The impact of GPS-enabled shortest path routing on mobility: a game theoretic approach

Théophile Cabannes (corresponding author)
École Polytechnique, Palaiseau, France
652 Sutardja Dai Hall, Berkeley, CA 94720-1710
email: theophile.cabannes@gmail.com

Marco Antonio Sangiovanni Vincentelli
École Polytechnique, Palaiseau, France
email: marco-antonio.sangiovanni-vincentelli@polytechnique.edu

Alexander Sundt
University of California, Berkeley
email: asundt@berkeley.edu

Hippolyte Signargout
École Normale Supérieure de Lyon, Lyon, France
email: hippolyte.signargout@ens-lyon.fr

Emily Porter
University of California, Berkeley
email: emilyporter@berkeley.edu

Vincent Fighiera
École Nationale Supérieure des Arts et Métiers, Paris, France
email: vincent.fighiera@gmail.com

Juliette Ugirumurera
Lawrence Berkeley National Lab, Berkeley
email: Ugirumurera@lbl.gov

Alexandre M. Bayen
Institute of Transportation Studies
Electrical Engineering and Computer Sciences
Civil and Environmental Engineering
University of California, Berkeley

July 31, 2017

ABSTRACT

This article investigates the impact of app use on traffic patterns. With ubiquity of traffic information and the increased use of routing apps, urban and suburban areas in the US have seen a recent rise in "cut-through" traffic and related congestion patterns. This increase is suspected to be both an instantaneous phenomenon (a natural response of routing apps to special events, accidents, or other problems reducing capacity locally in transportation networks) and a trend (progressive increase of such traffic over time, with a corresponding shift in demand on the transportation network). Data is presented that supports both theories on the I-210 corridor in the LA Basin. Then, two models (including a new one) are presented to capture these phenomena. The models are extensions of known user equilibrium formulations that take into account routing preferences of motorists when using apps. The ability of the models to capture the trends observed in field data (and often mentioned in the popular media) are validated on benchmark examples for the same corridor. Simulations are run in which the percentage of routing app users is progressively increased to demonstrate the impact of these apps on traffic patterns and, in particular, potential convergence to Nash equilibria solutions of traffic flow problems. Finally, large-scale simulations using the micro-simulation tool Aimsun are also shown to recreate these issues.

INTRODUCTION

Context of this work

The rise of traffic congestion in the US

Traffic congestion is increasing at alarming rates in US cities [1] which impacts economic growth and productivity, and also leads to increased fuel consumption and emissions. This costs the US billions in GDP every year [2]. In recent years, a new type of traffic congestion has emerged that did not previously exist: *cut-through* traffic. This phenomenon has become increasingly disruptive in modern cities and is garnering the attention of residents, governments, and the media [3]. This phenomenon is both instantaneous (i.e. for example the response to an accident, a special event, or a local disturbance to the transportation network), and a trend (i.e. the progressive modification of demand due to the increased use of this information by motorists).

Genesis of the problem

In the early ages of the mobile internet, routing apps appeared as a viable tool for the early adopters equipped with an in-vehicle navigation system or an aftermarket personal navigation device. As the volume of vehicles transmitting data progressively increased [4], coverage of the road network expanded dramatically, leading to the emergence of major traffic information and routing service providers: INRIX, HERE, Google, Apple, Waze, TomTom, etc. In the US today, at least 60M motorists use Google Maps/Waze, and 40M motorists use Apple [5]. With *Mobility as a Service* companies like Uber or Lyft using single routing apps for nearly all their drivers, routing app usage has increased dramatically and has radically disrupted mobility patterns in large cities. For example, in San Francisco, Uber and Lyft have 45,000 registered drivers, almost all of whom are using the same routing app (Google Maps/Waze or the native Uber maps). These recent trends have led to new congestion patterns that never existed before [6].

Description of the problem

The core of the problem is that mapping apps provide a shortest path from origin to destination. Despite numerous, well-publicized claims from routing service companies and elected officials [7], these algorithms do not provide a solution to congestion and, as is well known in economic theory, do not converge to socially optimal solutions [8, 9]. It has long been known that by providing "selfish" solutions to users competing for the same commodity, one can at best hope for a *Nash equilibrium*, sometimes also referred to as *user equilibrium* or *Wardrop equilibrium* in traffic theory [9, 10] (for dynamic traffic assignment, see [11]).

Scale and growth of the problem

While shortcuts were once known only to locals, these routes are now available to anyone using navigation apps. The article refers to people who use navigational applications to route themselves as app users and those who do not as non-app users. Recent news articles have reported on communities that are negatively affected by these applications. There are several negative externalities arising from cut-through traffic: increased travel times for residents, safety concerns for pedestrians in the affected areas, and public policy challenges. Local infrastructure, which is financed by and for local taxpayers, receives a higher demand due to cut-through traffic. For instance, during peak hour in the Los Angeles neighborhood of Sherman Oaks, a residential street reaches more than 1000% of its sustainable load [12]. People have resorted to taking actions such

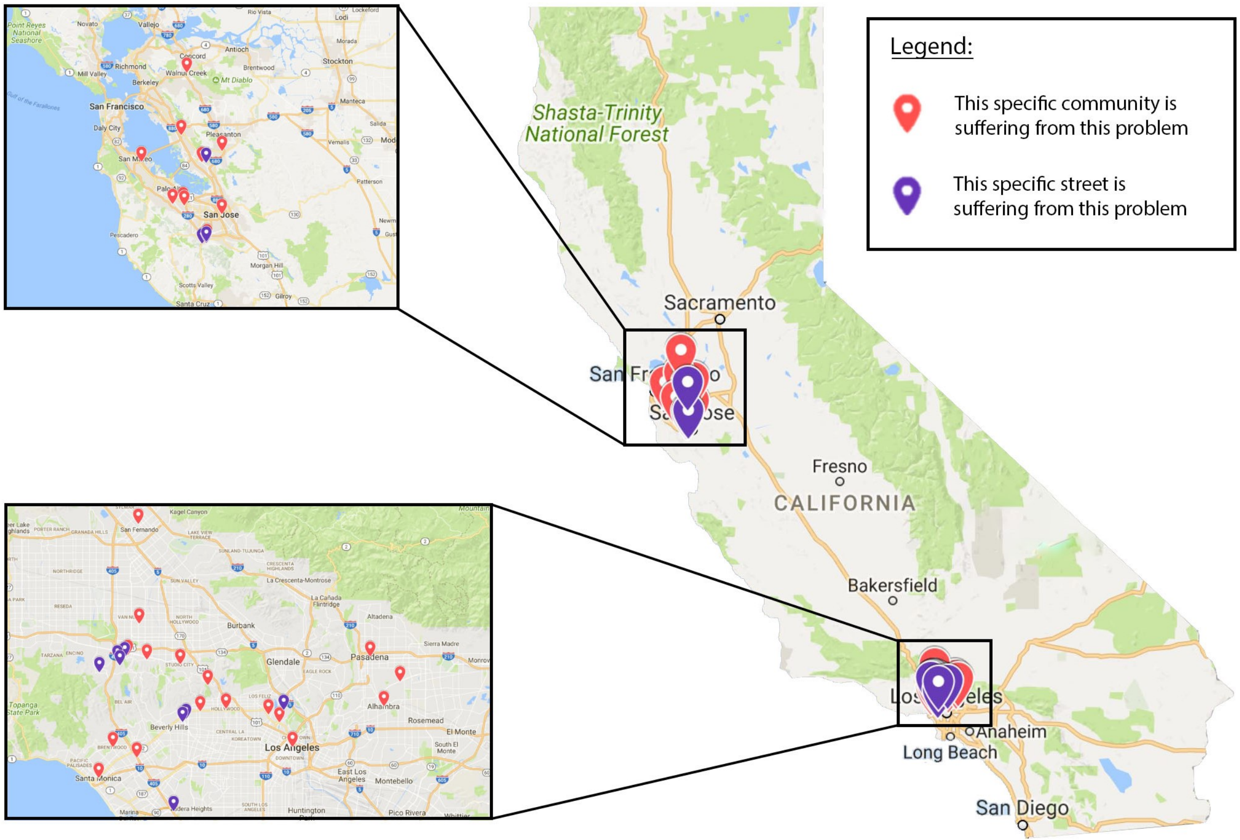


FIGURE 1 Locations where problems related to app- induced cut-through traffic have been reported in California in the popular press over the last two years, with actual location of either the problem or the complaint (when available).

as posting false reports to crowd-sourcing apps like Waze. A recent online petition targeting Waze and Google Maps for ruining Los Gatos residents' quality of life immediately gathered hundreds of supporters [13]. Local governments are also under pressure to implement countermeasures like lower speed limits, stop signs, and turn prohibitions to discourage excessive through-traffic [14]. Figure 1 shows a map of known locations where this problem has been significant enough to lead to discussion in the popular media.

Contributions of this article

This article focuses both on the instantaneous aspect of the cut-through traffic problem and long term trends. The key contributions of the article are as follows:

1. Experimental measurement of the phenomena during high congestion: For specific routes commonly suggested by routing apps, we use INRIX speed data on a specific day in LA to show that travel times on arterial roads and the I-210 are equalized during peak hours between Pasadena and Azusa. We also show that arterial road detours can be as much as 20% faster than the corresponding I-210 route.

2. Observation of trends over longer time periods: Using PeMS data (2013 to 2017) and

INRIX data (2014 to 2015), we also show that an increasing number of drivers might be using shortcuts, leading to a 3X flow increase on some off ramps over four years and a 14% decrease in speed on some arterial roads over one year.

3. Modeling: For analytical purposes, we use a previously developed model, the *cognitive cost model* [15], aimed at modeling the behavioral difference between app users and non-app users. We also introduce a new model, called the *restricted path choice model*, in which non-app users choose their route from a restricted path choice set.

4. High Performance Computing (HPC) implementation: We implement the Frank-Wolfe algorithm for the multi-class user equilibrium models presented above in an HPC environment, and parallelize the corresponding shortest-path computation step (the most time-consuming step in the algorithm). This significantly reduces the computation time of the user equilibrium assignment when run on the entire LA network.

5. Analysis of model results: Both models exhibit the occurrence of reroute patterns in high congestion situations and the overall rise in travel times on arterial roads as the number of app users increases. These results replicate the rerouting phenomena and the trends observed in Contributions 1 and 2.

6. Further validation in micro-simulations: The negative externalities imposed on the traffic system by navigation apps are further illustrated in micro-simulation, using TSS' Aimsun, in which each vehicle's ability to use app information is explicitly modeled, and the corresponding effects measured.

Throughout the article, we outline as needed the limitations of the data used, i.e. the fact that we do not have access to individual vehicles' trajectories and corresponding app routing requests. Our measurements are based on assumptions that at best confirm phenomena and trends, but do not prove them. Such certainty could only be obtained by having access to actual routing service provider data (which is not shared with the public by these companies).

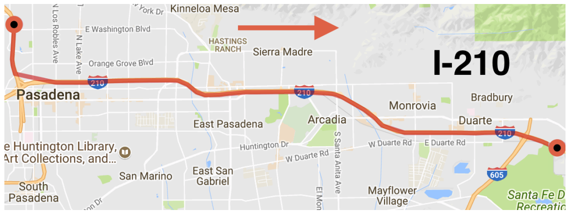
Organization of the article

The remainder of this article is organized as follows. We first present the analysis of field data for a subset of the I-210 corridor in the LA Basin and show evidence of the phenomena and trends of interest. We subsequently introduce the two models, the *cognitive cost model* [15], and the new *restricted path-choice model*. The ability of these models to exhibit the phenomena of interest is then demonstrated on numerical computations at scale in the LA Basin. Finally, we show the ability to capture these phenomena with dynamic models such as micro-simulators, with TSS' Aimsun.

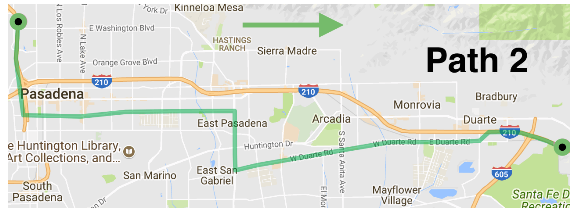
ANALYSIS OF REROUTES FROM LOOP AND ARTERIAL PROBE DATA

In this section, we use field data to quantify reroute phenomena for a highly congested day. We then present trends over a longer period to display increasing travel times on arterial roads due to an increase in arterial flow and a subsequent decrease in speed.

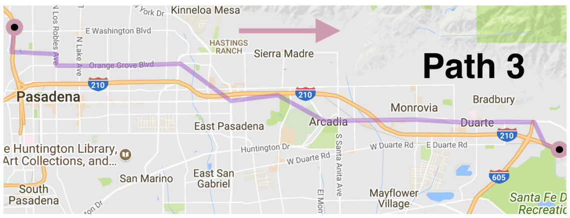
A sub-network of the LA Basin along the eastbound I-210 corridor is studied. Because of the geography of the corridor, paths parallel to the I-210 are good candidates for alternate routes.



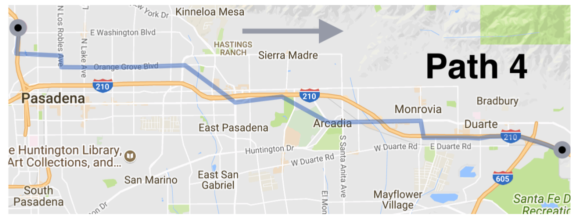
(a)



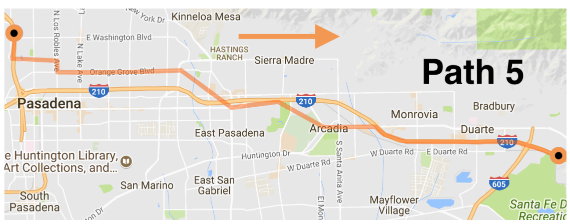
(b)



(c)



(d)



(e)

Path	I-210	2	3	4	5
Distance (mi)	13.7	15.8	14.8	14.6	14.5
Free flow travel time (min)	13	34	31	26	25

(f)

FIGURE 2 (a) I-210 freeway section considered for this work, with four alternative arterial paths (b, c, d, e). These paths have been chosen among the routinely suggested routes provided by Google Maps for the experiment because they include primarily arterial roads. (f) shows the distance and the free flow travel time of each path.

Five paths (see Figure 2) that lead from northwest Pasadena to Azusa (in northeast Los Angeles) are considered, one along I-210 and four alternative routes. These paths were chosen among the routinely suggested routes provided by Google Maps.

Data source

Data used in this analysis is taken from both INRIX and the Performance Measurement System (PeMS) [16]. INRIX data includes instantaneous link speeds (from which link travel times can be calculated) from 2014 through 2015. The travel times of the five considered paths have been computed from this data. We chose instantaneous travel time but could have chosen achieved travel time.

Travel time equalization during peak congestion

This section assesses the rise in travel time on the I-210 and the equalization of travel times between this freeway and the arterial routes parallel to it. Compared to average conditions shown in INRIX, the domain of interest shown in Figure 2 was highly congested during the evening peak hours (3:30 PM until 6:30 PM) on March 10, 2014.

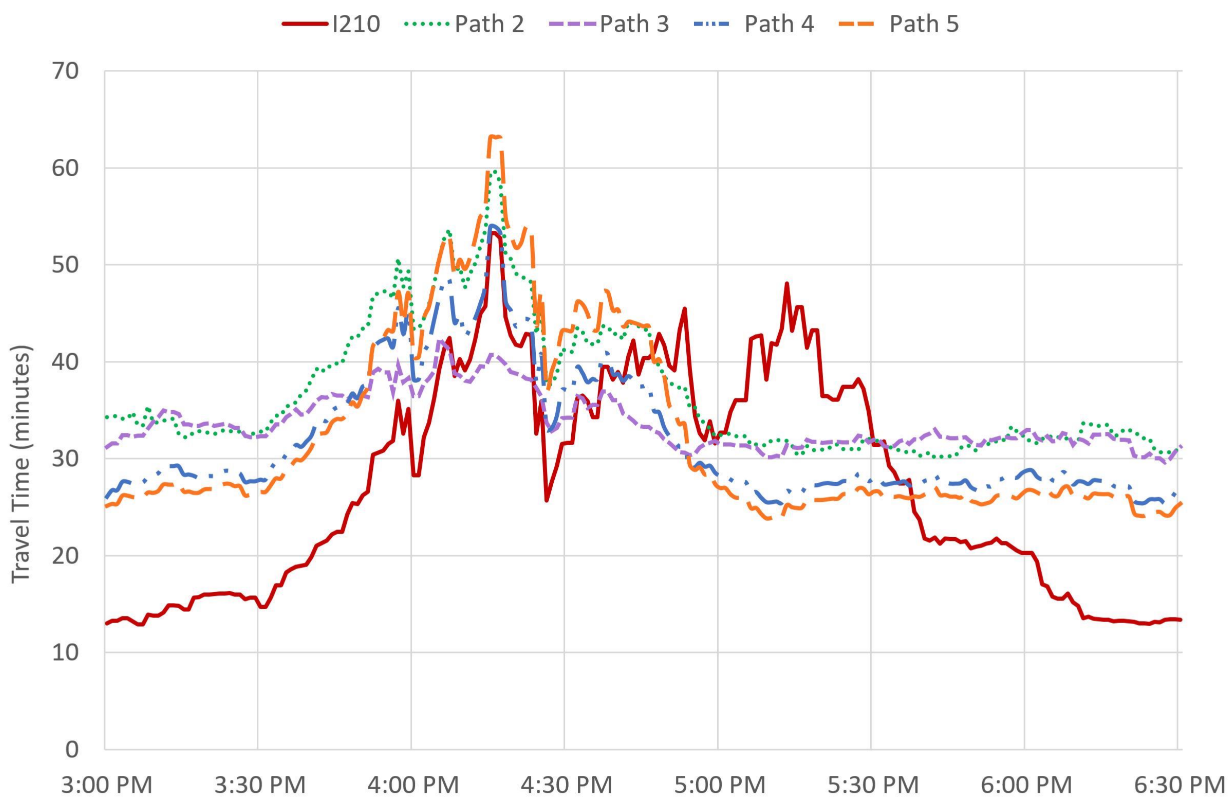


FIGURE 3 Evolution of travel times on the paths parallel to the I-210 considered in Figure 2 (March 10, 2014). At the beginning and end of the peak hour, the travel time on the freeway is close to the free flow travel time, and the freeway is faster than the arterial road routes. However, when the freeway travel time increases, the arterial detours become beneficial alternative routes (up to 20% faster at 4:20). This figure shows that in high congestion, drivers can reroute themselves to arterial roads in order to reduce their travel time, leading to travel time equalization among possible parallel routes.

As seen in Figure 3, at the start of the peak hour on March 10, 2014, around 4 PM, the travel time spikes, along with the travel times of the alternative paths. Corresponding with this spike is a decrease in the average difference in travel time between alternative paths and the I-210 freeway.

This increase and convergence of travel times is expected, as drivers using navigation applications are rerouted to arterial roads to minimize their own travel times.

Observation of yearly trends

In addition to observing the effect of navigational apps during peak hour of a single day, we examine trends on a larger time scale. As the number of app users increases, it is expected that the travel time on arterial streets will increase as well, due to increased flow rerouting around congestion on the freeway.

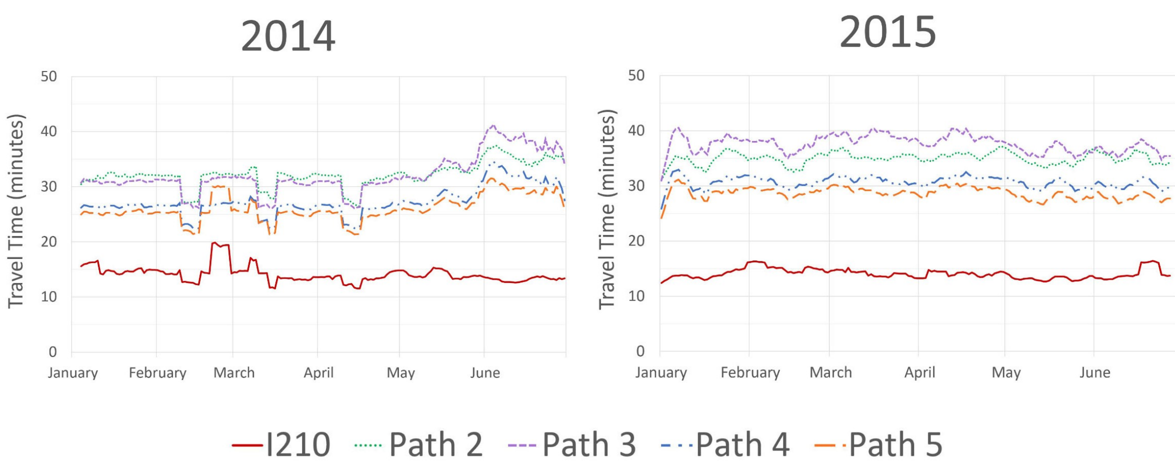
This phenomenon is observed in the evolution of the travel times shown in Figure 4a. The average travel time on the paths was computed during peak hours for each week from January to June in each year. As expected, in a one year time span, the travel times along the alternative paths increased by roughly 20%, i.e around five minutes. However, without direct access to Google Maps/Waze user data, we cannot be certain that rerouted cars actually used navigation apps. Possible alternate reasons for this increase could include demographic growth, urban activities development or other causes.

Figure 4a also shows that the I-210 is always faster than the arterial roads. The paths were suggested by Google Maps and chosen because they contained primarily arterial roads (which made computation of travel time easier). It is possible that the reroute paths suggested by apps are highly variable over time. These variable paths could include only portions of the arterial roads that are considered here. This would still explain the increase in the travel times of the alternative paths.

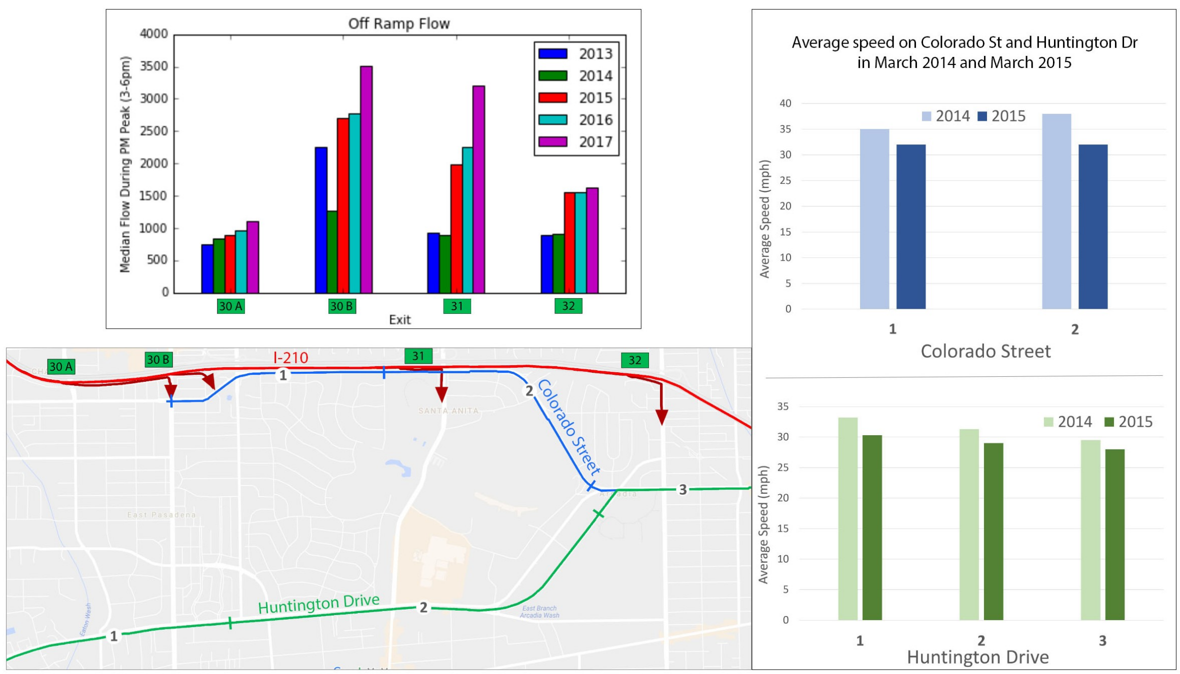
The travel time on the I-210 path oscillates around 15 minutes over the two years, remaining roughly constant over time. This occurrence can be explained by *latent demand* or by the very low marginal cost of extracting vehicles from the freeway (when a typical freeway lane capacity is around 2000 veh./h).

Additionally, PeMS data was analyzed, which consisted of flow data from inductive loop sensors embedded along the I-210 roadway (path 1) and its associated ramps. The median (chosen to reduce the effect of outliers) of the total evening peak flow was found over the weekdays of the month of March. Specifically, flows exiting the freeway were analyzed at 4 different off ramps and the yearly trend was examined. The selected off ramps include exits commonly suggested by applications like Google Maps (Michillinda Ave and Baldwin Ave exits), and the exits directly before and after them. All off ramps along this route were examined and the selected off ramps showed significant change as they were most often suggested by the applications.

A significant increase in flow using these off ramps is observed between 2013 and 2017, as shown in Figure 4b. Specifically at the exits for Michillinda Avenue and Baldwin Avenue, we see a 1.5- and 3-fold increase respectively over 4 years. While some of this increase can be explained by an increase in demand to Arcadia, it is unlikely that this can be explained solely by demand growth. Additionally, since these exits are often suggested by navigation-apps which have increased in popularity during the same time period, it is likely that the increased flow can be partially explained by app usage.



(a)



(b)

FIGURE 4 a) Evolution of the average travel time computed with INRIX data on the five paths during peak hours (4:30 to 5:30 PM) considered in 2014 and 2015, for each week from January to June. While the travel time on the I-210 remains roughly constant over two years, alternative paths suffer a 20% increase in travel time. The observed drops in 2014 are irregularities from data flaws. b) Evolution of the median off ramp flow from I-210 during March weekday peak hours and the evolution of speed on parallel paths. The significant increase in off ramp flow that occurs over the years coupled with the decrease in speed on parallel paths provides evidence in favor of app-induced arterial rerouting patterns.

MODEL OF INFORMATION-ENABLED ROUTING

This section uses two models to capture the previously described phenomena, a model from previously published work [15] and a new model introduced in this article. The ability of these models to reproduce the phenomena of interest is shown in the next section.

In numerous traffic network models, drivers are assumed to possess perfect information over the state of the network. The first model presented is the *cognitive cost model*, previously published in [15]. It separates drivers into two populations: those who use navigational applications to route themselves (app users) and those who do not (non-app users). In this model, non-app users incur a “cognitive cost” to access arterial roads. However, this cost does not depend on the particular journeys of non-app users, as they are uniformly discouraged from taking arterial roads. This type of modeling does not perfectly account for an actual lack of information on behalf of non-app users.

In order to extend the set of features encompassed by the aforementioned model, we also introduce a new mathematical approach that integrates this lack of information differently is presented - the *restricted path choice* model. This new approach considers the limited knowledge of non-app drivers by restricting their path-choice set, which is also dependent on the specific od pair of the trip.

We first introduce general network notation, which applies to both models.

Framework

Definition 4.1 (Network, paths, and demand)**.** Given a finite strongly connected directed graph $\mathcal{G}$ with vertex set $\mathcal{V}$ and edges set $\mathcal{E}$, i.e. $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, for each origin $o \in \mathcal{V}$ and destination $d \in \mathcal{V}$:

- Let $\mathcal{P}_{od}$ be the set of feasible paths without cycles from o to d.
- Let $\mathrm{r}_{od} \geq 0$ be the total number of vehicles that make the journey $o \to d$, per unit of time. We denote the demand matrix $\mathbf{r} = (\mathrm{r}_{od})_{o,d\in\mathcal{V}}$.

Definition 4.2 (Path flows)**.** For each path $\mathbf{p} \in \mathcal{P} = \bigcup\limits_{o,d\in\mathcal{V}} \mathcal{P}_{od}$:

- Let $h_{\mathbf{p}}$ be the flow of vehicles using path $\mathbf{p}$, in vehicles per unit of time (path flows). We denote the path flow vector $\mathbf{h} = (h_{\mathbf{p}})_{\mathbf{p}\in\mathcal{P}}$.
- We define $\delta_{\mathbf{p}} := (\delta_{\mathbf{p}}(e))_{e\in\mathcal{E}}$, where $\delta_{\mathbf{p}}(e) = \begin{cases} 1 \text{ if } e \in \mathbf{p} \\ 0 \text{ else} \end{cases}$
  This is the indicator vector of the links included in path $\mathbf{p}$. We denote the incidence matrix $\Delta = (\delta_{\mathbf{p}})_{\mathbf{p}\in\mathcal{P}}$

Definition 4.3 (Link flows)**.** For each link $e \in \mathcal{E}$, let f_e be the flow of vehicles using link e per unit of time (link flow). We denote the link flow vector $\mathbf{f} = (f_e)_{e\in\mathcal{E}}$.

Remark 4.1 (Static model)**.** Static equilibrium conditions are assumed. This assumption is commonly made in 15 minutes increments in the practitioner’s community. Therefore, for any path $\mathbf{p}$, $h_{\mathbf{p}}$ remains constant over time and $\mathbf{f} = \Delta\mathbf{h}$.

Definition 4.4 (Feasible assignment)**.** Given a demand matrix $\mathbf{r} \in \mathbb{R}_+^{|\mathcal{V}|\times|\mathcal{V}|}$:

- Let $\mathcal{H}_{\mathbf{r}} = \left\{ \mathbf{h},\ \mathbf{p} \in \mathcal{P} : h_{\mathbf{p}} \in \mathbb{R}_+,\ o, d \in \mathcal{V} : \sum_{\mathbf{p} \in \mathcal{P}_{od}} h_{\mathbf{p}} = \mathrm{r}_{od} \right\}$ be the set of feasible path flow allocations.
- Let $\mathcal{F}_{\mathbf{r}} = \{\Delta \mathbf{h},\ \mathbf{h} \in \mathcal{H}_{\mathbf{r}}\}$ be the set of feasible link flow allocations.

Definition 4.5 (Travel time)**.** For each link $e \in \mathcal{E}$, let t_e be the travel time on link e. We denote the travel time vector $\mathbf{t} = (t_e)_{e \in \mathcal{E}}$. For each path $\mathbf{p} \in \mathcal{P}$, we define $t^{\mathbf{p}}$ as the travel time on path $\mathbf{p}$ as the sum of the travel times on each link that is included in path $\mathbf{p}$, i.e. $t^{\mathbf{p}} = \delta_{\mathbf{p}}^{\intercal} \cdot \mathbf{t}$.

Remark 4.2 (Separability of travel time)**.** For each $\mathbf{f} \in \mathcal{F}_{\mathbf{r}}$, we denote $\mathbf{f} = (f_e)_{e \in \mathcal{E}}$ (link flow allocation). t_e is assumed to be only a function of f_e: $t_e(f_e)$. So $\mathbf{t}(\mathbf{f}) = (t_e(f_e))_{e \in \mathcal{E}}$.

Definition 4.6 (Cost function)**.** For every path allocation $\mathbf{h} \in \mathcal{H}_{\mathbf{r}}$, we define the cost function of each path $\mathbf{p} \in \mathcal{P}$ by $c_{\mathbf{p}}(\mathbf{h})$. We denote $\mathbf{c} = (c_{\mathbf{p}})_{\mathbf{p} \in \mathcal{P}}$.

Definition 4.7 (User equilibrium)**.** Given a traffic demand $\mathbf{r}$, a path allocation $\mathbf{h} \in \mathcal{H}_{\mathbf{r}}$ is a user equilibrium if and only if:

$$\forall o, d \in \mathcal{V},\ \forall \mathbf{p} \in \mathcal{P}_{od},\ h_{\mathbf{p}} \cdot (c_{\mathbf{p}}(\mathbf{h}) - \min_{\mathbf{q} \in \mathcal{P}_{od}} c_{\mathbf{q}}(\mathbf{h})) = 0 \tag{1}$$

Remark 4.3 (Wardrop's first condition [10])**.** At a **user equilibrium**, the travel time on all used routes between an od pair are equal, and less than those which would be experienced by a single vehicle on any unused route in the network.

Cognitive cost routing model

In the cognitive cost model [15], drivers are separated into two categories: app users and non-app users. App users have perfect knowledge of the traffic network. Their cost function is their travel time, as opposed to non-app users who are made to pay a multiplicative "cognitive cost" C for accessing arterial roads. This cognitive cost C is mathematically accounted for in the cost function of non-app users, $c_{\mathbf{p}}^{nr}$: $c_{\mathbf{p}}^{nr} = \delta_{\mathbf{p}}^{\intercal} \cdot (c_e^{nr})_{e \in \mathcal{E}}$ where:

$$\forall e \in \mathcal{E},\ c_e^{nr}(f_e) = \begin{cases} C \cdot t_e(f_e) & \text{if } e \text{ arterial road} \\ t_e(f_e) & \text{if } e \text{ freeway} \end{cases} \tag{2}$$

We denote $c_{\mathbf{p}}^{r}$ the cost function for a routed users on the path $\mathbf{p}$.

Definition 4.8 (Flow of routed and non-routed users, feasible path flow allocation)**.** For any $\mathbf{p} \in \mathcal{P}$, $h_{\mathbf{p}}^{r}$ and $h_{\mathbf{p}}^{nr}$ are the flow of routed users and non-routed users on the edge e, respectively. Thus, we have $h_{\mathbf{p}} = h_{\mathbf{p}}^{nr} + h_{\mathbf{p}}^{r}$. Likewise, $\mathcal{H}_{\mathbf{r}}^{r}$ and $\mathcal{H}_{\mathbf{r}}^{nr}$ are the feasible path flow allocation given the demand of app users, r_{od}^{r}, and non-app users, r_{od}^{nr}. We denote $\mathbf{h}^{r} = (h_{\mathbf{p}}^{r})_{\mathbf{p} \in \mathcal{P}} \in \mathcal{H}_{\mathbf{r}}^{r}$ and $\mathbf{h}^{nr} = (h_{\mathbf{p}}^{nr})_{\mathbf{p} \in \mathcal{P}} \in \mathcal{H}_{\mathbf{r}}^{nr}$ and we have $\mathrm{r}_{od} = \mathrm{r}_{od}^{r} + \mathrm{r}_{od}^{nr}$.

Property 4.1 (Cognitive cost variational inequality)**.** In this model, computing the user equilibrium (Wardrop's conditions) is equivalent to:
Finding $(\mathbf{h}^r, \mathbf{h}^{nr}) \in \mathcal{H}_{\mathbf{r}}^r \times \mathcal{H}_{\mathbf{r}}^{nr}$, such that:

$$\forall (\mathbf{g}^r, \mathbf{g}^{nr}) \in \mathcal{H}_{\mathbf{r}}^r \times \mathcal{H}_{\mathbf{r}}^{nr},\ c_{\mathbf{p}}^r(\mathbf{h})^\intercal \cdot (g_{\mathbf{p}}^r - h_{\mathbf{p}}^r) + c_{\mathbf{p}}^{nr}(\mathbf{h})^\intercal \cdot (g_{\mathbf{p}}^{nr} - h_{\mathbf{p}}^{nr}) \geq 0 \tag{3}$$

Remark 4.4. The equation above can be simplified. If one defines $c(\mathbf{h}) = (c^r(\mathbf{h}), c^{nr}(\mathbf{h}))^\intercal$, and $\tilde{\mathbf{h}} = (\mathbf{h}^r, \mathbf{h}^{nr})$, then the inequality is equivalent to :

$$\forall \mathbf{g} \in \mathcal{H}_{\mathbf{r}}^r \times \mathcal{H}_{\mathbf{r}}^{nr},\ c(\mathbf{h})^\intercal \cdot (\mathbf{g} - \tilde{\mathbf{h}}) \geq 0 \tag{4}$$

Remark 4.5. For any path $\mathbf{p}$, and for any $\mathbf{g} = (\mathbf{g}^r, \mathbf{g}^{nr}) \in \mathcal{H}_{\mathbf{r}}^r \times \mathcal{H}_{\mathbf{r}}^{nr}$, denote $g_{\mathbf{p}} = (g_{\mathbf{p}}^r, g_{\mathbf{p}}^{nr})$ and define $c_{\mathbf{p}}(\mathbf{h}) \otimes g_{\mathbf{p}} = c_{\mathbf{p}}^r(\mathbf{h}) \cdot g_{\mathbf{p}}^r + c_{\mathbf{p}}^{nr}(\mathbf{h}) \cdot g_{\mathbf{p}}^{nr}$

Proof: Suppose Wardrop's user equilibrium is reached at $\tilde{\mathbf{h}} = (\mathbf{h}^r, \mathbf{h}^{nr}) \in \mathcal{H}_{\mathbf{r}}^r \times \mathcal{H}_{\mathbf{r}}^{nr}$. Therefore, if $h_{\mathbf{p}}^r > 0$, then $c_{\mathbf{p}}^r(\mathbf{h}) = \pi_{o,d}^r := \min_{p \in \mathcal{P}} c_{\mathbf{p}}^r$, and if $h_{\mathbf{p}}^{nr} > 0$, then also $c_{\mathbf{p}}^{nr}(\mathbf{h}) = \pi_{o,d}^{nr} := \min_{p \in \mathcal{P}} c_{\mathbf{p}}^{nr}$, whenever $\mathbf{p}$ connects o to d. Hence, if $\pi_{o,d}$ denotes the vector $(\pi_{o,d}^r, \pi_{o,d}^{nr})$, then

$$\pi_{o,d} \otimes \tilde{h}_{\mathbf{p}} = c_{\mathbf{p}}(\mathbf{h}) \otimes \tilde{h}_{\mathbf{p}}$$

Let $\mathbf{g} := (g_{\mathbf{p}}^r, g_{\mathbf{p}}^{nr}) \in \mathcal{H}_{\mathbf{r}}^r \times \mathcal{H}_{\mathbf{r}}^{nr}$. Now, compute :

$$\begin{aligned}
c(\mathbf{h})^T \mathbf{g} &= \sum_{(o,d) \in \mathcal{V}} \sum_{\mathbf{p} \in \mathcal{P}_{od}} c_{\mathbf{p}}(\mathbf{h}) \otimes g_{\mathbf{p}} \\
&\geq \sum_{(o,d) \in \mathcal{V}} \pi_{o,d} \otimes \sum_{p \in \mathcal{P}_{od}} g_{\mathbf{p}} \\
&= \sum_{(o,d) \in \mathcal{V}} \sum_{\mathbf{p} \in \mathcal{P}_{od}} \pi_{o,d} \otimes \tilde{h}_{\mathbf{p}} \quad \text{because } \mathbf{h}, \mathbf{g} \in \mathcal{H}_{\mathbf{r}}^r \times \mathcal{H}_{\mathbf{r}}^{nr},\ \sum_{p \in \mathcal{P}_{od}} g_{\mathbf{p}}^\star = \sum_{p \in \mathcal{P}_{od}} h_{\mathbf{p}}^\star = \mathrm{r}_{od}^\star \\
&= \sum_{(o,d) \in \mathcal{V}} \sum_{\mathbf{p} \in \mathcal{P}_{od}} c_{\mathbf{p}}(\mathbf{h}) \otimes \tilde{h}_{\mathbf{p}} \quad \text{by the previous remark} \\
&= c(\mathbf{h})^T \tilde{\mathbf{h}}
\end{aligned}$$

Now suppose that we can find $\tilde{\mathbf{h}} = (\mathbf{h}^r, \mathbf{h}^{nr}) \in \mathcal{H}_{\mathbf{r}}^r \times \mathcal{H}_{\mathbf{r}}^{nr}$ that satisfies the variational inequality defined in Property 4.1. Suppose, to the contrary, that Wardrop's conditions are not reached, i.e. that there exists an origin o and a destination d, and a path $\mathbf{p}$ that connects o to d such that $c_{\mathbf{p}}^*(\mathbf{h}) > \pi_{o,d}^*$ and $h_{\mathbf{p}}^* > 0$, where $* = r$ or nr. Then define $\mathbf{g}$ as follows : take on the path $\mathbf{p}$, $g_{\mathbf{p}}^* = h_{\mathbf{p}}^* - \varepsilon$, and on a path $\mathbf{q}$ such that $c_{\mathbf{q}}^*(\mathbf{h}) = \pi_{o,d}$, take $g_{\mathbf{q}}^* = \varepsilon$; and any other component of $\mathbf{g}$ equal to the corresponding component of $\tilde{\mathbf{h}}$. If ε is small enough so that $\mathbf{g} \in \mathcal{H}_{\mathbf{r}}^r \times \mathcal{H}_{\mathbf{r}}^{nr}$, then it is easy to see that:

$$c(\mathbf{h})^\intercal (\mathbf{g} - \tilde{\mathbf{h}}) = \varepsilon (c_{\mathbf{q}}^*(\mathbf{h}) - c_{\mathbf{p}}^*(\mathbf{h})) = \varepsilon (\pi_{o,d}^* - c_{\mathbf{p}}^*(\mathbf{h})) < 0$$

This contradicts the variational inequality.

Restricted path choice routing model

In the restricted path choice routing model, app users possess perfect knowledge of the path set $\mathcal{P}_{od}$, as apps such as Google Maps or Waze can efficiently compute the set $\mathcal{P}_{od}$ and determine the optimal path (the path minimizing travel time) leading from o to d. On the other hand, non-app users do not have access to such extensive knowledge of the set $\mathcal{P}_{od}$: they tend to select routes empirically or let their navigation be determined by road signs. In this model, this heterogeneity is accounted for by reducing the path choice set of non-app users, $\mathcal{P}_{od}^{nr}$, to a subset of the path choice set of app users, $\mathcal{P}_{od}$, i.e. $\mathcal{P}_{od}^{nr} \subset \mathcal{P}_{od}$. If $\pi_{o,d}^{r}$ and $\pi_{o,d}^{nr}$ are the minimal travel times for the journey $o \to d$ for app users and non-app users respectively, the user equilibrium of the system is given by Wardrop's conditions:

$$
\begin{aligned}
h_{\mathbf{p}}^{r} \cdot \left(t^{\mathbf{p}}(\mathbf{h}) - \pi_{o,d}^{r}\right) &= 0 && \forall \mathbf{p} \in \mathcal{P}_{od} \\
h_{\mathbf{p}}^{nr} \cdot \left(t^{\mathbf{p}}(\mathbf{h}) - \pi_{o,d}^{nr}\right) &= 0 && \forall \mathbf{p} \in \mathcal{P}_{od}^{nr} \\
h_{\mathbf{p}}^{r} &\geq 0 && \forall \mathbf{p} \in \mathcal{P}_{od} \\
h_{\mathbf{p}}^{nr} &\geq 0 && \forall \mathbf{p} \in \mathcal{P}_{od}^{nr} \\
\pi_{o,d}^{r} &\geq 0 && \forall o, d \in \mathcal{V} \\
\pi_{o,d}^{nr} &\geq 0 && \forall o, d \in \mathcal{V} \\
\sum_{p \in \mathcal{P}_{od}^{nr}} h_{\mathbf{p}}^{nr} &= \mathrm{r}_{od}^{nr} && \forall o, d \in \mathcal{V} \\
\sum_{p \in \mathcal{P}_{od}} h_{\mathbf{p}}^{r} &= \mathrm{r}_{od}^{r} && \forall o, d \in \mathcal{V}
\end{aligned}
$$

Property 4.2 (Equivalence to the minimization problem)**.** Finding the user equilibrium of the system is equivalent to solving the minimization problem:

$$
\begin{aligned}
&\min_{\mathbf{f} \in \mathbb{R}^{|\mathcal{E}|}} \sum_{e \in \mathcal{E}} \int_0^{f_e} c_e(x) \cdot dx \\
\text{s.t.} \quad & \mathbf{f} = \sum_{\mathbf{p} \in \mathcal{P}} h_{\mathbf{p}}^{r} \cdot \delta_{\mathbf{p}} + \sum_{\mathbf{p} \in \mathcal{P}^{nr}} h_{\mathbf{p}}^{nr} \cdot \delta_{\mathbf{p}} \\
& \mathrm{r}_{od}^{r} = \sum_{\mathbf{p} \in \mathcal{P}_{od}} h_{\mathbf{p}}^{r} \\
& \mathrm{r}_{od}^{nr} = \sum_{\mathbf{p} \in \mathcal{P}_{od}^{nr}} h_{\mathbf{p}}^{nr}
\end{aligned}
\tag{5}
$$

Proof: A very similar variational inequality to that in Property 4.1 can be derived. Interpreting the $c(\mathbf{f})$ as a gradient yields the above conditions.

IMPLEMENTATIONS OF THE MODELS AT SCALE

Numerical solution methods

After running computations for a benchmark network, results from both models presented above are run on the entire LA basin network (including the I-210 corridor), so that the results presented here have as few peripheral effects as possible. The demand matrix considered in the present section was built from the Census Transportation Planning Products database, and based on the 2006-2010 American Community Survey Data (ctpp.transportation.org/Pages/5-Year-Data.aspx). The road network of Los Angeles was drawn from *OpenStreetMap* (OSM). The capacity of each link was determined by its category, which is provided by OSM (motorway, primary, secondary, tertiary). To obtain the numerical solutions below, we implement the Frank-Wolfe algorithm (a projected gradient descent which solves variational inequalities under given constraints) [17].

This algorithm was implemented in parallel on the Cori supercomputer at Lawrence Berkeley National Lab (http://www.lbl.gov) and has more than 60,000 compute cores that can operate simultaneously. Since the shortest-path calculations in Frank-Wolfe algorithm take more than 95% of the execution time, our HPC implementation focused on parallelizing the shortest-path computation as follows. We duplicated the network on 160 compute cores, and equally divided the od pairs among them. Thus, the 160 processes computed the shortest-paths for their assigned od pairs simultaneously. This is in contrast to the sequential Frank-Wolfe algorithm that calculates the shortest path for all od pairs, one at a time. This parallel Frank-Wolfe implementation reduced, on average, the computation time by a factor of 50 (with respect to the sequential implementation on Cori).

Results of user equilibrium models

The cognitive cost model and the restricted path choice model are implemented for two different scenarios. First, we modify the percentage of app users on a benchmark network, and then with a fixed percentage of app users we modify the capacity of the highway. Second, on the large scale LA network we modify the percentage of app users to analyze the impact of apps on travel times in the area of interest.

Accident on a benchmark network

The first scenario considered is implemented on a benchmark network, in which an accident blocks part of the highway. Three paths shown in Figure 5a, I-210 and two arterial roads, are considered.

The first computation (shown in Figures 5b and 5d) studies the impact of the number of app users on the traffic state. The highway capacity is reduced by half to 6,000 veh./h. The od demand is set to 20,000 veh./h. The results were computed with the cognitive cost model where the cognitive cost is set to 3,000 on both arterial roads. On this benchmark network, the implementation of the restricted path choice model, with non-app users only allowed to take the freeway, gives the same results.

As the ratio of app users increases, app users start rerouting to Arterial Road 2 (AR2), because it is faster than the congested highway. This transfer relieves the freeway but increases congestion on AR2. When the travel time on AR2 becomes as high as the travel time on Arterial Road 1 (AR1), app users start taking AR1 as well. Travel times stop evolving when app usage reaches 18%, which corresponds to a travel time equalization phenomenon: in these conditions, no app-user can reroute to decrease their travel time and Wardrop's first condition is reached.

The second benchmark computation (Figures 5c and 5e) studies the impact of the reduction of freeway capacity due to an accident. The computation was done with the cognitive cost model (demand set to 20,000 veh./h and cognitive cost set to 3,000). The ratio of app users is fixed to 25%.

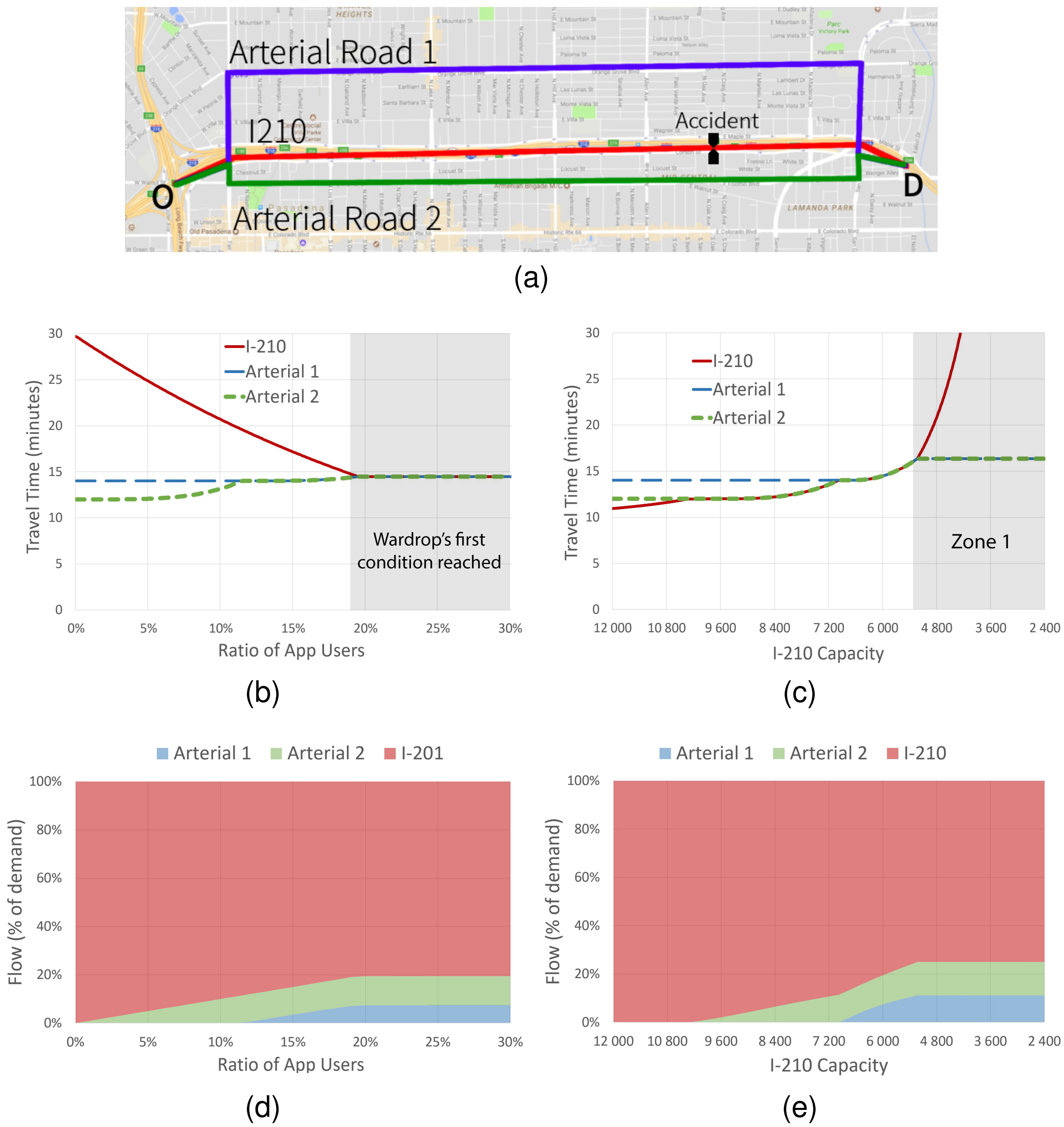


FIGURE 5 Benchmark scenario to validate the approach. (a) Possible reroutes along a segment of the I-210. Travel time (b) and distribution of flow on the I-210 and arterial roads (d) as a function of the percentage of app users. Once app usage reaches 18%, the travel time on the highway and on arterial roads are equal (i.e. Wardrop's first condition is reached). Travel time (c) and distribution of flow on the I-210 and arterial roads (e) as a function of the capacity reduction on the I-210. When the capacity is 12,000 veh./h the freeway is the fastest path. When the capacity is below 5,300 veh./h (denoted zone 1), the freeway is the longest path (all app users are on arterials while all non-app users are still on the freeway).

The results indicate that app users remain on the freeway while the capacity is higher than 10,000 veh./h at which point AR2 is as fast as the highway. App users allocate themselves among the fastest paths (I-210 and AR2) until the highway capacity drops below 6,600 veh./h and then all three paths are used. Travel times stop being equal for all paths around 5,300 veh./h because all app users have rerouted to arterial roads, while non-app users stay on the freeway, which gets increasingly congested (see zone 1 in Figure 5c).

The numerical computation show that the model is able to reproduce the phenomenon of cut-through traffic induced by shortest path routing. Once congestion spreads to arterial roads (due to an incident that causes high congestion on the freeway) travel times tend to equalize. Travel time equalization occurs as long as the number of app users is higher than the number of people routed onto arterials under Wardrop's first condition.

Increased app usage in a specific geographic area

The second scenario is implemented on the full Los Angeles network, with demand data collected from the American Community Survey, composed of 96,077 od pairs. We only vary the ratio of app users for a subset of od pairs, shown in Figure 6a (i.e. those with origins West of Pasadena and destinations East of Monrovia, which are likely to go through the I-210). For all other od pairs, we assume there are no app users in order to avoid considering rerouting phenomenon occurring outside the alternate routes that we are observing. We use the demand data for the full LA network to have realistic flows in the network while only assessing the effect of app usage on a sub-network.

A hybrid of the cognitive cost and restricted path choice models is used to compute the equilibrium flows for different ratios of app users. First, users with od pairs outside of the considered subset (see Figure 6a) are assigned to their shortest path under free flow. Then equilibrium flows for the subset of od pairs are computed using the cognitive cost model with different ratios of app users.

Figure 6b shows the evolution of travel times for three paths (from Figure 2). Travel time on the I-210 decreases rapidly with the increase of app users, while paths 2 and 3 remain stable. The travel time of path 2 increases with the first 30% of app users then decreases. As in the previous example, the reason for the initially high travel time is that the majority of drivers in the network travel via the I-210. As app users start to take path 2, the travel time along path 2 increases. When the percentage of app users is high enough, path 2 becomes too congested for new app users to take it, so app users are routed on other paths. As the shared links between path 2 and the I-210 are relieved of congestion by app users going elsewhere, the travel time for the I-210 and path 2 decrease. We do not see an increase in travel time for path 3 likely because users take other paths that we did not examine.

Figure 6c shows the evolution of travel times for the specific od pair corresponding to the start and end of these paths, for app and non-app users. The computation of the flows ensures user equilibrium, in which users of the same category (i.e. app and non-app) have the same travel time. Non-app users are required to take the I-210 therefore their travel time corresponds to that of the I-210 route. App users always take the shortest available path. Since the travel time of app users is strictly lower than those on paths 2 and 3 in Figure 6b, we conclude that these users are taking paths other than the two studied. Between 40% and 50% of app users, travel times start equalizing (as in Figure 5b) because enough people are using an arterial path for the highway to be as fast as any other route. Based on Figure 5b, one might expect that when travel times equalize they would

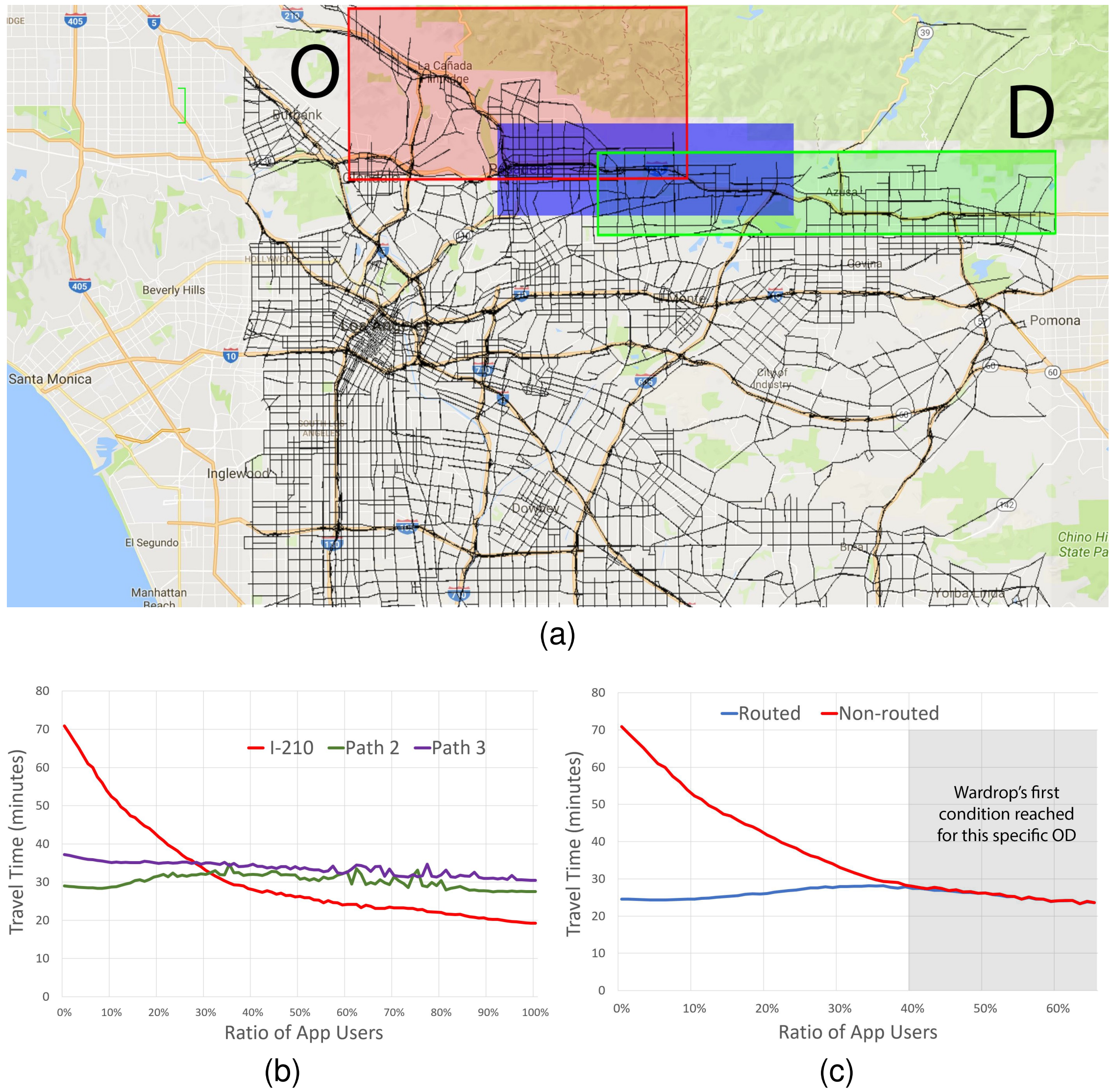


FIGURE 6 a) Network with the subset of *od* pairs. We consider *od* pairs with origin in the red area (at the traffic assignment zone granularity level), destination in the green area (same). Studied paths are in the blue area. b) Travel time trends on highway (I-210) and arterial roads (paths 2 and 3 in Figure 2) depending on the ratio of apps users. c) Travel times for both classes of users depending on ratio of app users. The same trends as those in Figure 5 can be observed: the travel times converge and equalize, with non-app users starting very slowly because of congestion on the highway, and app users travelling quickly since there are few users on the alternate routes.

also remain stable. This was the case under free flow because the path taken by non-app users was the shortest, therefore app users eventually stopped rerouting. In the present scenario, the I-210 may never be the fastest route for some od pairs. For this type of od pairs, app users will always reroute themselves outside the I-210. Thus, increasing the number of app users up to 100% will continue to decrease the travel time for the I-210.

These computations show the decrease of highway congestion with the increase of app users. The size of the network and the induced noise make it difficult to identify the right paths to study in order to get precise information about cut-through traffic. We still see that some arterial roads become congested to the point that they are as slow as the highway so app users are no longer interested in taking them. Finally, we note that on these specific paths, the reroute phenomena seems to make the system more efficient (based solely on travel times). This may not always be true, and one would have to check that the total travel time is indeed reduced. Regardless of the change in total travel time, the balance between time spent on arterials and time spent on the highway remains an equity issue (which is out of the scope of this article).

Further demonstration through microsimulations

While the previous models present several desirable features such as being analytical, compact, and implementable at scale, they are idealized and static. To further connect this work with practice, we have implemented the concepts embedded in these models into microsimulations (which integrate app usage at the individual vehicle level). This was completed using TSS' Aimsun on the I-210 corridor segment between Pasadena and Monrovia (Figure 7).

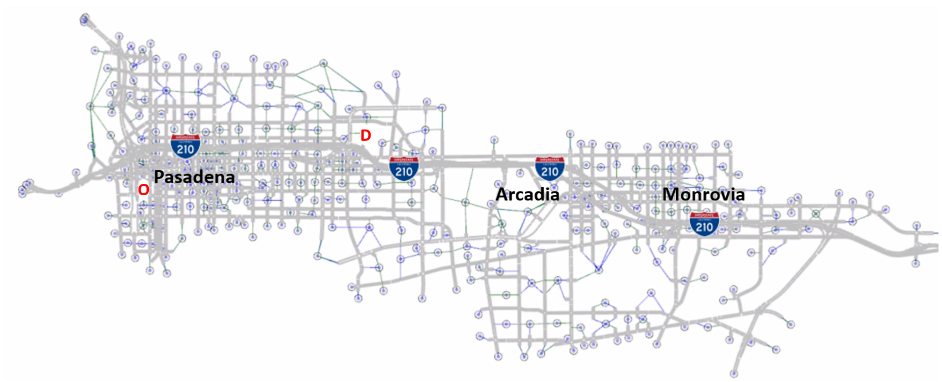
A subset of drivers is allowed to dynamically reroute based on information from the entire network (app users). App users select their shortest paths based on the state of the network. All other drivers (non app users) are routed based on static traffic assignment. The static traffic assignment is computed by the Frank-Wolfe algorithm. We expect to see non-app users spend more of their trip on the I-210, whereas app users may be encouraged to use alternate routes that make use of arterial roads to avoid highway congestion.

Experiments

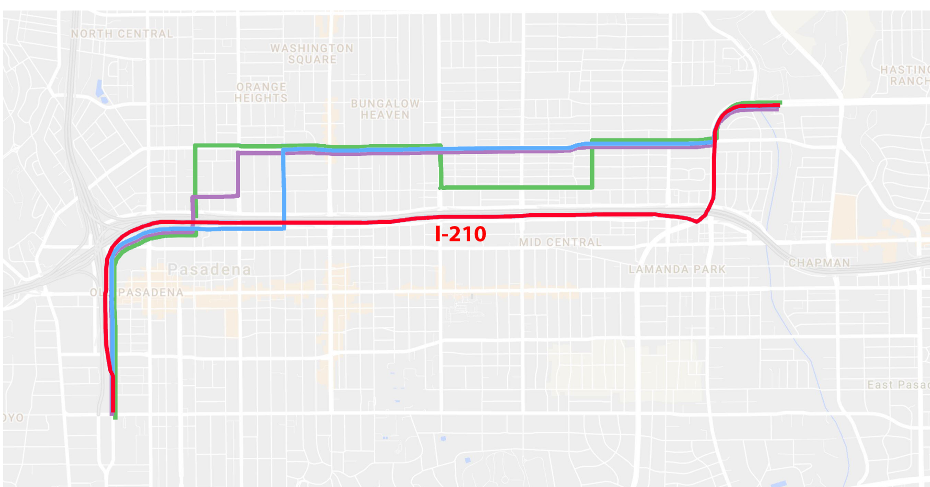
Background flow from a typical weekday (4:00 PM - 6:00 PM) is obtained from PeMS. Approximately 120,000 veh./h are routed through the network. The flow travels primarily Eastbound between 386 origins and destinations. From among the 386 ods, a specific pair, indicated in Figure 7a, was selected to determine the influence of real time information on their routing. We explore percentages of app users ranging from 25 to 75%. We only compare travel times of drivers (app users and non-app users) traveling between the specified od pair. Due to strategic routing decisions, app users spend less time traveling than the corresponding non-app users (Table 1 with average travel time denoted TT). The average travel time for app users increases as the number of app users increases. This is expected because there are more users taking arterial routes between the od pair, thus increasing the flow along these routes and increasing the travel time. Although the results show a general trend that confirms the observations made from data, the travel time between the selected od pair is relatively short (approximately 7 minutes via I-210 under free flow conditions) which means that small changes in the conditions on the road network can largely influence the obtained travel time results.

% app users	Avg. TT: app users [s]	Avg. TT: non-app users [s]	% TT saved for app users
25	744	1449	48
50	1179	1863	36
75	1288	1790	27

TABLE 1 Average achieved travel times over a 10-minute interval for app and non-app users between the *od* pair specified in Figure 7a. As app usage increases, the percent time savings for app users decreases, and the travel times between the I-210 and other routes are equalized.



(a)



(b)

FIGURE 7 (a) The selected *od* pair for the Aimsun experiments. (b) A selection of alternate paths that app users took instead of the I-210 and the I-210 route (taken by non-app users) shown in red.

CONCLUSIONS

This article investigated the impact of app use on traffic patterns. In particular, induced "cut-through" traffic was studied. First, through field data, we showed that both instantaneous phenomena (equalization of arterial and I-210 route travel times due to reroutes) and trends (progressive increase of travel time over longer periods) can be observed. While we do not possess the data to explicitly state that app usage was responsible for these facts, we strongly believe it to be the case. Subsequently, we used two models, including a new one, to capture the impacts of app-induced reroutes on traffic patterns. Simulations specific to the LA Basin and the I-210 corridor were run on benchmark networks, and at scale, to confirm the ability of the models to capture the phenomena of interest. Finally, to go beyond these idealistic models, micro-simulations were run using Aimsun, in which routing behavior is encoded at the vehicular level. These further confirm the trends we suspect from routing apps.

ACKNOWLEDGEMENTS

The authors wish to thank Anna Takayasu, Jerome Thai, Teddy Forscher, Dr. Alexander Keimer, Dr. Gabriel Gomes, and Dr. Xiaoye S. Li, for their help and fruitful discussions. The support of the California Department of Transportation and TSS are gratefully acknowledged.

REFERENCES

[1] D. Schrank, B. Eisele, T. Lomax, and J. Bak, "2015 urban mobility scorecard," 2015.

[2] S. Davis and S. Diegel, *Transportation Energy Data Book. Edition 22*. DIANE Publishing, 2002.

[3] J. Rogers, "LA traffic is getting worse and people are blaming the shortcut app Waze.," *Business Insider*, Dec. 2014.

[4] Cartesian, "The rise of mobile phones, 20 years of global adoption," 2015.

[5] A. Buczkowski, "The US mobile app report - Google maps app 64.5m users, applemaps 42m," *Geoawesomeness*, Oct. 2014.

[6] J. F. Rodriguez, "45,000 Uber and Lyft drivers may now operate in SF," *SF Examiner*, Nov. 2016.

[7] J. Geha, "Fremont: City takes steps to keep commuters on freeways and off its streets.," *East Bay Times*, Nov. 2016.

[8] J. Nash, "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[9] M. Patriksson, *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015.

[10] J. G. Wardrop, "Some theoretical aspects of road traffic research," in *ICE Proceedings: Engineering Divisions*, vol. 1, pp. 325–362, 1952.

[11] H. K. Lo and W. Y. Szeto, "A cell-based variational inequality formulation of the dynamic user optimal assignment problem," *Transportation Research Part B: Methodological*, vol. 36, no. 5, pp. 421–443, 2002.

[12] D. Bartholomew, "Sherman Oaks residents blame Waze navigation app for clogging streets.," *Los Angeles Daily News*, Jan. 2017.

[13] D. Moessinger, "Waze/Google maps ruining our neighborhoods!," *change.org*, June 2017.

[14] E. Weise, "Waze and other traffic dodging apps prompt cities to game the algorithms," *USAToday*, Mar. 2017.

[15] J. Thai, N. Laurent-Brouty, and A. Bayen, "Negative externalities of gps-enabled routing applications: A game theoretical approach," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 595–601, Nov 2016.

[16] P. Varaiya, "Freeway performance measurement system (pems), pems 6," tech. rep., California Center for Innovative Transportation, University of California, Berkeley, 2 2006.

[17] M. Fukushima, "A modified Frank-Wolfe algorithm for solving the traffic assignment problem," *Transportation Research Part B: Methodological*, vol. 18, pp. 169–177, April 1984.