DESIGNING BETTER TRANSPORT SERVICES THROUGH JOINT TRAFFIC FORECASTING AND NETWORK OPTIMIZATION

EMMA FREJINGER

Associate Professor, Dep. Computer Science and Operations Research, Université de Montréal

The presentation is based on joint work with

BERNARD GENDRON, LÉONARD MORIN AND MAËLLE ZIMMERMANN

Motivation and Scope

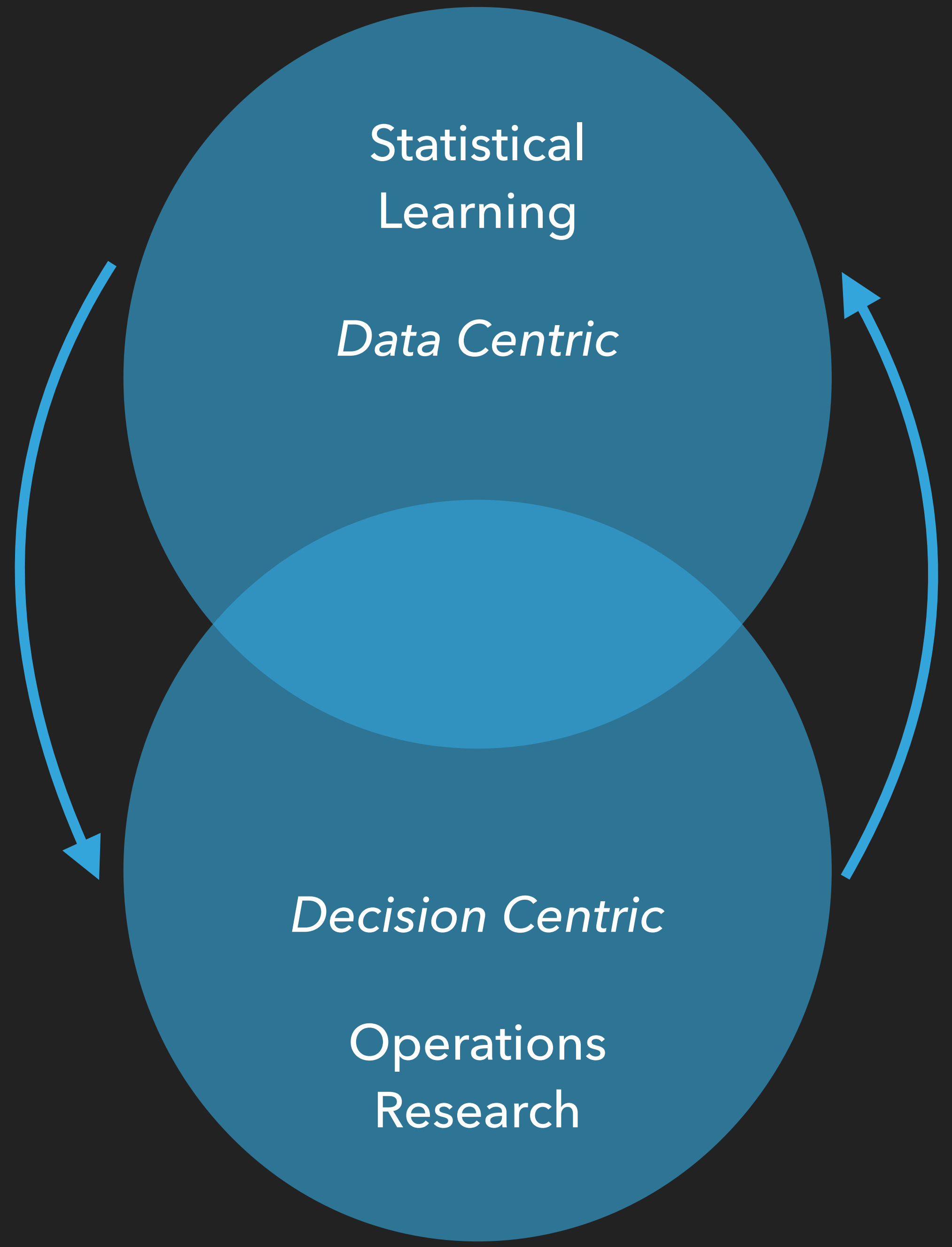
- Users in a transport system make choices according to their needs, preferences and budget which are heterogeneous in the population.

- In order to encourage users to adopt sustainable transport services, we need to design them taking user preferences into account.

- This leads to challenging optimization and forecasting problems and we propose models and algorithms to tackle them.

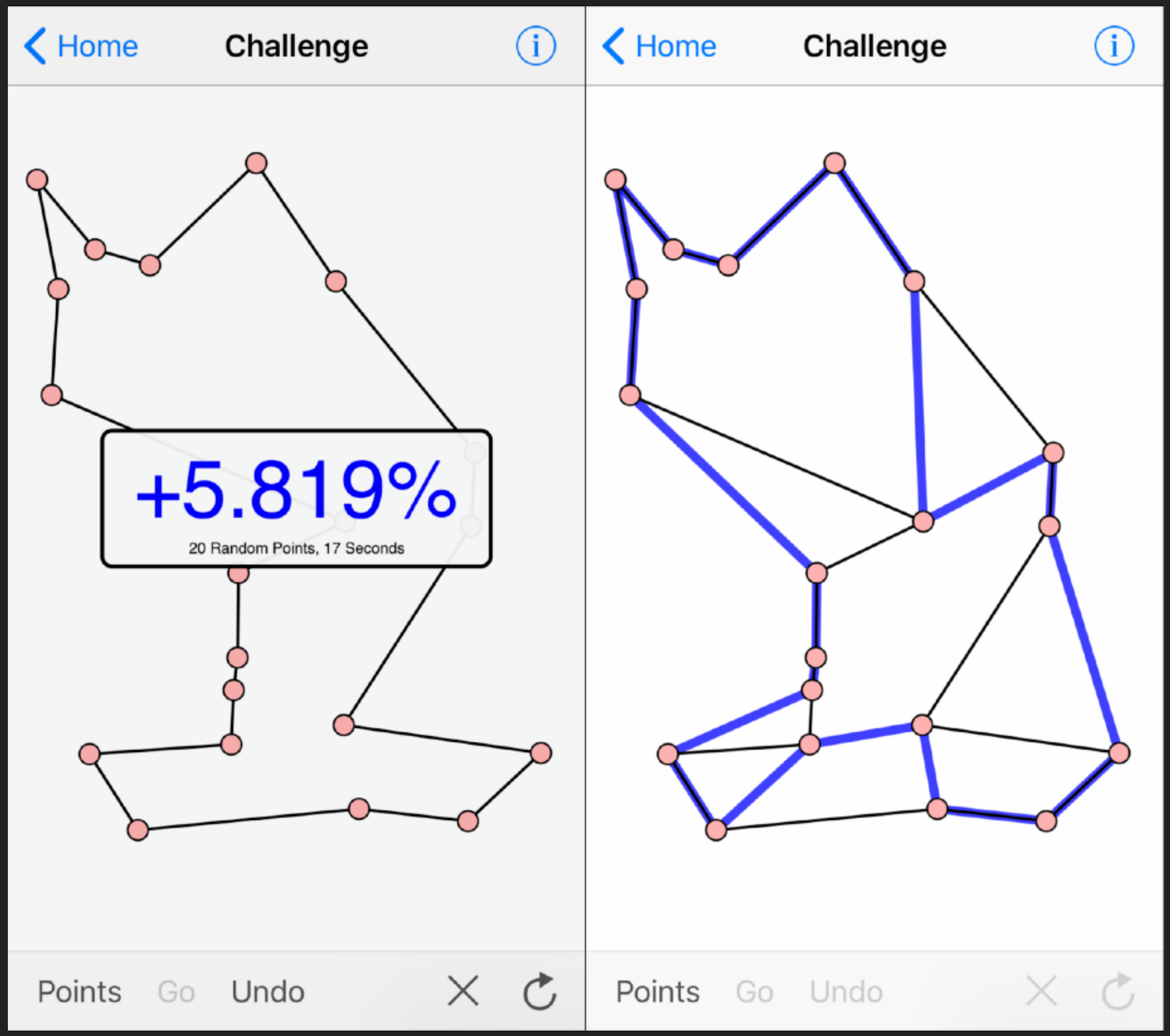
THIS TALK IS ABOUT A TOPIC AT THE INTERSECTION BETWEEN STATISTICAL LEARNING AND OPERATIONS RESEARCH

- Discrete optimization (within the field of operations research) is focused on solving complex decision problems that are too hard or too time consuming for humans to solve
- Combined with statistical learning: transforms forecasts into tangible decision support



MANY DISCRETE OPTIMIZATION PROBLEMS ARE VERY HARD

- The travelling salesman problem: Given a list of cities and distances between each pair, find the shortest route that visits each city and returns to the original city.
- Easy to understand but hard to solve
- TSP with 20 cities has 19!/2 = **60,822,550,000,000,000** solutions
- Effective algorithms exist to solve large instances

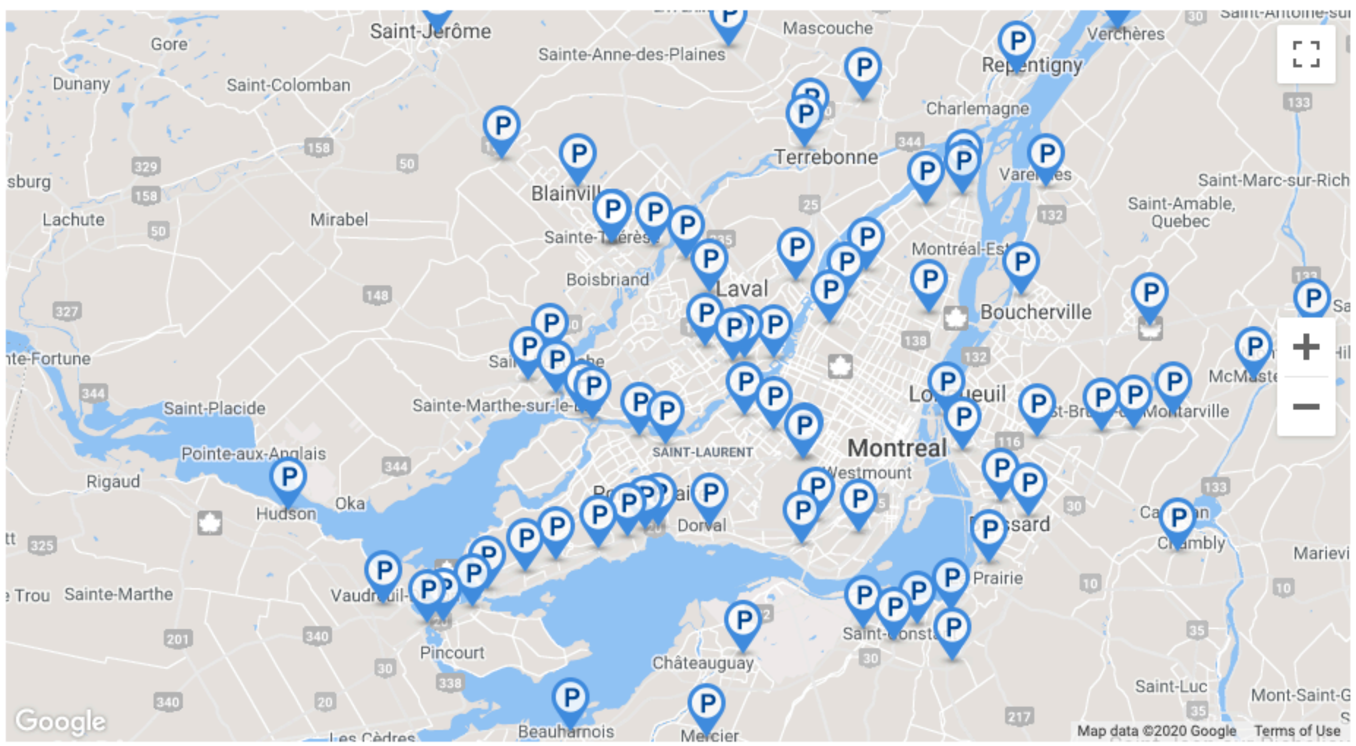


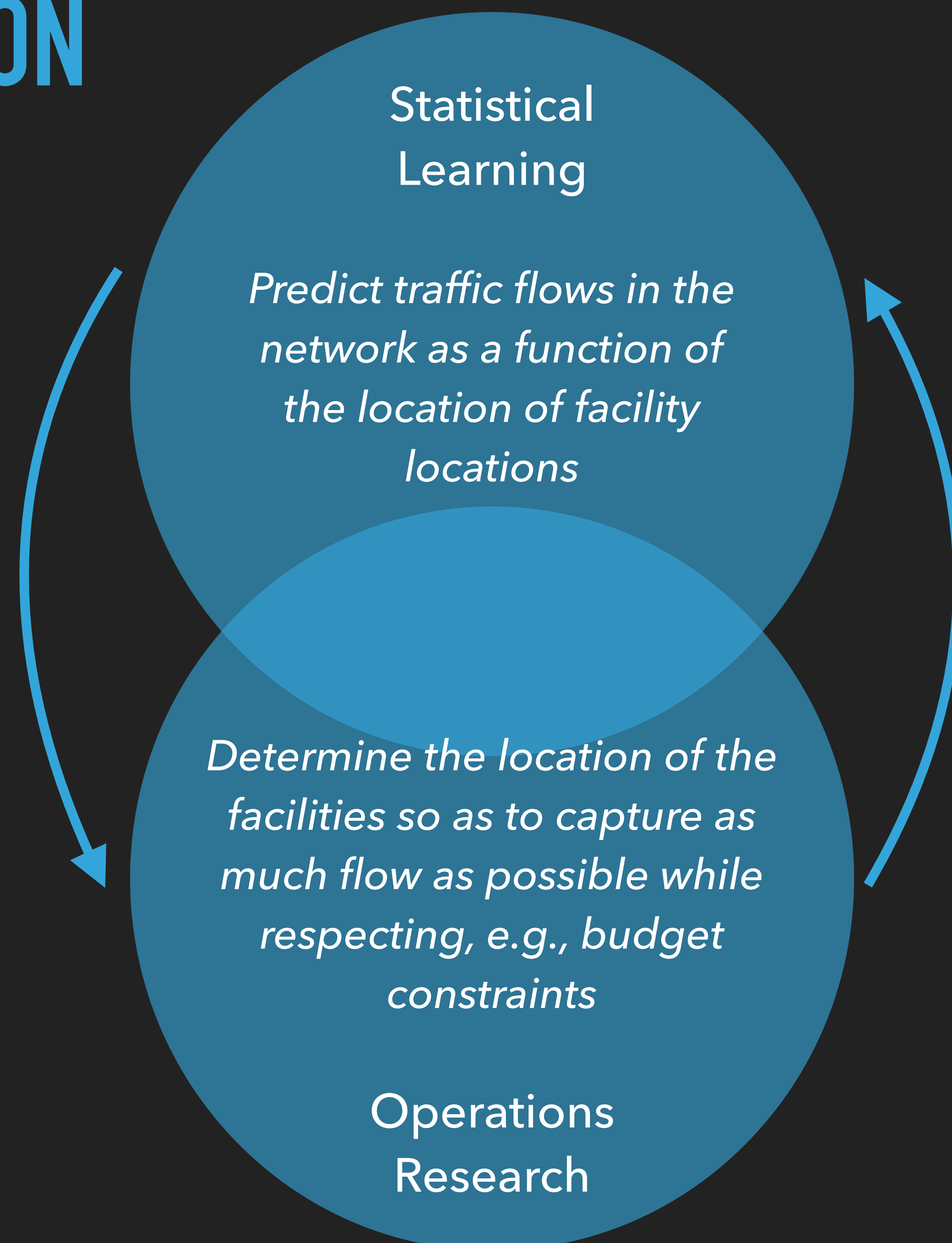
Don't get addicted!
Concorde TSP solver app

Book: In the pursuit of the travelling salesman: Mathematics at the limits of computation, William Cook, Princeton University Press, 2012.

EXAMPLE: FLOW CAPTURE
PARK-AND-RIDE FACILITY LOCATION

Park-and-ride facilities





EFFECTIVE ALGORITHMS AND COMPUTING POWER

- A wide range of real word applications rely on operations research methodologies: scheduling, vehicle routing, service network design, fleet management, …
- Impressive results over the past two decades: more than 265,000x algorithmic speedup! Dimitri Bertsimas (MIT) cites and overall 200 billion speedup between 1991 and 2014 for mixed integer programming solvers.

Most progress on mixed integer linear programs with known and fixed parameters - deterministic problems.

Problems that would have taken 7 years to solve in 1991, take one second now.

George Nemhauser
Georgia Institute of Technology

CPLEX and Gurobi solvers, assuming conservative 1000x machine speedup, 1991-2003.

STATISTICAL LEARNING + OPERATIONS RESEARCH

Interact

SL and optimization are integrated, interplay and learn by interaction with the environment.
Example: OR model learns to adapt to a changing environment.

Interplay

SL and optimization are integrated and interplay.
Example: bilevel optimization - flow capture

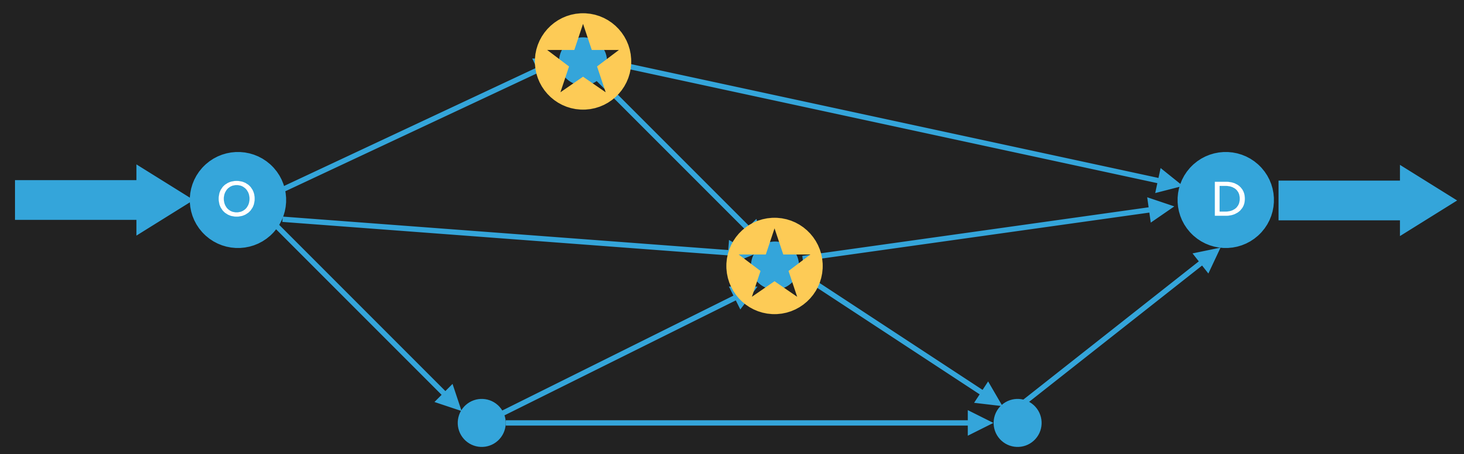
Understand

SL is used to characterize uncertainty in the environment and then treated as known.
Majority of models in practice, e.g., scheduling algorithms.

OUTLINE

- The flow capture problem in brief and why it is important to consider heterogeneous user preferences
- Bilevel optimization: challenging from both a forecasting and optimization perspective
- A short detour: a bilevel optimization perspective on the route choice modeling literature
- Flow capture under heterogeneous user behaviour in uncongested networks: high-level summary of key results
- Conclusion and future work

THE FLOW CAPTURE PROBLEM IN BRIEF



Travellers choose the path that minimizes travel time and the facilities reduce travel time, i.e., they make the corresponding paths more attractive.

The solution to the bilevel optimization problem depends on the demand model that distributes the flow in the network.

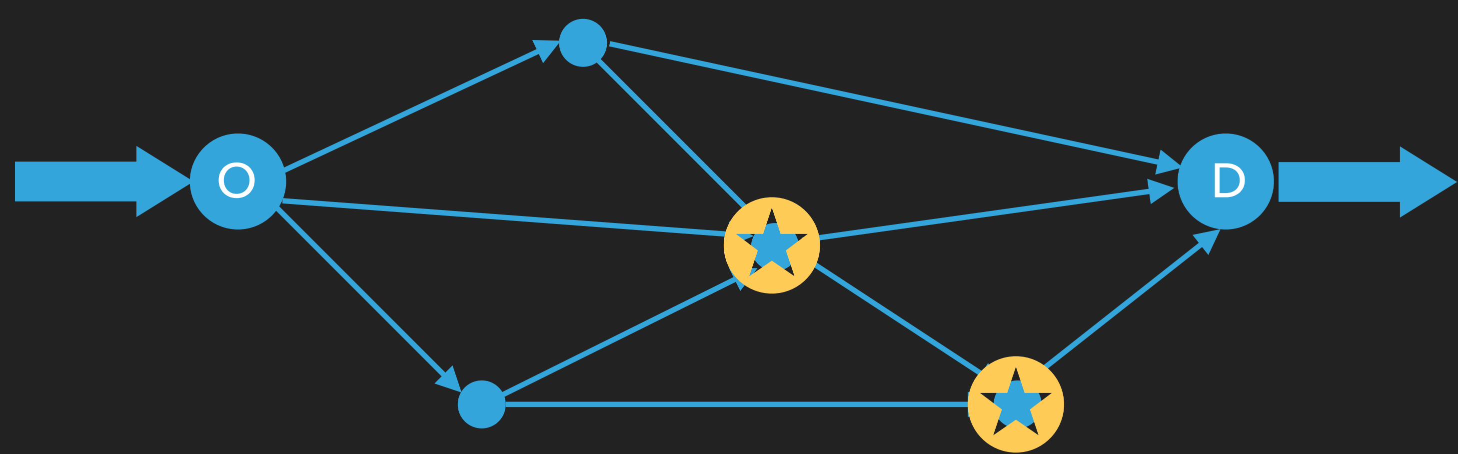
LEADER first selects on which arc to locate facilities in an uncongested network, anticipating the followers' reactions

Maximize captured flow while respecting budget constraint (here max. two resources)

Minimize travel time that depends on the leader's decisions

FOLLOWERS select the best path in reaction to leader decisions

THE FLOW CAPTURE PROBLEM IN BRIEF



Travellers have different preferences and empirical findings show that distributing flow according to a discrete choice model leads to a more accurate prediction of traffic flows.

Leader problem searches over the space of full space of solutions and followers adapt, the utilities hence depend on the upper-level decision variables.

Need for accurate forecasts of traffic flow given any facility location solution

LEADER first selects on which arc to locate facilities in an uncongested network, anticipating the followers' reactions

Maximize expected captured flow while respecting budget constraint (here max. two resources)

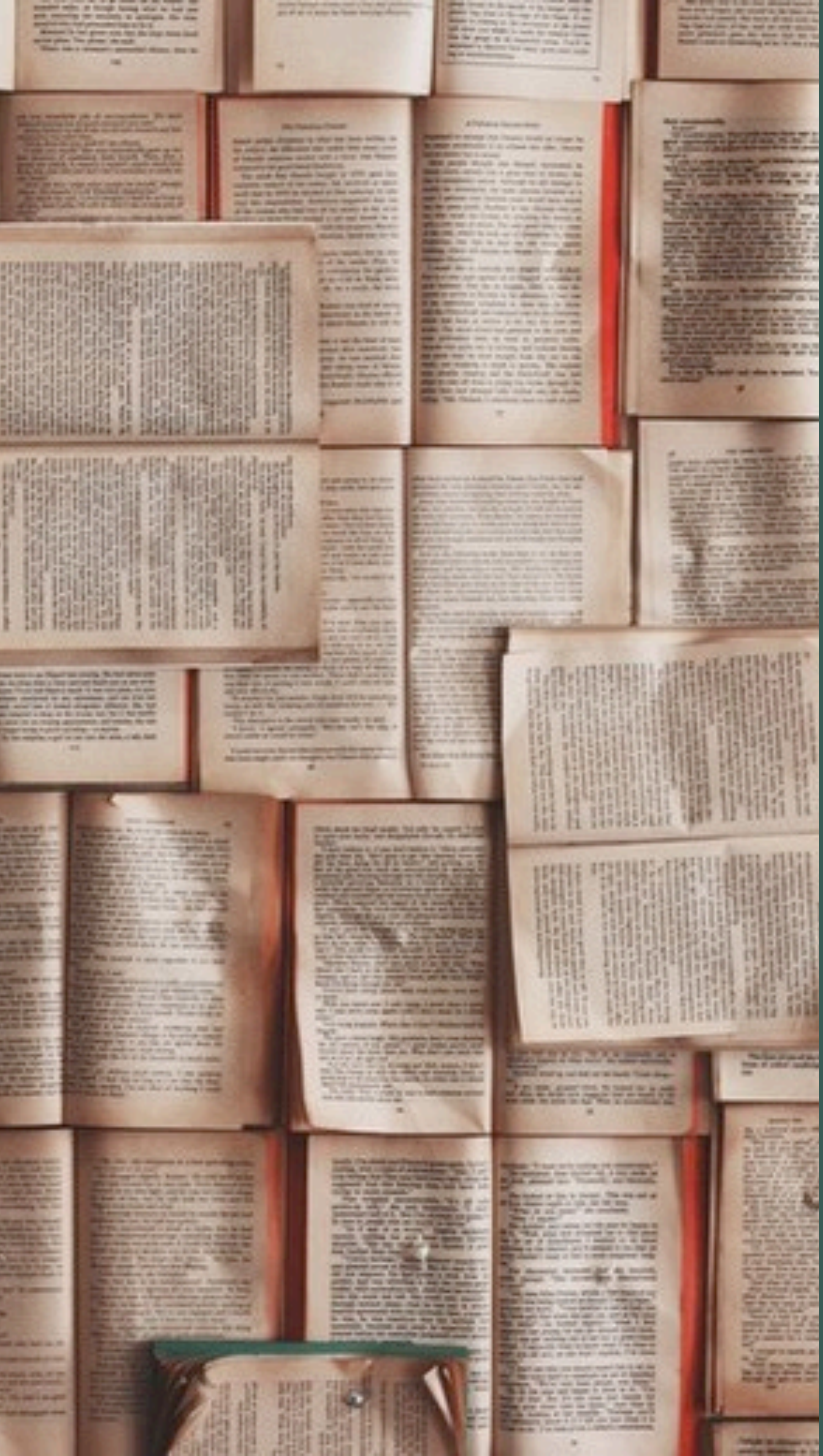
Maximize random utility that depends on the leader's decisions and travel time

FOLLOWER selects the best path in reaction to leader decisions

OPTIMIZATION AND FORECASTING CHALLENGES

- Bilevel formulations where the upper level is a discrete optimization problem and the follower choices are modelled by state-of-the-art choice models (random utility maximization models) are very challenging to solve: highly non-convex with combinatorial features
- The forecasting problem is also challenging: the traffic flow predictions should be accurate even for scenarios that have never been seen in any data
  - Structural models and computational challenges imposes restrictions on which model and model specifications that can be used

Our contribution: we bridge the gap between the state-of-the-art traffic flow prediction models and flow capture models and show that we can solve problems with heterogenous users in realistic size networks.



A bilevel optimization perspective on the route choice modelling literature

Zimmermann and Frejinger (2020)
A tutorial on recursive models for analyzing and predicting path choice behavior, EURO Journal on Transportation and Logistics, 9(2).

Frejinger and Zimmermann (2020)
Route Choice and Network Modeling
Prepared for International Encyclopedia of Transportation, Elsevier.

INTRODUCTION

- Route choice models are used to analyze and predict path choice behaviour
- Discrete choice models
  - Interpretation of model parameters
  - Prediction of traffic flow, often under new network scenarios for which there are no observations of path choices available

In the case of bilevel optimization: parameter interpretation, functional form of the deterministic utilities, the type of discrete choice model as well as the prediction accuracy for new network scenarios are important.

HIGH-LEVEL OVERVIEW

- Simplest model: deterministic shortest path with one common generalized cost function
  - Unrealistic: travellers have different preferences, measurement errors in attributes (e.g., travel time)
- Stochastic (discrete choice) model: travellers are random utility maximizers
- Learning - inverse problem: identify from observed path choices the utility function (aka generalized cost for shortest path problems) travellers seek to maximize
  - Deterministic utility: inverse shortest path problem
  - Random utility (assuming a family of distributions): learn parameters such that the likelihood of the sample of observed path choices is maximized (dynamic discrete choice models / inverse reinforcement learning)

RANDOM UTILITY MAXIMIZATION MODELS

Restricted network

Path-based models

Generate a restricted number of paths in each choice set for each OD. Ignore all other alternatives.

If the network changes, the **choice sets need to be updated**. Choice sets are difficult to validate and if they are not accurate, the predictions can be arbitrarily bad.

Unrestricted network

Path-based models

Assuming an unrestricted network, sample choice sets and correct utilities according to the sampling protocol.

Used to obtain unbiased estimates but unclear how to use for prediction.

Recursive (arc-based) models

No need to generate choice sets and predictions can be computed per O or D.

Expected traffic flows can be computed by solving a linear system.

If network changes, the value functions need to be recomputed or solve shortest path problems in a simulation approach.

In practice, restricting the network can have major impact on the results! Recursive models have important advantages in the context of bilevel optimization as it is possible to leverage the fact that we can compute shortest paths to predict traffic flow.

Flow Capture under Heterogeneous User Behaviour in Uncongested Networks

Leonard Morin, Emma Frejinger and Bernard Gendron

FLOW CAPTURE WITH STATE-OF-THE-ART TRAFFIC FLOW PREDICTION

- Problem: under limited budget, the leader wishes to deploy facilities of several types on the arcs in the network to capture the maximum amount of flow of all users, given that each facility intercepts a proportion of the flow
- Demand: flow can be indifferent, attracted and/or evasive to the resources
- Our contribution: close the gap between state-of-the-art flow capture and state-of-the-art route choice models
  - Flow is predicted using a nested recursive logit model in a simulation framework and can incorporate attracted / evasive / indifferent flows
  - We formulate a bilevel programming model, single-level reformulation and Benders decomposition (solved by Branch-and-Benders-Cut) that relies heavily on solving shortest path problems

MATHEMATICAL GYMNASTICS… AND WRITING THE FOLLOWER PROBLEM USING SIMULATED UTILITIES INSTEAD OF PROBABILITIES LEAD TO A NICE FORMULATION

LEADER

$$Z = \max_{y \in Y} \sum_{a \in A} \sum_{r \in R} \sigma_{ar} q_{ar} y_{ar} \left(\frac{1}{|S|} \sum_{s \in S} \sum_{k \in K} \sum_{l \in L} d_l^k x_{la}^{ks} \right)$$

$$x_l^{ks} \in \arg \min_{x_l^{ks} \in X_l^{ks}} \{ u_l^s(y, \psi^s) x_l^{ks} \}, \; s \in S, k \in K, l \in L,$$

$$X_l^{ks} = \left\{ x_l^{ks} \in \{0,1\}^{|A|} \mid \sum_{a \in F(n)} x_{la}^{ks} - \sum_{a \in B(n)} x_{la}^{ks} = e_n^k, n \in N \right\}$$

FOLLOWER

For fixed $y \in Y$, the follower problem decomposes by scenario, user categories and OD pair.

Deterministic shortest path problem that can be solved by Dijkstra.

Resource location and arc flow variables are only linked through bilinear terms in the objectives of the leader and the follower.

SINGLE-LEVEL REFORMULATIONS

- We use the properties of the model to derive single-level reformulations and a Benders decomposition method
- We can derive different MILP reformulations that can be solved by state-of-the-art solvers:
  - Replace objective of the follower by dual feasibility constraints and optimality conditions (complementary slackness (CS) conditions or strong duality (SD) constraint)
  - Lagrangian reformulation (L)
  - Linearize bilinear terms by the introduction of new variables and constraints

Model	$\mathcal{CS}$	$\mathcal{SD}$	$\mathcal{L}$
$\mathscr{M}_{\mathcal{CS}}$	✓		
$\mathscr{M}_{\mathcal{CS}-\mathcal{L}}$	✓		✓
$\mathscr{M}_{\mathcal{SD}}$		✓	
$\mathscr{M}_{\mathcal{SD}-\mathcal{L}}$		✓	✓
$\mathscr{M}_{\mathcal{L}}$			✓

BENDERS DECOMPOSITION METHOD

- Based on $\mathscr{M}_{\mathscr{L}}$ model
- Designed to solve large-scale problems
- Master problem: binary variables y
- Given $y \in Y$, the Benders subproblem decomposes by scenario, OD pair and user categories
- We omit all the technical details on the reformulation and the generation of optimality cuts
- Algorithm: Branch-and-Benders-Cut (BBC) - a single branch and cut tree

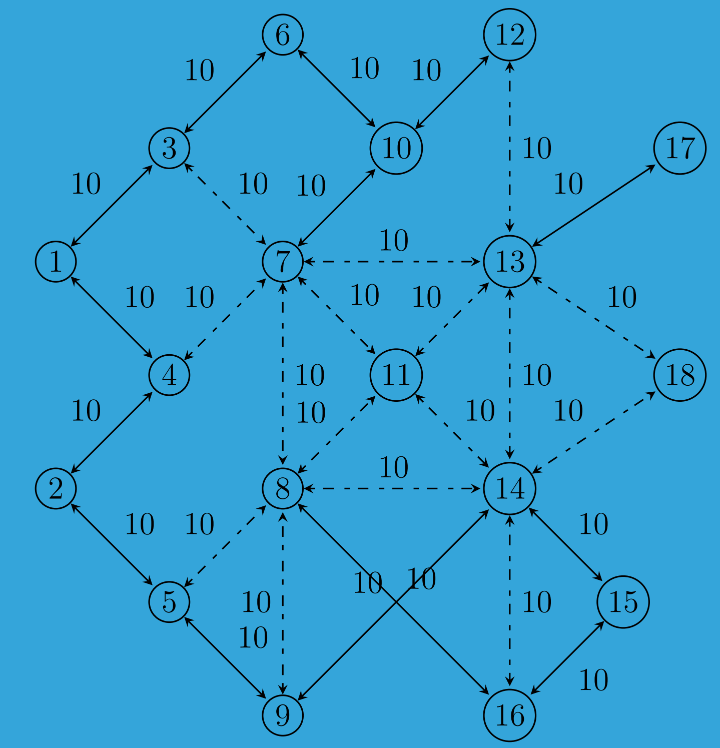
COMPUTATIONAL EXPERIMENTS

- Small network to study the impact of problem characteristics
- Winnipeg network to study the computing times
- Nested recursive logit (Mai et al., 2015) that allows for path utilities to be correlated (state of the art arc-based model)

$$u_{la}(y, t, \varepsilon) = \sum_{r \in R} \beta_{lar} \sigma_{ar} y_{ar} + \alpha_l t_a + \varepsilon_{la}$$

$$\varepsilon_{al} = \mathrm{GEV}(0{,}e^{\theta t_a}) + \kappa$$

RESULTS SMALL NETWORK



- 8 OD pairs, 32 candidate arcs
- 1 user category, 30 scenarios, 2 resource types
- Very Evasive (VE), mildly evasive (ME), mildly cooperative (MC), Very cooperative (VC)
- Four levels of budget: 10%, 25%, 50%,75% of candidate arcs

It is more difficult to prove optimality for cooperative than evasive users.

A budget of 25% leads to the most difficult instances to solve – if the budget is small, there are few possibilities and if budget is large, there are no tradeoffs to make.

The model that works the best depends on the user behaviour.

On these small instances the Benders decomposition method does not make a significant improvement.

RESULTS WINNIPEG NETWORK – EVASIVE USERS

- 1040 nodes and 2836 arcs
- 2 resource types
- Candidate arcs are selected in decreasing order of flows from the best known flow solution
- OD pairs are randomly sampled
- Budget fixed to 30%

The Benders decomposition method is required to solve these larger problems.

RESULTS WINNIPEG NETWORK – EVASIVE USERS

Stability (optimal value vs number of scenarios for instance 2)

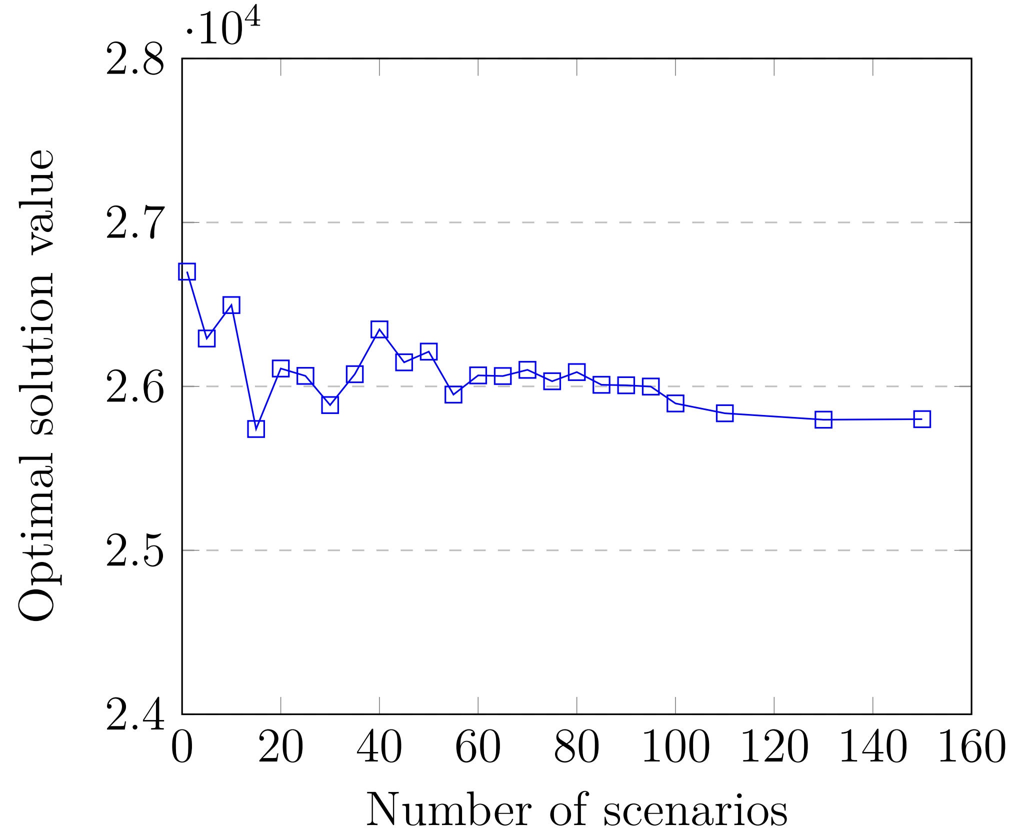
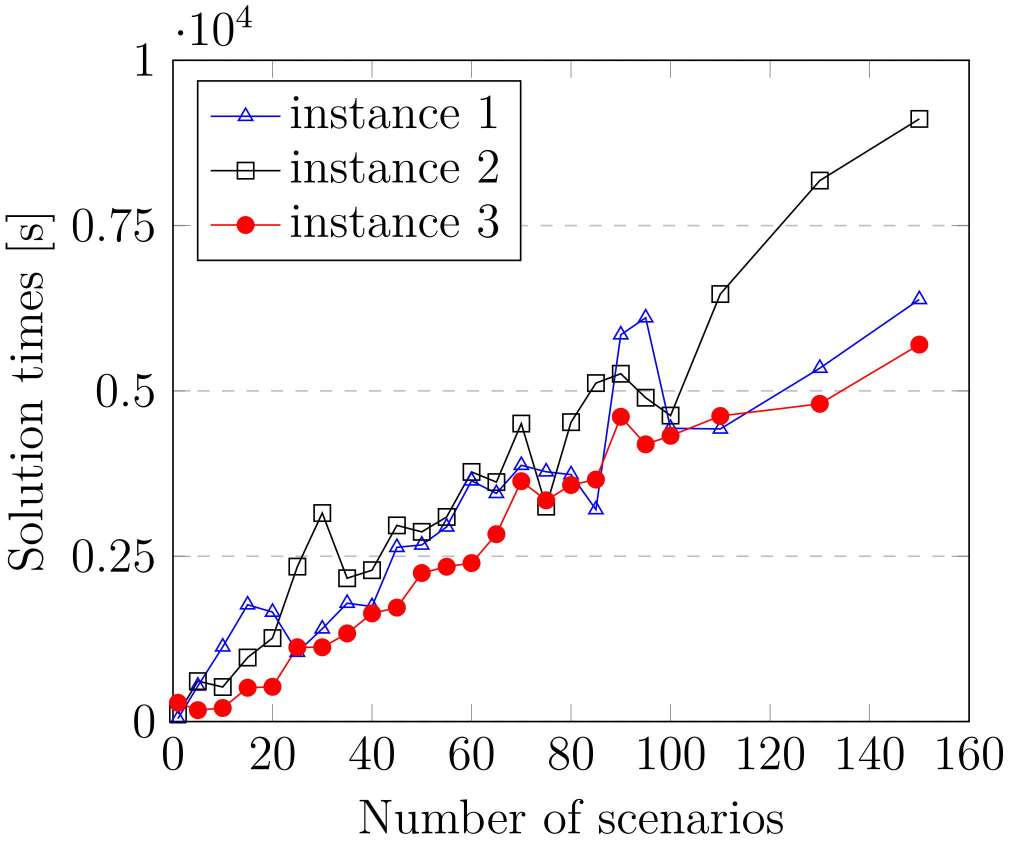


Fig. 9. Solution times vs number of scenarios



The number of scenarios required varies slightly over instances.

Computing time increases more or less linearly with the number of scenarios.

Results show promise to solve real large-scale problems.

Conclusion

Discrete optimization problems are important to a variety of applications. Often demand is assumed to be **fixed and known**, hence the solutions **ignore** that users in fact have **heterogeneous preferences**.

There is a need to **bridge the gap** between the state-of-the-art **discrete optimization** and the state-of-the-art **discrete choice models**.

We propose a **bilevel formulation** of a flow capture model along with an effective **solution approach**. They allow to compute **solutions that consider the hetereogenous preferences** of users predicted by a state-of-the-art path choice model.

Computational challenges can be addressed by leveraging the **structure** and **properties** of the models. In our case, relying on s**hortest path calculations** for choice predictions was crucial.

In **future work** we will use this approach for various **applications** and we will develop methodology for dealing with congested networks.

Thank you!

emma.frejinger@umontreal.ca