

Article

# Analyzing the Impact of Traffic Congestion Mitigation: From an Explainable Neural Network Learning Framework to Marginal Effect Analyses

Jianping Sun <sup>1,2</sup>, Jifu Guo <sup>1,2,\*</sup>, Xin Wu <sup>3,\*</sup>, Qian Zhu <sup>4</sup>, Danting Wu <sup>1</sup>, Kai Xian <sup>2</sup> and Xuesong Zhou <sup>2,3</sup>

<sup>1</sup> School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; sunjp2004@126.com (J.S.); 16120903@bjtu.edu.cn (D.W.)

<sup>2</sup> Beijing Transport Institute, Beijing 100073, China; xiank@bjtrc.org.cn (K.X.); xzhou74@asu.edu (X.Z.)

<sup>3</sup> School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85281, USA

<sup>4</sup> School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China; 15928665871@163.com

\* Correspondence: guojf@bjtrc.org.cn (J.G.); xinwu3@asu.edu (X.W.); Tel.: +1-(480)-4105902 (X.W.)

Received: 1 March 2019; Accepted: 7 May 2019; Published: 15 May 2019



**Abstract:** Computational graphs (CGs) have been widely utilized in numerical analysis and deep learning to represent directed forward networks of data flows between operations. This paper aims to develop an explainable learning framework that can fully integrate three major steps of decision support: Synthesis of diverse traffic data, multilayered traffic demand estimation, and marginal effect analyses for transport policies. Following the big data-driven transportation computational graph (BTCG) framework, which is an emerging framework for explainable neural networks, we map different external traffic measurements collected from household survey data, mobile phone data, floating car data, and sensor networks to multilayered demand variables in a CG. Furthermore, we extend the CG-based framework by mapping different congestion mitigation strategies to CG layers individually or in combination, allowing the marginal effects and potential migration magnitudes of the strategies to be reliably quantified. Using the TensorFlow architecture, we evaluate our framework on the Sioux Falls network and present a large-scale case study based on a subnetwork of Beijing using a data set from the metropolitan planning organization.

**Keywords:** computational graph; traffic demand estimation; congestion mitigation; marginal analyses; TensorFlow

## 1. Introduction

As the population size, economic growth, and personal travel activities continue to increase, traffic congestion in metropolitan areas remains one of the major concerns for urban transportation planning and management agencies [1]. Recurring traffic congestion is often caused by a regional imbalance involving excess demand and limited infrastructure, while nonrecurring congestion is mainly caused by various traffic incidents and severe weather conditions. Many advanced intelligent transportation systems (and more recently, smart transportation initiatives) have been developed to enable more sophisticated traffic demand management and reliable selection of effective congestion mitigation strategies.

The emerging traffic big data environment brings both opportunities and challenges to the core functions of traffic demand estimation and control. For example, Toole et al. (2015) demonstrated progress in using sampled mobile phone data to infer the most traveled routes, which are extremely

difficult to discover from traditional survey data [2]. By mining information from different data sources (e.g., traffic count sensors, mobile phone service records, floating car data, and automatic vehicle identification (AVI) data), planners and decision-makers hope to gain deep insight into human mobility patterns, which will accordingly increase the confidence level and reduce the uncertainty of transportation strategy evaluations [3–8]. Over the past several decades, many researchers have highlighted the need for a fully integrated connection from big data sources to pattern recognition and the deployment of final traffic demand/control strategies [9–12].

While it is important to extract and visualize spatial and temporal patterns from diverse data sources collected from advanced sensor networks, transportation planners and managers are more concerned with how to gain insight into the different causes of congestion because the benefits of different urban traffic congestion mitigation strategies must be better quantified in a data-rich environment. This expectation requires transformative advances not only within the traditional domains of traffic behavioral analysis and demand estimation, but also within the emerging field of big data itself [13,14]. Moreover, how to develop an explainable deep learning framework has been identified as one of the top 10 key challenges for the AI community [15].

Based on the above research needs, we attempt to bridge two important gaps between the three functional layers in a data-centric transportation planning and management framework: Models, data, and policy scenario evaluation, as shown in Figure 1.

1. The gap between layers 1 and 2 is related to the wide variety of available traffic data sources and the fact that core traffic network models are typically difficult to calibrate consistently.
2. The gap between layers 2 and 3 lies in the fact that traffic models contain many elements with limited certainty, whereas mission-critical scenario evaluation requires reliable current-state estimates and policy-sensitive forecasts.

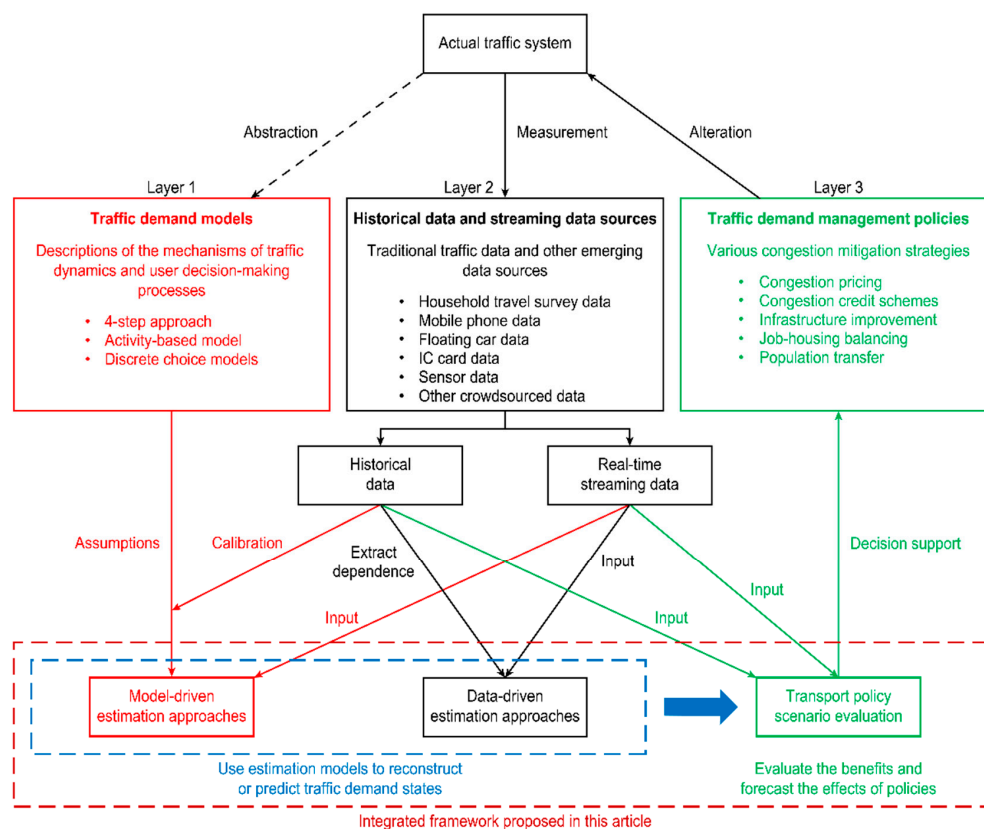


Figure 1. Three basic layers of a traffic demand management system.

## 2. Literature Review

A traffic demand management system is characterized by three modules: Traffic demand models, different data sources, and traffic demand management policies (e.g., congestion mitigation policies), as displayed in Figure 1. Both model-driven and data-driven approaches provide estimates of traffic demand states for evaluating the effects of congestion mitigation strategies in different scenarios.

### 2.1. Model-Driven Travel Demand Estimation Approaches

The first module includes conventional model-driven approaches for traffic demand estimation that are motivated by a priori knowledge of the kinds of decision-making that must be performed while traveling [16,17]. Traditionally, these decision-making processes are conceived as four-step choices, including trip generation, trip distribution, mode-split models, and traffic assignment [18]. Individual route choice behavior is assumed based on the user equilibrium condition, in which each user chooses the optimal route that minimizes his or her travel costs [19]. Discrete choice models are also used to estimate the modes and route choice behaviors of users [20]. Recently, a number of researchers have attempted to develop activity-based models to estimate users' trip chains [21] (see Figure 1).

One model-driven approach that focuses on network-wide demand estimation and prediction of origin-destination (OD) demand and route choice probabilities, as well as the resulting traffic network flow patterns, is based on "link proportions" obtained through static and dynamic traffic assignment (STA/DTA). However, in most of the literature, the four-step process and demand estimation are independent. Some studies simply combine STA/DTA with OD matrix estimation (ODME) using bi-level programming models [22–24], where the lower level is an STA/DTA model and the upper-level model uses econometric estimators to obtain estimates of the OD matrix.

It should be noted that the real-world deployment of model-driven methods faces technical and computational barriers. Different types of models are not consistent in their formulations of traffic demand. For example, dynamic network assignment for the management of information to travelers (DynaMIT) uses a deterministic queuing and speed model [25], dynamic traffic assignment and simulation for advanced network informatics (DYNASMART) adopts a Greenshield model [26], and the visual interactive system for transportation algorithms (VISTAs) use a cell transmission model [27]. Because the STA/DTA process provides "link proportions" as inputs for demand estimation, different models might lead to different estimates.

Conversely, the accuracy of STA/DTA models also depends heavily on the data available to calibrate the models (see Figure 1). In particular, DTA models require excessive data processing. The input of a DTA model includes demand-side inputs, the OD demand, supply-side inputs, and the network topology. The barriers involve the collection of representative data, the transformation of those data into the DTA input format, and the modeling and evaluation of control strategies.

One task undertaken in this paper is to develop a general framework that integrates the model-driven approach with the traffic demand estimation process using the forward and backward mechanism of computational graphs (CGs). Furthermore, the concept of CGs provides us with a comprehensive learning framework to coordinate the multiple traffic data sources required in traditional four-step methods.

### 2.2. Data-Driven Travel Demand Estimation Approaches

Beyond the use of traditional household survey data in existing models and studies [19,20,28], a number of data-driven traffic demand estimation approaches and online streaming data-driven models have been proposed (see Figure 1) that rely on emerging big data sources [2,28–35].

Inductive loops, radars, and cameras have become the predominant fixed vehicle detection devices in most cities due to their low unit equipment costs and relatively high performance [10,36,37]. Many existing ODME methods focus on the estimation of OD matrices using link counts derived from fixed sensor observations. It has been widely recognized that OD flow patterns are not unique because

of the non-uniqueness of the path flows and the limited observation data collected from sensors on links [38,39]. The in-car navigation Global Position System (GPS) technology has matured into a rapidly growing industry. Floating car data have been used in conjunction with link count data and video camera data to derive OD matrices and analyze route choices and trip length distributions [40,41]. Furthermore, mobile phone data make it possible to capture the characteristics of human mobility. Hao et al. [21] applied mobile phone data in activity-based models to generate user tours. Interestingly, Bonnel et al. [42] compared OD matrices that were separately generated from household surveys and mobile phone data and found a large disparity between the two outputs. In the existing literature, secondary data sources from social network services have also been used to study individual travel behaviors [43–45].

In data-driven traffic demand estimation models, it is extremely challenging to determine an overall cause–effect explanation based on the partially observable information acquired from multiple data sources. In addition, to achieve the full potential of data-assisted traffic congestion mitigation and demand management, it is critical to have a data mining platform that is sensitive to the requirements of urban planners in a heterogeneous data environment.

From the perspective of big data fusion, Wu et al. [46] proposed the **big data-driven transportation computational graph** (BTCG) framework as a fundamental mathematical modeling tool for performing multilayered traffic demand estimation based on the forward and backward propagation mechanism. The multilayered representation used in this framework structurally models four levels of travel demand variables, including trip generation, OD matrices, path flows, and link flows. Ma et al. (2019) extended the framework to the problem of multiclass traffic demand estimation [47]. This paper further investigates the application of CGs to facilitate data-assisted traffic congestion mitigation and traffic demand management.

### 2.3. Existing Congestion Mitigation Strategies

Traffic congestion can be defined in many ways. Because traffic demands change constantly, city congestion varies from day to day. However, approximately half of all congestion recurs every day in the same location. This paper focuses on investigating such recurrent congestion (see Figure 1). Traffic congestion can be reduced using various mitigation strategies. The most intuitive solution is to build new roadways to add new capacity. However, the famous Braess paradox shows that sometimes, new capacity may decrease utility for users [19]. Moreover, in metropolitan areas, the right-of-way is expensive; consequently, other mitigation options must be considered. The advent of ride-sharing services means that providing high-occupancy vehicle (HOV) lanes for vehicles with two or more passengers is also a feasible solution for expanding capacity.

Congestion pricing strategies [48–50] and credit schemes (e.g., tradable credit tickets) [51–56] based on market mechanisms are also effective tactics for mitigating congestion. Temporally, variable pricing strategies work to shift peak-hour traffic flows to off-peak periods. Spatially, these strategies incentivize people to consider using public transit or taking alternative routes to their destinations. Various pricing schemes have been investigated in the existing literature, such as trip-rate-based pricing, travel-distance-based pricing, and travel-time-based pricing [56–59].

Despite the success of congestion pricing in some cities (London has seen benefits from the implementation of congestion pricing in 2003) [60], toll schemes do not appeal to the public. Based on this fact, Wu et al. [61] indicated the importance of calculating the marginal external cost of one extra toll on links because congestion pricing may lead to a possible loss of overall social welfare. Wu et al. [61] recommended a Pareto-improving strategy to increase social benefit without increasing the travel expense for every stakeholder. However, a Pareto-efficient system is not always guaranteed. In most cases, it is important to develop an efficient method of calculating the marginal effect of one extra toll on each link and the tradeoff between benefits and costs.

Traffic congestion is also closely related to urban planning factors, such as job locations, land use, and house (rental) pricing. From the urban planning perspective, to reduce long-term congestion,

planners tend to implement population transfers and redesign residential locations and workplaces to balance employment supply and demand [62–66]; however, to the best of our knowledge, very few studies have comprehensively considered the marginal effects of macroscopic policies on the traffic volume on each link.

Another contribution of this paper is to extend the computational graph (CG) based learning framework to evaluate the marginal effects of different congestion mitigation strategies in a big data context.

#### 2.4. Outline

The remainder of this paper is organized as follows. Section 3 presents a general description of the overall system architecture and a conceptual illustration based on CGs. In Section 4, we propose a compact description of the traffic demand estimation problem based on the existing CG framework and extend the deep learning framework to the evaluation of the effects of congestion mitigation strategies by mapping those strategies onto CGs. Section 5 reports computational results obtained based on the Sioux Falls network and a subnetwork of Beijing to demonstrate the effectiveness and applicability of our framework. Section 6 concludes the paper and discusses some possible extensions of our work.

### 3. System Architecture and Conceptual Illustration

As background, in this section, we introduce the overall architecture of the “super-simulation system” that is designed for the new generation of data-driven traffic management platforms. It should be noted that the proposed CG-based learning framework for traffic demand estimation is only one of the engines of this “super-simulation” system. The architecture depicted in Figure 2 reveals the relationship between the CG-based demand estimation engine and the STA/DTA simulation engine.

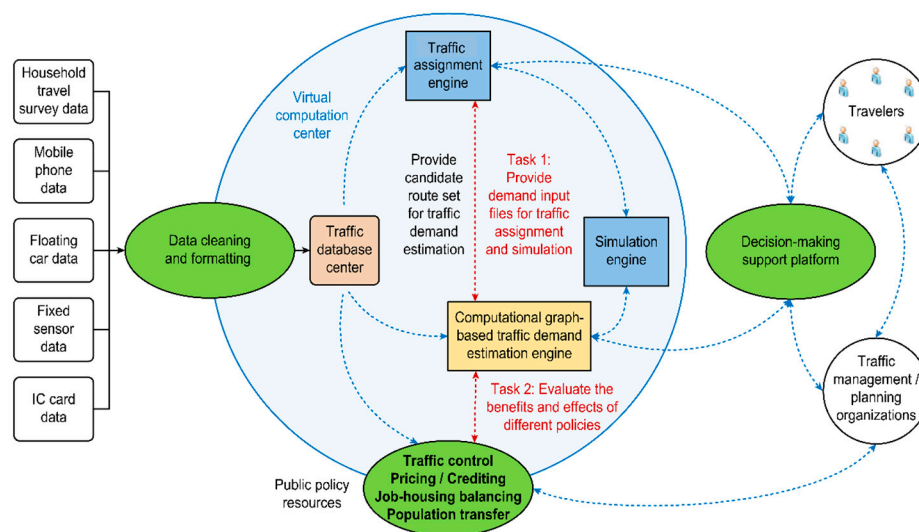


Figure 2. The system architecture of the “super-simulation” system.

In addition, we will briefly report on how big data sources and real-world congestion mitigation strategies interact with our traffic demand estimation engine (i.e., the CG-based learning framework). This discussion is followed by a comprehensive review of the concept of a CG and its mathematical relationship to discrete choice models, which shows that the CG representation is sufficiently flexible to be used in formulating various problems in the field of transportation modeling.

#### 3.1. System Architecture

The proposed CG-based learning framework is an important component of the overall system, as shown in Figure 2. A central traffic database is established that receives multiple types of traffic data,

including household survey data, mobile phone data, floating car data, loop detector data, and data from other possible sources. After being cleaned and formatted, the data are input into the CG-based traffic demand estimation engine and the traffic assignment engine. The outputs of these two engines can be viewed as the input to the simulation engine for visualizing the traffic system [67].

As shown in Figure 2, the CG-based traffic demand estimation engine and the traffic assignment engine (for STA/DTA) provide feedback to each other. On the one hand, the CG-based traffic demand estimation engine provides input files for the traffic assignment engine. On the other hand, the traffic assignment engine initializes the candidate route set for the traffic demand estimation engine. Zhuge et al. [68] reported on how to generate the candidate route set using a tree-based assignment (20 iterations are usually required to arrive at relatively stable conditions; see [68]). The CG-based traffic demand estimation engine should also be fully integrated with the transport policy resources. Then, we will be able to predict the possible demand patterns under different transport policies.

In summary, the essential tasks of the traffic demand estimation engine are as follows:

1. Estimate multilayered traffic demands (i.e., trip generation, trip distribution, and path/link flows).
2. Produce the input file for the traffic assignment engine.
3. Evaluate the effects of various transport policies.

Because this paper focuses on policy analysis, we do not consider the interaction between traffic assignment and traffic demand estimation. We invoke the STA engine only to generate the candidate physical route set. Furthermore, we consider only the private car mode in a regional network during rush hour. The traffic demand estimation engine will be extended in the future.

### 3.2. Computational Graphs and Marginal Effect Analyses

The CG concept serves as a basic description language for many machine learning methods, such as artificial neural networks [69]. In particular, CGs serve as the basis of a fine-grained framework that can be used to decompose complex composite functions into a sequence of nested mappings. Each mapping in a CG (represented by an edge) expresses an elementary operation involving only one or two arguments.

In the field of computer science, CGs provide a paradigm for developing machine learning systems in a big data environment because they enable the efficient and flexible management, operation, and control of data sources using intuitive graphical representations and elementary mathematical operations. Consequently, CGs have become basic building blocks for many current machine learning software packages [70–73]. For example, TensorFlow, developed by the Google Brain team, is a machine learning library based on a CG framework.

CGs are also related to traditional transportation demand modeling. Below, we present an illustrative example adapted from Koppelman and Bhat (2006) [20] in the context of discrete choice models to show how to formulate a mode-split model (a multinomial logit model) using a CG. We refer readers to a previous introduction to the relationship between CGs and deep learning methods [69]. Furthermore, Zhao et al. (2019) provide a good discussion of the relationship between machine learning and logit modeling [74].

#### 3.2.1. Comparison between Discrete Choice Modeling and Computational Graph Modeling Based on a Mode-Split Model

Consider an example with two available alternatives: Drive alone (DA) and transit (TR). The planners intend to encourage users to use transit services instead of private driving to mitigate congestion in the driving network. The following utility function implies that the decision-maker preferences are a function of the average income (INC) and travel time (TT). For simplicity, we do not consider interaction terms or constant terms:

$$U_{DA} = \beta_{DA,1}INC - \beta_2TT_{DA}, \quad (1)$$

$$U_{TR} = \beta_{TR,1}INC - \beta_2 TT_{TR}. \quad (2)$$

Then, the probability of choosing DA can be calculated using the following multinomial logit model:

$$P_{DA} = \frac{\exp(U_{DA})}{\exp(U_{TR}) + \exp(U_{DA})} = \frac{1}{1 + \exp(U_{TR} - U_{DA})}. \quad (3)$$

Table 1 shows an example in which  $\beta_{DA,1} = 0.004$ ,  $\beta_{TR,1} = 0$ , and  $\beta_2 = 0.02$  for an individual from a household with a \$50,000 annual income facing travel times of 30 and 50 min for DA and TR, respectively. The utility and probability calculations are shown in Table 1.

**Table 1.** Utility and probability calculations with transit as the base alternative.

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive alone	$U_{DA} = \beta_{DA,1}INC - \beta_2 TT_{DA} = 0.004 \times 50 - 0.02 \times 30$	-0.4	0.6703	$P_{DA} = 0.65$
Transit	$U_{TR} = \beta_{TR,1}INC - \beta_2 TT_{TR} = 0 \times 50 - 0.02 \times 50$	-1	0.3679	$P_{TR} = 0.35$
	$\beta_{DA,1} = 0.004; \beta_{TR,1} = 0; \beta_2 = -0.02$		$\Sigma = 1.0382$	

As indicated by Goodfellow et al. [65], a logit model for a binary mode choice can be viewed as the simplest type of neural network, with three layers corresponding to two steps of calculations.

1. The first layer is a stack of neurons that express the utility function of the differences between pairs of alternatives for predicting the DA probability:

$$U_{TR} - U_{DA} = (\beta_{TR,1} - \beta_{DA,1})INC - \beta_2(TT_{TR} - TT_{DA}). \quad (4)$$

Consider  $-\beta_1 = \beta_{TR,1} - \beta_{DA,1}$ ; then,

$$U_{TR} - U_{DA} = -\beta_1 INC - \beta_2(TT_{TR} - TT_{DA}) = -\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR} = \beta_2 x, \quad (5)$$

where  $x = (-INC, TT_{DA} - TT_{TR})$  can be viewed as a vector that includes both income and travel time.

2. The second layer applies the logistic sigmoid function,  $\sigma(U_{DA} - U_{TR})$ , as an activation function to squeeze the output of the linear utility function into the interval (0, 1).
3. The third layer calculates the probability of choosing DA:

$$P_{DA} = \sigma(U_{DA} - U_{TR}) = \frac{1}{1 + \exp(U_{TR} - U_{DA})} = \frac{1}{1 + \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR})}. \quad (6)$$

The mode-split model is a composite function consisting of both a linear utility function (Equation (5)) and the sigmoid function (Equation (6)). Figure 3 shows how to describe this function using a CG.

There are seven operations in the expression: Two additions, three multiplications, one exponential operation, and one reciprocal operation. We can describe the mode-split model using the inputs, the outputs, and the following six intermediary variables as vertexes in a CG:

$$a = -\beta_1 INC, \quad (7)$$

$$b = -\beta_2 TT_{TR}, \quad (8)$$

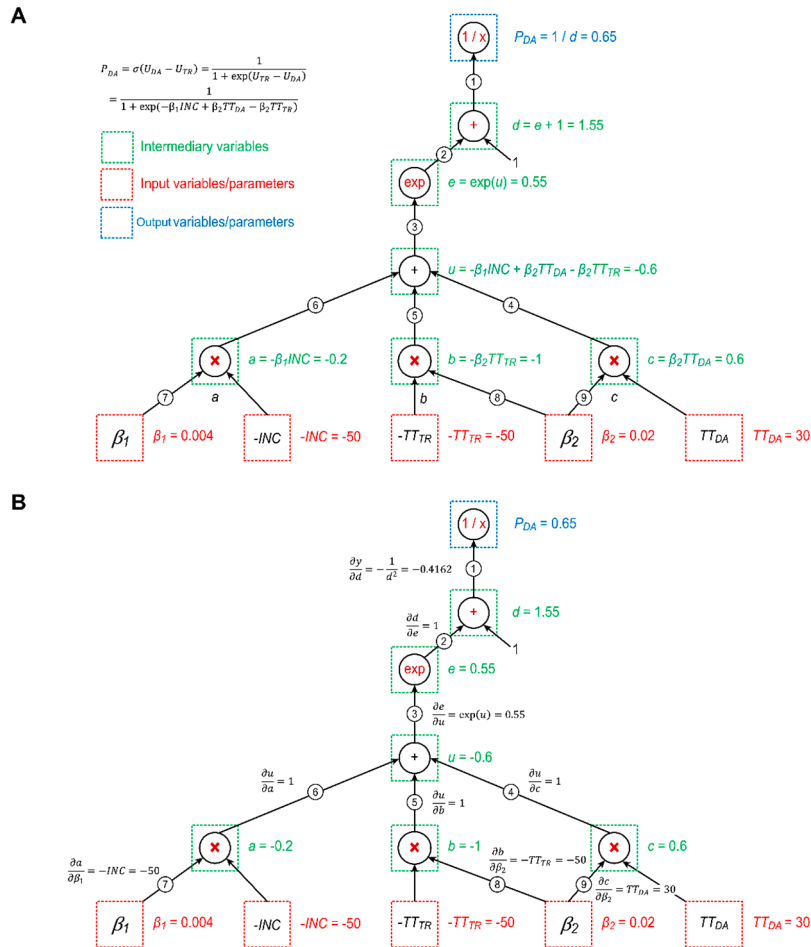
$$c = \beta_2 TT_{DA}, \quad (9)$$

$$u = a + b + c = -\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}, \quad (10)$$

$$e = \exp(u) = \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}), \tag{11}$$

$$d = u + 1 = \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}) + 1. \tag{12}$$

In the CG plotted in Figure 3, each edge corresponds to a derivative. For example, edge ① corresponds to the partial derivative of  $P_{DA}$  with respect to  $d$ , and edge ② corresponds to the partial derivative of  $d$  with respect to  $e$ . We evaluate the output value of  $P_{DA}$  by setting the input variables/parameters to certain values and then computing the vertexes progressively through the graph in the upward direction. Let us set  $\beta_1 = 0.004$ ,  $\beta_2 = 0.02$ ,  $INC = 50$  thousand dollars,  $TT_{DA} = 30$  min, and  $TT_{TR} = 50$  min. We obtain the same results as in Table 1 using the CG ( $P_{DA} = 0.65$ ).



**Figure 3.** An illustrative computational graph (CG) for a multinomial logit model for the function,  $\sigma(U_{DA} - U_{TR})$ ; (A) Forward propagation of CG; (B) Back propagation of CG.

To understand the partial derivatives in these cases, it is important to understand the chain rule in calculus. According to the chain rule, to calculate the partial derivative of  $P_{DA}$  with respect to any parameter, we need to sum over all possible paths from vertex  $P_{DA}$  to the vertexes of that variable/parameter in the graph while multiplying the derivatives on each edge on the same path. Examples are given below:

1.  $\frac{\partial P_{DA}}{\partial \beta_1}$  can be calculated by multiplying the partial derivatives on the path ①→②→③→⑥→⑦:

$$\frac{\partial P_{DA}}{\partial \beta_1} = \frac{\partial P_{DA}}{\partial d} \frac{\partial d}{\partial e} \frac{\partial e}{\partial u} \frac{\partial u}{\partial a} \frac{\partial a}{\partial \beta_1} = \left(-\frac{1}{d^2}\right) \exp(u) (-INC) = \frac{INC \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR})}{[\exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}) + 1]^2}. \tag{13}$$



2.  $\frac{\partial P_{DA}}{\partial \beta_2}$  can be calculated by summing the multiplied partial derivatives on the path ①→②→③→④→⑤ with the multiplied partial derivatives on the path ①→②→③→⑤→⑧:

$$\frac{\partial P_{DA}}{\partial \beta_2} = \frac{\partial P_{DA}}{\partial d} \frac{\partial d}{\partial e} \frac{\partial e}{\partial u} \frac{\partial u}{\partial b} \frac{\partial b}{\partial \beta_2} + \frac{\partial P_{DA}}{\partial d} \frac{\partial d}{\partial e} \frac{\partial e}{\partial u} \frac{\partial u}{\partial c} \frac{\partial c}{\partial \beta_2} = \left(-\frac{1}{d^2}\right) \exp(u) (TT_{DA} - TT_{TR}) = \frac{(TT_{TR} - TT_{DA}) \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR})}{[\exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}) + 1]^2}. \quad (14)$$

Typically, “loss errors” can be propagated to update the estimated variables, e.g.,  $\beta_1$  and  $\beta_2$ , across the CG using the stochastic gradient descent algorithm. This approach minimizes the sum of the Euclidean residuals for a set of individual observations labeled  $\bar{x}^m = (\overline{INC}^m, \overline{TT}_{DA}^m, \overline{TT}_{TR}^m)$  and binary samples,  $\bar{y}^m \in \{0, 1\}$ , where  $m = 1, 2, \dots, M$ :

$$F(\beta_1, \beta_2) = \min_{\beta_1, \beta_2} \frac{1}{2} \sum_{m=1}^M (\bar{P}_{DA}^m - \bar{y}^m)^2. \quad (15)$$

The gradients,  $\frac{\partial F(\beta_1, \beta_2)}{\partial \beta_1}$  and  $\frac{\partial F(\beta_1, \beta_2)}{\partial \beta_2}$ , can then easily be calculated easily as follows:

$$\frac{\partial F(\beta_1, \beta_2)}{\partial \beta_1} = \sum_{m=1}^M \frac{\partial F(\beta_1, \beta_2)}{\partial \bar{P}_{DA}^m} \frac{\partial \bar{P}_{DA}^m}{\partial \beta_1}, \quad (16)$$

$$\frac{\partial F(\beta_1, \beta_2)}{\partial \beta_2} = \sum_{m=1}^M \frac{\partial F(\beta_1, \beta_2)}{\partial \bar{P}_{DA}^m} \frac{\partial \bar{P}_{DA}^m}{\partial \beta_2}, \quad (17)$$

where  $\frac{\partial F(\beta_1, \beta_2)}{\partial \bar{P}_{DA}^m} = \bar{P}_{DA}^m - \bar{y}^m$ .

### 3.2.2. Marginal Effect Analyses Using a Computational Graph

Interestingly, the above formulations can equivalently be used to derive the marginal effect equations obtained with discrete choice models in [19]:

$$\frac{\partial P_{DA}}{\partial \beta_1} = \frac{INC \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR})}{[\exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}) + 1]^2} = INC \sigma(U_{DA} - U_{TR}) [1 - \sigma(U_{DA} - U_{TR})], \quad (18)$$

$$\frac{\partial P_{DA}}{\partial \beta_2} = \frac{(TT_{TR} - TT_{DA}) \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR})}{[\exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}) + 1]^2} = (TT_{TR} - TT_{DA}) \sigma(U_{DA} - U_{TR}) [1 - \sigma(U_{DA} - U_{TR})], \quad (19)$$

$$\frac{\partial P_{DA}}{\partial TT_{TR}} = \frac{-\beta_2 \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR})}{[\exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}) + 1]^2} = -\beta_2 \sigma(U_{DA} - U_{TR}) [1 - \sigma(U_{DA} - U_{TR})] \quad (20)$$

$$\frac{\partial P_{DA}}{\partial TT_{DA}} = \frac{\beta_2 \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR})}{[\exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}) + 1]^2} = \beta_2 \sigma(U_{DA} - U_{TR}) [1 - \sigma(U_{DA} - U_{TR})] \quad (21)$$

$$\frac{\partial P_{DA}}{\partial INC} = \frac{-\beta_1 \exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR})}{[\exp(-\beta_1 INC + \beta_2 TT_{DA} - \beta_2 TT_{TR}) + 1]^2} = -\beta_1 \sigma(U_{DA} - U_{TR}) [1 - \sigma(U_{DA} - U_{TR})]. \quad (22)$$

This relationship implies the potential importance of CGs in economics. Furthermore, the outstanding advantage of CGs is that, when a large number of parameters are involved in a real-world economic system, we can efficiently implement marginal analyses based on the principle of dynamic programming (DP) [69]. The DP principle can be applied in CGs to avoid duplicate calculations and sequentially update the derivatives along the “computing” paths in a backward fashion via the chain rule.

For example, in our case, for all three paths, we can first calculate:

$$\frac{\partial P_{DA}}{\partial d} \frac{\partial d}{\partial e} \frac{\partial e}{\partial u} = \sigma(U_{DA} - U_{TR})[1 - \sigma(U_{DA} - U_{TR})] = -0.4162 \times 1 \times 0.55 = -0.2284. \quad (23)$$

Then, we can reuse the intermediate result of  $-0.2284$  to calculate the following marginal values:

$$\frac{\partial P_{DA}}{\partial \beta_1} = -0.2284 \times (-50 \times 1) = 11.4217, \quad (24)$$

$$\frac{\partial P_{DA}}{\partial \beta_2} = -0.2284 \times (30 - 50) = 4.5687. \quad (25)$$

Provided that we estimate  $\beta_1 = 0.004$  and  $\beta_2 = 0.02$ , we can also use the intermediate result of  $-0.2284$  to calculate several important marginal values:

$$\frac{\partial P_{DA}}{\partial T_{TR}} = -0.2284 \times (-0.02) = 0.0046 = 0.46\%, \quad (26)$$

$$\frac{\partial P_{DA}}{\partial T_{DA}} = -0.2284 \times 0.02 = 0.0046 = -0.46\%, \quad (27)$$

$$\frac{\partial P_{DA}}{\partial INC} = -0.2284 \times (-0.004) = 0.0009 = 0.09\%. \quad (28)$$

These values imply that if the travel time of transit services decreases by 1 minute, then 0.46% more users will use the transit service system instead of private driving. Thus, the congestion in the driving network can be mitigated. By contrast, if the average income of users increases by 1 dollar, then 0.09% of users will use private driving instead of transit services.

#### 4. Congestion Mitigation Strategies Based on the Computational-Graph-Based Learning System

In Sections 4.1–4.3, we used a compact formulation to rewrite the multilayer CG representation proposed in [46], which captures three steps of the traditional four-step process: Trip generation, trip distribution, and path-flow-based traffic assignment. In contrast to [46], we described the methodology from the perspective of control theories instead of mathematical optimization. In Section 4.4, we attempt to map different congestion mitigation strategies on the CG representation as controllable inputs. Then, the marginal effects of these strategies can be calculated to reflect whether the corresponding control policies can increase the welfare of the transportation system.

##### 4.1. Variables

The three basic groups of estimated state variables used in the CG are as follows.

###### 1. Group I: Demand Variables

$\alpha$  indicates the trip generation variable vector, containing the trip generation results from all zones.  
 $\mathbf{q}$  indicates the trip distribution variable vector, containing the flow volume between all OD pairs.  
 $\mathbf{f}$  indicates the path flow variable vector, containing all flow volume on each path in the candidate route set generated by the traffic assignment engine.  
 $\mathbf{v}$  indicates the link flow variable vector, containing the flow volume on each link in the network.

To relate variables,  $\mathbf{f}$  and  $\mathbf{v}$ , we define  $\delta$  as a link-route incidence parameter matrix that indicates whether a route passes through a particular link.

###### 2. Group II: Proportional Variables

$\gamma$  indicates the OD split rate variable matrix expressing the rate at which each OD pair is selected from each traffic zone.

$\rho$  indicates the route choice proportion variable matrix, expressing the rate at which each route between each OD pair is selected.

### 3. Group III: Variables in Discrete Choice Models

In the field of traffic modeling,  $\rho$  can be determined by a multinomial logit model, which is the simplest type of stochastic network loading algorithm [18]. In the field of deep learning, the multinomial logit model is called the softmax function. The softmax function applies when we wish to represent a probability distribution over a discrete variable with multiple possible values. The softmax function can be regarded as a generalization of the logistic sigmoid function used in the illustrative example in the previous section:

$$\rho_r = \frac{\exp(U_r)}{\sum_{r \in P(w)} \exp(U_r)} \tag{29}$$

$$U_p = -\beta_{w,1}TC_r - \beta_{w,2}TT_r + \beta_w = -\beta_{w,1} \sum_{v \in V(p)} TC_v - \beta_{w,2} \sum_{v \in V(p)} TT_v + \beta_w, \tag{30}$$

where  $P(w)$  represents all candidate routes,  $r$ , between OD pairs, and  $V(r)$  represents the set of links that are passed through on route  $r$ . We express the softmax function (logit model) as follows:

$$\rho = \text{softmax}(\beta_1, \beta_2, \beta, TC, TT), \tag{31}$$

where  $TC$  and  $TT$  are the travel time parameter vector and the toll parameter vector, respectively.

$U_r$  denotes the utility of route  $r$ , where

$TC_r$  is the toll cost of the route, which is calculated by aggregating the link toll on each related link, and  $TT_r$  is the observed travel time of the route, which is calculated by approximately aggregating the observed link travel time for each related link.

In the logit model, we have three estimated variable vectors, as follows:

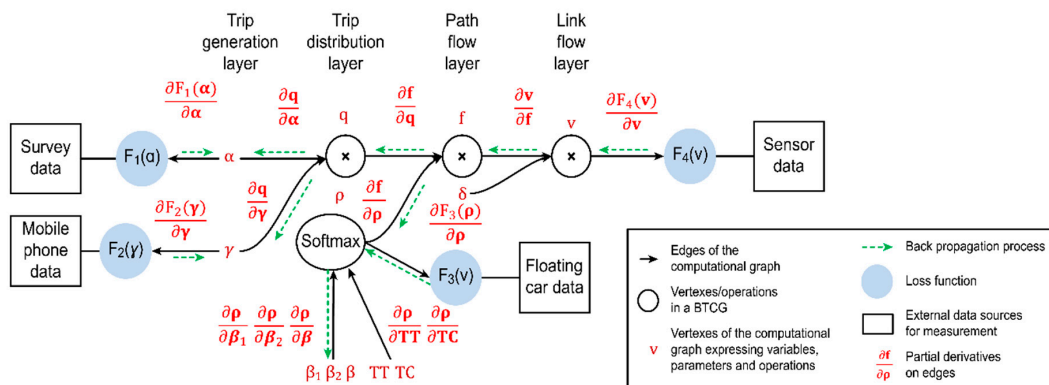
$\beta_1$  indicates the variable vector collecting all  $\beta_{w,1}$  between each OD pair,  $w$ .

$\beta_2$  indicates the variable vector collecting all  $\beta_{w,2}$  between each OD pair,  $w$ .

$\beta$  indicates the variable vector collecting all  $\beta_w$  between each OD pair,  $w$ .

#### 4.2. System Equations to Express the Computational Graph

The overall traffic demand estimation problem can be represented by the CG  $G(V_c, A_c)$  depicted in Figure 4. This CG can be generally viewed as a control system containing the following items.



**Figure 4.** Big data driven transportation computational graph (BTCG) to express a traffic demand estimation model.

## 1. State Variables

$V_c$  denotes the set of vertexes, including all state variables of the system, i.e., the estimated variable vectors and matrices listed in Section 4.1, and other parameters that can be expressed as scalars, vectors, matrices, or tensors.

## 2. State Transitions

$A_c$  denotes the set of directed edges that describe different operations in the traffic demand estimation model. If a variable,  $y$ , is computed by applying an operation to a variable,  $x$ , we draw a directed edge from the vertex representing  $x$  to the vertex representing  $y$  and annotate the vertex representing  $y$  with the name of the operation. Complicated functions are described by combining many elementary operations in a recursive fashion [69].

The graph in Figure 4 expresses the following four constraints to enforce flow conservation and capture user choice behaviors (layer-based state transitions):

$$\alpha \times \gamma = q, \quad (32)$$

$$q \times \rho = f, \quad (33)$$

$$f \times \delta = v, \quad (34)$$

$$\rho = \text{softmax}(\beta_1, \beta_2, \beta, \text{TC}, \text{TT}), \quad (35)$$

where the values of all flow variables are required to be nonnegative. The above constraints can also be viewed as a state transition model in a control system.

### 4.3. Mapping of Data Measurements on the Computational Graph

Traffic demand variable values are propagated forward through the CG in Figure 4 until they reach the output layers, which are connected to external data sources. The outputs of the CG can then be compared to reference measurements to calculate the loss errors. The CG simultaneously uses traffic demand estimation models (e.g., four-step processes, in the field of transportation modeling) and multiple measurements from different data sources to generate estimates of the system's variables (its states) that are better than the estimates obtainable when using only a single measurement type. As such, CGs can be viewed as a common data fusion method. Furthermore, in the era of big data, we can directly use multiple data samples and the stochastic gradient descent (SGD) algorithm to represent the noise in the system instead of assuming complex probability distributions.

The loss functions used in this paper are as follows:

#### 1. Measurements Associated with Trip Generation, $\bar{\alpha}$

We obtained a reference set of trip generation results for a zone using household survey data. The number of trips associated with this zone can be calculated by multiplying the population size by the rate of trips using private cars [28,75]. The population and trip rate can be calculated using the average numbers for different groups. Then, we have a reference set of trip generation results,  $\bar{\alpha}$ , and we can obtain the following loss function:

$$F_1(\alpha) = \frac{1}{2M_1} \|\alpha - \bar{\alpha}\|_2^2, \quad (36)$$

where  $M_1$  is the number of survey samples.

#### 2. Measurements Associated with the Origin-Destination Split Rates, $\bar{\gamma}$

Reference OD split rates can be generated from raw mobile phone data that match the given zoning system. Because it is technically difficult to identify the mode of transportation using mobile phone data, in this paper, we use mobile phone data to generate the reference OD split rates,  $\bar{\gamma}$ :

$$F_2(\gamma) = \frac{1}{2M_2} \|\gamma - \bar{\gamma}\|_2^2, \quad (37)$$

where  $M_2$  is the number of mobile phone samples. Notably, the locations of cellular base stations are generally not consistent with traffic zones. Hence, some matching assumptions must be made to derive OD splits from cellular records. Furthermore, a rule should also be specified for identifying stationary activities [14,41].

### 3. Measurements Associated with the Route Choice Proportions, $\bar{\rho}$

Floating car data can be matched to the driving network using matching algorithms [76,77]. Thus, we can obtain reference route choice proportions using floating car data:

$$F_3(\rho) = \frac{1}{2M_3} \|\rho - \bar{\rho}\|_2^2, \quad (38)$$

where  $M_3$  is the number of floating car samples. It should be noted that floating car data are sampled data. We should again specify a rule for identifying stationary activities and then match the GPS records to a map using a map matching algorithm (e.g., a hidden Markov chain algorithm) [76,77].

### 4. Measurements Associated with the Link Flows, $\bar{v}$

Sensor flow counts,  $\bar{v}$ , are collected from a subset of links:

$$F_4(v) = \frac{1}{2M_4} \|v - \bar{v}\|_2^2, \quad (39)$$

where  $M_4$  is the number of sensor data samples. Loop detector data are usually collected at regular time intervals (e.g., every 15 min). In this paper, we sum the counts for all lanes to generate the flow count on each link for each hour.

Overall, the CG can be expressed in terms of the following optimization problem:

$$\min Z = F_1(\alpha) + F_2(\gamma) + F_3(\rho) + F_4(v), \quad (40)$$

s.t. Equations (32)–(35).

Here, we simply assume that all four objective functions are equally important.

#### 4.4. Mapping of Congestion Strategies on the Layers of the Computational Graph

A CG can serve as a helpful tool to support the evaluation of a congestion mitigation strategy. The marginal effects of a policy can be calculated from the partial derivatives of a variable/parameter with respect to all other variables/parameters. That is, the gradients can be viewed as indicators of whether an imposed policy is reasonable. From the perspective of control theories, congestion mitigation strategies can be viewed as controllable inputs acting on traffic demand variables. Figure 5 shows how different strategies map to different variables in the CG. For example, pricing on paths/links can be viewed as controlling the route choice proportions,  $\rho$ , or the link flows,  $v$ . When planners aim to redesign urban layouts or relocate workplaces, the corresponding policies can similarly be viewed as controlling the variables,  $\gamma$  and  $q$ . Thus, we can define the **marginal effect** (ME) of such control as follows:

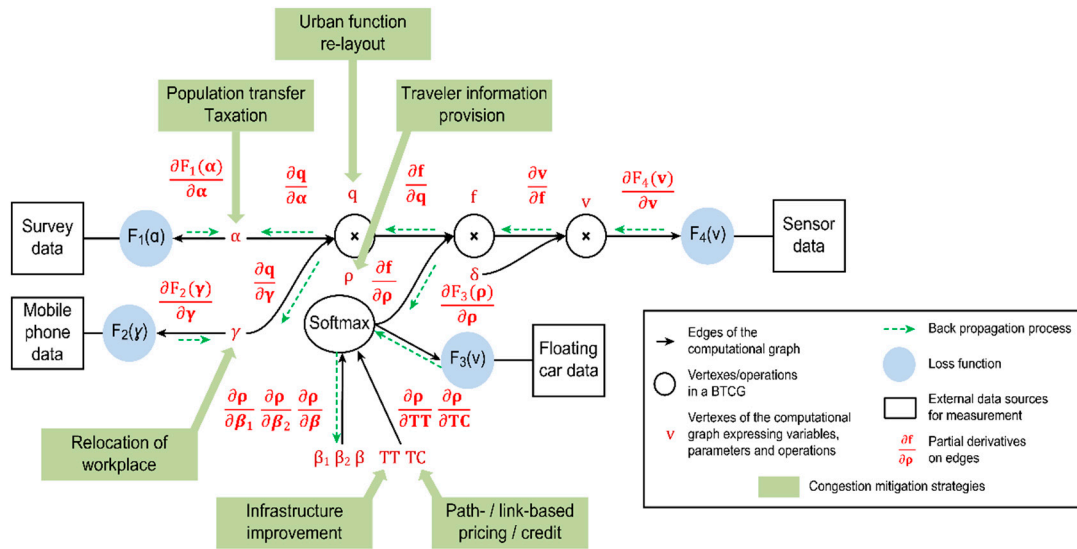


Figure 5. Different congestion mitigation strategies used to control different variables in the CG.

**Definition:** The marginal effect (ME) of a congestion mitigation strategy is defined as “the change in the overall negative utility of the system caused by a marginal change in the controlled variable”. In this paper, the negative utility is measured as the total travel time for all users in the transportation network. Let the vector field,  $F(\mathbf{v})$ , be a link travel time function mapping from  $\mathbb{R}^{|\mathcal{V}|}$  to  $\mathbb{R}^{|\mathcal{V}|}$ . Then, the ME can be calculated as:

$$ME = F(\mathbf{v} + \Delta) \cdot (\mathbf{v} + \Delta)^T - F(\mathbf{v}) \cdot (\mathbf{v})^T, \tag{41}$$

where  $\Delta$  represents the marginal changes in the link flows caused by the control (i.e., the policy). The physical meaning of the ME is the total reduction in “vehicle transport” on all links caused by the policy.

1. If  $ME \leq 0$ , the policy decreases the total travel time for users and has a positive effect.
2. If  $ME > 0$ , the policy increases the total travel time for users and has a negative effect.

One advantage of a CG is that it can directly yield the reciprocal of  $\Delta$ . Table 2 lists the congestion mitigation strategies considered and the variables they control in the CG. Table 2 also shows how to calculate  $1/\Delta$  based on the chain rule in calculus. Each  $1/\Delta$  in the table corresponds to a certain “pathway” propagated through the CG, as shown in Figures 4 and 5. The initial values of the variables before learning can be set with reference to typical historical data.

Table 2. Mapping of congestion mitigation strategies in the multilayer CG framework.

Policy	Layer in the CG	Purpose	Variable	$1/\Delta$
Population transfer/ taxation	Trip generation layer	Reduce the number of users in a zone Reduce users’ trip rates	$\alpha$	$\frac{\partial v}{\partial \alpha} = \frac{\partial v}{\partial f} \frac{\partial f}{\partial q} \frac{\partial q}{\partial \alpha}$
Urban functional re-layout	Trip distribution layer	Jobs-housing balance Reduce users’ travel distances	$q$	$\frac{\partial v}{\partial q} = \frac{\partial v}{\partial f} \frac{\partial f}{\partial q}$
Relocation of workplaces	Trip distribution layer	Jobs-housing balance Reduce users’ travel distances	$\gamma$	$\frac{\partial v}{\partial \gamma} = \frac{\partial v}{\partial f} \frac{\partial f}{\partial q} \frac{\partial q}{\partial \gamma}$
Traveler information provision	Path flow layer	Change users’ route choice behaviors	$\rho$	$\frac{\partial v}{\partial \rho} = \frac{\partial v}{\partial f} \frac{\partial f}{\partial \rho}$
Link-/path-based pricing/credit	Path/link flow layer	Change users’ route choice behaviors	TC	$\frac{\partial v}{\partial TC} = \frac{\partial v}{\partial f} \frac{\partial f}{\partial \rho} \frac{\partial \rho}{\partial TC}$
Infrastructure improvement	Path/link flow layer	Change users’ route choice behaviors	TT	$\frac{\partial v}{\partial TT} = \frac{\partial v}{\partial f} \frac{\partial f}{\partial \rho} \frac{\partial \rho}{\partial TT}$

#### 4.5. Algorithm

As shown in Figures 4 and 5, we can obtain a first-order partial derivative on each edge of the CG. Based on these derivatives, the loss errors (gradients) from the measurements are backpropagated

to each vertex starting from the outputs. The gradients are calculated using the chain rule. The updating process is widely used in the field of deep learning in the context of the backpropagation (BP) algorithm [69]. In this paper, we implement the algorithm as follows:

**Step 1.** Estimation:

**Step 1.1.** The forward passing step implements trip generation, trip distribution estimation, and traffic assignment.

**Step 1.2.** The backward propagation step updates the estimated variables using the SGD algorithm.

**Step 1.3.** Steps 1.1. and 1.2. are iteratively implemented until convergence is reached.

**Step 2.** Evaluation:

**Step 2.1.** Different mitigation strategies are evaluated by calculating their MEs on link volume.

Because our proposed learning framework can be regarded as a kind of control system, in Table 3, we compare the above algorithm with the standard Kalman filter (KF) (i.e., a linear Gaussian state-space model) [5,78–80]. The similarity between them is that they both provide a general framework for state estimation and data fusion. Both algorithms are recursive and follow a two-step process. The prediction step of the KF, corresponding to the forward passing process in Step 1.1 of our CG-based algorithm, produces the current estimates. The update process of the KF, corresponding to the backward propagation in Step 1.2 of our CG-based algorithm, updates the estimates based on the observed measurements. The difference between them is that they use different methods of updating their estimates. While the KF uses the Kalman optimal gain to update the state variables, the above algorithm implements a gradient-based update strategy. Furthermore, the KF applies a linear transition matrix to generate the next-stage estimates and uses covariance matrices to express the correlations between state variables. By contrast, our framework directly uses a CG to describe these relationships. Similar to other deep learning methods, our algorithm uses the SGD to capture the noise in the data, while the KF usually requires an assumed probability distribution.

It is important to note that the above algorithm can be easily implemented using the TensorFlow data programming architecture, which has been widely applied in many machine learning applications. To see the source code of our CG-based learning framework using the TensorFlow Python API, readers can refer to [81].

**Table 3.** Comparison of our CG-based learning framework and the Kalman Filter (KF).

Model	Computational Graph (CG)	Kalman Filter (KF)
State variables	Trip generation Trip distribution Path/link flows	Traffic state variables OD volume
Traffic observations	Multiple data sources	Time-varying observations
Algorithm process	Recursive forward/backward propagation	Recursive prediction and updating
Update method	Gradient	Kalman optimal gain
Noise	Stochastic gradient descent	Gaussian distribution
Control inputs	Policies imposed offline	External influences imposed online
Correlation between variables	Composite function	Covariance matrix
State transitions	Layer-based nonlinear transitions	Stage-based linear transitions

## 5. Numerical Examples

In this section, we present two numerical examples to illustrate the effectiveness of our methodology.

### 5.1. A Case Study Based on the Sioux Falls Network

We implemented our model based on multiple samples in the Sioux Falls network to validate the effectiveness of our framework. The input data (measurements) included two samples of hypothetical household survey data (trip generation results,  $\bar{\alpha}$ ), two samples of mobile phone data (OD split rates,  $\bar{\gamma}$ ), two samples of floating car data (route choice proportions,  $\bar{\rho}$ , for some routes), and three samples

of sensor data (link flow counts,  $\bar{v}$ ). We assumed that the data were collected during one hour of the morning peak time on several working days.

Notably, the accuracy of calibration will depend on the synthesized hypothetical data. To avoid discrepancies between different data sources, we designed one “seed” group of hypothetical data that satisfied the flow conservation constraints of Equations (36)–(39) in an ad hoc manner. The data set included the following:

1. One sample of survey data: Reference trip generation results for 5 zones (i.e., zones 1, 2, 7, 13, and 20).
2. One sample of mobile phone data: Reference OD split rates for 20 OD pairs (i.e., origin zones 1, 2, 7, 13, and 20 and destination zones 9, 11, 22, and 24).
3. One sample of floating car data: We enumerated all candidate paths between the 20 OD pairs, then randomly selected 7 of these paths and adopted assumed route choice proportions for them.
4. One sample of sensor data: We assigned assumed link counts to 7 links.

Then, we added some random perturbations to the “seed” samples to generate additional samples. The complete data set can be found in [81]. Figure 6 shows only the average values of our data set. All flows are expressed in units of vehicles/hour.

#### 5.1.1. Calibration Using Multiple Data Sources

Figure 7 shows the trend lines between the average estimated values and the average reference measurements. We find that reasonably satisfactory R-square values are achieved for all data sources. However, this finding does not imply that our proposed learning framework can always achieve good loss errors. When the inconsistency between different data sets increases, the obtained loss errors will worsen. However, this case study demonstrates that our proposed learning framework can simultaneously decrease the values of all four loss functions between the estimated values and the reference measurements. Figure 7 displays the convergence curves for the four different data sources. Because of the use of the SGD algorithm, the curves fluctuate, reflecting the random noise in the data. Figure 6 also displays four links (links (1, 3), (3, 12), (12, 13), and (13, 24)) with estimated volumes of >390 vehicles/hour (i.e., link flow/link capacity > 1.3). We find that sensors are not installed on any of these four links.

#### 5.1.2. Analysis of Congestion Components

Because the CG-based learning framework internally integrates the four-step process of transportation modeling and enables the estimation of the flows on all paths in the candidate path set, traffic managers can apply the results to analyze the individual components of the congestion on each link. Figure 8 displays the congestion pie charts for links (3, 12) and (13, 24).



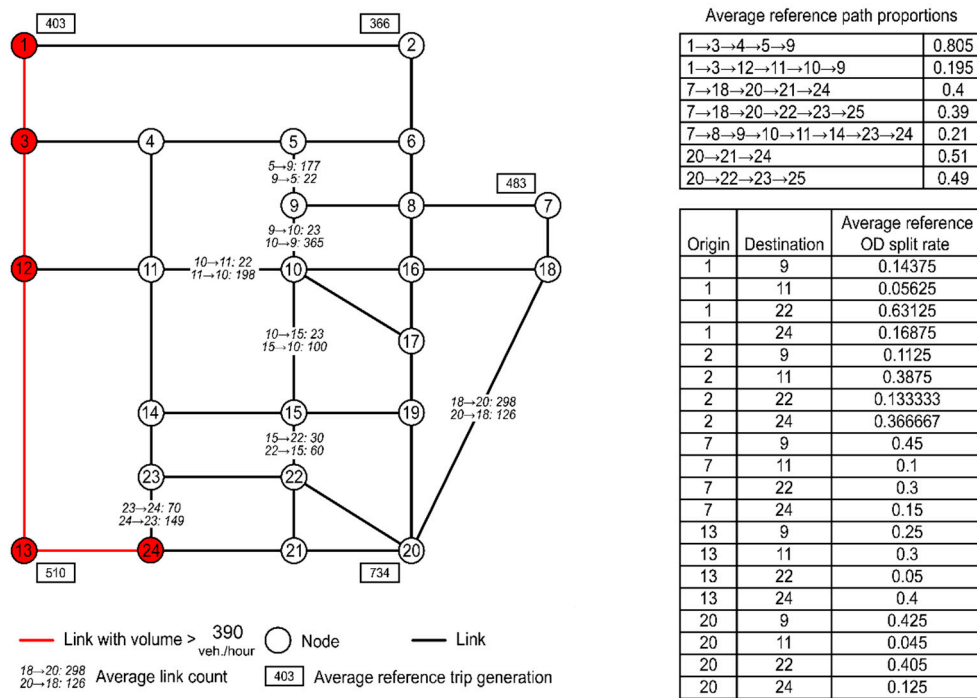


Figure 6. Basic information on the simplified Sioux Falls network (units of link flows: vehicles/hour).

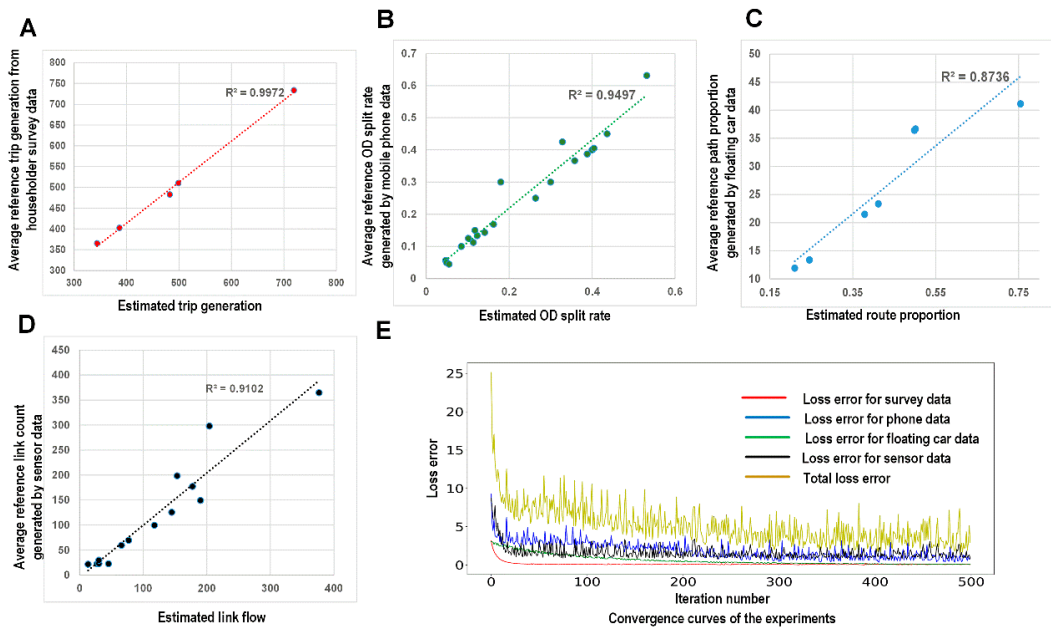
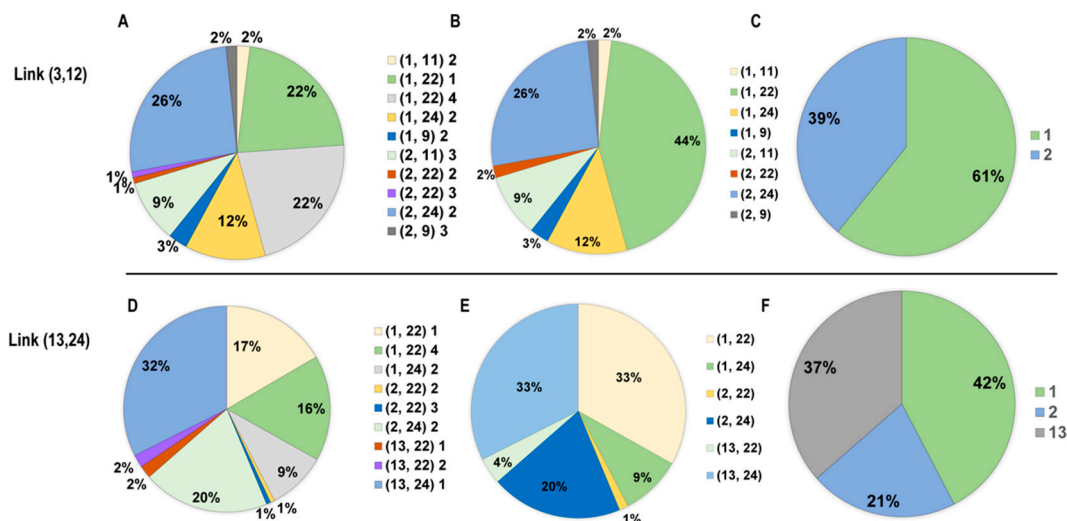


Figure 7. Linear regression lines between the corresponding estimated values and the average reference measurements from the various data sources (A) Household data, (B) Mobile phone data, (C) Floating car data, and (D) Sensor data; (E) The convergence curves of the loss functions for all data sources (units of link flows: vehicles/hour).



**Figure 8.** Conceptual illustration of different layers of congestion mitigation; (A) The path flows that contribute to the volume on link (3,12); (B) The OD volumes that contribute to the volume on link (3,12); (C) The trip generations that contribute to the volume on link (3,12); (D) The path flows that contribute to the volume on link (13,24); (E) The OD volumes that contribute to the volume on link (13,24); (F) The trip generations that contribute to the volume on link (13,24).

Table 4 further shows how the volume on link (3, 12) is composed of 10 path flows. As shown in Figure 8A and Table 4, the following three paths (bolded in Table 4) contribute the majority (a total of 70%) of the traffic volume on link (3, 12):

- 2→1→3→12→13→24.
- 1→3→12→13→24→21→22.
- 1→3→12→13→24→23→22.

**Table 4.** Analysis of congestion components for link (3, 12) (units: vehicles/hour).

From Zone	To Zone	Path Index	Node Sequence	Contributed Path Flow	Contributed OD Volume	Contributed Zone Production
1	9	2	1→3→12→11→10→9	13.4	13.4	
1	11	2	1→3→12→11	9.1	9.1	
1	22	1	<b>1→3→12→13→24→21→22</b>	<b>102</b>		282.9
1	22	4	<b>1→3→12→13→24→23→22</b>	<b>101.8</b>	203.7	
1	24	2	1→3→12→13→24	56.7	56.7	
2	9	3	2→1→3→12→11→10→9	7.8	7.8	
2	11	3	2→1→3→12→11	44.4	44.4	
2	22	2	2→1→3→12→13→24→23→22	4.2		182.5
2	22	3	2→1→3→12→13→24→21→22	4.2	8.4	
2	24	2	<b>2→1→3→12→13→24</b>	<b>122</b>	122	
Link volume					465.4	

Furthermore, by simply combining the corresponding path flows, it can be seen from Figure 8B and Table 4 that link (3, 12) is mainly passed by users of OD pairs (2, 24), (1, 22), and (1, 24) (26%, 44%, and 12%, respectively). Figure 8C and Table 4 show that 61% of the flow on link (3, 12) is generated from node 1. We can analyze the congestion components for link (13, 24) in a similar manner; see Table 5 and Figure 8D–F. The pie charts show that 37% of the flow on link (13, 24) is generated from node 13 and that most of the corresponding paths terminate at node 24. The detailed congestion components are presented in Table 4.

**Table 5.** Analysis of congestion components for link (13, 24) (units: vehicles/hour).

From Zone	To Zone	Path Index	Node Sequence	Contributed Path Flow	Contributed OD Volume	Contributed Zone Production
1	22	1	1→3→12→13→24→21→22	102		
1	22	4	1→3→12→13→24→23→22	101.8	203.7	
1	24	2	1→3→12→13→24	56.7	56.7	260.4
2	22	2	2→1→3→12→13→24→23→22	4.2	8.4	
2	22	3	2→1→3→12→13→24→21→22	4.2		130.3
2	24	2	2→1→3→12→13→24	122	122	
13	22	1	13→24→21→22	12.3	24.5	
13	22	2	13→24→23→22	12.3		224.1
13	24	3	13→214	199.6	198.6	
Link volume					612.1	

This congestion component analysis provides us with useful information about the sources of congestion. Knowledge of the congestion pie charts is very important for selecting the zones, OD pairs, or paths where congestion mitigation strategies will have the greatest effect. For example, if we wish to decrease the flows on link (3, 12), relocating some workplaces at node 22 to other zones might be an effective method.

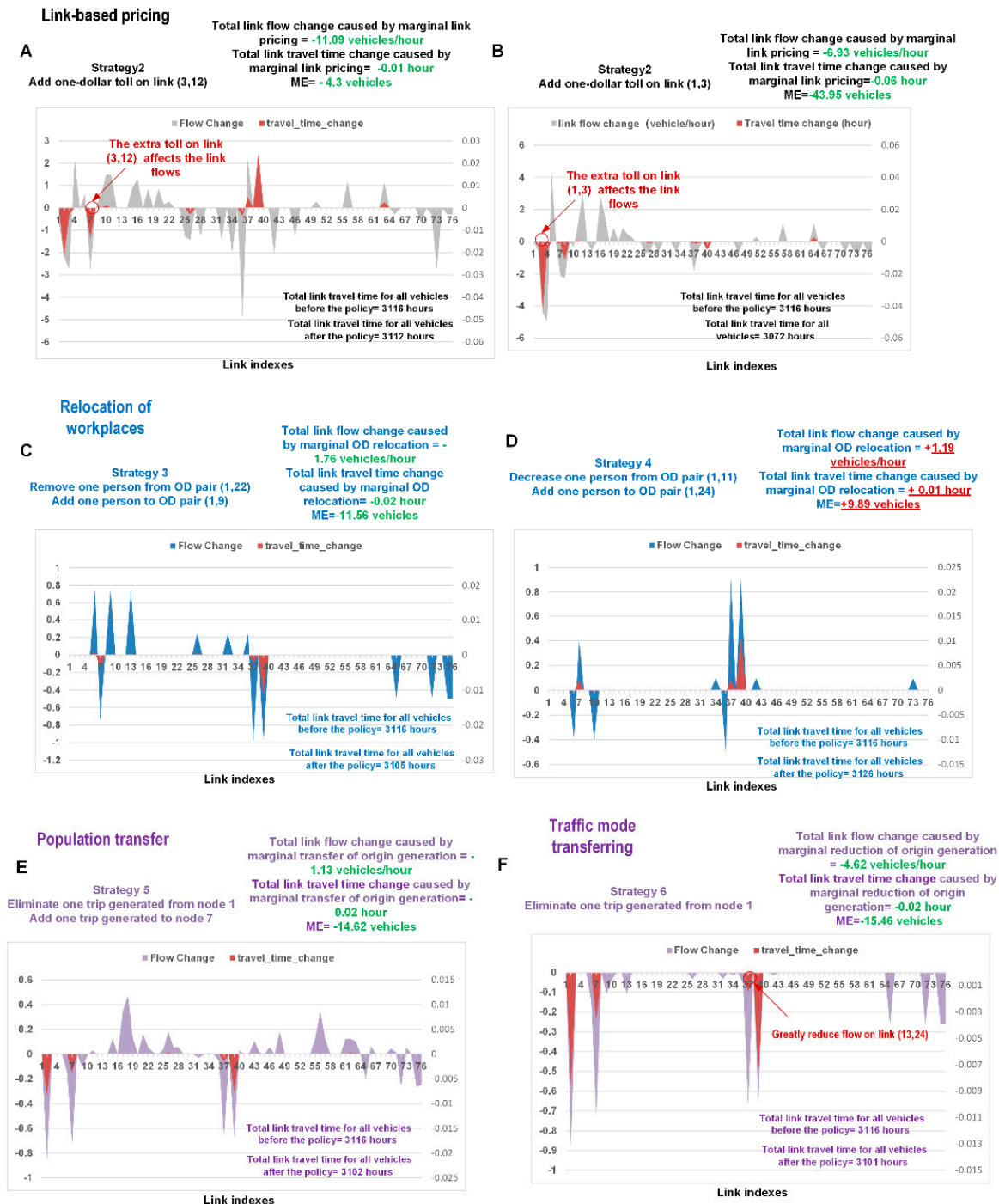
### 5.1.3. Marginal Effect Analysis

In this experiment, we aimed to evaluate the MEs of various congestion mitigation strategies. We used the Bureau of Public Roads (BPR)-form function as the travel time function,  $F(\mathbf{v})$ , in Equation (36). The current estimated link flows and the parameters of the BPR function are reported in Appendix A. Figure 9 shows the changes in the flows and travel times on each link. The MEs of the six strategies are displayed in Figure 9.

Several interesting observations can be found:

1. In Figure 9A,B, the toll successfully decreases the flows and travel times on links (3, 12) and (1, 3). It also reduces the total travel time on all links. Interestingly, although both links (3, 12) and (1, 3) are congested (465 vehicles/hour and 583 vehicles/hour, respectively), imposing a toll of one extra dollar on link (1, 3) produces more benefit than doing the same on link (3, 12). The former policy reduces the number of vehicles in the system by 44 (ME = 44 vehicles), while the latter results in a decrease of only 4 vehicles (ME = 4 vehicles). These findings demonstrate that similar pricing policies can have different effects.
2. As shown in Figure 9C, if one user changes his/her destination from node 24 to node 9, then the traffic flows on links (3, 12), (12, 13), and (13, 24) will decrease (by 0.7 vehicles/hour, 0.99 vehicles/hour, and 0.99 vehicle/hour, respectively). This figure justifies the importance of job-housing balancing in urban planning.
3. We also find that the impacts of the policies on the flows and travel times are complex, with some mitigation strategies potentially decreasing the overall welfare of the system. For the scenario depicted in Figure 9D, the ME corresponds to an increase of 9.89 vehicles in the system. The policy actually decreases the flow levels (by approximately 0.4 vehicles/hour) on links (3, 4), (4, 5), and (12, 11) (links 6, 9, and 36, respectively, in the plot); however, these three links are not approaching their capacities in their current states (257.3 vehicles/hour, 116 vehicles/hour, and 216 vehicles/hour, respectively; see the Appendix A). Unfortunately, the strategy also guides additional traffic flows (approximately 0.9 vehicles/hour) to links (12, 13) and (13, 24), which are already congested (391 vehicles/hour and 615 vehicles/hour, respectively). This is the reason why, in the short term, sometimes the functional relocation of a metropolitan area can sometimes lead to a worse result than before.
4. Figure 9E shows that the strategy of “population transfer” achieves good performance in relieving traffic congestion. As seen in Figure 9F, if methods are implemented to make fewer people from

zone 1 use private cars, this strategy will also apparently increase the overall utility of the system. In particular, these policies greatly reduce the flow on the congested link (13, 24). The reason for the beneficial effects of these policies can be identified from the congestion component pie chart shown in Figure 9: In total, 42% of the flow on link (13, 24) is generated from node 1.



**Figure 9.** Marginal effect (ME) analyses of different congestion mitigation strategies; (A) ME analyses of adding one-dollar toll on link (3, 12); (B) ME analyses of adding one-dollar toll on link (1, 3); (C) ME analyses of removing one person from OD pair (1,22) and adding one person to OD pair (1,9); (D) ME analyses of removing one person from OD pair (1,11) and adding one person to OD pair (1,24); (E) ME analyses of removing one person from zone 1 and adding one person to zone 7; (F) ME analyses of removing one person from zone 1.

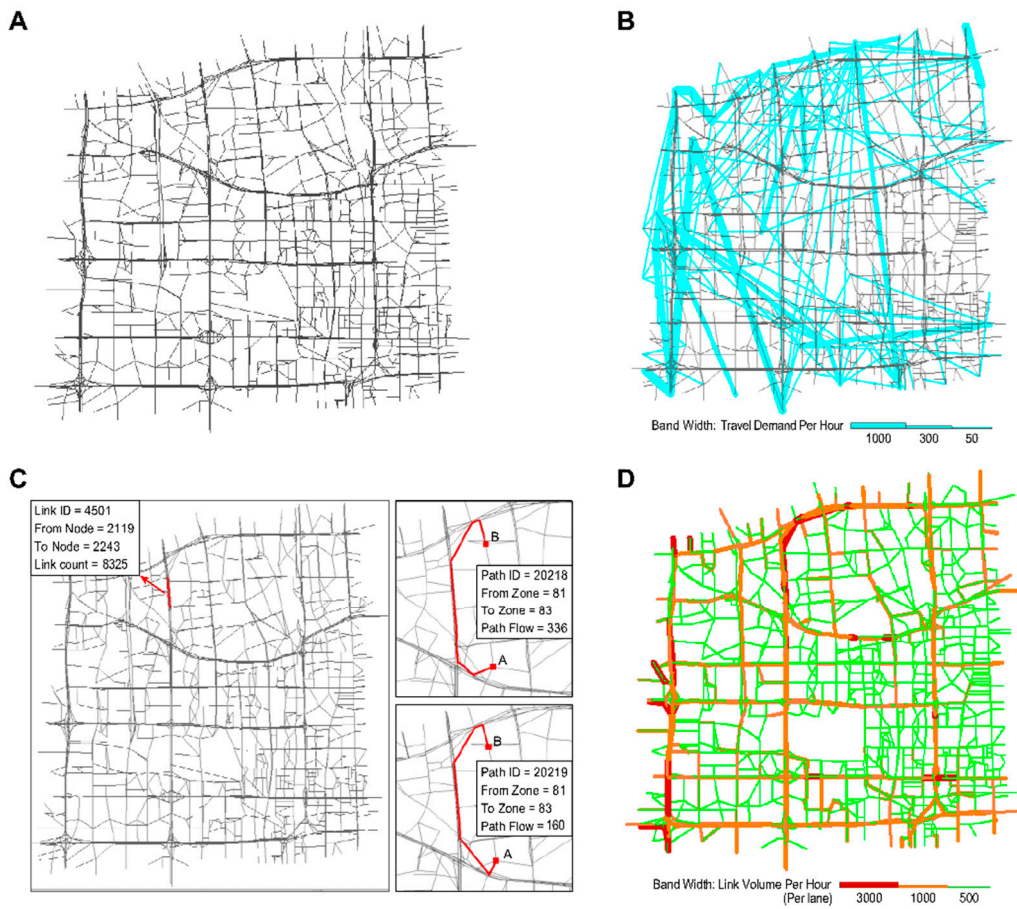
## 5.2. Application Study in a Beijing Subnetwork

To better demonstrate the applicability of the proposed CG-based learning framework, a median-scale experiment was conducted based on a subnetwork of Beijing with 2502 nodes, 236 zones, 14,967 OD pairs, 40,494 paths, and 5397 links. The traffic demand outside the subnetwork was merged in the traffic zones near the boundaries. We did not use mobile phone data or floating car data in this experiment. We utilized only the reference measurements for the trip generation results and OD split rates provided by the metropolitan planning organization. Loop detector data (from one hour on 113 links) from this organization were also applied. Hence, the objective function,  $\min F_1(\alpha) + F_2(\gamma) + F_4(v)$ , was used to integrate the data sources into the learning process.

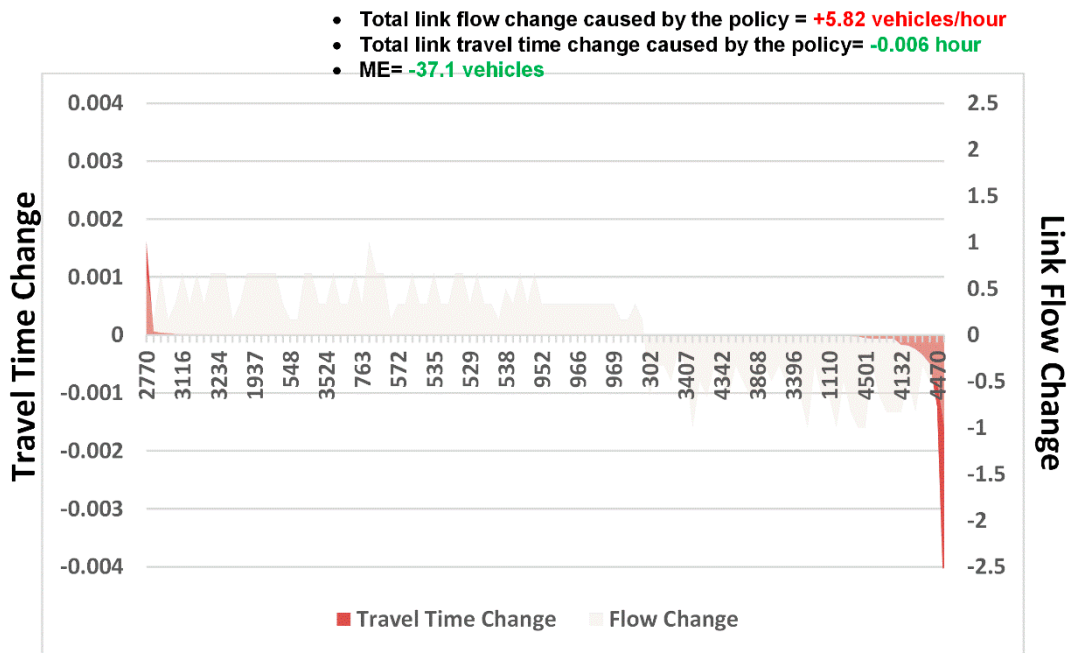
The maximum number of iterations was set to 1000, and the initial learning rate was set to 0.00001. The procedure was run under Linux on a Dell PowerEdge T630 tower server with two Intel Xeon Quad CPUs, eight 16 GB of RAM, and 512 GB of SSD storage. The number of variables in the CG can be estimated as follows: Number of neurons in each layer + number of connections between layers  $((236 + 14,967 + 40,494 + 5,397) + (14,967 + 40,494 + 40,494 \times 5397) = 116,555)$ . The total CPU time required for the learning process was 2 hours.

Several interesting results were obtained after the experiment:

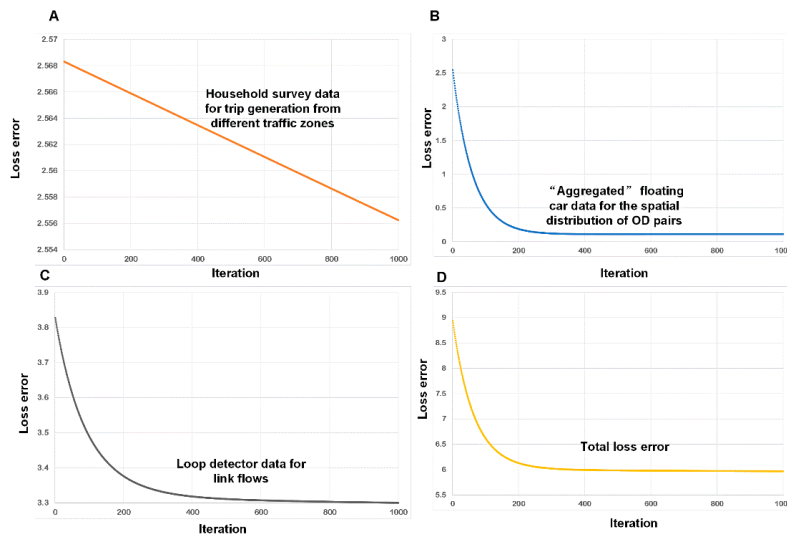
1. Estimated outputs: Figure 10A shows the physical network. The other panels in this figure illustrate the outputs estimated using different layers of the CG. Figure 10B displays the estimated trip distribution. As displayed in Figure 10D, the estimated link flows per hour per lane were also obtained using the proposed learning framework.
2. Congestion analysis: Figure 10C shows one of the most congested links (i.e., link 4501, from node 2119 to node 2243). The estimated link flow is 8325 vehicles/hour. Based on the CG, there are a total of 2157 paths passing through link 4501. Figure 10C displays the two paths that most strongly contribute to congestion on this link (paths 20218 and 20219). Furthermore, the flow on the link comes from 557 OD pairs and 102 traffic zones. Figure 13A,B display the top 30 OD pairs and trip-generating traffic zones that contribute the most to the flow volume on link 4501. We also labelled the OD pairs and traffic zones associated with the largest volume on the map. We find that the congestion on link 4501 is primarily caused by traffic demands for travel from the southern urban area to the northern subarea.
3. ME analysis: Because link 4501 and zone 83 are near several universities and companies in Beijing, one possible congestion mitigation strategy is to move some workplaces from zone 83 to zone 117. We can calculate the ME of this policy as follows. Figure 11 shows the changes in travel times and volumes on related links. Interestingly, this strategy actually increases the total volume on the links. However, it can reduce the travel times on certain highly congested links. The ME is  $-37.1$  vehicles, which implies that the policy can indeed mitigate congestion in the traffic system.
4. Calibration: The estimation processes using different data sources are shown in Figure 12. During the experiment, we normalized the estimates and the references to lie within the range of  $[0, 10]$ . We found that the different objective functions can be simultaneously estimated during the learning process.



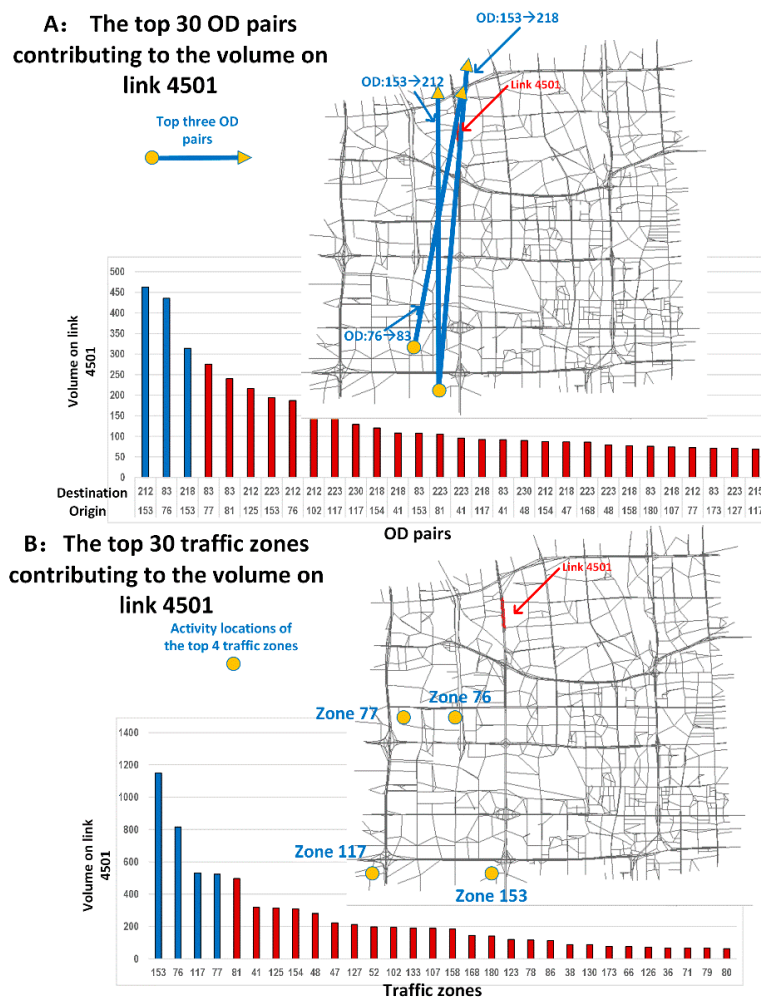
**Figure 10.** Experimental results based on the Beijing subnetwork; (A) physical network; (B) Estimated trip distribution; (C) The top 2 path flows contributing to the volume on link 4501; (D) Estimated link flows.



**Figure 11.** ME analysis of moving a workplace in zone 83 to zone 117.



**Figure 12.** Evolution of the error values of the loss functions over 1000 iterations; (A) the loss error of household survey data; (B) the loss error of floating car data; (C) the loss error of link counts; (D) the total loss error.



**Figure 13.** (A)The top 30 OD pairs in the subnetwork with the highest estimated contributions to the flow volume on link 4501; (B) The top 30 traffic zones in the subnetwork with the highest estimated contributions to the flow volume on link 4501.

## 6. Conclusions and Future Research Plans

In a traffic system, congestion and inefficiency are fundamentally related to the relative balance between demand and supply. A traffic system can operate efficiently if the traffic demand can be accurately estimated. This paper proposed an integrated framework that combines traditional model-driven traffic demand estimation with emerging data-driven approaches based on a CG framework. This CG-based learning framework might contribute to filling the two research gaps mentioned in Section 1. First, the framework can help to overcome challenges related to data mining and fusion when processing big data from multiple heterogeneous sources. Second, congestion mitigation strategies can be integrated with the CG to evaluate the benefits and costs of the corresponding policies. A real-world case study was also presented to demonstrate the applicability of the framework.

However, the proposed framework also has some limitations that should be addressed in the future. First, we anticipate that the proposed CG-based framework can be developed to describe more realistic traffic demands. We can extend the learning framework based on the fundamental unit method [28], that is:

$$\alpha = g \cdot t(p_1, p_2, \dots, p_n) \quad (42)$$

where  $g$  is a vector expressing the populations of different groups and the trip rate vector,  $t(p_1, p_2, \dots, p_n)$ , expresses the relationships between trip rates and various influencing factors  $((p_1, p_2, \dots, p_n))$ . The learning framework can be used to calibrate the related parameters. For example, suppose that the trip rate of a certain population group is impacted by the price of gas,  $p_1$ . Managers can impose gas taxes to reduce the trip rate of this group of people. When trip chain data are available, we can also apply a trip chain layer in place of the OD layer shown in 4 and reconstruct the CG based on activity-based models [80–84]. Second, the current framework has not fully considered several important traffic flow characteristics. We simply used the BPR function to describe the basic relationship between travel times and link flows, which is not accurate according to the fundamental diagrams in traffic flow theories. Finally, the only traffic mode considered in this research was driving in private cars. We also need to extend the CG by incorporating a mode-split layer or extending the path layer based on a supernetwork to capture both automobile and public transit networks [18]. This modification would allow managers to additionally impose strategies for adjusting the choices between different traffic modes. The final version of this demand estimation engine should serve as a decision-making support platform that can simulate transportation scenarios that cannot be practically realized on the basis of the closely interconnected relationship between traffic planners and travelers. It should have the capability of predicting and visualizing the evolution of multimodal traffic systems (including private cars, buses, and metro systems) and provide useful decision support advice for managers.

**Author Contributions:** Formal Analysis, J.S.; Validation, J.S.; Investigation, J.S.; Resources, J.S.; Conceptualization, X.Z.; Methodology, X.W., X.Z.; Data Curation, J.S., K.X.; Writing—Original Draft Preparation, X.W., D.W., Q.Z.; Writing—Review and Editing, J.S., X.W., X.Z.; Visualization, X.W., D.W.; Supervision, J.G.; Project Administration, J.G.; Funding Acquisition, J.G.

**Funding:** This research project, especially the acquisition of the large-scale Beijing test network and the various traffic data sets, has been supported through the Beijing International Cooperation Base for Science and Technology on Urban Transport and the Beijing Key Laboratory of Urban Traffic Operation Simulation and Decision Support (BZ0012). This work was also partially supported by the China Natural Science Funding under Grant 61731004 and was supported by the National Natural Science Foundation of China under project no. 71734004, titled “Research on advanced theories for urban transportation governance”. The authors from Arizona State University are partially funded by the National Science Foundation of the United States under NSF Grant No. CMMI 1538105 “Collaborative Research: Improving Spatial Observability of Dynamic Traffic Systems through Active Mobile Sensor Networks and Crowdsourced Data” and NSF Grant No. CMMI 1663657 “Real-time Management of Large Fleets of Self-Driving Vehicles Using Virtual Cyber Tracks”.

**Conflicts of Interest:** The authors declare that there are no conflict of interest.



## Appendix A

The current estimated link flows and parameters of the Bureau of Public Roads (BPR) function are shown in Table A1. The BPR function used to evaluate the MEs is as follows:

$$F(v) = \frac{\text{Link length}}{\text{Free flow speed}} \left( 1 + 1.5 \left( \frac{\text{Link flow}}{\text{Capacity}} \right)^4 \right) \quad (\text{A1})$$

**Table A1.** Current estimated link flows and parameters of the Bureau of Public Roads (BPR) function.

Link ID	Link Name	Estimated Flow (vehicles/hour)	Length (mile)	Free Flow Speed (miles/hour)	Capacity (vehicles/hour)
1	(1, 2)	0.94	6	60	300
2	(1, 3)	582.59	4	60	300
3	(2, 1)	242.51	6	60	300
4	(2, 6)	96.49	5	60	300
5	(3, 1)	0	4	60	300
6	(3, 4)	257.3	4	60	300
7	(3, 12)	465.36	4	60	300
8	(4, 3)	0	4	60	300
9	(4, 5)	115.66	2	60	300
10	(4, 11)	213.08	6	60	300
11	(5, 4)	71.45	2	60	300
12	(5, 6)	0	4	60	300
13	(5, 9)	177.53	5	60	300
14	(6, 2)	0	5	60	300
15	(6, 5)	119.96	4	60	300
16	(6, 8)	42.75	2	60	300
17	(7, 8)	244.9	3	60	300
18	(7, 18)	245.92	2	60	300
19	(8, 6)	66.22	2	60	300
20	(8, 7)	17	3	60	300
21	(8, 9)	158.5	10	60	300
22	(8, 16)	45.93	5	60	300
23	(9, 5)	13.36	5	60	300
24	(9, 8)	0	10	60	300
25	(9, 10)	27.45	3	60	300
26	(10, 9)	376.53	3	60	300
27	(10, 11)	45.45	5	60	300
28	(10, 15)	30.2	6	60	300
29	(10, 16)	0	5	60	300
30	(10, 17)	0	8	60	300
31	(11, 4)	0	6	60	300
32	(11, 10)	153.29	5	60	300
33	(11, 12)	0	6	60	300
34	(11, 14)	17.99	4	60	300
35	(12, 3)	140.07	4	60	300
36	(12, 11)	215.78	6	60	300
37	(12, 13)	390.75	3	60	300
38	(13, 12)	281.24	3	60	300
39	(13, 24)	614.86	4	60	300
40	(14, 11)	78.82	4	60	300
41	(14, 15)	0	5	60	300
42	(14, 23)	17.99	4	60	300
43	(15, 10)	117.69	6	60	300
44	(15, 14)	6.83	5	60	300
45	(15, 19)	0	4	60	300
46	(15, 22)	30.2	4	60	300
47	(16, 8)	0	5	60	300
48	(16, 10)	153.75	5	60	300
49	(16, 17)	0	2	60	300
50	(16, 18)	17.86	3	60	300
51	(17, 10)	0	8	60	300
52	(17, 16)	0	2	60	300
53	(17, 19)	0	2	60	300
54	(18, 7)	78.78	2	60	300
55	(18, 16)	125.67	3	60	300
56	(18, 20)	204.05	4	60	300
57	(19, 15)	58.78	4	60	300
58	(19, 17)	0	2	60	300
59	(19, 20)	0	4	60	300
60	(20, 18)	144.72	4	60	300
61	(20, 19)	58.78	4	60	300
62	(20, 21)	306.22	6	60	300
63	(20, 22)	333.84	5	60	300
64	(21, 20)	0	6	60	300
65	(21, 22)	292.64	2	60	300
66	(21, 24)	132.01	3	60	300
67	(22, 15)	65.74	4	60	300
68	(22, 20)	0	5	60	300
69	(22, 21)	0	2	60	300
70	(22, 23)	59.1	4	60	300
71	(23, 14)	71.99	4	60	300
72	(23, 22)	118.18	4	60	300
73	(23, 24)	77.09	2	60	300
74	(24, 13)	0	4	60	300
75	(24, 21)	118.42	3	60	300
76	(24, 23)	190.17	2	60	300

## References

1. Lomax, T.J.; Schrank, D.L. *The 2002 Urban Mobility Report*; Texas Transportation Institute, Texas A&M University: College Station, TX, USA, 2002.
2. Toole, J.L.; Colak, S.; Sturt, B.; Alexander, L.P.; Evsukoff, A.; González, M.C. The path most traveled: travel demand estimation using big data resources. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 162–177. [[CrossRef](#)]
3. Zhou, X.; Qin, X.; Mahmassani, H.S. Dynamic origin-destination demand estimation with multiday link traffic counts for planning applications. *Transp. Res. Record* **2003**, *1831*, 30–38. [[CrossRef](#)]
4. Zhou, X.; Mahmassani, H.S. Dynamic origin-destination demand estimation using automatic vehicle identification data. *IEEE Trans. Intell. Transp. Syst.* **2006**, *7*, 105–114. [[CrossRef](#)]
5. Zhou, X.; Mahmassani, H.S. A structural state space model for real-time traffic origin–destination demand estimation and prediction in a day-to-day learning framework. *Transp. Res. Part B Methodol.* **2007**, *41*, 823–840. [[CrossRef](#)]
6. Lu, C.-C.; Zhou, X.; Zhang, K. Dynamic origin–destination demand flow estimation under congested traffic conditions. *Transp. Res. Part C Emerg. Technol.* **2013**, *34*, 16–37. [[CrossRef](#)]
7. Asakura, Y.; Hato, E. Tracking survey for individual travel behaviour using mobile communication instruments. *Transp. Res. Part C Emerg. Technol.* **2004**, *12*, 273–291. [[CrossRef](#)]
8. Zhao, Y.; Kockelman, K.M. The propagation of uncertainty through travel demand models: an exploratory analysis. *Ann. Reg. Sci.* **2002**, *36*, 145–163. [[CrossRef](#)]
9. Han, K.; Yao, T.; Friesz, T.L. Lagrangian-based Hydrodynamic Model: Freeway Traffic Estimation. *arXiv* **2012**, arXiv:1211.4619.
10. Kachroo, P.; Sastry, S. Traffic assignment using a density-based travel-time function for intelligent transportation systems. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1438–1447. [[CrossRef](#)]
11. Alvarez, P.; Hadi, M.; Zhan, C. Data archives of intelligent transportation systems used to support traffic simulation. *Transp. Res. Record* **2010**, *2161*, 29–39. [[CrossRef](#)]
12. Kim, K.-O.; Rilett, L.R. Simplex-based calibration of traffic microsimulation models with intelligent transportation systems data. *Transp. Res. Record* **2003**, *1855*, 80–89. [[CrossRef](#)]
13. Caceres, N.; Wideberg, J.P.; Benitez, F.G. Deriving origin destination data from a mobile phone network. *IET Intell. Transp. Syst.* **2007**, *1*, 15–26. [[CrossRef](#)]
14. Calabrese, F.; Diao, M.; Di Lorenzo, G.; Ferreira, J.; Ratti, C. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transp. Res. Part C Emerg. Technol.* **2013**, *26*, 301–313. [[CrossRef](#)]
15. Available online: <http://usblogs.pwc.com/emerging-technology/top-10-ai-tech-trends-for-2018/> (accessed on 5 December 2017).
16. Seo, T.; Bayen, A.M.; Kusakabe, T.; Asakura, Y. Traffic state estimation on highway: A comprehensive survey. *Annu. Rev. Control* **2017**, *43*, 128–151. [[CrossRef](#)]
17. Van Zuylen, H.J.; Willumsen, L.G. The most likely trip matrix estimated from traffic counts. *Transp. Res. Part B Methodol.* **1980**, *14*, 281–293. [[CrossRef](#)]
18. Small, K.A.; Verhoef, E.T.; Lindsey, R. *The Economics of Urban Transportation*; Routledge: Abingdon, UK, 2007.
19. Sheffi, Y. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*; Prentice-Hall: Upper Saddle River, NJ, USA, 1985.
20. Koppelman, F.S.; Bhat, C. *A Self-Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*; U.S. Department of Transportation, Federal Transit Administration: Washington, DC, USA, 2006.
21. Hao, P.; Boriboonsomsin, K.; Wu, G.; Barth, M.J. Modal activity-based stochastic model for estimating vehicle trajectories from sparse mobile sensor data. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 701–711. [[CrossRef](#)]
22. Nguyen, S. *Estimating an OD Matrix from Network Data: A Network Equilibrium Approach*; Publication No. 87; Université de Montreal: Montreal, QC, Canada, 1977.
23. Willumsen, L.G. *Estimation of an O-D Matrix from Traffic Counts—A Review. Working Paper*; University of Leeds: Leeds, UK, 1978.
24. Tavana, H. *Internally-Consistent Estimation of Dynamic Network Origin-Destination Flows from Intelligent Transportation Systems Data Using Bi-Level Optimization*; The University of Texas: Austin, TX, USA, 2001.

25. Ben-Akiva, M.; Bierlaire, M.; Koutsopoulos, H.; Mishalani, R. DynaMIT: A Simulation-Based System for Traffic Prediction. In *DACCORD Short Term Forecasting Workshop*; Massachusetts Institute of Technology: Delft, The Netherlands, 1998.
26. Jayakrishnan, R.; Mahmassani, H.S.; Hu, T.Y. An evaluation tool for advanced traffic information and management systems in urban networks. *Transp. Res. C* **1994**, *2C*, 129–147. [[CrossRef](#)]
27. Ziliaskopoulos, A.K.; Waller, S.T. An Internet-based geographic information system that integrates data, models and users for transportation applications. *Transp. Res. Part C Emerg. Technol.* **2000**, *8*, 427–444. [[CrossRef](#)]
28. Patriksson, M. *The Traffic Assignment Problem: Models and Methods*; Dover Publications: Mineola, NY, USA, 2015.
29. Shi, Q.; Abdel-Aty, M. Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 380–394. [[CrossRef](#)]
30. Mudigonda, S.; Ozbay, K. Using Big Data and Efficient Methods to Capture Stochasticity for Calibration of Macroscopic Traffic Simulation Models. In *Celebrating 50 Years of Traffic Flow Theory*; Transportation Research Board: Portland, OR, USA, 2014; pp. 215–232.
31. Antoniou, C.; Barceló, J.; Breen, M.; Bullejos, M.; Casas, J.; Cipriani, E.; Ciuffo, B.; Djukic, T.; Hoogendoorn, S.; Marzano, V.; et al. Towards a generic benchmarking platform for origin–destination flows estimation/updates algorithms: design, demonstration and validation. *Transp. Res. Part C Emerg. Technol.* **2016**, *66*, 79–98. [[CrossRef](#)]
32. Ge, Q.; Fukuda, D. Updating origin–destination matrices with aggregated data of GPS traces. *Transp. Res. Part C Emerg. Technol.* **2016**, *69*, 291–312. [[CrossRef](#)]
33. Carrese, S.; Cipriani, E.; Mannini, L.; Nigro, M. Dynamic demand estimation and prediction for traffic urban networks adopting new data sources. *Transp. Res. Part C Emerg. Technol.* **2017**, *81*, 83–98. [[CrossRef](#)]
34. Hu, X.; Chiu, Y.; Villalobos, J.A.; Nava, E. A sequential decomposition framework and method for calibrating dynamic origin–destination demand in a congested network. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2790–2797. [[CrossRef](#)]
35. Yang, Y.; Fan, Y.; Wets, R.J.B. Stochastic travel demand estimation: improving network identifiability using multi-day observation sets. *Transp. Res. Part B Methodol.* **2018**, *107*, 192–211. [[CrossRef](#)]
36. Yang, H.; Yang, C.; Gan, L. Models and algorithms for the screen line-based traffic-counting location problems. *Comput. Oper. Res.* **2006**, *33*, 836–858. [[CrossRef](#)]
37. Qiu, T.Z.; Lu, X.Y.; Chow, A.H.F.; Shladover, S.E. Estimation of freeway traffic density with loop detector and probe vehicle data. *Transp. Res. Record* **2010**, *2178*, 21–29. [[CrossRef](#)]
38. Dafermos, S.; Nagurney, A. On some traffic equilibrium theory paradoxes. *Transp. Res. Part B Methodol.* **1984**, *18*, 101–110. [[CrossRef](#)]
39. Frederix, R.; Viti, F.; Corthout, R.; Tampère, C. New gradient approximation method for dynamic origin-destination matrix estimation on congested networks. *Transp. Res. Rec.* **2011**, *1*, 19–25. [[CrossRef](#)]
40. Nigro, M.; Cipriani, E.; Giudice, A.D. Exploiting floating car data for time-dependent origin–destination matrices estimation. *Intell. Transport. Syst.* **2018**, *22*, 157–174. [[CrossRef](#)]
41. Savrasovs, M.; Pticina, I. Methodology of OD Matrix Estimation Based on Video Recordings and Traffic Counts. *Procedia Eng.* **2017**, *178*, 289–297. [[CrossRef](#)]
42. Bonnel, P.; Hombourger, E.; Olteanu-Raimond, A.-M.; Smoreda, Z. Passive mobile phone dataset to construct origin-destination matrix: Potentials and limitations. *Transp. Res. Proced.* **2015**, *11*, 381–398. [[CrossRef](#)]
43. Yang, F.; Jin, P.J.; Cebelak, M.K.; Ran, B.; Walton, C.M. The application of venue-side location-based social networking (VS-LBSN) data in dynamic origin-destination estimation. *High Comm. Refug.* **2014**, *4*, 167–241. [[CrossRef](#)]
44. Wang, F.Y. Scanning the issue and beyond: real-time social transportation with online social signals. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 909–914. [[CrossRef](#)]
45. Jin, P.J.; Cebelak, M.K.; Yang, F.; Zhang, J.; Walton, C.M.; Ran, B. Location-based social networking data exploration into use of doubly constrained gravity model for origin-destination estimation. *Transp. Res. Rec.* **2014**, *2430*, 72–82. [[CrossRef](#)]
46. Wu, X.; Guo, J.; Xian, K.; Zhou, X. Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph. *Transp. Res. Part C Emerg. Technol.* **2018**, *96*, 321–346. [[CrossRef](#)]

47. Ma, W.; Pi, X.; Qian, S. Estimating multi-class dynamic origin-destination demand through a forward-backward algorithm on computational graphs. *arXiv* **2019**, arXiv:1903.04681.
48. Vickrey, W.S. Congestion theory and transport investment. *Am. Econ. Rev.* **1969**, *59*, 251–260.
49. Amirgholy, M.; Gao, H.O. Modeling the dynamics of congestion in large urban networks using the macroscopic fundamental diagram: User equilibrium, system optimum, and pricing strategies. *Transp. Res. Pt. B Methodol.* **2017**, *104*, 215–237. [[CrossRef](#)]
50. Zhu, S.J.; Du, L.Y.; Zhang, L. Rationing and pricing strategies for congestion mitigation: behavioral theory, econometric model, and application in Beijing. *Transp. Res. Pt. B Methodol.* **2013**, *57*, 210–224. [[CrossRef](#)]
51. Yang, H.; Tang, Y.L. Managing rail transit peak-hour congestion with a fare-reward scheme. *Transp. Res. Pt. B Methodol.* **2018**, *110*, 122–136. [[CrossRef](#)]
52. Zang, G.; Xu, M.; Gao, Z. High-occupancy vehicle lanes and tradable credits scheme for traffic congestion management: A bilevel programming approach. *Promet* **2018**, *30*, 1–10. [[CrossRef](#)]
53. Mizera, C. Congestion Mitigation: Programs and Strategies. In Proceedings of the 2007 Transportation Scholars Conference, Ames, IA, USA, 9 November 2007.
54. Tian, L.-J.; Yang, H.; Huang, H.-J. Tradable credit schemes for managing bottleneck congestion and modal split with heterogeneous users. *Transp. Res. Part E Logist. Transp. Rev.* **2013**, *54*, 1–13. [[CrossRef](#)]
55. Zhu, D.S.; Yang, H.; Li, C.M.; Wang, X.L. Properties of the multiclass traffic network equilibria under a tradable credit scheme. *Transp. Sci.* **2015**, *49*, 519–534. [[CrossRef](#)]
56. Daganzo, C.F.; Lehe, L.J. Distance-dependent congestion pricing for downtown zones. *Transp. Res. Pt. B Methodol.* **2015**, *75*, 89–99. [[CrossRef](#)]
57. Liu, Y.; Nie, Y. A credit-based congestion management scheme in general two-mode networks with multiclass users. *Netw. Spat. Econ.* **2017**, *17*, 681–711. [[CrossRef](#)]
58. Ramos, R.; Cantillo, V.; Arellana, J.; Sarmiento, I. From restricting the use of cars by license plate numbers to congestion charging: analysis for Medellin, Colombia. *Transp. Policy* **2017**, *60*, 119–130. [[CrossRef](#)]
59. Aboudina, A.; Abdelgawad, H.; Abdulhai, B.; Habib, K.N. Time-dependent congestion pricing system for large networks: integrating departure time choice, dynamic traffic assignment and regional travel surveys in the Greater Toronto Area. *Transp. Res. Part A Policy Pract.* **2016**, *94*, 411–430. [[CrossRef](#)]
60. Morton, C.; Lovelace, R.; Anable, J. Exploring the effect of local transport policies on the adoption of low emission vehicles: Evidence from the London congestion charge and hybrid electric vehicles. *Transp. Policy* **2017**, *60*, 34–46. [[CrossRef](#)]
61. Wu, D.; Yin, Y.; Lawphongpanich, S. Pareto-improving congestion pricing on multimodal transportation networks. *Eur. J. Op. Res.* **2011**, *210*, 660–669. [[CrossRef](#)]
62. Silva, J.; Morency, C.; Goulias, K.G. Using structural equations modeling to unravel the influence of land use patterns on travel behavior of workers in Montreal. *Transp. Res. Part A Policy Pract.* **2012**, *46*, 1252–1264. [[CrossRef](#)]
63. Zhou, J.; Wang, Y.; Schweitzer, L. Jobs/housing balance and employer-based travel demand management program returns to scale: Evidence from Los Angeles. *Transp. Policy* **2012**, *20*, 22–35. [[CrossRef](#)]
64. Zhou, J. From better understandings to proactive actions: Housing location and commuting mode choices among university students. *Transp. Policy* **2014**, *33*, 166–175. [[CrossRef](#)]
65. Peng, Z.-R. The jobs-housing balance and urban commuting. *Urban Stud.* **1997**, *34*, 1215–1235. [[CrossRef](#)]
66. Zhao, P.; Lü, B.; Roo, G.D. Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era. *J. Transp. Geogr.* **2011**, *19*, 59–69. [[CrossRef](#)]
67. Zhou, X.; Tong, L.; Mahmoudi, M.; Zhuge, L.; Yao, Y.; Zhang, Y.; Shi, T. Open-source VRPLite package for vehicle routing with pickup and delivery: A path finding engine for scheduled transportation systems. *Urban Rail Transit* **2018**, *4*, 68–85. [[CrossRef](#)]
68. Zhuge, L.; Li, W.; Guo, J.; Xian, K.; Wu, X.; Zhou, X. A Tree-Based Reoptimization Framework for Solving Traffic Assignment Problem in Rapid Decision Making Applications. In Proceedings of the 18th COTA International Conference of Transportation Professionals: Intelligence, Connectivity, and Mobility, CICTP 2018, Beijing, China, 5–8 July 2018; pp. 205–214.
69. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, UK, 2016.

70. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
71. Park, J.; Shim, K.; Lee, S.; Kim, M. Classification of Application Traffic Using Tensorflow Machine Learning. In Proceedings of the 2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS), Seoul, Korea, 27–29 September 2017; pp. 391–394.
72. Bergstra, J.; Bastien, F.; Breuleux, O. Theano: Deep Learning on GPUs with Python. In *Big Learn Workshop, NIPS'11*; Microtome Publishing: Granada, Spain, 2011; pp. 1–48.
73. GitHub. Calculus on Computational Graphs: Backpropagation. Available online: <http://colah.github.io/posts/2015-08-Backprop/> (accessed on 31 August 2015).
74. Zhao, X.; Yan, X.; Yu, A.; Van Hentenryck, P. Modeling stated preference for mobility-on-demand transit: A comparison of machine learning and logit models. *arXiv* **2018**, arXiv:1811.01315.
75. Hensher, D.A.; Button, K. *Handbook of Transport Modelling*; Emerald Group Publishing Limited: Bingley, UK, 2007.
76. Chen, C.; Ma, J.; Susilo, Y.; Liu, Y.; Wang, M. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transp. Res. Part C Emerg. Technol.* **2016**, *68*, 285–299. [[CrossRef](#)] [[PubMed](#)]
77. Tang, J.; Song, Y.; Miller, H.J.; Zhou, X. Estimating the most likely space–time paths, dwell times and path uncertainties from vehicle trajectory data: A time geographic method. *Transp. Res. Part C Emerg. Technol.* **2016**, *66*, 176–194. [[CrossRef](#)]
78. Lu, Z.; Rao, W.; Wu, Y.-J.; Guo, L.; Xia, J. A Kalman filter approach to dynamic OD flow estimation for urban road networks using multi-sensor data. *J. Adv. Transp.* **2015**, *49*, 210–227. [[CrossRef](#)]
79. Van Lint, H.; Djukic, T. Applications of Kalman filtering in traffic management and control. In *New Directions in Informatics, Optimization, Logistics, and Production*; Institute for Operations Research and the Management Sciences (INFORM): Catonsville, MD, USA, 2002; pp. 59–91.
80. Bertsekas, D.P. *Dynamic Programming and Optimal Control*; Athena Scientific: Belmont, MA, USA, 2000.
81. Available online: <https://github.com/Grieverwzn/Big-data-driven-computational-graph> (accessed on 23 March 2019).
82. Kitamura, R.; Chen, C.; Pendyala, R.M.; Narayanan, R. Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation* **2000**, *27*, 25–51. [[CrossRef](#)]
83. Pendyala, R.M.; Kitamura, R.; Kikuchi, A.; Yamamoto, T.; Fujii, S. Florida activity mobility simulator: Overview and preliminary validation results. *Transp. Res. Rec.* **2005**, *1921*, 123–130. [[CrossRef](#)]
84. Liu, J.; Kang, J.E.; Zhou, X.; Pendyala, R. Network-oriented household activity pattern problem for system optimization. *Transp. Res. Procedia* **2017**, *23*, 827–847. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).