Arc-based Traffic Assignment: Equilibrium Characterization and Learning

Chih-Yuan Chiu*[1], Chinmay Maheshwari*[1], Pan-Yang Su[1], and Shankar Sastry[1]

***Abstract*— Arc-based traffic assignment models (TAMs) are a popular framework for modeling traffic network congestion generated by self-interested travelers who sequentially select arcs based on their perceived latency on the network. However, existing arc-based TAMs either assign travelers to cyclic paths, or do not extend to networks with bidirectional arcs (or edges) between nodes. To overcome these difficulties, we propose a new modeling framework for stochastic arc-based TAMs. Given a traffic network with bidirectional arcs, we replicate its arcs and nodes to construct a directed acyclic graph (DAG), which we call the *Condensed DAG* (CoDAG) representation. Self-interested travelers sequentially select arcs on the CoDAG representation to reach their destination. We show that the associated equilibrium flow, which we call the *Condensed DAG equilibrium*, exists, is unique, and can be characterized as a strictly convex optimization problem. Moreover, we propose a discrete-time dynamical system that captures a natural adaptation rule employed by self-interested travelers to learn about the emergent congestion on the network. We show that the arc flows generated by this adaptation rule converges to a neighborhood of Condensed DAG equilibrium. To our knowledge, our work is the first to study learning and adaptation in an arc-based TAM. Finally, we present numerical results that corroborate our theoretical results.**

I. INTRODUCTION

Traffic assignment models (TAMs) [1–7] play a central role in congestion modeling for transportation networks, by informing crucial decisions about infrastructure investment, capacity management, and tolling for congestion regulation. The central dogma behind this modeling approach is that self-interested travelers select routes with minimal *perceived* latency (i.e., the Wardrop or user equilibrium), which can be modeled as deterministic [1, 2] or stochastic [3–7]. Empirical studies confirm that stochastic TAMs achieve greater success at interpreting congestion levels, compared to their deterministic counterparts [8].

There exist two dominant modeling paradigms in TAM: the route-based model [1, 5, 7, 9]—where each traveler makes a single choice between set of available routes from origin to destination—and the arc (or edge) based model [3, 10–13]—where the traveler sequentially makes routing decision at each node on the network, based on their perception of arc latencies. There are two major drawbacks of route-based models on real-world networks: route correlation and route enumeration. Specifically, the utility generated from different routes is correlated due to overlapping arcs on different routes. Moreover, exhaustive route enumeration is prohibitive in terms of computational cost, memory storage, and information acquisition, since the number of routes in a traffic network can be exponential in the number of arcs.

To avoid explicit route enumeration, Akamatsu [6] proposed the first arc-based stochastic TAM, which was further generalized by Baillon and Cominetti [3]. More recently, Fosgerau et al. and Mai et al. [4, 12] presented similar arc-based models based on dynamic discrete choice analysis, which are mathematically similar to the models proposed by Akamatsu [6] and Baillon and Cominetti [3]. However, these models suggest that travelers take cyclic routes with positive probability. To overcome this fundamental modeling challenge, Oyama et al. [14, 15] recently proposed various methods to explicitly avoid routing on cyclic routes. Unfortunately, these methods either do not apply beyond acyclic graphs [15] or lose accuracy by restricting the set of feasible routes [14]. Sequential arc selection models in network routing have also been studied by Calderone et al. [16, 17] where each arc selection is accompanied by stochastic transitions to the next arc, and a deterministic transition cost. This stands in contrast to the stochastic TAM literature, where transitions from arc to arc are assumed deterministic and the travel cost (latency) is assumed stochastic.

In this work, we propose an arc-based stochastic TAM that explicitly avoids cycles by considering routing on a directed acyclic graph derived from the original network, henceforth referred to as the *Condensed Directed Acyclic Graph* (CoDAG). The CoDAG representation duplicates an appropriate subset of nodes and arcs in the original network, to explicitly avoids cycles while preserving all feasible routes. Travelers sequentially select arcs on the CoDAG network at every intermediate node, based on perceived arc latencies. This route choice behavior is akin to the models prescribed by Akamatsu [6] and Baillon and Cominetti [3], but with routing occurring over the CoDAG associated with original network. We show that the corresponding equilibrium congestion pattern—which we term the *Condensed DAG equilibrium* (CoDAG equilibrium)—can be characterized as the unique minimizer of a strictly convex optimization problem.

Moreover, we propose a discrete-time dynamical system that captures a natural adaptation rule used by self-interested travelers who progressively learn towards equilibrium arc selections. In game theory literature, an equilibrium notion is only considered useful if there exists an adaptive learning scheme that allows self interested players to converge to it [18]. Despite significant research progress on both theoretical

*Equal contribution.

[1]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 (emails: {chihyuan_chiu, chinmay_maheshwari, pan_yang_su, sastry} at berkeley dot edu).

and algorithmic aspects of stochastic arc-based TAMs, to the best of our knowledge, there has been no research on adaptive learning schemes that ensure convergence to such equilibria. In recent years, adaptive learning schemes that converge to the equilibria in route-based TAMs have been extensively studied [19–24]. This line of research considers a population of self-interested travelers who repeatedly select routes by observing route latencies in past rounds of interaction. In this work, we extend this line of research to arc-based TAMs. In particular, we propose a discrete-time dynamics, in every round of which travelers update their arc selections at every node on the CoDAG network based on the past rounds of interaction. We prove that the emergent aggregate arc selection probabilities at every node (and the resulting congestion levels on each arc) globally asymptotically converge to a neighborhood of the CoDAG equilibrium.

To establish convergence, we appeal to the theory of stochastic approximation [25], which requires two conditions: (i) The vector field of the discrete-time dynamical system is Lipschitz, and (ii) The trajectories of an associated continuous-time dynamical system asymptotically converge to the CoDAG equilibrium. To prove the former, we establish recursive Lipschitz bounds for vector fields associated with every node. For the latter, we first construct a Lyapunov function using a strictly convex optimization objective associated with the CoDAG representation. We then show that the value of this Lyapunov function decreases along the trajectories of the continuous-time dynamical system. Our contributions are listed below:

1) We introduce a new arc-based traffic equilibrium concept—the *Condensed DAG equilibrium*—which overcomes some limitations of existing traffic equilibrium notions. Furthermore, we show that the Condensed DAG equilibrium is characterized by a solution to a strictly convex optimization problem.
2) We present, to the best of our knowledge, the first adaptive learning scheme in the context of stochastic arc-based TAM. Furthermore, we establish formal convergence guarantees for this learning scheme.
3) We validate our theoretical results on a simulated traffic network.

The rest of the paper is organized as follows. Section II introduces the setup considered in this paper, and defines the Condensed DAG representation. Section III defines the Condensed DAG equilibrium, and characterize it as a solution to a strictly convex optimization problem. Section IV presents discrete-time dynamics that converges to the Condensed DAG equilibrium and also provide a sketch of the proof. In Section V, we numerically study the convergence of the discrete-time dynamics on a simulated traffic network. Finally, Section VI presents concluding remarks and future work directions.

Notation: For each positive integer $n \in \mathbb{N}$, we denote $[n] := \{1, \cdots, n\}$. For each $i \in [n]$ in an Euclidean space $\mathbb{R}^n$, we denote by e_i the i-th standard unit vector.

II. CONDENSED DAG REPRESENTATION

A. Setup

Consider a traffic network represented by a directed graph $G_O = (I_O, A_O)$, possibly with bidirectional arcs, where I_O and A_O denote nodes and arcs, respectively. An example is depicted in Figure 1 (top left). Let the *origin nodes* and *destination nodes* be two disjoint subsets of nodes in G_O. Each traveler enters the network through an origin node to travel to a destination node, by sequentially selecting arcs at every intermediate node. This gives rise to congestion on each arc, which in turn decides the travel times. Specifically, each arc $\tilde{a} \in A_O$ is associated with a strictly increasing *latency function* $s_{\tilde{a}} : [0, \infty) \to [0, \infty)$, which gives for each arc the travel time as a function of traffic flow. To simplify our exposition, we assume that there is only one origin-destination tuple (o, d), although the results presented in this paper naturally extend to settings where the traffic network has multiple origin-destination pairs. We denote by g_o the demand of (infinitesimal) travelers who travel from the origin o to the destination d.

Remark 1: Arc selections made by travelers at different nodes are independent of one another and therefore if the underlying network has bidirectional edges, then sequential arc selection by a traveler can result in a cyclic route. For example, sequential arc selection in the original network shown on the top left in Figure 1, may lead a traveler to loop between i_2^O and i_3^O before reaching destination. To overcome this, we introduce a directed acyclic graph (DAG) representation of original graph G_O in the following subsections, called the *condensed DAG*. Sequential arc selections made on this network encodes the travel history by design and therefore avoids cyclic routes.

B. Preliminaries on DAG: Depth and Height

Before introducing condensed DAG representation, we first present the notions of *height* and *depth* of a DAG. These concepts are crucial for the construction and analysis of condensed DAGs in the following sections. For the exposition in this subsection, let G be a DAG with a single origin-destination pair (o, d). Furthermore, let $\mathbf{R}$ be the set of all acyclic routes in G which start at the origin node o and end at the destination node d.

*Definition 1 (**Depth**):* For each $r \in \mathbf{R}$ and $a \in r$, let $\ell_{a,r}$ denote the location of arc a in route r, i.e., a is the $\ell_{a,r}$-th arc in the route r, and with a slight abuse of notation, define: $\ell_a := \max_{r \in \mathbf{R}: a \in r} \ell_{a,r}$, We say that a is an ℓ_a*-th depth arc* in the Condensed DAG G. Moreover, we define the *depth of a node* $i \in I \backslash \{o\}$ by:

$$\bar{\ell}_i := \max_{a \in A_i^-} \ell_a$$

with $\bar{\ell}_o = 0$.

*Definition 2 (**Height**):* For each $r \in \mathbf{R}$ and $a \in r$, let $m_{a,r}$ denote the location of arc a in route r, i.e., a is the $(|r| - m_{a,r})$-th arc in route r, and with a slight abuse of notation, define: $m_a := \max_{r \in \mathbf{R}: a \in r} m_{a,r}$. We say that a is an m_a*-th height arc* in the Condensed DAG G . Moreover, we define the *height of a node* $i \in I \backslash \{d\}$ by:

$$\bar{m}_i := \max_{a \in A_i^+} m_a$$

with $\bar{m}_d = 0$.

C. Construction of Condensed DAG

For ease of description, we illustrate the construction through an example in Figure 1. We also present a pseudo-code to generate the condensed DAG representation.

A straightforward way to associate G_O with a DAG would be to brute-force enumerate all acyclic (simple) routes and construct a tree network by replicating arcs and nodes by the number of routes passing through them (see Figure 1, bottom). However, the resulting tree network may contain a significantly larger number of arcs and nodes compared with the original network. To ameliorate this, we present the *condensed DAG* representation (Figure 1, top right). The condensed DAG is formed by merging superfluous nodes and arcs in the tree network, while ensuring that the graph remains acyclic, and preserving the set of acyclic routes from the original network.

TABLE I: Arc correspondences between the graphs in Figure 1: The original network (top left), fully expanded tree (bottom), and the CoDAG (top right).

Original	Tree DAG	CoDAG
a_1^O	$a_1^T, a_2^T, a_3^T, a_4^T, a_5^T$	a_1^C
a_2^O	$a_6^T, a_7^T, a_8^T, a_9^T, a_{10}^T$	a_2^C
a_3^O	a_{12}^T, a_{13}^T	a_4^C
a_4^O	$a_{18}^T, a_{19}^T, a_{20}^T$	a_7^C
a_5^O	$a_{14}^T, a_{15}^T, a_{23}^T, a_{24}^T$	a_5^C, a_9^C
a_6^O	$a_{16}^T, a_{17}^T, a_{21}^T, a_{22}^T$	a_6^C, a_8^C
a_7^O	a_{11}^T, a_{25}^T	a_3^C, a_{10}^C
a_8^O	$a_{26}^T, a_{28}^T, a_{30}^T, a_{32}^T$	a_{11}^C
a_9^O	$a_{27}^T, a_{29}^T, a_{31}^T, a_{33}^T$	a_{12}^C

One can design a condensed DAG representation as follows:

(S1) Convert the original network G_O to a tree structure $G_T = (I_T, A_T)$, in which every branch emanating from the origin represents a route. Each node and arc is replicated by the number of acyclic routes that contains it. For every node i in G_T, compute the depth $\bar{\ell}_i$ and height $\bar{m}_i$ (see Definition 1-2).

(S2) Generate a partition P_T of I_T such that:

 (i) For each $X \in P_T$, all nodes in X are replicas of the same node in I_O that share the same height or depth in G_T.

 (ii) For any $X, Y \in P_T$, there exists no $i, i' \in X, j, j' \in Y$, such that $\bar{m}_j > \bar{m}_i$ and $\bar{m}_{j'} < \bar{m}_{i'}$

(S3) For each set element X of P_T, merge all nodes in X into a single node. Then, merge arcs which have same start and end nodes, and are replicas of the same edge in the original network G_O.

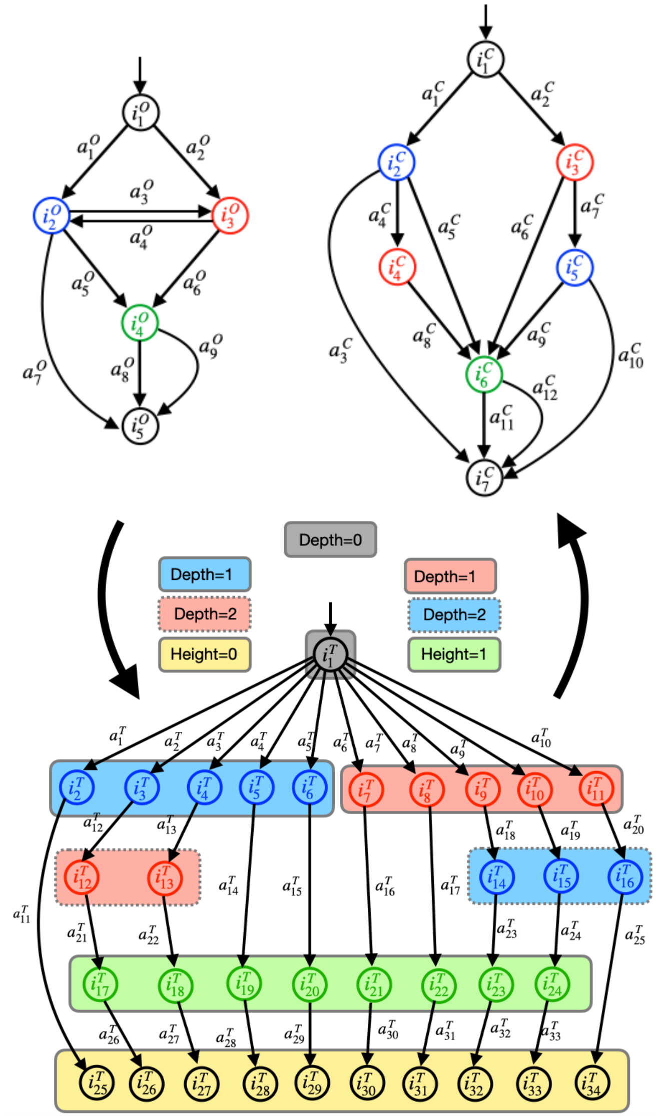

Fig. 1: Example of a single-origin single-destination original network G_O (top left, with superscript O), and its corresponding tree network (bottom, with superscript T) and condensed DAG G (top right, with superscript C). The blocks in G_T represent a partition P_T (see (S2)). The depth and height of nodes in every partition is denoted above G_T. Arc correspondences between the three networks are given by Table I, while node correspondences are indicated by color.

We refer to any graph generated via (S1)-(S3) as a *condensed DAG* (CoDAG) representation $G = (I, A)$ of the original network, where I and A are the set of nodes and arcs, respectively. By construction, the CoDAG representation explicitly avoids cyclic routes, and preserves all the acyclic routes from the original network. This is because the tree network constructed in (S1) preserves all acyclic routes from original network. Furthermore, the merging conditions stated in (S3) prohibit both the removal and the addition of routes.

Remark 2: A given traffic network with bidirectional arcs may yield several distinct CoDAG representations, any of which would be amenable to our analysis in subsequent sections. The development of an algorithmic procedure to compute a CoDAG with the least number of arcs or nodes is beyond the scope of this work.

Remark 3: The Condensed DAG representation G can be significantly smaller in size, compared to the tree network. There exists original networks whose corresponding tree representation G_T is exponentially larger than its corresponding CoDAG G. For example, consider network with nodes $i_1, \cdots, i_n$, with two directed arcs connecting i_k to i_{k+1}, for each $k \in [n-1]$. Here, the corresponding tree network would have 2^n arcs, while the CoDAG representation only has $2n$.

Remark 4: The arc-based TAM literature also considers modified representations of traffic networks with bidirectional arcs. For example, Oyama, Hara et al. [15, 26] considers the Network Generalized Extreme Value (NGEV) representations, which is similar to our CoDAG representation, but applies only to acyclic networks [15]. Thus, NGEV models cannot capture realistic traffic networks where almost all arcs are bidirectional. Meanwhile, Oyama, Hato et al. [14] considers the Choice Based Prism (CBP) representation, which prunes the available set of feasible routes to ameliorate computational inefficiency. While CBP explicitly avoids cyclic routes, it also removes some acyclic routes during the pruning process. In contrast, the CoDAG representation avoids this issue.

To conclude this section, we introduce some notation used throughout the rest of the paper. Recall that CoDAGs are formed by replicating the arcs in G_O. To describe this correspondence between arcs, we define $[\cdot] : A \to A_O$ to be a map from each CoDAG arc $a \in A$ to the corresponding arc $[a] \in A_O$. For each arc $a \in A$, let i_a and j_a denote the start and terminal nodes, and for each node $i \in I$, let $A_i^-, A_i^+ \subset A$ denote the set of incoming and outgoing arcs.

III. EQUILIBRIUM CHARACTERIZATION

In this section, we introduce the *condensed DAG (CoDAG) equilibrium* (Definition 3), which is based on the CoDAG representation of the original traffic network. Specifically, we show that the CoDAG equilibrium exists, is unique, and solves a strictly convex optimization problem (Theorem 1).

A. Condensed DAG Equilibrium

Below, we assume that every traveler knows G_O and has access to the same CoDAG representation of G_O. To avoid cyclic routes, we model travelers as performing sequential arc selection over the CoDAG representation $G = (I, A)$. The aggregate effect of the travelers' arc selections gives rise to the congestion on the network. Concretely, for each $a \in A$, let the *flow* or *congestion level* on arc a be denoted by w_a, and let the total flow on the corresponding arc in the original network be denoted, with a slight abuse of notation, by $w_{[a]} := \sum_{a' \in [a]} w_{a'}$. (Note that unlike existing TAMs, the latency of arcs in G can be coupled through the map $w_{[\cdot]}$, since multiple copies of the same arc in G_O may exist in G.) Then, the perceived latency of travelers on each arc $a \in A$ is described by:

$$\tilde{s}_{[a]}(w_{[a]}) := s_{[a]}(w_{[a]}) + \nu_a,$$

where ν_a is a zero-mean random variable. At each non-destination node $i \in I \backslash \{d\}$, travelers select among outgoing nodes $a \in A_i^+$ by comparing their perceived latencies-to-go $\tilde{z}_a : \mathbb{R}^{|A|} \to \mathbb{R}$, given recursively by:

$$\begin{aligned} \tilde{z}_a(w) &:= \tilde{s}_{[a]}(w_{[a]}) + \min_{a' \in A_{j_a}^+} \tilde{z}_{a'}(w), && j_a \neq d, \\ \tilde{z}_a(w) &:= \tilde{s}_{[a]}(w_{[a]}), && j_a = d, \end{aligned} \tag{1}$$

Consequently, the fraction of travelers who arrive at $i \in I \backslash \{d\}$ and choose arc $a \in A_i^+$ is given by:

$$P_{ij_a} := \mathbb{P}(\tilde{z}_a \leq \tilde{z}_{a'}, \forall a' \in A_i^+). \tag{2}$$

An explicit formula for the probabilities $\{P_{ij_a} : a \in A_i^+\}$ in terms of the statistics of $\tilde{z}_a$, is provided by the discrete-choice theory [27]. In particular, define $z_a(w) := \mathbb{E}[\tilde{z}_a(w)]$ and $\epsilon_a := \tilde{z}_a(w) - z_a(w)$, and define the latency-to-go at each node by:

$$\varphi_i(\{z_{a'}(w) : a' \in A_i^+\}) = \mathbb{E}\left[\min_{a' \in A_i^+} \tilde{z}_{a'}(w)\right], \tag{3}$$

Then, from discrete-choice theory [27]:

$$P_{ij_a} = \frac{\partial \varphi_i(z)}{\partial z_a}, \quad i \in I \backslash \{d\}, a \in A_i^+, \tag{4}$$

where, with a slight abuse of notation, we write $\varphi_i(z)$ for $\varphi_i(\{z_{a'} : a' \in A_i^+\})$. To obtain a closed form expression of φ, this work considers the *logit Markovian model* [3, 6], which assumes that the zero-mean noise ϵ is Gumbel-distributed with scale $\beta > 0$. Intuitively, $\beta > 0$ is an entropy parameter that models the degree to which the average traveler's perception of network latency is suboptimal. In this case, the corresponding latency-to-go at each node i in G is:

$$\varphi_i(z) = -\frac{1}{\beta} \ln\left(\sum_{a' \in A_i^+} e^{-\beta z_{a'}}\right). \tag{5}$$

Using (1) and (5), the expected minimum latency-to-go $z_a : \mathbb{R}^{|A|} \to \mathbb{R}$, associated with traveling on each arc $a \in A$, is given by:

$$z_a(w) = s_{[a]}\left(\sum_{\bar{a} \in [a]} w_{\bar{a}}\right) - \frac{1}{\beta} \ln\left(\sum_{a' \in A_{j_a}^+} e^{-\beta z_{a'}(w)}\right). \tag{6}$$

Note that (6) is well-posed, as z_a can be recursively computed along arcs of increasing height (Definition 2) from the destination back to the origin. For more details, please see Appendix B.

Against the preceding backdrop, we formally define the central equilibrium solution concept studied in this paper: the Condensed DAG Equilibrium (CoDAG Equilibrium).

*Definition 3 (**Condensed DAG Equilibrium**):* Given $\beta > 0$, a vector of arc-flow $\bar{w}^\beta \in \mathbb{R}^{|A|}$ is called a *Condensed DAG equilibrium* if, for each $i \in I \backslash \{d\}$, $a \in A_i^+$:

$$\bar{w}_a^\beta = \left(g_i + \sum_{a' \in A_i^+} \bar{w}_{a'}^\beta\right) \cdot \frac{\exp(-\beta z_a(\bar{w}^\beta))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta z_{a'}(\bar{w}^\beta))}, \tag{7}$$

where $g_i = g_o$ if $i = o$, $g_i = 0$ otherwise, and $w \in \mathcal{W}$, with:

$$\mathcal{W} := \Bigg\{ \bar{w}^\beta \in \mathbb{R}^{|A|} : \sum_{a \in A_i^+} \bar{w}_a^\beta = \sum_{a \in A_i^-} \bar{w}_a^\beta, \forall\, i \neq o, d, \quad (8)$$

$$\sum_{a \in A_o^+} \bar{w}_a^\beta = g_o,\ \bar{w}_a^\beta \geq 0, \forall\, a \in A \Bigg\}.$$

For any CoDAG equilibrium $\bar{w}^\beta$, the fraction of travelers at any node $i \in I \backslash \{d\}$ who selects an arc $a \in A_i^+$ is given by:

$$\bar{\xi}_a^\beta := \frac{\bar{w}_a^\beta}{\sum_{a' \in A_i^+} \bar{w}_{a'}^\beta}.$$

Remark 5: While the CoDAG equilibrium and Markov Traffic Equilibrium (MTE) share some similarities (see [3]), there also exist two main fundamental differences. First, by design, the CoDAG equilibrium does not yield cyclic routes with strictly positive probability (as is the case with the MTE). Second, unlike the MTE, congestion levels on arcs (which may be replicas of the same arc in G_O) in the CoDAG representation are coupled to each other. Therefore, MTE analysis does not extend to the CoDAG equilibrium in a straighforward manner.

B. Existence and Uniqueness of the CoDAG equilibrium

In this subsection, we show the existence and uniqueness of the CoDAG equilibrium. Specifically, we characterize this equilibrium as the unique minimizer of a strictly convex optimization problem over a compact set. First, for each $[a] \in A_O$, define:

$$f_{[a]}(w) := \int_0^{w_{[a]}} s_{[a]}(u)\, du, \quad (9)$$

and for each $i \in I \backslash \{d\}$, set:

$$\chi_i(w_{A_i^+}) := \sum_{a \in A_i^+} w_a \ln w_a - \left(\sum_{a \in A_i^+} w_a \right) \ln \left(\sum_{a \in A_i^+} w_a \right). \quad (10)$$

Finally, define $F : \mathcal{W} \to \mathbb{R}$ by:

$$F(w) = \sum_{[a] \in A_O} f_{[a]}(w) + \frac{1}{\beta} \sum_{i \neq d} \chi_i(w_{A_i^+}), \quad (11)$$

where $w_{A_i^+} \in \mathbb{R}^{|A_i^+|}$ denotes the components of w corresponding to arcs in A_i^+.

Theorem 1: The CoDAG equilibrium $\bar{w}^\beta \in \mathcal{W}$ exists, is unique, and is the unique minimizer of F over $\mathcal{W}$.

To prove Theorem 1, we first show that $F(\cdot)$ is strictly convex over $\mathcal{W}$ (Lemma 1). Therefore, F has a unique minimizer in $\mathcal{W}$. It then suffices to the CoDAG equilibrium definition (Definition 3) matches the Karush-Kuhn-Tucker (KKT) conditions for the optimization problem (11).

Lemma 1: The map $F : \mathcal{W} \to \mathbb{R}$ is strictly convex.

Proof: **(Proof Sketch)** It suffices to show that $f_{[a]}$ and χ_i are convex for each $[a] \in A_O$, $i \in I \backslash \{d\}$. Each $f_{[a]}$ is convex, since it is the composition of a convex function ($w \mapsto \sum_{a \in A_O} \int_0^{w_a} s_a(u) du$) with a linear function ($w_{[a]} := \sum_{a' \in [a]} w_{a'}$). Furthermore, we establish that for any $i \in I \backslash \{d\}$, $y_i \in \mathbb{R}^{|A_i^+|}$:

$$y_i^\top \nabla_w^2 \chi_i(w) y_i \geq 0,$$

where the equality holds if and only if y_i and $w_{A_i^+}$ are scalar multiples of one another. Strict convexity then follows by a contradiction argument showing that there exists at least one node $i \in I \backslash \{d\}$ such that $y_i^\top \nabla_w^2 \chi_i(w) y_i > 0$. ■

IV. LEARNING DYNAMICS

In this section, we propose a discrete-time dynamical system (PBR) which captures travelers' preferences for minimizing total travel time, as well as their perception uncertainties, while simultaneously learning about the emergent congestion on the network.

We leverage the constant step-size stochastic approximation theory to prove that these discrete-time dynamics converge to a neighborhood of the CoDAG equilibrium (Theorem 2). To this end, we first prove that the continuous-time counterpart to (PBR) globally asymptotically converges to the CoDAG equilibrium (Lemma 2). We then conclude the proof by verifying technical assumptions required to invoke results in stochastic approximation theory [25] (Lemma 3).

A. Discrete-time Dynamics

In this subsection, we present a discrete-time dynamical equations that captures the evolution of flows on the network as a result of learning and adaptation by self-interested travelers. More formally, at each discrete time step $n \geq 0$, g_o units of travelers arrive at the origin node o. At time step n, every traveler who reaches node $i \in I \backslash \{d\}$ selects some arc $a \in A_i^+$. For any $i \in I \backslash \{d\}, a \in A_i^+$, let $\xi_a[n]$ be the *aggregate arc selection probability*: the fraction of travelers at node i choosing arc a at time n. As a result of the arc selections made by every traveler, a flow of $W[n]$ is induced on the arcs as given below. For every $a \in A$:

$$W_a[n] = \left(g_{i_a} + \sum_{a' \in A_{i_a}^-} W_{a'}[n] \right) \cdot \xi_a[n], \quad (12)$$

where, as given in Definition 3, $g_{i_a} = g_o$ if $i_a = o$, and $g_{i_a} = 0$ otherwise.

At the end of each time step, every traveler reaches the destination and observes a noisy estimate of the latency-to-go[1] on every arc in the network (including ones they did not visit in that time step). Note that the latency-to-go for any arc is dependent on the congestion $W[n]$, which in turn depends on aggregate decisions taken by travelers (please refer to (12)). Based on the observed latencies, at time $n+1$, at every non-destination node $i \in I \backslash \{d\}$, a $\eta_i[n+1]$ fraction of travelers at node i switch to an arc with the minimum observed latency-to-go. Meanwhile, a $1 - \eta_i[n+1]$ fraction of travelers selects the same arc they selected at time step n. We assume that $\{\eta_i[n+1] \in \mathbb{R} : i \in I, n \geq 0\}$ are

[1] Latency-to-go realizations for different travelers are independent.

independent bounded random variables[2] in $[\underline{\mu}, \overline{\mu}]$, with $0 < \underline{\mu} < \mu < \overline{\mu} < 1$ and $\mathbb{E}[\eta_{i_a}[n+1]] = \mu$ for each node index $i \in I$ and discrete time index $n \geq 0$. To summarize, the dynamic evolution of arc selections by infinitesimal travelers is captured by the following evolution of $\xi[n]$. For every $i \in I\backslash\{d\}, a \in A_i^+$:

$$\xi_a[n+1] = \xi_a[n] + \eta_{i_a}[n+1] \cdot \left(-\xi_a[n] + P_{ij_a} \right),$$

where P_{ij_a} is defined in (2). Using (4) and (5), the previous equation can be rewritten as:

$$\begin{aligned} &\xi_a[n+1] \qquad \text{(PBR)} \\ =&\xi_a[n] + \eta_{i_a}[n+1] \\ &\cdot \left(-\xi_a[n] + \frac{\exp(-\beta[z_a(W[n])])}{\sum_{a' \in A_{i_a}^+} \exp(-\beta[z_{a'}(W[n])])} \right), \end{aligned}$$

The dynamics (PBR) bears close resemblance to perturbed best response dynamics in routing games [23], so we shall refer to (PBR) as *perturbed best response* dynamics.

We assume $\xi_a[0] > 0$ for each $a \in A$, i.e., each arc has some strictly positive initial traffic flow. This is reasonable, since the stochasticity in travelers' perception of network congestion ensures that each arc has a nonzero probability of being selected.

B. Convergence Results

Our main theorem establishes that the discrete-time dynamics (PBR) asymptotically converges to a neighborhood of the CoDAG equilibrium $\bar{w}^\beta$.

Theorem 2: Under the discrete-time flow dynamics (PBR), for each $\delta > 0$:

$$\limsup_{n \to \infty} \mathbb{E}\big[\|\xi[n] - \bar{\xi}^\beta\|_2^2\big] \leq O(\mu),$$
$$\limsup_{n \to \infty} \mathbb{P}\big(\|\xi[n] - \bar{\xi}^\beta\|_2^2 \geq \delta\big) \leq O\left(\frac{\mu}{\delta}\right).$$

To prove Theorem 2, we leverage the theory of constant step-size stochastic approximation [25]. This requires proving that the continuous-time dynamics corresponding to the discrete-time update (PBR), presented below, converges to the CoDAG equilibrium. For each arc $a \in A$:

$$\dot{\xi}_a(t) = -\xi_a(t) + \frac{\exp(-\beta \cdot z_a(w(t)))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta \cdot z_{a'}(w(t)))}, \qquad (13)$$

where $w(t)$ is the resulting arc flow associated with the arc selection probability $\xi(t)$, similar to (12):

$$w_a(t) = \xi_a(t) \cdot \left(g_{i_a} + \sum_{a' \in A_{i_a}^-} w_{a'}(t) \right). \qquad (14)$$

Lemma 2: Under the continuous-time flow dynamics (13) and (14), the traffic flow $w(t)$ globally asymptotically converges to the CoDAG equilibrium $\bar{w}^\beta$.

[2]The random variables $\{\eta_a[n] : a \in A, n \geq 0\}$ are assumed to be independent of travelers' perception uncertainties.

Proof: (**Proof Sketch**) Recall that $\bar{w}^\beta$ is the unique minimizer of the map $F : \mathcal{W} \to \mathbb{R}$, defined by (11). We show that F is a Lyapunov function for the continuous-time flow dynamics (20) induced by the arc selection dynamics (13). To this end, we first unwind the dynamics (13) and (14) to obtain the recursive relation:

$$\begin{aligned} \dot{w}_a(t) = &-\left(1 - \frac{\sum_{a' \in A_{i_a}^-} \dot{w}_{a'}(t)}{\sum_{\hat{a} \in A_{i_a}^+} w_{\hat{a}}(t)} \right) w_a(t) \\ &+ \sum_{a' \in A_{i_a}^-} w_{a'}(t) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta z_{a'}(w(t)))}. \end{aligned}$$

Then, we establish that along any trajectory starting on $\mathcal{W}$ and following the dynamics given by (13), we have for each $t \geq 0$:

$$\dot{F}(t) = \dot{w}(t)^\top \nabla_w F(w(t)) \leq 0.$$

The proof then follows from LaSalle's Theorem (see [28, Proposition 5.22]).

A detailed proof for Lemma 2 is deferred to Appendix C. ∎

Remark 6: On a technical level, the statement and proof technique of Theorem 2 share similarities with methods used to establish the convergence of best-response dynamics in potential games [23]. However, there exist crucial distinctions between the two approaches which render our problem more difficult. First, since the map F defined by (11) is not a potential function, the mathematical machinery of potential games cannot be directly applied. Moreover, the continuous-time flow dynamics (13) and (14) allow couplings between arbitrary arcs in the CoDAG. For more details, please see Appendix C.

Having established the global asymptotic convergence of the continuous-time dynamics (13) and (14) to the CoDAG equilibrium $\bar{w}^\beta$, it remains to verify the remaining technical conditions necessary to prove Theorem 2 via stochastic approximation theory. To this end, we rewrite the discrete ξ-dynamics (PBR) as a Markov process with a martingale difference term:

$$\xi_a[n+1] = \xi_a[n] + \mu\big(\rho_a(\xi[n]) + M_a[n+1]\big),$$

where $\rho_a : \mathbb{R}^{|A|} \times \mathbb{R}^{|A_O|} \to \mathbb{R}^{|A|}$ is given by:

$$\rho_a(\xi) := -\xi_a + \frac{\exp(-\beta \cdot z_a(w))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta \cdot z_{a'}(w))}, \qquad (15)$$

with $w \in \mathbb{R}^{|A|}$ defined arc-wise by $w_a = (g_{i_a} + \sum_{\hat{a} \in A_{i_a}^-} w_{a'}) \cdot \xi_a$, and:

$$M_a[n+1] := \left(\frac{1}{\mu} \eta_{i_a}[n+1] - 1 \right) \cdot \rho_a(\xi[n]). \qquad (16)$$

Here, $W_a[n] = \big(g_{i_a} + \sum_{a' \in A_{i_a}^-} W_{a'}[n]\big)$, as given by (12).

The following lemma bounds the magnitude of the discrete-time flow $W[n] \in \mathbb{R}^{|A|}$ and the martingale difference terms $M[n] \in \mathbb{R}^{|A|}$.

Lemma 3: Given initial flows $W[0]$ and arc selection probabilities $\xi[0]$:

1) For each $a \in A$: $\{M_a[n+1] : n \geq 0\}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_n := \sigma\big(\cup_{a\in A}(W_a[1], \xi[1], \cdots, W_a[n], \xi[n])\big)$.
2) There exist $C_w, C_m > 0$ such that, for each $a \in A$ and each $n \geq 0$, we have:

$$W_a[n] \in [C_w, g_o],$$
$$|M_a[n]| \leq C_m.$$

3) For each $a \in A$, the map ρ_a, given by (15), is Lipschitz continuous over the range of realizable flow and arc selection probability trajectories $\{W[n] : n \geq 0\}$ and $\{\xi[n] : n \geq 0\}$.

Proof: (**Proof Sketch**) The first part of Lemma 3 follows by observing that, with respect to $\mathcal{F}_n$, the only stochasticity in $M_a[n+1]$ originates from the input flows $\eta_{i_a}[n+1]$, which are i.i.d. The second part follows by invoking the flow continuity equations in (12) to recursively upper bound each $W_a[n]$ and $z_a(W[n])$, in increasing order of depth and height, respectively (since flows are propagated from origin to destination, and latency-to-go values are computed in the opposite direction). These bounds are then used to recursively establish upper and lower bounds for each $\xi_a[n]$, and consequently each $W[n]$, in order of increasing depth. Finally, the Lipschitz continuity of each ρ_a can be proved by establishing that ρ_a is continuously differentiable, with bounded derivatives over the compact domain defined by the bounds on $W[n]$ established in the second part of the lemma. For details, please see the proofs of Lemmas 5 and 6 in Appendix C. ■

V. EXPERIMENT RESULTS

In this section, we conduct numerical experiments to validate the theoretical analysis presented in Section IV. We show in simulation that, under (PBR), the traffic flows converge to a neighborhood of the condensed DAG equilibrium, as claimed by Theorem 2.

Consider the network presented in Figure 1, with affine edge-latency functions $s_{[a]}(w_{[a]}) = k_0 + k_1 w_{[a]}$ for each arc $a \in A$, where $k_0, k_1 > 0$ are simulation parameters provided in Table II. To validate Theorem 2, we evaluate and plot the traffic flow values $W_a[n]$ on each arc $a \in A$ and discrete time $n \geq 0$. Figure 3 illustrates that w converges to the condensed DAG equilibrium in approximately 15 iterations. Meanwhile, Figure 2 presents traffic flow values at the condensed DAG equilibrium (i.e., w^β) for the original network and condensed DAG. While travelers generally prefer routes of lower latency, each route has a nonzero level of traffic flow at equilibrium. The reason is that under the perturbed best response dynamics, users do not allocate all the traffic flow to the minimum-cost route, but instead distribute their traffic allocation more evenly.

VI. CONCLUSION AND FUTURE WORK

We present a new equilibrium concept for stochastic arc-based TAMs in which travelers are guaranteed to be routed on acyclic routes. Specifically, we construct a condensed DAG representation of original network, by replicating arcs and nodes to avoid cyclic routes, while preserving the set of feasible routes from the original network. We show that the proposed equilibrium can be characterized as the optimal solution of a strictly convex optimization problem. Furthermore, we propose adaptive learning dynamics for arc-based TAM that characterizes the evolution of flow generated by the simultaneous learning and adaptation of self-interested travelers. Additionally, we show that the learning dynamics converges to the equilibrium.

Interesting avenues of future research include: (i) Developing an equilibrium notion and corresponding convergent learning dynamics, for the case in which travelers can only access latency-to-go values on the routes they choose; and (ii) Developing dynamic tolling mechanisms to properly allocate equilibrium flows to induce social optimal loads.

TABLE II: Parameters for simulation

Notation	Default value
k_0	0, 1, 0, 1, 1, 0, 1, 1, 1 (ordered by edge index)
k_1	2, 1, 1, 1, 1, 1, 2, 2, 2 (ordered by edge index)
g_1	1
β	10
$\eta_{i_a}[n]$	0.9^n, $\forall a \in A, i \in I\backslash\{d\}$

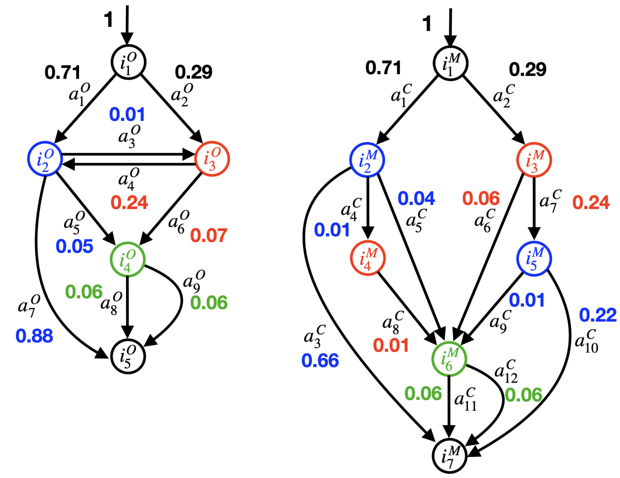

Fig. 2: Steady state traffic flow on each arc for an original network and condensed DAG. Flows on arcs emerging from same node are represented in same color.

REFERENCES

[1] Roberto Cominetti, Francisco Facchinei, and Jean B Lasserre. *Modern Optimization Modeling Techniques*. Springer Science & Business Media, 2012.
[2] John Glen Wardrop. "Some Theoretical Aspects of Road Traffic Research." In: *Proceedings of the institution of civil engineers* 1.3 (1952), pp. 325–362.
[3] Jean-Bernard Baillon and Roberto Cominetti. "Markovian Traffic Equilibrium". In: *Mathematical Programming* (Feb. 2008). DOI: 10.1007/s10107-006-0076-2.
[4] Mogens Fosgerau, Emma Frejinger, and Anders Karlstrom. "A Link-based Network Route Choice Model with Unrestricted Choice Set". In: *Transportation Research Part B: Methodological* 56 (2013), pp. 70–80.
[5] Carlos F Daganzo and Yosef Sheffi. "On Stochastic Models of Traffic Assignment". In: *Transportation science* 11.3 (1977), pp. 253–274.

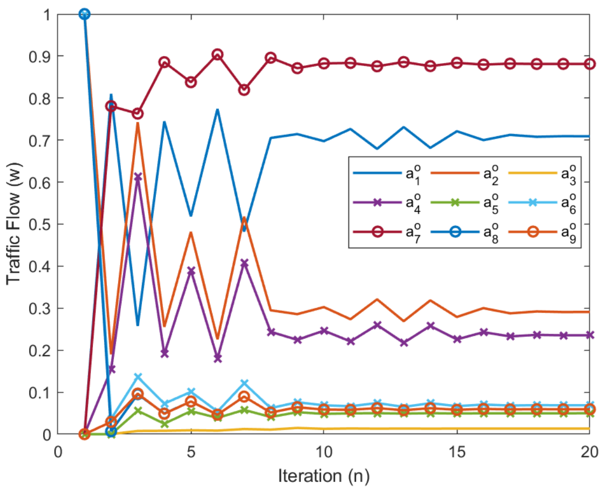

Fig. 3: Traffic flow $W[n]$ for the network in Fig. 2.

[6] Takashi Akamatsu. "Decomposition of Path Choice Entropy in General Transport Networks". In: *Transportation Science* 31.4 (Nov. 1997), pp. 349–362. DOI: 10.1287/trsc.31.4.349.

[7] Robert B. Dial. "A Probabilistic Multipath Traffic Assignment Model which Obviates Path Enumeration". In: *Transportation Research* 5.2 (1971), pp. 83–111. ISSN: 0041-1647.

[8] Yosef Sheffi and Warren Powell. "A Comparison of Stochastic and Deterministic Traffic Assignment over Congested Networks". In: *Transportation Research Part B: Methodological* 15.1 (1981), pp. 53–64.

[9] Tetsuo Yai, Seiji Iwakura, and Shigeru Morichi. "Multinomial Probit with Structured Covariance for Route Choice Behavior". In: *Transportation Research Part B: Methodological* 31.3 (1997), pp. 195–207.

[10] Maëlle Zimmermann and Emma Frejinger. "A Tutorial on Recursive Models for Analyzing and Predicting Path Choice Behavior". In: *EURO Journal on Transportation and Logistics* 9.2 (2020), p. 100004.

[11] Yuki Oyama and Eiji Hato. "A Discounted Recursive Logit Model for Dynamic Gridlock Network Analysis". In: *Transportation Research Part C: Emerging Technologies* 85 (2017), pp. 509–527.

[12] Tien Mai, Mogens Fosgerau, and Emma Frejinger. "A Nested Recursive Logit Model for Route Choice Analysis". In: *Transportation Research Part B: Methodological* 75 (2015), pp. 100–112.

[13] Tien Mai. "A Method of Integrating Correlation Structures for a Generalized Recursive Route Choice Model". In: *Transportation Research Part B: Methodological* 93 (2016), pp. 146–161.

[14] Yuki Oyama and Eiji Hato. "Prism-based Path Set Restriction for Solving Markovian Traffic Assignment Problem". In: *Transportation Research Part B: Methodological* 122 (2019), pp. 528–546.

[15] Yuki Oyama, Yusuke Hara, and Takashi Akamatsu. "Markovian Traffic Equilibrium Assignment Based on Network Generalized Extreme Value Model". In: *Transportation Research Part B: Methodological* 155 (2022), pp. 135–159.

[16] Dan Calderone and S Shankar Sastry. "Markov Decision Process Routing Games". In: *Proceedings of the 8th International Conference on Cyber-Physical Systems*. 2017, pp. 273–279.

[17] Dan Calderone and Shankar Sastry. "Infinite-horizon Average-cost Markov Decision Process Routing Games". In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2017, pp. 1–6.

[18] Drew Fudenberg and David K Levine. *The Theory of Learning in Games*. Vol. 2. MIT press, 1998.

[19] Walid Krichene, Benjamin Drighes, and Alexandre Bayen. "On the Convergence of no-Regret Learning in Selfish Routing". In: *International Conference on Machine Learning*. PMLR. 2014, pp. 163–171.

[20] Syrine Krichene, Walid Krichene, Roy Dong, and Alexandre Bayen. "Convergence of Heterogeneous Distributed Learning in Stochastic Routing Games". In: *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2015, pp. 480–487.

[21] Robert Kleinberg, Georgios Piliouras, and Éva Tardos. "Multiplicative Updates Outperform Generic no-Regret Learning in Congestion Games". In: *Proceedings of the forty-first annual ACM symposium on Theory of computing*. 2009, pp. 533–542.

[22] Chinmay Maheshwari, Kshitij Kulkarni, Manxi Wu, and S. Shankar Sastry. "Dynamic Tolling for Inducing Socially Optimal Traffic Loads". In: *2022 American Control Conference (ACC)*. 2022, pp. 4601–4607. DOI: 10.23919/ACC53348.2022.9867193.

[23] William H. Sandholm. *Population Games And Evolutionary Dynamics*. Economic Learning and Social Evolution, 2010.

[24] Emily Meigs, Francesca Parise, and Asuman Ozdaglar. "Learning Dynamics in Stochastic Routing Games". In: *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2017, pp. 259–266.

[25] Vivek Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

[26] Andrea Papola and Vittorio Marzano. "A Network Generalized Extreme Value Model for Route Choice Allowing Implicit Route Enumeration". In: *Computer-Aided Civil and Infrastructure Engineering* 28.8 (2013), pp. 560–580.

[27] M. E. Ben-Akiva. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge: MIT Press, 1985.

[28] Shankar Sastry. *Nonlinear Systems: Analysis, Stability, and Control*. Springer, 1999.

APPENDIX

Please use the following link to access a version with the appendix (https://drive.google.com/file/d/14Yh88gzhlqPO_e3PC_XSO54v1vOtuH6H/view?usp=sharing). If the hyperlinks do not work, the URL can be copied and pasted in a web browser; we apologize for the inconvenience. The authors will ensure that this link stays active.

Below, we present proofs omitted in the main paper due to space limitations.

A. Properties of Depth and Height

In the main text, we recursively defined some dynamical quantities, such as the time evolution of the traffic flows $w \in \mathbb{R}^{|A|}$ and the latency-to-go $z \in \mathbb{R}^{|A|}$, in a component-wise fashion, either from the origin of the Condensed DAG G towards the destination, or from the destination to the origin. To facilitate these recursive definitions, we require the following characterizations regarding the depths and heights of arcs in a Condensed DAG G.

1) Depth: First, we define the concept of depth of a directed acyclic graph (DAG), which will be crucial for the remaining exposition.

*Definition 4 (**Depth of a DAG**):* Given a DAG $G = (I, A)$ describing a single-origin single-destination traffic network, the *depth of* G, denoted $\ell(G)$, is defined by:

$$\ell(G) := \max_{a \in A} \ell_a$$

In this work, we consider only acyclic routes in traffic networks with finitely many edges, so we have $\ell(G) < \infty$. Moreover, the case $\ell(G) = 1$ corresponds to a parallel link network, for which the results of the following proposition have already been analyzed in [22]. Therefore, we assume below that $\ell(G) \geq 2$.

Proposition 1: Given a Condensed DAG $G = (I, A)$ with the route set $\mathbf{R}$:

1) For any $a \in A$, we have $\ell_a = 1$ if and only if $i_a = o$. Similarly, if $\ell_a = \ell(G)$, then $j_a = d$.
2) For any fixed $r \in \mathbf{R}$, and any $a, a' \in r$ with $\ell_{a,r} < \ell_{a',r}$, we have $\ell_a < \ell_{a'}$ i.e., arcs along a route have strictly increasing depth from the origin to the destination.
3) Fix any $a \in A$, and any $r \in \mathbf{R}$ containing a such that $\ell_{a,r} = \ell_a$. Then, for any $a' \in \mathbf{R}$ preceding a in r, we have $\ell_{a',r} = \ell_{a'}$.
4) For each depth $k \in [\ell(G)] := \{1, \cdots, \ell(G)\}$, there exists some $a \in A$ such that $\ell_a = k$.

Proof:

1) If $\ell_a \neq 1$, then $\ell_a \geq 2$, so there exists at least one route $r \in \mathbf{R}$ containing $a \in A$ such that $\ell_{a,r} \geq 2$. Thus, $i_a \neq o$ (otherwise the first $\ell_{a,r} - 1$ arcs of r would form a cycle). Conversely, if $i_a \neq o$, then no route $r \in \mathbf{R}$ contains $a \in A$ as its first arc, i.e., $\ell_{a,r} \geq 2$ for each $r \in \mathbf{R}$ containing a. Thus, $\ell_a = \max_{r \in \mathbf{R}: a \in r} \ell_{a,r} \geq 2$; in particular, $\ell_a \neq 1$. This establishes that $\ell_a = 1$ if and only if $i_a = o$.

   Now, suppose by contradiction that there exists some $a \in A$ such that $\ell_a = \ell(G)$ but $j_a \neq d$. Fix any $r \in \mathbf{R}$ such that $a \in r$ and $\ell_{a,r} = \ell_a$. Then a cannot be at the end of $\mathbf{R}$, since by definition, routes must end at d. Let $a' \in r$ be the arc immediately after a in r. Then $\ell_{a'} \geq \ell_{a',r} = \ell_{a,r} + 1 = \ell(G) + 1$, a contradiction to the definition of $\ell(G)$.
2) Fix $r \in \mathbf{R}$, $a, a' \in r$ such that $\ell_{a,r} < \ell_{a',r}$. If $\ell_a = 1$, then $\ell_{a'} \geq \ell_{a',r} > \ell_{a,r} = 1 = \ell_a$, and we are done. Suppose $\ell_a \geq 2$. By definition of ℓ_a, there exists some route r_2 such that $\ell_{a,r_2} = \ell_a$. Construct a new route $r_3 \in \mathbf{R}$ by replacing the first $\ell_{a,r}$ arcs of r with the first ℓ_{a,r_2} arcs of r_2. Then $\ell_{a'} \geq \ell_{a',r_3} = \ell_{a',r} - \ell_{a,r} + \ell_{a,r_2} > \ell_{a,r_2} = \ell_a$.
3) Fix any $a \in A$, and any $r \in \mathbf{R}$ containing a such that $\ell_{a,r} = \ell_a$. Suppose by contradiction that there exists some $a' \in \mathbf{R}$, preceding a in r, for which $\ell_{a'} \geq \ell_{a',r} + 1$. Then, by applying the second part of this lemma along the $(\ell_{a,r} - \ell_{a',r})$ arcs of $\mathbf{R}$ from a' to a, we find that $\ell_a \geq \ell_{a'} + (\ell_{a,r} - \ell_{a',r}) \geq \ell_{a,r} + 1 = \ell_a + 1$, a contradiction.
4) Fix any arc $a \in A$ with $\ell_a = \ell(G)$. Then there exists some $r \in \mathbf{R}$ containing a such that $\ell_{a,r} = \ell_a = \ell(G)$. It follows from the third part of this proposition that, for each $k \in [\ell(G)]$, the k-th arc in $\mathbf{R}$ is of depth k.

■

2) Height: Next, we define the concept of height of a directed acyclic graph (DAG), which will be crucial for the remaining exposition.

*Definition 5 (**Height of a DAG**):* Given a DAG $G = (I, A)$ describing a single-origin single-destination traffic network, the *height of* G, denoted $m(G)$, is defined by:

$$m(G) := \max_{a \in A} m_a$$

Since the traffic network under study is finite, and we consider only acyclic routes, we have $m(G) < \infty$. Moreover, the case $m(G) = 1$ corresponds to a parallel link network, for which the results of the following proposition have already been extensively analyzed in [22]. We will henceforth assume that $m(G) \geq 2$.

Proposition 2: Given an Condensed DAG $G = (I, A)$ with the route set $\mathbf{R}$:

1) For any $a \in A$, we have $m_a = 1$ if and only if $j_a = d$. Similarly, if $m_a = m(G)$, then $i_a = o$.
2) For any fixed $r \in \mathbf{R}$, and any $a, a' \in r$ with $m_{a,r} < m_{a',r}$, we have $m_a < m_{a'}$ i.e., arcs along a route from the origin to the destination have strictly decreasing depth.
3) Fix any $a \in A$, and any $r \in \mathbf{R}$ containing a such that $m_{a,r} = m_a$. Then, for any $a' \in \mathbf{R}$ following a in r, we have $m_{a',r} = m_{a'}$.
4) For each height $k \in [m(G)] := \{1, \cdots, m(G)\}$, there exists an arc $a \in A$ such that $m_a = k$.

The proof of Proposition 2 parallels that of Proposition 1, and is omitted for brevity.

B. Proofs of results in Section III

1) Proof of Lemma 1: Here, we establish Lemma 1, restated as follow: The map $F : \mathcal{W} \to \mathbb{R}$, as given below, is strictly convex.

$$F(w) := \sum_{[a] \in A_O} \int_0^{w_{[a]}} s_{[a]}(u)\, du + \frac{1}{\beta} \sum_{i \neq d} \Bigg[\sum_{a \in A_i^+} w_a \ln w_a - \Bigg(\sum_{a \in A_i^+} w_a \Bigg) \ln \Bigg(\sum_{a \in A_i^+} w_a \Bigg) \Bigg].$$

For convenience, we define $f_{[a]} : \mathcal{W} \to \mathbb{R}$, $\chi_i : \mathbb{R}^{|A_i^+|} \to \mathbb{R}$, $F : \mathcal{W} \to \mathbb{R}$ for each $[a] \in A_O$, $i \in I \backslash \{d\}$ by:

$$f_{[a]}(w) := \int_0^{w_{[a]}} s_{[a]}(u)\, du, \qquad \forall [a] \in A_O,$$

$$\chi_i(w_{A_i^+}) := \sum_{a \in A_i^+} w_a \ln w_a - \left(\sum_{a \in A_i^+} w_a \right) \ln \left(\sum_{a \in A_i^+} w_a \right), \quad \forall i \neq I\backslash\{d\},$$

where $w_{A_i^+} \in \mathbb{R}^{|A_i^+|}$ denotes the components of w corresponding to arcs in A_i^+. Then:

$$F(w) = \sum_{[a] \in A_0} f_{[a]}(w) + \frac{1}{\beta} \sum_{i \in I\backslash\{d\}} \chi_i^\beta(w).$$

Also, for convenience, define:

$$\mathcal{W}_s := \Bigg\{ w \in \mathbb{R}^{|A|} : \sum_{a \in A_i^+} w_a = \sum_{a \in A_i^-} w_a, \forall i \neq o, d, \quad (17)$$
$$\sum_{a \in A_o^+} w_a = 0. \Bigg\}.$$

Essentially, $\mathcal{W}_s$ is the tangent space of the linear manifold with boundary $\mathcal{W}$. Note that, using the notation described at the end of Section I, we can rewrite (17) as:

$$\mathcal{W}_s = \left\{ e_{A_i^-} - e_{A_i^+} : i \neq o, d \right\}^\perp \cap \left\{ e_{A_o^+} \right\}^\perp.$$

We can now establish the strict convexity of F.

We first establish the convexity of F. It suffices to show that $f_{[a]}$ and χ_i are convex for each $[a] \in A_O$, $i \in I\backslash\{d\}$. Note that each $f_{[a]}$ is convex since it is the composition of a convex function ($g(w) = \sum_{a \in A_0} \int_0^{w_a} s_a(u) du$) with a linear function ($w_{[a]} := \sum_{a' \in [a]} w_{a'}$). We show below that χ_i is convex, for each $i \in I\backslash\{d\}$.

Fix $i \in I\backslash\{d\}$. For any $a, a' \in A_i^+$ and each $w \in \mathcal{W}$:

$$\frac{\partial^2 \chi_i}{\partial w_a \partial w_{a'}}(w) = \frac{1}{w_a} \mathbf{1}\{a' = a\} - \frac{1}{\sum_{\bar{a} \in A_i^+} w_{\bar{a}}}.$$

Thus, for any $y \in \mathbb{R}^{|A_i^+|}$:

$$\begin{aligned}
& y^\top \nabla_w^2 \chi_i(w) y \\
&= \sum_{a, a' \in A_i^+} y_a y_{a'} \frac{\partial^2 \chi_i}{\partial w_a \partial w_{a'}}(w) \\
&= \sum_{a \in A_i^+} \frac{y_a^2}{w_a} - \frac{1}{\sum_{\bar{a} \in A_i^+} w_{\bar{a}}} \cdot \sum_{a, a' \in A_i^+} y_a y_{a'} \\
&= \frac{1}{\sum_{\bar{a} \in A_i^+} w_{\bar{a}}} \left(\sum_{\bar{a} \in A_i^+} w_{\bar{a}} \cdot \sum_{a \in A_i^+} \frac{y_a^2}{w_a} - \left(\sum_{a' \in A_i^+} y_{a'} \right)^2 \right) \\
&= \frac{1}{\sum_{\bar{a} \in A_i^+} w_{\bar{a}}} \left(\sum_{\bar{a} \in A_i^+} \left(\sqrt{w_{\bar{a}}}\right)^2 \cdot \sum_{a \in A_i^+} \left(\frac{y_a}{\sqrt{w_a}} \right)^2 \right. \\
&\qquad \left. - \left(\sum_{a' \in A_i^+} \sqrt{w_{a'}} \cdot \frac{y_{a'}}{\sqrt{w_{a'}}} \right)^2 \right) \\
&\geq 0, \qquad (18)
\end{aligned}$$

where the final inequality follows from the Cauchy-Schwarz inequality. Cauchy-Schwarz also implies that equality holds in (18) if and only if the vectors $(\sqrt{w_a})_{a \in A_i^+} \in \mathbb{R}^{|A_i^+|}$ and $(y_a/\sqrt{w_a})_{a \in A_i^+} \in \mathbb{R}^{|A_i^+|}$ are parallel, i.e., if $(y_a)_{a \in A_i^+}$ and $(w_a)_{a \in A_i^+}$ are scalar multiples of each other. This shows that χ_i is convex, and $dim(N(\nabla_w^2 \chi_i)) = 1$.

Second, suppose by contradiction that F is not strictly convex on $\mathcal{W}$. Then there exists some $\bar{w} \in \mathcal{W}$, $z \in \mathcal{W}_s \backslash \{0\}$ such that:

$$z^\top \nabla_w^2 F(\bar{w}) z = 0.$$

Since $\nabla_w^2 F(\bar{w})$ is symmetric positive semidefinite, this is equivalent to stating that z is in $N(\nabla_w^2 F(\bar{w}))$, the null space of $\nabla_w^2 F(\bar{w})$. Let A_z denote the set of arc indices for which z has a nonzero component, i.e.:

$$A_z := \{a' \in A : z_{a'} \neq 0\}.$$

Since z is not the zero vector, A_z is non-empty. Since there are a discrete and finite number of levels of G, there exists some $a \in A_z$ such that $\ell_a \leq \ell_{a'}$ for all $a' \in A_y$, i.e., $\ell_a = \min\{\ell_{a'} : a' \in A_y\}$. Without loss of generality, we consider the case $z_a > 0$ (if not, then replace z with $-z$, which would also be a nonzero vector in $N(\nabla_w^2 F(\bar{w}))$). We claim that $w_a \neq 0$, and that for all $a' \in A_{i_a}^+$:

$$z_{a'} = z_a \cdot \frac{w_{a'}}{w_a} \geq 0.$$

To see this, note that otherwise, the vectors $(z_a)_{a \in A_i^+} \in \mathbb{R}^{|A_i^+|}$ and $(w_a)_{a \in A_i^+}$ are not parallel, and so equality cannot be obtained in (18), i.e.,:

$$z^\top \nabla_w^2 \chi_i(\bar{w}) z > 0,$$

where, with a slight abuse of notation, we have defined $\chi_i(w) = \chi_i(A_i^+)$. As a result:

$$\begin{aligned}
& z^\top \nabla_w^2 F(\bar{w}) z \\
&= \sum_{[a] \in A} z^\top \nabla_w^2 f_{[a]}(\bar{w}) z + \frac{1}{\beta} \sum_{i' \neq d} z^\top \nabla_w^2 \chi_{i'}(\bar{w}) z \\
&\geq \frac{1}{\beta} z^\top \nabla_w^2 \chi_i(\bar{w}) z \\
&> 0,
\end{aligned}$$

a contradiction. Thus, $z_a > 0$, and $z_{a'} \geq 0$ for each $a' \in A_{i_a}^+$, so:

$$\sum_{a' \in A_{i_a}^+} z_{a'} > 0.$$

If $\ell_a = 1$, i.e., $i_a = o$, we arrive at a contradiction, since the fact that $z \in \mathcal{W}_s$ implies $\sum_{a' \in A_{i_a}^+} z_{a'} = 0$. If $\ell_a > 1$, we also arrive at a contradiction, since the fact that $z \in \mathcal{W}_s$ implies:

$$\sum_{\hat{a} \in A_{i_a}^-} z_{\hat{a}} = \sum_{a' \in A_{i_a}^+} z_{a'} > 0,$$

so there exists at least one $\ell_{\hat{a}} \in A_{i_a}^-$ with $z_{\hat{a}} > 0$. Then, by definition of $a \in A$, we have $\ell_a \leq \ell_{\hat{a}}$; this contradicts Proposition 1, Part 2, which implies that since $\hat{a} \in A_{i_a}^-$, there exists at least one arc containing $\hat{a}$ immediately before $a \in A$, and thus $\ell_{\hat{a}} \leq \ell_a - 1$. These contradictions complete the proof of the strict convexity of F on $\mathcal{W}$.

2) Proof of Theorem 1: We present the proof of Theorem 1, restated as follows: The Condensed DAG Equilibrium $\bar{w}^\beta \in \mathcal{W}$ exists, is unique, and is the unique optimal solution to the following convex optimization problem:

$$\begin{aligned} \min_{w \in \mathcal{W}} \ & \sum_{[a] \in A_0} \int_0^{w_{[a]}} s_{[a]}(u)\, dz \\ & + \frac{1}{\beta} \sum_{i \neq d} \Bigg[\sum_{a \in A_i^+} w_a \ln w_a \\ & \qquad - \Bigg(\sum_{a \in A_i^+} w_a \Bigg) \ln \Bigg(\sum_{a \in A_i^+} w_a \Bigg) \Bigg]. \end{aligned}$$

Proof: (**Proof of Theorem 1**) The following proof parallels that of Baillon, Cominetti [3, Theorem 2]. Recall that N denotes the set of nodes of the corresponding DAG. The Lagrangian $L : W \times \mathbb{R}^{|N|-1} \in \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ corresponding to the above optimization problem is:

$$\begin{aligned} & L(w, \mu, \lambda) \\ := & \sum_{[a] \in A_0} \int_0^{w_{[a]}} s_{[a]}(u)\, dz \\ & + \frac{1}{\beta} \sum_{i \neq d} \Bigg[\sum_{a \in A_i^+} w_a \ln w_a - \Bigg(\sum_{a \in A_i^+} w_a \Bigg) \ln \Bigg(\sum_{a \in A_i^+} w_a \Bigg) \Bigg] \\ & + \sum_{i \neq d} \mu_i \Bigg(g_i + \sum_{a' \in A_i^-} w_{a'} - \sum_{a' \in A_i^+} w_{a'} \Bigg) + \sum_{a \in A} \lambda_a w_a, \end{aligned}$$

with $g_i = g_o \cdot \mathbf{1}\{i = o\}$, where $\mathbf{1}\{\cdot\}$ is the indicator function that returns 1 if the input argument is true, and 0 otherwise. At optimum $(w^\star, \mu^\star) \in \mathcal{W} \times \mathbb{R}^{|N|-1}$, the KKT conditions give, for each $a \in A$:

$$\begin{aligned} 0 &= \frac{\partial L}{\partial w_a}(w^\star, \mu^\star) \\ &= s_{[a]}(w_{[a]}^\star) + \frac{1}{\beta} \ln \Bigg(\frac{w_a^\star}{\sum_{a' \in A_{i_a}^+} w_{a'}^\star} \Bigg) + \mu_{j_a}^\star - \mu_{i_a}^\star + \lambda_a, \\ 0 &= \lambda_a w_a, \qquad \forall\, a \in A. \end{aligned}$$

We claim that $(\hat{w}, \hat{\mu}) \in \mathcal{W} \times \mathbb{R}^{|N|-1}$, as given by the Condensed DAG equilibrium definition: For each $a \in A$, $i \in N$:

$$\begin{aligned} \hat{w}_a &= \Bigg(g_{i_a} + \sum_{a' \in A_{i_a}^-} \hat{w}_{a'} \Bigg) \cdot \frac{\exp(-\beta z_a(\hat{w}))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta z_{a'}(\hat{w}))}, \\ & \qquad \forall\, a \in A, \\ \hat{\mu}_i &= \varphi_i(z(\hat{w})) = -\frac{1}{\beta} \ln \Bigg(\sum_{a' \in A_i^+} e^{-\beta z_{a'}(\hat{w})} \Bigg), \\ & \qquad \forall\, i \in N, \\ \hat{\lambda}_a &= 0, \qquad \forall\, a \in A, \end{aligned}$$

satisfies the KKT conditions stated above. Indeed, we have $\hat{w}_a \geq 0$ for each $a \in A$, and:

$$\begin{aligned} & \frac{\partial L}{\partial w_a}(\hat{w}, \hat{\mu}, \hat{\lambda}) \\ =& s_{[a]}(\hat{w}_{[a]}) + \frac{1}{\beta} \ln \Bigg(\frac{\hat{w}_a}{\sum_{a' \in A_{i_a}^+} \hat{w}_{a'}} \Bigg) \\ & \qquad + \hat{\mu}_{j_a} - \hat{\mu}_{i_a} + \sum_{a \in A} \lambda_a \\ =& s_{[a]}(\hat{w}_{[a]}) + \frac{1}{\beta} \ln \Bigg(\frac{\exp(-\beta z_a(w))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta z_{a'}(w))} \Bigg) \\ & \qquad + \varphi_{j_a}(z) - \varphi_{i_a}(z) \\ =& s_{[a]}(\hat{w}_{[a]}) - z_a(w) + \varphi_{i_a}(z) + \varphi_{j_a}(z) - \varphi_{i_a}(z) \\ =& s_{[a]}(\hat{w}_{[a]}) + \varphi_{j_a}(z) - z_a(w) \\ =& 0, \end{aligned}$$

where the final equality follows from the definition of $(z_a)_{a \in A}$. ■

C. Proofs for Section IV

1) Proof of Lemma 2: We present the proof of Lemma 2, restated as follows: If $w(0) \in \mathcal{W}$, the continuous-time dynamical system (20) for the traffic flow $w(t)$ globally asymptotically converges to the corresponding Condensed DAG Equilibrium $\bar{w}^\beta \in \mathcal{W}$.

Proof: (**Proof of Lemma 2**) We recursively write the continuous-time evolution of the arc flows $w(\cdot)$ as follows, from (13) and (14). Recall that for any fixed $w \in \mathcal{W}$, at each non-destination node $i \in I \backslash \{d\}$, we have $\sum_{a' \in A_{i_a}^+} w_{a'} = \sum_{\hat{a} \in A_{i_a}^-} w_{\hat{a}}$. Thus, for each $a \notin A_o^+$:

$$\begin{aligned} & \dot{w}_a(t) \\ =& \dot{\xi}_a(t) \cdot \sum_{\hat{a} \in A_{i_a}^-} w_{\hat{a}}(t) + \xi_a(t) \cdot \sum_{\hat{a} \in A_{i_a}^-} \dot{w}_c(t) \\ =& \Bigg(-\xi_a(t) + \frac{\exp(-\beta z_a(w(t)))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta z_{a'}(w(t)))} \Bigg) \cdot \sum_{a' \in A_{i_a}^-} w_{a'}(t) \\ & \qquad + \xi_a(t) \cdot \sum_{\hat{a} \in A_{i_a}^-} \dot{w}_{\hat{a}}(t) \\ =& -w_a(t) + \sum_{\hat{a} \in A_{i_a}^-} w_{\hat{a}}(t) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta z_{a'}(w(t)))} \\ & \qquad + \frac{w_a(t)}{\sum_{a' \in A_{i_a}^+} w_{a'}(t)} \cdot \sum_{\hat{a} \in A_{i_a}^-} \dot{w}_{\hat{a}}(t) \end{aligned}$$

$$= -\left(1 - \frac{\sum_{\hat{a}\in A_{i_a}^-} \dot{w}_{\hat{a}}}{\sum_{a'\in A_{i_a}^+} w_{a'}}\right) w_a \quad (19)$$

$$+ \sum_{\hat{a}\in A_{i_a}^-} w_{\hat{a}}(t) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{a'\in A_{i_a}^+} \exp(-\beta z_{a'}(w(t)))},$$

for each $a \in A$. More formally, we define each component $h : \mathcal{W} \to \mathbb{R}^{|A|}$ recursively as follows. First, for each $a \in A_o^+$, we set:

$$h_a(w) := -w_a + g_o \cdot \frac{\exp(-\beta z_a(w))}{\sum_{a'\in A_o^+} \exp(-\beta z_{a'}(w))}.$$

Suppose now that, for some arc $a \in A$, the component $h_a : \mathcal{W} \to \mathbb{R}$ of h has been defined for each $\hat{a} \in A_{i_a}^-$. Then, we set:

$$h_a(w) := -\left(1 - \frac{\sum_{\hat{a}\in A_{i_a}^-} h_{\hat{a}}(w)}{\sum_{a'\in A_{i_a}^+} w_{a'}}\right) w_a + \sum_{a'\in A_{i_a}^-} w_{a'} \cdot \frac{\exp(-\beta z_a(w))}{\sum_{a'\in A_o^+} \exp(-\beta z_{a'}(w))}.$$

By iterating through the above definition forward through the Condensed DAG G from origin to destination, we can completely specify each h_a in a well-posed manner (For a more rigorous characterization of this iterative procedure, see Appendix A, Proposition 1). We then define the w-dynamics corresponding to the ξ-dynamics (13) by:

$$\dot{w} = h(w). \quad (20)$$

Now, recall the objective $F : \mathcal{W} \times \mathbb{R}^{|A_O|} \to \mathbb{R}$ of the optimization problem that characterizes $\bar{w}^\beta$, first stated in Theorem 1 as Equation (11), reproduced below:

$$F(w) := \sum_{[a]\in A_0} \int_0^{w_{[a]}} s_{[a]}(z)\, dz + \frac{1}{\beta} \sum_{i\neq d} \left[\sum_{a\in A_i^+} w_a \ln w_a - \left(\sum_{a\in A_i^+} w_a\right) \ln \left(\sum_{a\in A_i^+} w_a\right)\right].$$

Roughly speaking, our main approach is to show that F *is a Lyapunov function for the best-response dynamics in* (20). Let $\mathcal{W}_s$ denote the tangent space to $\mathcal{W}$, and let $\Pi_{\mathcal{W}_s}$ denote the orthogonal projection onto the linear subspace $\mathcal{W}_s$. Under the continuous-time flow dynamics (13) and (14), if $w \neq \bar{w}^\beta$:

$$\frac{d}{dt}(F \circ w)(t)$$
$$= \dot{w}(t)^\top \nabla_w F(w(t)) \quad (21)$$
$$= \dot{w}(t)^\top \Pi_{\mathcal{W}_s} \nabla_w F(w(t)) \quad (22)$$
$$= \dot{w}(t)^\top \Pi_{\mathcal{W}_s} \big(\nabla_w f(w(t)) + \nabla \chi^\beta(w(t))\big)$$
$$= \dot{w}(t)^\top \Pi_{\mathcal{W}_s} \Big(\big(s_{[a]}(w_{[a]}(t))\big)_{a\in A} + \nabla \chi^\beta(w(t))\Big) \quad (23)$$
$$= \dot{w}(t)^\top \Pi_{\mathcal{W}_s} \Bigg[-\nabla \chi^\beta \Bigg(\Bigg(\Bigg(g_{i_a} + \sum_{a'\in A_{i_a}^-} w_{a'}(t)\Bigg) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{\bar{a}\in A_i^+} \exp(-\beta z_{\bar{a}}(w(t)))}\Bigg)_{a\in A}\Bigg) + \nabla \chi^\beta(w(t))\Bigg] \quad (24)$$

$$= \dot{w}(t)^\top \Pi_{\mathcal{W}_s} \Bigg[-\nabla \chi^\beta \Bigg(\Bigg(\Bigg(g_{i_a} + \sum_{a'\in A_{i_a}^-} w_{a'}(t)\Bigg) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{\bar{a}\in A_i^+} \exp(-\beta z_{\bar{a}}(w(t)))}\Bigg)_{a\in A}\Bigg) + \nabla \chi^\beta \Bigg(\Bigg(\Bigg(1 - \frac{\sum_{a'\in A_{i_a}^-} h_{a'}(w(t))}{\sum_{\hat{a}\in A_{i_a}^+} w_{\hat{a}}(t)}\Bigg) \cdot w_a(t)\Bigg)_{a\in A}\Bigg)\Bigg] \quad (25)$$

$$= \Bigg[\Bigg(-\Bigg(1 - \frac{\sum_{a'\in A_{i_a}^-} h_{a'}(w(t))}{\sum_{\hat{a}\in A_{i_a}^+} w_{\hat{a}}(t)}\Bigg) w_a(t) + \Bigg(g_{i_a} + \sum_{a'\in A_{i_a}^+} w_{a'}(t)\Bigg) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{\bar{a}\in A_{i_a}^+} \exp(-\beta z_{\bar{a}}(w(t)))}\Bigg)_{a\in A}\Bigg]^\top \quad (26)$$
$$\Bigg[-\nabla \chi^\beta \Bigg(\Bigg(\Bigg(g_{i_a} + \sum_{a'\in A_{i_a}^-} w_{a'}(t)\Bigg) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{\bar{a}\in A_i^+} \exp(-\beta z_{\bar{a}}(w(t)))}\Bigg)_{a\in A}\Bigg) + \nabla \chi^\beta \Bigg(\Bigg(\Bigg(1 - \frac{\sum_{a'\in A_{i_a}^-} h_{a'}(w(t))}{\sum_{\hat{a}\in A_{i_a}^+} w_{\hat{a}}(t)}\Bigg) \cdot w_a(t)\Bigg)_{a\in A}\Bigg)\Bigg]$$
$$< 0. \quad (27)$$

We explain the equalities (21) = (22), (23) = (24), (24) = (25), and (26) = (27) below.

a) Verifying (21) = (22)*:* From the equations leading up to (19), we have, for each $w \in \mathcal{W}$:

$$\dot{w}_a(t)$$
$$= h_a(w(t))$$
$$= -w_a(t) + \Bigg(g_{i_a} + \sum_{a'\in A_{i_a}^-} w_{a'}(t)\Bigg) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{a'\in A_{i_a}^+} \exp(-\beta z_{a'}(w(t)))} + \xi_a(t) \cdot \sum_{a'\in A_{i_a}^-} h_{a'}(w(t))$$
$$= -w_a(t) + \Bigg(g_{i_a} + \sum_{a'\in A_{i_a}^-} w_{a'}(t)\Bigg) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{a'\in A_{i_a}^+} \exp(-\beta z_{a'}(w(t)))} + \xi_a(t) \cdot \sum_{a'\in A_{i_a}^-} \dot{w}_{a'}(t).$$

Fix any node $i \in I$ in the Condensed DAG, and consider the sum of the above equation over the arc subset A_i^+:

$$\sum_{\hat{a}\in A_i^+} \dot{w}_{\hat{a}}(t) = -\sum_{\hat{a}\in A_i^+} w_{\hat{a}}(t) + \left(g_i + \sum_{a'\in A_{i_a}^-} w_{a'}(t)\right)\cdot 1 + 1\cdot \sum_{a'\in A_{i_a}^-} \dot{w}_{a'}(t).$$

Rearranging terms, we obtain:

$$\frac{d}{dt}\left(\sum_{\hat{a}\in A_i^+} w_{\hat{a}} - \sum_{a'\in A_{i_a}^-} w_{a'} - g_i\right) = -\left(\sum_{\hat{a}\in A_i^+} w_{\hat{a}} - \sum_{a'\in A_{i_a}^-} w_{a'} - g_i\right).$$

Since $w(0) \in \mathcal{W}$ by assumption, we have the initial condition $(\sum_{\hat{a}\in A_i^+} w_{\hat{a}} - \sum_{a'\in A_{i_a}^-} w_{a'} - g_i)(0) = 0$ for the above linear time-invariant differential equation. We thus conclude that, for each $t \geq 0$:

$$\sum_{\hat{a}\in A_i^+} w_{\hat{a}}(t) - \sum_{a'\in A_{i_a}^-} w_{a'}(t) - g_i = 0.$$

Since this holds for any arbitrary node $i \in I$, we have $w(t) \in \mathcal{W}$ for all $t \geq 0$.

b) Verifying (23) = (24)*:* We will show that:

$$\Pi_{\mathcal{W}_s}\left[\left(s_{[a]}(w_{[a]}(t))\right)_{a\in A} + \nabla\chi^\beta\left(\left(\left(g_{i_a} + \sum_{a'\in A_{i_a}^-} w_{a'}(t)\right) \cdot \frac{\exp(-\beta z_a(w(t)))}{\sum_{\bar{a}\in A_i^+}\exp(-\beta z_{\bar{a}}(w(t)))}\right)_{a\in A}\right)\right] = 0, \tag{28}$$

which would a fortiori establish the desired claim (23) = (24). To do so, first note that, for each $i \neq d$, $a \in A_i^+$:

$$\frac{\partial \chi^\beta}{\partial w_a}(w) = \frac{1}{\beta}\cdot\left[\ln w_a + 1 - \ln\left(\sum_{a\in A_i^+} w_a\right) - 1\right] \tag{29}$$
$$= \frac{1}{\beta}\ln\left(\frac{w_a}{\sum_{a\in A_i^+} w_a}\right).$$

Thus, we have:

$$\begin{aligned}
&\frac{\partial \chi^\beta}{\partial w_a}\left(\left(\left(g_{i_a} + \sum_{a'\in A_{i_a}^-} w_{a'}\right) \cdot \frac{\exp(-\beta z_a(w))}{\sum_{\bar{a}\in A_{i_a}^+}\exp(-\beta z_{\bar{a}}(w))}\right)_{a\in A}\right)\\
&= \frac{1}{\beta}\ln\left(\frac{\exp(-\beta z_a(w))}{\sum_{\bar{a}\in A_i^+}\exp(-\beta z_{\bar{a}}(w))}\right)\\
&= -z_a(w) - \frac{1}{\beta}\ln\left(\sum_{\bar{a}\in A_{i_a}^+}\exp(-\beta z_{\bar{a}}(w))\right)\\
&= -z_a(w) + \varphi_{i_a}(w).
\end{aligned}$$

Concatenating these partial derivatives to form the gradient, we can now verify (28) by observing that:

$$\begin{aligned}
&\Pi_{\mathcal{W}_s}\Bigg[\left(s_{[a]}(w_{[a]})\right)_{a\in A} + \nabla\chi^\beta\left(\left(\left(g_{i_a} + \sum_{a'\in A_{i_a}^-} w_{a'}\right) \cdot \frac{\exp(-\beta z_{\hat{a}}(w))}{\sum_{\bar{a}\in A_i^+}\exp(-\beta z_{\bar{a}}(w))}\right)_{\hat{a}\in A}\right)_{a\in A}\Bigg]\\
&= \Pi_{\mathcal{W}_s}\left(s_{[a]}(w_{[a]}) - z_a(w) + \varphi_{i_a}(w)\right)_{a\in A}\\
&= \Pi_{\mathcal{W}_s}\left(\varphi_{i_a}(w) - \varphi_{j_a}(w)\right)_{a\in A}\\
&= \Pi_{\mathcal{W}_s}\left[\sum_{a\in A}\varphi_{i_a}(w)e_a - \sum_{a\in A}\varphi_{j_a}(w)e_a\right]\\
&= \Pi_{\mathcal{W}_s}\Bigg[-\sum_{\hat{a}\in A_d^-}\varphi_{j_{\hat{a}}}(w)e_{\hat{a}} + \sum_{a'\in A_o^+}\varphi_{i_{a'}}(w)e_{a'} + \sum_{i\neq\{o,d\}}\left(\sum_{a'\in A_i^+}\varphi_i(w)e_{a'} - \sum_{\hat{a}\in A_i^-}\varphi_i(w)e_{\hat{a}}\right)\Bigg]\\
&= \Pi_{\mathcal{W}_s}\left[0 + \varphi_o(w)e_{A_o^+} + \sum_{i\neq\{o,d\}}\varphi_i(w)\left(e_{A_i^-} - e_{A_i^+}\right)\right]\\
&= 0,
\end{aligned}$$

where the last equality follows by definition of $\Pi_{\mathcal{W}_s}$.

c) Verifying (24) = (25)*:* We will show that:

$$\nabla\chi^\beta(w) = \nabla\chi^\beta\left(\left(\left(1 - \frac{\sum_{a'\in A_{i_a}^-}\dot{w}_{a'}}{\sum_{\hat{a}\in A_{i_a}^+} w_{\hat{a}}}\right)\cdot w_a\right)_{a\in A}\right),$$

which is equivalent to showing that (24) = (25). From (29), we have for each $a \in A$:

$$\begin{aligned}
&\frac{\partial \chi^\beta}{\partial w_a}\left(\left(\left(1 - \frac{\sum_{a'\in A_{i_a}^-} h_{a'}(w)}{\sum_{\hat{a}\in A_{i_a}^+} w_{\hat{a}}}\right)\cdot w_a\right)_{a\in A}\right)\\
&= \frac{1}{\beta}\ln\left(\frac{\left(1 - \frac{\sum_{a'\in A_{i_a}^-} h_{a'}(w)}{\sum_{\hat{a}\in A_{i_a}^+} w_{\hat{a}}}\right)w_a}{\sum_{\bar{a}\in A_{i_a}^+}\left(1 - \frac{\sum_{a'\in A_{i_{\bar{a}}}^-} h_{a'}(w)}{\sum_{\hat{a}\in A_{i_{\bar{a}}}^+} w_{\hat{a}}}\right)w_{\bar{a}}}\right)\\
&= \frac{1}{\beta}\ln\left(\frac{\left(1 - \frac{\sum_{a'\in A_{i_a}^-} h_{a'}(w)}{\sum_{\hat{a}\in A_{i_a}^+} w_{\hat{a}}}\right)w_a}{\left(1 - \frac{\sum_{a'\in A_{i_a}^-} h_{a'}(w)}{\sum_{\hat{a}\in A_{i_a}^+} w_{\hat{a}}}\right)\cdot\sum_{\bar{a}\in A_{i_a}^+} w_{\bar{a}}}\right)\\
&= \frac{1}{\beta}\ln\left(\frac{w_a}{\sum_{\bar{a}\in A_{i_a}^+} w_{\bar{a}}}\right)\\
&= \frac{\partial \chi^\beta}{\partial w_a}(w).
\end{aligned}$$

The second equality above follows because, for each $\bar{a} \in A_{i_a}^+$, we have $i_{\bar{a}} = i_a$. This verifies that (24) = (25). ■

d) Verifying (26) = (27)*:* Suppose $\frac{d}{dt}(F \circ w) = 0$ at some $\tilde{w} \in \mathcal{W}$. From (26), and by the definition of the convex function χ:

$$
\begin{aligned}
&0 \\
&= \frac{d}{dt}(F \circ w) \\
&= \sum_{i \in I \backslash \{d\}} \Bigg(\Bigg[-\left(1 - \frac{\sum_{\hat{a} \in A_{i_a}^-} h_{\hat{a}}(\tilde{w})}{\sum_{a' \in A_{i_a}^+} \tilde{w}_{a'}}\right) \tilde{w}_a \\
&+ \left(g_{i_a} + \sum_{a' \in A_{i_a}^+} \tilde{w}_{a'}\right) \cdot \frac{\exp(-\beta \cdot z_a(\tilde{w}))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta \cdot z_{a'}(\tilde{w}))} \Bigg]_{a \in A} \Bigg)^\top \\
&\frac{1}{\beta} \Bigg(\nabla \chi_i^\beta \Bigg(\Bigg[\left(1 - \frac{\sum_{\hat{a} \in A_{i_a}^-} h_{\hat{a}}(\tilde{w})}{\sum_{a' \in A_{i_a}^+} \tilde{w}_{a'}}\right) \tilde{w}_a \Bigg]_{a \in A} \Bigg) \\
&\quad - \nabla \chi_i^\beta \Bigg(\Bigg[\left(g_{i_a} + \sum_{a' \in A_{i_a}^+} \tilde{w}_{a'}\right) \\
&\qquad \cdot \frac{\exp(-\beta \cdot z_a(\tilde{w}))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta \cdot z_{a'}(\tilde{w}))} \Bigg]_{a \in A} \Bigg),
\end{aligned}
$$

where, for each $i \in I \backslash \{d\}$, the convex map $\chi_i^\beta : \mathbb{R}^{|A_i^+|} \to \mathbb{R}$ is defined by:

$$
\begin{aligned}
&\chi_i^\beta(\{w_a : a \in A_i^+\}) \\
&= \sum_{a \in A_i^+} w_a \ln w_a - \left(\sum_{a \in A_i^+} w_a\right) \ln \left(\sum_{a \in A_i^+} w_a\right)
\end{aligned}
$$

The convexity of each χ_i^β implies that each of the above summands must be non-positive; since they sum to 0, each summand must be 0.

Now, for each $i \in I \backslash \{d\}$ and each $w \in \mathbb{R}^{A_i^+}$, we have $N(\nabla^2 \chi_i^\beta(w)) = \text{span}\{w\}$. In words, χ_i^β increases linearly only along rays emanating from the origin. In the context of the above summands, this implies that, for each $i \in I \backslash \{d\}$, there exists constants $K_i \in \mathbb{R}$ such that, for each $a \in A_i^+$:

$$
\begin{aligned}
&K_i \left(1 - \frac{\sum_{\hat{a} \in A_{i_a}^-} h_{\hat{a}}(\tilde{w})}{\sum_{a' \in A_{i_a}^+} \tilde{w}_{a'}}\right) \tilde{w}_a \\
&= \left(g_{i_a} + \sum_{a' \in A_{i_a}^+} \tilde{w}_{a'}\right) \cdot \frac{\exp(-\beta \cdot z_a(\tilde{w}))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta \cdot z_{a'}(\tilde{w}))}.
\end{aligned}
$$

By definition of $h : \mathcal{W} \to \mathbb{R}^{|A|}$, for each $a \in A_o^+$:

$$
\begin{aligned}
h_a(\tilde{w}) &= -\tilde{w}_a + g_o \cdot \frac{\exp(-\beta z_a(\tilde{w}))}{\sum_{a' \in A_o^+} \exp(-\beta z_{a'}(\tilde{w}))} \\
&= (K_o - 1)\tilde{w}_a
\end{aligned}
$$

and for each $a \in A_i^+$ with $i \neq o$:

$$
\begin{aligned}
h_a(w) &:= -\left(1 - \frac{\sum_{\hat{a} \in A_{i_a}^-} h_{\hat{a}}(\tilde{w})}{\sum_{a' \in A_{i_a}^+} \tilde{w}_{a'}}\right) \tilde{w}_a \\
&\quad + \sum_{a' \in A_{i_a}^-} \tilde{w}_{a'} \cdot \frac{\exp(-\beta z_a(\tilde{w}))}{\sum_{a' \in A_o^+} \exp(-\beta z_{a'}(\tilde{w}))} \\
&= (K_o - 1)\left(1 - \frac{\sum_{\hat{a} \in A_{i_a}^-} h_{\hat{a}}(\tilde{w})}{\sum_{a' \in A_{i_a}^+} \tilde{w}_{a'}}\right) \tilde{w}_a.
\end{aligned}
$$

By the flow continuity equations:

$$
0 = \sum_{a' \in A_o^+} h_{a'}(\tilde{w}) = (K_o - 1) \cdot \sum_{a' \in A_i^+} \tilde{w}_{a'} = (K_o - 1) g_o,
$$

so $K_o = 1$, and thus $h_a(\tilde{w}) = 0$ for each $a \in A_o^+$. Now, suppose there exists some $m \in [m(G)-1]$ such that $h_a(\tilde{w}) = 0$ for each $a \in A$ such that $m_a \leq m$. Then, for each $a \in A$ such that $m_a = m+1$, the flow continuity equations give:

$$
\begin{aligned}
0 &= \sum_{a' \in A_i^+} h_{a'}(\tilde{w}) - \sum_{\hat{a} \in A_i^-} h_{\hat{a}}(\tilde{w}) \\
&= \sum_{a' \in A_i^+} h_{a'}(\tilde{w}) \\
&= (K_{i_a} - 1) \cdot \sum_{a' \in A_{i_a}^+} \tilde{w}_{a'}.
\end{aligned}
$$

Thus, $K_{i_a} = 1$, so $h_a(\hat{w}) = 0$. This completes the recursion step, and shows that $h(\hat{w}) = 0$, i.e., $\hat{w} = \bar{w}^\beta$.

In summary, we established that the map F strictly decreases along any trajectory that starts in $\mathcal{W} \backslash \{\bar{w}^\beta\}$ and follows the best-response dynamics (20). The convergence of the dynamics (20) to the Condensed DAG equilibrium (3) now follows by invoking either Sandholm, Corollary 7.B.6 [23], or Sastry, Proposition 5.22 and Theorem 5.23 (LaSalle's Principle and its corollaries) [28].

2) Proof of Lemma 3: To prove Lemma 3, we require the following results. We first establish bounds on the trajectory of discrete-time and continuous-time traffic flow dynamics.

Lemma 4:

1) Consider the discrete-time dynamics:

$$
Y[n+1] = (1 - \delta[n])Y[n] + \delta[n]F[n],
$$

where, for each $n \geq 0$, we have $\delta[n] \in (0,1)$ and $Y[0], F[n] \in [a,b]$ for some $a, b \in \mathbb{R}$ satisfying $a < b$. Then $Y[n] \in [a,b]$ for each $n \geq 0$.

2) Consider the continuous-time dynamics:

$$
\dot{y}(t) = -y(t) + f(t),
$$

where, for each $t \geq 0$, we have $y(0), f(t) \in [a,b]$ for some $a, b \in \mathbb{R}$ satisfying $a < b$. Then $y(t) \in [a,b]$ for each $t \geq 0$.

Proof:

1) Suppose there exists some $N \geq 0$ such that $Y[n] \in [a,b]$ for each $n \leq N$. (Since $Y[0] \in [a,b]$ by assumption, this is certainly true for $n = 0$). Then:

$$
\begin{aligned}
Y[n+1] &= (1 - \delta[n])Y[n] + \delta[n]F[n] \\
&\in [(1 - \delta[n]) \cdot a + \delta[n] \cdot a, (1 - \delta[n]) \cdot b \\
&\qquad + \delta[n] \cdot b] \\
&= [a,b].
\end{aligned}
$$

This completes the induction step, and thus the proof of this part of the lemma.

2) We compute:

$$\frac{d}{dt}(e^t y(t)) = e^t\big(\dot{y}(t) + y(t)\big) = e^t f(t).$$

Integrating from time 0 to time t, we have, for each $t \geq 0$:

$$e^t y(t) - y(0) = \int_0^t e^\tau f(\tau)\, d\tau.$$

Rearranging terms, we obtain, for each $t \geq 0$:

$$\begin{aligned} y(t) &= e^{-t}y(0) + e^{-t}\int_0^t e^\tau f(\tau)\, d\tau \\ &\in \big[e^{-t}a + (1-e^{-t})a, e^{-t}b + (1-e^{-t})b\big] \\ &= [a,b]. \end{aligned}$$

■

Before proceeding, we rewrite the discrete ξ-dynamics (PBR) as a Markov process with a martingale difference term:

$$\begin{aligned} &\xi_a[n+1] \\ &= \xi_a[n] + \mu\Bigg(- \xi_a[n] + \frac{\exp(-\beta\big[z_a(W[n])\big])}{\sum\limits_{a' \in A_{i_a}^+} \exp(-\beta\big[z_{a'}(W[n])\big])} \\ &\qquad + M_a[n+1]\Bigg), \end{aligned}$$

with:

$$\begin{aligned} &M_a[n+1] \\ &:= \left(\frac{1}{\mu}\eta_{i_a}[n+1] - 1\right) \\ &\quad\cdot\left(- \xi_a[n] + \frac{\exp(-\beta\big[z_a(W[n])\big])}{\sum_{a' \in A_{i_a}^+} \exp(-\beta\big[z_{a'}(W[n])\big])}\right). \end{aligned}$$

The following lemma bounds the magnitude of the discrete-time flow $W[n] \in \mathbb{R}^{|w|}$ and the martingale difference terms $M[n] \in \mathbb{R}^{|w|}$.

Lemma 5: The discrete-time dynamics (PBR) and (12) satisfy:

1) For each $a \in A$: $\{M_a[n+1] : n \geq 0\}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_n := \sigma\big(\cup_{a\in A}(W_a[1], \xi[1], \cdots, W_a[n], \xi[n])\big)$.
2) There exist $C_w, C_m > 0$ such that, for each $a \in A$ and each $n \geq 0$, we have:

$$\begin{aligned} W_a[n] &\in [C_w, g_o], \\ |M_a[n]| &\leq C_m. \end{aligned}$$

The continuous-time dynamics (13) and (14) satisfy:

3) For each $a \in A$, $n \geq 0$:

$$w_a(t) \in [C_w, g_o].$$

Proof:

1) We have:

$$\begin{aligned} &\mathbb{E}[M_a[n+1]|\mathcal{F}_n] \\ &= \left(\frac{1}{\mu}\mathbb{E}[\eta_{i_a}[n+1]] - 1\right) \\ &\quad\cdot\left(-\xi_a[n] + \frac{\exp(-\beta\big[z_a(W[n])\big])}{\sum_{a' \in A_{i_a}^+} \exp(-\beta\big[z_{a'}(W[n])\big])}\right) \\ &= 0. \end{aligned}$$

2) We separate the proof of this part of the lemma into the following steps.
   - First, we show that for each $a \in A$, $n \geq 0$, we have $\xi_a[n] \in (0,1]$.

     Fix $a \in A$ arbitrarily. Then $\xi_a[0] \in (0,1]$ by assumption, and for each $n \geq 0$:

     $$\frac{\exp(-\beta\big[z_a(W[n])\big])}{\sum_{a' \in A_{i_a}^+} \exp(-\beta\big[z_{a'}(W[n])\big])} \in (0,1],$$

     since the exponential function takes values in $(0,\infty)$. Thus, by Lemma 4, we have $\xi_a[n] \in (0,1]$ for each $n \geq 0$.
   - Second, we show that for each $a \in A$, $n \geq 0$, we have $W_a[n] \in (0, g_o]$.

     Note that (12), together with the assumption that $W[0] \in \mathcal{W}$, implies that $W[n] \in \mathcal{W}$ for each $n \geq 0$. Now, fix $a \in A$, $n \geq 0$ arbitrarily. Let $\mathbf{R}(a) \subseteq \mathbf{R}$ denote the set of all routes passing through a, and for each $r \in \mathbf{R}(a)$, let $a_{r,k}$ denote the k-th arc in r. Then, by the conservation of flow encoded in $\mathbf{R}$:

     $$\begin{aligned} W_a[n] &= g_o \cdot \sum_{r \in \mathbf{R}(a)} \prod_{k=1}^{|r|} \xi_{a_{r,k}} \\ &\leq g_o \cdot \sum_{r \in \mathbf{R}} \prod_{k=1}^{|r|} \xi_{a_{r,k}} \\ &= g_o. \end{aligned}$$

     Similarly, since $\xi_a[n] \in (0,1]$ for each $a \in A$, $n \geq 0$, we have:

     $$W_a[n] = g_o \cdot \sum_{r \in \mathbf{R}(a)} \prod_{k=1}^{|r|} \xi_{a_{r,k}} > 0.$$
   - Third, we show that there exists $C_z > 0$ such that $|z_a(W[n])| \leq C_z$ for each $a \in A$, $n \geq 0$. Fix $a \in A_d^- = \{a \in A : m_\alpha = 1\}$ arbitrarily. Then, from (6):

     $$\begin{aligned} z_a(w) &= s_{[a]}(w_{[a]}) \in [0, s_{[a]}(g_o)], \\ \Rightarrow |z_a(w)| &\leq s_{[a]}(g_o) := C_{z,1}. \end{aligned}$$

     Now, suppose that at some height $k \in [m(G)-1]$, there exists some $C_{z,k} > 0$ such that, for each $n \geq 0$, and each $a \in A$ satisfying $m_a \leq k$ and each $n \geq 0$, we have $|z_a(w)| \leq C_{z,k}$. Then, for each $n \geq 0$, and each $a \in A$ satisfying $m_a = k+1$ (at least one such $a \in A$ must exist, by Proposition 2, Part 4):

     $$z_a(w) = s_{[a]}(w_{[a]}) - \frac{1}{\beta}\ln\left(\sum_{a' \in A_{j_a}^+} e^{-\beta \cdot z_{a'}(w)}\right)$$

$$\leq s_{[a]}(g_o) - \frac{1}{\beta}\ln\left(|A_{j_a}^+|e^{-\beta\cdot C_z}\right)$$
$$= s_{[a]}(g_o) + C_z,$$

and:

$$z_a(w) = s_{[a]}(w_{[a]}) - \frac{1}{\beta}\ln\left(\sum_{a'\in A_{j_a}^+} e^{-\beta\cdot z_{a'}(w)}\right)$$
$$\geq 0 + 0 - \frac{1}{\beta}\ln\left(|A_{j_a}^+|e^{\beta\cdot C_z}\right)$$
$$= -\frac{1}{\beta}\ln|A| - C_z,$$

from which we conclude that:

$$|z_a(w)| \leq \max\left\{s_{[a]}(g_o) + C_z, \frac{1}{\beta}\ln|A| + C_z\right\}$$
$$:= C_{z,k+1},$$

with $C_{z+1} \geq C_z$. This completes the induction step, and the proof is completed by taking $C_z := C_{z,m(G)}$.

- Fourth, we show that there exists some $C_\xi > 0$ such that $\xi_a[n] \geq C_\xi$ for each $a \in A$, $n \geq 0$.

  Define:

$$C_\xi := \min\left\{\min_{a'\in A}\xi_{a'}[0], \frac{1}{|A|}e^{-2\beta C_z}\right\} \in (0,1).$$

  By definition of C_ξ, we have $\xi_a[0] \geq C_\xi$. Moreover, for each $n \geq 0$, we have:

$$\frac{\exp(-\beta\big[z_a(W[n])\big])}{\sum_{a'\in A_{i_a}^+}\exp(-\beta\big[z_{a'}(W[n])\big])}$$
$$\geq \frac{e^{-\beta C_z}}{|A_{i_a}^+|\cdot e^{\beta C_z}}$$
$$\geq \frac{1}{|A|}e^{-2\beta C_z}$$
$$\geq C_\xi.$$

  Thus, by Lemma 4, $\xi_a[n] \geq C_\xi$ for each $n \geq 0$.

- Fifth, we show that there exists $C_w > 0$ such that, for each $a \in A$, $n \geq 0$, we have $W_a[n] \geq C_w$.

  Fix $a \in A$, $n \geq 0$. Let $\mathbf{R}(a)$ denote the set of all routes in the Condensed DAG containing a, and let $r \in \mathbf{R}(a)$ be arbitrarily given. By unwinding the recursive definition of $W_a[n]$ from the flow dispersion probability values $\{\xi_a[n] : a \in A, n \geq 0\}$, we have:

$$W_a[n] = g_o \cdot \sum_{\substack{r'\in\mathbf{R}\\ a\in r'}}\prod_{a'\in r'}\xi_{a'}[n]$$
$$\geq g_o \cdot \prod_{a'\in r}\xi_{a'}[n]$$
$$\geq g_o \cdot (C_\xi)^{|r|}$$
$$\geq g_o \cdot (C_\xi)^{\ell(G)}$$
$$:= C_w.$$

- Sixth, we show that there exists $C_m > 0$ such that, for each $a \in A$, $n \geq 0$, we have $M_a[n] \geq C_m$.

  Define, for convenience, $C_\mu := \max\{\overline{\mu} - \mu, \mu - \underline{\mu}\}$. Since $\eta_{i_a}[n] \in [\underline{\mu}, \overline{\mu}]$, we have from (16) that for each $a \in A$, $n \geq 0$:

$$M_a[n+1]$$
$$= \left(\frac{1}{\mu}\eta_{i_a}[n+1] - 1\right)\cdot\Bigg(-\xi_a[n] + \frac{\exp(-\beta\big[z_a(W[n])\big])}{\sum_{a'\in A_{i_a}^+}\exp(-\beta\big[z_{a'}(W[n])\big])}\Bigg).$$

  Thus, by the triangle inequality:

$$|M_a[n+1]| \leq \frac{1}{\mu}C_\mu\cdot(1+1) = \frac{2}{\mu}C_\mu := C_m.$$

3) We separate the proof of this part of the lemma into the following steps.
   - First, we show that for each $a \in A$, $t \geq 0$, we have $\xi_a(t) \in (0,1]$.

     Fix $a \in A$. By assumption, $\xi_a(0) \in (0,1]$, and at each $t \geq 0$:

$$\frac{\exp(-\beta z_a(w))}{\sum_{a'\in A_{i_a}^+}\exp(-\beta z_{a'}(w))} \in (0,1].$$

     Thus, by Lemma 4, we conclude that $\xi_a(t) \in (0,1]$ for each $t \geq 0$.
   - Second, we show that $w_a(t) \in [0, g_o]$ for each $t \geq 0$.

     The proof here is nearly identical to the proof that $W_a[n] \in (0, g_o)$ in the second bullet point of the second part of this lemma, and is omitted for brevity.
   - Third, we show that $|z_a(w_a(t))| \leq C_z$ for each $t \geq 0$.

     The proof here is nearly identical to the proof that $|z_a(W_a[n])| \leq C_z$ in the fourth bullet point of the second part of this lemma, and is omitted for brevity.
   - Fourth, we show that there exists some $C_\xi > 0$ such that $\xi_a(t) \geq C_\xi$ for each $a \in A$, $t \geq 0$.

     Define:

$$C_\xi := \min\left\{\min\{\xi_{a'}(0) : a' \in A\}, \frac{1}{|A|}e^{-2\beta C_z}\right\}$$
$$\in (0,1).$$

     By definition of C_ξ, we have $\xi_a(0) \geq C_\xi$. Moreover, for each $n \geq 0$, we have:

$$\frac{\exp(-\beta\big[z_a(W[n])\big])}{\sum_{a'\in A_{i_a}^+}\exp(-\beta\big[z_{a'}(W[n])\big])}$$
$$\geq \frac{e^{-\beta C_z}}{|A_{i_a}^+|\cdot e^{\beta C_z}}$$
$$\geq \frac{1}{|A|}e^{-2\beta C_z}$$
$$\geq C_\xi.$$

     Thus, by Lemma 4, we have $\xi_a(t) \geq C_\xi$ for each $t \geq 0$.
   - Fifth, we show that there exists $C_w > 0$ such that, for each $a \in A$, $t \geq 0$, we have $w_a(t) \geq C_w$.

The proof here is nearly identical to the proof that $W_a[n] \geq C_w$ in the fourth bullet point of the second part of this lemma, and is omitted for brevity.

■

The lemma below establishes the final part of Lemma 3. Below, we restrict the domains of the maps $\bar{\xi}^\beta$ and z_a to reflect the bounds of the traffic flow trajectory w established in the above lemma, i.e., $\bar{\xi}^\beta, z_a : \mathcal{W}' \to \mathbb{R}$, with the flow restricted to:

$$\mathcal{W}' := \mathcal{W} \cap [C_w, g_o]^{|A|}$$

and the toll restricted to $[0, C_p]^{|A_O|}$.

Lemma 6: The continuous-time dynamics (20) satisfies:

1) For each $a \in A$, the restriction of the latency-to-go map $z_a : \mathcal{W} \to \mathbb{R}^{|A_O|} \to \mathbb{R}$ to $\mathcal{W}'$ is Lipschitz continuous.
2) The map from the probability transitions $\xi \in \prod_{i \in I \setminus \{d\}} \Delta(A_i^+)$ and the traffic flows $w \in \mathcal{W}$ is Lipschitz continuous.
3) For each $a \in A$, the restriction of the continuous dynamics transition map $\rho_a : \mathbb{R}^{|A|} \times \mathbb{R}^{|A_O|} \to \mathbb{R}^{|A|}$, defined recursively by:

$$\rho_a(\xi) := -\xi_a + \frac{\exp(-\beta z_a(w))}{\sum_{a' \in A_{i_a}^+} \exp(-\beta z_{a'}(w))}$$

to $\mathcal{W}'$ is Lipschitz continuous.

Proof:

1) We shall establish the Lipschitz continuity of z_a, for each $a \in A$, by providing uniform bounds on its partial derivatives across all values of its arguments $w \in \mathcal{W}'$.

The proof follows by induction on the height index $k \in [m(G)]$. For each $a \in A$, let $\tilde{z}_a : \mathbb{R}^{|A|} \to \mathbb{R}$ be the continuous extension of $z_a : \mathcal{W} \to \mathbb{R}$ to the Euclidean space $\mathbb{R}^{|A|}$ containing $\mathcal{W}$. By definition of Lipschitz continuity, if $\tilde{z}_a$ is Lipschitz for some $a \in A$, then so is z_a. For each $a \in A_d^- = \{a \in A : m_a = 1\}$ and any $w \in \mathbb{R}^{|A|}$:

$$\tilde{z}_a(w) = s_{[a]}(w_{[a]}).$$

Thus, for any $\hat{a} \in A$, and any $w \in \mathbb{R}^{|A|}$, $p \in \mathbb{R}^{|A_O|}$:

$$\frac{\partial \tilde{z}_a}{\partial w_{\hat{a}}}(w) = \frac{ds_{[a]}}{dw}(w_{[a]}) \cdot \mathbf{1}\{\hat{a} \in [a]\} \in [0, C_{ds}].$$

We set $C_{z,1} := C_{ds}$.

Now, suppose that there exists some depth $k \in [m(G) - 1]$ and some constant $C_{z,k} > 0$ such that, for any $a \in A$ satisfying $m_a \leq k$, and any $w \in \mathcal{W}$, $n \geq 0$, the map $\tilde{z}_a : \mathbb{R}^{|A|} \to \mathbb{R}$ is continuously differentiable, with:

$$\left| \frac{\partial \tilde{z}_a}{\partial w_{\hat{a}}}(w) \right| \leq C_{z,k}.$$

Continuing with the induction step, fix $a \in A$ such that $m_a = k + 1$ (there exists at least one such link, by Proposition 1, Part 4). From Proposition 1, Part 2, we have $m_{a'} \leq k$ for each $a' \in A_{i_a}^+$. Thus, the induction hypothesis implies that, for any $\hat{a} \in A$:

$$\tilde{z}_a(w) = s_{[a]}(w_{[a]}) - \frac{1}{\beta} \sum_{a' \in A_{j_a}^+} e^{-\beta z_{a'}(w)}.$$

Computing partial derivatives with respect to each component of w, we obtain:

$$\frac{\partial \tilde{z}_a}{\partial w_{\hat{a}}}(w) = \frac{ds_{[a]}}{dw}(w_{[a]}) \cdot \mathbf{1}\{\hat{a} \in [a]\} + \sum_{a' \in A_{j_a}^+} e^{-\beta \tilde{z}_{a'}(w)} \cdot \frac{\partial \tilde{z}_{a'}}{\partial w_{\hat{a}}}(w),$$

$$\Rightarrow \left| \frac{\partial \tilde{z}_a}{\partial w_{\hat{a}}}(w) \right| \leq C_{ds} + |A| \cdot C_{z,k}.$$

We can complete the induction step by taking $C_{z,k+1} := C_{ds} + |A| \cdot C_{z,k}$.

This establishes that, for each $a \in A$, the map z_a is continuously differentiable, with partial derivatives uniformly bounded by a uniform constant, $C_z := C_{z,m(G)}$. This establishes the Lipschitz continuity of the map z_a for each $a \in A$, and thus proves this part of the proposition.

2) Recall that the map from traffic distributions probabilities (ξ) to traffic flows (w) is given as follows, for each $a \in A$. Recall that $\mathbf{R}(a)$ denotes the set of all routes in the Condensed DAG that contain the arc a:

$$w_a = \left(g_{i_a} + \sum_{\hat{a} \in A_i^-} w_a \right) \cdot \xi_a = g_o \cdot \sum_{r \in \mathbf{R}(a)} \prod_{k=1}^{|r|} \xi_{a_{r,k}},$$

where $a_{r,k}$ denotes the k-th arc along a given route $r \in \mathbf{R}$, for each $k \in |r|$. Thus, the map from ξ to w is continuously differentiable. Moreover, the domain of this map is compact; indeed, for each $a \in A$, we have $\xi_a \in [0, 1]$, and for each non-destination node $i \neq d$, we have $\sum_{a \in A_i^+} \xi_a = 1$. Therefore, the map $\xi \mapsto w$ has continuously differentiable derivatives with magnitude bounded above by some constant uniform in the compact set of realizable probability distributions ξ. This is equivalent to stating that the map $\xi \mapsto w$ is Lipschitz continuous.

3) Above, we have established that the maps z_a and $\xi \mapsto w$ are Lipschitz continuous. Since the addition and composition of Lipschitz maps is Lipschitz, it suffices to verify that the map $\hat{\rho} : \mathbb{R}^{|A|} \to \mathbb{R}^{|A|}$, defined elementwise by:

$$\hat{\rho}_a(z) := \frac{e^{-\beta z_a}}{\sum_{a' \in A_{i_a}^+} e^{-\beta z_{a'}}}, \qquad \forall\, a \in A$$

is Lipschitz continuous. We do so below by computing, and establishing a uniform bound for, its partial derivatives. For each $\hat{a} \in A$:

$$\frac{\partial \hat{\rho}_a}{\partial z_{\bar{a}}} = \frac{1}{(\sum_{a' \in A_{i_a}^+} e^{-\beta z_{a'}})^2}$$

$$
\begin{aligned}
&\cdot \Bigg(\sum_{a' \in A_{i_a}^+} e^{-\beta z_{a'}} \cdot (-\beta) e^{-\beta z_a} \cdot \frac{\partial z_a}{\partial z_{\bar{a}}} \\
&\qquad - e^{-\beta z_a} \cdot \sum_{a' \in A_{i_a}^+} (-\beta) e^{-\beta z_{a'}} \frac{\partial z_{a'}}{\partial z_{\bar{a}}} \Bigg) \\
&= -\frac{e^{-\beta z_a}}{\sum_{a' \in A_{i_a}^+} e^{-\beta z_{a'}}} \cdot \beta \cdot \frac{\partial z_a}{\partial z_{\hat{a}}} \\
&\quad + \frac{\beta e^{-\beta z_a}}{(\sum_{a' \in A_{i_a}^+} e^{-\beta z_{a'}})^2} \cdot \sum_{a' \in A_{i_a}^+} e^{-\beta z_{a'}} \frac{\partial z_{a'}}{\partial z_{\bar{a}}}.
\end{aligned}
$$

Observe that:

$$
\begin{aligned}
\sum_{a' \in A_{i_a}^+} e^{-\beta z_{a'}} \frac{\partial z_{a'}}{\partial z_{\bar{a}}} &= \sum_{a' \in A_{i_a}^+} e^{-\beta z_{a'}} \cdot \mathbf{1}\{a' = \hat{a}\} \\
&\leq \max_{a' \in A_{i_a}^+} e^{-\beta z_{a'}}.
\end{aligned}
$$

This, together with triangle inequality, then gives:

$$
\left| \frac{\partial \hat{\rho}_a}{\partial z_{\bar{a}}} \right| = \beta + \beta = 2\beta.
$$

This concludes the proof for this part of the proposition.

■

We present the proof of Theorem 2, restated as follows: For any $\delta > 0$:

$$
\begin{aligned}
\limsup_{n \to \infty} \mathbb{E}\big[\|\xi[n] - \bar{\xi}^\beta\|_2^2\big] \leq O(\mu), \\
\limsup_{n \to \infty} \mathbb{P}\big(\|\xi[n] - \bar{\xi}^\beta\|_2^2 \geq \delta\big) \leq O\left(\frac{\mu}{\delta}\right).
\end{aligned}
$$

3) Proof of Theorem 2: Here, we conclude the proof of Theorem 2.

Proof: (**Proof of Theorem 2**) Lemma 5 asserts that $M[n]$ is bounded (uniformly in $n \geq 0$), while Lemma 6 establishes that $\rho : \mathbb{R}^{|A|} \to \mathbb{R}$ is Lipschitz continuous. The proof of Theorem 2 now follows by applying the stochastic approximation results in Borkar [25], Chapters 2 and 9. ■