Commuting and internet traffic congestion

Berliant, Marcus

Washington University in St. Louis

24 May 2021

Commuting and Internet Traffic Congestion*

Marcus Berliant[†‡]

May 2021

Abstract

We examine the fine microstructure of commuting in a game-theoretic setting with a continuum of commuters. Commuters' home and work locations can be heterogeneous. A commuter transport network is exogenous. Traffic speed is determined by link capacity and by local congestion at a time and place along a link, where local congestion at a time and place is endogenous. The model can be reinterpreted to apply to congestion on the internet. We find sufficient conditions for existence of equilibrium, that multiple equilibria are ubiquitous, and that the welfare properties of morning and evening commute equilibria differ on a generalization of a directed tree. JEL numbers: L86, R41 Keywords: Commuting; Internet traffic; Congestion externality; Efficient Nash equilibrium

*The author is grateful to David Boyce, whose address at the Regional Science Association International North American Meetings in Savannah incited the author (who was also his discussant) to write this paper. He is grateful to the Helen R. Whiteley Center for its kind hospitality. Alex Anas, Richard Arnott, Gilles Duranton, Jan Eeckhout, Ignacio Esponda, Mogens Fosgerau, Jonathan Hall, Eren Inci, Kamhon Kan, T. John Kim, Lewis Kornhauser, Alessandro Lizzeri, Bill Neilson, Kirill Pogorelskiy, Tony Smith, Taka Tabuchi, Matt Turner, Erik Verhoef and Ping Wang contributed interesting comments. The author is grateful to seminar audiences at the Public Economic Theory meetings in Galway, the Summer Meetings of the Econometric Society in Tokyo, the Kuhmo-Nectar Conference in Stockholm, the RSAI meetings in Miami, the Institute of Economic Research at Kyoto University, the University of Tokyo, Academia Sinica, Washington University in St. Louis, UC - Davis, the Seventh Kansas Workshop on Economic Theory, and the CARE - Colloquium for comments. Washington University generously provided sabbatical support. The author remains responsible for any opinions or errors herein. This work is dedicated to the memory of Bill Sandholm.

†Department of Economics, Washington University, Campus Box 1208, 1 Brookings Drive, St. Louis, MO 63130-4899 USA. Fax: (314) 935-4156, e-mail: berliant@wustl.edu

‡Institute of Economic Research, Kyoto University and California Institute of Technology.

1 Introduction

1.1 Motivation

Commuting is a ubiquitous feature of the urban economy. Although the classic literature has answered some basic questions in the field, such as whether equilibrium commuting patterns are generally efficient, surprisingly some very important questions remain open. Can traffic be improved simply by equilibrium selection rather than through congestion pricing? In contrast with most of the literature, our model says that multiple equilibria are to be expected, so this question has content. Do models without an explicit time clock give us an accurate picture of traffic, in the sense that they can approximate behavior in a truly dynamic model? Schrank et al (2019, p. 7, Exhibit 6) give evidence that traffic delays vary greatly by time, whereas Malone et al (2017, Figure 7(a)) give analogous evidence for the internet. Finally, if travel delay depends on endogenous local congestion rather than exogenous bottlenecks, what does equilibrium look like?

There is an important application of our model to traffic and congestion on the internet. Instead of cars, packets of information move over the network, each with a given origin and destination. Both positive and normative questions concerning route choice and departure time can be addressed with our model.[1] Interestingly, both the car and internet congestion literatures began with discrete models (at different times), and eventually moved to continuous flow models for tractability reasons. Congestion on the internet can be viewed as either packet loss or delay.

The economic models employed in the commuting literature are often very special and unrealistic; a literature review will be provided in the next subsection. One class of models features identical commuters, a very simple network structure (for example a home, a workplace and one link between them), and an exogenous bottleneck that results in queuing of traffic. It is not known to what degree the results derived in the literature rely on these or other strong simplifying assumptions that generally provide a reduced form viewpoint. In contrast, we study a new class of more natural models that allows arbitrary heterogeneity in both commuters and network structure (for example allowing cross-commuting), where congestion is endogenous and traffic slows in response

[1] In general, one user will send out many packets. However, if these represent a negligible proportion of the total number of packets, coordination of the strategy choices for these packets is the same as no coordination for our purposes.

to congestion relative to road capacity. In the last subsection of the introduction, we will provide simple examples that display the contrast between the existing literature and our class of models.

There are important differences in implications between our framework and the existing literature, mainly due to the detailing of fine microstructure in our work. What we mean by fine microstructure is not only a game where both route and departure time are strategic choices of players, but also that time-dependent events, such as cars catching up with others and slowing down, can happen in the course of traversing a link. The reduced forms, such as an exogenous congestion function (that gives delay time as a function of traffic), used elsewhere are generally not supported by this microstructure, leading to different results. Our model employs a microfounded, evolving congestion concept that is suggested by the transportation engineering literature. Thus, the conditions sufficient for existence of equilibrium are markedly different. As we shall see in the examples, it is quite natural to have multiple equilibria in our framework, whereas the goal of the existing literature is often to prove that equilibrium is unique. Finally, as we shall illustrate, equilibria in our framework are qualitatively very different from those derived in the rest of the literature, mainly due to the fine microstructure.

1.2 Five Related Literatures

Before proceeding to our examples and analysis, we discuss the basic literature on congestion. We divide this literature into 5 components: the transportation economics literature, the game-theoretic literature on congestion externalities, the transportation engineering literature, the mathematics of conservation laws, and the electrical engineering literature on internet congestion.[2] We discuss these in turn. Our work is at the junction of all of these literatures. In contrast with our work, the first two literatures tend not to study dynamic micro behavior along roads. The second two literatures take individual behavior as fixed, so the models are mechanical. The last literature tends not to examine Nash equilibrium, but rather other positive or normative ideas.

The older literature on transportation economics deals with models with no time clock or with just one route or bottleneck where traffic queues. Beckmann et al. (1956) provide a model of rush hour where traffic flows are constant. They analyze optimum and equilibrium in a stylized model with no explicit

[2] These literatures tend not to cite each other, rendering literature reviews labor-intensive and occasionally puzzling, due to terminological differences.

time clock, but with a representative commuter. Vickrey (1963, 1969)[3] provided the classical analysis of congestion externalities, pricing, and infrastructure investment. The basic economic problem detailed is that the choice to commute, of its timing, and of its route by one commuter affects the commuting time of others. Although the marginal time cost of one additional commuter on another is small, when the marginal time cost of an additional commuter is aggregated across all commuters, the cost of (and optimal toll for) the externality can be large. Arnott et al. (1993) examine primarily welfare under various pricing schemes when there is only one route or bottleneck, but allow elastic trip demand and use continuous time. Traffic does not slow down due to congestion, but rather queues at a bottleneck with limited capacity. In their conclusions (p. 177), they note: "In the context of rush hour traffic congestion, for example, models should be developed which derive hypercongestion (traffic-jam situations) from driving behavior, solve for equilibrium on a congested network, and account for heterogeneity among users..." This is what we attempt.

The contemporary literature on transportation economics uses the terminology "dynamic traffic assignment problem" for the kind of model we shall construct. Merchant and Nemhauser (1978) initiate the modern literature by proposing a discrete time model with a single destination node where events in a link of the transport network at a given time, namely the number of cars entering the link, the cost of traversing the link as a function of traffic, and the number of cars exiting a link as a function of traffic, are all exogenous black boxes. They provide an example and examine algorithms for finding a social optimum. Ross and Yinger (2000) embed a model of point congestion similar to ours in a classic urban monocentric city model with both land consumption and a symmetric radial road network. This is similar to a simple network with only one commuting corridor. Traffic flow is continuous but not necessarily smooth. They show that the only equilibrium in a general urban equilibrium version of a commuting model with continuous departure times and flow congestion but no bottlenecks is an unreasonable one with a never ending rush hour. As we shall explain below, by allowing a large but finite number of departure times and randomizing departures over small intervals between these discrete departure times, with some effort we can overcome these difficulties. In our context, traffic flow might not be continuous. Konishi (2004) considers existence, uniqueness and efficiency of Nash equilibrium primarily in a static

[3] In the first of these papers, the automobile is called "our rubber-tired sacred cow."

model but also in a discrete time dynamic model with a simple network, employing Schmeidler's (1973) theorem[4] as we do. He uses bottlenecks whereas we use speed reductions resulting from congestion. Konishi's work is quite complementary to ours, as we are not concerned with the issues he addresses, namely existence of equilibrium in static models with a finite number of commuters, conditions sufficient for uniqueness of equilibrium in static models with a continuum of commuters, and existence and uniqueness of equilibrium in dynamic models of simple networks with exogenous bottlenecks.

An independent, modern literature in transportation economics examines necessary conditions at a Nash equilibrium for the dynamic traffic assignment problem. Heydecker and Addison (2005) consider what happens along a link as a black box, and derive such a condition. Of course, if such a black box is made more specific, the necessary condition can be refined. Zhang and Zhang (2010) use a bottleneck model and obtain a more specialized condition.

In their survey, de Palma and Fosgerau (2011, p. 208) conclude: "The extension of the dynamic model to large networks remains a difficult problem. So far, existence and uniqueness of equilibrium have not been established (in spite of many attempts)."

The game-theoretic literature on externalities, for example Sandholm (2001), has the potential to be useful in our context. However, the strong symmetry assumptions used, that yield strong and interesting conclusions, exclude almost all of the games of interest to us. For example, they exclude the simple special case of our model where there are two nodes called home and work with one link between them, but two departure times. Hofbauer and Sandholm (2007) study congestion games with a continuum of players, but their assumptions on congestion rule out the type of dynamic micro-interaction along a link that is the focus of our work. Sandholm (2007) considers an evolutionary approach to setting optimal tolls in the case where there is a *finite* number of *identical* commuters (so they have the same home and work locations) modified by an idiosyncratic preference component, without the symmetry assumption but with further structure on the evolutionary process.[5] Hu (2010) explores Nash equilibrium with continuous departures for a single commuting corridor

[4] To apply Schmeidler's work to obtain Nash equilibrium in pure strategies, it is important that the set of pure strategies be finite. In our model, the interpretation is that the set of departure time strategies is finite.

[5] It is also interesting to inquire how tolls would be implemented in practice in these models, since in theory the toll is based on the overall strategy chosen, namely the route and/or departure time. Would toll booths along the route be able to implement this?

for one morning rush hour. It is shown that with a specific dynamic for equilibrium selection, the equilibrium exists and is unique. As we shall illustrate in the last subsection of the introduction, multiple equilibria are quite natural in models of commuting.

The transportation engineering literature is naturally concerned more with practical traffic issues than with the questions we pose; see, for example, Daganzo (2008). Typically this literature takes the behavior of individuals, namely their choice of routes and departure times, as exogenous. Thus, Nash equilibrium is not studied.[6] For example, Zhu and Marcotte (2000) use predetermined (exogenous) departure times. The closest relative to our model in this literature is the cell transmission version of the Lighthill-Whitman-Richards (LWR) model; see Daganzo (2008) section 4.4.6. There are some important differences. First, the LWR model takes departures as exogenous and possibly smooth, whereas we do not. Second, like most models of traffic, the LWR model employs queues or bottlenecks when there is congestion. In contrast, we assume that traffic slows as a function of traffic density. These two important differences express themselves as differences in the equilibrium behavior of the models.[7] More recent examples include Han et al (2013)[8] and Han et al (2015).

Turning next to the mathematics literature, our mathematical problem on one link boils down to a conservation law coupled with a discontinuous differential equation. Even with just one link between an origin and destination with *exogenous* departure times and homogeneous commuters, existence and uniqueness of the resulting traffic pattern is a difficult question that requires interesting assumptions and techniques to resolve. A major issue is the existence and uniqueness of behavior of the system when the initial conditions can be discontinuous. This is important to us, as we don't want to place restrictions on the joint behavior of individuals when we eventually consider Nash equilibrium. The mathematics were introduced in Bressan (2000, chapter 6)

[6]For example, the first appearance of a utility function in Daganzo (2008) is at the bottom of p. 315. The body of the book ends at the top of p. 319.

[7]In other parts of the transportation engineering literature, existence and uniqueness of Nash equilibrium is studied in the context of a bottleneck model, using an S-shaped wish curve (defining ideal bottleneck exit times). In these models, it is unclear what happens if an atom of commuters arrives at the bottleneck at the same time, or if the fragile condition of an S-shaped wish curve is violated - the complement appears to be open and dense in the set of wish curves.

[8]In addition to queues, this work also features a highly non-standard notion of Nash equilibrium.

and Garavello and Piccoli (2006); that work is based on Bressan (1988) and Bressan and Shen (1998).[9] The key paper for our purposes is the seminal work of Strub and Bayen (2006), who remark in their conclusions (p. 564), "However, results are still lacking in order to generalize our approach to a real highway network. For such a network, PDEs are coupled through boundary conditions, which makes the problem harder to pose."[10] Once we have introduced notation and concepts, we shall remark further on both related literature in mathematics and alternative approaches to solving the induced mathematical problem.[11] An important contribution of Strub and Bayen (2006) is actually the *definition* of a solution to the mathematical problem of determining flow in the one link system with *exogenous* departures, since there were issues of either existence or uniqueness with many of the previous attempts. The technical difficulties in the literature are partly the result of working with functions of bounded variation with a two dimensional domain: time and distance. The (discontinuous) conservation law tells us that cars are not lost over a link, with initial condition zero cars on the link and boundary conditions corresponding to the departure of cars. The conservation law is coupled with a (discontinuous) differential equation that gives progress of a car over the link. An important mathematical problem is relating properties of functions on two dimensions that are of bounded variation to their variation on each dimension separately. Part of our mathematical contribution here is to relate the solution of the conservation law of Strub and Bayen (2006) to a

[9] Although the motivation for Bressan (2000) is the simple traffic problem with one home location, one work location, and one link, the mathematical problem solved in this book is *different* from the economic problem that motivates it. This will cause us some headaches. In particular, the initial condition used in the book is the traffic at various locations along the link at time 0, trivially 0 in our model. Traffic is not allowed to enter the link after time 0. We are much more interested in boundary conditions that, for an arbitrary time, give the traffic entering a link at *location* 0. Nevertheless, the mathematics introduced in this book are very useful.

[10] There are many challenges that we must address to extend their results from one link to many. For example, it is difficult to prove that the link exit density has the same properties as the link entry density, that is used as the entry density for another link. A secondary challenge is that boundary conditions are formulated in terms of density (cars per mile) when they should be formulated in terms of volume (cars per hour). Although we take the proper approach for boundary conditions using volume, the technicalities can be simplified some if we were to use density.

[11] We note in frustration that much of the literature cited here is motivated by mathematics rather than economics. Beyond Strub and Bayen (2006), to our knowledge there is no result that applies directly even to the case of two nodes and one link.

Carathéodory solution to the discontinuous ordinary differential equation by applying Biles et al (2014). To accomplish this, we employ techniques initially developed by Friedrich et al (2018), using properties of the Godunov scheme, to obtain bounds on total variation of the solution to the conservation law for each of time and space separately.[12]

In the end, we are able to embed the more elementary framework of Strub and Bayen (2006) in a model with an arbitrary transport network, heterogeneous commuters and endogenous choice of departure times and routes, examining Nash equilibrium as well as Pareto optimum. Unfortunately, we cannot apply their results directly, but must open up the details of their clever proof.

The final literature related to our work is the literature on internet congestion. Although we interpret our model as traffic on roads for consistency of exposition, it applies as well to packets on links in the internet. A fine survey of this literature can be found in Jacobsson (2008). Due to the complexity of the discrete model, a continuous model was developed by Kelly et al (1998), forming a foundation for our work. Much of the literature has a focus on exogenous departures and routes, not Nash equilibrium. Other parts of the literature, such as Kelly et al (1998), focus on steady states of the dynamic model with congestion pricing, or what we call a static model with congestion pricing. There is likely an unexplored relationship with potential games, as represented for example by Sandholm (2007).

1.3 Preview

In summary, the main difference between our work and most of the literature is that we use the fine microstructure from transportation engineering and the mathematics of conservation laws to address more macro economic questions. We do not use exogenous departure and route choices, nor do we employ bottlenecks or queues. Instead we allow endogenous choice of departure times and routes, but require that traffic slow down as a function of endogenous congestion on an arbitrary transportation network. To our knowledge, this represents a new class of models of commuting that has fewer black boxes (such as delay functions in the standard literature) and, more importantly, different properties compared with others.

Although the notation used to describe the models formally is burdensome, we will give examples and intuition for the results in addition to the techni-

[12]Since most of the examples of conservation laws come from physics, these results might also be of use there.

calities. We formulate both a static model, where time plays no role, and a dynamic model, where it does play a role. We assume that commuters have an inelastic demand for one trip per day to work. Future work should extend this to elastic demand.

Our results and the outline of the balance of the paper are as follows. Although classical results concerning Nash equilibrium and Pareto optimum are replicated in our context, we highlight novelties. In the next subsection of the introduction, we detail and preview our results with minimal notation by using the simplest example, a network with two nodes and one link where all commuters live at one node and commute to their jobs at the other. In Section 2, we give our notation and specify the general static (timeless) and dynamic models. At this point, we prove classical results in our context, but also find assumptions sufficient for existence and uniqueness of a traffic pattern over time and across links given a set of boundary conditions (corresponding to a fixed strategy profile) in the dynamic model. Moving on to Nash equilibrium, we find conditions sufficient to prove it exists, and show that it is generally not unique. Section 3 gives our applications. First, we show that the static model cannot be viewed as a reduced form of the dynamic model, where time is explicit. Then we study the welfare properties of Nash equilibrium in the context of a generalization of a tree network in the dynamic model. Nash equilibrium of the morning commute will generally be inefficient, whereas there exists a Nash equilibrium of the evening commute that is efficient. Finally, Section 4 gives our conclusions. All proofs are contained in an Appendix.

1.4 Example

1.4.1 Our Basic Model

We begin with a simple example to illustrate how the model works and the intuition behind our results. Consider measure 1 commuters uniformly distributed on the interval $[0, 1]$ with nodes 1 and 2. The length of the link between the two nodes is 1. Each commuter commutes from node 1 to node 2 each day. For simplicity, we consider only the morning rush hour at this juncture. Denote the capacity of the link in terms of volume (cars per hour) by $x \in \mathbb{R}_+$. Suppose that the time it takes to travel the link at the speed limit is $t(1, 2) = 1$. In the static model, the travel time is given by 1 if the average number of travellers does not exceed free flow capacity x of the road, and by $\frac{1}{x}$ otherwise. This means that if road link capacity is exceeded, then traffic slows down in proportion to the ratio of excess commuters to capacity, $\max\{1, \frac{1}{x}\}$. For example, if $x = 1/2$, then the travel time for a commuter on the link is 2. There really are no choices here for the commuters or a social planner optimizing efficiency, since the route is fixed and the model is static; there are no departure times to be chosen.

Now consider a dynamic version of the model. Route choice is still fixed, but departure (and consequent arrival) times are a choice variable of the commuters. We model departure times in $\mathbb{R}_+$, and we call the required arrival time at the destination node 2 (say 9 AM) $\tau^A \in \mathbb{R}_+$. There is no penalty for arriving at work early, but the penalty for arriving at work late is ∞.[13] This is mainly for illustration. We shall consider more general penalties for both early and late arrival in the remaining sections. They add some complications.

Again, in this simple model there is no route choice. But there is a choice of departure time. First, we illustrate how, for any choice of departure times by all commuters, the travel time to the destination node 2 can be computed. It is assumed that the latter is minimized by each individual commuter at a Nash equilibrium (given the choices of others), and the social planner maximizes a utilitarian welfare function that is minus the integral of commuting times subject to the arrival constraint.

The speed of a particular cohort of commuters who depart at the same time is computed as follows. Begin with the local density of commuters on the road at a particular place on the route and at a particular time. This local density at a given place and time is computed as the limit of neighborhoods on the road of total (measure of) commuters in the neighborhood divided by the one dimensional size of that neighborhood. The limit is taken as the length of the neighborhood goes to zero. The result will be the density of commuters (with respect to distance) at that place and time. Then, as in the static model, traffic slows down in proportion to the ratio of excess commuter density to capacity, where capacity is in terms of volume, namely commuters per hour. In terms of notation, for our examples $\upsilon = \min\left\{1, \frac{x}{f}\right\}$, where υ is speed and f is density.

An example will help illustrate. Again consider the commuters uniformly distributed on $[0, 1]$. Suppose that all the commuters at 0 depart at time 0, all the commuters at 1 depart at time 1, and so forth. Set the arrival time $\tau^A = 2$. We compute traffic speeds (in this case, the arrival time constraint

[13] Notice that for this example, if there is positive probability of arriving late to work, the payoff is $-\infty$ and the commuter is indifferent among all strategies with positive probability of arriving late.

will not bind). With these departure times, when road capacity is high so that $x \geq 1$, then capacity does not bind. The unit interval of commuters moves from origin to destination at full speed and perfect synchrony, and the local density of traffic is always 1 except for commuters with labels 0 and 1. The density around them is $\frac{1}{2}$ since there is nobody on one side of them (for example the commuters with label 0 have nobody in front of them). But this does not alter their speed, since they are already at the speed limit. In theory, at least, commuters can catch up with those ahead of them (if the ones ahead are travelling slower) and slow themselves down.

What if $x < 1$? We consider two simple patterns; for more concreteness, for example one can take $x = 1/2$. Set the arrival time $\tau^A = \frac{2}{x}$. First, suppose that commuters depart uniformly in the time interval $[0, \frac{1}{x}]$ with volume x and density 1. Traffic slows down by a factor of $\frac{1}{x}$ relative to the no congestion case; thus, traffic speed for the commuters is uniform at x, traffic density (cars per mile) is 1, and traffic volume (cars per hour) is speed multiplied by density, or x.[14] It takes $\frac{1}{x}$ time to traverse the link, so the last commuters (labelled 1) reach the destination at $\frac{2}{x}$. Call this the *congested commuting pattern*.

Now consider the same general departure pattern as in the preceding paragraph, with commuters labelled 0 beginning travel at time 0, whereas commuters labelled 1 begin their trip at time $\frac{1}{x}$. Again set the arrival time $\tau^A = \frac{2}{x}$. The local density of commuters on the route is x, so all commuters travel at the speed limit 1, and volume is x. Thus, travel time for all commuters is 1. The last commuter arrives at time $\frac{1}{x} + 1 < \frac{2}{x}$. Call this the *uncongested commuting pattern*.[15]

These two simple commuting patterns, or strategy profiles, serve to illustrate the computation of local density, speed, and volume. Of course, the calculations can be much more complicated in, for example, more intricate commuting networks or for more intricate departure patterns. The simple patterns also serve to illustrate the important role played by arrival time. It is rather evident that for the fixed arrival time as specified at $\tau^A = \frac{2}{x}$, these

[14] Sharp readers will notice that the density in front of the commuters departing at time 0 is 0. But such commuters represent a set of measure 0.

[15] There are more uncongested commuting patterns, for example when $x = 1/2$ and departures are uniform on times $[0, 3]$. The volume of departures at each time is $1/3$. The density f with respect to distance is $1/3$ in locations and at times where there are commuters. Therefore, speed $v = \min(1, x/f) = 1$. The travel time of each commuter is 1. The last commuter arrives at 4.

strategy profiles are Nash equilibria.[16,17] Notice that all commuters reach work by the arrival time τ^A for either pattern, but travel time is longer for the congested commuting pattern. Thus, welfare can differ across dynamic commuting patterns even for this simple example. It is evident that the uncongested commuting pattern Pareto dominates the congested commuting pattern.[18] Finally, it is vital to see that more than one departure density can be consistent with a given departure volume.

1.4.2 The Classical Model with Queues

A crucial comparison is between our model, with endogenous congestion and speed, and the classical models of the literature that use queues. We argue that the equilibrium (or even disequilibrium) behavior of our model is different and much more realistic, illustrated as follows.

First, consider the model detailed previously, but for simplicity set $x = \frac{1}{2}$. For the congested commuting pattern, it takes each commuter time 2 to traverse the link. The first commuter arrives at time 2, whereas the last arrives at time 4. For the uncongested commuting pattern, it takes each commuter time 1 to traverse the link. The first commuter arrives at time 1, whereas the last commuter arrives at time 3.

We turn next to a model with queuing. There are many variations on the bottleneck model, particularly in continuous time. For example, Arnott et al. (1993) assume that it takes no time to get from home to a bottleneck, and that after exiting the bottleneck, the commuter immediately arrives at work. The variation we use is closer to our model, and is due to Zhang and Zhang (2010). A link consists of two parts, a main body first and then a queue at the end. The main body has infinite capacity so traffic flows at the speed limit independent of any congestion. The queue or bottleneck at the end of the main body operates with limited capacity, using a first-in-first-out principle. For our particular example, it takes time 1 for any commuter to traverse the main body (independent of congestion), and the queue allows volume 1 to exit

[16] There are actually many other Nash equilibrium patterns associated with this example that feature no congestion. We focus on these two patterns for simplicity.

[17] Rath (1994, 1998) studies refinements in a general framework that applies to our model. Both the congested commuting pattern and the uncongested commuting pattern are proper and perfect Nash equilibria in this example.

[18] It is important to point out that our basic example of a network is series-parallel and even extension-parallel, but some Nash equilibria of the dynamic model are not weakly Pareto efficient.

the queue at any given time.

What does Nash equilibrium with a queue look like for this example? Everyone leaves as soon as possible (at time 0), arrives at the bottleneck at time 1, and the last commuter leaves the bottleneck at time 2. Here, we assume that if everyone arrives at the bottleneck at the same time, the order in which they proceed is random. This equilibrium does not resemble at all the ones obtained using our model of endogenous congestion.

1.4.3 Comparison of the Static and Dynamic Models

Consider next the comparison of the static with the dynamic model. The first pattern, the congested commuting pattern, we study for the case $x < 1$ seems to be the analog of the static case, since traffic speed is constricted. But the second, uncongested pattern does not seem to have an analog. Thus, the static and dynamic models have different Nash equilibrium predictions. Moreover, if the dynamic analog of the static equilibrium is the congested commuting pattern, it is Pareto dominated by another pattern present in the dynamic model but disallowed by the static model.

In fact, we can say more. In section 3.1 below, we describe how to extend this example so that there is *no* equilibrium of the dynamic model even remotely resembling the equilibrium of the static model.

With the model specified as we have outlined, generally a Nash equilibrium in pure strategies or a pure strategy optimum might not exist. So in what follows, for the dynamic model, we must simplify the problem. This is accomplished by using a fixed, finite set of possible departure times that divide equally the time scale in the model. When commuters choose a departure time, they are distributed uniformly over the interval with midpoint their chosen departure time, and length equal to the distance between allowable departure times. With this structure, a Nash equilibrium in pure strategies and a pure strategy optimum exist. Moreover, for our example, the congested and uncongested commuting patterns we have specified are Nash equilibria of the model, and the uncongested commuting pattern is Pareto optimal. Notice that in this simple example, there is a Nash equilibrium that is Pareto optimal. This will be generalized in Section 3.2.

What follows below just makes the ideas behind our simple examples formal and general, for instance allowing an arbitrary commuting network where commuters have various different origins and destinations. But we wish to emphasize that this work is just a first step in this interdisciplinary research, that we will make assumptions that limit the scope of the results, and that much work remains. Our goal here is to clarify the issues, some of which are subtle, that span several literatures.

2 The Commuting Model

Readers who wish to understand the content of the work through examples only can focus on Examples 1-4 below and then skip to section 4.

2.1 The Static Model: Equilibrium and Optimum

Here we lay out the details of a game with an atomless measure space (continuum) of players; a finite set of nodes at which the players live, or to which they commute, or through which they commute; and a finite set of transport links between the nodes with exogenous capacity.

To begin, the measure space of commuters is given by $(C, \mathcal{C}, \mu)$ where C is the set of commuters, $\mathcal{C}$ is a σ-algebra on C, and μ is a positive, non-atomic measure.[19] We assume that singletons of the form $\{c\}$ for $c \in C$ are in $\mathcal{C}$; that for all $c \in C$, $\mu(\{c\}) = 0$; and $0 < \mu(C) < \infty$.

The origins and destinations in the commuting network are given by a finite set of *nodes*, denoted by $m, n = 1, 2, ...N$. Let $\mathcal{N} = \{1, 2, ..., N\}$. The commuting network itself is given by a finite set of links between nodes. The capacity of any direct link (with no intermediate nodes) between nodes m and n is given by $x_{mn} \in [0, \infty]$, whereas $x_{nn} = \infty$. If a direct link between nodes m and n does not exist, then $x_{mn} = 0$. The units attached to x_{mn} are cars per hour, or volume.

What remains is to specify the strategies and payoffs of the commuters. In the static game, there is no choice of time of departure or arrival. There is only route choice. We assume that each commuter has a fixed origin node and a fixed destination node, with inelastic demand for exactly one trip between the origin and destination. Thus, there is an exogenous, measurable *origin map* $O : C \to \mathcal{N}$ and an exogenous, measurable *destination map* $D : C \to \mathcal{N}$. Notice that there can be heterogeneity among commuters in origins and destinations. This will create heterogeneity in the reduced form utility functions of the commuters.

[19] Skorokhod's theorem implies that we could, with some loss of generality, restrict attention to the unit interval with Lebesgue measure.

Let π_k be the map that projects a vector onto its coordinate k. A *route*, denoted by r, is a vector of integer length ℓ no less than 2. Next we define the set of all routes, $\mathfrak{R}$:

$$\begin{aligned}\mathfrak{R}^\ell &= \left\{r \in \mathcal{N}^\ell \mid \text{for } i = 1, 2, ..., \ell - 1, x_{\pi_i(r)\pi_{i+1}(r)} > 0\right\} \\ \mathfrak{R} &\equiv \bigcup_{\ell=2}^{\infty} \mathfrak{R}^\ell\end{aligned}$$

To avoid trivial situations, we assume that if there is a positive measure of commuters with a particular origin and destination, that there is some route between the nodes. A *commuting length map* is a measurable map $l : C \to \{2, 3, ...\}$. A *commuting route structure* is a pair (l, R) where l is a commuting length map and R is a measurable map $R : C \to \mathfrak{R}$ such that for $i = 1, 2, ..., l(c)-1$, $x_{\pi(i)\pi(i+1)} > 0$, and almost surely for $c \in C$, $\pi_1(R(c)) = O(c)$ and $\pi_{l(c)}(R(c)) = D(c)$.

Given a commuting route structure (l, R), its *flow* $f \in \mathbb{R}_+^{N^2}$ is given by the measure of commuters or cars using the link, $f(m, n) = \mu(\{c \in C \mid \exists k \in \{1, 2, ..., l(c)-1\}$ with $\pi_k(R(c)) = m$ and $\pi_{k+1}(R(c)) = n\})$ for $m, n = 1, 2, ...N$. We assume that the length of the link between nodes m and n is $\lambda(m, n) \geq 0$ for $m, n = 1, 2, ...N$. If the link is congested, then the travel time increases. For our examples, it increases in proportion to the excess of commuters above free flow capacity x_{mn}, $\max\left\{1, \frac{f(m,n)}{x_{mn}}\right\}$.[20] Here, x_{mn} has units cars per hour. For instance, if the number of commuters is twice the capacity of a link, then the travel time is doubled. We ask that the reader bear this special case in mind, since we use it in all of our examples to give concrete intuition.

More generally, we can allow traffic to slow down according to any well-behaved function of the number of commuters at a distance on a link and link capacity. Therefore, we specify the function $\upsilon : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_{++}$ where $\upsilon(f, x)$ is the speed of traffic with usage f on a link with capacity x. We assume that for fixed x, υ is continuous and non-increasing in f. For our

[20] There is an issue of normalization here, namely whether f is divided by λ or not. In essence, it depends on whether a link that is twice as long is half as congested for the same number of commuters on the link. This depends on the interpretation of the static model, whether congestion is viewed as a pulse of commuters or whether they are uniformly spread out over the link. In this paper, we take the view that in the model without time, twice as many commuters on a link results in twice the congestion, no matter the length of the link. However, if one takes the view that length of the link matters, the result is simply division of our f by λ, and this makes no essential difference in the the results we obtain. As we show in Section 3.1, interpretation of the static model is difficult.

examples, average speed on a link mn is given by link length divided by time spent on link, namely $\upsilon(f, x_{mn}) \equiv \frac{\lambda(m,n)}{\max\left\{1, \frac{f}{x_{mn}}\right\}} = \lambda(m,n) \cdot \min\left\{1, \frac{x_{mn}}{f}\right\}$.

Although it is difficult to discuss travel time in a model that is inherently atemporal, equilibrium in this model is to be viewed as a sort of steady state. Under this interpretation, f is the measure of commuters (repeatedly) passing through the link on their route.

This should not be construed as an endorsement of the static model.[21] In fact, the static model has many flaws. However, it is the dominant model in the literature and used in many of the papers we have cited, so we must discuss it for the purpose of comparison. The particular functional form we use for commuting time in our examples is a slight generalization of the one typically used for the Braess paradox in the static model.

The *time cost of a commuting structure* (l, R) *for commuter* c is

$$\theta(l, R, c) = \sum_{\{(m,n)\in\mathcal{N}\times\mathcal{N} \mid \pi_i(R(c))=m, \pi_{i+1}(R(c))=n \text{ for some } 0\leq i\leq l(c)-1\}} \frac{\lambda(m,n)}{\upsilon(f(m,n)), x_{mn})} \tag{1}$$

Thus, $-\theta$ is the objective or payoff function for each commuter. The *utilitarian welfare function for the static model* is

$$U(l, R) = -\int_C \theta(l, R, c) d\mu(c)$$

A *Nash equilibrium of the static model* is a commuting structure (l, R) such that almost surely for $c \in C$, there is no route r of length ℓ for commuter c such that

$$\theta(l, R, c) > \sum_{\{(m,n)\in\mathcal{N}\times\mathcal{N} \mid \pi_i(r)=m, \pi_{i+1}(r)=n \text{ for some } 0\leq i\leq \ell-1\}} \frac{\lambda(m,n)}{\upsilon(f(m,n), x_{mn})}$$

Existence of Nash equilibrium in pure strategies can be proved by applying Schmeidler (1973, Theorems 1 and 2). Rosenthal (1973) proves that a Nash equilibrium in pure strategies exists even when there is a finite number of commuters. Sandholm (2001) shows that equilibrium exists *and is unique* under additional conditions, primarily that speed is strictly decreasing in link usage f.

Next we prove (informally) that an optimum exists. The problem can easily be reduced to optimization of the utilitarian welfare function over a

[21] I come to bury the static model, not to praise it.

compact set as follows. Notice first that there is a finite number of types of commuters, defined by their origin-destination pairs. Instead of using route choice for each commuter, employ as control variables the measure of each type following each route. Thus, the social planner controls a finite number of variables in a compact set using a continuous objective, so a maximum is attained.

Example 1: We note that due to the congestion externality, the Nash equilibria are unlikely to be Pareto (or utilitarian) optimal. To see this informally, consider an example with 3 nodes where each link has length 1. All commuters travel between nodes 1 and 3. There is a direct route, and an alternate route that runs via node 2; see Figure 1. The alternate route takes longer than the direct route for each fixed number of commuters below capacity because it requires a longer distance of travel. For example, each road has capacity 1 and takes 1 unit of time to cross when running below capacity, so the longer route uses 2 units of time when running below capacity, whereas the shorter route takes 1 unit of time when running below capacity.

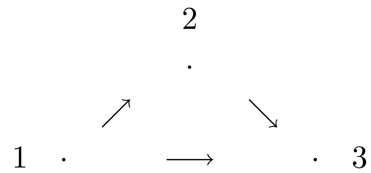

Figure 1: Nash equilibrium is not Pareto optimal

Suppose that there is measure $\frac{5}{2}$ of commuters. A Nash equilibrium of this model has the direct route running above capacity, with measure 2 commuters using it for a total travel time of 2, and the indirect route running below capacity (.5 measure, with a total travel time of 2) such that the travel time to work for each commuter is the same. To create a Pareto improvement over the Nash equilibrium, simply move some commuters (say measure .5) from the direct to the indirect route. The travel time on the indirect route (namely 2) is the same as at the Nash equilibrium, even for the commuters switched to that route, whereas the travel time for those on the direct route decreases (to 1.5).[22]

[22] Notice that in this example, even though the Nash equilibrium is not Pareto efficient, it is weakly Pareto efficient.

2.2 The Dynamic Model: Equilibrium and Optimum

The basics of the dynamic model are the same as those for the static model. To differentiate the notation, we will add "dynamic" to the names and add time τ as an argument of functions. In the dynamic model, each commuter chooses both a departure time (from their origin node) and a route. Routes were discussed in the previous subsection. We allow a commuter to depart at any time $\tau^d \in [0, T]$. As we shall see shortly, it is important that this set be bounded.

It is vital to reinterpret some of the variables used in the static model in terms of dynamics. In the static model, f represents the measure of commuters using a link. In the dynamic model, f will represent the *density* of cars per unit distance at a particular place and time on the link. The free flow capacity for our examples, x_{mn}, will continue to have units cars per hour.

A *dynamic commuting route structure* is a triple (τ^d, l, R) where $\tau^d : C \to [0, T]$ is a measurable function giving departure times for all commuters, l is a commuting length map and R is a measurable map $R : C \to \mathfrak{R}$ such that almost surely for $c \in C$, $\pi_1(R(c)) = O(c)$ and $\pi_{l(c)}(R(c)) = D(c)$.

At this juncture, there is an issue concerning the detail in which we model congestion on each link in the dynamic model. It varies in the literature we have cited. The simplest way to model this is to look only at average congestion on a link. More complicated is to assume that as traffic ebbs and flows, the congestion at the end of the link determines traffic speed on the entire link. The most detailed model allows cars to catch up with each other over the course of a link. We use this most detailed model, but assume that link capacity is constant across the link. This is without loss of generality, provided that capacity changes only a finite number of times on a link. In that case, we just add more nodes and links with different capacities in series.

We shall define commuter progress from origin to destination through a differential equation in distance. But first we must define progress on each component of a route in a dynamic route structure. Fix a dynamic route structure (τ^d, l, R). The basic idea is this. From departure time to the end of the first link, we follow the differential equation for congestion for the first link, and then begin on the second link, and so forth. A crucial assumption made here and in most of the literatures we cite is that cars cannot pass. For notational simplicity, for $i = 1, ..., l(c)$, define $\tau_i(c)$ to be the time that node $\pi_i(R(c))$ is reached. Evidently, $\tau_1(c) = \tau^d(c)$.

Given a dynamic commuting route structure (τ^d, l, R), we shall associate with it a function $\delta_{mn}(\tau_m(c), \tau)$ that gives as its value the distance travelled on link mn by commuter c at time τ who begins travel on link mn at time $\tau_m(c)$.[23] In the end, this function will be increasing in its second argument but decreasing in its first argument. Does such a function exist, and is it unique? Fix such a function $\widehat{\delta}_{mn}$. To ease notation, compute inductively

$$\tau_{i+1}(c) = \inf\{\tau' > 0 \mid \widehat{\delta}_{\pi_i(R(c))\pi_{i+1}(R(c))}(\tau_i(c), \tau') = \lambda(\pi_i(R(c)), \pi_{i+1}(R(c))\} \tag{2}$$

Then we can compute its *density at time τ on link mn at distance Δ*, called $\widehat{f} : \mathcal{N}^2 \times \mathbb{R}^2_+ \to \mathbb{R}_+$, and written as $\widehat{f}(m, n, \tau, \Delta)$.[24] It is the density of commuters (per unit distance) at time τ and at distance Δ along link mn, and it is given by the (possibly discontinuous) partial differential equation or *conservation law*:

$$\frac{\partial \widehat{f}(m, n, \tau, \Delta)}{\partial \tau} + \frac{\partial \Phi(\widehat{f}(m, n, \tau, \Delta))}{\partial \Delta} = 0 \tag{3}$$

where

$$\Phi(m, n, \tau, \Delta) \equiv \upsilon\left(\widehat{f}(m, n, \tau, \Delta), x_{mn}\right) \cdot \widehat{f}(m, n, \tau, \Delta) \tag{4}$$

is defined to be the *flux*.[25] The flux is the volume of commuters passing through a point per unit of time. We abuse notation slightly and sometimes write

$$\Phi_{mn}(f) = \upsilon\left(f, x_{mn}\right) \cdot f$$

For our example, note that $\Phi_{mn} = \min\{x_{mn}, f\}$.

Equation (3) is actually *the fundamental conservation law of transportation economics* applied to this model. As explained in Bressan (2000, equation 1.2), if we fix an interval of locations on a link, the measure of commuters inside this interval can only change over time from inflows into the interval from the left and outflows from the interval to the right. Another interpretation of equation (3) states that the change with respect to time in commuter density at a given place and time can be found by looking at the change in the flux (commuters per hour) at preceding locations nearby.

Next we compute

$$\frac{\partial \widehat{\delta}_{mn}(\tau_m(c), \tau)}{\partial \tau} = \upsilon\left(\widehat{f}(m, n, \tau, \widehat{\delta}_{mn}(\tau_m(c), \tau)), x_{mn}\right) \tag{5}$$

[23] A formal definition of this function will be given in (6) below.

[24] In terms of notation, f will be a scalar representing an arbitrary value of the density, whereas $\widehat{f}$ is a density function.

[25] The literatures we have cited use inconsistent definitions for the terms "flow" and "flux" in this context. Here we are using definitions from the mathematics of conservation laws.

This describes the progress made by commuters on each link of the entire dynamic commuting route structure for any time τ. Equation (5) is the coupled discontinuous differential equation discussed in the introduction.

Unfortunately, the coupled system defining $\widehat{f}$ and $\widehat{\delta}$, namely (2), (3), and (5), is technically challenging. The reason is that we cannot restrict τ^d, the function defining the departure strategies of players, beyond assuming that it is a measurable function. Each individual makes a choice, and this is not necessarily coordinated. Discontinuities in departures can result in discontinuities in $\partial\widehat{\delta}_{mn}/\partial\tau$ that rule out our ability to use standard techniques from the theory of ordinary differential equations as well as the contraction mapping theorem. Instead, we use Biles et al (2014).

Even if we can retrieve a well-defined $\widehat{\delta}_{mn}$ for each τ^d function, the issue then becomes the fact that there might not exist a Nash equilibrium in pure strategies, since the space of pure strategies is a continuum. Schmeidler (1973) relies heavily on the fact that the number of pure strategies available to players is finite.

We address the problems of discontinuities in boundary conditions and an infinite number of strategies at once by simplifying the dynamic model. Fix $\overline{\tau}$ where $T/\overline{\tau}$ is an even integer, and define the departure strategy space to be $\{\overline{\tau}, 3\overline{\tau}, ..., (T/\overline{\tau} - 1)\overline{\tau}\}$. This makes the strategy space finite. We assume that all the commuters who choose, say, $\overline{\tau}$ will be randomly and uniformly distributed on $(0, 2\overline{\tau})$, those who choose the strategy $3\overline{\tau}$ will be randomly and uniformly distributed on $(2\overline{\tau}, 4\overline{\tau})$, and so forth. The examples in the introduction fit this framework because they use a uniform distribution of departure times. Moreover, they survive as Nash equilibria no matter how fine the grid, even in the limit where commuters can choose their precise departure time. We view $\overline{\tau}$ as a commuter's intended departure time, where the actual departure time is the intended time plus a small random variable. The constant $\overline{\tau}$ can be small.[26]

Having addressed these initial challenges, we come upon another that is

[26] The classical problem with addressing this issue is that, in the limit with no restrictions on departure time choice, the Nash equilibrium might be in mixed strategies. It is doubtful that an equilibrium in mixed strategies is useful in the applied context here. In fine but finite departure grids, this corresponds to extreme oscillations (say between 0 and 1) in adjacent grid elements. The weak limit would be a strategy profile that is constant at 1/2. Utilities could easily be weakly discontinuous, and thus the Nash equilibrium would be in mixed strategies. An alternative, used in the stochastic games literature, is to perturb utilities. In our context, this does not seem very natural.

generated by the mathematics of conservation laws. Consider the simple one link model used in the introduction where link capacity is 1; the cohort departing at times $(0, 2\overline{\tau})$ has low density, say $1/2$, and hence high speed; and the cohort departing at times $(2\overline{\tau}, 4\overline{\tau})$ has high density, say $3/2$. Common sense and observation of the real world says that the first cohort will begin at high speed whereas the second cohort will begin at lower speed. A distance gap with no commuters will form between them and expand along the link. This is not only the common sense solution, but will also obtain in the solution proposed in the literature if there is an $\varsigma > 0$ start time gap between the cohorts.

Surprisingly and unfortunately, that is not the solution in the case where there is no such starting time gap. The solution in this case involves the initiation of a third step between the first two, with further step initiation possibly following. There is no distance gap between the cohorts. This can be found in unnumbered equations in Bressan (2000, p. 110, Case 2) and Strub and Bayen (2006, p. 559). More detail can be found in Section 4.5 of Bressan (2000), in particular Figure 4.5 and *especially Figure 4.6.* Thus, there is a discontinuity in the solution along a link as $\varsigma \to 0$. Since we do not consider the solution at $\varsigma = 0$ to be the right one, we are forced to take limits of solutions as $\varsigma \to 0$.[27]

It should be obvious by now that we cannot simply apply the LWR model, that relies on the mathematics of conservation laws, as suggested in the transportation engineering literature.

We begin by giving the intuition for speed calculations, and then provide a formal proof of existence and uniqueness of the function $\widehat{f}$, from which everything else can be calculated. For example, $\partial\widehat{\delta}_{mn}/\partial\tau$ can be calculated from (5) once $\widehat{f}$ is known. The intuition we give for the evolution of the system over time and the speed of commuters on a link will be justified because Theorem 1 will tell us that $\widehat{f}$ is unique on $(0, \overline{t}) \times (0, \lambda(m, n))$ for a given departure pattern, and the evolution of the system and speeds we propose will be a solution.

For speed calculations, it is useful to define some concepts. A *threshold* is a location on the network where the speed of commuters is different on the two sides of the threshold at a given time. An important example of a threshold is a node. Of course, a node is a form of a stationary threshold, since it doesn't

[27] As we shall see later in the paper, in Theorem 1 specifically, given departure times and routes, the solution $\widehat{f}$ to the problem we pose for each link is unique almost surely *on the interior* of the product of the time and distance (on the link) domains. This indeterminacy issue arises on the boundary, namely where distance on the link is zero.

move over time. Next we will investigate thresholds that move, appear and disappear. An example of a threshold of this type is the boundary between two cohorts, where a *cohort* is defined as a group of commuters with the same route and departure time choices.

Fix a dynamic commuting route structure (τ^d, l, R). Let $\widehat{\tau}^d(c, \tau') = \tau^d(c) + \tau'$, where τ' is a random variable uniformly distributed on $(-\overline{\tau}, \overline{\tau})$, denote the actual departure time of commuter c, that differs from the chosen departure time $\tau^d(c)$ by at most $\overline{\tau}$ as described just above. To reduce the notational burden, we shall generally suppress the second argument (τ') in any function $\widehat{\tau}$. Then $\widehat{\tau}_1(c) = \widehat{\tau}^d(c)$. In general, given $\widehat{\tau}_i$, we will define inductively $\widehat{\tau}_{i+1}$. Fix any origin node m and destination node $n \neq m$. On each segment mn, define a subset of commuters who travel together on link mn as:

$$\begin{aligned}\alpha_{mn}(c) \equiv \{c' \in C \mid \tau^d(c') = \tau^d(c);\ \text{for some } i \leq l(c),\\ \pi_1(R(c)) = \pi_1(R(c')), \ldots, \pi_{i-1}(R(c)) = \pi_{i-1}(R(c'));\\ \pi_{i-1}(R(c)) = \pi_{i-1}(R(c')) = m,\ \pi_i(R(c)) = \pi_i(R(c')) = n\}\end{aligned}$$

Assume for now that Φ_{mn} is strictly increasing in f. Then the *default speed* for commuter c is given by

$$S_{mn}(c) = \upsilon\left(\Phi_{mn}^{-1}\left(\mu(\alpha_{mn}(c))/2\overline{\tau}\right), x_{mn}\right)$$

The appearance of Φ_{mn}^{-1} here will be discussed in detail in Remark 6 below. At this point, it is useful simply to note that we must translate departure volume (cars per hour) to departure density (cars per mile). The literatures we have surveyed do not account for this.

The default speed might be counterfactual, but it is a useful construct. At the default speed, intervals of commuters never overlap with each other. When they never overlap, the time on this link is exactly $\lambda(m,n)/S_{mn}(c)$, so $\widehat{\tau}_{i+1}(c) = \widehat{\tau}_i(c) + \lambda(m,n)/S_{mn}(c)$. Similarly, $\delta_{mn}(\widehat{\tau}_i(c), \tau) = S_{mn}(c) \cdot [\tau - \widehat{\tau}_i(c)]$ where $\pi_i(R(c)) = m$. But there are two other possibilities beyond this first case. The second case is when commuters using different routes blend with each other or separate beginning at a node; this is actually a generalization of the concept of default speed. The third case is if a segment of commuters catches up with another along a link. We consider each of these in turn.

The second case that is possible in the model is when commuters using different routes blend or separate at a node. For the case where they separate, if they are not combined with commuters using other routes, they move at the default speed on the link. But this is just to give intuition. To ease notation, define $\pi_0(R(c)) \equiv \pi_1(R(c))$ for all routes R and almost all commuters c, and define $\widehat{\tau}_0(c) \equiv \widehat{\tau}_1(c)$ for almost all commuters c. Formally, defining the set of commuters approaching link mn from link $m'm$ at the same time:[28]

$$\begin{aligned}\gamma_{m'mn}(c,\epsilon) \equiv \{c' \in C \mid \widehat{\tau}_{j-1}(c') \in (\widehat{\tau}_i(c)-\epsilon, \widehat{\tau}_i(c)+\epsilon);\\ \pi_i(R(c)) = m,\, \pi_{i+1}(R(c)) = n;\\ \pi_j(R(c')) = m,\, \pi_{j+1}(R(c')) = n \text{ and } \pi_{j-1}(R(c')) = m' \;\}\end{aligned}$$

The speed of commuters is given by:

$$S^*_{mn}(c) = v\left(\Phi^{-1}_{mn}\left(\sum_{m'\in\mathcal{N}} \lim_{\epsilon\to 0} \frac{\mu(\gamma_{m'mn}(c,\epsilon))}{2\epsilon}\right), x_{mn}\right)$$

Provided that they don't catch up with anyone else, their time on the link is exactly $\lambda(m,n)/S^*_{mn}(c)$, so $\widehat{\tau}_{i+1}(c) = \widehat{\tau}_i(c)+\lambda(m,n)/S^*_{mn}(c)$ whereas $\delta_{mn}(\widehat{\tau}_i(c),\tau) = S_{mn}(c)\cdot[\tau-\widehat{\tau}_i(c)]$ where $\pi_i(R(c)) = m$.

On each segment mn, we say that *commuter c catches up with commuter c' on link mn* if

$$\begin{gathered}\pi_i(R(c)) = \pi_j(R(c')) = m,\, \pi_{i+1}(R(c)) = \pi_{j+1}(R(c')) = n\\ \widehat{\tau}_j(c') < \widehat{\tau}_i(c)\\ \frac{\lambda(m,n)}{S^*_{mn}(c)-S^*_{mn}(c')} < \widehat{\tau}_i(c) - \widehat{\tau}_j(c')\end{gathered}$$

The slower commuter c', who is unaffected, continues on at the same speed as before the faster one c catches them. If commuter c catches up with commuter c' on link mn, define the *catch up* time, for $\pi_i(R(c)) = \pi_j(R(c')) = m$, $\pi_{i+1}(R(c)) = \pi_{j+1}(R(c')) = n$, as $\tau^* = \widehat{\tau}_i(c)+\frac{S^*_{mn}(c')\cdot[\widehat{\tau}_j(c)-\widehat{\tau}_i(c')]}{S^*_{mn}(c)-S^*_{mn}(c')}$. At the first time when a member of one cohort (defined above) catches up with a member of another cohort along a link, a new threshold is created at this time and distance. As it crosses the threshold, the traffic in the faster cohort slows down to the speed of the cohort immediately in front of them by increasing its density at the threshold to match that in the slower cohort. Thus, for all $c'' \in C$ with $\pi_k(R(c'')) = m$, $\pi_{k+1}(R(c'')) = n$, then $\widehat{\tau}_{k+1}(c'') = \tau^* + \frac{\lambda(m,n)-\delta_{mn}(\widehat{\tau}_i(c),\tau^*)}{S^*_{mn}(c')}$ whereas $\delta_{mn}(\widehat{\tau}_k(c''),\tau) = \delta_{mn}(\widehat{\tau}_i(c),\tau^*) + [\tau-\tau^*]\cdot S^*_{mn}(c')$ for all $\tau > \tau^*$ on this link.

To abbreviate notation, let $\widehat{f}(m,n,\tau,\Delta)^- \equiv \lim_{\epsilon\to 0}\widehat{f}(m,n,\tau,\Delta-\epsilon)$ and $\widehat{f}(m,n,\tau,\Delta)^+ \equiv \lim_{\epsilon\to 0}\widehat{f}(m,n,\tau,\Delta+\epsilon)$. A threshold on link mn is defined by

[28] The notational simplification allows us to model those beginning their first link as $m' = m$.

$(\tau, \Delta) \in [0, T] \times [0, \lambda(m, n)]$ such that $\widehat{f}(m, n, \tau, \Delta)^- \neq \widehat{f}(m, n, \tau, \Delta)^+$. The threshold itself moves along the link at speed

$$\upsilon\left(\widehat{f}(m, n, \tau, \Delta)^+\right) - \frac{\widehat{f}(m, n, \tau, \Delta)^- \cdot \upsilon\left(\widehat{f}(m, n, \tau, \Delta)^-\right)}{\widehat{f}(m, n, \tau, \Delta)^+}.$$

We shall remark on this further after a formal statement of the first result.

Here we have given heuristic details of the behavior of commuters at a departure node and along a link. What we will need for the formal analysis is how the departure density and its composition at the start of a link relate to the exit density and composition at the end of a link. This will be given formally in equation (19) of the appendix.

To prepare for the first result, let us make explicit the assumptions we will use.

Assumption 1: For each fixed x_{mn}, speed $0 < \upsilon(f, x_{mn}) < \infty$ is Lipschitz continuous and non-increasing in f.

Assumption 1 means that car speed with no congestion is bounded, speed is a continuous (though not necessarily smooth) function of congestion, and speed does not increase with more cars. As an alternative to assuming that υ is Lipschitz, we could directly assume that Φ is Lipschitz, as that is what we use. But since both f and υ are bounded (see below after Assumption 2), υ Lipschitz implies that Φ is Lipschitz.

Next, we need some preparation for Assumption 2. Eventually, we will need a bound on the total variation of boundary conditions at the start of a link that is uniform across links. The purpose is to have a compact space that we will use to find a fixed point. A sufficient (and weakly necessary) condition is a hierarchy of links that we will specify next. Let the set of links be denoted by:

$$\mathcal{L} \equiv \{(m, n) \in \mathcal{N} \times \mathcal{N} \mid m \neq n\}$$

We postulate a complete preorder on $\mathcal{L}$ denoted by $\succeq$, with its asymmetric part denoted by $\succ$. Recall that $\mathfrak{R}$ is the set of all possible routes. Next, we shall restrict routes to $\mathcal{R} \subseteq \mathfrak{R}$. For $\ell = 2, 3, ..., N$, define

$$\mathcal{R}^\ell \equiv$$
$$\left\{r \in \mathfrak{R}^\ell \mid \text{For all } i = 2, 3, ..., \ell - 1,\ (\pi_i(r), \pi_{i+1}(r)) \succ (\pi_{i-1}(r), \pi_i(r)),\ \pi_i(r) \neq \pi_j(r) \text{ for } i \neq j\right\}$$

Assumption 2: Routes r are restricted to:[29]

$$r \in \mathcal{R} \equiv \bigcup_{\ell=2}^{N} \mathcal{R}^{\ell}$$

There are two pieces to this assumption. First, we have restricted route length to N or less. In fact, all that is needed is a finite upper bound on route length. We choose N for simplicity. The assumption that nodes are not repeated along a route makes indexing progress along the route easy. These assumptions are made mainly to keep notation simple.

The second piece is more interesting. Let us begin with the mathematics. The purpose of this assumption is to provide a uniform upper bound on total variation (across time) of boundary or entry conditions for the node at the start of a link. Without this upper bound, we lose both compactness of the space of initial conditions and the ability to solve the differential equation (5). We need compactness for a fixed point theorem, and the ability to solve the differential equation in order to compute travel times and payoffs.

To obtain such an upper bound, we must examine behavior when cohorts merge at a node and travel the next link together. Variation in density in one cohort can be transmitted to the other at the initial node. For example, consider two cohorts that merge at a node. If one has a constant entry density over time, but the other has either an increase or decrease in entry density, the density of the first cohort will generally not be constant once it enters the link. Thus, total variation can build up. Even if commuters don't travel in circles, the variation that is transmitted can build up along links that form a closed loop. So to prevent this, we impose a hierarchy on links.

Turning next to the economics of this assumption, it means that commuters (or packets for the internet) must not be travelling on links that form circles. However, travel in opposite directions on links or routes is fine; in fact, this is common for both the internet and commuting applications. For example, if there is a central business district, then one way to satisfy the assumption is to have commuters from each suburb travel towards it during morning rush hour and away from it in the evenings. Circular roads or links forming a circle are also fine, as long as the circle is not completed by overlapping commuters. In the context of the internet, the assumption provides a warning concerning the

[29] Technically, Assumption 2 means that a dynamic route structure is restricted: $R(c) \in \mathcal{R}$ almost surely.

potential build up of total variation in circles on the internet, even if no set of packets travels in a circle, due to the transmission of total variation across cohorts that travel the same link at the same time.

All of our examples (including a generalization of directed trees used in section 3.2) satisfy this assumption.

Milchtaich (2006) is a classic examination of efficiency of Nash equilibrium in the static model, so it important to relate assumptions in that work to our Assumption 2 in particular. The major distinction is that our Assumption 2 applies to directed links, whereas the assumptions in that paper refers to nodes. In particular, links between two nodes in opposite directions are different in our framework. For example, Proposition 3 of Milchtaich (2006) states that: "A two-terminal network G is series-parallel if and only if the vertices can be indexed in such a way that, along each route, they have increasing indices." Theorem 1 of that paper states: "Braess's paradox does not occur in a two-terminal network G if and only if G is series-parallel." It is obvious that the classical Braess paradox example, given in Figure 1 of that paper, satisfies our Assumption 2 but is not series-parallel.

Turning next to analysis of the system, there are two immediate, useful consequences of bounding the commuting route length by N. First, the set of routes that are possible for a commuter to choose, henceforth called $\mathcal{R}$, is finite. Second, we can examine bounds on our endogenous function $\widehat{f}(m,n,\tau,\Delta)$. Evidently, $\widehat{f}(m,n,\tau,\Delta) \geq 0$. Now consider upper bounds. An upper bound for departure density is $\Phi_{mn}^{-1}\left(\frac{\mu(C)}{2\overline{\tau}}\right)$. But it is useful to have a uniform bound on density beyond departure density. As we have seen, when one cohort of commuters catches up with a slower cohort ahead of it, this cohort of commuters slows down by building up density so it is the same as that of the slower cohort. Thus, this does not change the upper bound on density. Where density can build up is at nodes, where cohorts can combine. It is important to note also that boundary conditions at the origin of any route are stated in terms of volume (cars per hour) rather than density (cars per mile). Thus, an upper bound on endogenous density is given by the maximal density: $\overline{f} = N \cdot \max_{m,n} \Phi_{mn}^{-1}\left(\frac{\mu(C)}{2\overline{\tau}}\right)$.

Definition: Let $\bar{t}$ be an upper bound on the time it will take until the last commuter reaches the end of their route:

$$\bar{t} = T + N \cdot \max_{m,n\in\mathcal{N},\, m\neq n}\left[\frac{\lambda(m,n)}{\upsilon\left(\overline{f}, x_{mn}\right)}\right]$$

This time will be finite as long as $v > 0$.

At this point, there is an important but technical issue that must be addressed. We shall use Schauder's theorem[30] to show that for any choice of strategies by commuters, namely the choice of route and departure time for each, the density on each link of commuters in space and time as well as total commuting time are well-defined, namely such a density exists and is unique. This requires some continuity of commuting times in initial conditions on a link. Moreover, we employ Schmeidler's theorem to prove existence of Nash equilibrium for the dynamic commuting game. One of the requirements of Schmeidler's results is that utility is continuous (in the weak topology on L^1) in the strategy profile of all commuters. For the dynamic model as stated, there is an important type of discontinuity that must be addressed.

The discontinuity is related to moving thresholds. In particular, if a threshold moves backward through a node, a discontinuity in commuting times and payoffs can result. Consider the following example represented in Figure 2:

$$\begin{array}{ccccc} & & & \cdot & C \\ A & & B & \nearrow & \\ \cdot & \Longrightarrow & \cdot & & \\ & & & \searrow & \\ & & & \cdot & D \end{array}$$

Figure 2: A discontinuity

Traffic moves from left to right, with origin at node A through a node represented by B. After passing through the node, some traffic heads up and to the right on route ABC, whereas other traffic heads down and to the right on route ABD. Suppose that after passing through node B, traffic heading down and to the right on route ABD travels at high speed, and the traffic volume and density are steady. Suppose further that a large, slow cohort passes through the node and heads up and to the right on route ABC, but is followed along the same route by a faster cohort that catches up to the slower one along the upper right link BC, after node B. Thus, a threshold is formed and the faster cohort slows down to match the speed and density of the slower one. If the volume of this faster cohort is so large that the threshold[31] backs up along the upper right link BC and through node B to the left link AB, we

[30]See Smart (1974).

[31]Notice that the cars themselves always have positive speed.

claim that a discontinuity in the speed and payoffs of traffic heading down and to the right on route ABD can occur. The speed of the steady traffic heading down and to the right is reduced to the speed of traffic at the threshold, thus increasing in a discrete manner its density and the time needed to travel the link down and to the right, BD. This can happen despite the fact that much of the traffic on the first link AB proceeds up and to the right, because the density of traffic using the lower link BD jumps up when the threshold passes backward through node B to the left link AB.

A sufficient (but not necessary) condition to prevent this type of discontinuity would be one that prevents thresholds from moving backwards, whether through a node or not. Thus, we assume:

Assumption 3: $\Phi_{mn}(f)$ is a non-decreasing function of f.

This assumption prevents thresholds from moving backwards, because it says that the volume of consumers moving past a given point will not decrease if density goes up. The direction of movement of thresholds is governed by local volume (cars per unit time), not by local density (cars per unit distance). With this additional assumption to address the discontinuity, commuting times and payoff functions will be shown to be continuous in strategies.

For our examples, note that $\Phi_{mn}(f) = \min\{x_{mn}, f\}$ satisfies this assumption.

Actually, there is a *much more compelling reason* than the one given above to impose Assumption 3. One of the main points of Strub and Bayen (2006) is that boundary conditions for the conservation law (on the one link they use) can only be satisfied in a weak, not a strong sense. To see this, consider the case of a strictly concave flux function where flux is decreasing somewhere. Next we rewrite equation (4) of their paper, giving the weak left boundary condition on the link, in our notation:

$$\begin{cases} f(m,n,0,\tau) = \rho_{mn}(\tau) \text{ or} \\ \frac{d\Phi_{mn}(f(m,n,0,\tau))}{df} \leq 0 \text{ and } \frac{d\Phi_{mn}(\rho_{mn}(\tau))}{df} \leq 0 \text{ or} \\ \frac{d\Phi_{mn}(f(m,n,0,\tau))}{df} \leq 0 \text{ and } \frac{d\Phi_{mn}(\rho_{mn}(\tau))}{df} \geq 0 \text{ and } \Phi_{mn}\left(f(m,n,0,\tau)\right) \leq \Phi_{mn}\left(\rho_{mn}(\tau)\right) \end{cases}$$

The first line would be considered the usual boundary condition, stating that the density of cars at distance zero and time τ on a link is equal to the density entering the link. That line alone is called the strong boundary condition. The other two options involve the *potential loss of cars* and volume at the entry node for the link. Example 2 of their paper shows that the strong boundary condition, namely the one given on the first line, cannot always be satisfied.[32] In their example, departure density is higher than specified by the boundary condition, but volume is lower. In fact, this example also justifies first proving existence of a solution in the case of a strictly increasing flux function, and turning later to the case of a weakly increasing flux function.[33]

In this framework, when density f increases, $\nu(f)$ weakly decreases whereas $\Phi_{mn}(f)$ weakly increases. Thus, speed and volume are (weakly) monotonically related. This rules out *some* interesting cases of *hypercongestion*, where there are multiple speeds that will accommodate the same volume. Although it is more complicated, we conjecture that an extension of the model could accommodate these cases by eliminating Assumption 3 and allowing payoff discontinuities at nodes using the following technique. Strub and Bayen (2006) allow a boundary condition at the end of a link as well as the beginning of a link, and this can combined with Khan's (1989) generalization of Schmeidler (1973) to upper semicontinuous payoff functions to obtain existence of Nash equilibrium. However, as Strub and Bayen (2006, section 2) note, boundary conditions on both ends of a link can cause inconsistencies (or gridlock) in the density on a link, and the possibility of no solution. That is a second reason why they use a *weak* formulation of boundary conditions that allows violation of strong boundary conditions under certain circumstances. The extension to allow more general flux functions would require weak boundary conditions, and thus the potential loss of cars.

A final issue concerning a decreasing flux function is that standard concave flux functions that are decreasing for high traffic densities generally have a density f^* at which $\Phi_{mn}(f^*) = 0$, and traffic comes to a halt. This could be permanent and part of an equilibrium, for example if a high enough density leaves home at the first opportunity.

Unlike the entire extant literature, we state origin departure boundary conditions (cars entering a link per hour) in terms of volume rather than density (cars per mile). This is a very important distinction. A strategy profile of commuters determines initial volume, namely departures per hour, and not departure density, unless density is completely determined by volume. In general, both departure volume and departure density must be specified in Nash equilibrium.

[32] The boundary condition at the right endpoint of the link is immaterial to this example.

[33] Of course, it would be possible to use the weak boundary condition and allow more general flux functions.

Definitions: Let

$$\Omega \equiv \left\{ (\widehat{\tau}, \tau) \in \left[0, \overline{t}\right]^2 \mid \tau \geq \widehat{\tau} \right\} \tag{6}$$

$$\mathcal{D}_{mn} \equiv \big\{ \delta_{mn} : \Omega \to [0, \lambda(m,n)] \text{ measurable} \mid \text{for } (\widehat{\tau}, \tau), (\widehat{\tau}', \tau') \in \Omega\text{: } \delta_{mn}(\widehat{\tau}, \widehat{\tau}) = 0,$$
$$\left| \delta_{mn}(\widehat{\tau}, \tau) - \delta_{mn}(\widehat{\tau}', \tau') \right| \leq \upsilon(0) \cdot \left(\left| \widehat{\tau} - \widehat{\tau}' \right| + |\tau - \tau'| \right) \big\}$$

$$\mathcal{D} \equiv \prod_{m,n=1,\ m \neq n}^{N} \mathcal{D}_{mn}$$

We use square block metric for the Lipschitz condition as a matter of convenience.

The following definition comes from Strub and Bayen (2006), adapted to our context. Interpretations immediately follow the definitions. For further discussion, see also Bressan (2000).

Definition: A collection of measurable functions $\left\{ \widehat{f}(m,n,\cdot,\cdot), \widehat{\delta}_{mn} \right\}_{m,n=1,\ m \neq n}^{N}$, *where* $\widehat{f}(m,n,\cdot,\cdot) : [0,\overline{t}] \times [0, \lambda(m,n)] \to [0, \overline{f}]$ *and* $\widehat{\delta}_{mn} \in \mathcal{D}_{mn}$*, is called a* ***solution*** *to the conservation law (3) with initial and boundary conditions if, for every* m *and* n ($m \neq n$)*, for every* $k \in \mathbb{R}$*, for every* C^1 *function* $\varphi_{mn} : (0, \overline{t}) \to \mathbb{R}_+$ *with compact support, for every* C^1 *function* $\psi_{mn} : (0,\overline{t}) \times (0, \lambda(m,n)) \to \mathbb{R}_+$ *with compact support, the following hold:*

$$0 \leq \int_0^{\lambda(m,n)} \int_0^{\overline{t}} \left| \widehat{f}(m,n,\tau',\delta') - k \right| \cdot \frac{\partial \psi_{mn}(\tau',\delta')}{\partial \tau} \tag{7}$$
$$+ sign\left(\widehat{f}(m,n,\tau',\delta') - k \right) \cdot \left[\Phi_{mn}\left(\widehat{f}(m,n,\tau',\delta') \right) - \Phi_{mn}(k) \right] \cdot \frac{\partial \psi_{mn}(\tau',\delta')}{\partial \delta} d\tau' d\delta'$$

and there exist $2(N-1)^2$ *sets of Lebesgue measure zero:* $E_{mn}^0 \subseteq [0, \lambda(m,n)]$, $E_{mn}^L \subseteq [0, \overline{t}]$, such that for all $m, n = 1, ..., N$, $m \neq n$,

$$\lim_{\tau \to 0,\ \tau \notin E_{mn}^0} \int_0^{\lambda(m,n)} \left| \widehat{f}(m,n,\tau,\delta') \right| d\delta' = 0$$

$$\lim_{\delta \to 0,\ \delta \notin E_{mn}^L} \int_0^{\overline{t}} L_{mn}\left(\widehat{f}(m,n,\tau',\delta), \rho_{mn}(\tau') \right) \varphi(\tau') d\tau' = 0$$

where

$$L_{mn}(a,b) \equiv \sup_{\kappa \in I(a,b)} \left(sign(a-b) \cdot [\Phi_{mn}(a) - \Phi_{mn}(\kappa)] \right)$$
$$I(a,b) \equiv [\inf(a,b), \sup(a,b)] \tag{8}$$
$$\widehat{\delta}_{mn}(\widehat{\tau}, \tau) = \int_{\widehat{\tau}}^{\tau} \upsilon\left(\widehat{f}(m,n,\tau', \widehat{\delta}_{mn}(\widehat{\tau}, \tau')), x_{mn} \right) d\tau'$$

$$\rho_{mn}(\tau) \equiv \Phi_{mn}^{-1}\left(\frac{\mu(\{c \in C \mid \pi_1(R(c)) = m,\ \pi_2(R(c)) = n,\ \left|\tau^d(c) - \tau\right| < \overline{\tau}\})}{2\overline{\tau}} + \sum_{m'=1,\ m'\neq m}^{N} \Phi_{m'm}\left(\widehat{f}(m', m, \tau, \lambda(m', m))\right)\right) \tag{9}$$

Remark 1: The crucial but subtle connection between the functions $\widehat{f}$ and $\widehat{\delta}$ is through equation (9), called the boundary condition, and definition (2). Condition (9) gives entry into a link by those just departing from their origin node and those continuing their travel through the node from other links. We shall (temporarily) make the assumption that Φ_{mn} is strictly increasing, so that its inverse is well-defined.

Remark 2: What we call a solution is actually a refinement of other solution concepts used in the literature that are more obviously related to equation (3). The least restrictive of these is the concept of distributional solution, followed by the more restrictive weak solution. The (yet more restrictive) solution concept we use is generally called an entropy weak solution in the literature. Motivation for using this solution is that although we have existence theorems for all of these solution concepts, uniqueness holds only for the entropy weak solution concept. There is also intuition for the refinement in terms of stability, usually called admissibility conditions, in the mathematics literature we have cited.

Remark 3: It is important to provide at least a heuristic explanation, part of the folklore in the literature, about why this represents a solution to the partial differential equation or conservation law (3), since there is no obvious connection between the partial differential equation and what we call a solution.[34] Suppose that ψ can be chosen so that $\frac{\partial \psi_{mn}}{\partial \tau}$ is close to an indicator function for some set in $[0, \overline{t}] \times [0, \lambda(m, n)]$ and $\frac{\partial \psi_{mn}}{\partial \delta}$ is close to an indicator function for that same set multiplied by $\frac{1}{v\left(\widehat{f}(m,n,\tau,\delta), x_{mn}\right)}$, so that we can focus on the integrand in inequality (7). If we can choose another function so that these derivatives are close to -1 multiplied by these functions,[35] then

[34] Evidently, this is one of the barriers to entering this literature.

[35] Notice that these restrictions are on the derivatives of ψ_{mn} rather than on ψ_{mn} itself, so it is possible to make the derivatives negative while satisfying the non-negativity constraint on ψ. To be consistent with compact support, this requires an approximation to jumps in the derivative for example close to $\tau = 0$ or $\delta = 0$.

inequality (7) implies:

$$\left|\widehat{f}(m,n,\tau,\delta)-k\right|+$$
$$sign\left(\widehat{f}(m,n,\tau,\delta)-k\right)\cdot\left[\Phi_{mn}\left(\widehat{f}(m,n,\tau,\delta)\right)-\Phi_{mn}(k)\right]\cdot\frac{1}{\upsilon\left(\widehat{f}(m,n,\tau,\delta),x_{mn}\right)}=0$$

Dividing by $sign\left(\widehat{f}(m,n,\tau,\delta)-k\right)$, we obtain

$$\left(\widehat{f}(m,n,\tau,\delta)-k\right)+\left[\Phi_{mn}\left(\widehat{f}(m,n,\tau,\delta)\right)-\Phi_{mn}(k)\right]\cdot\frac{1}{\upsilon\left(\widehat{f}(m,n,\tau,\delta),x_{mn}\right)}=0$$

Now choose $k_h=\widehat{f}(m,n,\tau-\frac{1}{h},\delta)$ for $h=1,2,3,...$ Then dividing by $\frac{1}{h}$ and taking limits as $h\to\infty$ yields

$$\frac{\partial\widehat{f}(m,n,\tau,\delta)}{\partial\tau}+\Phi'_{mn}\left(\widehat{f}(m,n,\tau,\delta)\right)\cdot\frac{\partial\widehat{f}(m,n,\tau,\delta)}{\partial\tau}\cdot\frac{1}{\upsilon\left(\widehat{f}(m,n,\tau,\delta),x_{mn}\right)}=0$$

This expression is the same as (3).

Theorem 1: Suppose that υ satisfies Assumption 1 (so that Φ_{mn} is Lipschitz) and that feasible routes are restricted to satisfy Assumption 2. Suppose further that for all $m,n\in\mathcal{N}$ $(m\neq n)$ flux Φ_{mn} satisfies $f-f'\leq\xi\cdot[\Phi_{mn}(f)-\Phi_{mn}(f')]$ for $f>f'$, where $\xi>0$,[36] *and that both $\Phi_{mn}^{-1}\left(\frac{\mu(C)}{2\tau}\right)\neq\emptyset$*[37] *and $\Phi_{mn}(0)=0$. Then to each dynamic commuting route structure (τ^d,l,R), there corresponds a unique solution $\left\{\widehat{f}(m,n,\cdot,\cdot),\widehat{\delta}_{mn}\right\}_{m,n=1,\ m\neq n}^{N}$.*[38]

Remark 4: The case where Φ_{mn} is weakly increasing and Lipschitz, as in the examples, will be dealt with when existence of equilibrium is considered. For technical reasons, it is easiest to consider this case as a limit of the cases where Φ_{mn} satisfies the conditions of Theorem 1. Bressan (2000, p. 2) suggests an example where Φ is strictly increasing:

$$\Phi(f)=a_1\left(\ln\frac{a_2}{f}\right)f\qquad(0\leq f\leq a_2)$$

[36] A sufficient condition is: Φ_{mn} is C^1 with $\Phi'_{mn}>0$.

[37] This condition is sufficient, but not necessary, as illustrated by the example in the introduction.

[38] Given the definitions of a solution to the conservation law both here and in the mathematics literature, the solution $\widehat{f}(m,n,\cdot,\cdot)$ is only unique up to sets of measure zero in $[0,\overline{t}]\times[0,\lambda(m,n)]$. This is important but rarely noted.

Remark 5: One issue concerning our system is how we define a solution. Our system in $\widehat{f}$ is generally rather discontinuous, so it requires special treatment. There are alternatives to the technique we use, which we consider to be the most straightforward given our framework. One such alternative is to assume that the flux function Φ_{mn} is smooth and either strictly convex or strictly concave.[39] The conservation law is then called *strictly hyperbolic*; see Bressan (2000), particularly section 10.2. We can then define a *Filippov solution* (Filippov, 1973) to this problem, that was introduced into economics by Ito (1979).[40] Colombo and Marson (2003) and particularly Marson (2004) can be applied to obtain existence and uniqueness of a solution.[41] However, we do not place further restrictions on the flux.

Remark 6: It is important to discuss the assumption $\Phi_{mn}^{-1}\left(\frac{\mu(C)}{2\overline{\tau}}\right) \neq \emptyset$. We are taking departure strategies as times of departure. This leads naturally to boundary conditions for the initial link of a route that are phrased in terms of cars per hour, or traffic volume. To start cars on a route, we must rephrase this in terms of density, cars per mile, so that the conservation law (3) can be applied. Since volume at zero density must be zero by definition (4), this assumption ensures that for every volume that is possible as a boundary condition for the initial link, there is a density that will generate it. In the case where the density $\Phi_{mn}^{-1}(\cdot)$ is not unique, the density selected from the inverse must be specified as part of the equilibrium concept.

Remark 7: There is an interesting conceptual issue regarding the transition between links on a route at nodes. Depending on what one wants to conserve in passing from one link to the next, either volume (cars per hour) or density (cars per mile), the transition could be different or even impossible. We note that the entire literature takes the position that it is density, not volume, that should be conserved, since boundary conditions are always phrased in terms of density. Consider, for example, $\nu(f) = \frac{x}{f}$. Then volume is constant at x, whereas density can be any positive number. For this example, transitions between links with differing x that preserve volume could

[39] Notice that the triangular flux function, often used in transportation engineering, is neither smooth nor strictly concave.

[40] Formally speaking, we could introduce the general definition of a Filippov solution and then show that there exists one with finite total variation, but here we follow Columbo and Marson (2003) and Marson (2004) who skip this step because this fact is already well-known.

[41] In fact, Strub and Bayen (2006) use a strictly concave flux function in their application in section 5 to the I-210 in Los Angeles. Thus, they could have used a Filippov solution instead of a weak entropy solution for their application.

be impossible. Nevertheless, we take the position here that it is volume that is conserved when passing through a node, as given in (9), as it seems more natural and appropriate. Thus, speed functions such as $\nu(f) = \frac{x}{f}$ are excluded by the assumption $\Phi_{mn}(0) = 0$.[42]

The proof of Theorem 1 can be found in the Appendix. Formally, we prove that for given departure times and route choices the system behavior given by $\left\{ \left(\widehat{f}(m,n,\cdot,\cdot), \widehat{\delta}_{mn}(\cdot,\cdot) \right) \right\}_{m,n=1}^{N}$ exists and is uniquely defined. To accomplish this, we apply Schauder's theorem in a slightly unorthodox manner to the set of boundary conditions for each node, where the boundary conditions lie in the space of functions of bounded variation with respect to time. More precisely, the boundary conditions give the density of cars at the start of a link at a particular time, as in (9) or $\widehat{f}(m,n,\cdot,0)$. We could alternatively use volume for the fixed point instead of density, but since volume and density are in one-to-one correspondence under the assumptions of Theorem 1, there is little difference. An added complication in the proof is that although we consider only step functions as admissible boundary conditions at the start of any route, we do not know *a priori* (before Theorem 1 is proved) that the profile of arrivals at the end of a link, used as part of the boundary condition for the next link on a route, is a step function.[43] Thus, to prove that a solution exists, we must define the fixed point map for any boundary condition of bounded variation, not just step functions.

Next we examine existence of Nash equilibrium in pure strategies in our dynamic model.

The *time cost of a dynamic commuting structure* (τ^d, l, R) *for commuter* c is $\int_{-\overline{\tau}}^{\overline{\tau}} \frac{\widehat{\tau}_{l(c)}(c,\tau) - (\tau^d(c)+\tau)}{2\overline{\tau}} d\tau$. This is the expected time cost taken over all perturbations of departure time.

Fix an arrival time $\tau^A \in [0,\infty]$. Next we introduce the *arrival penalty function* $P : \mathbb{R}_+ \to \mathbb{R}_+$. To give intuition, think of $\tau = \tau_{l(c)}(c)$. The general arrival penalty is a function given by

$$P(\tau) \geq 0 \text{ where } P(\tau^A) = 0$$

Next, we turn to some examples. In the introduction we required that:

$$\text{Almost surely for } c \in C,\ \widehat{\tau}_{l(c)}(c) \leq \tau^A$$

[42] There might also be some interesting, unexplored duality between links/density on the one hand, and nodes/volume on the other.

[43] With links of varying length and capacity, even these step functions can be very complex.

Thus, $P(\tau) = 0$ for $\tau \leq \tau^A$ whereas $P(\tau) = \infty$ for $\tau > \tau^A$. It is actually more common in the literature to use an asymmetric linear penalty function; see Arnott et al (1993). Such a specification will be used in Example 2 below. We can allow further generalization, for example heterogeneous required arrival times τ^A, but at the cost of messier notation. Note that in the framework with a finite number of departure times, the penalty is actually the *expectation* of P for the given choice of strategy, since commuters are randomly assigned using a uniform distribution over a small departure time interval. We make this precise in equation (10) next.

The individual payoff function for the dynamic model is thus:

$$u(c; \tau^d, l, R) \equiv -\int_{-\overline{\tau}}^{\overline{\tau}} \frac{\widehat{\tau}_{l(c)}(c, \tau') - (\tau^d(c) + \tau') + P(\widehat{\tau}_{l(c)}(c, \tau'))}{2\overline{\tau}} d\tau' \tag{10}$$

The *utilitarian welfare function for the dynamic model* is

$$U(\tau^d, l, R) =$$
$$-\int_C \int_{-\overline{\tau}}^{\overline{\tau}} \frac{[\widehat{\tau}_{l(c)}(c, \tau') - (\tau^d(c) + \tau') + P(\widehat{\tau}_{l(c)}(c, \tau'))]}{2\overline{\tau}} d\tau' d\mu(c)$$

A *Nash equilibrium in pure strategies of the dynamic model* is a dynamic commuting structure (τ^d, l, R) with associated solution[44] $\left\{\widehat{f}(m, n, \cdot, \cdot), \widehat{\delta}_{mn}\right\}_{m,n=1,\ m\neq n}^{N}$ such that almost surely for $c \in C$, there is no route r of length $\ell \leq N$ and departure time $\tau^{d\prime}$ for commuter c such that, computing arrival times $\widehat{\tau}'$ as in Theorem 1 for the new route and departure time,

$$\int_{-\overline{\tau}}^{\overline{\tau}} \frac{\widehat{\tau}_{l(c)}(c, \tau') - (\tau^d(c) + \tau') + P(\widehat{\tau}_{l(c)}(c, \tau'))}{2\overline{\tau}} d\tau' > \int_{-\overline{\tau}}^{\overline{\tau}} \frac{\widehat{\tau}'_{\ell}(c, \tau') - (\tau^{d'} + \tau') + P(\widehat{\tau}'_{\ell}(c, \tau'))}{2\overline{\tau}} d\tau'$$

We note that due to the congestion externality, the Nash equilibria are unlikely to be Pareto (or utilitarian) optimal. Example 2 below will make this precise.

Next, in Theorem 2, we shall prove existence of Nash equilibrium in pure strategies for our model with discrete and finite departure times by applying Schmeidler (1973, Theorems 1 and 2). For the model with a continuum of departure time strategies, we can only obtain existence of ϵ-equilibrium in pure strategies. It is also worth noting that since Φ_{mn} will not be required to be

[44] In the case where Φ_{mn} is not strictly increasing, the definition of a solution must be modified slightly as detailed just below.

strictly increasing, we must modify (9) to:

$$\rho_{mn}(\tau) \in \Phi_{mn}^{-1}\left(\frac{\mu(\{c \in C \mid \pi_1(R(c)) = m,\ \pi_2(R(c)) = n,\ \left|\tau^d(c) - \tau\right| < \overline{\tau}\})}{2\overline{\tau}} + \sum_{m'=1,\ m' \neq m}^{N} \Phi_{m'm}\left(\widehat{f}(m', m, \tau, \lambda(m', m))\right)\right)$$

This adds another layer of complication to Nash equilibrium in the case where Φ_{mn} is not strictly increasing. We shall return to this important issue after Example 2, where it will be made less abstract.

Theorem 2: Under Assumptions 1-3, if the penalty function P is continuous[45] *and for all $m, n \in \mathcal{N}$ $(m \neq n)$ $\Phi_{mn}^{-1}\left(\frac{\mu(C)}{2\overline{\tau}}\right) \neq \emptyset$ and $\Phi_{mn}(0) = 0$, there exists a Nash equilibrium in pure strategies.*

One can prove that a utilitarian optimum exists for the discrete departure time model under the assumptions of Theorem 2. Instead of looking at a continuum of individual strategies, give the social planner the control variables that are the measure of commuters using each route at each departure time. The control vector is finite-dimensional. Assume, to begin, that Φ_{mn} is strictly increasing in f. Under the assumptions used in Theorem 1, densities and utility levels are well-defined for each departure and route strategy profile. In the proof of Theorem 2, found in the Appendix, it is shown that destination arrival times are continuous in the departure and route strategy profile. Thus, the utilitarian objective is continuous as a function of the measure of commuters using each route and departure time, so an optimum exists.

Consider next the case where Φ_{mn} is non-decreasing in f. As usual, take a sequence of initial conditions converging to the supremum. These initial conditions are in terms of volumes and routes, but there exists associated departure densities (per mile instead of per minute) associated with these volumes such that the supremum is approached. In the proof of Theorem 1, the only use made of Φ_{mn} strictly increasing in f is to prove that $\widehat{f}$ is unique, so there is an associated sequence of densities such that the optimum is approached. Following the remainder of the proof of Theorem 2 (that proves continuity of the objective in the strategy profile), and applying the dominated convergence theorem, the optimum will be achieved in the limit.

[45] Although some of our examples, such as the one in the introduction, feature a discontinuous P, a nearby continuous P with sufficiently steep slope just after the arrival time would work just as well, but would distract from the point of the example.

Example 2: What does Nash Equilibrium look like in the case of a linear penalty function? This is important for applications, as much of the literature uses such a specification. It is actually quite interesting. Suppose that

$$P(\tau) = \begin{cases} \eta \cdot (\tau^A - \tau) \text{ if } \tau^A \geq \tau \\ \zeta \cdot (\tau - \tau^A) \text{ if } \tau \geq \tau^A \end{cases}$$

where η, $\zeta > 0$. Notice that the late work arrival penalty ζ should generally be larger than the early arrival penalty η. It would be logical to flip these coefficients for the commute home, but in some cases it might also make sense to use departure penalties rather than arrival penalties for the commute home.

(a) To fix ideas, we consider the example from the introduction, with one link and two nodes, modified for this penalty function. Capacity of the link is $x = 1$, whereas travel time on the uncongested link is 1. At a Nash equilibrium, utility must be equalized across commuters, for otherwise everyone will imitate the happiest ones only. Fortunately for urban economists, this is a familiar condition. There is mass $\frac{3}{2}$ of identical commuters. Consider an example with 2 departure times, $\frac{1}{2}$ and $\frac{3}{2}$. Those who choose departure time $\frac{1}{2}$ actually leave at a random time distributed uniformly between 0 and 1, whereas those who choose departure time $\frac{3}{2}$ actually leave at a random time distributed uniformly between 1 and 2. Let $\tau^A = \frac{7}{2}$ and $\eta = \zeta \leq \frac{1}{3}$. It will turn out that in a Nash equilibrium, the commuters who choose departure time $\frac{1}{2}$ travel at the speed limit, whereas the commuters who leave at time $\frac{3}{2}$ travel slower and arrive later. The volume Φ of departures will be $\frac{1}{2}$ on $[0, 1]$ but 1 on $[1, 2]$. Suppose the density of commuters who choose departure time $\frac{1}{2}$ is $\frac{1}{2}$, whereas the (endogenous) density of commuters who leave at time $\frac{3}{2}$ is called $w' > 1$. For those who choose departure time $\frac{1}{2}$, their travel time is 1 whereas their *expected* early arrival penalty is 2η. For those departing at time $\frac{3}{2}$, their travel time is w' whereas their *expected* early arrival penalty is $\eta \cdot (\frac{7}{2} - (w' + \frac{3}{2}))$. Setting these negative utilities equal to each other, we obtain $w' = \frac{1}{1-\eta}$. Notice that, similar to Example 1, we can create a Pareto improvement by reducing the density of agents departing at time $\frac{3}{2}$ from w' to $\frac{3}{4}$. Their volume decreases to $\frac{3}{4}$ and their speed increases to 1. Density and volume for those departing at time $\frac{1}{2}$ increase from $\frac{1}{2}$ to $\frac{3}{4}$, but their speed remains 1. The payoff to those departing at time $\frac{1}{2}$ remains $-(1 + 2\eta)$, but the payoff to those departing at time $\frac{3}{2}$ increases to $-(1 + \eta)$. This disrupts the equal utility condition.

(b) To examine determinacy of Nash equilibrium, consider next the case where the total measure of consumers is 2. Nash equilibrium volume will be

constant at 1 on $[0,2]$. Density and speed are 1 for commuters departing on $[0,1]$. Density for commuters departing on $[1,2]$ is $\frac{1}{1-\eta}$, so speed is $1-\eta$. More generally, keeping volume the same, we can increase the density on $[0,1]$ to $\beta \geq 1$, $\beta \leq \frac{3}{2} - \frac{\eta}{1-\eta}$. Then the payoff to commuters departing in that interval is $-(\beta(1-\eta)+3\eta)$. Nash equilibrium density for departures on $[1,2]$ will be $\beta + \frac{\eta}{1-\eta}$, so there is a continuum of Nash equilibria indexed by β. The equilibria are Pareto ranked, and the equilibrium with $\beta = 1$ is best. To examine how equilibrium depends on the grid, we modify the example to allow any choice of departure time. Take $\eta = \zeta \leq \frac{1}{5}$. Nash equilibrium volume will again be constant at 1 on $[0,2]$, but starting density will vary; it is $1+\frac{\eta}{1-\eta}\cdot\tau$ for $\tau \in [0,2]$. The total cost of travel for a commuter departing at time τ inclusive of penalty is $1+\frac{\eta}{1-\eta}\cdot\tau+\eta(\frac{7}{2}-\tau-(1+\frac{\eta}{1-\eta}\tau)) = 1+\frac{5}{2}\eta+\tau(\frac{\eta}{1-\eta}-\eta-\frac{\eta^2}{1-\eta}) = 1+\frac{5}{2}\eta$, independent of departure time τ. If the last arrival is to occur exactly at $\frac{7}{2}$ (so the constraint on departure times does not bind), take $\eta = \frac{1}{5}$. More generally, other equilibria have departure density as $\beta+\frac{\eta}{1-\eta}\cdot\tau$ for $\beta \geq 1$. There remains a large number of equilibria. The payoff to every agent is $-(\beta\,(1-\eta)+\frac{7}{2}\eta)$, so equilibria are again Pareto ranked. If the last arrival is to occur exactly at $\frac{7}{2}$, then β and η satisfy $\beta = \frac{3}{2} - \frac{2\eta}{1-\eta}$.

(c) To explore further the nature of equilibria for this example, notice that if we allowed departures at time $5/2$, each commuter would unilaterally deviate to this departure time, since travel time would be 1 whereas the penalty would be 0, so net utility would be higher than equilibrium utility. One way to rectify this is to impose a arrival penalty of at least $1/2$ for all arrivals at or after time $7/2$. This would allow the equilibria detailed in (b) above to persist even under more general departure times.[46] An alternative to this penalty function would employ a larger mass of agents and a single peaked departure density with the peak at time 2. In that case, the late arrival penalty ζ trades off against a lower density for later departure times. The commuters departing after time 2 will catch up to the commuters departing before them and slow down. However, their commuting time will be shorter than those departing at time 2 due to a lower initial density. Calculations for this modification are quite messy either with continuous or discrete departure times, as they involve moving thresholds.[47]

[46] The reader will notice that this is not quite true, since the choice of departure time $\frac{5}{2}$ implies randomization over the departure time interval $[2,3]$. But inclusion of randomization in this argument simply requires that the penalty for late arrival be increased to at least 1.

[47] An interesting proposal for equilibrium selection for the model in part (b) is to allow departures at negative times, and take the limit as the allowable negative departure times

Note that the static model cannot generate equilibria that feature departure density monotone increasing in time.

The dynamic commuting game and the Nash equilibrium concept in our context has more in common with those of a generalized game than standard games; see Debreu (1952) for the origin of generalized games. The reason is that, given a strategy profile for all players, departure volumes and routes are determined, but if volume and density are not one to one, then departure density can be any density that yields the given departure volume. This is highlighted in Example 2. In other words, the choice of (pure) strategy by all players does not completely determine their payoff. The way to address this issue is to include a departure density (consistent with departure volume) in the equilibrium concept, as we have done, so that commuters can calculate their payoff for any strategy they choose given the aggregate departure volume and density. An individual commuter's choice does not alter either the departure volume or density. This is a big advantage of using a non-atomic model of commuters.

3 Applications

3.1 Can the Static Equilibrium be Supported by a Dynamic Equilibrium?[48]

Given identical exogenous data for the static and dynamic commuting games, are Nash equilibrium densities in the static and dynamic models the same? In other words, is the static model a reduced form of the dynamic model? This is important for addressing the issue of whether the static model makes sense. For if the answer to this question is negative, then there should be no interest in the static model, since its equilibrium behavior is different from the analogous dynamic model, and the real world is dynamic.

For simplicity, we return now to the examples used in the introduction, namely where there is no penalty for early arrival and an infinite penalty for late arrival. One could imagine that the static model represents some sort of steady state of the dynamic model, where commuters are introduced at constant volumes and densities at all the nodes, and the densities in the links

tends to zero. This would select the equilibrium where the.departures at time 0 must travel at the maximum speed.

[48]The ideas in this subsection owe much to Anas (2007) and to discussions with Alex Anas.

are constant over time. But with a fixed arrival time (say 9 AM), a steady state does not make sense. The time profile of equilibrium departures will generally not be constant over time, since everyone must get to work by the arrival time. Even if arrival time varied by commuter, one would not expect to see a steady state attained.

For brevity in examining the equilibria of the static and dynamic models, we use the example in the introduction and compare equilibrium speeds and travel times in the two models.

Example 3: Take $x = \frac{1}{2}$ for the example in the introduction, and take the arrival time to be $\tau^A = 3$. For the static model, speed is $\frac{1}{2}$ and time on the link for each commuter is 2. For the dynamic model, the congested commuting pattern is no longer a Nash equilibrium, because the last commuter arrives at time 4. The uncongested commuting pattern remains a Nash equilibrium, but features speed 1 and travel time 1.[49]

There is a clear trade-off in constructing the dynamic model, the point of this short subsection. Do we want a model closer to the static model in terms of equilibrium, or do we want a model closer to reality?

Verhoef (1999) studies a similar problem in a very different class of models, and concludes (p. 365) that, "For static models of peak demand, it was argued that for such models to be dynamically consistent, rather heroic assumptions on the pattern of scheduling costs have to be made."

3.2 Welfare Properties of Nash Equilibrium

Equilibrium selection is an important issue in one shot congestion games with Pigouvian congestion taxes. Under such taxes, there can be multiple Nash equilibria, only some of which are efficient.[50] As remarked in the introduction, Sandholm (2007) shows that with a finite number of commuters, an evolutionary process, and Pigouvian taxes, the outcome will be efficient. A major limitation of this work is the assumption of a common utility function with idiosyncratic perturbations, which seems to rule out heterogeneous origins and destinations.

[49] Without an arrival time, it's easy to argue that neither the static nor the dynamic model is a reasonable model of the morning commute.

[50] We do not provide an example here, both because they are available in the literature (for more macro models) and because, as will be apparent from Theorem 3, examples in our framework with non-constant (or non-zero) Pigouvian taxes will have relatively complicated route structures. For instance, a one link example won't work.

Although that approach is clearly interesting, we take a completely different approach here, motivated by our examples. A major advantage of our approach is that we can compare non-trivial commutes (home to work) with their reverses (work to home), to our knowledge absent in the literature. As we wish to focus on departure times rather than routes in the dynamic model, we discuss the following restrictions:[51]

Definitions: *An* ***outbound shrubbery network*** *is a set of routes* $\mathcal{R}^{out} \subseteq \mathfrak{R}$ *such that for any* $r \in \mathfrak{R}^{\ell} \cap \mathcal{R}^{out}$, $r' \in \mathfrak{R}^{\ell'} \cap \mathcal{R}^{out}$, $\pi_1(r) = \pi_1(r')$ *and there do not exist* $1 < i < \ell$ *and* $1 < i' < \ell'$ *with* $\pi_{i-1}(r) \neq \pi_{i'-1}(r')$ *and* $\pi_i(r) = \pi_{i'}(r')$. *An* ***inbound shrubbery network*** *is a set of routes* $\mathcal{R}^{in} \subseteq \mathfrak{R}$ *such that for any* $r \in \mathfrak{R}^{\ell} \cap \mathcal{R}^{in}$, $r' \in \mathfrak{R}^{\ell'} \cap \mathcal{R}^{in}$, $\pi_{\ell}(r) = \pi_{\ell'}(r')$ *and there do not exist* $1 < i < \ell$ *and* $1 < i' < \ell'$ *with* $\pi_i(r) = \pi_{i'}(r')$ *and* $\pi_{i+1}(r) \neq \pi_{i'+1}(r')$.

These are generalizations of inbound and outbound (directed) tree networks. The difference is simply that we allow only the outermost node to be the same for two or more branches, possibly forming a loop. Simple examples will be given below.

In terms of commuting, an inbound shrubbery network might be a reasonable model of commuting from home to work, whereas an outbound shrubbery network might be a reasonable model of commuting from work to home. In terms of electronic networks, this might not be a good model of the internet, but tree structures are often used in local area networks. The property of interest for an outbound shrubbery network is preventing mergers of routes at nodes where traffic continues together along the next link.

Why are we introducing a restrictive condition like this? It is one way to sort out the efficiency properties of Nash equilibrium in our dynamic model. What is perhaps strange but interesting is that on a two way network, commuting to work may be inefficient, whereas commuting to home might be efficient. In other words, reversing the commute on a directed network can change the efficiency properties of Nash equilibrium.

Notice that both types of shrubbery networks satisfy Assumption 2: $\mathfrak{R} \cap \mathcal{R}^{out} \subseteq \mathcal{R}$, $\mathfrak{R} \cap \mathcal{R}^{in} \subseteq \mathcal{R}$.

We wish to examine the similarities and differences between commuting from home to work and commuting from work to home. Since networks are arbitrary in our general framework, we focus on shrubbery, and begin our analysis with an example. Most of the intuition can be gleaned from this

[51] Extensions will be discussed at the end of this subsection.

example. What is important for our purposes is asymmetry.

Example 4:[52] First, consider the commute from a common home location O on the right to a common work location D on the left, via through either node A or node B followed by a merge at node C, as represented in Figure 3:

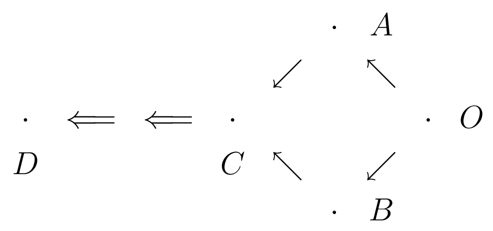

Figure 3: Example of an inbound shrubbery network

For expositional clarity, consider the path OAC to be just one link, and similarly for OBC; we have inserted nodes A and B only to distinguish between the two paths. Suppose that rush hour is from time 0 to time 1 with two possible departure times: 1/4 and 3/4, where commuters choosing the first departure time are uniformly distributed over actual departure times $[0, 1/2]$ and commuters choosing the second departure time are uniformly distributed over actual departure times $[1/2, 1]$. The length of links OAC and OBC is 1, whereas the length of link CD is 2. Speed on links between nodes O and C is given by $\min\left\{\frac{1}{f}, 1\right\}$. Speed on the link between nodes C and D is given by $\min\left\{\frac{1}{f}, 1\right\} + \frac{1}{8}$. There is no arrival time penalty; it's not very natural when comparing a commute and its reverse, though the examples and theorem can likely be extended in this direction. There is measure 2 commuters travelling from node O to node D.

Let's first examine Nash equilibrium. Consider the symmetric strategy profile where measure 1/2 commuters choose route OAD and choose departure time 1/4 and thus are uniformly distributed with volume 1 over $[0, 1/2]$, whereas measure 1/2 commuters choose route OAD and choose departure time 3/4 and thus are uniformly distributed with volume 1 over $[1/2, 1]$. As always, we must be careful about how volume (cars per hour) translates into density

[52]Notice that the network used in this example is extension parallel. For the static model, the Nash equilibrium is weakly Pareto efficient. For the dynamic model, it depends on the direction of commute.

(cars per mile), particularly at nodes. In this case, we set density $f = 1$, so speed is 1. Similarly, commuters choosing route OBD are split: measure $1/2$ choose departure time $1/4$ and are uniformly distributed over $[0, 1/2]$ with volume 1, whereas measure $1/2$ choose departure time $3/4$ and are uniformly distributed over $[1/2, 1]$ with volume 1. Again, density $f = 1$. At the merge node C, volume is 2. On the link CD, set density $f = 8$, so speed is $\frac{1}{4}$. Each commuter experiences a total travel time of 9: travel time is 1 on the initial link, and 8 on the link between node C and node D.

Next consider the following slightly asymmetric strategy profile that will not be a Nash equilibrium. The volume and density departure schedule for commuters who use link OAC remains the same as above. Commuters who use link OBC will have the following departure schedule. Measure $\frac{1}{2}$ depart at time $1/4$. For those who depart at time $1/4$, volume is 1, the departure density is 2, and initial speed is $\frac{1}{2}$. Measure $\frac{1}{2}$ depart at time $3/4$. For those who depart at time $3/4$, volume is 1, the departure density is 2, so initial speed is $\frac{1}{2}$. The first cohort to arrive at node C will be those using link OAC who depart in $[0, 1/2]$, and who arrive at node C at times in the interval $[1, 3/2]$. Next are the commuters using link OAC who depart in $[1/2, 1]$, and who arrive at node C at times in the interval $[3/2, 2]$. The next commuters to arrive are the commuters using link OBC who depart in the interval $[1/2, 1]$ and who arrive at node C at times in the interval $[2, 5/2]$. Finally, the cohort of commuters using link OBC in $[1/2, 1]$ arrive at node C at times in the interval $[5/2, 3]$. Notice that the overlap in arrival times at node C is of measure zero. At the merge node C, volume is 1, in contrast with the Nash equilibrium, where it is 2. So traffic can travel faster along this segment; set density $f = \frac{8}{9}$ and speed at $\frac{9}{8}$. Therefore, in the end, travel time for all commuters using link OAC is $1 + \frac{9}{4} = \frac{13}{4} < 9$, whereas travel time for the commuters using link OBC is $2 + \frac{9}{4} = \frac{17}{4} < 9$. Clearly, this strategy profile Pareto dominates the Nash equilibrium strategy profile, but is not a Nash equilibrium itself since commuters using link OBC receive a lower utility level than those using link OAC.[53]

Next we reverse the commute. The structure of permissible departure times and speed as a function of density are the same. The new diagram is represented in Figure 4:

[53]If the utility function is linear in money, the Pareto efficient allocation could be implemented by charging a toll at node A that is equal to the difference in travel time between the two routes.

$$
\begin{array}{cccccccc}
 & & & & & \cdot & A & \\
 & & & & \nearrow & & \searrow & \\
\cdot & \Longrightarrow & \Longrightarrow & \cdot & & & & \cdot\ O \\
D & & & C & \searrow & & \nearrow & \\
 & & & & & \cdot & B &
\end{array}
$$

Figure 4: Example of an outbound shrubbery network

A Nash equilibrium and utilitarian optimal strategy profile has measure 1/2 of each type (for a total of 1) departing work at node D at each of the two departure times. The departure volume is 2, the departure density is 8, whereas initial speed is $\frac{1}{4}$. Travel time on link DC is 8. Volume on the second link (either CAO or CBO) is 1 whereas density is 1, so the speed is 1. Time spent on the second link is 1, so the total travel time of each commuter is 9. The next theorem makes this example more general:

Theorem 3: With an outbound shrubbery network, suppose that for all $r, r' \in \mathfrak{R}^\ell \cap \mathcal{R}^{out}$, $\pi_{\ell-2}(r) = \pi_{\ell-2}(r')$, $\pi_{\ell-1}(r) \neq \pi_{\ell-1}(r')$, $\pi_\ell(r) = \pi_\ell(r')$ *implies* $x_{\pi_{\ell-2}(r)\pi_{\ell-1}(r)} = x_{\pi_{\ell-2}(r')\pi_{\ell-1}(r')}$ *and* $x_{\pi_{\ell-1}(r)\pi_\ell(r)} = x_{\pi_{\ell-1}(r')\pi_\ell(r')}$. *Under Assumption 1, assuming*[54] $\frac{\partial v}{\partial f} < 0$, *for all* $m, n \in \mathcal{N}$ $(m \neq n)$, $\Phi^{-1}_{mn}\left(\frac{\mu(C)}{T}\right) \neq \emptyset$,[55] $\Phi_{mn}(0) = 0$, *and* $P = 0$, *there is a Pareto optimal*[56] *strategy profile that is also a Nash equilibrium. Thus, there exists an efficient Nash equilibrium.*

The proof of Theorem 3 is in the Appendix. Thus, under these additional assumptions, efficiency can be achieved not through taxes, but by equilibrium selection. Prisoners' dilemma problems are ruled out by the structure of the game, specifically these additional assumptions. The first half of Example 4 is an inbound shrubbery network that has no efficient Nash equilibrium, so an analog of Theorem 3 for an inbound shrubbery network is impossible.

Since the simple example with two nodes and one link from the introduction is trivially an outbound shrubbery network, it cannot be true that all Nash

[54] Sharp-eyed readers will notice that this condition is not satisfied by Example 4. We can weaken this condition to: $\frac{\partial v}{\partial f}_+ |_{\overline{g}} < 0$, where $+$ denotes the right-hand derivative of the function and $|_{\overline{g}}$ denotes evaluation of the derivative at the departure time profile that assigns equal volume and density to all departure times and routes for a given type of commuter.

[55] As in Theorem 1, this condition is sufficient but not necessary.

[56] Although we have not defined it formally, Pareto optimum is the usual concept in our context of a continuum of agents.

equilibria are efficient. For such a simple example as well as for outbound shrubbery networks (the evening commute) more generally, congestion pricing is unnecessary if equilibrium can be selected, for example by using flow control. In contrast, congestion pricing seems necessary for other commutes.

The result can likely be extended, for example allowing limited asymmetry in the final links of routes.

Finally, the larger implications of this subsection are important for the comparison of the dynamic and static frameworks. It is hard to imagine an analog of Example 4 in a static model. As should be apparent from the examples, the normative properties of the dynamic and static models as well as the number and variety of equilibria differ. A great deal of work remains to be done on empirical and experimental approaches to the comparison of the models, for example in equilibrium selection.

4 Conclusions

We have asked and answered several questions about commuting using two models, one static and one dynamic. For each model, we have shown that a Nash equilibrium in pure strategies exists for the one shot game, that a Pareto optimum exists, and that Nash equilibrium is generally not Pareto optimal. Beyond that, we have shown that all Nash equilibria of the static model can look very different from any Nash equilibrium of the dynamic model. Since the static model features behavior unlike the dynamic one, we reject the former as a reduced form of the latter and stick with the dynamic model. Finally, we have examined the welfare properties of Nash equilibrium in the particular case of a shrubbery network, and found that equilibrium might not be efficient for the morning commute, but under some conditions there always is an efficient Nash equilibrium for the evening commute. Thus, congestion pricing is more important for the morning commute, whereas equilibrium selection (perhaps via flow control) is more important for the evening commute. Further effort should be devoted to discovering the welfare properties of Nash equilibrium on specific directed networks. In sum, *what we have shown is that a model of congestion using microfounded behavior has very different properties from the reduced form models used in the literature.*

Our commuting model can be reinterpreted as a model of internet congestion. In this context, local area networks often have a tree or shrubbery structure, so for example the results on efficiency of Nash equilibrium and the consequences for congestion tolls can be reinterpreted in this framework. Much work remains to flesh out the application to the internet. At the micro level, routes are chosen by the TCP/IP software as a proxy for the user, though the user chooses the time of day. At the macro level, whereas we have only considered small users, it is likely that the supply side involves strategic and competitive large players, such as internet service providers and content providers. Moreover, there is likely asymmetric information, for example the reason why internet speed might be slow might be unknown to some end users. Malone et al (2017) offers some interesting insights into these issues. Questions about equilibrium and optimum, like those put forth here, should be addressed.

A natural question is: Are the Nash equilibria of the model stable under finer departure grids? The equilibria that have a uniform distribution of departures and arrival penalties survive no matter how fine the time grid is, even in the limit when commuters can choose their precise departure time. This includes the examples in the introduction, Example 3 and Example 4. It also includes the cases addressed in Theorem 3, with an outbound shrubbery network but no arrival time.

For simple examples, the Nash equilibria of our model can be solved analytically. For more complex examples, the proof of Theorem 1 indicates that a numerical solution technique involves nesting the solution of a discontinuous system of differential equations inside a fixed point solution algorithm.

In a companion paper to this research, Berliant (2020) examines the set of Nash equilibria in the infinitely repeated versions of both the static and dynamic commuting games, and the folk theorem is used to obtain these large sets. There we present some preliminary evidence from the shutdown of an expressway in St. Louis that commuters do not always play one shot Nash equilibrium. We also discuss the application of the anti-folk theorem to our specific game, namely conditions under which the Nash equilibria of the infinitely repeated game are the Nash equilibria of the one shot game.

Self-driving cars would represent another interesting application of the model. Given the detailed microstructure of the model, a centralized system of self-driving cars could compute and implement an efficient allocation. A more decentralized system could have some cars that are self-driving and others that are driven by humans. Nash equilibrium could be explored in this context. A useful reference for these issues, with an emphasis on ride-sharing and tolls, is Ostrovsky and Schwarz (2018).

In the same vein, the supply side of our model is passive. We have already mentioned how strategic behavior in the internet application on the supply side might be studied. Similarly, platform markets for ride services in the commuting application would allow strategic behavior, for example a monopoly. This would affect not only the characteristics of Nash equilibrium, but equilibrium selection as well. One might view ride matching and routing as a constrained planning problem. This can also be applied to bus times and routes, where the local government has market power.

Our model could be extended to allow elastic demand for travel to or from work. The extension of the model to allow land markets and endogenous choice of household residence and job location would also be interesting. The extension to multiple lanes of traffic and passing would be useful. In part, this can be accomplished by introducing more links between a pair of nodes, as in Example 4, but this alteration does not allow lane changes.

The dynamic model should be applied to real world commuting. Since it can accommodate an arbitrary (exogenous) route structure, it has both positive and normative content, especially regarding Pareto improvements. For example, it can be used to perform cost benefit analysis with respect to changing infrastructure and mass transit. More specifically, it could be used to examine adding lanes (increasing capacity x_i), adding public transit to reduce road demand, and tolling links. A prerequisite would be to incorporate elastic demand for trips into the model, since all of these alterations to the model could have large effects on demand. A first step would be the calculation of comparative statics in each of the exogenous variables.

References

[1] Anas, A., 2007. "The Trips-to-Flows Riddle in Static Traffic Equilibrium: How to drive a BMW?" Unpublished manuscript.

[2] Arnott, R., A. de Palma and R. Lindsey, 1993. "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand." *American Economic Review* 83, 161-179.

[3] Beckmann, M., C.B. McGuire and C.B. Winsten, 1956. *Studies in the Economics of Transportation.* Yale University Press: New Haven.

[4] Berliant, M., 2020. "Daily Commuting." Working Paper. https://mpra.ub.uni-muenchen.de/100169/1/MPRA_paper_100169.pdf

[5] Biles, D.C., M. Federson and R. López Pouso, 2014. "A Survey of Recent Results for the Generalizations of Ordinary Differential Equations." *Abstract and Applied Analysis* http://dx.doi.org/10.1155/2014/260409.

[6] Bressan, A., 1988. "Unique Solutions for a Class of Discontinuous Differential Equations." *Proceedings of the American Mathematical Society* 104, 772-778.

[7] Bressan, A., 2000. *Hyperbolic Systems of Conservation Laws: The One-Dimensional Cauchy Problem*. Oxford University Press: Oxford.

[8] Bressan, A. and W. Shen, 1998. "Uniqueness for Discontinuous ODE and Conservation Laws." *Nonlinear Analysis* 34, 637-652.

[9] Colombo, R.M. and A. Marson, 2003. "A Hölder Continuous ODE Related To Traffic Flow." *Proceedings of the Royal Society of Edinburgh* 133A, 759-772.

[10] Daganzo, C.F., 2008. *Fundamentals of Transportation and Traffic Operations.* Emerald Group Publishing: Bingley, UK.

[11] de Palma, A. and M. Fosgerau, 2011. "Dynamic Traffic Modeling." In *A Handbook of Transport Economics*, edited by A. de Palma, R. Lindsey, E. Quinet and R. Vickerman. Edward Elgar: Cheltenham, UK. Ch. 9, pp. 188-212.

[12] Debreu, G., 1952. "A Social Equilibrium Existence Theorem." *Proceedings of the National Academy of Sciences* 38, 886-893.

[13] Filippov, A.F., 1973. *Differential Equations with Discontinuous Right-hand Sides*. Kluwer: Dordrecht, Netherlands.

[14] Friedrich, J., O. Kolb and S. Göttlich, 2018. "A Godunov Type Scheme for a Class of Scalar Conservation Laws with Non-local Flux." Working paper.

[15] Garavello, M. and B. Piccoli, 2006. *Traffic Flow on Networks: Conservation Laws Model*. American Institute of Mathematical Sciences: Springfield, MO.

[16] Han, K., T.L. Friesz and T. Yao, 2013. "Existence of Simultaneous Route and Departure Choice Dynamic User Equilibrium." *Transportation Research Part B* 53, 17-30.

[17] Han, K., B. Piccoli and T.L. Friesz, 2015. "Continuity of the Path Delay Operator for LWR-Based Network Loading with Spillback." Working Paper.

[18] Heydecker, B.G. and J.D. Addison, 2005. "Analysis of Dynamic Traffic Equilibrium with Departure Time Choice." *Transportation Science* 39, 39-57.

[19] Hofbauer, J. and W.H. Sandholm, 2007. "Evolution in Games with Randomly Disturbed Payoffs." *Journal of Economic Theory* 132, 47-69.

[20] Hu, D., 2010. "Equilibrium and Dynamics of the Discrete Corridor Problem." Unpublished manuscript.

[21] Ito, T., 1979. "A Filippov Solution of a System of Differential Equations with Discontinuous Right-Hand Sides." *Economics Letters* 4, 349-354.

[22] Jacobsson, K., 2008. *Dynamic Modeling of Internet Congestion Control.* Doctoral Thesis, KTH School of Electrical Engineering: Stockholm.

[23] Kelly, F.P., A.K. Maulloo and D.K.H. Tan, 1998. "Rate Control for Communication Networks: Shadow Prices, Proportional Fairness, and Stability." *Journal of the Operational Research Society* 49, 237-252.

[24] Khan, M.A., 1989. "On Cournot-Nash Equilibrium Distributions for Games with a Nonmetrizable Action Space and Upper Semicontinuous Payoffs." *Transactions of the American Mathematical Society* 315, 127-146.

[25] Konishi, H., 2004. "Uniqueness of User Equilibrium in Transportation Networks with Heterogeneous Commuters." *Transportation Science* 38, 315-330.

[26] Malone, J.B., A. Nevo and J.W. Williams, 2017. "The Tragedy of the Last Mile: Economic Solutions to Congestion in Broadband Networks." Unpublished manuscript.

[27] Marson, A., 2004. "Nonconvex Conservation Laws and Ordinary Differential Equations." *Journal of the London Mathematical Society* 69, 428-440.

[28] Merchant, D.K. and G.L. Nemhauser, 1978. "A Model and an Algorithm for the Dynamic Traffic Assignment Problems." *Transportation Science* 12, 183-199.

[29] Milchtaich, I., 2006. "Network Topology and the Efficiency of Equilibrium." *Games and Economic Behavior* 57, 321-346.

[30] Ostrovsky, M. and M. Schwarz, 2018. "Carpooling and the Economics of Self-Driving Cars." NBER Working Paper 24349.

[31] Rath, K.P., 1994. "Some Refinements of Nash Equilibria of Large Games." *Games and Economic Behavior* 7, 92-103.

[32] Rath, K.P., 1998. "Perfect and Proper Equilibria of Large Games." *Games and Economic Behavior* 22, 331-342.

[33] Rosenthal, R.W., 1973. "A Class of Games Possessing Pure-Strategy Nash Equilibria." *International Journal of Game Theory* 2, 65-67.

[34] Ross, S.L. and J. Yinger, 2000. "Timing Equilibria in an Urban Model with Congestion." *Journal of Urban Economics* 47, 390-413.

[35] Sandholm, W.H., 2001. "Potential Games with Continuous Player Sets." *Journal of Economic Theory* 97, 81-108.

[36] Sandholm, W.H., 2007. "Pigouvian Pricing and Stochastic Evolutionary Implementation." *Journal of Economic Theory* 132, 367–382.

[37] Schmeidler, D., 1973. "Equilibrium Points of Non-Atomic Games." *Journal of Statistical Physics* 7, 295-301.

[38] Schrank, D., B. Eisele and T. Lomax, 2019. *2019 Urban Mobility Report.* Texas A&M Transportation Institute: College Station.

[39] Smart, D.R., 1974. *Fixed Point Theorems.* Cambridge University Press: Cambridge, UK.

[40] Strub, I.S. and A.M. Bayen, 2006. "Mixed Initial-Boundary Value Problems for Scalar Conservation Laws: Application to the Modeling of Transportation Networks." In *Hybrid Systems: Computation and Control*, Lecture Notes in Computer Science Volume 3927, edited by J. Hespanha and A. Tiwari, pp. 552-567. Springer Verlag: Berlin.

[41] Verhoef, E., 1999. "Time, Speeds, Flows and Densities in Static Models of Road Traffic Congestion and Congestion Pricing." *Regional Science and Urban Economics* 29, 341-369.

[42] Vickrey, W., 1963. "Pricing in Urban and Suburban Transport." *American Economic Review* 53, 452-465.

[43] Vickrey, W., 1969. "Congestion Theory and Transport Investment." *American Economic Review* 59, 251-261.

[44] Zhang, X. and H.M. Zhang, 2010. "Simultaneous Departure Time/Route Choices in Queuing Networks and a Novel Paradox." *Networks and Spatial Economics* 10, pp. 93-112.

[45] Zhu, D. and P. Marcotte, 2000. "On the Existence of Solutions to the Dynamic User Equilibrium Problem." *Transportation Science* 34, 402-414.

5 Appendix: Proofs

5.1 Proof of Theorem 1

Preliminaries: We want to find a unique fixed point in initial conditions at the start of a link over time, that we will call g, and progress along a link that we have defined as δ. The main issue is consistency of the commuting pattern with boundary values on links, namely the density of departures along a link from a node over time.[57] These initial conditions are partly exogenous, due to the fixed choice of departure times and routes for Theorem 1 (in contrast with Theorem 2), and partly endogenous, for nodes along a commuter's route that are not the point of origin. So we employ a fixed point on this data; it will be in a subspace of functions of bounded variation.

We have already defined the space where δ lives; see (6). Notice that for Theorem 1, flux Φ_{mn} is one to one. Recalling that each permissible route can go through a given node at most once (see Assumption 2), next we define the space of all possible boundary conditions, $\mathcal{G}$:

Definitions: Let ϕ_{mn} *be the Lipschitz constant for* Φ_{mn} and let $\phi \equiv \max_{m,n} \phi_{mn}$. *Fix* z, $0 < z \leq \frac{1}{\phi}$. *For* $g : [0, \bar{t}] \rightarrow \mathbb{R}_+$, *the total variation*

[57] For the proof of Theorem 1, we have assumed that volume is strictly increasing in density, so initial conditions can be phrased in terms of either.

norm is defined as:

$$TV(g) \equiv \sup \left\{ \sum_{k=0}^{K} |g(t_k) - g(t_{k-1})| \,\middle|\, K \geq 1,\, t_k \in [0, \bar{t}],\, t_0 < t_1 < \cdots < t_K \right\}$$

Next, define the lower bound on departure density different from zero:[58]

$$\underline{g} \equiv \min \left\{ \Phi_{mn}^{-1} \left(\frac{\mu(\{c \in C \mid R(c)) = r,\ \left|\tau^d(c) - \tau\right| < \overline{\tau}\})}{2\overline{\tau}} \right) \middle| m, n = 1, 2, ..., N; \right.$$

$$\left. r \in \mathcal{R}, \tau \in [0, \bar{t}],\ \pi_1(R(c)) = m, \pi_2(R(c)) = n, \mu(\{c \in C \mid R(c)) = r, \left|\tau^d(c) - \tau\right| < \overline{\tau}\}) > 0 \right\}$$

$$> 0$$

Number the equivalence classes of links defined by the relation $\succeq$, from the bottom class up, using the index $h = 1, 2, ..., H \leq N^2 - N$. Define $\chi_1 \equiv (T/\overline{\tau} + 1) \cdot \overline{f}$ and define inductively

$$\chi_{h+1} \equiv 2 \cdot |\mathcal{R}| \cdot (T/\overline{\tau} + 1) \cdot \overline{f} + \frac{\overline{f} \cdot \xi^2 \cdot \phi}{z} + \chi_h \cdot \left(\phi^2 \cdot \xi^2 + \frac{N \cdot \xi \cdot \phi \cdot \overline{\Phi}}{\underline{\Phi}} \right) \tag{11}$$

$$\text{where } \overline{\Phi} \equiv \max_{m,n} \Phi_{mn}\left(\overline{f}\right) \text{ and } \underline{\Phi} \equiv \min_{m,n} \Phi_{mn}\left(\underline{g}\right)$$

for $h = 1, 2, ..., H - 1$. For link mn, define $h(mn)$ as the equivalence class to which it is assigned. Define:

$$\mathcal{G}_m^r \equiv \begin{cases} \left\{ \Phi_{mn}^{-1} \left(\frac{\mu(\{c \in C | R(c) = r,\ \left|\tau^d(c) - \tau\right| < \overline{\tau}\})}{2\overline{\tau}} \right) \right\} \text{ if } \pi_1(r) = m,\ \pi_2(r) = n; \\ \left\{ g_m^r(\cdot) \text{ measurable on } [0, \bar{t}] \mid 0 \leq g_m^r(\cdot) \leq \overline{f},\ g_m^r(\bar{t}) = 0,\ TV(g_m^r(\cdot)) \leq \chi_{h(mn)}^r \right\} \\ \qquad \text{if } \pi_i(r) = m,\ \pi_{i+1}(r) = n \text{ for some } i > 1 \end{cases}$$

Finally, define:

$$\mathcal{G} \equiv \prod_{n=1}^{N} \prod_{\{r \in \mathcal{R} | \pi_i(r) = n \text{ for some } i\}} \mathcal{G}_n^r$$

We shall be searching for a fixed point in $\mathcal{G}$. So the next step is to define the map from $\mathcal{G}$ into itself.

We begin by fixing some $g \in \mathcal{G}$. For reasons explained in the text below equation (5), we alter it by inserting a small gap of size $\varsigma > 0$ after every downward jump, namely where $g_m^r(\tau^-) > g_m^r(\tau^+)$ To keep notation from

[58] The reason we can define $\underline{g}$ in this way is that, given that volume is strictly increasing in density, density cannot decrease from the start of a route, though it can increase.

getting out of hand, we will index g by ς only at the end when we let $\varsigma \to 0$.[59] The 2 in (11) accounts for the insertion of these gaps.

Definitions: For each link mn, $m \neq n$, we define

$$g(m,n,\cdot) \equiv \sum_{\{r \in \mathcal{R} | \pi_i(r)=m,\ \pi_{i+1}(r)=n \text{ for some } i \geq 1\}} g_n^r(\cdot)$$

To simplify notation, define

$$\chi_{mn} \equiv \chi_{h(mn)}$$

Notice that $TV(g(m,n,\cdot)) \leq \chi_{mn}$. After some preparation, we shall define the map $\mathcal{T} : \mathcal{G} \to \mathcal{G}$. We will call $\mathcal{T}(g) \equiv \widehat{g}$. Next we begin preparations for defining this map.

Given the initial condition

$$\widehat{f}(m,n,0,\Delta) = 0 \ \forall \ \Delta \geq 0$$

and the left boundary condition on each link mn, $m \neq n$: $\widehat{f}(m,n,\tau,0) = g(m,n,\tau)$, Strub and Bayen (2006) yields existence of a unique solution (as we have defined it) called $\widehat{f}(m,n,\tau,\Delta)$ of bounded variation on $(0,\bar{t})\times(0,\lambda(m,n))$. We must be a little careful here, specifically at the right boundary $\lambda(m,n)$. Although they only use the solution on $(0,\bar{t}) \times (0,\lambda(m,n))$, as they remark, it is in fact defined on $[0,\bar{t}] \times [0,\lambda(m,n)]$. All we need is that it is defined on $(0,\bar{t}) \times (0,\lambda(m,n)]$. Next, to make the right boundary condition non-binding, we simply set (in their notation) $\rho_b(t) = 0$. Then the right boundary condition becomes vacuous.[60] The initial (in contrast with the boundary) condition is: at time 0, the density of traffic along the link is 0. Only the left boundary condition will apply in a significant way.

Next we define a unique $\delta_{mn}(\widehat{\tau},\tau) \in \mathcal{D}_{mn}$ associated with $\widehat{f}$. To accomplish this, we shall apply Biles et al (2014) Theorem 1 to the (discontinuous) ordinary differential equation:[61]

$$\frac{\partial \delta_{mn}(\widehat{\tau},\tau)}{\partial \tau} = \upsilon\left(\widehat{f}(m,n,\tau,\delta_{mn}(\widehat{\tau},\tau)), x_{mn}\right) \tag{12}$$

[59]As detailed in the main text, one might prefer the use of the classical solution to the conservation law when there are no gaps inserted. In this case, less dense traffic does not separate itself from denser traffic that follows. Then we can set $\varsigma = 0$.

[60]In fact, this is where we use Assumption 3 (or the stronger version in the statement of Theorem 1), implying that traffic congestion does not backup onto a link at the endpoint of that link. In particular, we ignore behavior outside the link when we solve the conservation law for traffic on a link.

[61]Our first attempts, before finding Biles et al (2014), tried to apply Bressan (1988, Theorem 1). It is of some interest to see why the latter result cannot be applied. For

This will require us to delve a little into the clever proof of existence of a solution $\widehat{f}$ used by Strub and Bayen (2006)[62] in order to integrate it with the structure of Biles et al (2014).[63] These ideas will also be useful shortly in order to prove that $\widehat{g} \in \mathcal{G}$.

We know from Strub and Bayen (2006), p. 560, that for each $\Delta \in [0, \lambda(m,n)]$, $\widehat{f}(m,n,\cdot,\Delta)$ is of bounded variation. But for our purposes, it will be useful to prove the stronger assertion: For each $\Delta \in [0, \lambda(m,n)]$,

$$TV(\widehat{f}(m,n,\cdot,\Delta)) \leq \chi_{mn}.$$

That is next on the agenda.[64]

Strub and Bayen (2006) use an approximation, generally called the Godunov approximation, to construct the solution that we call $\widehat{f}(m,n,\cdot,\cdot)$. In their notation, they consider only one link and thus drop m and n. To reduce notation, we also drop these indexes temporarily. The discrete approximation they use is called ρ_i^s, where s denotes a time cell and i denotes a location cell, and where i and s are integers. Specifically, the cells are defined as follows:

$$\begin{aligned} I_i &= \left[\frac{\lambda}{\mathcal{M}} \cdot \left(i - \frac{1}{2}\right), \frac{\lambda}{\mathcal{M}} \cdot \left(i + \frac{1}{2}\right)\right] \\ J_s &= \left[\frac{\lambda}{\mathcal{M}} z \cdot \left(s - \frac{1}{2}\right), \frac{\lambda}{\mathcal{M}} z \cdot \left(s + \frac{1}{2}\right)\right] \end{aligned}$$

where λ is the length of the link, $\mathcal{M}$ denotes the number of location cells ($i = 1, 2, ..., \mathcal{M}$), s indexes time cells ($s = 1, 2, ..., \mathcal{K}$) where $\mathcal{K}$ is the smallest integer strictly larger than $\frac{\bar{t}\mathcal{M}}{\lambda z}$, and $z > 0$ is an arbitrary constant. The cell sizes tend to zero ($\mathcal{M} \to \infty$) as the approximation converges. It is important to note that, from the uniqueness result of Strub and Bayen (2006), the limit is actually independent of choice of z.

---

that Theorem, the natural upper bound on speed is the limit on speed as density tends to zero. But then, given a simple discontinuity in $\widehat{f}$, where $\widehat{f}$ is for example piecewise constant, we can leapfrog back and forth over the discontinuity using changes in both time and distance. This creates, in the terminology of Bressan (1988), unbounded variation in the cone Γ^M. Thus, the result cannot be applied, since bounded variation in the cone Γ^M is a key assumption of the Theorem.

[62] The keys to this proof are the Godunov construction and the Courant-Friedrichs-Lewy (CFL) condition.

[63] Since there are notational conflicts between the two papers as well as with our notation, integration requires some notational changes.

[64] A method for proving this, different from the one we use, would directly employ the fact that $\widehat{f}(m,n,\cdot,\Delta)$ is of bounded variation for each $\Delta \in [0, \lambda(m,n)]$, with possibly different bounds across Δ; then show that there is a uniform bound.

The boundary condition is given by:

$$\rho_i^0 = \frac{\mathcal{M}}{\lambda z} \cdot \int_{J_s} g(m,n,\tau) d\tau$$

The next issue, both difficult and important, is to show that $TV(\widehat{f}(m,n,\cdot,\Delta)) \leq \chi_{mn}$. This is important because we must show that the exit density from a link (as a function of time) has a uniform bound on variation so that we have compactness and we can apply a fixed point theorem. It is a stronger requirement than simply showing that the exit density has bounded variation for each given g. The reason this issue is difficult is due to the Godunov scheme. As Friedrich et al (2018, p. 8) note, "In particular, the Godunov type scheme also does not fit into the classical assumptions of total variation diminishing (TVD) schemes, as the total variation may slightly increase (as it is the same for the analytical solution)." Although we wish we could directly apply their results on bounded variation in section 3.3 of their paper, we cannot. Our framework is simpler (as we use local flux),[65] but the big hazard with the mathematics literature on conservation laws that also applies here is that they address the initial value problem rather than the boundary value problem. So we must alter their clever argument substantially.

To begin, the key equation system from Strub and Bayen (2006, p. 559) is as follows:[66]

$$\begin{cases} \rho^s_{i+\frac{1}{2}} \text{ is an element } f \text{ of } I(\rho^s_i, \rho^s_{i+1}) \text{ such that } sign(\rho^s_{i+1} - \rho^s_i) \cdot \Phi(f) \text{ is minimal} \\ \rho_i^{s+1} = \rho_i^s - z \cdot \left(\Phi(\rho^s_{i+\frac{1}{2}}) - \Phi(\rho^s_{i-\frac{1}{2}}) \right) \end{cases} \tag{13}$$

Then noting that in our particular traffic context, $\rho^s_{i+\frac{1}{2}} = \rho^s_i$ and $\rho^s_{i-\frac{1}{2}} = \rho^s_{i-1}$, and employing (13),

$$\begin{aligned} \rho^{s+1}_{i+1} - \rho^s_{i+1} &= \rho^s_{i+1} - z \cdot \left(\Phi(\rho^s_{i+1}) - \Phi(\rho^s_i)\right) - \rho^{s-1}_{i+1} + z \cdot \left(\Phi(\rho^{s-1}_{i+1}) - \Phi(\rho^{s-1}_i)\right) \\ &= z \cdot \left[\Phi(\rho^s_i) - \Phi(\rho^{s-1}_i)\right] + \left[\rho^s_{i+1} - \rho^{s-1}_{i+1}\right] + z \cdot \left[\Phi(\rho^{s-1}_{i+1}) - \Phi(\rho^s_{i+1})\right] \end{aligned} \tag{14}$$

Now consider two cases: $\rho^s_{i+1} \geq \rho^{s-1}_{i+1}$ and $\rho^s_{i+1} < \rho^{s-1}_{i+1}$. In the first case, using (14),

$$\begin{aligned} \rho^{s+1}_{i+1} - \rho^s_{i+1} &= z \cdot \left[\Phi(\rho^s_i) - \Phi(\rho^{s-1}_i)\right] + \left[\rho^s_{i+1} - \rho^{s-1}_{i+1}\right] + z \cdot \left[\Phi(\rho^{s-1}_{i+1}) - \Phi(\rho^s_{i+1})\right] \\ &= z \cdot \left[\Phi(\rho^s_i) - \Phi(\rho^{s-1}_i)\right] + \left|\rho^s_{i+1} - \rho^{s-1}_{i+1}\right| - z \cdot \left|\Phi(\rho^s_{i+1}) - \Phi(\rho^{s-1}_{i+1})\right| \end{aligned}$$

[65] It is easy to see how our work can be generalized to non-local flux, in the spirit of Friedrich et al (2018).

[66] The definition of I can be found in (8).

Using the fact that $z\phi \leq 1$, $\left|\rho_{i+1}^{s} - \rho_{i+1}^{s-1}\right| - z \cdot \left|\Phi(\rho_{i+1}^{s}) - \Phi(\rho_{i+1}^{s-1})\right| \geq 0$. Then taking absolute values,

$$\left|\rho_{i+1}^{s+1} - \rho_{i+1}^{s}\right| \leq z \cdot \left|\Phi(\rho_{i}^{s}) - \Phi(\rho_{i}^{s-1})\right| + \left|\rho_{i+1}^{s} - \rho_{i+1}^{s-1}\right| - z \cdot \left|\Phi(\rho_{i+1}^{s}) - \Phi(\rho_{i+1}^{s-1})\right|$$

Turning to the second case, multiplying both sides of (14) by -1,

$$\begin{aligned}\rho_{i+1}^{s} - \rho_{i+1}^{s+1} &= z \cdot \left[\Phi(\rho_{i}^{s-1}) - \Phi(\rho_{i}^{s})\right] + \left[\rho_{i+1}^{s-1} - \rho_{i+1}^{s}\right] + z \cdot \left[\Phi(\rho_{i+1}^{s}) - \Phi(\rho_{i+1}^{s-1})\right] \\ &= z \cdot \left[\Phi(\rho_{i}^{s-1}) - \Phi(\rho_{i}^{s})\right] + \left[\rho_{i+1}^{s-1} - \rho_{i+1}^{s}\right] - z \cdot \left[\Phi(\rho_{i+1}^{s-1}) - \Phi(\rho_{i+1}^{s})\right]\end{aligned}$$

Taking absolute values, in a similar fashion we obtain:

$$\left|\rho_{i+1}^{s} - \rho_{i+1}^{s+1}\right| \leq z \cdot \left|\Phi(\rho_{i}^{s-1}) - \Phi(\rho_{i}^{s})\right| + \left|\rho_{i+1}^{s-1} - \rho_{i+1}^{s}\right| - z \cdot \left|\Phi(\rho_{i+1}^{s-1}) - \Phi(\rho_{i+1}^{s})\right|$$

Thus, the following holds in either case:

$$\left|\rho_{i+1}^{s+1} - \rho_{i+1}^{s}\right| \leq z \cdot \left|\Phi(\rho_{i}^{s-1}) - \Phi(\rho_{i}^{s})\right| + \left|\rho_{i+1}^{s-1} - \rho_{i+1}^{s}\right| - z \cdot \left|\Phi(\rho_{i+1}^{s-1}) - \Phi(\rho_{i+1}^{s})\right|$$

Next, summing terms:

$$\begin{aligned}\sum_{s=1}^{\mathcal{M}-1}\sum_{i=0}^{\mathcal{K}-1}\left|\rho_{i+1}^{s+1} - \rho_{i+1}^{s}\right| &\leq z \cdot \sum_{s=1}^{\mathcal{M}-1}\sum_{i=0}^{\mathcal{K}-1}\left|\Phi(\rho_{i}^{s}) - \Phi(\rho_{i}^{s-1})\right| \\ &\quad + \sum_{s=1}^{\mathcal{M}-1}\sum_{i=0}^{\mathcal{K}-1}\left|\rho_{i+1}^{s} - \rho_{i+1}^{s-1}\right| - z \cdot \sum_{s=1}^{\mathcal{M}-1}\sum_{i=0}^{\mathcal{K}-1}\left|\Phi(\rho_{i+1}^{s-1}) - \Phi(\rho_{i+1}^{s})\right|\end{aligned}$$

Moving the second set of terms on the right hand side to the left and eliminating common elements,

$$\sum_{i=0}^{\mathcal{K}-1}\left|\rho_{i+1}^{\mathcal{M}} - \rho_{i+1}^{\mathcal{M}-1}\right| - \sum_{i=0}^{\mathcal{K}-1}\left|\rho_{i+1}^{1} - \rho_{i+1}^{0}\right| \leq z \cdot \left(\sum_{s=1}^{\mathcal{M}-1}\left|\Phi(\rho_{0}^{s}) - \Phi(\rho_{0}^{s-1})\right| - \sum_{s=1}^{\mathcal{M}-1}\left|\Phi(\rho_{\mathcal{K}}^{s}) - \Phi(\rho_{\mathcal{K}}^{s-1})\right|\right)$$

Therefore,

$$z \cdot \sum_{s=1}^{\mathcal{M}-1}\left|\Phi(\rho_{\mathcal{K}}^{s}) - \Phi(\rho_{\mathcal{K}}^{s-1})\right| + \sum_{i=0}^{\mathcal{K}-1}\left|\rho_{i+1}^{\mathcal{M}} - \rho_{i+1}^{\mathcal{M}-1}\right| \leq z \cdot \sum_{s=1}^{\mathcal{M}-1}\left|\Phi(\rho_{0}^{s}) - \Phi(\rho_{0}^{s-1})\right| + \sum_{i=0}^{\mathcal{K}-1}\left|\rho_{i+1}^{1} - \rho_{i+1}^{0}\right|$$

Now if we have chosen $\bar{t}$ large enough so that all commuters have arrived at their destinations (at the positive minimal speed) before that time, then for $\mathcal{M}$ sufficiently large, $\sum_{i=0}^{\mathcal{K}-1}\left|\rho_{i+1}^{\mathcal{M}} - \rho_{i+1}^{\mathcal{M}-1}\right| = 0$. In addition, the only location for which $\left|\rho_{i+1}^{1} - \rho_{i+1}^{0}\right| > 0$ is $i = 0$, and then it is bounded by $\overline{f}$. Hence,

$$z \cdot \sum_{s=1}^{\mathcal{M}-1}\left|\Phi(\rho_{\mathcal{K}}^{s}) - \Phi(\rho_{\mathcal{K}}^{s-1})\right| \leq z \cdot \sum_{s=1}^{\mathcal{M}-1}\left|\Phi(\rho_{0}^{s}) - \Phi(\rho_{0}^{s-1})\right| + \overline{f}$$

and therefore,

$$\sum_{s=1}^{\mathcal{M}-1} \left|\Phi(\rho_{\mathcal{K}}^{s}) - \Phi(\rho_{\mathcal{K}}^{s-1})\right| \leq \sum_{s=1}^{\mathcal{M}-1} \left|\Phi(\rho_{0}^{s}) - \Phi(\rho_{0}^{s-1})\right| + \frac{\overline{f}}{z}$$

Applying the Lipschitz conditions,

$$\begin{aligned}\frac{1}{\xi} \cdot \sum_{s=1}^{\mathcal{M}-1} \left|\rho_{\mathcal{K}}^{s} - \rho_{\mathcal{K}}^{s-1}\right| &\leq \sum_{s=1}^{\mathcal{M}-1} \left|\Phi(\rho_{\mathcal{K}}^{s}) - \Phi(\rho_{\mathcal{K}}^{s-1})\right| \\ &\leq \sum_{s=1}^{\mathcal{M}-1} \left|\Phi(\rho_{0}^{s}) - \Phi(\rho_{0}^{s-1})\right| + \frac{\overline{f}}{z} \\ &\leq \phi \cdot \sum_{s=1}^{\mathcal{M}-1} \left|\rho_{0}^{s} - \rho_{0}^{s-1}\right| + \frac{\overline{f}}{z}\end{aligned}$$

Summarizing,

$$\sum_{s=1}^{\mathcal{M}-1} \left|\rho_{\mathcal{K}}^{s} - \rho_{\mathcal{K}}^{s-1}\right| \leq \phi \cdot \xi \cdot \sum_{s=1}^{\mathcal{M}-1} \left|\rho_{0}^{s} - \rho_{0}^{s-1}\right| + \frac{\overline{f} \cdot \xi}{z} \tag{15}$$

In fact, this inequality holds not just for location cell $\mathcal{K}$, but for any location $\Delta \in (0, \lambda(m,n))$, by setting the limit of the various sums to the location cell containing Δ, called $i(\Delta, \mathcal{M})$, rather than to $\mathcal{K} - 1$. Recall that the entry density for this link is:

$$TV(\widehat{f}(m,n,\cdot,0)) \leq \chi_{mn}$$

and thus $\sum_{s=1}^{\mathcal{M}-1} \left|\rho_{0}^{s} - \rho_{0}^{s-1}\right| \leq \chi_{mn}$. Strub and Bayen (2008) show that a subsequence of ρ, which is implicitly indexed by $\mathcal{M}$, converges strongly (and thus almost surely) in $L^1\left((0, \lambda(m,n)) \times \left(0, \overline{t}\right)\right)$ to the unique solution that is of bounded variation. Since this is not pointwise convergence (and $\{\lambda(m,n)\} \times \left[0, \overline{t}\right]$ is of measure 0 in $[0, \lambda(m,n)] \times \left[0, \overline{t}\right]$), it is possible that $\sum_{s=1}^{\mathcal{M}-1} \left|\rho_{\mathcal{K}}^{s} - \rho_{\mathcal{K}}^{s-1}\right|$ does not converge to $TV\left(\widehat{f}(m,n,\cdot,\lambda(m,n))\right)$. To obtain the exit density from this link, take a sequence $\{\Delta_k\}_{n=1}^{\infty}$ with $\Delta_k < \lambda(m,n)$ and $\lim_{n\to\infty} \Delta_k = \lambda(m,n)$, and for each k construct the sequence $\left\{\rho_{i(\Delta_k,\mathcal{M})}^{s}\right\}_{s=1}^{\infty}$. Then using (15), $\sum_{s=1}^{\mathcal{M}-1} \left|\rho_{i(\Delta_k,\mathcal{M})}^{s} - \rho_{i(\Delta_k,\mathcal{M})}^{s-1}\right|$ is (uniformly) bounded. Applying Helly's theorem, for each fixed k construct the function $\widetilde{f}(m,n,\cdot,\Delta_k)$ as a pointwise limit of a subsequence of $\left\{\rho_{i(\Delta_k,\mathcal{M})}^{s}\right\}_{s=1}^{\infty}$ as $\mathcal{M} \to \infty$. Then apply Helly's theorem again to obtain the pointwise limit of a subsequence of $\widetilde{f}(m,n,\cdot,\Delta_k)$ as $k\to\infty$, and call this density $\widehat{f}(m,n,\cdot,\lambda(m,n))$. This exit density (as a function of time) will form the basis for entry density on succeeding links. Notice that

$$TV(\widehat{f}(m,n,\cdot,\lambda(m,n)))\leq\phi\cdot\xi\cdot\chi_{mn}+\frac{\overline{f}\cdot\xi}{z}. \tag{16}$$

By remark 2.1 of Bressan (2000), we can take it to be right continuous in t.

Next, we examine whether this exit density is unique, at least among functions of bounded variation that satisfy $TV(f(m,n,\cdot,\Delta))\leq\phi\cdot\xi\cdot\chi_{mn}+\frac{\overline{f}\cdot\xi}{z}$. Suppose that there are two different exit limits of bounded variation; call them $\widetilde{f}(m,n,\cdot,\lambda(m,n))$ and $\widehat{f}(m,n,\cdot,\lambda(m,n))$. Now we already know from Strub and Bayen (2006) that $\widetilde{f}(m,n,\cdot,\cdot)=\widehat{f}(m,n,\cdot,\cdot)$ a.s. (t,Δ). The next argument parallels Strub and Bayen (2006, pp. 558-559) where they argue that their solution is unique; we abuse notation slightly, as in Strub and Bayen (2006, p. 558), and write ψ as a function of location only, independent of time.[67] We know for C^1 $\varphi:(0,\overline{t})\to\mathbb{R}_+$ with compact support and C^1 $\psi:(0,\lambda(m,n))\to\mathbb{R}_+$ with compact support,

$$\begin{aligned}&\int_0^{\lambda(m,n)}\int_0^{\overline{t}}\left|\widetilde{f}(m,n,t,\Delta)-\widehat{f}(m,n,t,\Delta)\right|\psi(\Delta)\varphi'(t)+\\&sign\left(\widetilde{f}(m,n,t,\Delta)-\widehat{f}(m,n,t,\Delta)\right)\cdot\left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right)-\Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right)\cdot\psi'(\Delta)\varphi(t)d\Delta dt\\&\geq 0\end{aligned} \tag{17}$$

For φ approximating the indicator function of $[0,\overline{t}]$, we have:

$$\begin{aligned}&\limsup_{t\to 0}\int_0^{\lambda(m,n)}\left|\widetilde{f}(m,n,t,\Delta)-\widehat{f}(m,n,t,\Delta)\right|\psi(\Delta)d\Delta\\&-\liminf_{t\to\overline{t}}\int_0^{\lambda(m,n)}\left|\widetilde{f}(m,n,t,\Delta)-\widehat{f}(m,n,t,\Delta)\right|\psi(\Delta)d\Delta\\&\geq-\int_0^{\lambda(m,n)}\int_0^{\overline{t}}sign\left(\widetilde{f}(m,n,t,\Delta)-\widehat{f}(m,n,t,\Delta)\right)\cdot\\&\left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right)-\Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right)\cdot\psi'(\Delta)d\Delta dt\end{aligned}$$

[67] Implicitly, this uses the idea that the indicator function on $[0,\overline{t}]$ can be approximated by C^1 functions with compact support on $(0,\overline{t})$.

Taking ψ to approximate the indicator function of $[0, \lambda(m,n)]$,

$$\begin{aligned}
\geq \ & \lim \sup_{\Delta \to \lambda(m,n)} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot \\
& \left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt \\
& - \lim \inf_{\Delta \to 0} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot \\
& \left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt
\end{aligned}$$

In sum, we have:

$$\begin{aligned}
& \lim \sup_{t \to 0} \int_0^{\lambda(m,n)} \left|\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right| d\Delta \\
& - \lim \inf_{t \to \bar{t}} \int_0^{\lambda(m,n)} \left|\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right| d\Delta \\
\geq \ & \lim \sup_{\Delta \to \lambda(m,n)} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot \\
& \left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt \\
& - \lim \inf_{\Delta \to 0} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot \\
& \left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt
\end{aligned}$$

From the conditions on links at times 0 and $\bar{t}$, the left hand side (the first two terms) are zero, we obtain:

$$\begin{aligned}
& \lim \inf_{\Delta \to 0} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot \\
& \left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt \\
\geq \ & \lim \sup_{\Delta \to \lambda(m,n)} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot \\
& \left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt
\end{aligned}$$

As in Strub and Bayen (2006, p. 558), the left hand side is 0. The right hand side is non-negative (recall that flux Φ is increasing in density). Hence,

$$\lim \sup_{\Delta \to \lambda(m,n)} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot \left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt$$

$$= 0$$

Now since Φ is strictly increasing in f, we know that

$$\lim \sup_{\Delta \to \lambda(m,n)} \left\| \widetilde{f}(m,n,\cdot,\Delta) - \widehat{f}(m,n,\cdot,\Delta) \right\|_{L^1} = 0,$$

implying that $\widetilde{f}(m,n,\cdot,\lambda(m,n)) = \widehat{f}(m,n,\cdot,\lambda(m,n))$ a.s. (t). Both $\widetilde{f}(m,n,\cdot,\lambda(m,n))$ and $\widehat{f}(m,n,\cdot,\lambda(m,n))$ are of bounded variation, so by Lemma 2.1 and Remark 2.1 of Bressan (2000), by taking right continuous versions, they are in fact equal.

Although we used the argument just above to obtain a well-defined exit density, if we replace $\lambda(m,n)$ with an arbitrary distance $\overline{\Delta}$, $0 < \overline{\Delta} < \lambda(m,n)$, the same argument applies and we have that for any sequence $\left\{\overline{\Delta}_k\right\}_{k=1}^{\infty}$ with $\lim_{k\to\infty} \overline{\Delta}_k = \overline{\Delta}$, $\lim_{k\to\infty} \widehat{f}(m,n,\cdot,\overline{\Delta}_k) = \widehat{f}(m,n,\cdot,\overline{\Delta})$ a.s. (t), where

$$TV(\widehat{f}(m,n,\cdot,\overline{\Delta})) \le \phi \cdot \xi \cdot \chi_{mn} + \frac{\overline{f} \cdot \xi}{z}.$$

Taking the right continuous version, it follows that $\lim_{k\to\infty} \widehat{f}(m,n,\cdot,\overline{\Delta}_k) = \widehat{f}(m,n,\cdot,\overline{\Delta})$ pointwise.

The next step in our analysis is to examine existence and uniqueness of Carathéodory solutions to (12), given that we have a unique solution to the conservation law. Theorem 1 of Biles et al (2014) employs as one sufficient condition that the number of discontinuities in $\widehat{f}(m,n,\tau,\cdot)$ is countable for each τ. To prove this, we first show that for each τ, $TV(\widehat{f}(m,n,\tau,\cdot)) < \infty$. We shall repeat some of the arguments above with the space domain in place of the time domain.

From (13),

$$\rho_i^{s+1} - \rho_{i-1}^{s+1} = \rho_i^s - \rho_{i-1}^s - z \cdot \left(\Phi(\rho_i^s) - \Phi(\rho_{i-1}^s)\right) + z \cdot \left(\Phi(\rho_{i-1}^s) - \Phi(\rho_{i-2}^s)\right)$$

Again, we consider two cases: $\rho_i^s \ge \rho_{i-1}^s$ and $\rho_i^s < \rho_{i-1}^s$. In the first case,

$$\rho_i^{s+1} - \rho_{i-1}^{s+1} = \left|\rho_i^s - \rho_{i-1}^s\right| - z \cdot \left|\Phi(\rho_i^s) - \Phi(\rho_{i-1}^s)\right| + z \cdot \left(\Phi(\rho_{i-1}^s) - \Phi(\rho_{i-2}^s)\right)$$

Taking absolute values and using the fact that $0 < z \le \frac{1}{\phi}$, implying $\left|\rho_i^s - \rho_{i-1}^s\right| - z \cdot \left|\Phi(\rho_i^s) - \Phi(\rho_{i-1}^s)\right| \ge 0$,

$$\left|\rho_i^{s+1} - \rho_{i-1}^{s+1}\right| \le \left|\rho_i^s - \rho_{i-1}^s\right| - z \cdot \left|\Phi(\rho_i^s) - \Phi(\rho_{i-1}^s)\right| + z \cdot \left|\Phi(\rho_{i-1}^s) - \Phi(\rho_{i-2}^s)\right|$$

In the second case, $\rho_i^s < \rho_{i-1}^s$, and

$$\rho_i^{s+1} - \rho_{i-1}^{s+1} = \rho_i^s - \rho_{i-1}^s - z \cdot \left(\Phi(\rho_i^s) - \Phi(\rho_{i-1}^s)\right) + z \cdot \left(\Phi(\rho_{i-1}^s) - \Phi(\rho_{i-2}^s)\right)$$

Multiplying both sides by -1,

$$\rho_{i-1}^{s+1} - \rho_i^{s+1} = \rho_{i-1}^s - \rho_i^s - z \cdot \left(\Phi(\rho_{i-1}^s) - \Phi(\rho_i^s)\right) + z \cdot \left(\Phi(\rho_{i-2}^s) - \Phi(\rho_{i-1}^s)\right)$$

Taking absolute values,

$$\left|\rho_i^{s+1} - \rho_{i-1}^{s+1}\right| \leq \left|\rho_i^s - \rho_{i-1}^s\right| - z \cdot \left|\Phi(\rho_i^s) - \Phi(\rho_{i-1}^s)\right| + z \cdot \left|\Phi(\rho_{i-1}^s) - \Phi(\rho_{i-2}^s)\right|$$

In either case,

$$\left|\rho_i^{s+1} - \rho_{i-1}^{s+1}\right| \leq \left|\rho_i^s - \rho_{i-1}^s\right| - z \cdot \left|\Phi(\rho_i^s) - \Phi(\rho_{i-1}^s)\right| + z \cdot \left|\Phi(\rho_{i-1}^s) - \Phi(\rho_{i-2}^s)\right|$$

Next, fix $\mathcal{M}'$ integer, $1 \leq \mathcal{M}' \leq \mathcal{M}$. Summing terms:

$$\begin{aligned}\sum_{s=0}^{\mathcal{M}'-1}\sum_{i=1}^{\mathcal{K}-1}\left|\rho_i^{s+1} - \rho_{i-1}^{s+1}\right| \quad \leq \quad & \sum_{s=0}^{\mathcal{M}'-1}\sum_{i=1}^{\mathcal{K}-1}\left|\rho_i^s - \rho_{i-1}^s\right| - z \cdot \sum_{s=0}^{\mathcal{M}'-1}\sum_{i=1}^{\mathcal{K}-1}\left|\Phi(\rho_i^s) - \Phi(\rho_{i-1}^s)\right| \\ & +z \cdot \sum_{s=0}^{\mathcal{M}'-1}\sum_{i=1}^{\mathcal{K}-1}\left|\Phi(\rho_{i-1}^s) - \Phi(\rho_{i-2}^s)\right|\end{aligned}$$

Moving the first set of terms on the right hand side to the left and eliminating common elements,

$$\sum_{i=1}^{\mathcal{K}-1}\left|\rho_i^{\mathcal{M}'} - \rho_{i-1}^{\mathcal{M}'}\right| - \sum_{i=1}^{\mathcal{K}-1}\left|\rho_i^0 - \rho_{i-1}^0\right| \leq -z \cdot \sum_{s=0}^{\mathcal{M}'-1}\left|\Phi(\rho_{\mathcal{K}-1}^s) - \Phi(\rho_{\mathcal{K}-2}^s)\right| \leq 0$$

Now $\left|\rho_i^0 - \rho_{i-1}^0\right| = 0$ except for $i = 1$, and then it is bounded by $\overline{f}$. Hence $TV(\widehat{f}(m,n,\tau,\cdot)) < \infty$, and by Bressan (2000, Lemma 2.1), $\widehat{f}(m,n,\tau,\cdot)$ has only countably many discontinuities.

With this fact in hand, Biles et al (2014) Theorem 1 and the discussion following for the scalar case implies there exists a Carathéodory solution to (12), and the set of all such solutions form a funnel. We take the minimum (speed) of these, since we do not allow passing. Using the result in that paper, it is unique.

Next, we apply the arguments elaborated above to define, and discover properties of, the map $\mathcal{T} : \mathcal{G} \to \mathcal{G}$. Let $\widehat{f}(m,n,\cdot,\cdot)$ be the unique solution to the conservation law on link mn with initial condition 0 and boundary condition $g(m,n,\cdot)$.

Let $\delta_{mn}(\widehat{\tau},\tau)$ be the corresponding (unique) solution to the differential equation (12). Define:

$$\widehat{\tau}_{mn}(\tau) = \delta_{mn}^{-1}(\cdot,\tau)(\lambda(m,n)) \tag{18}$$

Notice that since speed $\upsilon > 0$, $\delta_{mn}(\widehat{\tau}, \tau)$ is strictly decreasing in $\widehat{\tau}$, so $\widehat{\tau}_{mn}(\tau)$ is well-defined.

With this preparation, we can define the image $\mathcal{T}(g) = \widehat{g}$, that will depend on both g, through the solution on a link $\widehat{f}$ as defined above, and δ, through its inverse image $\widehat{\tau}$.

$$\widehat{g}_n^r(\tau) \equiv \begin{cases} \Phi_{mn}^{-1}\left(\frac{\Phi_{mn}(g_m^r(\widehat{\tau}_{mn}(\tau)))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau))\right)} \cdot \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))\right)\right) \\ \qquad \text{if } \exists\ i \geq 3 \text{ with } \pi_{i-1}(r) = m,\ \pi_i(r) = n; \\ \Phi_{mn}^{-1}\left(\frac{\mu(\{c \in C \mid R(c)) = r,\ \left|\tau^d(c) - \tau\right| < \overline{\tau}\})}{2\overline{\tau}}\right) \text{ if } \pi_1(r) = m,\ \pi_2(r) = n \end{cases} \tag{19}$$

The argument that $\widehat{g} \in \mathcal{G}$ is as follows.

First, for the case $\pi_1(r) = m$, $\pi_2(r) = n$, by definition

$$TV(\widehat{g}_n^r(\tau)) = TV\left(\Phi_{mn}^{-1}\left(\frac{\mu(\{c \in C \mid \pi_1(R(c)) = n,\ \left|\tau^d(c) - \tau\right| < \overline{\tau}\})}{2\overline{\tau}}\right)\right) \leq (T/\overline{\tau}+1)\cdot\overline{f}$$

In all other cases,

$$TV(\widehat{g}^r_n(\cdot)) = TV\left(\Phi^{-1}_{mn}\left(\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(\tau))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(\tau))\right)}\right.\right.$$

$$\left.\left.\cdot\Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))\right)\right)\right)$$

$$= \sup_{K\geq 1,\ t_k\in[0,\bar{t}],\ t_0<t_1<\cdots<t_K}\left\{\sum_{k=1}^{K}\left|\Phi^{-1}_{mn}\left(\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(t_k))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(t_k))\right)}\right.\right.\right.$$

$$\left.\cdot\Phi_{mn}\left(\widehat{f}(m,n,t_k,\lambda(m,n))\right)\right)$$

$$-\Phi^{-1}_{mn}\left(\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(t_{k-1}))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(t_{k-1}))\right)}\right.$$

$$\left.\left.\left.\cdot\Phi_{mn}\left(\widehat{f}(m,n,t_{k-1},\lambda(m,n))\right)\right)\right|\right\}$$

$$\leq \sup_{K\geq 1,\ t_k\in[0,\bar{t}],\ t_0<t_1<\cdots<t_K}\left\{\sum_{k=1}^{K}\xi\left|\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(t_k))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(t_k))\right)}\right.\right.$$

$$\cdot\Phi_{mn}\left(\widehat{f}(m,n,t_k,\lambda(m,n))\right)$$

$$\left.\left.-\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(t_{k-1}))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(t_{k-1}))\right)}\cdot\Phi_{mn}\left(\widehat{f}(m,n,t_{k-1},\lambda(m,n))\right)\right|\right\}$$

$$\leq \sup_{K\geq 1,\ t_k\in[0,\bar{t}],\ t_0<t_1<\cdots<t_K}\xi\left\{\sum_{k=1}^{K}\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(t_k))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(t_k))\right)}\right.$$

$$\left.\cdot\left|\Phi_{mn}\left(\widehat{f}(m,n,t_k,\lambda(m,n))\right)-\Phi_{mn}\left(\widehat{f}(m,n,t_{k-1},\lambda(m,n))\right)\right|\right\}$$

$$+\sup_{K\geq 1,\ t_k\in[0,\bar{t}],\ t_0<t_1<\cdots<t_K}\xi\left\{\sum_{k=1}^{K}\left|\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(t_k))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(t_k))\right)}\right.\right.$$

$$\left.\left.-\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(t_{k-1}))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(t_{k-1}))\right)}\right|\cdot\Phi_{mn}\left(\widehat{f}(m,n,t_{k-1},\lambda(m,n))\right)\right\}$$

$$\leq \xi\phi TV(\widehat{f}(m,n,\cdot,\lambda(m,n)))$$

$$+\sup_{K\geq 1,\ t_k\in[0,\bar{t}],\ t_0<t_1<\cdots<t_K}\xi\left\{\sum_{k=1}^{K}\left|\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(t_k))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(t_k))\right)}\right.\right.$$

$$\left.\left.-\frac{\Phi_{mn}\left(g^r_m(\widehat{\tau}_{mn}(t_{k-1}))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}}\Phi_{mn}\left(g^{r'}_m(\widehat{\tau}_{mn}(t_{k-1}))\right)}\right|\right\}\cdot\Phi_{mn}\left(\overline{f}\right)\quad(20)$$

To simplify this expression further, we focus on the second term. For notational brevity, define:

$$\Xi \equiv \frac{\frac{1}{\left[\sum_{\{r'\in\mathcal{R}|\exists\ i \text{ s.t. } \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)\right]}}{\frac{1}{\left[\sum_{\{r'\in\mathcal{R}|\exists\ i \text{ s.t. } \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)\right]}}.$$

$$\begin{aligned}
&\left|\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_k))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)}\right.\\
&\left.-\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)}\right|\\
= {}& \Xi\cdot\left|\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)\right]\cdot\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_k))\right)\right.\\
&\left.-\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right)\cdot\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)\right]\right|\\
\leq {}& \Xi\cdot\left(\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)\right]\right.\\
&\cdot\left|\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_k))\right)-\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right)\right|\\
&+\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right)\cdot\left|\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)\right]\right.\\
&\left.\left.-\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)\right]\right|\right)\\
\leq {}& \frac{\left|\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_k))\right)-\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right)\right|}{\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)\right]}\\
&+\frac{1}{\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)\right]}\cdot\\
&\left|\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)\right]\right.\\
&\left.-\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)\right]\right|
\end{aligned}$$

$$
\begin{aligned}
\leq \quad & \frac{1}{\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)\right]} \cdot \\
& \Bigg(\quad |\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_k))\right) - \Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right)| + \\
& \Bigg|\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)\right] \\
& - \left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)\right]\Bigg|\Bigg)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& \Bigg|\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_k))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)} \\
& - \frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)}\Bigg|
\end{aligned}
$$

$$
\begin{aligned}
\leq \quad & \frac{1}{\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)\right]} \cdot \\
& \Bigg(\quad |\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_k))\right) - \Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right)| \\
& + \Bigg|\left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)\right] \\
& - \left[\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)\right]\Bigg|\Bigg)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& \Bigg|\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_k))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right)} \\
& - \frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i\ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right)}\Bigg|
\end{aligned}
$$

$$\leq \left(\max \left\{ \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right) \right], \right. \right.$$

$$\left. \left. \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right) \right] \right\} \right)^{-1}$$

$$\cdot \left(\left| \quad \Phi_{mn} \left(g_m^{r}(\widehat{\tau}_{mn}(t_{k-1})) \right) - \Phi_{mn} \left(g_m^{r}(\widehat{\tau}_{mn}(t_k)) \right) \right| \right.$$

$$+ \left| \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right) \right] \right.$$

$$\left. \left. - \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right) \right] \right| \right) \tag{21}$$

The key point from the last inequality is that as long as

$$\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right) > 0$$

or

$$\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right) > 0,$$

then

$$\max \left\{ \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right) \right], \right.$$

$$\left. \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right) \right] \right\} \geq \Phi_{mn} \left(\underline{g} \right)$$

If both are 0, then we can ignore this term in the calculations of $TV(\widehat{g}_n^r(\widehat{\tau}_{mn}(\cdot)))$ and $TV\left(\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\cdot))\right)\right)$, since the difference is 0, so this term in the calculation of TV is 0.

Therefore, from (20),

$$TV(\widehat{g}_n^r(\cdot)) \leq \xi \phi TV(f(m, n, \cdot, \lambda(m, n)))$$

$$+\xi \Phi \left(\overline{f} \right) \sup_{K \geq 1, \ t_k \in [0, \overline{t}], \ t_0 < t_1 < \cdots < t_K} \left\{ \sum_{k=1}^{K} \left| \frac{\Phi_{mn} \left(g_m^{r}(\widehat{\tau}_{mn}(t_k)) \right)}{\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right)} \right. \right.$$

$$\left. \left. - \frac{\Phi_{mn} \left(g_m^{r}(\widehat{\tau}_{mn}(t_{k-1})) \right)}{\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn} \left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right)} \right| \right\}$$

From (16) and (21),

$$\leq \phi^2 \cdot \xi^2 \cdot \chi_{mn} + \frac{\overline{f} \cdot \xi^2 \cdot \phi}{z}$$

$$+\xi \Phi\left(\overline{f}\right) \sup \left\{ \left(\max \left\{ \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right) \right], \right. \right. \right.$$

$$\left. \left. \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right) \right] \right\} \right)^{-1} \cdot$$

$$\left(\sum_{k=1}^{K} \left| \Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_{k-1}))\right) - \Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(t_k))\right) \right| \right.$$

$$+ \sum_{k=1}^{K} \left| \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_k))\right) \right] \right.$$

$$\left. \left. - \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))\right) \right] \right| \right)$$

$$\left. \Bigm| \ K \geq 1, \, t_k \in [0, \overline{t}], \, t_0 < t_1 < \cdots < t_K \right\}$$

$$\leq \phi^2 \cdot \xi^2 \cdot \chi_{mn} + \frac{\overline{f} \cdot \xi^2 \cdot \phi}{z}$$

$$+ \frac{\xi \cdot \phi \cdot \Phi\left(\overline{f}\right)}{\Phi\left(\underline{g}\right)} \left[TV(g_m^r(\widehat{\tau}_{mn}(\cdot))) + TV \left(\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\cdot)) \right) \right]$$

$$\leq \phi^2 \cdot \xi^2 \cdot \chi_{mn} + \frac{\overline{f} \cdot \xi^2 \cdot \phi}{z}$$

$$+ \frac{\xi \cdot \phi \cdot \Phi\left(\overline{f}\right)}{\Phi\left(\underline{g}\right)} \left[TV(g_m^r(\cdot)) + TV \left(\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\cdot) \right) \right]$$

To bound this expression, note that if $\pi_1(r) = m$, $TV(g_m^r(\cdot)) \leq (T/\overline{\tau} + 1) \cdot \overline{f} = \chi_1$. More generally, if there is some i with $\pi_i(r) = m$, $\pi_{i+1}(r) = n$, $TV(g_m^r(\cdot)) \leq \chi_{mn}$ and $TV\left(\sum_{\{r' \in \mathcal{R} | \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\cdot)\right) \leq (N-1) \cdot \chi_{mn}$. Hence by (11), for any $r \in \mathcal{R}$ with $\pi_{i+2}(r) = n'$, $TV(\widehat{g}_n^r(\cdot)) \leq \chi_{h(nn')}$ and $\widehat{g} \in \mathcal{G}$.

The set $\mathcal{G}$ is obviously convex as a product of convex sets. Imposing the L^1 norm topology on each component $\mathcal{G}_n^r$, Helly's theorem implies that $\mathcal{G}_n^r$ is compact and hence $\mathcal{G}$ is compact as a product of compact sets. What remains is to show that $\mathcal{T}$ is continuous. Here we use intensively its definition (19).

Let $\{g(\cdot)_q\}_{q=1}^{\infty} \subseteq \mathcal{G}$ where $\lim_{q\to\infty} g(\cdot)_q = g(\cdot)$, and thus $\lim_{q\to\infty} g_n^r(\cdot)_q = g_n^r(\cdot)$ for all $r \in \mathcal{R}$ and n such that $\pi_i(r) = n$ for some i. We must show that $\lim_{q\to\infty} \mathcal{T}(g(\cdot)_q) = \mathcal{T}(g(\cdot))$. To prove this, we must examine each node in each admissible route independently. So let us focus on node n (subscript) in route r (superscript) for the calculations.

Let $\widehat{f}(m,n,\cdot,\cdot)_q$ be the (unique) solution to the boundary value problem with boundary conditions given by

$$g(m,n,\cdot)_q \equiv \sum_{\{r\in\mathcal{R}|\pi_i(r)=m,\ \pi_{i+1}(r)=n \text{ for some } i\geq 1\}} g_n^r(\cdot)_q$$

Let $\widehat{\tau}_{mn}(\tau)_q$ be the corresponding solution to (18). Next we show that in L^1, $\widehat{f}(m,n,\cdot,\cdot) = \lim_{q\to\infty} \widehat{f}(m,n,\cdot,\cdot)_q$ exists and is a solution at boundary conditions $g(\cdot)$. The proof traces back through Strub and Bayen's (2006) proof that a solution exists, detailed above, and uses an interchange of limits as follows. The boundary condition at each link mn for the Godunov approximation is given by:

$$\begin{aligned} \rho_{0,q}^s &= \frac{\mathcal{M}}{\lambda(m,n)\cdot z}\int_{J_s} g(m,n,\cdot)_q dt \\ \rho_0^s &= \frac{\mathcal{M}}{\lambda(m,n)\cdot z}\int_{J_s} g(m,n,\cdot) dt \end{aligned}$$

Evidently, $\rho_{0,q}^s \to \rho_0^s$ by the dominated convergence theorem. All of the pieces of the proof in Strub and Bayen (2006) rely on $\rho_{0,q}^s$ as well as equalities or weak inequalities. So if they hold for every element of the sequence, they also hold for the limit. Thus, $\widehat{f}(m,n,\cdot,\cdot) = \lim_{q\to\infty} \widehat{f}(m,n,\cdot,\cdot)_q$ exists and is the (unique) solution at initial conditions $g(\cdot)$.[68]

Next, we check continuity of the sequence of solutions to the differential

[68] An alternative proof uses the definition of a solution as given in section 2.2. The sequence satisfies the definition, and what is to be shown is that the limit satisfies it. Since there is convergence in norm, the tricky part is dealing with the exceptional sets of measure zero.

equation implied by the sequence of solutions to the conservation law.

$$\begin{aligned}
&\int_0^{\bar{t}} \Bigg| \Phi_{mn}^{-1} \left(\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau))\right)} \right. \\
&\left. \cdot \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))\right) \right) \\
&-\Phi_{mn}^{-1} \left(\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q\right)} \right. \\
&\left. \cdot \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))_q\right) \right) \Bigg| d\tau \\
\leq\ & \xi \int_0^{\bar{t}} \Bigg| \frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau))\right)} \cdot \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))\right) \\
&-\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q\right)} \cdot \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))_q\right) \Bigg| d\tau \\
\leq\ & \xi \int_0^{\bar{t}} \Bigg| \frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau))\right)} \cdot \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))\right) \\
&-\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q\right)} \cdot \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))\right) \Bigg| \\
&+\Bigg| \frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q\right)} \cdot \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))\right) \\
&-\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q\right)} \cdot \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))_q\right) \Bigg| d\tau
\end{aligned}$$

$$\begin{aligned}
&\leq \xi \int_0^{\bar{t}} \Bigg| \frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau))\right)} \\
&-\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q\right)} \Bigg| \cdot \Phi_{mn}\left(\overline{f}\right) \\
&+\left| \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))\right) - \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))_q\right) \right| d\tau \\
&= \Phi_{mn}\left(\overline{f}\right) \cdot \xi \int_0^{\bar{t}} \Bigg| \frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau))\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau))\right)} \qquad (22) \\
&-\frac{\Phi_{mn}\left(g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q\right)}{\sum_{\{r'\in\mathcal{R}|\text{for some } i \ \pi_i(r')=m,\ \pi_{i+1}(r')=n\}} \Phi_{mn}\left(g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q\right)} \Bigg| d\tau \\
&+\xi \int_0^{\bar{t}} \left| \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))\right) - \Phi_{mn}\left(\widehat{f}(m,n,\tau,\lambda(m,n))_q\right) \right| d\tau
\end{aligned}$$

We consider each of the two terms in (22) separately. For the first term, note that $\dot{\delta}$ is bounded above by $v\left(0, x_{mn}\right) < \infty$, so $d\widehat{\tau}_{mn}(\tau)_q/d\tau$ and $d\widehat{\tau}_{mn}(\tau)/d\tau$ are both bounded away from 0 by $1/v\left(0, x_{mn}\right)$. Hence, sets of measure 0 in time τ are mapped to sets of measure 0 in the images of $\widehat{\tau}_{mn}(\cdot)_q$ and $\widehat{\tau}_{mn}$. Using Ascoli's theorem and passing to a subsequence if necessary, $\widehat{\tau}_{mn}(\cdot)_q \rightarrow \widehat{\tau}_{mn}(\cdot)$ uniformly. For if not, then $\lim_{q\rightarrow\infty} \widehat{\tau}_{mn}(\cdot)_q \neq \widehat{\tau}_{mn}(\cdot)$, and there are two solutions to the differential equation (12), a contradiction. Since $\lim_{q\rightarrow\infty} g_m^r(\cdot)_q = g_m^r(\cdot)$ in L^1 norm, the convergence is a.s. Hence $g_m^r(\widehat{\tau}_{mn}(\cdot)_q)_q \rightarrow g_m^r(\widehat{\tau}_{mn}(\cdot))$ a.s. By Lebesgue's dominated convergence theorem, the first term converges to 0.

For the second term in (22), recall that $\widehat{f}(m, n, \cdot, \lambda(m, n))$ is defined uniquely. Now suppose that $\lim_{q\rightarrow\infty} \widehat{f}(m, n, \cdot, \lambda(m, n))_q \neq \widehat{f}(m, n, \cdot, \lambda(m, n))$. Then by Helly's theorem, we can find a subsequence of $\{f(m, n, \cdot, \lambda(m, n))_q\}_{q=1}^{\infty}$ converging to, say, $\widetilde{f}(m, n, \cdot, \lambda(m, n)) \neq \widehat{f}(m, n, \cdot, \lambda(m, n))$, where convergence is pointwise and $\widetilde{f}(m, n, \cdot, \lambda(m, n))$ is of bounded variation. By a uniqueness argument given above, it must be that $\widehat{f}(m, n, \cdot, \lambda(m, n))$ is not the exit density for a solution. From (17),

$$\begin{gathered}
\int_0^{\lambda(m,n)} \int_0^{\bar{t}} \left| \widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right| \psi(\Delta)\varphi'(t) \\
+sign\left(\widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta)\right) \cdot \left(\Phi\left(\widetilde{f}(m, n, t, \Delta)\right) - \Phi\left(\widehat{f}(m, n, t, \Delta)\right)\right) \cdot \psi'(\Delta)\varphi(t) d\Delta dt \\
\geq 0
\end{gathered}$$

For φ approximating the indicator function of $[0, \bar{t}]$, we have:

$$\begin{aligned}
&\limsup_{t\rightarrow 0} \int_0^{\lambda(m,n)} \left| \widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right| \psi(\Delta) d\Delta \\
&- \liminf_{t\rightarrow \bar{t}} \int_0^{\lambda(m,n)} \left| \widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right| \psi(\Delta) d\Delta \\
&\geq - \int_0^{\lambda(m,n)} \int_0^{\bar{t}} sign\left(\widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta)\right) \cdot \\
&\left(\Phi\left(\widetilde{f}(m, n, t, \Delta)\right) - \Phi\left(\widehat{f}(m, n, t, \Delta)\right)\right) \cdot \psi'(\Delta) d\Delta dt
\end{aligned}$$

Taking ψ to approximate the indicator function of $[0, \lambda(m,n)]$,

$$\geq \lim \sup_{\Delta \to \lambda(m,n)} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot$$
$$\left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt$$
$$- \lim \inf_{\Delta \to 0} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot$$
$$\left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt$$

In sum, we have:

$$\limsup_{t \to 0} \int_0^{\lambda(m,n)} \left|\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right| \psi(\Delta) d\Delta -$$
$$\liminf_{t \to \bar{t}} \int_0^{\lambda(m,n)} \left|\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right| \psi(\Delta) d\Delta$$
$$\geq \lim \sup_{\Delta \to \lambda(m,n)} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot$$
$$\left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt$$
$$- \lim \inf_{\Delta \to 0} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot$$
$$\left(\Phi\left(f(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt$$

Since the left hand side (the first two terms) are zero, we obtain:

$$\lim \inf_{\Delta \to 0} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot$$
$$\left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt$$
$$\geq \lim \sup_{\Delta \to \lambda(m,n)} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot$$
$$\left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt$$

As in Strub and Bayen (2006, p. 558), the left hand side is 0. The right hand side is non-negative (recall that flux Φ is strictly increasing in density). Hence,

$$\lim \sup_{\Delta \to \lambda(m,n)} \int_0^{\bar{t}} sign\left(\widetilde{f}(m,n,t,\Delta) - \widehat{f}(m,n,t,\Delta)\right) \cdot \left(\Phi\left(\widetilde{f}(m,n,t,\Delta)\right) - \Phi\left(\widehat{f}(m,n,t,\Delta)\right)\right) dt$$
$$= 0$$

Now since Φ is strictly increasing in f, then we know that

$$\lim \sup_{\Delta \to \lambda(m,n)} \left\| \widetilde{f}(m,n,\cdot,\Delta) - \widehat{f}(m,n,\cdot,\Delta) \right\|_{L^1} = 0,$$

implying that $\widetilde{f}(m,n,\cdot,\lambda(m,n)) = \widehat{f}(m,n,\cdot,\lambda(m,n))$ a.s. (t). Both $\widetilde{f}(m,n,\cdot,\lambda(m,n))$ and $\widehat{f}(m,n,\cdot,\lambda(m,n))$ are of bounded variation, so by Lemma 2.1 and Remark 2.1 of Bressan (2000), by taking right continuous versions, they are in fact equal, a contradiction. Therefore, the second term in (22) converges to zero, so the whole expression converges to zero, and we have continuity of $\mathcal{T}$.

Next we let $\varsigma \to 0$. Existence of a limit solution (in terms of exit densities) follows from Helly's theorem. Uniqueness follows from the fact that for each $\varsigma > 0$ the Godunov approximation converges in L^1 to a limit: The approximations differ only by L^1 distance at most $\varsigma \cdot \chi_{mn}$ on link mn. So if there are two solutions with different Godunov approximation subsequences converging in L^1 to different limits with positive L^1 distance between them, then choosing ς small enough, we obtain a contradiction.

Next, apply Schauder's theorem to the space $\mathcal{G}$ with the L^1 norm and the mapping $\mathcal{T}$. This yields existence of at least one fixed point. To show that it is unique, find the earliest time at which the two solutions diverge. Observe that for given boundary conditions, behavior within a link is well-defined. So if two solutions exist and the earliest divergence between them occurs within a link, we have a contradiction. Thus, the divergence must occur at a node. Finding the earliest time at which such a divergence occurs, the boundary conditions must be ill-defined, a contradiction.

∎

5.2 Proof of Theorem 2

Proof: A mixed strategy is a measurable map $y : C \to [0,1]^{|\mathcal{R}| \times (T/\overline{\tau}-1)}$. We use the notation y^i to denote a vector component of y, so we impose the obvious condition $\sum_{i=1}^{|\mathcal{R}| \times (T/\overline{\tau}-1)} y^i(c) = 1$ almost surely in c.

First, we can define a *strategy distribution* as $\int_C y \equiv \prod_{i=1}^{|\mathcal{R}| \times (T/\overline{\tau}-1)} \int_C y^i(c) d\mu$. Second, we notice that the proof of Theorem 1 does not use the exact dynamic commuting route structure, but rather the strategy distribution induced by a dynamic route structure. In other words, the proof of Theorem 1 implies that for any given strategy distribution, there exists a unique traffic pattern. Information about which commuter plays each strategy is irrelevant.

Third, we define the utility of a commuter for a mixed strategy and any strategy distribution. Fix $c \in C$. The utility function $u(c; \tau^d, l, R)$ was given in (10). For pure strategy i corresponding to $l(c), R(c), \tau^d(c)$, this is written as $\widehat{u}^i(c, \int_C y) = u(c; \tau^d, l, R)$. We have argued that in the end the traffic pattern depends only on the strategy distribution. For technical reasons, it is useful here to define $u^i(c, \int_C y) \equiv -\infty$ if $\pi_1(R(c)) \neq O(c)$ or $\pi_{l(c)}(R(c)) \neq D(c)$; utility was undefined for this circumstance. Then for commuter $c \in C$, we can write the utility from the use of pure strategy i (a route and time of departure) given an aggregate strategy profile $\int_C y$, as $\widehat{u}^i(c, \int_C y)$ and $\widehat{u}(c, y) = \prod_{i=1}^{|\mathcal{R}| \times (T/\overline{\tau} - 1)} \widehat{u}^i(c, \int_C y)$, where the dynamic route structure (τ^d, l, R) generates the strategy distribution $\int_C y$. For this to be well-defined, we are using the fact that the utility will depend only on the strategy distribution generated by the dynamic commuting route structure, and the fact that the dynamic route structure can now be chosen arbitrarily subject to the strategy distribution since we no longer stick to the requirement that the origin and destination nodes are pre-specified. Finally, we can define the utility of commuter c from using mixed strategy $y(c)$ by $y(c) \cdot \widehat{u}(c, y)$.

It is clear from this set of definitions that our model satisfies two of the assumptions of Schmeidler (1973), namely the measurability assumption (b) and the fact that utility depends only on the strategy distribution, not on individual strategies. Assumption (a), regarding the continuity of $\widehat{u}$ in its second argument, remains to be verified.

We take a sequence of mixed strategies $\{y_q\}_{q=1}^{\infty}$ such that $\lim_{q \to \infty} y_q = y$ in the L^1 weak topology, and prove that for each $c \in C$, $\lim_{q \to \infty} \widehat{u}(c, y_q) = \widehat{u}(c, y)$. Our hypothesis implies $\lim_{q \to \infty} \int_C y_q = \int_C y$. Let $g \in \mathcal{G}$ be the fixed point associated with the strategy profile y,[69] and let $g_q \in \mathcal{G}$ be the fixed point associated with the strategy profile y_q. Thus, we have an associated sequence $\{g(\cdot)_q\}_{q=1}^{\infty} \subseteq \mathcal{G}$ where for each q, $g(\cdot)_q = \mathcal{T}(g(\cdot)_q)$. Since $\mathcal{G}$ is compact, there is a converging subsequence. Now pass to *any* converging subsequence, call it $\left\{g(\cdot)_{q_p}\right\}_{p=1}^{\infty} \subseteq \mathcal{G}$, where $\lim_{p \to \infty} g(\cdot)_{q_p} = \widehat{g}(\cdot)$. By continuity of $\mathcal{T}$, $\widehat{g}(\cdot) = \mathcal{T}(\widehat{g}(\cdot))$. Hence, $\widehat{g} = g$, and $\lim_{q \to \infty} g_q = g$. We use an analogous argument below for both density and progress along a link.

[69] Although y represents a mixed strategy, as we have noted, all that matters is the the distribution over routes and departure times, so arrival times can be found uniquely for each mixed strategy profile using Theorem 1.

Define

$$\mathcal{F}_{mn} \equiv \Big\{\widehat{f}(m,n,\cdot,\cdot) \text{ measurable on } [0,\bar{t}] \times [0,\lambda(m,n)]$$
$$\mid 0 \le \widehat{f}(m,n,\cdot,\cdot) \le \overline{f} \text{ a.s., } \widehat{f}(m,n,0,\Delta) = 0 \ \forall\ \Delta \ge 0\Big\}$$

Then we can define:

$$\mathcal{F} \equiv \prod_{m,n=1,\ m\neq n}^{N} \mathcal{F}_{mn}$$

We denote a typical element of $\mathcal{F}$ by $\widehat{f} = \left(\widehat{f}(m,n,\cdot,\cdot)\right)_{m,n=1,\ m\neq n}^{N}$.

Now for each q there exists a unique solution $\widehat{f}_q \in \mathcal{F}$ associated with boundary conditions g_q. There is also a unique density $\widehat{f} \in \mathcal{F}$ associated with g. Impose the weak* topology on the densities as a subset of L^∞. Applying the Banach-Alaoglu theorem, there is a converging subsequence. Now pass to any converging subsequence, call it $\left\{\widehat{f}(\cdot)_{q_p}\right\}_{p=1}^{\infty} \subseteq \mathcal{F}$, where $\lim_{p\to\infty} \widehat{f}(\cdot)_{q_p} = \widetilde{f}(\cdot)$, and where convergence is pointwise a.s. in (τ,Δ). As in the proof of Theorem 1, it must be that $\widehat{f} = \widetilde{f}$ a.s.(t) for each fixed Δ and a.s.(Δ) for each fixed t.

Now for each q there exists a unique solution $\delta_q \in \mathcal{D}$ associated with flow $\widehat{f}_q$. There is also a unique solution $\delta \in \mathcal{D}$ associated with $\widehat{f}$. Impose the uniform topology on the solutions as a subset of C_0. Applying Ascoli's theorem, there is a converging subsequence. Now pass to any converging subsequence, call it $\{\delta(\cdot)_{q_p}\}_{p=1}^{\infty} \subseteq \mathcal{D}$, where $\lim_{p\to\infty} \delta(\cdot)_{q_p} = \widehat{\delta}(\cdot)$.

Next define $\widehat{e}(m,n,\widehat{\tau},\tau)_{q_p} \equiv \widehat{f}(m,n,\tau,\delta_{mn}(\widehat{\tau},\tau)_{q_p})_{q_p}$. The function $\widehat{e}$ follows a cohort that begins at $\widehat{\tau}$ along link mn. Since density can only rise along a link,

$$TV\left(\widehat{f}(m,n,\cdot,\delta_{mn}(\widehat{\tau},\cdot)_{q_p})_{q_p}\right) \le \overline{f}$$

So applying Helly's theorem and passing to a further subsequence if necessary,

$$\lim_{q\to\infty} \widehat{e}(m,n,\widehat{\tau},\tau)_{q_p} = \widehat{e}(m,n,\widehat{\tau},\tau) \tag{23}$$

where convergence is pointwise in τ and $TV\left(\widehat{e}(m,n,\cdot,\widehat{\tau})\right) \le \overline{f}$.

So for each p,

$$\frac{\partial \delta_{mn}(\widehat{\tau},\tau)_{q_p}}{\partial \tau} = \upsilon\left(f(m,n,\tau,\delta(\widehat{\tau},\tau)_{q_p})_{q_p}, x_{mn}\right) = \upsilon\left(\widehat{e}(m,n,\widehat{\tau},\tau)_{q_p}, x_{mn}\right)$$

so

$$\lim_{p\to\infty} \frac{\partial \delta_{mn}(\widehat{\tau},\tau)_{q_p}}{\partial \tau} = \upsilon\left(\widehat{e}(m,n,\widehat{\tau},\tau), x_{mn}\right) \tag{24}$$

Next suppose that $\lim_{p\to\infty} \frac{\partial\delta(\widehat{\tau},\tau)_{q_p}}{\partial\tau} \neq \frac{\partial\widehat{\delta}(\widehat{\tau},\tau)}{\partial\tau}$ on a set of positive measure in τ. Hence, by the fundamental theorem of calculus and Lebesgue's dominated convergence theorem, there exists τ' such that

$$\begin{aligned}
\widehat{\delta}_{mn}(\widehat{\tau},\tau') &= \int_0^{\tau'} \frac{\partial\widehat{\delta}_{mn}(\widehat{\tau},\tau)}{\partial\tau} d\tau = \int_0^{\tau'} \lim_{p\to\infty} \frac{\partial\delta_{mn}(\widehat{\tau},\tau)_{q_p}}{\partial\tau} d\tau \\
&= \lim_{p\to\infty} \int_0^{\tau'} \frac{\partial\delta_{mn}(\widehat{\tau},\tau)_{q_p}}{\partial\tau} d\tau \neq \int_0^{\tau'} \frac{\partial\widehat{\delta}_{mn}(\widehat{\tau},\tau)}{\partial\tau} d\tau = \widehat{\delta}_{mn}(\widehat{\tau},\tau')
\end{aligned}$$

This is obviously a contradiction. So $\lim_{p\to\infty} \frac{\partial\delta_{mn}(\widehat{\tau},\tau)_{q_p}}{\partial\tau} = \frac{\partial\widehat{\delta}_{mn}(\widehat{\tau},\tau)}{\partial\tau}$ a.s. In fact, from (24), continuity of υ, and (23) we know convergence is pointwise in τ. Hence,

$$\frac{\partial\widehat{\delta}_{mn}(\widehat{\tau},\tau)}{\partial\tau} = \upsilon\left(\widehat{e}(m,n,\widehat{\tau},\tau), x_{mn}\right)$$

From (6) we know that $\delta_{mn}(\widehat{\tau},\widehat{\tau}) = \widehat{\delta}_{mn}(\widehat{\tau},\widehat{\tau}) = 0$, so by uniqueness of the solution to (5), $\widehat{\delta}_{mn}(\cdot) = \delta_{mn}(\cdot)$.

Fix a route r of length ℓ and a departure time $\widehat{\tau}$. Define

$$\begin{aligned}
\tau^*_{mn}(\widehat{\tau}) &\equiv \min\left\{0 \leq \tau \leq \bar{t} \mid \delta_{mn}(\widehat{\tau},\tau) = \lambda(m,n)\right\} \\
&= \delta^{-1}_{mn}(\lambda(m,n))(\widehat{\tau})
\end{aligned}$$

The function $\tau^*_{mn}(\widehat{\tau})$ can be viewed as a relatively simple implicit function. Since $\frac{\partial\widehat{\delta}_{mn}(\widehat{\tau},\tau)}{\partial\tau} \geq \upsilon(\overline{f}, x_{mn}) > 0$ and $\frac{\partial\widehat{\delta}_{mn}(\widehat{\tau},\tau)}{\partial\widehat{\tau}} \leq -\upsilon(0, x_{mn}) < 0$, $\tau^*_{mn}(\widehat{\tau})$ is well-defined, strictly increasing, and continuous.[70] Now let τ^d and $\tau^{d\prime}$ be origin departure time choices for route r, and let τ and τ' be associated perturbations, where $\widehat{\tau} = \tau^d + \tau$ and $\widehat{\tau}' = \tau^{d\prime} + \tau'$. Thus, arrival time at the final destination can be written as: $\widehat{\tau}_\ell\left(\tau^d + \tau'\right) = \tau^*_{\pi(\ell-1),\pi(\ell)}\left(\tau^*_{\pi(\ell-2),\pi(\ell-1)}\left(\cdots\tau^*_{\pi(1),\pi(2)}\left(\tau^d+\tau'\right)\cdots\right)\right)$. Define

$$\begin{aligned}
\Upsilon_r \equiv \Bigg\{ & \widehat{\tau}_\ell : [0,T] \to [0,\bar{t}] \text{ measurable} \mid \left|\widehat{\tau}_\ell\left(\tau^d+\tau\right) - \widehat{\tau}_\ell\left(\tau^{d\prime}+\tau'\right)\right| \\
&\leq \left(\max_{m,n}\left\{\frac{\upsilon(0,x_{mn})}{\upsilon(\overline{f},x_{mn})}\right\}\right)^{\ell-1} \cdot \left|\tau^d + \tau - \tau^{d\prime} - \tau'\right| \Bigg\}
\end{aligned}$$

By Ascoli's theorem, Υ_r is a compact subset of C_0.

For each q there is a unique $\delta(\cdot)_q$ and thus a unique $\widehat{\tau}_\ell(\cdot)_q$. There is also a unique $\widehat{\tau}_\ell(\cdot) \in \Upsilon_r$ associated with $\delta(\cdot)$. So there is a converging

[70] This can either be proved directly or a non-C^1 version of the implicit function theorem can be used.

subsequence associated with $\{\delta(\cdot)_q\}_{q=1}^{\infty}$. Now take any converging subsequence of $\left\{\widehat{\tau}_\ell(\cdot)_q\right\}_{q=1}^{\infty}$, call it $\left\{\widehat{\tau}_\ell(\cdot)_{q_p}\right\}_{p=1}^{\infty}$. It has a limit: $\widehat{\tau}'_\ell(\cdot) \in \Upsilon_r$. Suppose that $\widehat{\tau}'_\ell(\cdot) \neq \widehat{\tau}_\ell(\cdot)$. Now since $\left\{\delta(\cdot)_{q_p}\right\}_{p=1}^{\infty}$ converges uniformly to $\delta(\cdot)$, for each $\tau^d + \tau' \in [0, T]$, $\lim_{p\to\infty} \widehat{\tau}_\ell\left(\tau^d + \tau'\right)_{q_p} = \widehat{\tau}_\ell\left(\tau^d + \tau'\right)$, so in fact $\widehat{\tau}'_\ell(\cdot) = \widehat{\tau}_\ell(\cdot)$, a contradiction.

Apply Schmeidler (1973), theorems 1 and 2, there exists a Nash equilibrium in pure strategies.

Finally, consider the case where Φ is non-decreasing (instead of strictly increasing) in f, and as always $\Phi(f) \equiv v(f) \cdot f$. Let $\widehat{v}(f) = v(f) + \epsilon$, where $\epsilon > 0$ is small. Then since $v(f)$ is non-increasing in f, so is $\widehat{v}(f)$. Moreover, $\widehat{\Phi}(f) \equiv \widehat{v}(f) \cdot f = (v(f) + \epsilon) \cdot f = \Phi(f) + \epsilon \cdot f$, so $\widehat{\Phi}(f)$ is strictly increasing in f. Apply our results to the modified game using $\widehat{v}(f)$ and $\widehat{\Phi}(f)$ to obtain an equilibrium in pure strategies for each ϵ. As the number of strategies is actually finite, we can find an accumulation point of the strategy profile as $\epsilon \to 0$. Using continuity of the payoffs (as demonstrated above), by a standard argument the accumulation point is an equilibrium profile for $\epsilon = 0$.

∎

5.3 Proof of Theorem 3

Proof: Given an outward bound shrubbery network, the only possible route choice is at the third to last node on a route. The strategy profile we propose as a Pareto efficient Nash equilibrium is to distribute each type of commuter, where type is defined as an origin-destination pair, uniformly across all departure times and across possible routes from that type's origin to destination. This creates equal volume. We choose the minimal density consistent with this volume (since volume is continuous, such a density exists). Clearly this is a Nash equilibrium, as all commuters of a given type have the same travel time and thus receive the same utility. Now suppose that there is a strategy profile that Pareto dominates the Nash equilibrium profile. Thus, it must be that there is some departure time and route that has a higher than average volume. Since volume $\Phi(f) = f \cdot v(f)$, and $v(f)$ is strictly decreasing, for this departure time, there is some type that has a higher than average density (where the average is over departure times and routes for this type). The commuters of this type with this departure time will have a longer commute than at Nash equilibrium, contradicting that the alternative strategy profile Pareto dominates the Nash equilibrium profile. ∎