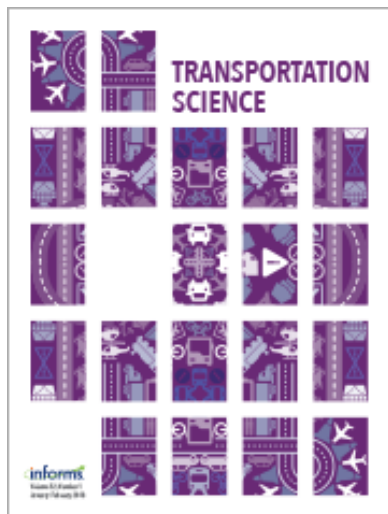




## Transportation Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### On the Value of Dynamism in Transit Networks

J. Carlos Martínez Mori, M. Grazia Speranza, Samitha Samaranayake

#### To cite this article:

J. Carlos Martínez Mori, M. Grazia Speranza, Samitha Samaranayake (2023) On the Value of Dynamism in Transit Networks. Transportation Science

Published online in Articles in Advance 10 Jan 2023

. <https://doi.org/10.1287/trsc.2022.1193>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2023, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# On the Value of Dynamism in Transit Networks

J. Carlos Martínez Mori,<sup>a,\*</sup> M. Grazia Speranza,<sup>b</sup> Samitha Samaranyake<sup>a,c</sup>

<sup>a</sup>Center for Applied Mathematics, Cornell University, Ithaca, New York 14853; <sup>b</sup>Department of Economics and Management, University of Brescia, Brescia, Italy; <sup>c</sup>School of Civil and Environmental Engineering, Cornell University, Ithaca, New York 14853

\*Corresponding author

Contact: [jm2638@cornell.edu](mailto:jm2638@cornell.edu), <https://orcid.org/0000-0002-6714-221X> (JCMM); [grazia.speranza@unibs.it](mailto:grazia.speranza@unibs.it),

<https://orcid.org/0000-0002-8893-5227> (MGS); [samitha@cornell.edu](mailto:samitha@cornell.edu) (SS)

Received: July 13, 2022

Revised: October 19, 2022

Accepted: December 8, 2022

Published Online in Articles in Advance:  
January 10, 2023

<https://doi.org/10.1287/trsc.2022.1193>

Copyright: © 2023 INFORMS

**Abstract.** The rise of on-demand mobility technologies over the past decade has sparked interest in the integration of traditional transit and on-demand systems. One of the main reasons behind this is the potential to address a fundamental trade-off in transit: the ridership versus coverage dilemma. However, unlike purely fixed systems or purely on-demand systems, integrated systems are not well understood; their planning and operational problems are significantly more challenging, and their broader implications are the source of a heated debate. Motivated by this debate, we introduce the dynamicity gap, a general concept that quantifies the attainable benefit of allowing (but not requiring) dynamic components in the response strategy to a multistage optimization problem. Although computing the dynamicity gap exactly may be intractable, we develop an analytical framework with which to approximate it as a function of problem input parameters. The framework allows us to certify the value of dynamism (i.e., a dynamicity gap greater than one) for certain combinations of problem input parameters. We showcase our approach with two sets of computational experiments, from which we gain both qualitative and quantitative insights about the settings in which the integration of transit and on-demand systems may certifiably be a worthwhile investment.

**Funding:** This work was partially supported by the National Science Foundation [Grants DMS-1839346 and CNS-1952011]. Part of this research was performed while the authors were visiting the Institute for Pure and Applied Mathematics, which is supported by the National Science Foundation [Grant DMS-1925919].

**Supplemental Material:** The online appendix is available at <https://doi.org/10.1287/trsc.2022.1193>.

**Keywords:** transit and on-demand systems • network design • stochastic optimization

## 1. Introduction

The design of transit systems is a classical yet persistently challenging problem. Part of the difficulty stems from the complexity of the application domain—it has multiple interacting components such as the physical infrastructure network (e.g., the network of bus lanes), the operational network (e.g., the set of bus lines), timetables, and crew and fleet schedules. To add to this, there may be multiple objectives and/or quality of service targets, often without a clear mathematical description and in conflict with one another. As a result, there is no single “global” optimization problem, and if there was, it would not be tractable. Rather, the design of transit systems is typically decomposed into a sequence of steps, starting from the bare bones—the design of the physical infrastructure network—and continuing on toward increasingly operational considerations. Even when the process is decomposed, the optimization problems arising in each step are usually NP-hard (see Desaulniers and Hickman 2007 and Schöbel 2012 for an overview of the transit system design process).

The rise of on-demand mobility technologies over the past decade has sparked interest in the integration

of traditional transit and on-demand systems—the number of *microtransit* (i.e., high-capacity on-demand shuttles) pilot programs conducted by transit agencies across the United States is a testament to this (see Westervelt et al. 2018 for a compilation of experiences). One of the main reasons behind this is the potential for microtransit to address a fundamental trade-off in transit: the *ridership versus coverage* dilemma. It is well known that, given a limited budget, transit networks that maximize ridership and transit networks that maximize coverage (e.g., the geographical service area) tend to be vastly different (see Walker 2012 for a practitioner-oriented discussion). Intuitively, integrated systems may bridge this gap by letting each subsystem do what it does best; transit should focus on ridership, microtransit should extend coverage as a first/last mile service, and the two should be jointly optimized.

However, unlike purely fixed systems or purely on-demand systems, integrated systems are not well understood; their planning and operational problems are significantly more challenging, and their broader implications are the source of a heated debate. Some transportation researchers and practitioners have suggested that

**Table 1.** Summary of Notation

| Symbol                                                             | Description                                                                                                        |
|--------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| $G = (V, E)$                                                       | Underlying graph topology                                                                                          |
| $T \in \mathbb{N}$                                                 | Number of stages                                                                                                   |
| $\delta > 0$                                                       | Stage duration                                                                                                     |
| $\mathcal{I} := \{I_1, I_2, \dots, I_k\} \subseteq 2^{V \times V}$ | Possible input scenarios (also known as travel demand realizations)                                                |
| $I_i \in \mathcal{I}$                                              | $i$ th possible input scenario                                                                                     |
| $I^t \in \mathcal{I}$                                              | Input scenario during the $t$ th stage                                                                             |
| $\mathcal{P}_i \subseteq \{0, 1\}^{2m}$                            | Set of integrated networks configurations that can serve $I_i$                                                     |
| $\mathcal{P}^t \subseteq \{0, 1\}^{2m}$                            | Set of integrated networks configurations that can serve $I^t$                                                     |
| $(x, z_i) \in \mathcal{P}_i$                                       | Integrated static $x$ and dynamic $z_i$ network that can serve $I_i$                                               |
| $(x, z^t) \in \mathcal{P}^t$                                       | Integrated static $x$ and dynamic $z^t$ network that can serve $I^t$                                               |
| $\mathcal{D}$                                                      | Probability distribution over $\mathcal{I}$                                                                        |
| $p_i$                                                              | $\Pr_{\mathcal{D}}[I^t = I_i]$ for every $t \in [T]$                                                               |
| $c_s := c_s(\delta) \in \mathbb{R}_{\geq 0}^m$                     | Per-stage cost vector for static network (with usage duration $\delta$ )                                           |
| $c_d := c_d(\delta) \in \mathbb{R}_{\geq 0}^m$                     | Per-stage cost vector for dynamic network (with usage duration $\delta$ )                                          |
| $c \in \mathbb{R}_{\geq 0}^m$                                      | Cost vector for a single run of the static network                                                                 |
| $\eta > 0$                                                         | Surcharge coefficient                                                                                              |
| $\delta_s > 0$                                                     | Transit headway                                                                                                    |
| $\delta_d > 0$                                                     | Microtransit batching interval                                                                                     |
| $\alpha \geq 1$                                                    | Dynamicity gap (see (2) for definition)                                                                            |
| $\theta > 0$                                                       | Relative cost coefficient (in Remark 1, let $\delta := \delta_d$ and $\theta = \eta \cdot (\delta_s / \delta_d)$ ) |
| $\theta^+ := \alpha(1)$                                            | Dynamicity gap when $\theta = 1$ (see Theorem 1)                                                                   |

on-demand systems can complement traditional transit (e.g., Feigon and Murphy 2016, Shaheen and Chan 2016, Alonso-González et al. 2018, Hall, Palsson, and Price 2018, Stiglic et al. 2018, Liu and Ouyang 2021). At the same time, others have raised concerns about or even flat-out dismissed the supposed benefits (e.g., Walker 2012, 2018; Rayle et al. 2016; Westervelt et al. 2018; Merlin 2019).

Motivated by this debate, we introduce the *dynamicity gap*, a general concept that quantifies the attainable benefit of allowing (but not requiring) dynamic components in the response strategy for a multistage optimization problem. We study the dynamicity gap within the context of the strategic planning of transit infrastructure networks: the first step in the transit system design process and arguably, the most decisive one because all subsequent steps depend on it. However, we note that the concept is more generally applicable in domains where goals can be met through a combination of static and dynamic (i.e., stage-specific) decisions. Our contributions are threefold.

1. As a conceptual contribution, we introduce the dynamicity gap.
2. Computing the dynamicity gap exactly may be intractable. Therefore, as a methodological contribution, we develop an analytical framework with which to approximate it as a function of the problem input parameters. The framework allows us to certify the value of dynamism (i.e., a dynamicity gap greater than one) for certain combinations of input parameters.
3. We study the dynamicity gap within the context of the strategic planning of transit infrastructure networks. We pose the design of integrated transit networks as a

multistage network design problem and showcase our analytical framework with two sets of computational experiments. As a scientific contribution, we provide both qualitative and quantitative insights about the settings in which the integration of transit and on-demand systems may certifiably be a worthwhile investment.

As we formalize our study, we point to Table 1 for a summary of notation.

### 1.1. Design of Transit Networks

We study the dynamicity gap within the context of centrally designed integrated transit networks. To this end, we first describe the *Steiner forest problem*, the prototypical problem in network design. We focus on this abstraction because it lends itself to mathematical analysis and comprehensive experimentation that we believe to be useful at the level of the strategic planning of transit infrastructure networks. One can in principle enhance it with operational features such as capacity constraints, detour constraints (i.e., ensuring travel demand is met through relatively direct routes), fleet rebalancing constraints, pricing, and consumer choice models; by extending the notion of picking edges to picking “lines” (i.e., picking paths in  $G$ ); by limiting the number of line transfers passengers can take; and so on. However, we note that producing and efficiently solving the resulting models are research areas in and of themselves (see Bertsimas, Ng, and Yan 2020, 2021; Luo, Samaranayake, and Banerjee 2021 for some recent representative work).

In the Steiner forest problem, we are given a connected graph  $G = (V, E)$  with costs  $c : E \rightarrow \mathbb{R}_{\geq 0}$  and a collection  $I \subseteq V \times V$  of origin-destination pairs. The problem is to find a minimum cost subset  $X \subseteq E$  of edges supporting a

path between each origin-destination pair. In the context of transit,  $G$  represents the underlying graph topology (e.g., a road network),  $I$  represents the travel demand, and  $X$  represents the installed network. More generally, we aim to distinguish between two types of installed networks: a transit network and a microtransit network. Therefore, we extend the Steiner forest abstraction by allowing us to pick two subsets  $X, Z \subseteq E$  of edges such that (s.t.)  $X \cup Z$  supports a path between each origin-destination pair. We treat  $X$  as a transit network and  $Z$  as a microtransit network so that  $X \cup Z$  is an *integrated network*—this distinction becomes much more meaningful in Section 1.2 and beyond, where we consider a multistage version of the problem. Going forward, we represent subsets  $X, Z \subseteq E$  of edges by their characteristic vectors  $x, z \in \{0,1\}^m$  where  $m = |E|$ .

## 1.2. Dynamicity Gap

We consider a multistage version of the design of integrated transit networks, wherein the transit network is static, whereas the microtransit network is dynamic. The temporal planning horizon is partitioned into  $T \in \mathbb{N}$  stages indexed by  $[T] := \{1, 2, \dots, T\}$ . It is implicit that each stage has a (say uniform) duration  $\delta > 0$ ; we elaborate on this shortly. Let  $\mathcal{I} := \{I_1, I_2, \dots, I_k\} \subseteq 2^{V \times V}$  be the collection of possible travel demand realizations (i.e., the collection of possible sets of origin-destination pairs), which we also refer to as input scenarios. For each  $t \in [T]$ , let  $I^t \in \mathcal{I}$  be the travel demand during the  $t$ th stage and  $\mathcal{P}^t$  be the finite and nonempty set of integrated network configurations that can serve it. Let  $x \in \{0,1\}^m$  be decision variables corresponding to the static network (e.g., transit) paid for at cost  $c_s := c_s(\delta) \in \mathbb{R}_{\geq 0}^m$  on every stage. Let  $z^1, z^2, \dots, z^T \in \{0,1\}^m$  be decision variables corresponding to the dynamic network (e.g., microtransit) over the stages, each paid for at cost  $c_d := c_d(\delta) \in \mathbb{R}_{\geq 0}^m$ . For  $a, b \in \mathbb{R}^m$ , let  $a \cdot b = \sum_{j=1}^m a_j b_j$  denote their dot product. Then, the design of integrated transit networks can be posed as a multistage optimization problem of the form

$$\begin{aligned} \min_{x, z^1, z^2, \dots, z^T} \quad & \sum_{t=1}^T (c_s \cdot x + c_d \cdot z^t) \\ \text{s.t.} \quad & (x, z^t) \in \mathcal{P}^t, \quad \forall t \in [T]. \end{aligned} \quad (1)$$

The constraints  $(x, z^t) \in \mathcal{P}^t$  for each  $t \in [T]$  ensure the static network  $x$  and the dynamic network  $z^t$  together serve the travel demand  $I^t$  during the  $t$ th stage.

We emphasize the general dependency of the per-stage costs  $c_s := c_s(\delta)$  and  $c_d := c_d(\delta)$  on the stage duration  $\delta$ . If a system operates at a timescale different from  $\delta$ , it is crucial that its per-stage cost is appropriately prorated. To illustrate this, let  $c \in \mathbb{R}_{\geq 0}^m$  encode the cost of a single dispatch of the static system (e.g., the cost per vehicle mile times the total vehicle miles covered by transit in a single dispatch of all routes) and  $\eta \cdot c$  encode

the cost of a single dispatch of the dynamic system, for some  $\eta > 0$ . The surcharge coefficient  $\eta$  captures the notion that static systems and dynamic systems have different operational costs on a per-mile basis, independent of their relative frequencies (e.g., accounting only for backend costs such as fuel, labor, and use of software). We allow the systems to have different timescales by distinguishing between the transit headway  $\delta_s > 0$  (i.e., the time interval between subsequent transit dispatches) and the microtransit batching interval  $\delta_d > 0$  (i.e., the time interval over which incoming travel demands are aggregated and microtransit routes reoptimized). Then, we prorate the per-stage costs of transit and microtransit as  $c_s = (\delta/\delta_s) \cdot c$  and  $c_d = (\delta/\delta_d) \cdot \eta \cdot c$ , respectively. For example, if the stages are of duration  $\delta = 1$  minute, a transit system with headway  $\delta_s = 10$  minutes incurs only 1/10th of its dispatch cost on any given stage. Going forward, we tie the stage duration to the microtransit batching interval so that  $\delta := \delta_d$ .

Let  $\text{OPT}$  denote the cost of an optimal solution to (1) and  $\text{OPT}^\Sigma$  denote the cost of an optimal static solution to (1): that is, one in which we additionally require  $z^1 = z^2 = \dots = z^T = \mathbf{0}$ , where  $\mathbf{0}$  refers to the zero vector. We define the *dynamicity gap*  $\alpha$  of (1) as the unitless coefficient

$$\alpha := \frac{\text{OPT}^\Sigma}{\text{OPT}} \geq 1. \quad (2)$$

Large values of  $\alpha$  indicate large gains from introducing dynamism. Conversely, values of  $\alpha$  close or equal to one indicate little to no gains from introducing dynamism. In this way, the dynamicity gap quantifies the value of dynamism.

## 1.3. Summary of Results

The dynamicity gap  $\alpha$  quantifies the value of dynamism, but computing it involves solving (1), which may be intractable. Moreover, we observe from (1) that  $\alpha$  is influenced by implicit and explicit parameters such as the costs  $c_s$  and  $c_d$ , the stage duration  $\delta$ , and the relationship between the sets of feasible configurations  $\mathcal{P}^1, \mathcal{P}^2, \dots, \mathcal{P}^T$ .

The overarching goal of this work is to *parametrically* study the behavior of  $\alpha$  without the need of solving the underlying multistage optimization problem. To this end, we assume  $c_s = c$  for some  $c \in \mathbb{R}_{\geq 0}^m$  and  $c_d = \theta \cdot c$  for some relative cost coefficient  $\theta > 0$ . That is, we restrict our analysis to problems of the form

$$\begin{aligned} \min_{x, z^1, z^2, \dots, z^T} \quad & \sum_{t=1}^T (c \cdot x + \theta \cdot c \cdot z^t) \\ \text{s.t.} \quad & (x, z^t) \in \mathcal{P}^t, \quad \forall t \in [T]. \end{aligned} \quad (3)$$

However, when tying this form back to transit, we assume  $\delta := \delta_d$ , and as described in Section 1.2, we prorate the per-stage costs as  $c_s = (\delta/\delta_s) \cdot c$  and  $c_d = \eta \cdot c$ .

Then, for any fixed  $\delta_s, \delta_d > 0$ , form (1) becomes

$$\begin{aligned} \min_{x, z^1, z^2, \dots, z^T} \quad & \sum_{t=1}^T \left( \frac{\delta_d}{\delta_s} \cdot c \cdot x + \eta \cdot c \cdot z^t \right) \\ \text{s.t.} \quad & (x, z^t) \in \mathcal{P}^t, \quad \forall t \in [T] \\ & = \frac{\delta_d}{\delta_s} \cdot \min_{x, z^1, z^2, \dots, z^T} \sum_{t=1}^T \left( c \cdot x + \eta \cdot \frac{\delta_s}{\delta_d} \cdot c \cdot z^t \right) \\ & \text{s.t.} \quad (x, z^t) \in \mathcal{P}^t, \quad \forall t \in [T]. \end{aligned} \quad (4)$$

Note that this is a scaled version of (3) with  $\theta = \eta \cdot (\delta_s/\delta_d)$ . In this way, we embed any dependency on  $\eta, \delta_s, \delta_d$  within  $\theta$ , which then captures both the difference in service frequency and the difference in operational costs on a per-mile basis.

Our analytical contributions around (3) are with respect to the case in which the input scenarios  $I^1, I^2, \dots, I^T$  are independent and identically distributed (i.i.d.) with probability distribution  $\mathcal{D}$  over  $\mathcal{I}$ . This constitutes a specialization because in our definition (2) of dynamicity gap, the sequence of input scenarios *need not* be stochastic (where we have distributional information over input parameters) or even uncertain (as in robust optimization, where we only *uncertainty set* information over input parameters). We first show that if the input scenarios  $I^1, I^2, \dots, I^T$  are sampled i.i.d. from a probability distribution  $\mathcal{D}$  over  $\mathcal{I}$  and moreover,  $T \rightarrow \infty$ , then we can reformulate almost surely (a.s.) the horizon-normalized version of the multistage Problem (3), wherein we scale the objective function by  $1/T$ , as the two-stage stochastic problem:

$$\begin{aligned} \min_{x, z_1, z_2, \dots, z_k} \quad & c \cdot x + \sum_{i=1}^k p_i \cdot \theta \cdot c \cdot z_i \\ \text{s.t.} \quad & (x, z_i) \in \mathcal{P}_i, \quad \forall i \in [k]. \end{aligned} \quad (5)$$

Recall  $\mathcal{I} := \{I_1, I_2, \dots, I_k\}$  is the collection of possible input scenarios. For each  $i \in [k]$ , let  $I_i \in \mathcal{I}$  be the  $i$ th input scenario, and let  $\mathcal{P}_i$  be the finite and nonempty set of integrated network configurations that can serve it. Let  $p_i = \Pr_{\mathcal{D}}[I^t = I_i]$  for every  $t \in [T]$ . Let  $x \in \{0,1\}^m$  be decision variables corresponding to the first-stage network, and for each  $i \in [k]$ , let  $z_i \in \{0,1\}^m$  be decision variables corresponding to the second-stage network under input scenario  $I_i$ . Then, the constraints  $(x, z_i) \in \mathcal{P}_i$  for each  $i \in [k]$  ensure the first-stage network  $x$  and the second-stage network  $z_i$  together serve the input scenario  $I_i$ . This intuitive result is closely related to the convergence of the sample average approximation (SAA) method shown by Kleywegt, Shapiro, and Homem-de Mello (2002).

As a corollary, in this case, the dynamicity gap  $\alpha$  of (3) reduces a.s. to the dynamicity gap of (5): the ratio between the cost of an optimal static solution to (5), that is one in which we additionally require  $z_1 = z_2 = \dots = z_k = \mathbf{0}$ , and the cost of an optimal solution to (5). This equivalence

allows us to treat the dynamicity gap  $\alpha := \alpha(\theta)$  as a function  $\alpha : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 1}$  of the relative cost coefficient  $\theta$ . In this way, our second and main analytical contribution is a certificate of the value of dynamism (i.e., a certificate that  $\alpha(\theta) > 1$ ) whenever the relative cost coefficient does not exceed a certain value. Although this certificate is not tight in general, we illustrate in Remark 2 (see Section 3) that producing it does not require solving the two-stage stochastic problem, and thus, it is (relatively) tractable.

**Theorem 1.** Suppose  $I^1, I^2, \dots, I^T$  are sampled i.i.d. from a probability distribution  $\mathcal{D}$  over  $\mathcal{I}$  and moreover,  $T \rightarrow \infty$ . Let  $\theta^+ := \alpha(1)$  be the dynamicity gap of (5) when  $\theta = 1$ —equivalently a.s., the dynamicity gap of (3) when  $\theta = 1$ . For  $\theta > 0$ , we a.s. have

$$\alpha(\theta) \geq \max \left\{ \frac{\theta^+}{\theta}, 1 \right\}. \quad (6)$$

Our choice of notation  $\theta^+ := \alpha(1)$  (as opposed to  $\alpha^+ := \alpha(1)$ ) follows from the way we use (6); it implies that if the relative cost coefficient  $\theta$  satisfies  $\theta < \theta^+$ , then  $\alpha(\theta) > 1$ . We can strengthen this result to estimate  $\alpha(\theta)$  to any arbitrary precision, provided we solve a finite number of two-stage stochastic problems (Theorem 5).

**Remark 1.** Tying this result back to transit, under the transformation from (1) to (4) wherein  $\delta := \delta_d$ ,  $c_s = (\delta_d/\delta_s) \cdot c$ , and  $c_d = \eta \cdot c$ , the condition  $\theta < \theta^+$  is equivalent to  $\eta \cdot (\delta_s/\delta_d) < \theta^+$ . To see this, note that for any fixed  $\delta_s, \delta_d > 0$ , we have

$$\begin{aligned} \alpha &:= \frac{\text{OPT}^\Sigma}{\text{OPT}} = \frac{\frac{\delta_d}{\delta_s} \cdot \min_x \sum_{t=1}^T c \cdot x}{\frac{\delta_d}{\delta_s} \cdot \min_{x, z^1, z^2, \dots, z^T} \sum_{t=1}^T \left( c \cdot x + \eta \cdot \frac{\delta_s}{\delta_d} \cdot c \cdot z^t \right)} \\ &= \frac{\min_x \sum_{t=1}^T c \cdot x}{\min_{x, z^1, z^2, \dots, z^T} \sum_{t=1}^T \left( c \cdot x + \eta \cdot \frac{\delta_s}{\delta_d} \cdot c \cdot z^t \right)} \\ &\quad \text{s.t.} \quad (x, z^t) \in \mathcal{P}^t, \quad \forall t \in [T] \end{aligned}$$

By Theorem 1, the last ratio is a.s. greater than one whenever  $\theta = \eta \cdot (\delta_s/\delta_d) < \theta^+$ , where  $\theta^+ := \alpha(1)$  is computed for the special case in which  $\theta = \eta \cdot (\delta_s/\delta_d) = 1$ —for example, if  $\eta = 1$  and  $\delta_s = \delta_d$ .

We view this as a quick, high-level rule of thumb giving a green light for the full-blown integrated transit system design process; given a microtransit batching interval  $\delta_d > 0$  and a probability distribution  $\mathcal{D}$  over input scenarios  $\mathcal{I}$  of duration  $\delta := \delta_d$ , we set  $\eta \cdot (\delta_s/\delta_d) = 1$  to (relatively) tractably compute  $\theta^+ := \alpha(1)$ , with which we can certify the value of microtransit for

certain combinations of transit frequency  $\delta_s$  and surcharge coefficient  $\eta$ , namely whenever  $\eta \cdot \delta_s < \theta^+ \cdot \delta_d$ .

By (6), higher values of  $\theta^+$  lead to a larger  $\theta$  regime wherein the dynamicity gap  $\alpha(\theta)$  is certifiably greater than one. Therefore, as our third contribution, we use  $\theta^+$  as a proxy measure for the value of dynamism and study how it is influenced by other parameters implicit in (3) within the context of integrated transit networks. We conduct two sets of computational experiments.

1. *Stylized experiments.* We conduct exhaustive stylized experiments involving a multistage version of the Steiner tree problem as an elemental abstraction for the design of integrated transit networks. Our goal is to gain qualitative insights about the effects of the underlying network topology and the demand distribution on the value of dynamism for network design. Our experiments suggest that microtransit may be valuable if demand concentrates in downtown areas as a function of their prominence yet still appears sparingly in suburban areas. Conversely, they suggest that microtransit may be unnecessary in cities with high spatial segregation of urban functions (e.g., residential, commercial, industrial), wherein there is no prominent downtown area and demand concentrates in peripheral areas.

2. *Realistic experiments.* We conduct more realistic experiments involving a multistage version of the Steiner forest problem and publicly available data from New York City. Our goal is to showcase how we can use our framework to obtain quantitative estimates of parameter combinations under which dynamism is certifiably valuable and to moreover gain qualitative insights about the effects of the transit headway  $\delta_s$ , microtransit batching interval  $\delta_d$ , and surcharge coefficient  $\eta$ , as well as the passengers' tolerance to en route detours (with respect to the shortest path in the underlying road network), on the value of dynamism. For example, under the assumptions of our model, if the transit headway is 10 minutes, microtransit is batched every 6 minutes, and passengers tolerate detours incurring up to a 25% en route travel time increase, then dynamism is certifiably valuable whenever  $\eta \leq 1.25$ . At a qualitative level, our experiments suggest that, for any fixed transit headway  $\delta_s$ , the value of dynamism increases with the microtransit batching interval  $\delta_d$ —that is, assuming microtransit passengers tolerate long wait times relative to the existing transit headway. They moreover suggest slight gains from increased passenger tolerance to en route detours, particularly for small microtransit batching intervals. These observations can be explained as follows; tolerance to detours enhances resource sharing for small  $\delta_d$ , whereas for large  $\delta_d$ , resource sharing is naturally enhanced by the number of travel demands per stage without the need of increased detours. Lastly, our experiments highlight the road segments where a static transit network might be most useful given the historical distribution of travel

demand. Informally, we rank road segments by the frequency (over the stages) with which they appear as part of the nonanticipatory installed network, thereby providing some indication about their relative importance as potential trunk lines and hence, about whether service on them should be installed statically or dynamically. We believe this can be leveraged in a subsequent step of the transit system design process: the design of the operational network. In particular, we believe frequently used road segments can be combined to obtain a good data-driven initial set of candidate lines for line planning via column generation (see Borndörfer, Grötschel, and Pfetsch 2007, Gattermann, Harbering, and Schöbel 2017).

#### 1.4. Organization

The remainder of this paper is organized as follows. In Section 2, we outline related work. In Section 3, we develop our analytical framework, including the reformulation of the multistage problem as a two-stage stochastic problem as well as Theorem 1. In Section 4, we describe our computational experiments and summarize our findings. In Section 5, we make concluding remarks.

## 2. Related Work

In Section 2.1, we outline work related to (3) as an abstract formulation. In Section 2.2, we outline related work around transit and on-demand systems.

### 2.1. Optimization Under Uncertainty

Formulation (3) resembles that of paradigms of optimization under uncertainty. In *robust optimization*, the decision maker produces a solution that is feasible for all input scenarios. In *two-stage stochastic optimization with recourse*, the decision maker leverages distributional information about the problem input to produce first-stage decisions. In the second stage, an input scenario is realized, and the decision maker produces a feasible solution by complementing their first-stage decisions with second-stage recourse actions. The objective is to minimize the expected total cost. In *two-stage adaptive optimization*, the decision maker similarly produces first-stage decisions and second-stage recourse actions, except the objective is to minimize the worst-case total cost.

The dynamicity gap measures the attainable benefit of allowing (but not requiring) dynamic components in the response strategy to a multistage optimization problem. In this sense, it continues a line of work dedicated to measuring the potential benefit of solving the “true” problem at hand, compared with solving a simplified version of it. An early example of this approach is from Birge (1982) within the context of two-stage stochastic optimization. Birge (1982) defines the *value of the stochastic solution* as the difference between the

objective value attainable with first-stage decisions obtained through the *expected value problem* (in which stochastic parameters are replaced by their expectation) and the objective value of the two-stage stochastic problem. We note that this approach necessitates that the expected value problem encodes an instance of the problem at hand, which need not be the case if the unknown parameters are required to take integer values. More recent examples are from Bertsimas and Goyal (2010), who study the relative quality of robust solutions for two-stage stochastic and adaptive optimization problems. They define the *stochasticity gap* as the ratio between the cost of an optimal solution to the robust problem and the cost of an optimal solution to the two-stage stochastic problem. Similarly, they define the *adaptability gap* as the ratio between the cost of an optimal solution to the robust problem and the cost of an optimal solution to the two-stage adaptive problem. Although the dynamicity gap is similar in form to these gaps, it differs fundamentally in that (3) need not be stochastic (where we have distributional information over input parameters) or even uncertain (where we have *uncertainty set* information over input parameters). (See Ben-Tal, El Ghaoui, and Nemirovski 2009 for a detailed explanation of the difference between the stochastic and robust optimization paradigms.) In Online Appendix A, we delineate how the dynamicity gap and stochasticity gap differ as concepts even under the conditions of our two-stage reformulation, although for the specific form of (5) we consider, the two quantities evaluate identically whenever  $\theta \geq 1$ .

Bertsimas and Goyal (2010) show, among other structural results, that the stochasticity gap is at most two when both the uncertainty set over the right-hand side of the constraints and its distribution are *symmetric* and there are no integer decision variables in the second stage. Bertsimas, Goyal, and Sun (2011) study the impact of broader geometric properties of the uncertainty set and its distribution, such as symmetry, on the relative quality of static and finitely adaptable (i.e., nearly static) solutions for multistage stochastic and adaptive optimization problems. Bertsimas, Goyal, and Lu (2015) give a tight characterization of the adaptability gap for two-stage linear packing problems under a general class of uncertainty sets, including settings in which the robust solution is optimal. Awasthi, Goyal, and Lu (2019) study another general class of uncertainty sets, for which they give both a logarithmic hardness of approximation result for the adaptive optimization problem and an approximation guarantee for the robust solution (i.e., a bound on the adaptability gap), which is furthermore tight up to a constant in certain settings.

Another related but distinct measure is the *adaptivity gap*, introduced by Dean, Goemans, and Vondrák (2008). They consider stochastic problems where solutions are built incrementally via a sequence of decisions, each of

which incrementally instantiates the problem input. Then, the adaptivity gap measures the relative benefit of adapting the sequence of decisions in response to past realizations.

Our approach differs from that in the references in that we study the behavior of the dynamicity gap  $\alpha$  and related measures as a function of the relative cost coefficient  $\theta$  and other implicit input parameters. Capturing these parameters analytically easily becomes unwieldy, and thus, we naturally take an experimental approach. The purpose of our analytical framework is to support experimentation by introducing tractable measures that act as principled proxies for the value of dynamism.

The algorithmic aspects of two-stage stochastic optimization have been studied extensively. The main challenge in two-stage stochastic optimization is that, generally speaking, an explicit representation of the underlying distribution over input scenarios may be exponentially large. Kleywegt, Shapiro, and Homem-de Mello (2002) show the convergence of the SAA method, a natural Monte Carlo simulation-based approach. Ravi and Sinha (2006) give approximation algorithms for several problems assuming polynomially many input scenarios. Immorlica et al. (2004) consider several problems where the input scenario is determined by a set of active clients and give approximation algorithms when the clients are activated independently and the cost between stages differs by a constant factor—this is the type of setting we consider in our first set of experiments. Gupta et al. (2004) provide approximation algorithms for several problems using the same proportionality assumption together with black box access to the input scenario distribution. Gupta et al. (2005) extend this framework to multistage optimization, wherein the recourse actions become increasingly expensive, and to a setting in which the relative cost coefficient depends on the input scenario. Shmoys and Swamy (2006) also give approximation algorithms under the black box model, but they do not require the costs between stages to be proportional. Their method involves a dedicated version of the ellipsoid method to solve the arising linear programming relaxations, followed by a simple rounding scheme.

## 2.2. Transit and On-Demand Systems

The design of transit systems is a mature area of research; we point the reader to Desaulniers and Hickman (2007) and Schöbel (2012) for an overview. This is also the case for the operation of on-demand mobility; we point the reader to Toth and Vigo (2014) and Alonso-Mora et al. (2017) for commonly used techniques.

The design and operation of integrated systems are an area of increasing interest. We briefly describe some representative work. Archetti, Speranza, and Weyland (2018) conduct a simulation study suggesting on-demand systems can offer user-favorable service even under the presence of direct travel and transit as alternatives. Stiglic et al.

(2018) give an optimization model for the operations of integrated systems given an existing transit network. Their base model is restricted to trips starting in a suburban area, going into the city center, and in which at most two passengers can share a vehicle (both passengers need to connect through the same transit station). They find that integrated systems can enhance mobility and increase transit ridership. Steiner and Irnich (2020) give a strategic planning optimization model to integrate on-demand services within an existing transit network. Given a travel demand realization, their model decides on existing transit lines to maintain or suspend and what areas to cover with the on-demand system along with transfer points to transit (using an approximate model for on-demand costs), as well as passenger route assignment. Liu and Ouyang (2021) give an approximate analytic model for the joint optimization of integrated systems on a square region of the plane. Their objective is to minimize the system-wide cost per passenger (e.g., accounting for costs per vehicle mile, costs per vehicle hour, road congestion). They assume a constant passenger arrival rate and uniformly distributed origin-destination pairs. They restrict the transit network to grids of uniform spacing and the on-demand system to operate within independent “local” squares, both determined as part of the optimization. They find that integrated systems generally outperform purely fixed or purely on-demand systems, except possibly on very small or very congested cities. Périer et al. (2021) consider the joint optimization of transit lines and single-occupancy vehicle routes so as to maximize a measure of welfare subject to a budget constraint. Given a collection of travel demands and pool of candidate transit lines, they develop a  $(1 - 1/e - \epsilon)$ -approximation algorithm when no transit-to-transit transfers are allowed.

### 3. Analytical Framework

We first outline some technical assumptions. As mentioned in Section 1.1, we represent integrated network configurations by their characteristic vectors in  $\{0,1\}^m$ . We assume without loss of generality that any feasible integrated network  $x, z \in \{0,1\}^m$  satisfies  $x_j + z_j \in \{0,1\}$  for all  $j \in [m]$ —because the edge costs are nonnegative, if  $x_j + z_j = 2$ , we can set  $z_j = 0$  and maintain origin-destination connectivity at no greater cost. In particular, we assume that for each  $t \in [T]$ ,  $(x, z^t) \in \mathcal{P}^t$ , and  $j \in [m]$ , we have  $x_j + z_j^t \in \{0,1\}$ , meaning  $\mathcal{P}^t \subseteq \{0,1\}^{2m}$ . Similarly, we assume that for each  $i \in [k]$ ,  $(x, z_i) \in \mathcal{P}_i$ , and  $j \in [m]$ , we have  $x_j + z_{i,j} \in \{0,1\}$ , where  $z_{i,j}$  is the  $j$ th entry of  $z_i$ , meaning  $\mathcal{P}_i \subseteq \{0,1\}^{2m}$ . We assume throughout that the feasible configurations  $\mathcal{P}^1, \mathcal{P}^2, \dots, \mathcal{P}^T$  in (3) are nonempty and finite. Recall  $p_i = \Pr_{\mathcal{D}}[I^t = I_i]$  for every  $t \in [T]$ . We assume that  $p_i > 0$  for all  $i \in [k]$ , as otherwise, we remove  $I_i$  from  $\mathcal{I}$ . Lastly, we assume that  $p_i < 1$  for all  $i \in [k]$ , as otherwise, we have a deterministic single-stage problem.

### 3.1. Two-Stage Reformulation

In this section, we show that if the input scenarios  $I^1, I^2, \dots, I^T$  are sampled i.i.d. from a probability distribution  $\mathcal{D}$  over  $\mathcal{I}$  and moreover,  $T \rightarrow \infty$ , then we can a.s. reformulate the horizon-normalized version of the multistage Problem (3), wherein we scale the objective function by  $1/T$ , as the two-stage stochastic Problem (5). As a corollary, in this case, the dynamicity gap of (3) reduces a.s. to the dynamicity gap of (5).

For fixed  $T \in \mathbb{N}$  and  $I^1, I^2, \dots, I^T$ , let  $\mathcal{I}(T) := \{I_i \in \mathcal{I} : (\exists t \in [T])[I^t = I_i]\}$  be the set of input scenarios observed at least once. Note that for any  $t, t' \in [T]$  with  $I^t = I^{t'} = I_i$ , we have  $\mathcal{P}^t = \mathcal{P}^{t'} = \mathcal{P}_i$ . Therefore, we rewrite the objective value of (3) as

$$\min_{(x, z_1, z_2, \dots, z_k) \in \mathcal{P}(\mathcal{I}(T))} \{\hat{g}_T(x, z_1, z_2, \dots, z_k)\},$$

where

$$\hat{g}_T(x, z_1, z_2, \dots, z_k) := \sum_{i=1}^k |\{t \in [T] : I^t = I_i\}| (c \cdot x + \theta \cdot c \cdot z_i)$$

aggregates the stage costs by the number of occurrences of each input scenario and

$$\begin{aligned} \mathcal{P}(\mathcal{I}(T)) &:= \{(x, z_1, z_2, \dots, z_k) \in \{0,1\}^{m \times (k+1)} \\ &: (\forall I_i \in \mathcal{I}(T))[(x, z_i) \in \mathcal{P}_i]\} \end{aligned}$$

is the set of integrated network configurations that can serve the set  $\mathcal{I}(T)$  of observed input scenarios. We normalize  $\hat{g}_T$  by the horizon  $T$  and write

$$\hat{v}_T := \min_{(x, z_1, z_2, \dots, z_k) \in \mathcal{P}(\mathcal{I}(T))} \{\hat{h}_T(x, z_1, z_2, \dots, z_k)\}, \quad (7)$$

where  $\hat{h}_T(x, z_1, z_2, \dots, z_k) := \frac{1}{T} \hat{g}_T(x, z_1, z_2, \dots, z_k)$ .

Similarly, we rewrite the objective value of (5) as

$$v := \min_{(x, z_1, z_2, \dots, z_k) \in \mathcal{P}(\mathcal{I})} \{h(x, z_1, z_2, \dots, z_k)\}, \quad (8)$$

where

$$h(x, z_1, z_2, \dots, z_k) := c \cdot x + \sum_{i=1}^k p_i \cdot \theta \cdot c \cdot z_i$$

and

$$\begin{aligned} \mathcal{P}(\mathcal{I}) &:= \{(x, z_1, z_2, \dots, z_k) \in \{0,1\}^{m \times (k+1)} \\ &: (\forall I_i \in \mathcal{I})[(x, z_i) \in \mathcal{P}_i]\}. \end{aligned}$$

We relate the horizon-normalized version of (3)–(5) through the relation between (7) and (8). First, we show that the feasible regions of (7) and (8) are equal a.s. as  $T \rightarrow \infty$ .

**Proposition 1.** *If the input scenarios  $I^1, I^2, \dots, I^T$  are sampled i.i.d. from a probability distribution  $\mathcal{D}$  over  $\mathcal{I}$ , then  $\mathcal{I}(T) = \mathcal{I}$  a.s. as  $T \rightarrow \infty$ .*

**Proof.** For any  $T \in \mathbb{N}$ , we have  $\mathcal{I}(T) \subseteq \mathcal{I}$  because for every  $t \in [T]$ ,  $I^t = I_i$  for some  $I_i \in \mathcal{I}$ . It remains to show

$\mathcal{I} \subseteq \mathcal{I}(T)$  a.s. as  $T \rightarrow \infty$ . Note that for any  $I_i \in \mathcal{I}$  and any  $T \in \mathbb{N}$ , we have  $\Pr_{\mathcal{D}}[I_i \notin \mathcal{I}(T)] = (1 - p_i)^T$ . Therefore,

$$\begin{aligned} \sum_{T=1}^{\infty} \Pr_{\mathcal{D}}[\mathcal{I} \not\subseteq \mathcal{I}(T)] &\leq \sum_{T=1}^{\infty} \sum_{i=1}^k \Pr_{\mathcal{D}}[I_i \notin \mathcal{I}(T)] \\ &= \sum_{T=1}^{\infty} \sum_{i=1}^k (1 - p_i)^T = \sum_{i=1}^k \sum_{T=1}^{\infty} (1 - p_i)^T, \end{aligned}$$

where the inequality holds by the union bound and the second equality holds by the monotone convergence theorem (recall  $p_i > 0$  for every  $i \in [k]$ ). In particular,  $\sum_{T=1}^{\infty} \Pr_{\mathcal{D}}[\mathcal{I} \not\subseteq \mathcal{I}(T)] < \infty$ , and the Borel–Cantelli lemma implies  $\Pr_{\mathcal{D}}[\{\mathcal{I} \not\subseteq \mathcal{I}(T) \text{ infinitely often}\}] = 0$ .  $\square$

As a consequence, we have that, under the conditions of Proposition 1,

$$\hat{v}_T = \min_{(x, z_1, z_2, \dots, z_k) \in \mathcal{P}(T)} \{\hat{h}_T(x, z_1, z_2, \dots, z_k)\} \quad (9)$$

a.s. as  $T \rightarrow \infty$ . Note that the only difference between Expressions (7) and (9) is in the feasible set.

For ease of notation, let  $\mathcal{P}$  denote  $\mathcal{P}(\mathcal{I})$ . For any  $\epsilon > 0$ , let  $\mathcal{P}^\epsilon$  be the nonempty set of solutions to (8) within an  $\epsilon$ -additive term from optimality and  $\hat{\mathcal{P}}_T^\epsilon$  be the nonempty set of solutions to (7) within an  $\epsilon$ -additive term from optimality. In particular,  $\mathcal{P}^0$  is the set of optimal solutions to (8), and  $\hat{\mathcal{P}}_T^0$  is the set of optimal solutions to (7). We use Proposition 1 to show the objective value, and the sets of approximately optimal solutions to (7) converge a.s. as  $T \rightarrow \infty$  to their counterparts in (8). The remainder of our argument largely replicates, within the context of our formulation, the argument of Kleywegt, Shapiro, and Homem-de Mello (2002) for the convergence of the SAA method. We include a proof in Online Appendix B for completeness.

**Theorem 2.** *If the input scenarios  $I^1, I^2, \dots, I^T$  are sampled i.i.d. from a probability distribution  $\mathcal{D}$  over  $\mathcal{I}$ , then*

1.  $\hat{v}_T \rightarrow v$  a.s. as  $T \rightarrow \infty$  and
2. for any  $\epsilon \geq 0$ ,  $\hat{\mathcal{P}}_T^\epsilon \subseteq \mathcal{P}^\epsilon$  a.s. as  $T \rightarrow \infty$ .

This result has two immediate yet key consequences, one operational and one analytical.

**Corollary 1.** *Under the conditions of Theorem 2, we can a.s. extend an optimal solution  $(x, z_1, z_2, \dots, z_k)$  to (5) to an optimal solution to (3) by setting  $x$  as an ex ante static response and following the natural ex post dynamic policy for every stage  $t \in \mathbb{N}$ : if  $I^t = I_i$ , we respond with  $z^t = z_i$  so that  $(x, z^t) = (x, z_i) \in \mathcal{P}_i = \mathcal{P}^t$ . This holds because normalizing the objective function of (3) does not change the set of optimal solutions.*

**Corollary 2.** *Under the conditions of Theorem 2, the dynamicity gap  $\alpha$  of (3) reduces a.s. to the dynamicity gap of (5).*

Namely, if  $\text{OPT}_{(5)}$  denotes the cost of an optimal solution to (5) and  $\text{OPT}_{(5)}^\Sigma$  denotes the cost of an optimal static solution to (5), that is one in which we additionally require  $z_1 = z_2 = \dots = z_k = \mathbf{0}$ , then

$$\alpha = \frac{\text{OPT}_{(5)}^\Sigma}{\text{OPT}_{(5)}} \quad \text{a.s.}$$

In other words, we may study the limit behavior of  $\alpha$  with a framework built around (5), which is more suitable for parametric analysis. In light of this, for the remainder of this work we assume the conditions of Theorem 2 hold, drop the a.s. notation, and let  $\text{OPT}$  and  $\text{OPT}^\Sigma$  be with respect to (5) instead of the original Problem (3).

### 3.2. Estimating the Value of Dynamism

For ease of notation, let  $\mathcal{P} = \mathcal{P}(\mathcal{I})$ , and let  $\Pi(\theta)$  refer to (5) explicitly parametrized by the relative cost coefficient  $\theta$  and implicitly parametrized by the distribution  $\mathcal{D}$  over  $\mathcal{I}$ . Let  $h_\theta$  and  $\text{OPT}(\theta)$  refer to the objective function and objective value of  $\Pi(\theta)$ , respectively. Similarly, let  $\Pi^\Sigma(\theta)$  refer to the static version of (5) explicitly parametrized by  $\theta$  and  $\text{OPT}^\Sigma(\theta)$  refer to its objective value. Recall  $\mathbf{0}$  refers to the zero vector. We begin by showing a monotonicity property with respect to the relative cost coefficient.

**Lemma 1.** *For  $0 < \theta_1 \leq \theta_2 < \infty$  and  $(x, z_1, z_2, \dots, z_k) \in \mathcal{P}$ , we have  $h_{\theta_1}(x, z_1, z_2, \dots, z_k) \leq h_{\theta_2}(x, z_1, z_2, \dots, z_k)$ . If, furthermore,  $(x, z_1, z_2, \dots, z_k) \in \mathcal{P}^\Sigma$ , then the inequality holds at equality.*

**Proof.** Note that  $h_{\theta_1}(x, z_1, z_2, \dots, z_k) = c \cdot x + \sum_{i=1}^k p_i \cdot \theta_1 \cdot c \cdot z_i \leq \sum_{i=1}^k p_i \cdot \theta_2 \cdot c \cdot z_i = h_{\theta_2}(x, z_1, z_2, \dots, z_k)$ , where the inequality holds because  $c \geq \mathbf{0}$  and because  $(x, z_1, z_2, \dots, z_k) \in \mathcal{P}$  implies  $z_1, z_2, \dots, z_k \geq \mathbf{0}$ . If  $(x, z_1, z_2, \dots, z_k) \in \mathcal{P}^\Sigma$ , then  $z_1 = z_2 = \dots = z_k = \mathbf{0}$ , and we have equality.  $\square$

**Corollary 3.** *For  $0 < \theta_1 \leq \theta_2 < \infty$ , we have (i)  $\text{OPT}(\theta_1) \leq \text{OPT}(\theta_2)$ , (ii)  $\text{OPT}^\Sigma(\theta_1) = \text{OPT}^\Sigma(\theta_2)$ , and (iii)  $\alpha(\theta_2) \leq \alpha(\theta_1)$ .*

**Proof.** Let  $(x, z_1, z_2, \dots, z_k) \in \mathcal{P}$  be an optimal solution to  $\Pi(\theta_2)$ , and note that it is feasible for  $\Pi(\theta_1)$ . Then,  $\text{OPT}(\theta_1) \leq h_{\theta_1}(x, z_1, z_2, \dots, z_k) \leq h_{\theta_2}(x, z_1, z_2, \dots, z_k) = \text{OPT}(\theta_2)$ , where the second inequality holds by Lemma 1.  $\text{OPT}^\Sigma(\theta_1) = \text{OPT}^\Sigma(\theta_2)$  holds because  $z_1 = z_2 = \dots = z_k = \mathbf{0}$  implies  $(x, z_1, z_2, \dots, z_k) \in \mathcal{P}^\Sigma$  is an optimal solution to  $\Pi^\Sigma(\theta_2)$  if and only if it is an optimal solution to  $\Pi^\Sigma(\theta_1)$ . Together, these facts imply  $\alpha(\theta_2) = \text{OPT}^\Sigma(\theta_2)/\text{OPT}(\theta_2) = \text{OPT}^\Sigma(\theta_1)/\text{OPT}(\theta_2) \leq \text{OPT}^\Sigma(\theta_1)/\text{OPT}(\theta_1) = \alpha(\theta_1)$ .  $\square$

Next, we provide a parametric upper bound on  $\text{OPT}(\theta)$  that takes as input any collection  $\chi_1, \chi_2, \dots, \chi_k \in \{0, 1\}^m$  of nonanticipatory feasible solutions to input scenarios  $I_1, I_2, \dots, I_k$ .

**Theorem 3.** Let  $\chi_1, \chi_2, \dots, \chi_k \in \{0,1\}^m$  be such that  $(\chi_i, \mathbf{0}) \in \mathcal{P}_i$  for all  $i \in [k]$ . Then, for any  $\theta > 0$ , we have

$$\text{OPT}(\theta) \leq \sum_{j=1}^m c_j \cdot \min \left\{ 1, \theta \cdot \left( \sum_{i \in [k]: \chi_{i,j}=1} p_i \right) \right\},$$

where  $c_j$  is the  $j$ th entry of  $\mathbf{c} \in \mathbb{R}_{\geq 0}^m$  in (5) and  $\chi_{i,j}$  is the  $j$ th entry of the  $i$ th vector  $\chi_i$ . If in addition,  $\chi_i = \mathbf{x} + \mathbf{z}_i$  for all  $i \in [k]$ , where  $(\mathbf{x}, \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k) \in \mathcal{P}$  is an optimal solution to  $\Pi(\theta)$ , then the inequality holds at equality.

**Proof.** Let  $S = \{j \in [m] : \min\{1, \theta \cdot (\sum_{i \in [k]: \chi_{i,j}=1} p_i)\} = 1\}$ . Define  $\xi \in \{0,1\}^m$ , where  $\xi_j = 1$  if and only if  $j \in S$ . Similarly, for each  $i \in [k]$ , define  $\zeta_i \in \{0,1\}^m$ , where  $\zeta_{i,j} = 1$  if and only if  $\chi_{i,j} = 1$  yet  $j \notin S$ . By construction,  $(\xi, \zeta_i) \in \mathcal{P}_i$  for each  $i \in [k]$ . Therefore,  $(\xi, \zeta_1, \zeta_2, \dots, \zeta_k) \in \mathcal{P}$  and

$$\begin{aligned} \text{OPT}(\theta) &\leq \sum_{j=1}^m \left( c_j \cdot \xi_j + \sum_{i=1}^k p_i \cdot \theta \cdot c_j \cdot \zeta_{i,j} \right) \\ &= \sum_{j=1}^m c_j \cdot \left( \xi_j + \theta \cdot \sum_{i=1}^k p_i \cdot \zeta_{i,j} \right) \\ &= \sum_{j=1}^m c_j \cdot \min \left\{ 1, \theta \cdot \left( \sum_{i \in [k]: \chi_{i,j}=1} p_i \right) \right\}. \end{aligned}$$

Now, suppose  $(\mathbf{x}, \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k) \in \mathcal{P}$  is an optimal solution to  $\Pi(\theta)$  and  $\chi_i = \mathbf{x} + \mathbf{z}_i$  for all  $i \in [k]$ . Note that  $\chi_i \in \{0,1\}^m$  for all  $i \in [k]$  by assumption (the first paragraph of Section 3). First, take any decision  $j \in [m]$  with  $x_j = 1$ , meaning it is chosen statically. Then,  $j \in S$ , and its cost  $c_j$  is accounted for exactly. Conversely, take any decision  $j \in [m]$  with  $x_j = 0$ , meaning it is chosen dynamically. This decision is paid for with cost multiplied by a factor of  $\theta$  with probability  $\sum_{i \in [k]: \chi_{i,j}=1} p_i$ . Suppose  $\theta \cdot \sum_{i \in [k]: \chi_{i,j}=1} p_i > 1$ . Then, setting  $x_j = 1$  and  $z_{i,j} = 0$  for all  $i \in [k]$  maintains feasibility at strictly lower cost, contradicting optimality. Therefore, its expected cost  $c_j \cdot \theta \cdot (\sum_{i \in [k]: \chi_{i,j}=1} p_i)$  is accounted for exactly.  $\square$

**Remark 2.** If  $\theta = 1$  and more generally, if  $0 < \theta \leq 1$ , there is no benefit in making static decisions. We can see this from the fact that, for any such  $\theta$ , the following holds:

$$\sum_{j=1}^m c_j \cdot \min \left\{ 1, \theta \cdot \left( \sum_{i \in [k]: \chi_{i,j}=1} p_i \right) \right\} = \sum_{j=1}^m c_j \cdot \theta \cdot \left( \sum_{i \in [k]: \chi_{i,j}=1} p_i \right).$$

Therefore, to compute the objective value of (5) with  $\theta = 1$ , it is sufficient to, individually for each input scenario  $I_i \in \mathcal{I}$ , compute an optimal nonanticipatory solution with characteristic vector  $\chi_i$  (i.e., a minimum cost  $\chi_i$  such that  $(\chi_i, \mathbf{0}) \in \mathcal{P}_i$ ) and aggregate their costs weighted by the probabilities  $p_i$  for  $i \in [k]$ . This is *relatively* tractable in the sense that it does not involve solving a two-stage stochastic problem, but a collection

of independent, deterministic single-stage optimization problems (e.g., although the single-stage problems may remain NP-hard and the collection may be large, their computation can be parallelized). In the case of the Steiner forest problem, we refer the reader to Ljubić (2021) for a survey of state-of-the-art solution techniques.

We moreover use the following technical result, which we prove in Online Appendix B.

**Lemma 2.** Let  $\chi_1, \chi_2, \dots, \chi_k \in \{0,1\}^m$  be such that  $(\chi_i, \mathbf{0}) \in \mathcal{P}_i$  for all  $i \in [k]$ . Then, for  $0 < \theta_1 \leq \theta_2 < \infty$ , we have

$$\begin{aligned} &\sum_{j=1}^m c_j \cdot \min \left\{ 1, \theta_2 \cdot \left( \sum_{i \in [k]: \chi_{i,j}=1} p_i \right) \right\} \\ &\leq \frac{\theta_2}{\theta_1} \cdot \sum_{j=1}^m c_j \cdot \min \left\{ 1, \theta_1 \cdot \left( \sum_{i \in [k]: \chi_{i,j}=1} p_i \right) \right\}. \end{aligned}$$

In what follows, we use the results obtained thus far to show that if we evaluate  $\alpha := \alpha(\theta)$  at a finite number  $\kappa$  of points  $1 = \theta_1 < \theta_2 < \dots < \theta_\kappa < \infty$ , we can produce estimates  $\hat{\alpha}^-, \hat{\alpha}^+ : \mathbb{R}_{\geq 1} \rightarrow \mathbb{R}_{\geq 1}$  such that, for any  $\theta \in \mathbb{R}_{\geq 1}$ , we have  $\hat{\alpha}^-(\theta) \leq \alpha(\theta) \leq \hat{\alpha}^+(\theta)$ . The estimates  $\hat{\alpha}^-, \hat{\alpha}^+$  are a concatenation of local estimates, similar to a step function, interpolating between the evaluations  $\alpha(\theta_1), \alpha(\theta_2), \dots, \alpha(\theta_\kappa)$ . Let  $\mathbb{1}$  denote the indicator function.

**Theorem 4.** Let  $1 = \theta_1 < \theta_2 < \dots < \theta_\kappa < \theta_{\kappa+1} = \infty$ . Let  $\hat{\alpha}^- : \mathbb{R}_{\geq 1} \rightarrow \mathbb{R}_{\geq 1}$ , where

$$\hat{\alpha}^-(\theta) = \sum_{\ell=1}^{\kappa} \max\{\theta_\ell \cdot \alpha(\theta_\ell)/\theta, \alpha(\theta_{\ell+1})\} \cdot \mathbb{1}_{\{\theta_\ell \leq \theta < \theta_{\ell+1}\}},$$

and let  $\hat{\alpha}^+ : \mathbb{R}_{\geq 1} \rightarrow \mathbb{R}_{\geq 1}$ , where

$$\hat{\alpha}^+(\theta) = \sum_{\ell=1}^{\kappa} \alpha(\theta_\ell) \cdot \mathbb{1}_{\{\theta_\ell \leq \theta < \theta_{\ell+1}\}}.$$

Then, for any  $\theta \geq 1$ , we have

$$\hat{\alpha}^-(\theta) \leq \alpha(\theta) \leq \hat{\alpha}^+(\theta).$$

**Proof.** Let  $\theta \geq 1$ , and note that  $\theta_\ell \leq \theta < \theta_{\ell+1}$  holds for exactly one  $\ell \in [\kappa]$ —pick such  $\ell$ . If  $\max\{\theta_\ell \cdot \alpha(\theta_\ell)/\theta, \alpha(\theta_{\ell+1})\} = \alpha(\theta_{\ell+1})$ , then  $\hat{\alpha}^-(\theta) = \alpha(\theta_{\ell+1}) \leq \alpha(\theta)$ , where the inequality holds by Corollary 3 and  $\theta < \theta_{\ell+1}$ . Otherwise,  $\max\{\theta_\ell \cdot \alpha(\theta_\ell)/\theta, \alpha(\theta_{\ell+1})\} = \theta_\ell \cdot \alpha(\theta_\ell)/\theta$ , and we have

$$\begin{aligned} \hat{\alpha}^-(\theta) &= \frac{\theta_\ell \cdot \alpha(\theta_\ell)}{\theta} = \frac{\theta_\ell}{\theta} \cdot \frac{\text{OPT}^\Sigma(\theta_\ell)}{\text{OPT}(\theta_\ell)} = \frac{\theta_\ell}{\theta} \cdot \frac{\text{OPT}^\Sigma(\theta)}{\text{OPT}(\theta_\ell)} \\ &= \frac{\theta}{\theta_\ell} \cdot \frac{\text{OPT}^\Sigma(\theta)}{\sum_{j=1}^m c_j \cdot \min\{1, \theta_\ell \cdot (\sum_{i \in [k]: \chi_{i,j}=1} p_i)\}} \\ &\leq \frac{\text{OPT}^\Sigma(\theta)}{\sum_{j=1}^m c_j \cdot \min\{1, \theta \cdot (\sum_{i \in [k]: \chi_{i,j}=1} p_i)\}} \leq \frac{\text{OPT}^\Sigma(\theta)}{\text{OPT}(\theta)} \\ &= \alpha(\theta), \end{aligned}$$

where the third equality holds by Corollary 3,  $\chi_1, \chi_2, \dots, \chi_k \in \{0,1\}^m$  are such that  $(\chi_i, \mathbf{0}) \in \mathcal{P}_i$  for all  $i \in [k]$  and

Theorem 3 holds at equality for  $\theta_\ell$ , the first inequality holds by Lemma 2, and the second inequality holds by Theorem 3. Lastly,  $\hat{\alpha}^+(\theta) = \alpha(\theta_\ell) \geq \alpha(\theta)$ , where the inequality holds by Corollary 3 and  $\theta \geq \theta_\ell$ .  $\square$

As we state next, the bounds can be made arbitrarily tight by systematically evaluating  $\alpha$  at sufficiently many points. We prove the following result in Online Appendix B.

**Theorem 5.** Let  $\epsilon > 0$ . Let  $\theta_1 = 1$  and  $\theta_{\ell+1} = (1 + \epsilon) \cdot \theta_\ell$  for all  $i \in \mathbb{N}$ . Let  $\kappa = \arg \min_{\ell \in \mathbb{N}} \alpha(\theta_\ell) < 1 + \epsilon$ . Then, Theorem 4, given  $\theta_1, \theta_2, \dots, \theta_\kappa$ , yields  $\hat{\alpha}^-$  and  $\hat{\alpha}^+$  such that for any  $\theta \geq 1$ , we have  $\hat{\alpha}^+(\theta) < (1 + \epsilon) \cdot \hat{\alpha}^-(\theta)$ .

In light of the monotonicity result in Corollary 3, we define  $\theta^* := \arg \min_{\theta > 0} \{\alpha(\theta) = 1\}$  as the critical relative cost coefficient after which dynamism is no longer valuable. We then obtain the following corollary to Theorem 5.

**Corollary 4.** Let  $\kappa$  be as in the statement of Theorem 5. Then,  $\theta_\kappa \leq \theta^*$ .

Theorem 5 gives a systematic way of estimating the dynamicity gap as a function of the relative cost coefficient  $\theta$  to any arbitrary precision by evaluating it at a finite set of points. However, evaluating it at any given point generally involves solving the two-stage stochastic Problem (5), which may be intractable. As an alternative, Remark 2 points out that evaluating the dynamicity gap at  $\theta_1 = 1$  is *relatively* tractable and that this is sufficient to obtain an (albeit weaker) lower bound on the critical relative cost coefficient  $\theta^*$ . This is Theorem 1, which we finally prove.

**Proof of Theorem 1.** For  $\theta \geq 1$ , the inequality  $\alpha(\theta) \geq \max\{\theta^+/ \theta, 1\}$  follows from the special case of Theorem 4 with  $\kappa = 1$ . For  $0 < \theta < 1$ , we use the equality in Remark 2. The inequality  $\theta^+ \leq \theta^*$  follows from  $\max\{\theta^+/ \theta, 1\} > 1$  as long as  $\theta^+ / \theta > 1$ : that is, for any  $0 < \theta < \theta^+$ .  $\square$

## 4. Computational Experiments

We now use the framework developed in Section 3 to perform computational experiments around (abstractions of) the design of integrated transit networks. Recall the advantage of the bound in Theorem 1 over computing the dynamicity gap exactly is that it is relatively tractable, as stated in Remark 2. In particular, using  $\theta^+ := \alpha(1)$  as a proxy for measure for the value of dynamism allows us to swiftly run comprehensive experiments. Because  $\theta^+ \leq \theta^*$ , where  $\theta^* := \arg \min_{\theta > 0} \{\alpha(\theta) = 1\}$  is the critical relative cost coefficient, we can certify that the dynamicity gap is greater than one whenever  $0 < \theta < \theta^+$ . Therefore, our goal is to understand how various input parameters implicit in (3), equivalently (5) under the conditions of Theorem 1, influence  $\theta^+$ . We conduct two sets of experiments.

1. In Section 4.1, we conduct stylized experiments involving a multistage version of the Steiner tree problem as an elemental abstraction. Our goal is to gain qualitative insights about the effects of the network topology and demand distribution on the value of dynamism for network design.

2. In Section 4.2, we conduct more realistic experiments involving a multistage version of the Steiner forest problem and publicly available data from New York City. Our goal is to showcase how we can use our framework to obtain quantitative estimates of parameter combinations under which dynamism is certifiably valuable and to moreover gain qualitative insights about the effects of the transit headway  $\delta_s$ , microtransit batching interval  $\delta_d$ , and surcharge coefficient  $\eta$  (see Section 1.2 for descriptions), as well as the passengers' tolerance to en route detours, on the value of dynamism.

### 4.1. Stylized Experiments

**4.1.1. Setup.** We consider a multistage version of the Steiner tree problem as the most elemental abstraction for the design of integrated transit networks. In the Steiner tree problem, we are given a graph  $G = (V, E)$  and a set  $I \subseteq V$  of terminals. The problem is to find a minimum cost set of edges connecting every pair in  $I$ . The possible input scenarios  $\mathcal{I}$  correspond to the possible terminal sets. We point the reader to Online Appendix C for an integer linear programming formulation of this problem. In its multistage version, the connectivity requirements of each stage are met through a combination of edges  $X \subseteq E$  picked statically and edges  $Z^t \subseteq E$  picked dynamically.

We consider all 995 unweighted connected simple graphs on  $2 \leq n \leq 7$  nodes. Such a list has been compiled by Read and Wilson (1998) and is retrievable in `python` through the `networkx` package of Hagberg, Swart, and Chult (2008). These are admittedly small graphs, but this is what enables us to run exhaustive experiments—the number of such graphs grows exponentially in  $n$ , and the Steiner tree problem is well known to be NP-hard.

We consider probability distributions  $\mathcal{D}$  over  $\mathcal{I}$  arising from independent Bernoulli trials on the nodes. For each  $u \in V$ , let  $q_u := \Pr[u \in I]$  be the probability that  $u$  is a terminal. This yields a probability distribution  $\mathcal{D}$  over  $\mathcal{I}$  with

$$p_i = \Pr_{\mathcal{D}}[I = I_i] = \prod_{u \in I_i} q_u \prod_{u \in V \setminus I_i} (1 - q_u) \quad (10)$$

for all  $i \in [k]$ . Different choice of parameters  $0 \leq q_u \leq 1$  for  $u \in V$  yields different distributions. We test three different rules to generate these parameters.

1. If  $q_u = 1/2$  for each  $u \in V$ , then Equation (10) yields  $p_i = 1/2^n$  for each  $i \in [k]$ . This is the uniform distribution over  $\mathcal{I}$ , which we denote by  $\mathcal{U}$ .

For  $u, v \in V$ , let  $\ell(u, v)$  be the shortest-path length between  $u$  and  $v$  (e.g., with respect to costs  $c : E \rightarrow \mathbb{R}_{\geq 0}$ ).

The *closeness centrality* of a node  $u \in V$ , denoted by  $C(u)$ , is given by  $C(u) := (n-1)/\sum_{v \in V} \ell(u,v)$ . This centrality measure, first introduced by Bavelas (1950), characterizes a node as “central” if it is close to all other nodes. We say a node is “peripheral” if it is not central. Intuitively, if  $G = (V, E)$  represents a road network, nodes in a prominent downtown area are “central,” whereas nodes in a suburban area are “peripheral.”

2. If  $q_u = C(u) \cdot n/(n-1)$  for each  $u \in V$ , then Equation (10) yields a distribution biased toward terminal sets consisting of “central” nodes yet still supported on “peripheral” nodes. We denote this distribution by  $\mathcal{D}^{+\text{cent}}$ . Intuitively, we think of  $\mathcal{D}^{+\text{cent}}$  as a distribution where demand concentrates in downtown areas as a function of how “central” they are but still arises sparingly elsewhere.

3. Conversely, if  $q_u = 1 - C(u) \cdot n/(n-1)$  for each  $u \in V$ , then Equation (10) yields a distribution biased toward terminal sets consisting of “peripheral” nodes yet still supported on “central” nodes. We denote this distribution by  $\mathcal{D}^{-\text{cent}}$ . Intuitively, we think of  $\mathcal{D}^{-\text{cent}}$  as a distribution where demand concentrates in suburban areas as a function of how “peripheral” they are but still arises sparingly elsewhere.

Lastly, to study the effects of the underlying topology, we characterize graphs through three different measures of connectivity.

1. The *average degree* of a graph  $G$ , denoted by  $\bar{d}(G)$ , is the average degree over all nodes. Formally,  $\bar{d}(G) := \sum_{u \in V} |\mathcal{N}(u)| / n = 2m/n$ , where  $\mathcal{N}(u) \subseteq V$  is the set of neighbors of  $u \in V$ .

2. The *average node connectivity* of a graph  $G$ , first introduced by Beineke, Oellermann, and Pippert (2002) and denoted by  $\bar{\kappa}(G)$ , is the average, over all pairs of nodes, of the maximum number of internally node-disjoint paths connecting them. Formally,  $\bar{\kappa}(G) := \sum_{u,v \in V: u \neq v} \kappa_G(u,v) / \binom{n}{2}$ , where  $\kappa_G(u,v)$  is the maximum number of internally node-disjoint paths connecting  $u$  and  $v$  in  $G$ .

3. Let  $L_G$  be the Laplacian matrix of a graph  $G$  and  $\lambda_1, \lambda_2, \dots, \lambda_n$  be its eigenvalues, counting multiplicities, in decreasing order. The *algebraic connectivity* of  $G$ , denoted by  $a(G)$ , is the second smallest eigenvalue of  $L_G$  counting multiplicities. That is,  $a(G) := \lambda_{n-1}$ . It holds that  $a(G) > 0$  if and only if  $G$  is connected. It, moreover, holds that  $a(G) \leq n$ , with the inequality holding at equality if and only if  $G$  is the complete graph on  $n$  nodes.

We note that these connectivity measures are related in subtle ways. For example, Das (2018) has shown that  $a(G) - \bar{d}(G) \geq 4 - n - 4/n$  and has moreover characterized when the inequality holds at equality. See Newman (2018) for a comprehensive description of these measures.

**4.1.2. Results.** For very small graphs, we derive analytic expressions for  $\alpha(\theta)$ . There is one connected graph with  $n = 2$ , namely the complete graph  $K_2 = (V, E)$  with

$V = \{1, 2\}$  and  $E = \{\{1, 2\}\}$ . The collection of possible terminal sets is  $\mathcal{I} = \{I_1, I_2, I_3, I_4\}$ , where  $I_1 = \emptyset$ ,  $I_2 = \{1\}$ ,  $I_3 = \{2\}$ , and  $I_4 = \{1, 2\}$ . Similarly, there are two connected graphs with  $n = 3$ , namely the path  $P_3 = (V, E)$  with  $V = \{1, 2, 3\}$  and  $E = \{\{1, 2\}, \{2, 3\}\}$  and the complete graph  $K_3 = (V, E)$  with  $V = \{1, 2, 3\}$  and  $E = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$ . The collection of possible terminal sets is  $\mathcal{I} = \{I_1, I_2, \dots, I_8\}$ , where  $I_1 = \emptyset$ ,  $I_2 = \{1\}$ ,  $I_3 = \{2\}$ ,  $I_4 = \{3\}$ ,  $I_5 = \{1, 2\}$ ,  $I_6 = \{2, 3\}$ ,  $I_7 = \{1, 3\}$ , and  $I_8 = \{1, 2, 3\}$ . We summarize our expressions in Table 2 and point the reader to Online Appendix C for derivations.

Table 2 showcases how we may fix a set of characteristic vectors forming an input to Theorem 3 to obtain analytic estimates expressions for  $\alpha(\theta)$ . For any fixed  $\theta$  and distribution  $\mathcal{D}$ , we may produce exact estimates by exhausting all possible characteristic vectors in Theorem 3. Indeed, the expressions in Table 2 are exact for  $\theta = 1$  and the uniform distribution  $\mathcal{U}$ , in which case  $\theta^+ = 4$  on  $K_2$ ,  $\theta^+ = 2\sqrt{6}$  on  $P_3$ , and  $\theta^+ = 3.2$  on  $K_3$  (see Online Appendix C).

For larger graphs, we transition to a computational study. In Figure 1, we present scatterplots of  $\theta^+$  as a function of graph connectivity measures for graphs on  $n = 7$  nodes and different distributions  $\mathcal{D}$  over  $\mathcal{I}$ . We point to Online Appendix C for similar figures for graphs on  $4 \leq n \leq 6$  nodes and error curves (in gap form) of the bound in Theorem 1 relative to the exact value of  $\alpha(\theta)$  for  $\theta > 1$ .

Figure 1 shows medium to strong correlation between graph connectivity and the value of dynamism. For the distribution  $\mathcal{D}^{+\text{cent}}$  biased toward “central” nodes, dynamism tends to be more valuable on sparsely connected graphs. This can be explained as follows; in well-connected graphs, a large proportion of nodes are highly “central,” in which case a large proportion of nodes are “almost always” (in a colloquial sense of the term) terminals under  $\mathcal{D}^{+\text{cent}}$ . Given that road networks are far from being complete graphs, this supports the notion that microtransit may be valuable if demand concentrates in downtown areas as a function of their prominence yet still appears sparingly in suburban areas. Conversely, for the uniform distribution  $\mathcal{U}$  and the distribution  $\mathcal{D}^{-\text{cent}}$  biased toward “peripheral” nodes, dynamism tends to be more valuable on well-connected graphs. This can be explained as follows; in sparsely connected graphs, a large proportion of edges are utilized under most input scenarios (especially if terminals are likely to be on the “periphery,” as is the case for both  $\mathcal{U}$  and  $\mathcal{D}^{-\text{cent}}$ ), rendering dynamism unnecessary. This suggests that in cities with high spatial segregation of urban functions (e.g., residential, commercial, industrial), wherein there is no prominent “downtown” and demand concentrates in “peripheral” areas, microtransit may be unnecessary.

## 4.2. Realistic Experiments

**4.2.1. Setup.** We now consider a more realistic abstraction for the design of integrated transit networks—the

**Table 2.** Analytic Expressions for  $\alpha(\theta)$  for Our Multistage Steiner Tree Problem on Graphs  $K_2$ ,  $P_3$ , and  $K_3$  and a Generic Distribution  $\mathcal{D}$  over  $\mathcal{I}$ 

|                  | $K_2$                                                | $P_3$                                                                                               | $K_3$                                                                                                                    |
|------------------|------------------------------------------------------|-----------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| $\alpha(\theta)$ | $= \max\left\{1, \frac{1}{p_4 \cdot \theta}\right\}$ | $= \frac{2}{\min\{1, (p_5 + p_7 + p_8) \cdot \theta\} + \min\{1, (p_6 + p_7 + p_8) \cdot \theta\}}$ | $\geq \frac{2}{\min\{1, (p_5 + p_8) \cdot \theta\} + \min\{1, (p_6 + p_8) \cdot \theta\} + \min\{1, p_7 \cdot \theta\}}$ |

Steiner forest problem and its multistage version, as described in Sections 1.1 and 1.2, respectively. We obtain a crowdsourced graph  $G = (V, E)$  representing the Manhattan road network through the osmnx package of Boeing (2017). The nodes  $V$  represent intersections, and the edges  $E$  represent road segments weighted by length  $c : E \rightarrow \mathbb{R}_{\geq 0}$  in meters. We treat length as a proxy for both operating cost and travel time. We represent travel demand with taxi trip records from June 2016, available for download from the NYC Taxi and Limousine Commission (2021).

To bring this abstraction closer to reality and to reduce the number of variables needed in our integer linear programming formulations via multicommodity flows, we impose detour constraints on pairwise connectivity. Namely, if  $I^t$  is the travel demand during the  $t$ th stage,  $(u, v) \in I^t$ , and the shortest-path length between  $u$  and  $v$  in  $G$  with respect to costs  $c : E \rightarrow \mathbb{R}_{\geq 0}$  is  $\ell(u, v)$ , then the shortest-path length between  $u$  and  $v$  in the integrated network during the  $t$ th stage must be less than or equal to  $\rho \cdot \ell(u, v)$  for some allowable detour factor  $\rho \geq 1$ .

We preprocess the raw data as follows. For tractability purposes, we focus on a subset of Manhattan roughly south of the Flatiron Building and prune  $G$  accordingly. We turn  $G$  into a simple undirected graph after deleting any self-loops, bidirecting every edge, and removing any duplicates. We delete any nodes of unit degree and contract any edges shorter than 30 meters. To account for lower speed limits and lower traffic light priority on streets (roughly traversing  $G$  from east to west) compared with avenues (roughly traversing  $G$  from south to north), we augment the length of edges labeled as “residential” or as “unclassified” by a factor of 1.5—road class labels are part of the crowdsourced data obtained via osmnx. We focus on trips starting on weekdays between 7:00 a.m. and 8:00 a.m. We match the geographical start of a trip (encoded by latitude and longitude) to the nearest node in  $G$  and discard the trip if the Euclidean distance exceeds 250 meters. We do the same with the geographical end of a trip. We discard any trips shorter than 1,000 meters as these are unlikely to take place in transit.

Recall from Section 1.2 that  $\delta$  is the stage duration, that  $\delta_d$  is the microtransit batching interval, and that we match  $\delta := \delta_d$ . Given any fixed  $\delta_d$ , we distribute the trips into bins of uniform duration  $\delta := \delta_d$  based on their start timestamp. The trips assigned to each bin constitute the input scenario of each stage—we assume these satisfy the i.i.d. condition of Theorem 2. Because the data are

finite, the number of stages depends on the choice of  $\delta_d$ . For example, because there were 22 weekdays in July 2016, we have  $(60/1) \cdot 22 = 1,320$  stages for  $\delta_d = 1$  minute but only  $(60/15) \cdot 22 = 88$  stages for  $\delta_d = 15$  minutes. Because trips correspond to the same hourly interval on weekdays, for binning purposes we focus on  $\delta_d$ , a divisor of 60 minutes.

We aim to use  $\theta^+ := \alpha(1)$  as a proxy for the critical relative cost coefficient  $\theta^* := \arg \min_{\theta > 0} \{\alpha(\theta) = 1\}$ . We compute OPT(1) by solving each stage independently in a nonanticipatory manner, as justified in Remark 2. We do so with a 5% optimality tolerance and a time-out of  $\max\{10, \delta_d\}$  minutes. If there are  $T \in \mathbb{N}$  stages, we let  $p_i = |\{t \in [T] : I^t = I_i\}|/T$  for all  $i \in [k]$ . However, for our scale of  $G$ , computing OPT $^\Sigma$  remains challenging as the Steiner forest problem is NP-hard. Therefore, as a polynomial time solvable approximation, we use the length of a minimum spanning tree of  $G$ . We justify this as follows.

**Proposition 2.** Let  $\hat{\theta}^+ := \ell(\text{MST}(G))/\text{OPT}(1)$ , where  $\ell(\text{MST}(G))$  denotes the length of a minimum length spanning tree of  $G$ . If the input scenarios  $I^1, I^2, \dots, I^T$  are sampled i.i.d. from a probability distribution  $\mathcal{D}$  over  $\mathcal{I}$  with  $p_i > 0$  for all  $i \in [k]$ , then a.s. as  $T \rightarrow \infty$ , we have  $\theta^+ \geq \hat{\theta}^+$ .

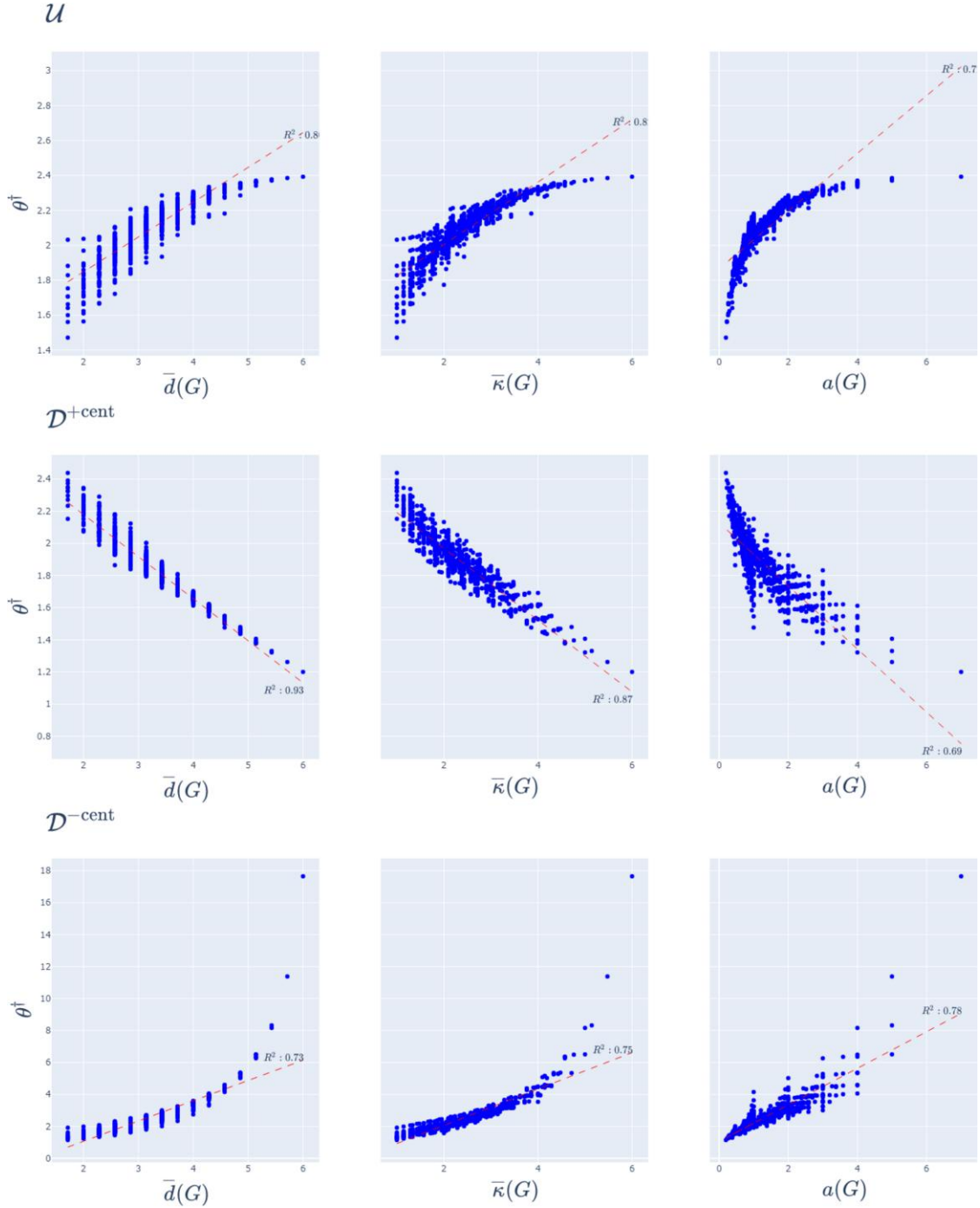
**Proof.** By Proposition 1,  $\mathcal{I}(T) = \mathcal{I}$  a.s. as  $T \rightarrow \infty$ . In particular,  $V \times V \in \mathcal{I}(T)$  a.s. as  $T \rightarrow \infty$ , which requires any optimal static solution to contain a spanning tree. By nonnegativity of the edge lengths, any circuit-creating edges (which may exist because of the detour constraints) can be removed at no additional cost. Then,  $\theta^+ := \text{OPT}^\Sigma(1)/\text{OPT}(1) \geq \ell(\text{MST}(G))/\text{OPT}(1)$ .  $\square$

An alternative proof with more realistic assumptions (but heavier notation) observes that, as  $T \rightarrow \infty$ , we a.s. observe at least one travel demand for each possible origin-destination pair in the city. Therefore, the optimal static solution should at least span all of  $V$ .

**4.2.2. Results.** Recall the condition  $\eta \cdot (\delta_s/\delta_d) < \theta^+$  certifying the value of dynamism in Remark 1. For any fixed stage duration  $\delta := \delta_d$  and any fixed allowable detour factor  $\rho$ , our experiments use Proposition 2 to compute the lower bound  $\hat{\theta}^+$  on  $\theta^+$ —this allows us to certify the value of dynamism whenever  $\eta \cdot (\delta_s/\delta_d) < \hat{\theta}^+$ . If  $\delta_s = \delta_d$ , the condition reduces to  $\eta < \hat{\theta}^+$ . More generally, for  $\delta_s \neq \delta_d$ , the condition reduces to  $\eta < \hat{\theta}^+ \cdot (\delta_d/\delta_s)$ .

Figure 2 shows the term  $\hat{\theta}^+ \cdot (\delta_d/\delta_s)$  as a function of  $\delta_d$ ,  $\delta_s$ , and  $\rho$  for  $\delta_d \leq \delta_s$ —which is to say that the

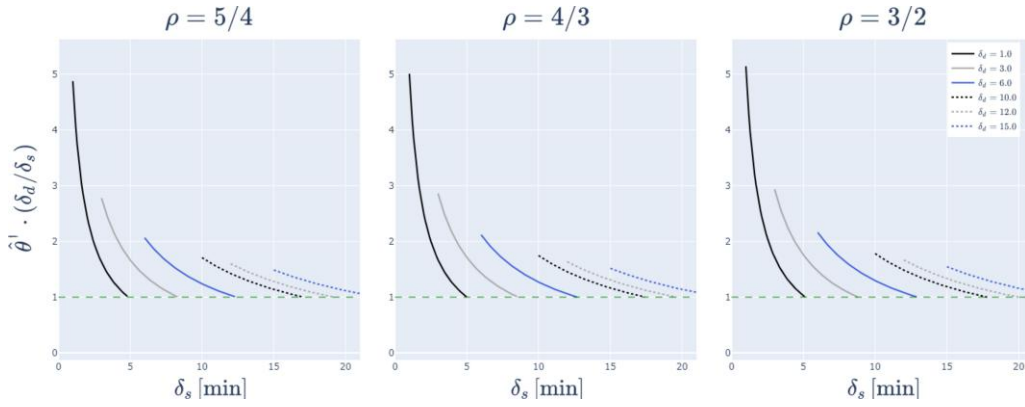
**Figure 1.** (Color online) Plot of  $\theta^+ := \alpha(1)$  as a Function of Graph Connectivity Measures for the Multistage Steiner Tree Problem on Graphs on  $n = 7$  Nodes and Different Distributions  $\mathcal{D}$  over  $\mathcal{I}$ : the Uniform Distribution  $\mathcal{U}$ , the Distribution  $\mathcal{D}^{+\text{cent}}$  Biased Toward “Central” Nodes, and the Distribution  $\mathcal{D}^{-\text{cent}}$  Biased Toward “Peripheral” Nodes



Note. The dashed lines correspond to linear fits.

microtransit batching interval less than or equal to the transit headway. In this way, we obtain quantitative estimates on the parameter combinations under which dynamism is certifiably valuable. For example, the curves suggest that if the microtransit batching interval

and the transit headway are each 15 minutes, then on-demand integration is worthwhile as long as the surcharge coefficient  $\eta$  is less than around 1.5. We caution that these experiments, although more realistic than those in Section 4.1, are still based on an abstraction that does

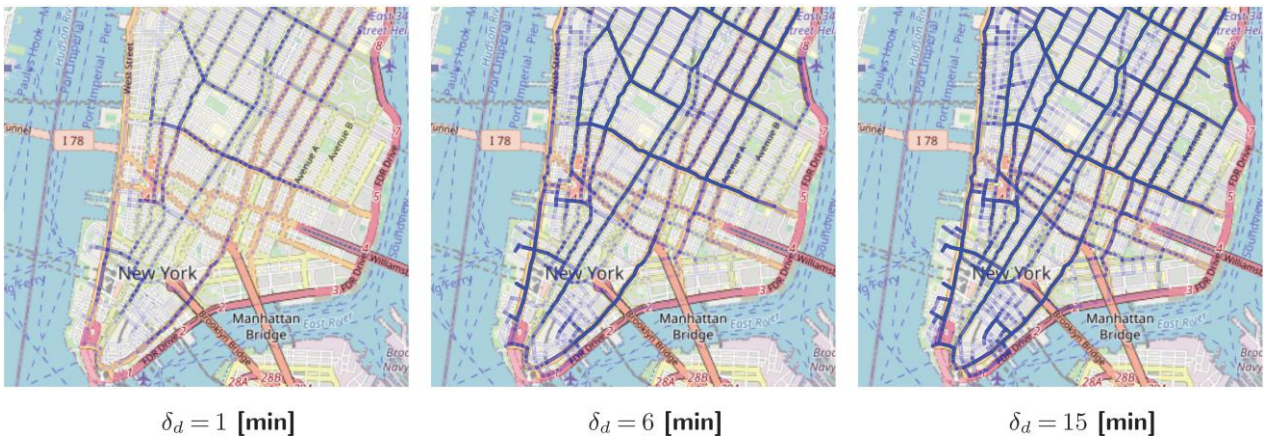
**Figure 2.** (Color online)  $\hat{\theta}^+ \cdot (\delta_d/\delta_s)$  as a Function of  $\delta_d$ ,  $\delta_s$ , and  $\rho$ 

*Notes.* The left, center, and right panels correspond to  $\rho$  equal to  $5/4$ ,  $4/3$ , and  $3/2$ , respectively. Within each panel, each curve corresponds to a different choice of  $\delta := \delta_d$ , as indicated by the legend. For fixed  $\delta_d$ ,  $\delta_s$ , and  $\rho$ , we certify the value of dynamism for surcharge coefficient  $\eta$  less than the value along the corresponding curve. For example, if  $\delta_d = 6$  and  $\rho = 5/4$  (i.e., the curve third from the left in the left panel) and moreover,  $\delta_s = 10$ , we can certify the value of dynamism for  $\eta \leq 1.25$  (i.e., the vertical axis value of the curve at the horizontal axis value of 10).

not capture factors such as demand-side effects, fleet size, vehicle capacities, and rebalancing. Nevertheless, the figure provides robust qualitative insights. First, for any fixed transit headway  $\delta_s$ , the value of dynamism increases with the microtransit batching interval  $\delta_d$  assuming passengers tolerate longer waiting times relative to the transit headway. This effect is amplified for small  $\delta_s$ , where the surcharge coefficient  $\eta$  can be large—for very small  $\delta_s$ , transit operations are already very costly, in which case microtransit can be valuable even if the surcharge coefficient is very large (again, assuming passengers tolerate long waiting times relative to the transit headway). We moreover observe slight gains from increased passenger tolerance to detours as captured by  $\rho$ , particularly for small  $\delta_d$ . Namely, if we treat  $\hat{\theta}^+ := \hat{\theta}^+(\delta_d, \rho)$  as a function  $\hat{\theta}^+ : \mathbb{R}_{>0} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 1}$ , the plots show that for fixed  $\delta_d$  and  $\rho_1 \geq \rho_2$ , we have

$\hat{\theta}^+(\delta_d, \rho_1) \geq \hat{\theta}^+(\delta_d, \rho_2)$ . These effects can be explained as follows; for small  $\delta_d$ , there are fewer requests per stage, and so, travel demands are met with more direct, less shared paths. In this case, increasing  $\rho$  enhances sharing, thereby reducing costs. For large  $\delta_d$ , there are more requests per stage, and so, travel demands are more likely to overlap, naturally enhancing sharing without the need of increasing  $\rho$ . In other words, the longer customers wait to be served by the dynamic system, the cheaper it is for the system to offer them shared yet direct travel.

The effects of  $\delta_d$  are further evidenced in Figure 3. Recall we compute an optimal nonanticipatory solution with characteristic vector  $\chi^t$  for input scenario  $I^t \in \mathcal{I}$  (i.e., a minimum cost  $\chi^t$  such that  $(\chi^t, 0) \in \mathcal{P}^t$ ) individually for each stage  $t \in [T]$ . For each  $j \in [m]$  corresponding to the  $j$ th edge, we compute the frequency  $|\{t \in [T] : \chi_j^t = 1\}|/T$  with which it appears as part of the

**Figure 3.** (Color online) We Compute an Optimal Nonanticipatory Solution Independently for Each Stage  $t \in [T]$ , with  $\rho = 5/4$  and Different Choices of  $\delta := \delta_d$  in Minutes

*Notes.* The darker a road segment, the higher its frequency (over the stages) as part of the installed network. (Left panel)  $\delta_d = 1$  (minutes). (Center panel)  $\delta_d = 6$  (minutes). (Right panel)  $\delta_d = 15$  (minutes).

nonanticipatory solutions. In other words, we rank road segments by the frequency (over the stages) with which they appear as part of the installed network and thereby, their importance as potential trunk lines. We observe that the smaller  $\delta := \delta_d$  is, the fewer high rank road segments there are. However, these few road segments are precisely the best candidates for forming the static network; they appear as part of the installed network in most stages despite the fact that there are few requests per stage for small  $\delta_d$ —they are the ones that enable sharing. We believe this can be leveraged in a subsequent step of the transit system design process: the design of the operational network. In particular, we believe frequently used road segments, especially those that are frequently used for small  $\delta := \delta_d$ , can be combined to obtain a good data-driven initial set of candidate lines for line planning via column generation (see Borndörfer, Grötschel, and Pfetsch 2007, Gattermann, Harbering, and Schöbel 2017).

## 5. Conclusions

Our goal with this work is to provide a principled and tractable analytical framework with which to study the value of dynamism, as quantified by the dynamicity gap and related measures. Our main practical motivation is the ongoing debate regarding the value of on-demand integration in transit systems. We showcase our framework with two sets of computational experiments involving high-level abstractions of integrated transit systems. Our abstractions are by no means an exact representation of the real world; producing and solving such models are research problems in and of themselves. However, we believe they capture the essence of the real-world problem sufficiently well to provide qualitative insight about the conditions under which on-demand integration might be most valuable. We hope this style of characterization enables accessible insight for both researchers and practitioners: given the problem at hand, leverage knowledge about the input parameters to quickly assess whether dynamism is worthwhile investment.

Going forward, we are interested in expanding our framework and experiments to handle enhanced models with more realistic operational features and moreover, in studying the value of dynamism in settings not necessarily related to network design. We are also interested in studying the value of dynamism while relaxing a crucial assumption in our framework, namely that the input scenarios are sampled i.i.d. on every stage. Although it may be much more challenging to obtain convergence results, we believe capturing stage dependence and stage transition costs (e.g., in our context, microtransit rebalancing) would significantly enhance our models.

## Acknowledgments

The authors thank the anonymous referees for their careful review and constructive commentary.

## References

- Alonso-González MJ, Liu T, Cats O, Van Oort N, Hoogendoorn S (2018) The potential of demand-responsive transport as a complement to public transport: An assessment framework and an empirical evaluation. *Transportation Res. Record* 2672(8): 879–889.
- Alonso-Mora J, Samaranayake S, Wallar A, Frazzoli E, Rus D (2017) On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proc. Natl. Acad. Sci. USA* 114(3):462–467.
- Archetti C, Speranza MG, Weyland D (2018) A simulation study of an on-demand transportation system. *Internat. Trans. Oper. Res.* 25(4):1137–1161.
- Awasthi P, Goyal V, Lu BY (2019) On the adaptivity gap in two-stage robust linear optimization under uncertain packing constraints. *Math. Programming* 173(1):313–352.
- Bavelas A (1950) Communication patterns in task-oriented groups. *J. Acoustical Soc. America* 22(6):725–730.
- Beineke LW, Oellermann OR, Pippert RE (2002) The average connectivity of a graph. *Discrete Math.* 252(1–3):31–45.
- Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization* (Princeton University Press, Princeton, NJ).
- Bertsimas D, Goyal V (2010) On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Math. Oper. Res.* 35(2):284–305.
- Bertsimas D, Goyal V, Lu BY (2015) A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Math. Programming* 150(2):281–319.
- Bertsimas D, Goyal V, Sun XA (2011) A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Math. Oper. Res.* 36(1):24–54.
- Bertsimas D, Ng YS, Yan J (2020) Joint frequency-setting and pricing optimization on multimodal transit networks at scale. *Transportation Sci.* 54(3):839–853.
- Bertsimas D, Ng YS, Yan J (2021) Data-driven transit network design at scale. *Oper. Res.* 69(4):1118–1133.
- Birge JR (1982) The value of the stochastic solution in stochastic linear programs with fixed recourse. *Math. Programming* 24(1): 314–325.
- Boeing G (2017) OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Systems* 65:126–139.
- Borndörfer R, Grötschel M, Pfetsch ME (2007) A column-generation approach to line planning in public transport. *Transportation Sci.* 41(1):123–132.
- Das KC (2018) Proof of conjecture involving algebraic connectivity and average degree of graphs. *Linear Algebra Appl.* 548:172–188.
- Dean BC, Goemans MX, Vondrák J (2008) Approximating the stochastic knapsack problem: The benefit of adaptivity. *Math. Oper. Res.* 33(4):945–964.
- Desautniers G, Hickman MD (2007) Public transit. Barnhart C, Laporte G, eds. *Handbook in Operations Research and Management Science*, vol. 14 (Elsevier, Amsterdam), 69–127.
- Feigon S, Murphy C (2016) Shared mobility and the transformation of public transit. tCRP Research Report No. 188, Transportation Research Board, Washington, DC.
- Gattermann P, Harbering J, Schöbel A (2017) Line pool generation. *Public Transport* 9(1):7–32.
- Gupta A, Pál M, Ravi R, Sinha A (2004) Boosted sampling: Approximation algorithms for stochastic optimization. *Proc. Thirty-Sixth Annual ACM Sympos. Theory Comput.* (Association for Computing Machinery, New York), 417–426.
- Gupta A, Pál M, Ravi R, Sinha A (2005) What about Wednesday? Approximation algorithms for multistage stochastic optimization. Chekuri C, Jansen K, Rolim JDP, Trevisan L, eds. *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques. APPROX RANDOM 2005*, Lecture Notes in Computer Science, vol. 3624 (Springer, Berlin), 86–98.

- Hagberg A, Swart PS, Chult D (2008) Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Laboratory, Los Alamos, NM.
- Hall JD, Palsson C, Price J (2018) Is Uber a substitute or complement for public transit? *J. Urban Econom.* 108:36–50.
- Immorlica N, Karger D, Minkoff M, Mirrokni VS (2004) On the costs and benefits of procrastination: Approximation algorithms for stochastic combinatorial optimization problems. *Proc. Fifteenth Annual ACM-SIAM Sympos. Discrete Algorithms* (Society for Industrial and Applied Mathematics), 691–700.
- Kleywegt AJ, Shapiro A, Homem-de Mello T (2002) The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.* 12(2):479–502.
- Liu Y, Ouyang Y (2021) Mobility service design via joint optimization of transit networks and demand-responsive services. *Transportation Res. Part B Methodological* 151:22–41.
- Ljubić I (2021) Solving Steiner trees: Recent advances, challenges, and perspectives. *Networks* 77(2):177–204.
- Luo Q, Samaranayake S, Banerjee S (2021) Multimodal mobility systems: Joint optimization of transit network design and pricing. *Proc. ACM/IEEE 12th Internat. Conf. Cyber-Physical Systems*, 121–131.
- Merlin LA (2019) Transportation sustainability follows from more people in fewer vehicles, not necessarily automation. *J. Amer. Planning Assoc.* 85(4):501–510.
- Newman M (2018) *Networks* (Oxford University Press, Oxford, UK).
- NYC Taxi and Limousine Commission (2021) TLC trip record data. Accessed February 27, 2021, <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>.
- Périvier N, Hssaine C, Samaranayake S, Banerjee S (2021) Real-time approximate routing for smart transit systems. *Proc. ACM Measurement Anal. Comput. Systems*, (Association for Computing Machinery, New York), 73–74.
- Ravi R, Sinha A (2006) Hedging uncertainty: Approximation algorithms for stochastic optimization problems. *Math. Programming* 108(1):97–114.
- Rayle L, Dai D, Chan N, Cervero R, Shaheen S (2016) Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy* 45:168–178.
- Read R, Wilson R (1998) *An Atlas of Graphs* (Clarendon, Oxford, UK).
- Schöbel A (2012) Line planning in public transportation: Models and methods. *OR Spectrum* 34(3):491–510.
- Shaheen S, Chan N (2016) Mobility and the sharing economy: Potential to facilitate the first-and last-mile public transit connections. *Built Environ.* 42(4):573–588.
- Shmoys DB, Swamy C (2006) An approximation scheme for stochastic linear programming and its application to stochastic integer programs. *J. ACM* 53(6):978–1012.
- Steiner K, Irnich S (2020) Strategic planning for integrated mobility-on-demand and urban public bus networks. *Transportation Sci.* 54(6):1616–1639.
- Stiglic M, Agatz N, Savelsbergh M, Gradisar M (2018) Enhancing urban mobility: Integrating ride-sharing and public transit. *Comput. Oper. Res.* 90:12–21.
- Toth P, Vigo D (2014) *Vehicle Routing: Problems, Methods, and Applications* (SIAM, Philadelphia).
- Walker J (2012) *Human Transit: How Clearer Thinking About Public Transit Can Enrich Our Communities and Our Lives* (Island Press, Washington, DC).
- Walker J (2018) Microtransit: What I think we know. *Human Transit*. Accessed June 27, 2021, <https://humantransit.org/2018/02/microtransit-what-i-think-we-know.html>.
- Westervelt M, Huang E, Schank J, Borgman N, Fuhrer T, Peppard C, Narula-Woods R (2018) UpRouted: Exploring microtransit in the United States. Technical report, Eno Center for Transportation, Washington, DC.