively. It follows that

$$\frac{\varphi_T(0, x_0)}{\varphi_T(N, x_N)} = \sum_{y \in \mathcal{X}^{N+1}(x_0, x_N)} \delta_1(y_0) \left[\exp\left(-\frac{l(y)}{T} \right) \right]$$

$$= \sum_{y \in \mathcal{X}^{N+1}(x_0, x_N)} \left[\exp\left(-\frac{l(y)}{T} \right) \right] \quad (28)$$

where again $\mathcal{X}^{N+1}(x_0, x_N)$ denotes the family of paths joining x_0 and x_N in N time periods.

Proof of i): Let $y = (y_0 = x_0, y_1, \ldots, y_{N-1}, y_N = x_N)$ be such that $l(y) > l_m(x_0, x_N)$. Then

$$\mathfrak{M}_T^*(y) = \frac{\varphi_T(N, x_N)}{\varphi_T(0, x_0)} \left[\exp\left(-\frac{l(y)}{T} \right) \right].$$

By (28), we have $\frac{\varphi_T(0, x_0)}{\varphi_T(N, x_N)} \ge \exp(-\frac{l_m(x_0, x_N)}{T})$. Hence,

$$\mathfrak{M}_T^*(y) = \frac{\varphi_T(N, x_N)}{\varphi_T(0, x_0)} e^{-\frac{l(y)}{T}} \le e^{-\frac{l(y) - l_m(x_0, x_N)}{T}}.$$

Since $l(y) - l_m(x_0, x_N) > 0$, the right-hand side tends to zero as $T \searrow 0$.

Proof of ii): For $T \nearrow +\infty$, $\exp(-\frac{l(y)}{T})$ tends to 1 for all paths $y \in \mathcal{X}^{N+1}(x_0, x_N)$. Since $\frac{\varphi_T(N, x_N)}{\varphi_T(0, x_0)}$ does not depend on the specific
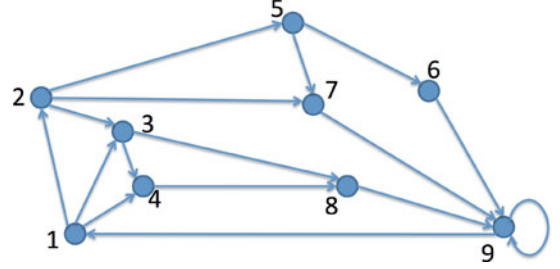

Fig. 1.   Network topology.

path in $\mathcal{X}^{N+1}(x_0, x_N)$ (it is just a normalization like the partition function), we conclude that as T tends to infinity, $\mathfrak{M}_T^*$ tends to the uniform distribution on $\mathcal{X}^{N+1}(x_0, x_N)$.

Proof of iii): Note that Problem 2 is feasible when $l_m(x_0, x_N) \le \bar{L}$ holds. By standard Lagrangian duality theory, there exists a Lagrangian multiplier $\lambda \in [0, \infty]$ such that the maximizer of the corresponding Lagrangian (17) over $\mathcal{P}(\nu_0, \nu_N)$ is the solution of Problem 2.[2] On the other hand, maximizing (17) over $\mathcal{P}(\nu_0, \nu_N)$ is equivalent to solving Problem 3 with $T = 1/\lambda$. □

Remark 1: Let us interpret l_{ij} as the cost of transporting a unit mass over the link ij. Then $L(P)$ is the *expected cost* corresponding to the transport plan P. For $T = 0$, the free energy functional reduces to $L(P)$ as our problem amounts to a discrete OMT problem [38]. In this, one seeks minimum cost paths—a combinatorial problem which can also be formulated as a linear programming problem [4]. Precisely as in the diffusion case [10]–[12], we also see that when the "heat bath" temperature is close to 0, the solution of the SBP is close to the solution of the discrete OMT problem (claim i) of Theorem 4). Since for the former an efficient iterative algorithm is available [22], we see that also in this discrete setting the SBP provides a valuable computational approach to solving OMT problems. This is illustrated in Section VI through an academic example. It should also be observed that the measure $\mathfrak{M}_T^*(\delta_1, \delta_n)$ is just a "Boltzmann measure" on the subset of $\mathcal{X}^{N+1}$ of paths originating in x_0 and terminating in x_N. Thus, the above proof is analogous to the classical one for P_T^*.

VI. Examples

Consider the graph in Fig. 1. We seek to transport a unit mass from node 1 to node 9 in $N = 3$ and 4 steps. We first consider the case where the costs of all the edges are equal to 1. Here we add a zero cost self-loop at 9, i.e., $l_{99} = 0$. The shortest path from node 1 to 9 is of length 3 and there are three such paths, which are $1 - 2 - 7 - 9$, $1 - 3 - 8 - 9$, and $1 - 4 - 8 - 9$. If we want to transport the mass with a minimum number of steps, we may end up using one of these three paths. To achieve robustness, we apply the Schrödinger bridge framework. Since

---

[2]Actually, using (28), it is easy to see that $L(\mathfrak{M}_T^*) = \mathbb{E}_{\mathfrak{M}_T^*}[l(Y)]$ is a strictly increasing function of T. Indeed,

$$\frac{\partial \mathbb{E}_{\mathfrak{M}_T^*}[l(Y)]}{\partial T} = \frac{1}{T^2} \operatorname{Var}_{\mathfrak{M}_T^*}[l(Y)]$$

where $l(Y) = \sum_{t=0}^{N-1} l_{Y_t Y_{t+1}}$ and $Y = (Y_0, Y_1, \ldots, Y_N)$ is the Markov chain. In view of Points 1 and 2, we conclude that $\mathbb{E}_{\mathfrak{M}_T^*}[l(Y)]$ bijectively maps $[0, +\infty]$ onto

$$\left[l_m(x_0, x_N), \frac{1}{|\mathcal{X}^{N+1}(x_0, x_N)|} \sum_{y \in \mathcal{X}^{N+1}(x_0, x_N)} l(y) \right].$$

all the three feasible paths have equal length, we get a transport plan with equal probabilities using all these three paths, regardless of the choice of temperature T. The evolution of mass distribution is given by

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 2/3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

where the four rows of the matrix show the mass distribution at time step $t = 0, 1, 2, 3$, respectively. As we can see, the mass spreads out first and then goes to node 9. When we allow for more steps $N = 4$, the mass spreads even more before reassembling at node 9, as follows, for $T = 1$:

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.4705 & 0.3059 & 0.2236 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.0823 & 0.0823 & 0.1645 & 0 & 0.2236 & 0.4473 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.0823 & 0.0823 & 0.1645 & 0.6709 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}.
$$

There are 7 feasible paths of length 4, which are $1 - 2 - 7 - 9 - 9, 1 - 3 - 8 - 9 - 9, 1 - 4 - 8 - 9 - 9, 1 - 2 - 5 - 6 - 9, 1 - 2 - 5 - 7 - 9, 1 - 3 - 4 - 8 - 9$, and $1 - 2 - 3 - 8 - 9$. The amount of mass traveling along these paths are

$$0.2236, 0.2236, 0.2236, 0.0823, 0.0823, 0.0823, 0.0823.$$

The first three are the most probable paths. This is consistent with Proposition 1 since they are the paths with minimum length. If we change the temperature T, the flow changes. The set of most probable paths, however, remains invariant. In particular, when $T = 0.1$, the flow concentrates on the most probable set (effecting OMT-like transport), as follows:

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.3334 & 0.3333 & 0.3333 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.3334 & 0.6666 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}.
$$

Now we change the graph by setting the length of edge $(7, 9)$ as 2, that is, $l_{79} = 2$. When $N = 3$ steps are allowed to transport a unit mass from node 1 to node 9, the evolution of mass distribution for the optimal transport plan, for $T = 1$, is given by

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.1554 & 0.4223 & 0.4223 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.1554 & 0.8446 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}.
$$

The mass travels through paths $1 - 2 - 7 - 9, 1 - 3 - 8 - 9$, and $1 - 4 - 8 - 9$, but unlike the first case, the transport plan does not take equal probability for these three paths. Since the length of the edge $(7, 9)$ is larger, the probability that the mass takes this path becomes smaller. The plan does, however, assign equal probability to the two paths $1 - 3 - 8 - 9$ and $1 - 4 - 8 - 9$ with minimum length, that is, these are the most probable paths. The evolutions of mass for $T = 0.1$

and $T = 100$ are

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

and

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.3311 & 0.3344 & 0.3344 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.3311 & 0.6689 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

respectively. We observe that, when $T = 0.1$ the flow assigns almost equal mass to the three available paths, while, when $T = 100$, the flow concentrate on the most probable paths $1 - 3 - 8 - 9$ and $1 - 4 - 8 - 9$. This is clearly a consequence of Theorem 4.

VII. Conclusions and Outlook

In this paper, we considered transportation over strongly connected, directed graphs. The development built on our earlier work [13]. More specifically, we introduced as measure of efficiency for a given transportation plan the average path length (cost) in, e.g., (13), and as a measure of robustness the entropy (14). This allowed us to explore efficient robust transport plans via solving corresponding optimization problems. Important insights gained in this paper include the results on certain invariances of the Schrödinger's bridges—the "iterated bridge" invariance property and the invariance of the "most probable path." We explained their relevance for efficient robust transport over networks. We also considered the dependence of the optimal transportation schedule on the temperature parameter following similar ideas from statistical physics. In this, we highlighted the connection between the SBP and OMT. Specifically, the solution of the SBP near-zero temperature is an approximation to the solution of the corresponding OMT problem. The relevance of the conceptual framework developed here for assessing robustness of real-world networks (e.g., communication networks [47], biological [2], [40], and financial [41]) will be the subject of future work.

References

[1] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, 2000.

[2] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. London, U.K.: Chapman and Hall, 2006.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures (Lectures in Mathematics)*, 2nd ed. Basel, Switzerland: Birkhäuser-Verlag, 2008.

[4] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows*. Hoboken, NJ, USA: Wiley, 2011.

[5] A.-L. Barabàsi, *Network Science*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[6] N. K. Olver, "Robust network design," Ph.D. dissertation, Dept. Math. Statist., McGill Univ., Montreal, QC, Canada, 2010.

[7] L. Arnold, V. M. Gundlach, and L. Demetrius, "Evolutionary formalism for products of positive random matrices," *Ann. Appl. Probab.*, vol. 4, no. 3, pp. 859–901, 1994.

[8] A. Blaquière, "Controllability of a Fokker-Planck equation, the Schrödinger system, and a related stochastic optimal control (revised version)," *Dyn. Control*, vol. 2, no. 3, pp. 235–253, 1992.

[9] R. Brockett, "Notes on the control of the Liouville equation," in *Control of Partial Differential Equations*. New York, NY, USA: Springer-Verlag, 2012, pp. 101–129.

[10] Y. Chen, T. T. Georgiou, and M. Pavon, "On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint," *J. Optim. Theory Appl.*, vol. 169, no. 2, pp. 671–691, 2016.

[11] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal transport over a linear dynamical system," *IEEE Trans. Automat. Control*, vol. 62, no. 5, pp. 1558–2523, May 2017.

[12] Y. Chen, T. T. Georgiou, and M. Pavon, "Entropic and displacement interpolation: A computational approach using the Hilbert metric," *SIAM J. Appl. Math.*, vol. 76, no. 6, pp. 2375–2396, 2016.

[13] Y. Chen, T. T. Georgiou, M. Pavon, and A. Tannenbaum, "Robust transport over networks," *IEEE Trans. Automat. Control*, vol. 62, no. 9, pp. 4675–4682, Sep. 2017.

[14] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[15] C. Imre, "Sanov property, generalized I-projection and a conditional limit theorem," *Ann. Probab.*, vol. 12, pp. 768–793, 1984.

[16] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2292–2300.

[17] P. Dai Pra, "A stochastic control approach to reciprocal diffusion processes," *Appl. Math. Optim.*, vol. 23, no. 1, pp. 313–329, 1991.

[18] P. Dai Pra and M. Pavon, "On the Markov processes of Schrödinger, the Feynman–Kac formula and stochastic control," in *Realization and Modelling in System Theory*. New York, NY, USA: Springer-Verlag, 1990, pp. 497–504.

[19] L. Demetrius and T. Manke, "Robustness and network evolution: An entropic principle," *Phys. A, Statist. Mech. Appl.*, vol. 346, no. 3, pp. 682–696, 2005.

[20] D. Durr and A. Bach, "The Onsager-Machlup functional as Lagrangian for the most probable path of a diffusion process," *Commun. Math. Phys.*, vol. 60, is. 1978, pp. 153–170.

[21] R. Filliger, M.-O. Hongler, and L. Streit, "Connection between an exactly solvable stochastic optimal control problem and a nonlinear reaction-diffusion equation," *JOTA*, vol. 137, no. 3, pp. 497–505, 2008.

[22] T. T. Georgiou and M. Pavon, "Positive contraction mappings for classical and quantum Schrödinger systems," *J. Math. Phys.*, vol. 56, no. 3, 2015, Art. no. 033301.

[23] A. Galichon, S. Kominers, and S. Weber, "The nonlinear Bernstein-Schrödinger equation in economics," in *Proc. 2nd Conf. Netw. Geometric Sci. Inf.*, 2015, pp. 51–59.

[24] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[25] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. Amsterdam, The Netherlands: North-Holland, 1989.

[26] C. T. Ireland and S. Kullback, "Contingency tables with given marginals," *Biometrika*, vol. 55, pp. 179–188, 1968.

[27] B. Jamison, "Reciprocal processes," *Probab. Theory Relat. Fields*, vol. 30.1, pp. 65–86, 1974.

[28] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Phys. Rev. Lett.*, vol. 87, no. 19, 2001, Art. no. 198701.

[29] C. Léonard, "From the Schrödinger problem to the Monge-Kantorovich problem," *J. Funct. Anal.*, vol. 262, pp. 1879–1920, 2012.

[30] C. Léonard, "A survey of the Schroedinger problem and some of its connections with optimal transport," *Discrete Continuous Dyn. Syst. A*, vol. 34, no. 4, pp. 1533–1574, 2014.

[31] B. C. Levy, R. Frezza, and A. J. Krener, "Modeling and estimation of discrete-time Gaussian reciprocal processes," *IEEE Trans. Automat. Control*, vol. 35, no. 9, pp. 1013–1023, Sep. 1990.

[32] T. Mikami, "Monge's problem with a quadratic cost by the zero-noise limit of h-path processes," *Probab. Theory Related Fields*, vol. 129, pp. 245–260, 2004.

[33] T. Mikami and M. Thieullen, "Optimal transportation problem by stochastic optimal control," *SIAM J. Control Optim.*, vol. 47, no. 3, pp. 1127–1139, 2008.

[34] L. Onsager and S. Machlup, "Fluctuations and irreversible proceses, I–II," *Phys. Rev.*, vol. 91, pp. 1505–1512; 1512–1515, 1953.

[35] M. Pavon, *Lecture notes, Univ. Padova, Padova, Padua, Italy*, 2015. [Online]. Available: http://www.math.unipd.it/~pavon/Teaching_files/notes.pdf

[36] M. Pavon and F. Ticozzi, "Discrete-time classical and quantum Markovian evolutions: Maximum entropy problems on path space," *J. Math. Phys.*, vol. 51, no. 4, 2010, Art. no. 042104.

[37] M. Pavon and A. Wakolbinger, "On free energy, stochastic control, and Schrödinger processes," in *Modeling, Estimation and Control of Systems With Uncertainty*. New York, NY, USA: Springer-Verlag, 1991, pp. 334–348.

[38] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory*, vol. 1. New York, NY, USA: Springer Sci. Bus. Media, 1998.

[39] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[40] R. Sandhu *et al.*, "Graph curvature for differentiating cancer networks," *Sci. Rep.*, vol. 5, 2015, Art. no. 12323, doi: 10.1038/srep12323.

[41] R. Sandhu, T. Georgiou, and A. Tannenbaum, "Ricci curvature: An economic indicator for market fragility and systemic risk," *Sci. Adv.*, vol. 2, 2016, Art. no. e1501495, doi: 10.1126/sciadv.1501495.

[42] K. Savla, G. Como, and M. A. Dahleh, "Robust network routing under cascading failures," *IEEE Trans. Netw. Sci. Eng.*, vol. 1, no. 1, pp. 53–66, Jan.–Jun. 2014.

[43] E. Schrödinger, "Über die Umkehrung der Naturgesetze," *Sitzungsberichte der Preuss Akad. Wissen., Phys. Math. Klasse*, pp. 144–153, 1931.

[44] Y. Takahashi and S. Watanabe, "The probability functionals (onsager-machlup functions) of diffusion processes," in *Stochastic Integrals (Lecture Notes Math.)*, vol. 851, Berlin, Germany: Springer-Verlag, 1980, pp. 433–463.

[45] C. Villani, *Topics in Optimal Transportation*, vol. 58. Providence, RI, USA: AMS, 2003, vol. 58.

[46] C. Villani, *Optimal Transport: Old and New*, vol. 338. New York, NY, USA: Springer-Verlag, 2008.

[47] C. Wang, E. Jonckheere, and R. Banirazi, "Wireless network capacity versus Ollivier-Ricci curvature under heat diffusion (HD) protocol," *Proc. Amer. Control Conf.*, 2013, pp. 3536–3541.

[48] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[49] L. B. White and F. Carravetta, "Optimal smoothing for finite state hidden reciprocal processes," *IEEE Trans. Automat. Control*, vol. 56, no. 9, pp. 2156–2161, Sep. 2011.