UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Padua Research Archive - Institutional Repository

Robust Transport over Networks

(Article begins on next page)

Robust transport over networks

Yongxin Chen, Tryphon Georgiou, Michele Pavon and Allen Tannenbaum

Abstract—We consider transportation over a strongly connected, directed graph. The scheduling amounts to selecting transition probabilities for a discrete-time Markov evolution which is designed to be consistent with initial and final marginal constraints on mass transport. We address the situation where initially the mass is concentrated on certain nodes and needs to be transported over a certain time period to another set of nodes, possibly disjoint from the first. The evolution is selected to be closest to a *prior* measure on paths in the relative entropy sense–such a construction is known as a Schrödinger bridge between the two given marginals. It may be viewed as an atypical stochastic control problem where the control consists in suitably modifying the prior transition mechanism. The prior can be chosen to incorporate constraints and costs for traversing specific edges of the graph, but it can also be selected to allocate *equal probability to all paths of equal length connecting any two nodes* (i.e., a uniform distribution on paths). This latter choice for prior transitions relies on the so-called *Ruelle-Bowen random walker* and gives rise to scheduling that tends to utilize all paths as uniformly as the topology allows. Thus, this Ruelle-Bowen law ($\mathfrak{M}_{RB}$) taken as prior, leads to a transportation plan that tends to lessen congestion and ensures a level of robustness. We also show that the distribution $\mathfrak{M}_{RB}$ on paths, which attains the maximum entropy rate for the random walker given by the topological entropy, can itself be obtained as the time-homogeneous solution of a maximum entropy problem for measures on paths (also a Schrödinger bridge problem, albeit with prior that is not a probability measure). Finally we show that the paradigm of Schrödinger bridges as a mechanism for scheduling transport on networks can be adapted to graphs that are not strongly connected, as well as to weighted graphs. In the latter case, our approach may be used to design a transportation plan which effectively compromises between robustness and other criteria such as cost. Indeed, we explicitly provide a robust transportation plan which assigns *maximum probability* to *minimum cost paths* and therefore compares favorably with Optimal Mass Transportation strategies.

I. Introduction

Transport over networks has been the focus of a rapidly expanding literature due to its intrinsic relevance in a wide range of applications that include power transmission, traffic, financial transactions, biological systems and so on [1], [2], [3], [4]. Furthermore, the topic relates to a host of other questions pertaining to the connectivity of graphs and the relative significance of their nodes as in the Google PageRank problem [5] and the study of interaction between genes in biological networks [6].

Our starting point is an important insight on the relation between the topological structure of a network and the entropy rate of a random walker on the graph [7], [8]. As it turns out, there is a unique way to specify transition probabilities at each node in such a way that all paths of equal length joining any two particular nodes

Y. Chen is with the Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota MN 55455, USA; chen2468@umn.edu

T.T. Georgiou is with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, USA; tryphon@uci.edu

M. Pavon is with the Dipartimento di Matematica "Tullio Levi Civita", Università di Padova, via Trieste 63, 35121 Padova, Italy; pavon@math.unipd.it

A. Tannenbaum is with the Department of Computer Science, Stony Brook University, Stony Brook, NY 11794: allen.tannenbaum@stonybrook.edu

have equal probability. Thereby, a measure is placed on the family of paths between graph nodes that maximizes the entropy rate of a random walker, and this is a characteristic of the network. So far, the use of this concept has been to assign significance to each node in relation to the corresponding occupancy stationary distribution (*centrality measures*).

The focus in our paper is on how to schedule transportation plans across a network. The novel framework that we propose is that of the so-called Schrödinger bridge problem, where a flow is specified in agreement with an initial and a final marginal distribution on the nodes while, at the same time, the probability law on the paths is the closest possible to a prior in the relative entropy sense. The Ruelle-Bowen random walk provides a natural notion of "uniform" prior which gives equal importance to all paths. As a result, the transportation flow that is selected to agree with specified initial and final marginals tends to spread across all available paths as much as possible given the topological structure of the network. Thereby, such a flow leads to relatively low probability of conflict and congestion, and ensures a certain degree of inherent robustness of the transport plan. It is well appreciated that, typically, robustness, efficiency and cost are conflicting criteria when designing networks.

By extending our approach to weighted graphs, we show that the choice of a prior distribution may be used to ensure that the resulting transportation attains a satisfactory compromise between robustness and other criteria such as cost. Indeed, we exhibit a *robust* transportation plan which assigns *maximum probability* to all *minimum cost paths*. It appears attractive when compared to Optimal Mass Transportation strategies which are not necessarily robust and where the minimum cost of transportation between any two nodes is supposed to be given. Thus, the approach to scheduling transport based on Schrödinger bridges affords great flexibility. Moreover, it appears computationally attractive in view of the iterative algorithm proposed in [9].

The paper is outlined as follows. In Section II, we present the solution to a general Schrödinger bridge problem (SBP), where the prior measure is not necessarily a probability measure, as a straightforward extension of the results in [10], [9]. Section III is devoted to solutions of the SBP with equal initial and final marginals which have a *time-invariant* transition mechanism so that they admit invariant measures. We establish the surprising result (Theorem III.4) that there is only one such bridge. This measure on paths can be constructed generalizing a classical result by Parry [11]. In Section IV, considering the special case of a prior transition given by the adjacency matrix, we describe the most important features of the Ruelle-Bowen random walker along the lines of [7]. We observe that this measure $\mathfrak{M}_{RB}$ on trajectories can be viewed as a solution to a "time-homogeneous" Schrödinger bridge problem where the prior transition mechanism is given by the adjacency matrix. Section V describes our procedure to produce a robust transportation plan over a given strongly connected network: We take the Ruelle-Bowen distribution $\mathfrak{M}_{RB}$ as prior in a Schrödinger bridge problem with prescribed initial and final marginals. We also prove that the optimal transportation can also be obtained in one step by taking the

adjacency matrix as prior transition mechanism (Proposition V.2). In Section VI, we outline our approach in the cases of weighted and not strongly connected graphs. Finally, in Section VII we illustrate our approach on a simple unweighted and weighted graph.

II. THE DISCRETE SCHRÖDINGER BRIDGE PROBLEM

We first describe the "ingredients" of the discrete Schrödinger Bridge problem (SBP) considered in [10], [9]. In fact, we will consider a slight generalization, where the "prior" is not necessarily a probability law. The goal is to determine a time-evolution of probability distributions $\nu_t(\cdot)$ having support on a discrete space

$$\mathcal{X} = \{1, \ldots, n\},$$

e.g., the nodes of a network, over a time-indexing set

$$\mathcal{T} = \{0, 1, \ldots, N\}$$

in a way such that it matches the specified marginal distributions $\nu_0(\cdot)$ and $\nu_N(\cdot)$ and the resulting random evolution is closest to the "prior" in a suitable sense. Regarding notation, we use $\mu_t(\cdot)$, $\nu_t(\cdot)$ for distributions, where typically, μ relates to a "prior" law while ν represents a "new" distribution with end-points specified and obtained by solving the SBP.

The prior law is induced by the Markovian evolution

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

for nonnegative distributions $\mu_t(\cdot)$ over $\mathcal{X}$ with $t \in \mathcal{T}$. Throughout, we assume that $m_{ij} \geq 0$ for all indices $i, j \in \mathcal{X}$ and for simplicity, for the most part, that the matrix

$$M = [m_{ij}]_{i,j=1}^n$$

does not depend on t. In this case, we will often assume that all entries of M^N are positive. The rows of the transition matrix M do not necessarily sum up to one, in which case the "total transported mass" is not necessarily preserved. This is the case, in particular, of a Markov chain with "creation" and "killing". In fact, M may simply encode the topological structure of a directed network with m_{ij} being zero or one, depending whether a certain transition is allowed.

The evolution (1), together with measure $\mu_0(\cdot)$, which we assume positive on $\mathcal{X}$, i.e.,

$$\mu_0(x) > 0 \text{ for all } x \in \mathcal{X}, \quad (2)$$

induces a measure $\mathfrak{M}$ on $\mathcal{X}^{N+1}$ as follows. It assigns to a path[1] $x = (x_0, x_1, \ldots, x_N) \in \mathcal{X}^{N+1}$ the value

$$\mathfrak{M}(x_0, x_1, \ldots, x_N) = \mu_0(x_0) m_{x_0 x_1} \cdots m_{x_{N-1} x_N}, \quad (3)$$

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

$$\mu_t(x_t) = \sum_{x_0, x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_N} \mathfrak{M}(x_0, x_1, \ldots, x_N), \quad t \in \mathcal{T}.$$

The "prior" distribution $\mathfrak{M}$ on the space of paths may be at odds with a pair of specified marginals ν_0 and ν_N in that one or possibly both,

$$\mu_0(x_0) \neq \nu_0(x_0), \quad \mu_N(x_N) \neq \nu_N(x_N).$$

[1] Here a path is a sequence of adjacent nodes, which is different to most literature where a path does not pass through the same node more than once.

We denote by $\mathcal{P}(\nu_0, \nu_N)$ the family of probability distributions on $\mathcal{X}^{N+1}$ having the prescribed marginals. We seek a distribution in this set which is closest to the prior $\mathfrak{M}$ in a suitable entropic sense. To this end, let us first recall the definition of relative entropy for probability distributions.

Definition II.1. *The* Relative Entropy *between the probability distributions P and Q is*

$$\mathbb{D}(P\|Q) := \begin{cases} \sum_x P(x) \log \frac{P(x)}{Q(x)}, & \mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q), \\ +\infty, & \mathrm{Supp}(P) \not\subseteq \mathrm{Supp}(Q), \end{cases}$$

where, by definition, $0 \cdot \log 0 = 0$ and the summation is over the common set where they are supported.

As is well known, $\mathbb{D}(P\|Q)$ is not symmetric and does not satisfy the triangle inequality. It does, however, satisfy $\mathbb{D}(P\|Q) \geq 0$ and $\mathbb{D}(P\|Q) = 0$ if and only if $Q = P$, see, e.g., [12]. It can also be extended to positive measures that are not probability distributions. In fact, it is quite common to consider Q to be a uniform measure that may not be a probability measure, such as the Lebesgue measure or the stationary Wiener measure [13]. Naturally, while the value of $\mathbb{D}(P\|Q)$ may turn out negative due to miss-match of scaling, the relative entropy is always jointly convex. We view the prior $\mathfrak{M}$ (specified by M and μ_0) in a similar manner, and consider the Schrödinger Bridge problem:

Problem II.2. *Determine*

$$\mathfrak{M}^*[\nu_0, \nu_N] = \operatorname{argmin}\{\mathbb{D}(P\|\mathfrak{M}) \mid P \in \mathcal{P}(\nu_0, \nu_N)\}. \quad (4)$$

Provided all entries of M^N are positive, the problem has a solution, which is unique due to strict convexity. This is stated next.

Theorem II.3. *For a nonnegative square matrix M such that M^N has all positive entries and positive measures ν_0 and ν_N on $\mathcal{X}$, there exist nonnegative functions $\varphi(\cdot)$ and $\hat{\varphi}(\cdot)$ on $[0, N] \times \mathcal{X}$ satisfying for $t \in [0, N-1]$ the system*

$$\varphi(t, i) = \sum_j m_{ij} \varphi(t+1, j), \quad (5a)$$

$$\hat{\varphi}(t+1, j) = \sum_i m_{ij} \hat{\varphi}(t, i), \quad (5b)$$

with the boundary conditions

$$\varphi(0, x_0) \cdot \hat{\varphi}(0, x_0) = \nu_0(x_0) \quad (5c)$$

$$\varphi(N, x_N) \cdot \hat{\varphi}(N, x_N) = \nu_N(x_N), \quad (5d)$$

for all $x_0, x_N \in \mathcal{X}$. Moreover, the solution $\mathfrak{M}^*[\nu_0, \nu_N]$ to Problem II.2 is unique and obtained by

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

where[2]

$$\pi_{ij}(t) := m_{ij} \frac{\varphi(t+1, j)}{\varphi(t, i)}. \quad (6)$$

Equation (6) *specifies* one-step transition probabilities *that are well defined.*

Proof. The argument in [10, Theorem 4.1] and [9, Section III] applies verbatim to this setting which is slightly more general in that M does not prescribe a probability kernel. The system (5a-5d) is known as a

[2] Here we use the convention that $0/0 = 0$.

Schrödinger system. The existence of solution is shown in [9, Section III] by establishing that the composition

$$\hat{\varphi}(0,x_0) \overset{(M^T)^N}{\longrightarrow} \hat{\varphi}(N,x_N) \overset{(5d)}{\longrightarrow} \varphi(N,x_N) \longrightarrow \dots$$

$$\dots \overset{M^N}{\longrightarrow} \varphi(0,x_0) \overset{(5c)}{\longrightarrow} (\hat{\varphi}(0,x_0))_{\text{next}} \qquad (7)$$

is contractive in the Hilbert metric [14], [15]. The fact that $\pi_{ij}(t)$ in (6) satisfy $\sum_j \pi_{ij}(t) = 1$ follows from (5a). $\qquad \square$

Notice that φ and $\hat{\varphi}$ are unique up to multiplication of φ by a positive constant and division of $\hat{\varphi}$ by the same constant. This is a direct consequence of the proof above as the Hilbert metric is a metric on the projective space. The statement of the theorem is analogous to results for the classical Schrödinger system (5) of diffusions see[13]. The requirement that M^N has only positive entries can be slightly relaxed and replaced by a suitable condition that guarantees existence of solution for the particular ν_0 and ν_N. The case when M is time-varying can also be readily established along the lines of [10, Theorem 4.1] and [9, Theorem 2].

Finally, to simplify the notation, let $\varphi(t)$ and $\hat{\varphi}(t)$ denote the column vectors with entries $\varphi(t,i)$ and $\hat{\varphi}(t,i)$, respectively, with $i \in \mathcal{X}$. In matrix form, (5a), (5b) and (6) read

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

and

$$\Pi(t) = [\pi_{ij}(t)] = \text{diag}(\varphi(t))^{-1} M \text{diag}(\varphi(t+1)). \qquad (8b)$$

III. TIME-HOMOGENEOUS BRIDGES

In this section, we consider the case of Schrödinger bridge problems when the marginals are identical, namely, $\nu_0 = \nu_N = \nu$. In particular, we are interested in the case when the solution of the SBP corresponds to a time-homogeneous Markov evolution. Note that, from Theorem II.3, $\mathfrak{M}^*[\nu,\nu]$ is in general time inhomogeneous. We first recall the following celebrated result on the spectral properties of nonnegative matrices [16].

Theorem III.1 (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;*
  iv) *v is the only non-negative eigenvector of A;*
  v) *let $B = [b_{ij}]$ be an $n \times n$ matrix with nonnegative entries. If $a_{ij} \le b_{ij}$, $\forall i,j \le n$ and $A \ne B$, then $\lambda_A < \lambda_B$.*

Since the nonnegative matrix M is such that M^N has only positive entries, by the above Perron-Frobenius Theorem, M has a unique positive eigenvalue λ_M which is equal to the spectral radius. Let ϕ and $\hat{\phi}$ be the corresponding right and left eigenvectors and denote their entries by $\phi(x)$ and $\hat{\phi}(x)$ with $x \in \mathcal{X}$, respectively. Then both have only positive entries. We normalize ϕ and $\hat{\phi}$ so that

$$\sum_{x \in \mathcal{X}} \phi(x)\hat{\phi}(x) = 1.$$

This leads to a special probability distribution

$$\bar{\nu}(x) = \phi(x)\hat{\phi}(x). \qquad (9)$$

It turns out that $\bar{\nu}$ is the only probability measure such that the associated SBP has a time-homogeneous solution; we shall name it the *time-homogeneous bridge* associated with M. It admits the following variational characterization.

Proposition III.2. *Let M be a nonnegative matrix such that M^N has only positive entries, and $\mathfrak{M}$ the measure on $\mathcal{X}^{N+1}$ given by (3) with μ_0 satisfying (2). Then the solution to the Schrödinger bridge problem*

$$\mathfrak{M}^*[\bar{\nu},\bar{\nu}] = \text{argmin}\{\mathbb{D}(P\|\mathfrak{M})|P \in \mathcal{P}(\bar{\nu},\bar{\nu})\}, \qquad (10)$$

where $\bar{\nu}$ is as in (9), has the time-invariant transition matrix

$$\bar{\Pi} = \lambda_M^{-1} \text{diag}(\phi)^{-1} M \text{diag}(\phi) \qquad (11)$$

and invariant measure $\bar{\nu}$.

Proof. Since ϕ and $\hat{\phi}$ are the right and left eigenvectors of M associated with eigenvalue λ_M, the nonnegative functions φ and $\hat{\varphi}$ defined by

$$\varphi(t,x) = \lambda_M^t \phi(x), \quad \hat{\varphi}(t,x) = \lambda_M^{-t}\hat{\phi}(x)$$

satisfy the Schrödinger system (5). By Theorem II.3, the solution $\mathfrak{M}^*[\bar{\nu},\bar{\nu}]$ of the Schrödinger bridge problem (10) then has the transition matrix (see (8b))

$$\begin{aligned}\bar{\Pi} &= \text{diag}(\varphi(0))^{-1} M \text{diag}(\varphi(1)) \\ &= \lambda_M^{-1}\text{diag}(\phi)^{-1} M \text{diag}(\phi),\end{aligned}$$

which is exactly (11). Moreover, since

$$\bar{\Pi}^T \bar{\nu} = \lambda_M^{-1}\text{diag}(\phi)M^T\hat{\phi} = \bar{\nu},$$

it follows that $\bar{\nu}$ is the corresponding invariant measure. $\qquad \square$

In particular, notice that $\mathfrak{M}^*[\bar{\nu},\bar{\nu}]$, and its extension to infinite paths $x = (x_0,x_1,x_2,\dots)$ through (11), is *stationary*. Indeed, we have the following more general result which is of independent interest.

Proposition III.3. *Let $P \in \mathcal{P}(\nu,\nu)$ be a Markovian measure on $\mathcal{X}^{N+1}$ having time-invariant transition matrix Π. Then ν is invariant for Π, i.e. $\Pi^T\nu = \nu$.*

Proof. Let $\Pi^T\nu = m$. Then

$$d_H(\nu,m) = d_H((\Pi^T)^N\nu, (\Pi^T)^N m) \le \lambda d_H(\nu,m)$$

where d_H is the Hilbert distance [14], [9] and $\lambda < 1$ is the contraction ratio of the map $(\Pi_\nu^T)^N$. Since both ν and m are probability distributions, it follows that $m = \nu$ and ν is invariant. $\qquad \square$

We show next that, under mild assumptions, there is only one time-homogeneous bridge between equal marginals. In the following result, we shall use the following notation. As before, let $\mathfrak{M}$ be given by (3) with μ_0 satisfying (2). We denote by $\mathfrak{M}(2N)$ the unique extension of $\mathfrak{M}$ to all of the discrete interval $[0,2N]$ by its time-invariant transition mechanism. We also denote by $\mathfrak{M}^*[\nu,\nu](2N)$ the Schrödinger bridge with prior $\mathfrak{M}(2N)$ and equal marginals ν at times $t = 0$ and $t = 2N$.

Theorem III.4. *Let M be a nonnegative matrix such that M^N has only positive entries. Let ν be a probability measure. Suppose $N > 1$ and that the transition matrix Π_ν of $\mathfrak{M}^*[\nu,\nu](2N)$ does not depend on time. Then $\nu = \bar{\nu}$ as in (9) and $\Pi_\nu = \bar{\Pi}$ as in (11).*

Proof. Let $\varphi_\nu(t) = M\varphi_\nu(t+1)$ be the space-time harmonic function associated to the minimizer $\mathfrak{M}^*[\nu, \nu]$. Suppose first that M has only positive entries and consider times $t = N-2, N-1, N$. By (8) and the time invariance of Π_ν, we must have

$$\Pi_\nu = \mathrm{diag}(\varphi_\nu(N-2))^{-1} M \, \mathrm{diag}(\varphi_\nu(N-1))$$
$$= \mathrm{diag}(\varphi_\nu(N-1))^{-1} M \, \mathrm{diag}(\varphi_\nu(N)).$$

It follows that

$$M = D_\nu(N-1) M D_\nu(N)^{-1},$$

where

$$D_\nu(t) = \mathrm{diag}(\varphi_\nu(t)) \, \mathrm{diag}(\varphi_\nu(t-1))^{-1}$$
$$= \mathrm{diag}(d_1^\nu(t), \ldots, d_n^\nu(t))$$

is diagonal for all t. Hence,

$$m_{ij} = e_i^T M e_j = d_i(N-1) m_{ij} d_j(N)^{-1}, \quad \forall i, j.$$

Varying j for a fixed i, since $m_{ij} \neq 0$, we get that $D(N)$ is a scalar matrix, say λI, not dependent on t and $\varphi(N)$ is a right eigenvector of M. By the Perron-Frobenius Theorem, it follows that $\varphi(N)$ corresponds to λ_M. It readily follows that $\hat{\varphi}(0)$ is an eigenvector of M^T with positive entries corresponding to the same eigenvalue λ_M. By (5c)-(5d), ν is equal to $\bar{\nu}$.

A similar argument establishes the result when M has merely nonnegative entries. Indeed, looking at the N-step transition matrix $\Pi_\nu^{(N)} = \Pi_\nu^N$ on the time intervals $[0, N]$ and $[N, 2N]$ the same argument as in the full positive case gives that $\varphi(2N)$ is a right eigenvector of M^N with positive entries. But so is φ. By Theorem III.1, iv), they can be taken to be equal. $\qquad\square$

Consider now the following special case. We have a strongly connected, aperiodic directed graph $(\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V} = \{1, 2, \ldots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Let A be the adjacency matrix of the graph so that $a_{ij} = 1$ if there is an edge from i to j and $a_{ij} = 0$ otherwise. Then, there exists N such that A^N has all positive entries. As we shall see in the next section, the Schrödinger bridge problem (10) just considered with $M = A$ as prior transition turns out to have as solution the Ruelle-Bowen measure $\mathfrak{M}_{\mathrm{RB}}$ [7, Section III]. This probability measure has a number of useful properties, in particular it gives the same probability to paths of the same length between any two given nodes. All of this is discussed in the next section.

IV. THE RUELLE-BOWENS RANDOM WALK

In this section, we follow closely the beautiful paper [7] by Delvenne and Libert, which explains the Ruelle-Bowens (RB) random walk. The RB random walk amounts to a Markovian evolution on a directed graph that assigns equal probabilities to all paths of equal length between any two nodes. The motivation of [7] was to assign a natural invariant probability to nodes based on relations that are encoded by a graph, and thereby determine a *centrality measure*, akin to Google Page ranking, yet more robust and discriminating. Our motivation is quite different. The RB random walk provides a uniform distribution on paths. Therefore, it represents a natural distribution to serve as prior in the SBP in order to achieve a maximum spreading of the mass transported over the available paths. In this section, besides reviewing basics on the RB random walk, we show that the RB distribution is itself a solution to the Schrödinger bridge problem II.2.

We consider a strongly connected, directed graph

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}).$$

The idea in Google Page ranking the nodes is based on a random walk where a jump takes place from one node to any of its neighbors with equal probability. The alternative proposed in [7] is an *entropy ranking*, based on the stationary distribution of the RB random walk [11], [17]. The transition mechanism is such that it induces a uniform distribution on *paths* of equal length joining any two nodes. This distribution is characterized as the one maximizing the *entropy rate* [12] for the random walker. Let us briefly recall the relevant concept. The Shannon entropy for paths of length t is at most

$$\log |\{\text{paths of length } t\}|.$$

Hence, the entropy rate is bounded by the *topological entropy rate*

$$H_\mathcal{G} = \limsup_{t\to\infty}[\log |\{\text{paths of length } t\}|/t].$$

Here $|\{\cdot\}|$ denotes the cardinality of a set. Notice that $H_\mathcal{G}$ only depends on the graph $\mathcal{G}$ and not on the probability distribution on paths. More specifically, if A denotes the adjacency matrix of the graph, the number of paths of length t is the sum of all the entries of A^t. Thus, it follows that $H_\mathcal{G}$ is the logarithm of the spectral radius of A, namely the maximum of the absolute values of the eigenvalues of A, that is

$$H_\mathcal{G} = \log(\lambda_A). \tag{12}$$

We next construct the Rulle-Bowen random walk. Let A as in the Perron-Frobenius Theorem III.1 and let u and v be its left and right eigenvectors[3] with positive entries corresponding to λ_A, so that

$$A^T u = \lambda_A u, \quad Av = \lambda_A v. \tag{13}$$

Suppose u and v are chosen so that

$$\langle u, v \rangle := \sum_i u_i v_i = 1.$$

As in the previous section, it is readily seen that their entrywise multiplication

$$\nu_{RB}(i) = u_i v_i \tag{14}$$

defines a probability distribution which is invariant under the transition matrix

$$R = [r_{ij}], \quad r_{ij} = \frac{v_j}{\lambda_A v_i} a_{ij}. \tag{15}$$

that is,

$$R^T \nu_{RB} = \nu_{RB}. \tag{16}$$

If A in (13) represents the adjacency matrix A of a graph, then the transition matrix R in (15) together with the stationary measure ν_{RB} in (14), define the *Ruelle-Bowen path measure*

$$\mathfrak{M}_{\mathrm{RB}}(x_0, x_1, \ldots, x_N) := \nu_{RB}(x_0) r_{x_0 x_1} \cdots r_{x_{N-1} x_N}. \tag{17}$$

Proposition IV.1. *The measure $\mathfrak{M}_{\mathrm{RB}}$ (17) assigns probability $\lambda_A^{-t} u_i v_j$ to any path of length t from node i to node j.*

Proof. Starting from the stationary distribution (14), and in view of (15), the probability of a path ij is

$$u_i v_i \left(\frac{1}{\lambda_A} v_i^{-1} v_j \right) = \frac{1}{\lambda_A} u_i v_j,$$

[3]We are now conforming to notation in [7] for ease of comparison. Hence we use u and v rather than $\hat{\phi}$ and ϕ.

assuming that node j is accessible from node i in one step. Likewise, the probability of the path ijk is

$$u_i v_i \left(\frac{1}{\lambda_A} v_i^{-1} v_j \right) \left(\frac{1}{\lambda_A} v_j^{-1} v_k \right) = \frac{1}{\lambda_A^2} u_i v_k$$

independent of the intermediate state j, and so on. Thus, the claim follows. □

The striking property of $\mathfrak{M}_{\mathrm{RB}}$ is that induces a uniform probability measure on paths of equal length between any two given nodes. We quote from [7] "Since the number of paths of length t is of the order of λ_A^t (up to a factor) the distribution on paths of fixed length is uniform up to a factor (which does not depend on t). Hence the Shannon entropy of paths of length t grows as $t \log \lambda_A$, up to an additive constant. The entropy rate of this distribution is thus $\log \lambda_A$ which is optimal" by the expression for $H_{\mathcal{G}}$ in (12).

The analysis also shows that the Ruelle-Bowen distribution is the solution of the particular SBP where the "prior" transition mechanism is given by the adjacency matrix! This observation is apparently new and beautifully links the topological entropy rate to a maximum entropy problem on path space. We state next this special case of Proposition III.2.

Proposition IV.2. *Let A be the adjacency matrix of a strongly connected aperiodic graph $\mathcal{G}$. Let $\mathfrak{M}$ the nonnegative measure on $\mathcal{X}^{N+1}$ given by (3) with $M = A$ and μ_0 satisfying (2). Then, the Ruelle-Bowen measure $\mathfrak{M}_{\mathrm{RB}}$ (17) solves the SBP (10) with marginals $\nu_0 = \nu_N = \nu_{RB}$.*

V. ROBUST TRANSPORT OVER NETWORKS

Once again we consider a strongly connected, directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with n vertices. We identify node 1 as a *source* and node n as a *sink* and seek to transport a unit mass from 1 to n in at most N steps. The task is formalized by setting an initial marginal distribution $\nu_0(x) = \delta_1(x)$ Kronecker's delta. Similarly, the final distribution is $\nu_N(x) = \delta_n(x)$. Generally, we seek a transportation plan which is *robust* and avoids *congestion* as much as the topology of the graph permits. This latter feature of the transportation plan will be achieved in this section indirectly, without explicitly bringing into the picture the capacity of each edge (this is done in Section VI). With these two key specifications in mind, we like to control the flux so that the initial mass *spreads as much as possible* on the feasible paths joining vertices 1 and n in N steps before reconvening at time N in vertex n. We shall achieve this by constructing a suitable Markovian transition mechanism. As we want to allow for the possibility that all or part of the mass reaches node n at some time less than N, we always include a loop in node n so that our adjacency matrix A always has $a_{nn} = 1$. We observed in the previous section that the Ruelle-Bowen $\mathfrak{M}_{\mathrm{RB}}$ measure on paths can be obtained as the solution of the maximum entropy problem when the "prior transition matrix" is the adjacency matrix. Since $\mathfrak{M}_{\mathrm{RB}}$ gives equal probability to paths joining two specific vertices, it is natural to use it as a prior in a new maximum entropy problem with marginals δ_1, δ_n so as to achieve the spreading of the probability mass on the feasible paths joining the source with the sink. Thus, we consider the following maximum entropy problem

Problem V.1. *Determine*

$$\mathfrak{M}^*[\delta_1, \delta_n] = \operatorname{argmin}\{\mathbb{D}(P\|\mathfrak{M}_{\mathrm{RB}})|P \in \mathcal{P}(\delta_1, \delta_n)\}.$$

By Theorem II.3, the optimal, time varying transition matrix $\Pi^*(t)$ of the above problem is given, recalling the notations in (8), by

$$\Pi^*(t) = \operatorname{diag}(\varphi(t))^{-1} R \operatorname{diag}(\varphi(t+1)), \quad (18)$$

where

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

with the boundary conditions

$$\varphi(0,x)\hat{\varphi}(0,x) = \delta_1(x), \quad \varphi(N,x)\hat{\varphi}(N,x) = \delta_n(x) \quad (19)$$

for all $x \in \mathcal{X}$. In view of (15), if we define

$$\varphi_v(t) := \lambda_A^{-t} \operatorname{diag}(v)\varphi(t), \quad \hat{\varphi}_v(t) := \lambda_A^t \operatorname{diag}(v)^{-1}\hat{\varphi}(t),$$

then we have

$$\varphi_v(t) = A\varphi_v(t+1), \quad \hat{\varphi}_v(t+1) = A^T \hat{\varphi}_v(t), \quad t = 0, \dots, N-1.$$

Moreover,

$$\varphi_v(t,x)\hat{\varphi}_v(t,x) = \varphi(t,x)\hat{\varphi}(t,x), \quad t = 0, \dots, N-1, \quad x \in \mathcal{X}.$$

Here, again, A is the adjacency matrix of $\mathcal{G}$ and v is the right eigenvector corresponding to the spectral radius λ_A.

The above analysis provides another interesting way to express $\mathfrak{M}^*[\delta_1, \delta_n]$; it also solves the Schrödinger bridge problem with the same marginals δ_1 and δ_n while different prior transition matrix A, the adjacency matrix. Thus, we can replace the two-step procedure by a single bridge problem. This is summarized in the following proposition.

Proposition V.2. *Let A be the adjacency matrix of a strongly connected aperiodic graph $\mathcal{G}$, $\mathfrak{M}$ the nonnegative measure on $\mathcal{X}^{N+1}$ given by (3) with $M = A$ and μ_0 satisfying (2), then, the solution $\mathfrak{M}^*[\delta_1, \delta_n]$ of Problem V.1 also solves the Schrödinger bridge problem*

$$\min\{\mathbb{D}(P\|\mathfrak{M})|P \in \mathcal{P}(\delta_1, \delta_n)\}. \quad (20)$$

The iterative algorithm of [9, Section III] can now be based on (20) to efficiently compute the transition matrix of the optimal robust transport plan $\mathfrak{M}^*[\delta_1, \delta_n]$.

Remark V.3. *Finally, observing that if A^N has also zero entries, the robust transport described in this section may still be feasible provided there is at least one path of length N joining node 1 with node n, i.e., $(A^N)_{1n} > 0$.*

As we discussed in the beginning of this section, the intuition to use $\mathfrak{M}_{RB}$ as a prior is to achieve the spreading of the probability on all the feasible paths connecting the source and the sink. It turns out this is indeed the case; the solution $\mathfrak{M}^*[\delta_1, \delta_n]$ of Problem V.1 assigns equal probability to all the feasible paths of lengths N joining the source 1 with the sink n. To see this, by (18), the probability of the optimal transport plan $\mathfrak{M}^*[\delta_1, \delta_n]$ assigns on path $x = (x_0, x_1, \dots, x_N)$ is

$$\mathfrak{M}^*[\delta_1, \delta_n](x) = \delta_1(x_0)\prod_{t=0}^{N-1} r_{x_t x_{t+1}}\frac{\varphi(t+1, x_{t+1})}{\varphi(t, x_t)}$$

$$= \delta_1(x_0)\frac{\varphi_v(N, x_N)}{\varphi_v(0, x_0)}\prod_{t=0}^{N-1} a_{x_t x_{t+1}}.$$

Observing that $\prod_{t=0}^{N-1} a_{x_t x_{t+1}} = 1$ for feasible path and 0 otherwise, and $\delta_1(x_0)\varphi_v(N, x_N)/\varphi_v(0, x_0)$ depends only on the boundary points x_0, x_N, we conclude that $\mathfrak{M}^*[\delta_1, \delta_n]$ assigns equal probability to all the feasible paths. Moreover, there are $(A^N)_{1n}$ feasible paths of length N connecting nodes 1 and n. Thus we establish the following.

Proposition V.4. $\mathfrak{M}^*[\delta_1, \delta_n]$ *assigns probability* $1/(A^N)_{1n}$ *to each of all the feasible paths of length* N *connecting* 1 *and* n.

VI. GENERALIZATION: NOT STRONGLY CONNECTED AND WEIGHTED GRAPHS

Consider again a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with n vertices. We associate to the edge ij an "energy" $U_{ij} \geq 0$. We study the following two specific cases (and their combination):

a) Graphs that are not strongly connected: We consider the same problem as in the previous section but the graph is not strongly connected. Following [7], we can give a large positive energy U_0 to non existing links (this kind of "teleportation" is employed in the random walk of the Google Page rank algorithm to avoid getting stuck in absorbing states) and energy $U_{ij} = 0$ to existing links. Then the adjacency matrix A is replaced by the matrix

$$B = [b_{ij}] = [\exp(-U_{ij})].$$

The matrix B has all positive entries. Hence, we can apply the Perron-Forbenius theorem. Let u and v be left and right eigenvectors with positive entries of the matrix B corresponding to the spectral radius λ_B of B, so that

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

Suppose that u and v are chosen so that $\langle u, v \rangle = \sum_i u_i v_i = 1$. Then μ_U given by

$$\mu_U(i) = u_i \cdot v_i \tag{21}$$

is a probability distribution which is invariant for the transition matrix

$$R_U = \lambda_B^{-1} \operatorname{diag}(v)^{-1} B \operatorname{diag}(v), \tag{22}$$

namely

$$R_U^T \mu_U = \mu_U. \tag{23}$$

The corresponding path space measure $\mathfrak{M}_U$ is no longer uniform on paths of equal length. Indeed, the probability of the path ($i = x_0, x_1, \ldots, x_{t-1}, j = x_t$) is

$$\lambda_B^{-t} \exp(-\sum_{\ell=0}^{t-1} U_{x_\ell x_{\ell+1}}) u_i v_j.$$

However, it is the minimum *free energy rate* (topological pressure in thermodynamics) distribution attaining the maximum value of $-F = -\bar{U} + S$ given by $\log \lambda_B$ and has therefore the form of a *Boltzmann distribution*, see [7, Section IV] for details. Notice that, as soon as there are virtual links, $B \neq A$. By statement v) in Theorem III.1, we then have $\log \lambda_A < \log \lambda_B$. Namely, the topological entropy has increased in accordance to our intuition. The expected total path energy of a path of length t is precisely $t \cdot \bar{U}$.

Again, as in Proposition IV.2, we have a special case of Proposition III.2. Namely, the measure $\mathfrak{M}_U$ is the solution of a SBP where the prior $\mathfrak{M}$ is a Markovian measure on $\mathcal{X}^{N+1}$ as in (3) but with transition mechanism given by $M = B$. If U_0 is very large, most of the transportation will occur on the real edges. We can then take $\mathfrak{M}_U$ as the prior distribution in a maximum entropy problem as in Section V obtaining again through the solution $\mathfrak{M}_U^*[\delta_1, \delta_n]$ a robust transportation plan from node 1 to node n.

b) Weighted graphs: The quantities U_{ij} may represent the cost of transporting a unit of mass on that edge or may be inversely

proportional to capacity of the link, etc. The measure $\mathfrak{M}_U$ in this case may be far from uniform since it takes into account costs/capacities of the links. Again we can set up a maximum entropy problem with $\mathfrak{M}_U$ as prior obtaining a transport $\mathfrak{M}_U^*$ which compromises between the need to be robust and the cost/capacities of the different paths joining the source and the sink. For instance, if $U_{ij} = c_{ij}$, the cost of transporting a unit of goods on the link ij, is large, the solution to the maximum entropy problem with send less mass through this link provided the topology of the graph allows for alternative routes. In this case, low cost and robustness of the transportation plan may be effectively conjugated. Indeed, we have the following striking result which generalizes Propositions IV.1 and V.4.

Theorem VI.1. $\mathfrak{M}^*[\delta_1, \delta_n](x)$ *assigns equal probability to paths* $x \in \mathcal{X}^{N+1}$ *of equal cost. In particular, it assigns maximum and equal probability to minimum cost paths.*

Proof. For a path $x = (x_0, x_1, \ldots, x_N)$, we have

$$\mathfrak{M}^*[\delta_1, \delta_n](x) = \delta_1(x_0) \frac{\varphi_v(N, x_N)}{\varphi_v(0, x_0)} \prod_{t=0}^{N-1} b_{x_t x_{t+1}}$$

$$= \delta_1(x_0) \frac{\varphi_v(N, x_N)}{\varphi_v(0, x_0)} \exp[-\sum_{t=0}^{N-1} U_{x_t x_{t+1}}]. \tag{24}$$

Observe once more that $\delta_1(x_0) \frac{\varphi_v(N, x_N)}{\varphi_v(0, x_0)}$ does not depend on the particular path joining x_0 and x_N. Since $\sum_{t=1}^{N-1} U_{x_t x_{t+1}}$ is the total cost of the path, the conclusion now follows. $\square$

In the discrete optimal mass transport (OMT) problem, one usually (e.g., see [18]) seeks to first identify the least costly path(s) $(x_0, x_1^*, \ldots, x_{N-1}^*, x_N)$ from any starting node $x_0 \in \mathcal{X}$ to any ending node x_N, along with the corresponding end-point cost for a unit mass[4],

$$C_{x_0 x_N} = \min_{x_1^*, \ldots, x_{N-1}^*} \left(U_{x_0 x_1^*} + \ldots + U_{x_{N-1}^* x_N} \right).$$

This is a *combinatorial problem* but can also be cast as a linear program [19]. Having a solution to this first problem, the OMT problem can then be recast as the linear program

$$\min_q \left\{ \sum_{x_0, x_N} q_{x_0, x_N} C_{x_0 x_N} \mid q_{x_0, x_N} \geq 0, \tag{25} \right.$$

$$\left. \sum_{x_0} q_{x_0, x_N} = \nu_N(x_N), \sum_{x_N} q_{x_0, x_N} = \nu_0(x_0) \right\}.$$

The solution to (25) is the transport plan q_{x_0, x_N} which dictates the portion of mass that is to be sent from x_0 to x_N along the corresponding least costly path $(x_0, x_1^*, \ldots, x_{N-1}^*, x_N)$. Alternatively, the OMT problem can be directly cast as a linear program in as many variables as there are edges [19].

An apparent shortcoming of the OMT formalism is the "rigidity" of the transportation to utilize only paths with minimal cost from starting to ending node. The transport provided by Theorem VI.1, which readily generalizes to any two marginals ν_0 and ν_N, provides an attractive alternative to the OMT approach: Minimum cost paths all have maximum probability, but some of the mass is also transported on alternative paths thereby ensuring a certain amount of robustness of the transportation plan. Also notice that the Schrödinger bridge

---

[4]We assume a self loop for each node with zero cost, i.e., $U_{xx} = 0$ for each $x \in \mathcal{X}$.

measure $\mathfrak{M}_U^*[\delta_1, \delta_n]$ determines, as a by-product, the minimum cost paths!

The argument provided at the end of the previous section (see Proposition V.2) shows once more that $\mathfrak{M}_U^*[\delta_1, \delta_n]$ can be obtained in both of the above cases in one step as solution to the Schrödinger bridge problem with the same marginals δ_1 and δ_n and prior transition matrix B.
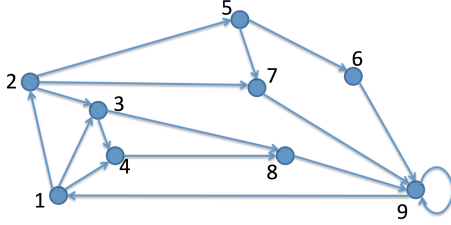

Fig. 1: Network topology

VII. EXAMPLES

We present a simple academic example to illustrate our method. Consider the graph in Figure 1. We seek to transport a unit mass from node 1 to node 9 in $N = 3$ and 4 steps. We add a self loop at node 9, i.e., $a_{99} = 1$, to allow for transport paths with different step sizes.

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 minimum number of steps, we may end up using one of these three paths. This is not so robust. On the other hand, if we apply the Schrödinger bridge framework with the RB measure $\mathfrak{M}_{\mathrm{RB}}$ as the prior, then we get a transport plan with equal probabilities using all these three paths. 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 shown below

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4/7 & 2/7 & 1/7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/7 & 1/7 & 2/7 & 0 & 1/7 & 2/7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/7 & 1/7 & 2/7 & 3/7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now we change the graph by adding a cost on the edge $(7, 9)$ so that the weighted adjacency matrix B has $b_{79} = 0.5$ and coincides with the adjacency matrix in the other entries. 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 is given by

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/5 & 2/5 & 2/5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/5 & 4/5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The mass travels through paths $1-2-7-9$, $1-3-8-9$ and $1-4-8-9$, but unlike the unweighted case, the transport plan doesn't take equal probability for these three paths Since we added a cost on the edge $(7, 9)$, the probability that the mass takes this path becomes smaller. The plan does, however, assign equal probability to the two minimum cost paths $1-3-8-9$ and $1-4-8-9$ in

agreement with Theorem VI.1. Suppose now we allow for more steps $N = 4$ and change the B matrix as follows: for all existing links, the "energy" is 0.7 excepting that $b_{79} = 0.5$ and $b_{99} = 0.9$. Here, transporting on any edge is expensive. It is, however, more expensive to transverse link $(7, 9)$ and less expensive to let the mass sit at the sink node 9. The evolution of the mass distribution is now

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5042 & 0.3173 & 0.1785 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1388 & 0.1388 & 0.2380 & 0 & 0.1275 & 0.3569 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1388 & 0.0992 & 0.2776 & 0.4844 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}.$$

We observe that almost one half of the mass (0.4844) reaches node 9 in three steps, and then sits there, travelling on the three shortest paths $1-2-7-9$, $1-3-8-9$ and $1-4-8-9$. As before, more mass (0.1785) travels on the two minimum cost paths $1-3-8-9$ and $1-4-8-9$ in agreement with Theorem VI.1, whereas 0.1275 travels on the more expensive, minimum length path $1-2-7-9$. There are now several other ways the mass can reach node 9 in 4 steps. Our robust transportation plan takes full advantage of them, transporting more that one half of the total mass along these alternative paths.

Finally, we consider the case where the graph is not strongly connected deleting links $(1, 4)$, $(2, 7)$ and $(9, 1)$ in Figure 1. Again we want to transport a unit mass from node 1 to node 9. In order to do this, we add an artificial energy U_0 to each non existing link as discussed in Section VI. We display the results for $N = 4$ steps. When we take $U_0 = 2$, the evolution of mass is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0415 & 0.4079 & 0.3416 & 0.0326 & 0.0462 & 0.0326 & 0.0326 & 0.0326 & 0.0326 \\ 0.0270 & 0.0349 & 0.1740 & 0.1477 & 0.2330 & 0.0603 & 0.0603 & 0.1614 & 0.1014 \\ 0.0116 & 0.0152 & 0.0199 & 0.0242 & 0.0163 & 0.1709 & 0.1709 & 0.2641 & 0.3069 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We can see that there is quite a portion of mass traveling along virtual (non existing) edges. If we increase the value to $U_0 = 8$, then the mass evolution becomes

$$\begin{bmatrix} 1.0000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0001 & 0.5995 & 0.4000 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0001 \\ 0.0000 & 0.0000 & 0.2000 & 0.1999 & 0.3994 & 0.0002 & 0.0002 & 0.1999 & 0.0004 \\ 0.0000 & 0.0000 & 0.0000 & 0.0001 & 0.0000 & 0.1999 & 0.1999 & 0.3995 & 0.2007 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}.$$

The portion of mass traveling along non existing edges is negligible. Eventually, all the mass would be transported along feasible paths and in the limit the mass evolution (flow) is given by the rows of

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3/5 & 2/5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/5 & 1/5 & 2/5 & 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/5 & 1/5 & 2/5 & 1/5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

VIII. CONCLUSIONS

In this paper, we have proposed a novel approach to design a robust transportation plan on a given directed graph. It is based on a sort of generalized maximum entropy problem (Schrödinger bridge) for measures on paths of the given network. Taking as prior measure the Ruelle-Bowen-Parry random walker, the solution naturally tends to spread the mass on all available routes joining the source and the sink. Hence, the resulting transport appears robust with respect to links/nodes failure. This approach can be adapted to graphs that are not strongly connected, as well as to weighted graphs. In the latter case, it can be used to effectively compromise between robustness and cost. Indeed, we exhibit a *robust* transportation plan which assigns *maximum probability* to *minimum cost paths* and therefore appears attractive when compared with Optimal Mass Transportation approaches. Since the transport plan is computed as a Schrödinger bridge, for which an efficient iterative algorithm is available, our procedure also appears to be computationally attractive.

In this paper, in order to avoid obscuring the fundamental ideas and to keep the paper at a reasonable length, we have chosen to present the essential features of our approach without touching on a number of related fascinating topics. For instance, in this paper *robustness* of a transport plan simply means that, in case of failure of certain links (e.g. due to congestion) or nodes, most of the mass will anyway reach the target nodes. There are, however, other notions of robustness in graph theory [1], [20], [21], [6], [22], some related to entropic principles [23], [24].

When weights represent costs, our approach of Section VI compromising between minimization and robustness can be further compared to *Optimal Mass Transport* (OMT) over graphs [25], where only cost matters, and entropically regularized OMT-schemes [26], [27]. In discrete OMT, however, the cost function is supposed to be given, although computing it is typically an intractable problem for large networks.

Also, it is apparent that choosing the uniform as prior distribution in the maximum entropy problem such as in Section V we obtain a spreading of trajectories over which the transport occurs similar to the one in Optimal Mass Transport (OMT) on manifolds with positive Ricci-Curbastro *curvature* [28]. On discrete spaces and graphs, similar notions of curvature have been defined by Ollivier [29], [30]. They capture robustness and connectedness, convexity of entropy, and are related to the spectral gap [31], [32]. Their relevance in applications is discussed in, e.g., [33], [21], [6], [22]. It is therefore natural to investigate the precise connection between the role of the prior in random evolutions such as those studied in this paper and deterministic evolution on discrete curved spaces. These fascinating topics deserve investigation and will be addressed elsewhere.

REFERENCES

[1] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: Percolation on random graphs," *Physical Review Letters*, vol. 85, no. 25, p. 5468, 2000.

[2] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[3] D. M. Scott, D. C. Novak, L. Aultman-Hall, and F. Guo, "Network robustness index: a new method for identifying critical links and evaluating the performance of transportation networks," *Journal of Transport Geography*, vol. 14, no. 3, pp. 215–227, 2006.

[4] G. Cabanes, E. van Wilgenburg, M. Beekman, and T. Latty, "Ants build transportation networks that optimize cost and efficiency at the expense of robustness," *Behavioral Ecology*, vol. 26, no. 1, pp. 223–231, 2014.

[5] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks*, vol. 56, pp. 3825–3833, 2012.

[6] R. Sandhu, T. Georgiou, E. Reznik, L. Zhu, I. Kolesov, Y. Senbabaoglu, and A. Tannenbaum, "Graph curvature for differentiating cancer networks," *Scientific reports*, vol. 5, 2015.

[7] J.-C. Delvenne and A.-S. Libert, "Centrality measures and thermodynamic formalism for complex networks," *Physical Review E*, vol. 83, no. 4, p. 046117, 2011.

[8] H. Kitano, "Towards a theory of biological robustness," *Molecular Systems Biology*, vol. 3, no. 1, p. 137, 2007.

[9] T. T. Georgiou and M. Pavon, "Positive contraction mappings for classical and quantum Schrödinger systems," *Journal of Mathematical Physics*, vol. 56, no. 3, p. 033301, 2015.

[10] M. Pavon and F. Ticozzi, "Discrete-time classical and quantum markovian evolutions: Maximum entropy problems on path space," *Journal of Mathematical Physics*, vol. 51, no. 4, p. 042104, 2010.

[11] W. Parry, "Intrinsic Markov chains," *Transactions of the American Mathematical Society*, vol. 112, no. 1, pp. 55–66, 1964.

[12] T. Cover and J. Thomas, *Elements of information theory*. Wiley, 2012.

[13] H. Föllmer, "Random fields and diffusion processes," in *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*. Springer, 1988, pp. 101–203.

[14] G. Birkhoff, "Extensions of Jentzsch's theorem," *Transactions of the American Mathematical Society*, vol. 85, no. 1, pp. 219–227, 1957.

[15] P. J. Bushell, "Hilbert's metric and positive contraction mappings in a Banach space," *Arch. Rat. Mech. Anal.*, vol. 52, pp. 330–338, 1973.

[16] R. Horn and C. Johnson, *Matrix analysis*. Cambridge Univ. Press, 2012.

[17] D. Ruelle, *Thermodynamic formalism: the mathematical structure of equilibrium statistical mechanics*. Cambridge University Press, 2004.

[18] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory*. Springer Science & Business Media, 1998, vol. 1.

[19] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear programming and network flows*. John Wiley & Sons, 2011.

[20] A. Jamakovic and S. Uhlig, "On the relationship between the algebraic connectivity and graph's robustness to node and link failures," in *Next Generation Internet Networks, 3rd EuroNGI Conf. on*, 2007, pp. 96–102.

[21] C. Wang, E. Jonckheere, and R. Banirazi, "Wireless network capacity versus ollivier-ricci curvature under heat-diffusion (hd) protocol," in *American Control Conference, 2014*, 2014, pp. 3536–3541.

[22] R. Sandhu, T. Georgiou, and A. Tannenbaum, "Market fragility, systemic risk, and ricci curvature," *arXiv preprint arXiv:1505.05182*, 2015.

[23] L. Arnold, V. M. Gundlach, and L. Demetrius, "Evolutionary formalism for products of positive random matrices," *The Annals of Applied Probability*, pp. 859–901, 1994.

[24] L. Demetrius and T. Manke, "Robustness and network evolution: an entropic principle," *Physica A: Statistical Mechanics and its Applications*, vol. 346, no. 3, pp. 682–696, 2005.

[25] C. Léonard, "Lazy random walks and optimal transport on graphs," *arXiv preprint arXiv:1308.0226*, 2013.

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

[27] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, "Iterative bregman projections for regularized transportation problems," *SIAM Journal on Scientific Computing*, vol. 37, no. 2, pp. A1111–A1138, 2015.

[28] C. Villani, *Optimal transport: old and new*. Springer, 2008, vol. 338.

[29] Y. Ollivier, "Ricci curvature of markov chains on metric spaces," *Journal of Functional Analysis*, vol. 256, no. 3, pp. 810–864, 2009.

[30] ——, "A survey of Ricci curvature for metric spaces and markov chains," *Probabilistic approach to geometry*, vol. 57, pp. 343–381, 2010.

[31] F. Bauer, J. Jost, and S. Liu, "Ollivier-Ricci curvature and the spectrum of the normalized graph Laplace operator," *arXiv:1105.3803*, 2011.

[32] J. Jost and S. Liu, "Ollivier's Ricci curvature, local clustering and curvature-dimension inequalities on graphs," *Discrete & Computational Geometry*, vol. 51, no. 2, pp. 300–322, 2014.

[33] R. Banirazi, E. Jonckheere, and B. Krishnamachari, "Heat diffusion algorithm for resource allocation and routing in multihop wireless networks," in *Global Communications Conference (GLOBECOM), 2012 IEEE*. IEEE, 2012, pp. 5693–5698.