Analytics for Better Urban Cycling

Bo Lin
(soon) National University of Singapore

eMERGE Seminar, UC Berkeley
March 2nd, 2026

Cycling has become increasingly popular

Cycling can be stressful

Bike lanes help to alleviate cycling stress

Toronto looking to build 100 km of brand new bike lanes across the city

Laura Hanrahan | Dec 7 2021, 10:19 am

Analytics for better urban cycling

2020 2021 2023 2025

Analytics for better urban cycling

NSERC Alliance Grants

2020 — 2021 — 2023 — 2025

Phase I	Phase II	Phase III
Assessment	**Planning**	**Applicability**

Analytics for better urban cycling

NSERC Alliance Grants

Assessment framework


U of T News

Follow U of T News

Toronto's COVID-19 bike lane expansion boosted access to jobs, retail: U of T study

2020 **2021** **2023** **2025**

Phase I
Assessment

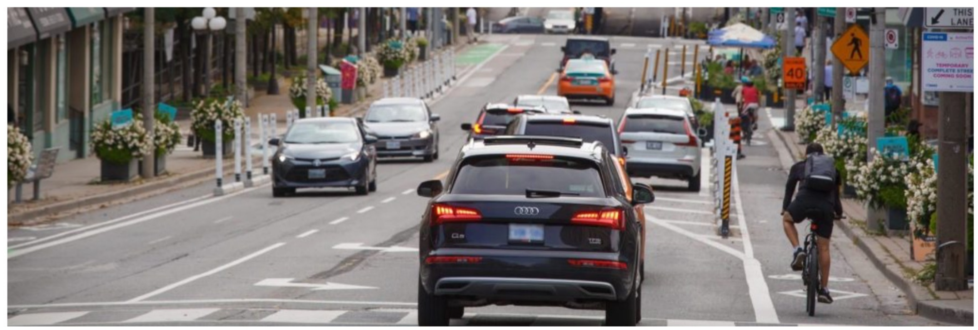
Analytics for better urban cycling

NSERC Alliance Grants

Assessment framework

Yonge Street (implemented)

ActiveTO Midtown Yonge Complete Street



2020 — 2021 — 2023 — 2025

Phase I
Assessment

Analytics for better urban cycling

NSERC Alliance Grants

Assessment framework

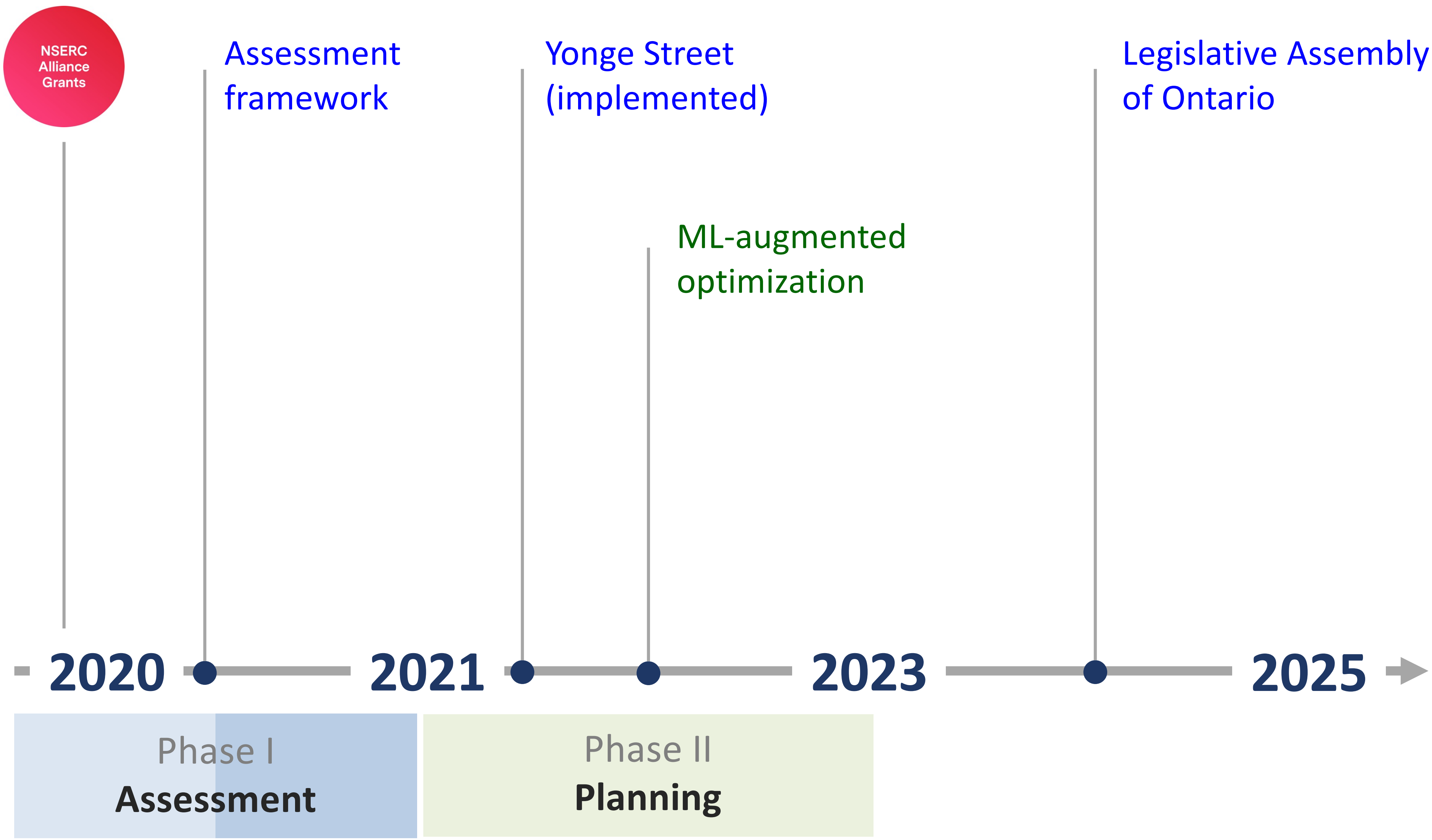
Yonge Street (implemented)

Legislative Assembly of Ontario

2020 **2021** **2023** **2025**

Phase I
Assessment

Analytics for better urban cycling



Analytics for better urban cycling

NSERC Alliance Grants

Assessment framework

Yonge Street (implemented)

Legislative Assembly of Ontario

Toronto's 2025—2029 plan

2020 2021 2023 2025

Phase I
Assessment

Phase II
Planning

Analytics for better urban cycling

NSERC Alliance Grants

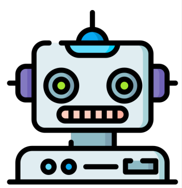
Assessment framework

Younge Street (implemented)

Legislative Assembly of Ontario

ML-augmented optimization

Toronto's 2025—2029 plan



AutoLTS

Cycling stress

AutoLTS

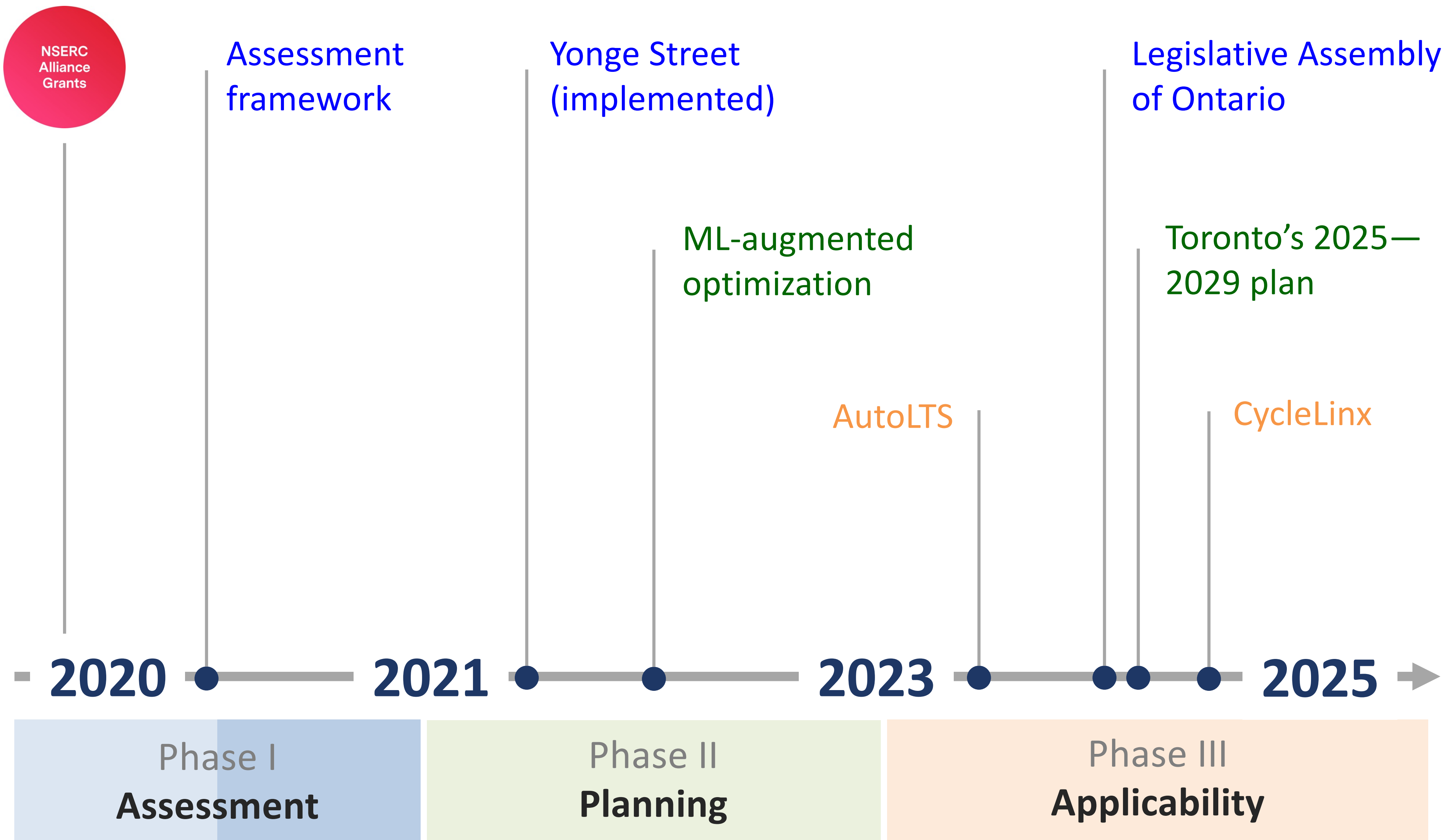
2020 **2021** **2023** **2025**

Phase I
Assessment

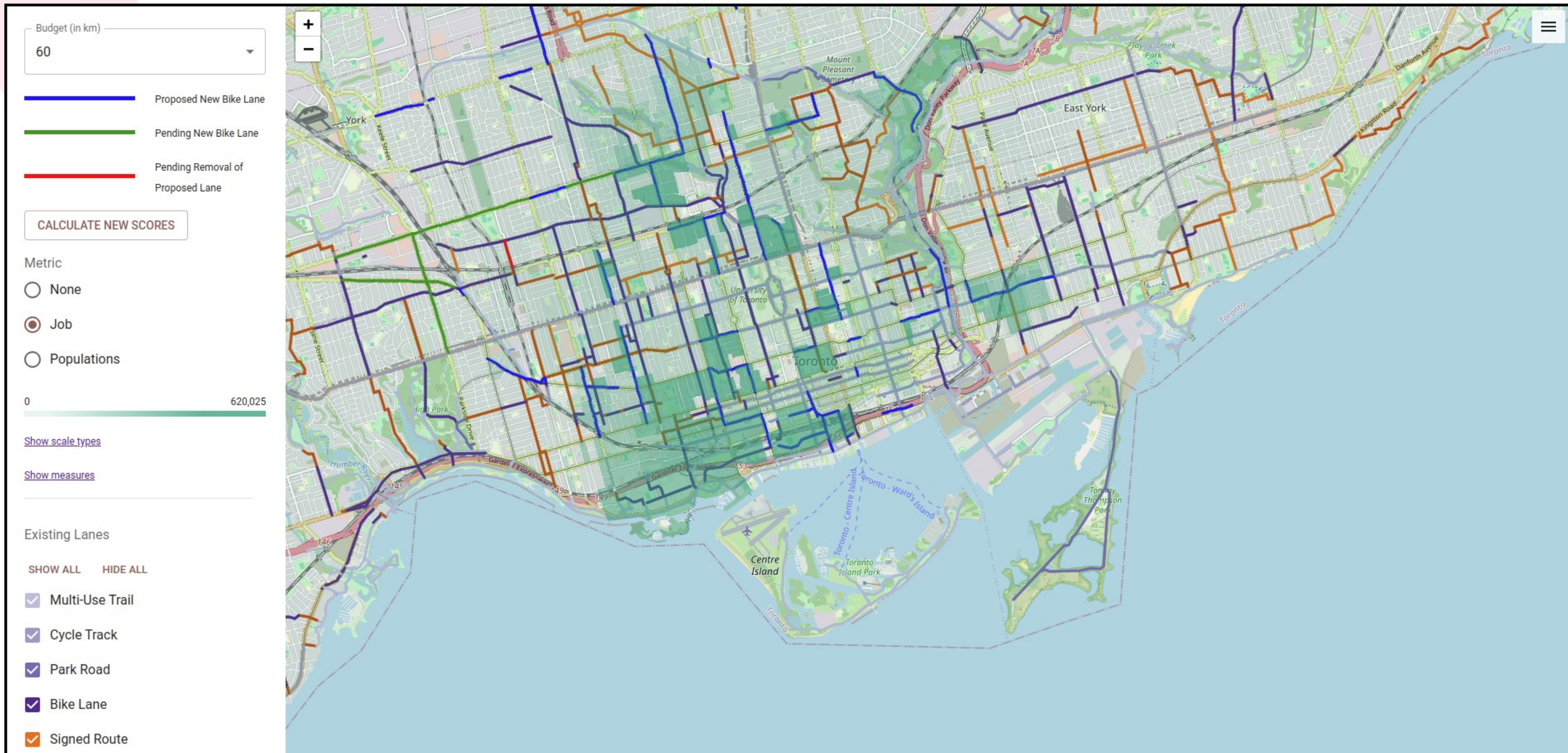
Phase II
Planning

Phase III
Applicability

Analytics for better urban cycling



Analytics for better urban cycling



slative Assembly
ntario

CycleLinx

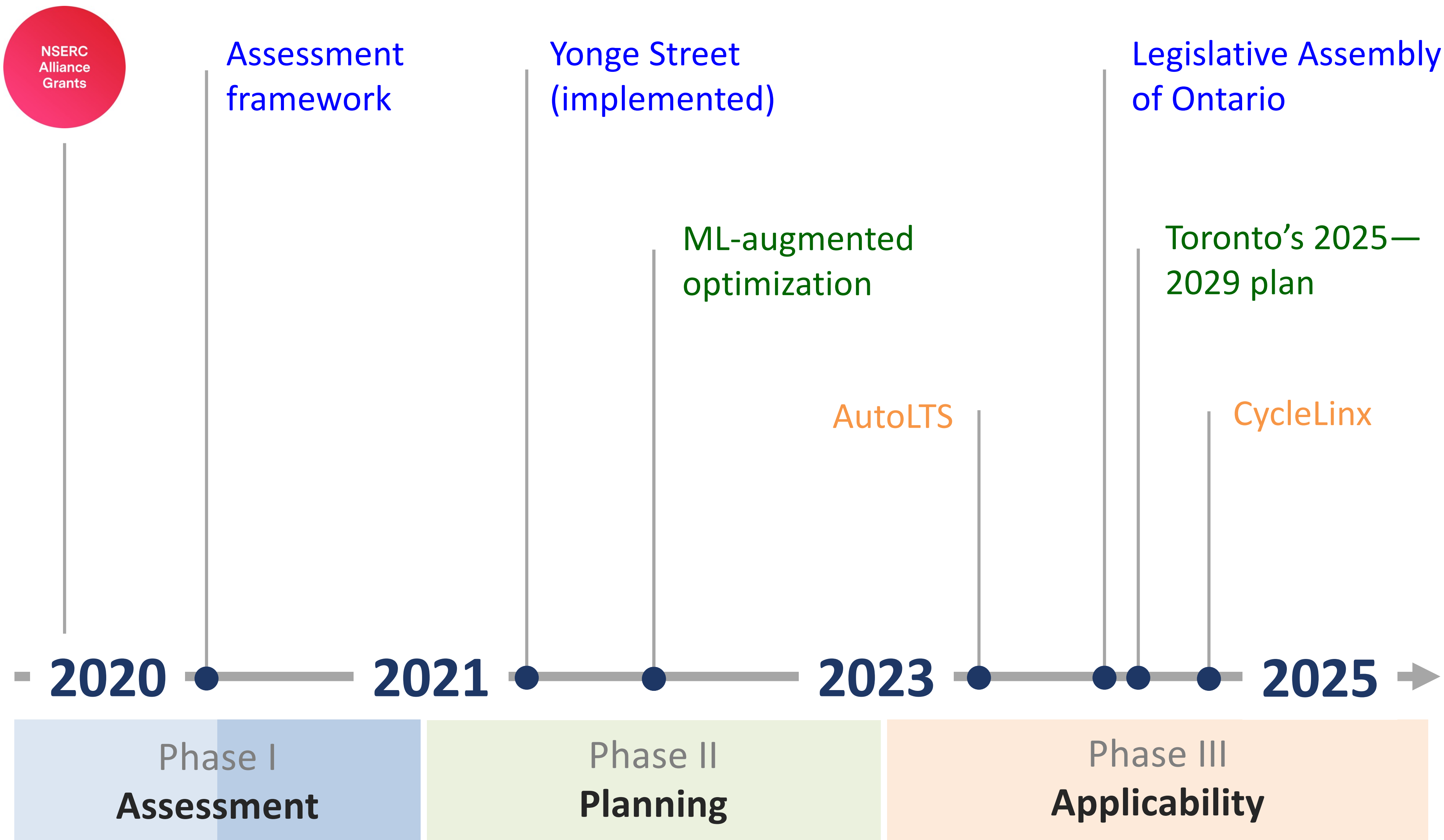
2020 2021 2023 2025

Phase I
Assessment

Phase II
Planning

Phase III
Applicability

Analytics for better urban cycling



Cycling network design as a bilevel problem

Predictive of cycling mode choice
(Imani, Miller, Saxe, 2019)

leader

follower

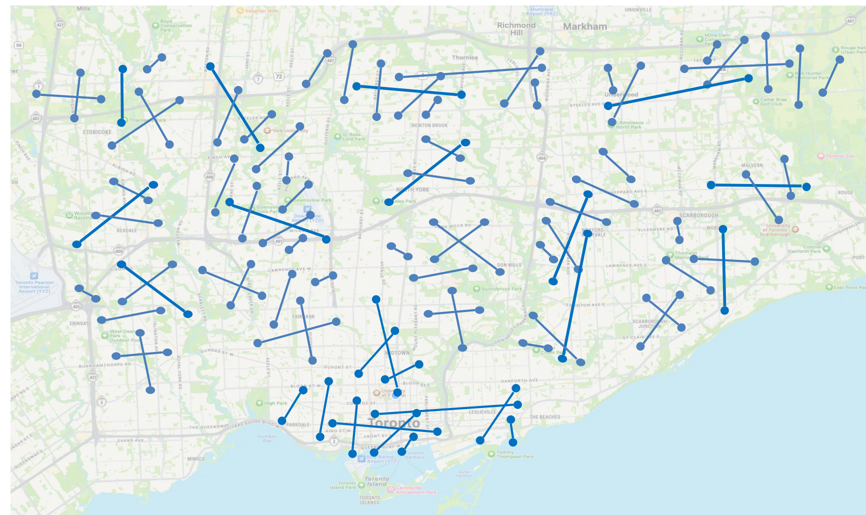
Maximize $\sum_{(o,d)\in\mathcal{F}} S^{od}(\mathbf{x})$

Network design $\mathbf{x} \in \mathcal{X}$

Low-stress shortest path $(\mathbf{x})$

...

x 1 million!



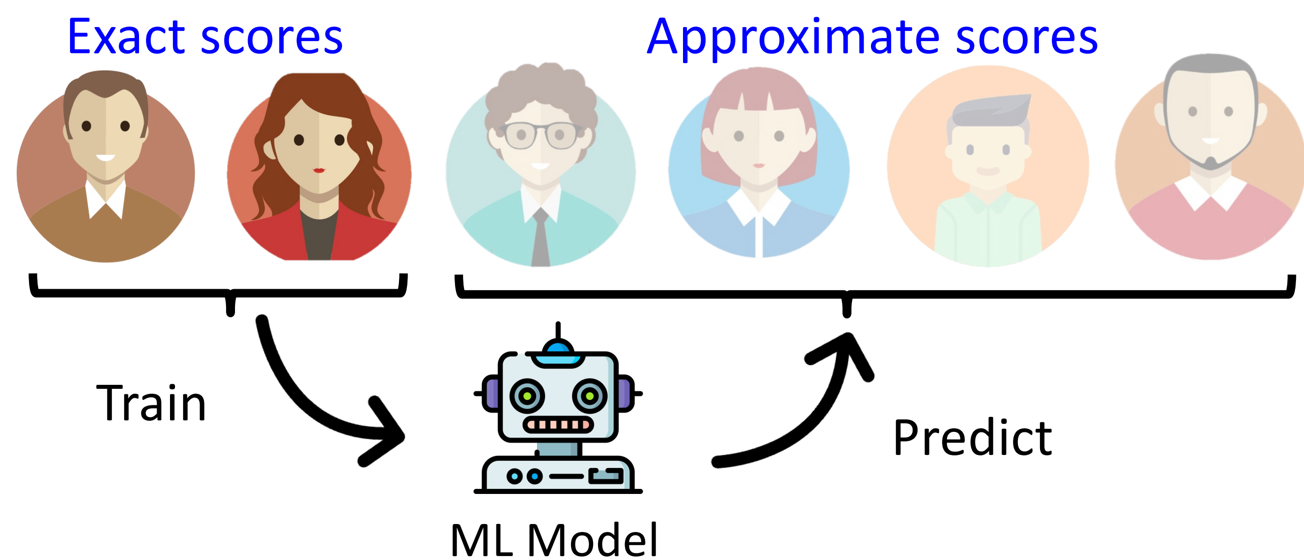
Bilevel Program	Two-Stage Stochastic Program
Leader	1st-stage decision maker
Followers	2nd-stage decision makers (in different scenarios)

How to deal with one million followers?

Sampling!

Tractable, but lose sight of the "bigger picture"

Can we capture other followers (w/o routing) ?



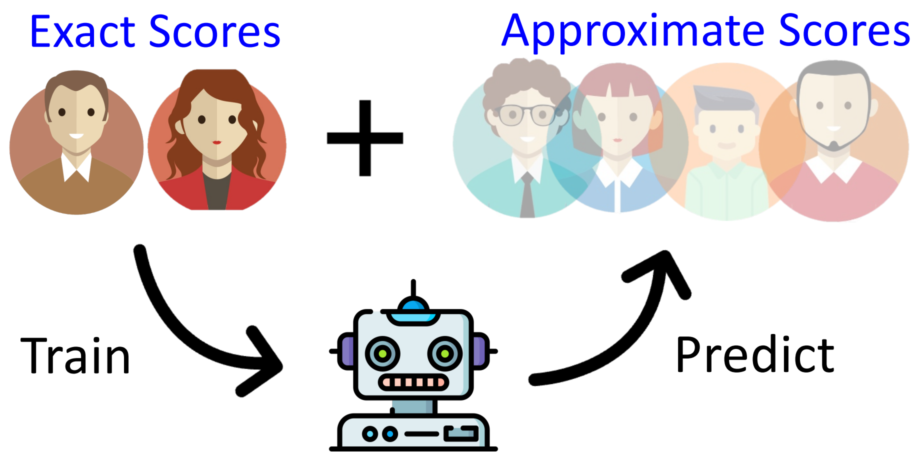
***Key contribution**: ML-Augmented optimization*

- *New paradigm of integrating ML and OPT*
- *Strong performance and theoretical guarantees*

Our Approach

The ML-Augmented Model

$$\underset{\mathbf{x}}{\text{maximize}}$$



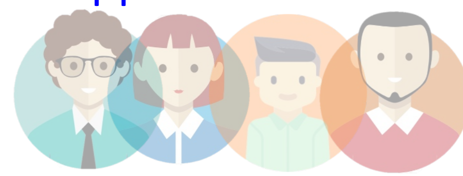
Our Approach

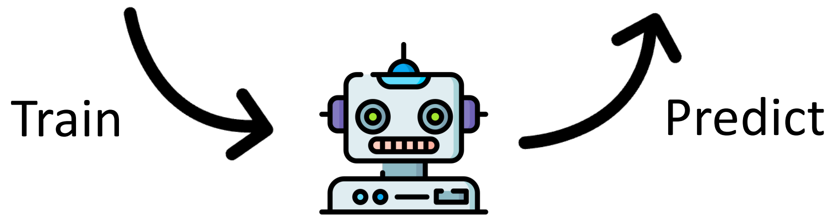
The ML-Augmented Model

Exact Scores

Approximate Scores

$$\underset{\mathbf{x}}{\text{maximize}} \sum_{(o,d)\in\mathcal{T}} S^{od}(\mathbf{x}) +$$





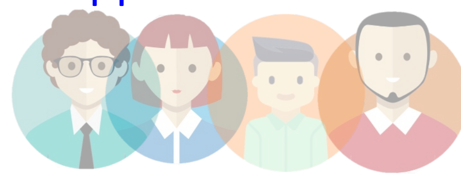
Our Approach

The ML-Augmented Model

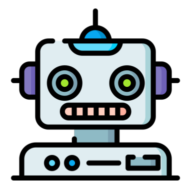
Exact Scores

Approximate Scores

$$\underset{\mathbf{x}}{\text{maximize}} \sum_{(o,d)\in\mathcal{T}} S^{od}(\mathbf{x}) +$$



Train



Predict

$$\mathbf{w}^{\intercal}\mathbf{f}^{od} \longrightarrow S^{od}(\mathbf{x})$$

ML model parameters

features (OD locations, etc.)

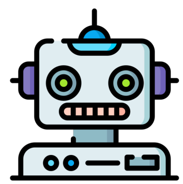
Our Approach

The ML-Augmented Model

Exact Scores Approximate Scores

$$\underset{\mathbf{x}}{\text{maximize}} \sum_{(o,d)\in\mathcal{T}} S^{od}(\mathbf{x}) + \sum_{(o,d)\in\mathcal{F}\setminus\mathcal{T}} \mathbf{w}^\intercal \mathbf{f}^{od}$$

Train



Predict

$$\mathbf{w}^\intercal \mathbf{f}^{od} \longrightarrow S^{od}(\mathbf{x})$$

ML model parameters

features (OD locations, etc.)

Our Approach

The ML-Augmented Model

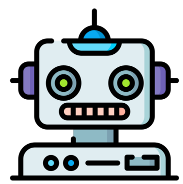
Exact Scores

Approximate Scores

$$\underset{\mathbf{x}}{\text{maximize}} \sum_{(o,d)\in\mathcal{T}} S^{od}(\mathbf{x}) + \sum_{(o,d)\in\mathcal{F}\setminus\mathcal{T}} \mathbf{w}^\intercal \mathbf{f}^{od}$$

Train

Sync



Predict

$$\mathbf{w}^\intercal \mathbf{f}^{od} \longrightarrow S^{od}(\mathbf{x})$$

ML model parameters

features (OD locations, etc.)

Our Approach

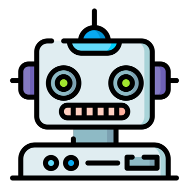
The ML-Augmented Model

Exact Scores Approximate Scores

$$\underset{\boxed{\mathbf{x},\mathbf{w}}}{\text{maximize}} \sum_{(o,d)\in\mathcal{T}} S^{od}(\mathbf{x}) + \sum_{(o,d)\in\mathcal{F}\setminus\mathcal{T}} \mathbf{w}^\intercal \mathbf{f}^{od}$$

Simultaneous Optimization & Training

Train



Predict

$$\mathbf{w}^\intercal \mathbf{f}^{od} \longrightarrow S^{od}(\mathbf{x})$$

ML model parameters

features (OD locations, etc.)

Our Approach

The ML-Augmented Model

Exact Scores Approximate Scores

$$\underset{\boxed{\mathbf{x},\mathbf{w}}}{\text{maximize}} \sum_{(o,d)\in\mathcal{T}} S^{od}(\mathbf{x}) + \sum_{(o,d)\in\mathcal{F}\setminus\mathcal{T}} \mathbf{w}^\intercal \mathbf{f}^{od}$$

Simultaneous Optimization & Training

$$\frac{1}{|\mathcal{T}|} \sum_{(o,d)\in\mathcal{T}} \left| S^{od}(\mathbf{x}) - \mathbf{w}^\intercal \mathbf{f}^{od} \right| \leq \bar{L}$$

L1 Training loss is bounded

Our Approach

The ML-Augmented Model

Exact Scores / Approximate Scores

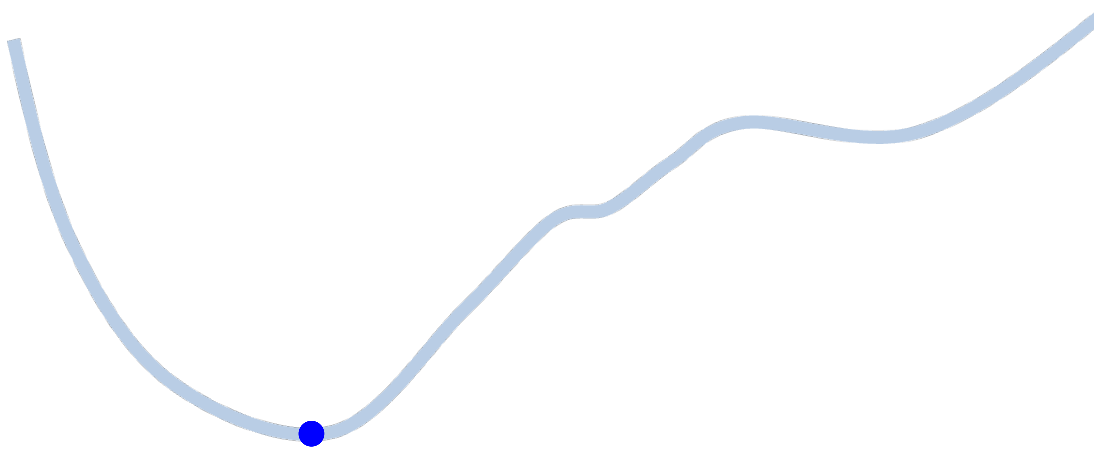
$$\underset{\boxed{\mathbf{x},\mathbf{w}}}{\text{maximize}} \sum_{(o,d)\in\mathcal{T}} S^{od}(\mathbf{x}) + \sum_{(o,d)\in\mathcal{F}\setminus\mathcal{T}} \mathbf{w}^\intercal \mathbf{f}^{od}$$

Simultaneous Optimization & Training

$$\frac{1}{|\mathcal{T}|}\sum_{(o,d)\in\mathcal{T}} \left|S^{od}(\mathbf{x}) - \mathbf{w}^\intercal \mathbf{f}^{od}\right| \le \bar{L}$$

L1 Training loss is bounded

Training loss



$\mathbf{w}$

Our Approach

The ML-Augmented Model

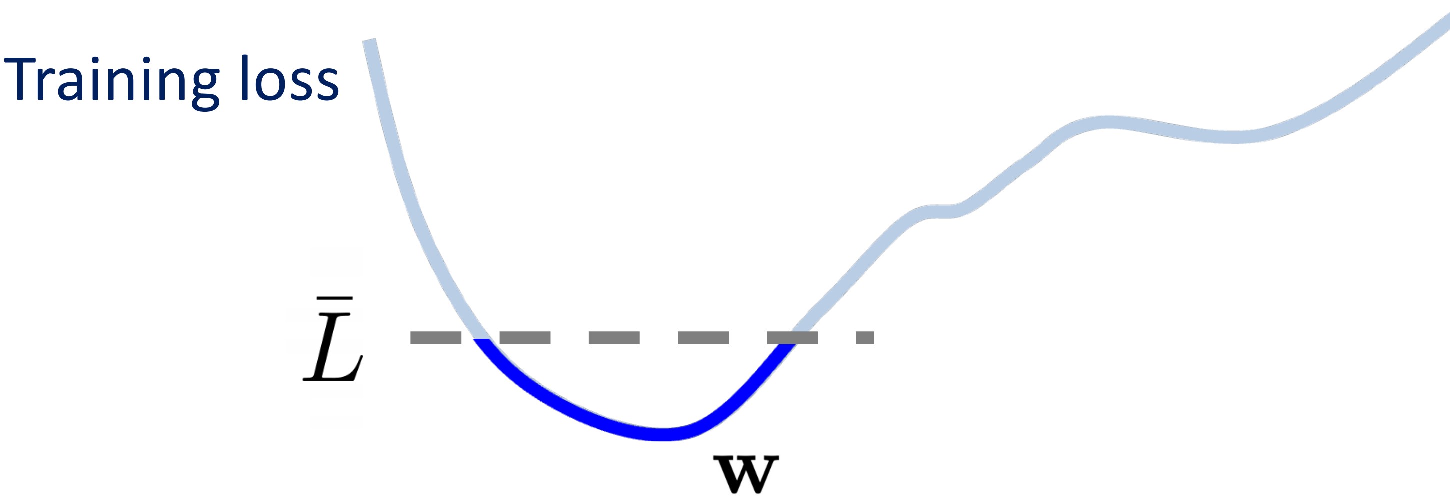
Exact Scores / Approximate Scores

$$\underset{\boxed{\mathbf{x},\mathbf{w}}}{\text{maximize}} \sum_{(o,d)\in\mathcal{T}} S^{od}(\mathbf{x}) + \sum_{(o,d)\in\mathcal{F}\setminus\mathcal{T}} \mathbf{w}^\intercal \mathbf{f}^{od}$$

Simultaneous Optimization & Training

$$\frac{1}{|\mathcal{T}|}\sum_{(o,d)\in\mathcal{T}} \left|S^{od}(\mathbf{x}) - \mathbf{w}^\intercal \mathbf{f}^{od}\right| \leq \bar{L}$$

L1 Training loss is bounded



Our Approach

The ML-Augmented Model

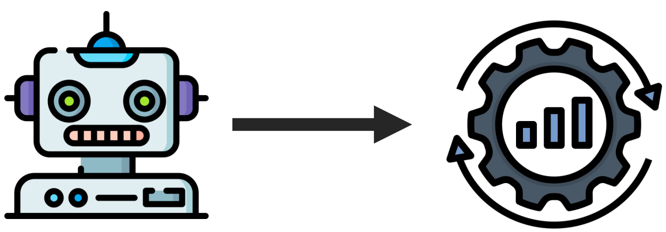
Exact Scores — Approximate Scores

$$\underset{\boxed{\mathbf{x},\mathbf{w}}}{\text{maximize}} \sum_{(o,d)\in\mathcal{T}} S^{od}(\mathbf{x}) + \sum_{(o,d)\in\mathcal{F}\setminus\mathcal{T}} \mathbf{w}^\intercal \mathbf{f}^{od}$$

Simultaneous Optimization & Training

$$\frac{1}{|\mathcal{T}|}\sum_{(o,d)\in\mathcal{T}} \left|S^{od}(\mathbf{x}) - \mathbf{w}^\intercal \mathbf{f}^{od}\right| \le \bar{L}$$

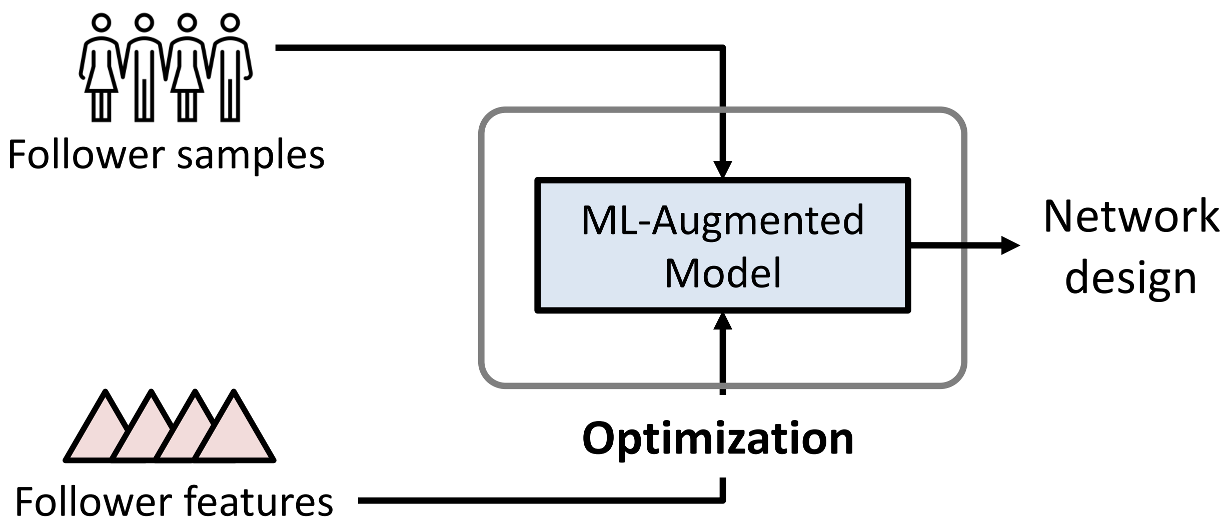
L1 Training loss is bounded



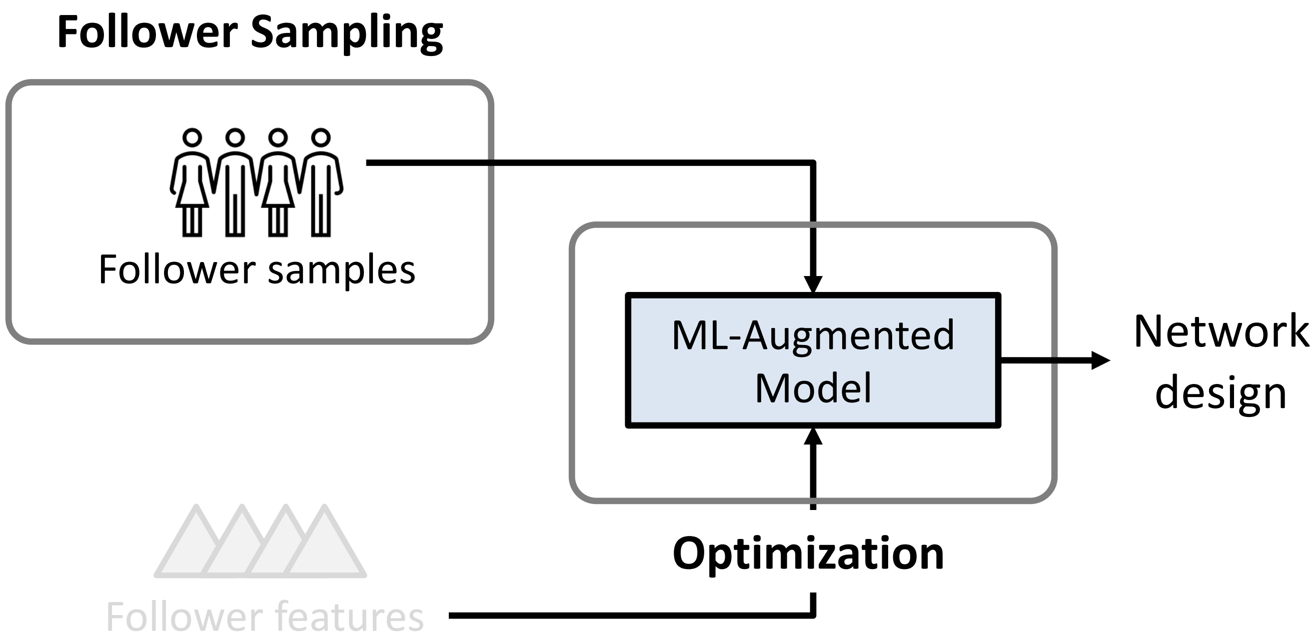
VS

Predict, then optimize (sequentially) — Predict and optimize (simultaneously)

Roadmap

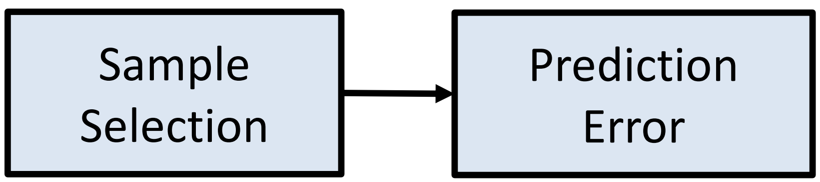

Follower samples
ML-Augmented Model
Network design
Optimization
Follower features

Roadmap


Follower Sampling
Follower samples
ML-Augmented Model
Network design
Optimization
Follower features

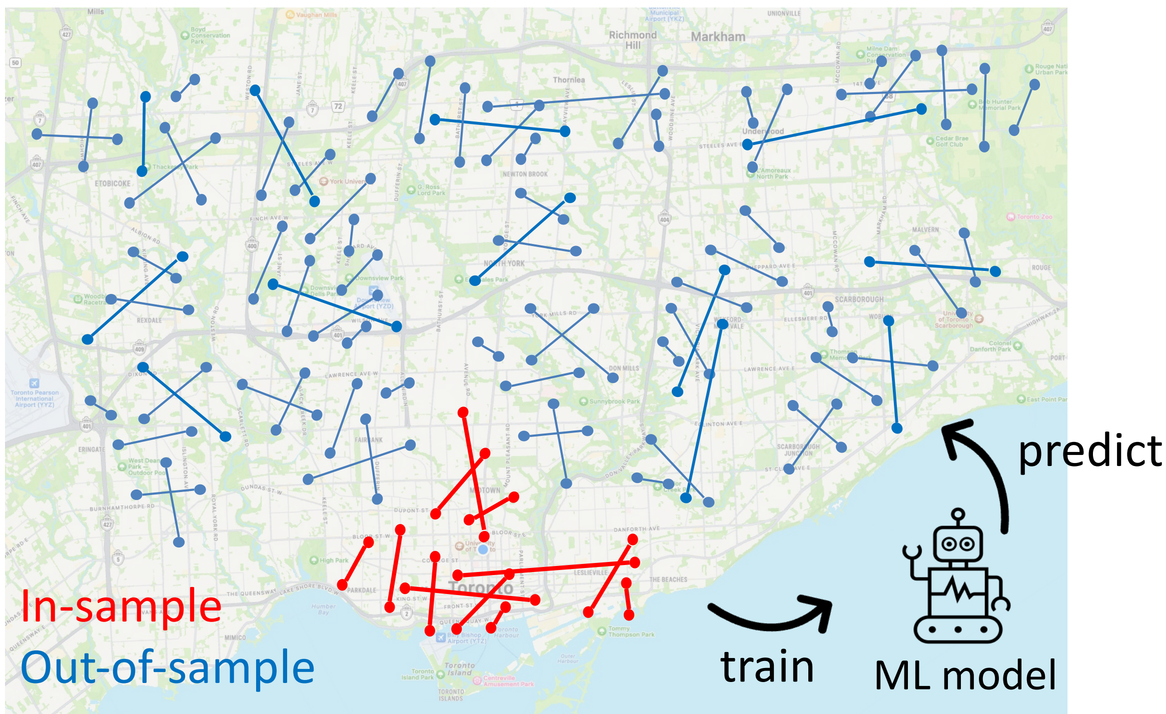
Why is sample selection so important?

Why is sample selection so important?



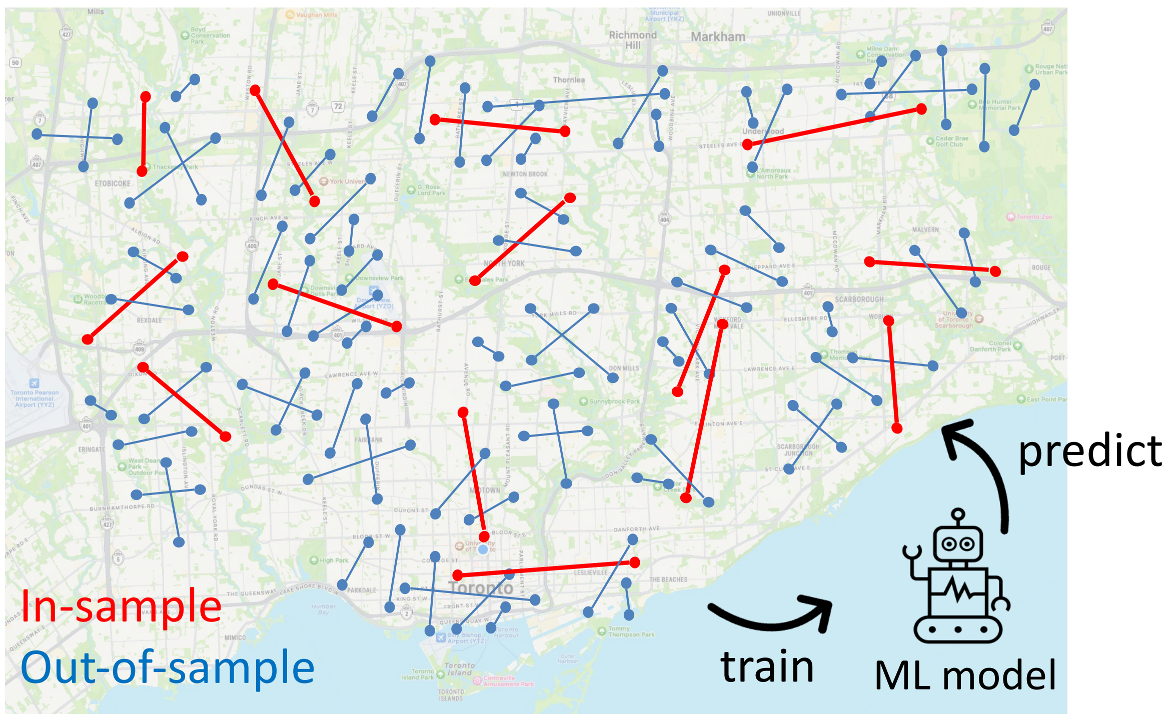
Why is sample selection so important?

Sample Selection → Prediction Error



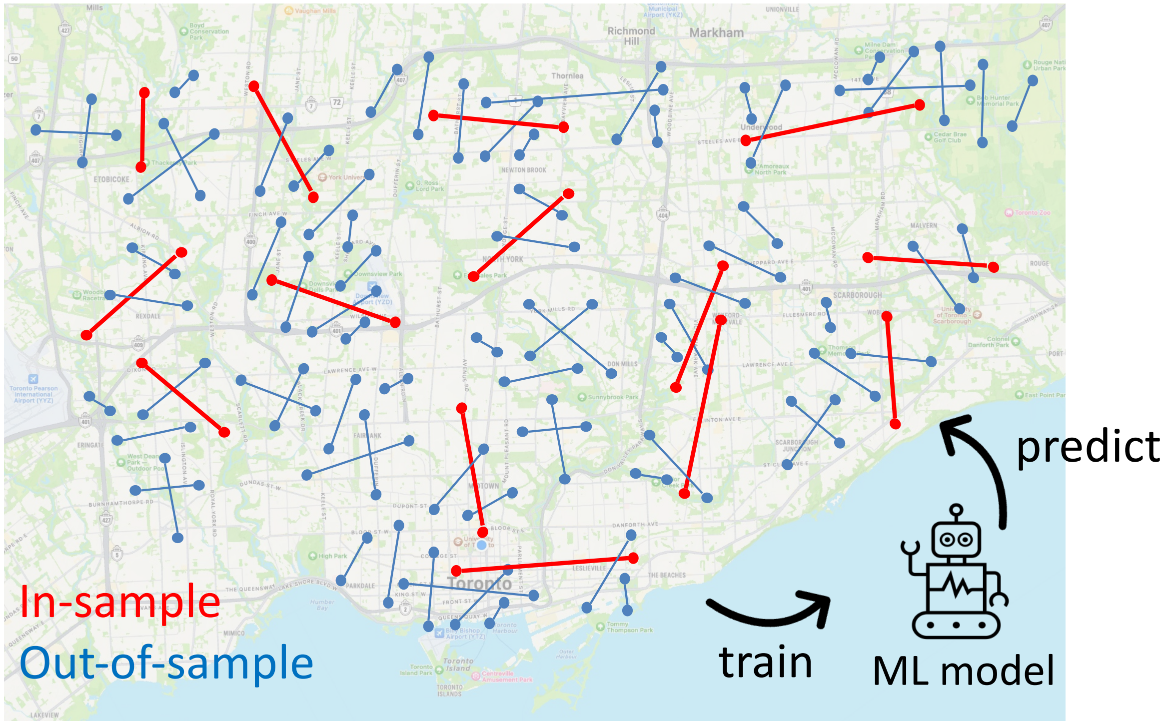
Why is sample selection so important?

Sample Selection → Prediction Error



Why is sample selection so important?

Sample Selection → Prediction Error → Approximation Error → Solution Quality

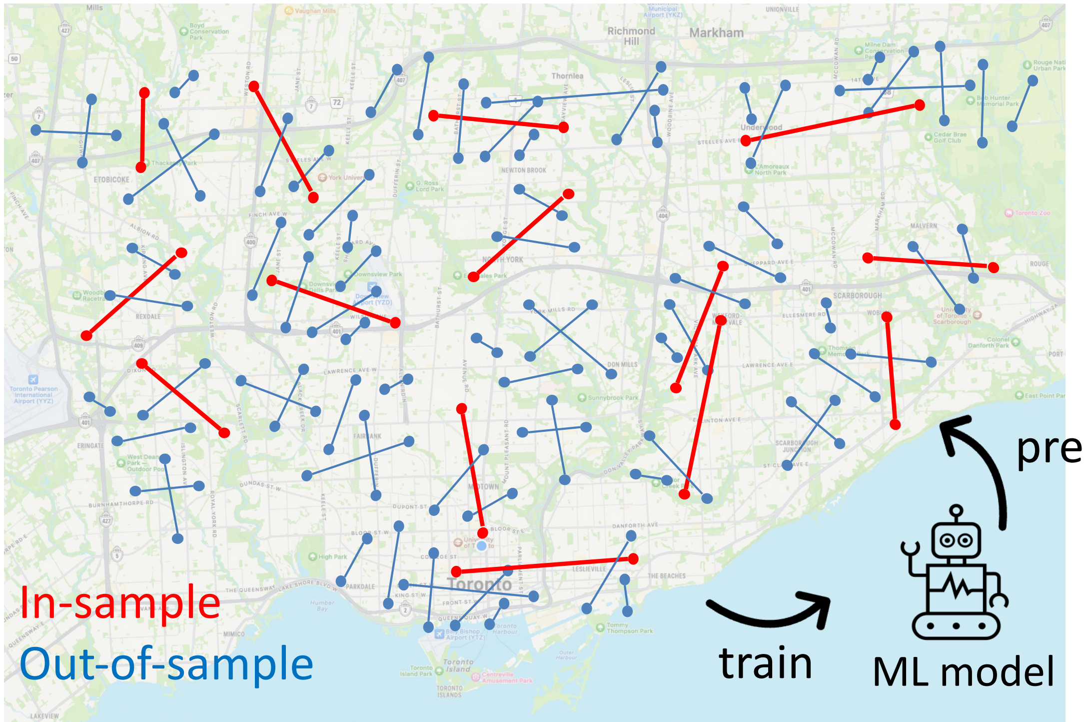


Why is sample selection so important?

Sample Selection → Quantify → Solution Quality

Prediction Error

Approximation Error



How does follower sampling affect solution quality?

Theorem 2 (informal)

optimality gap as evaluated on the original objective $\leqslant E(\mathcal{T})$

sample

How does follower sampling affect solution quality?

Theorem 2 (informal)

optimality gap $\leq$

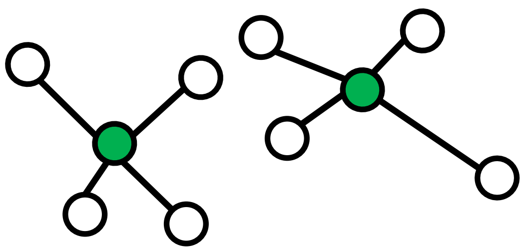
$$2\bar{Q}\bar{L} + 2\bar{Q}(\lambda+\mu)\sum_{s\in\mathcal{F}\setminus\mathcal{T}} d(\mathbf{f}^s, \mathbf{f}^{v(s)}) + \sqrt{4\bar{Q}^2\bar{L}^2\left[|\mathcal{F}\setminus\mathcal{T}| + \sum_{t\in\mathcal{T}}(m^t_{\mathcal{F}\setminus\mathcal{T}})^2\right]\log(1/\gamma)}$$

1st term: **training loss**

2nd term: **distance to the nearest in-sample follower**

3rd term: **# of out-of-sample followers assigned**

Feature Space



○ out-of-sample follower ($F\setminus T$)

● in-sample follower (T)

Follower Sampling Problem

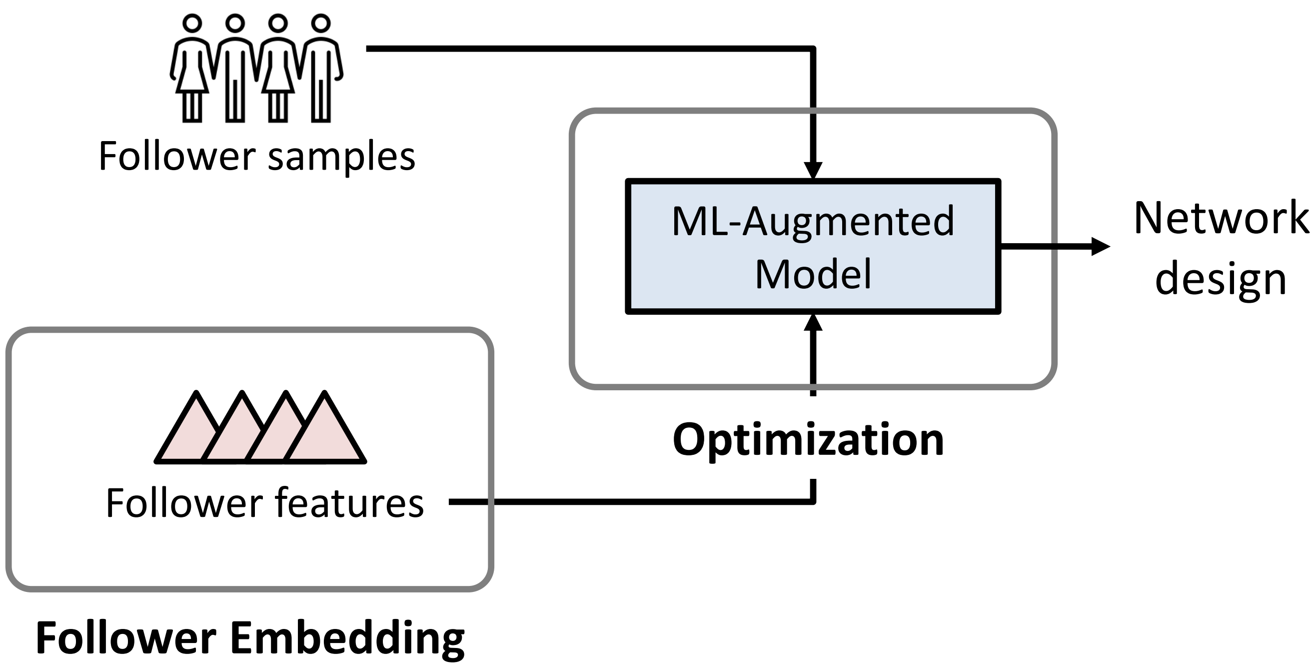
minimize total distance (2nd term)

subject to sample size $\leq p$

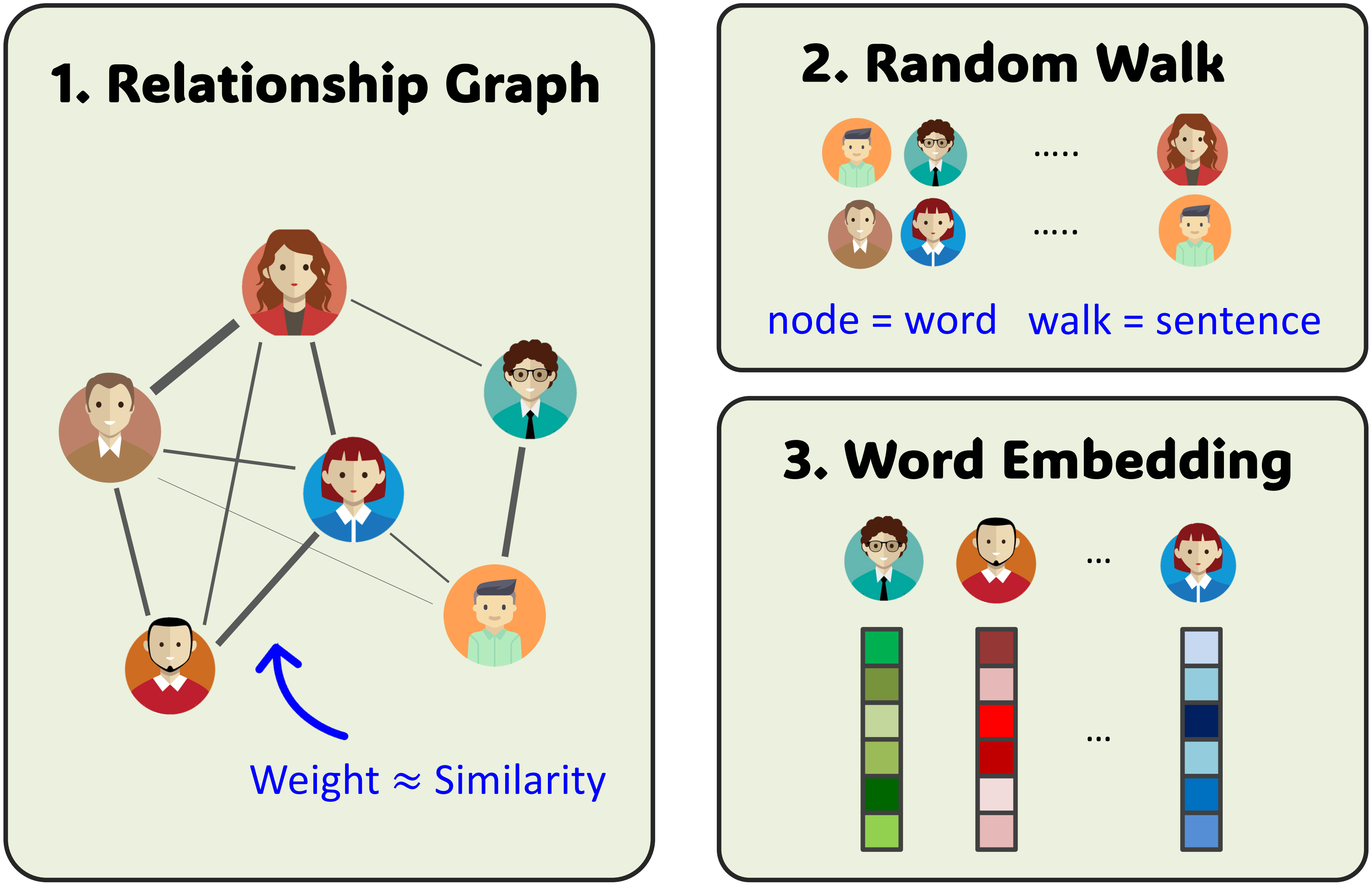
\# assigned to each $\leq d$ (3rd term)

Balanced p-median problem

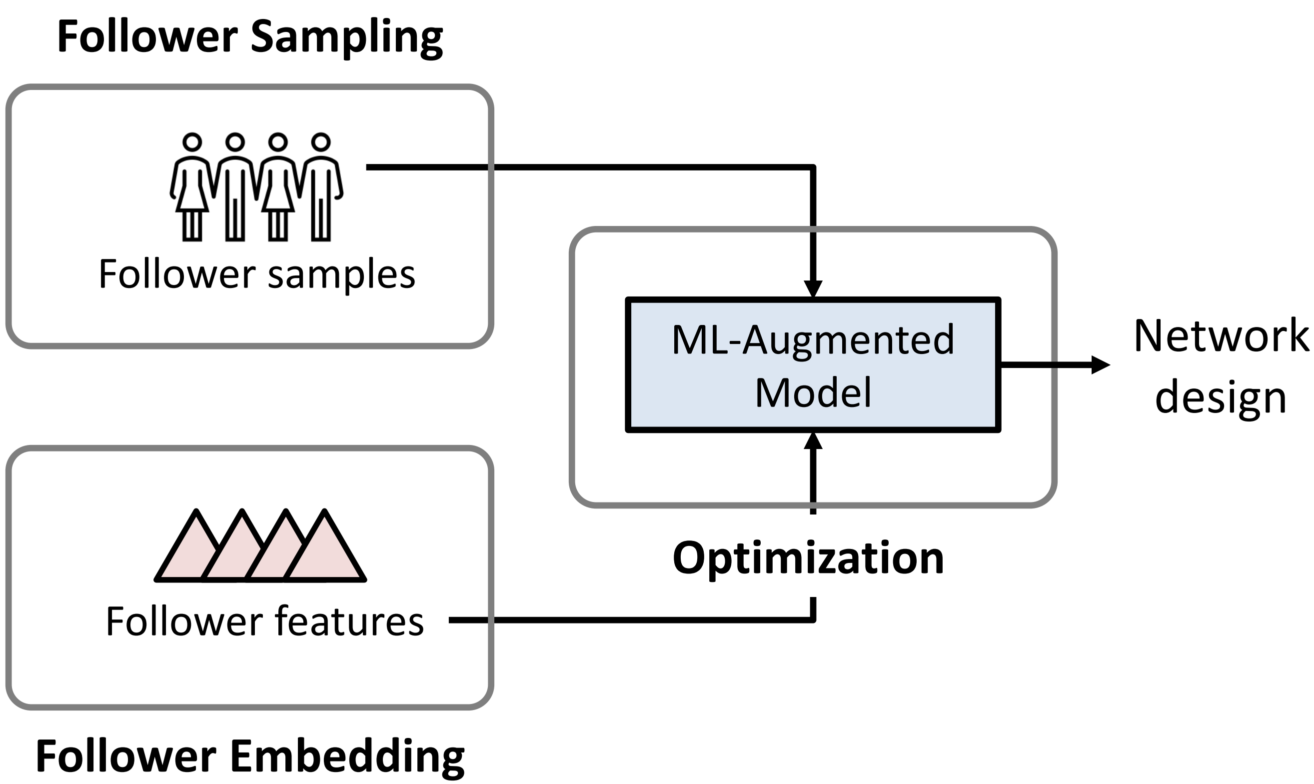
Roadmap


Follower samples
ML-Augmented Model
Network design
Optimization
Follower features
Follower Embedding

Follower embedding



Roadmap

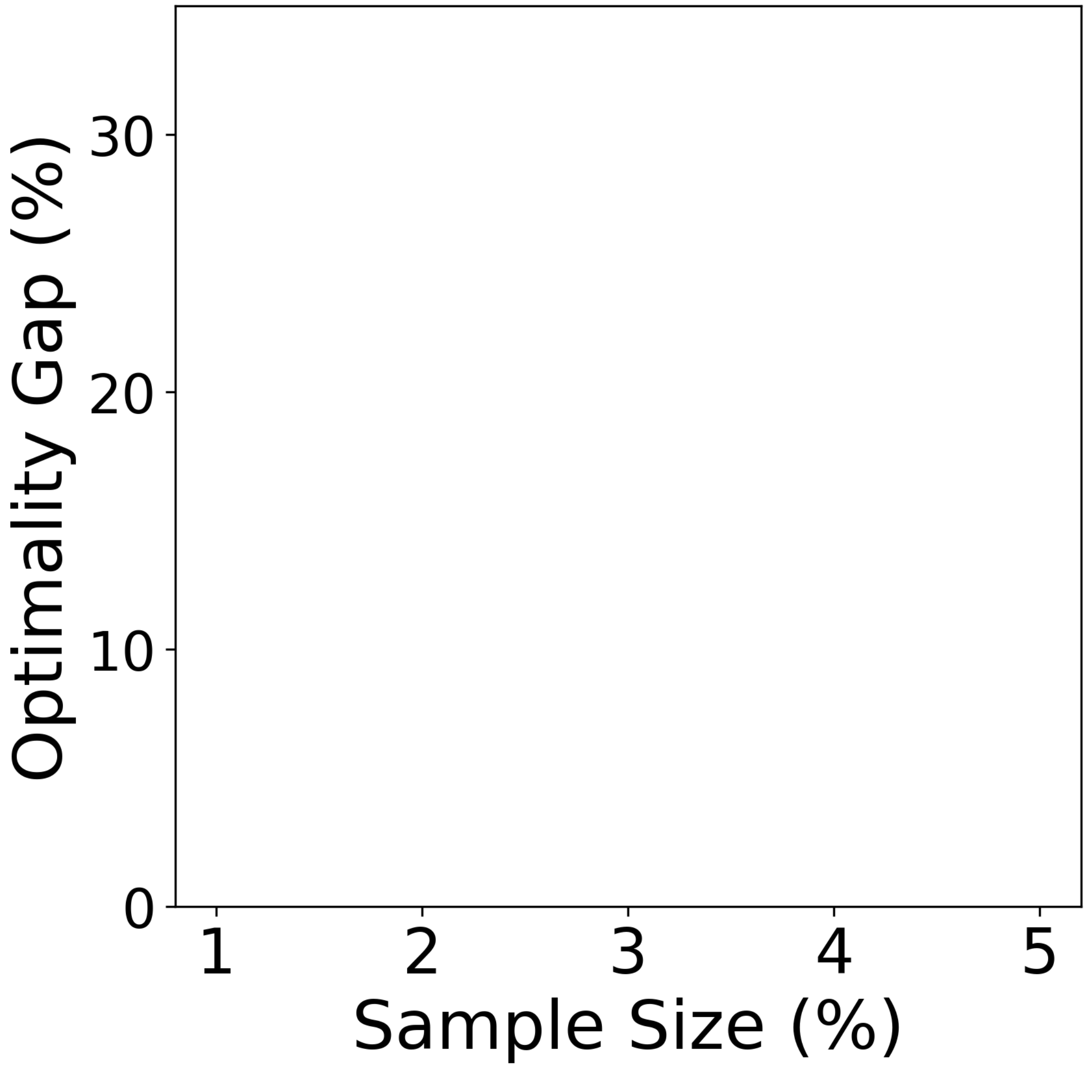

Follower Sampling
Follower samples
ML-Augmented Model
Network design
Optimization
Follower features
Follower Embedding

Synthetic Instances: Solution Quality

- 3,526 OD pairs, 1824 edges, 373 nodes

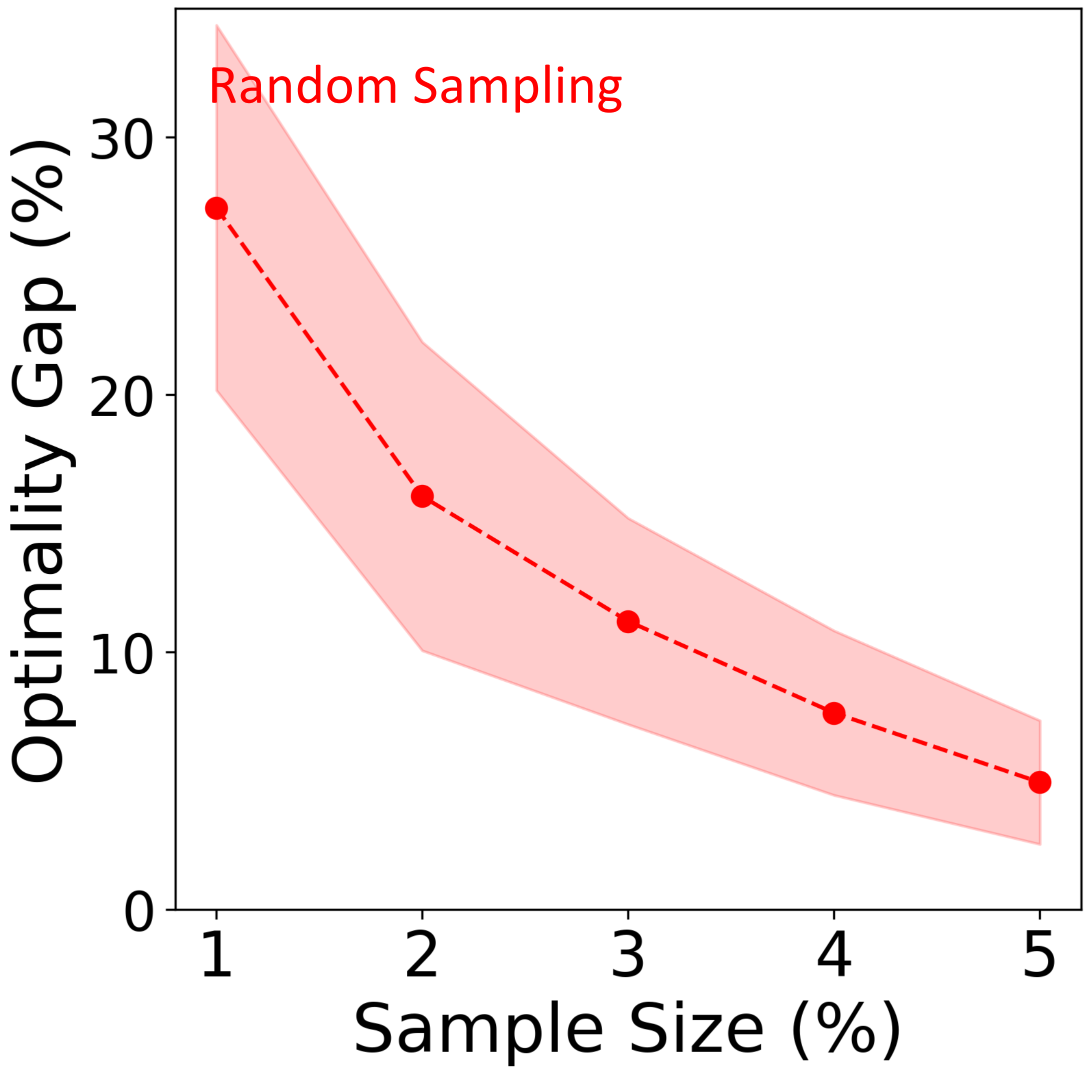
Synthetic Instances: Solution Quality

- 3,526 OD pairs, 1824 edges, 373 nodes



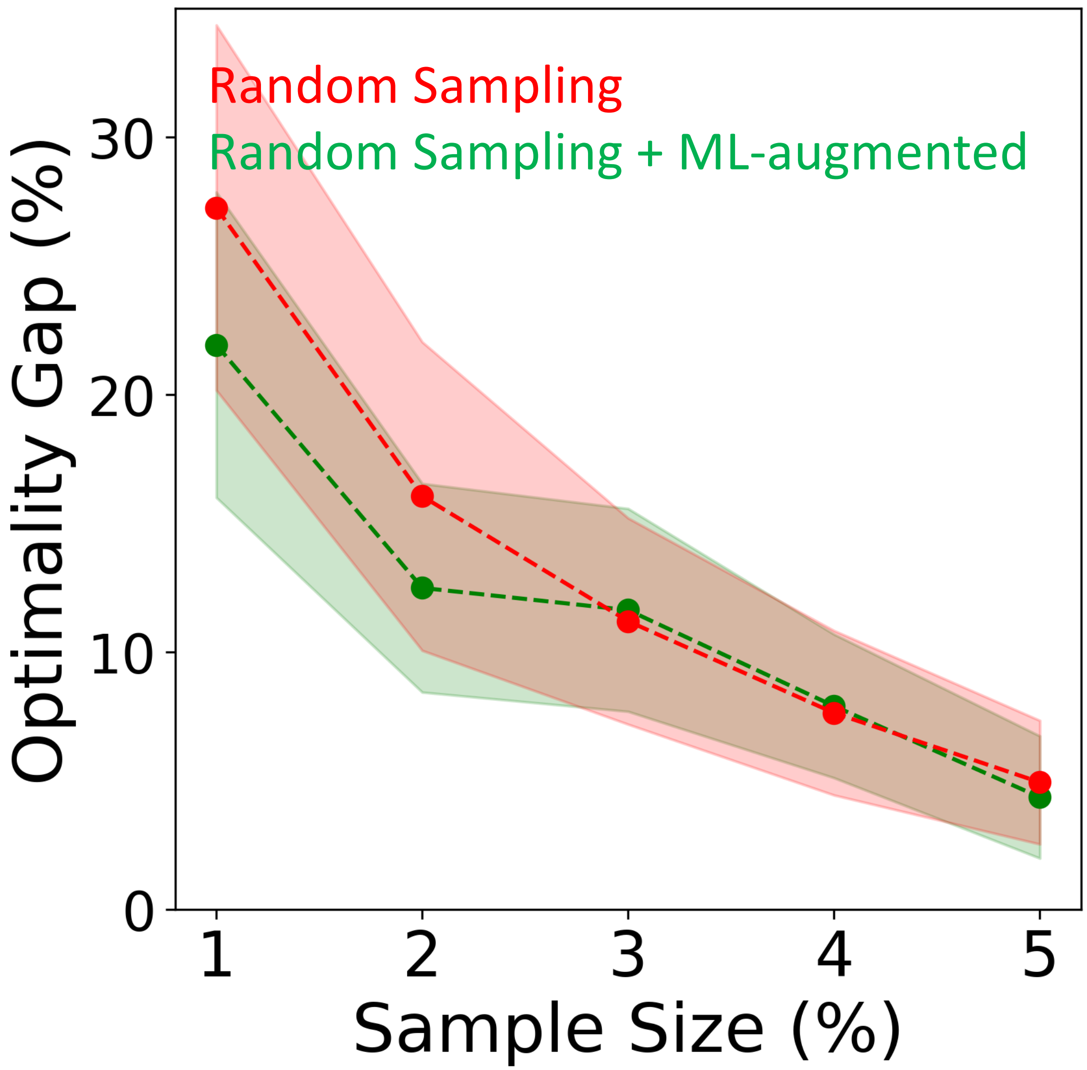
Synthetic Instances: Solution Quality

- 3,526 OD pairs, 1824 edges, 373 nodes



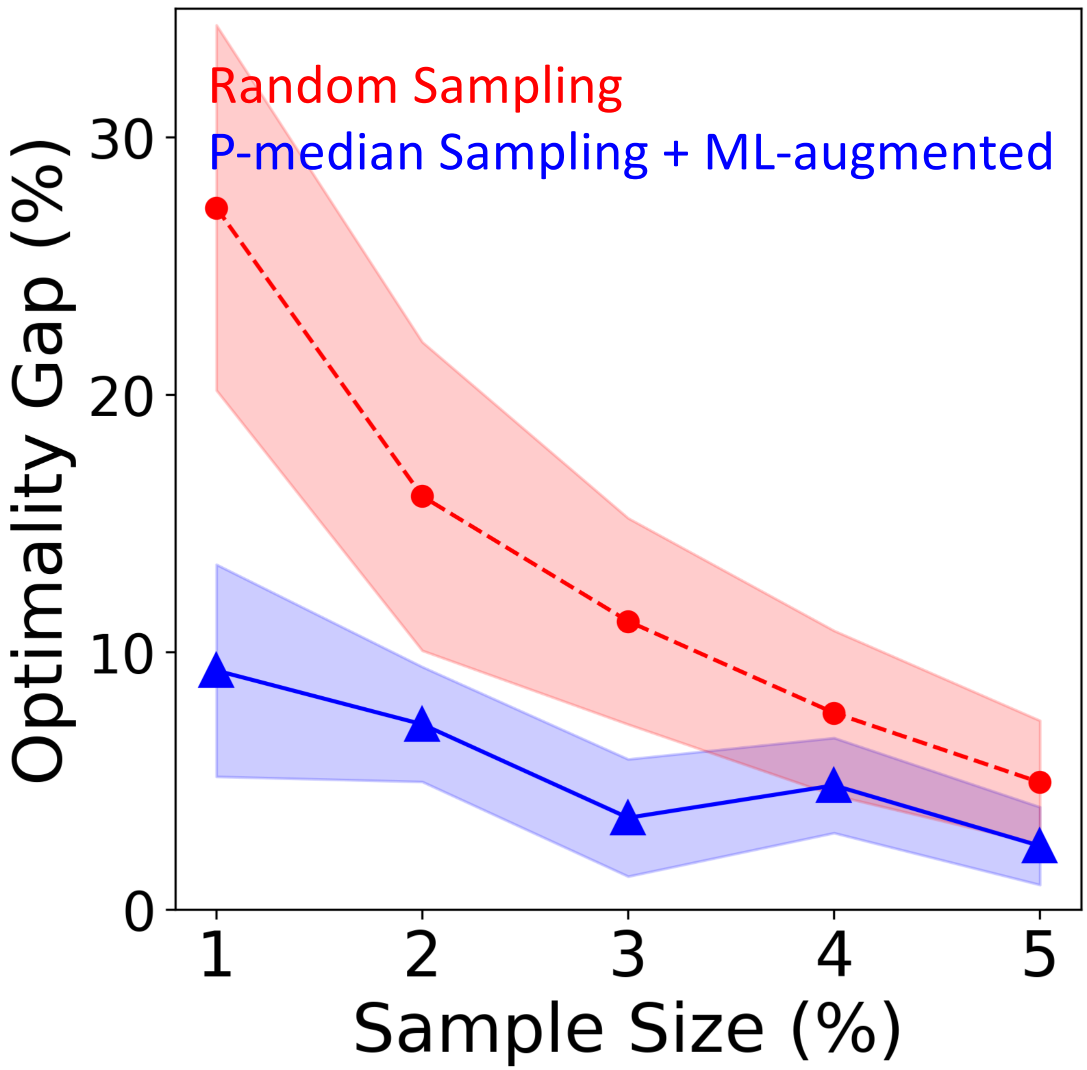
Synthetic Instances: Solution Quality

- 3,526 OD pairs, 1824 edges, 373 nodes



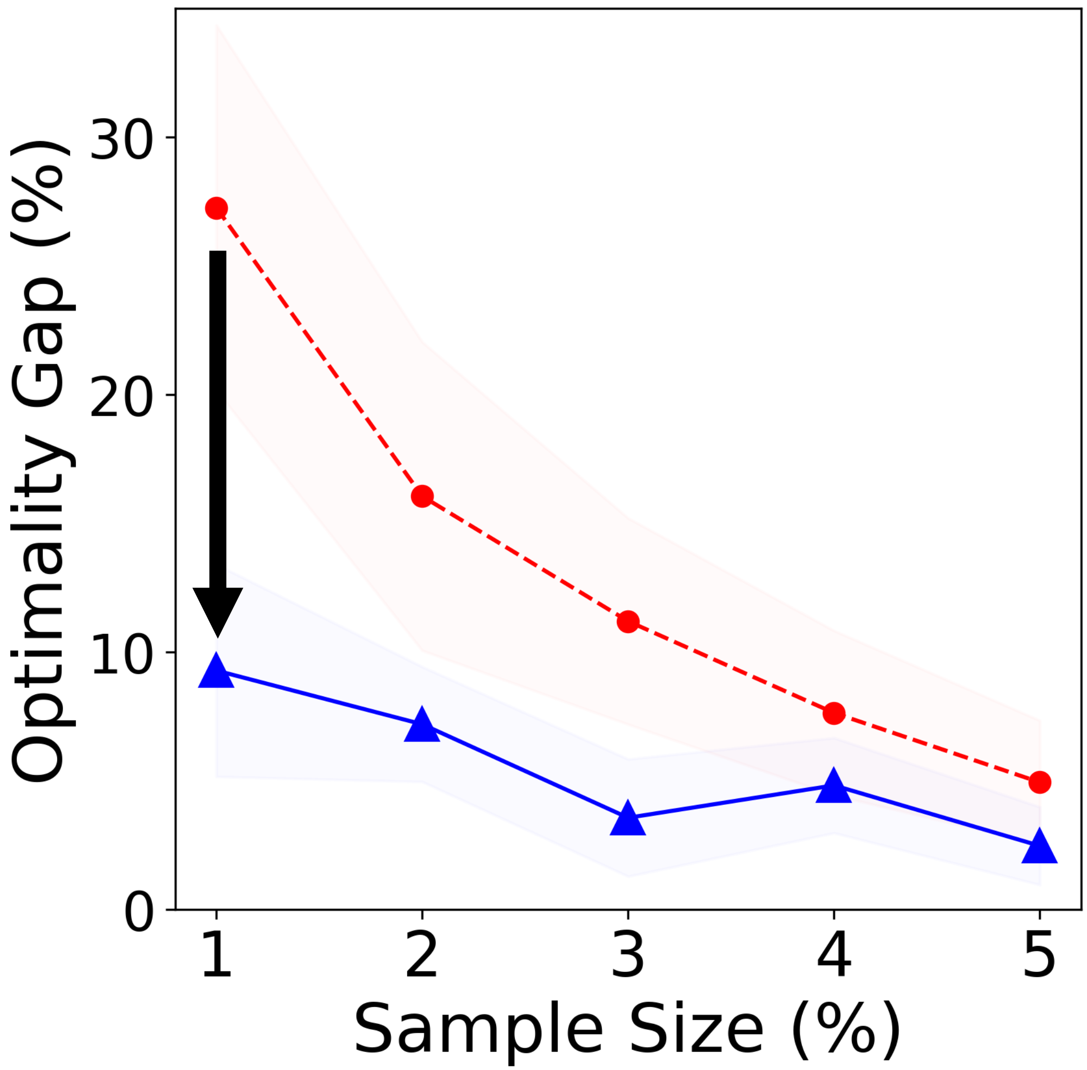
Synthetic Instances: Solution Quality

- 3,526 OD pairs, 1824 edges, 373 nodes

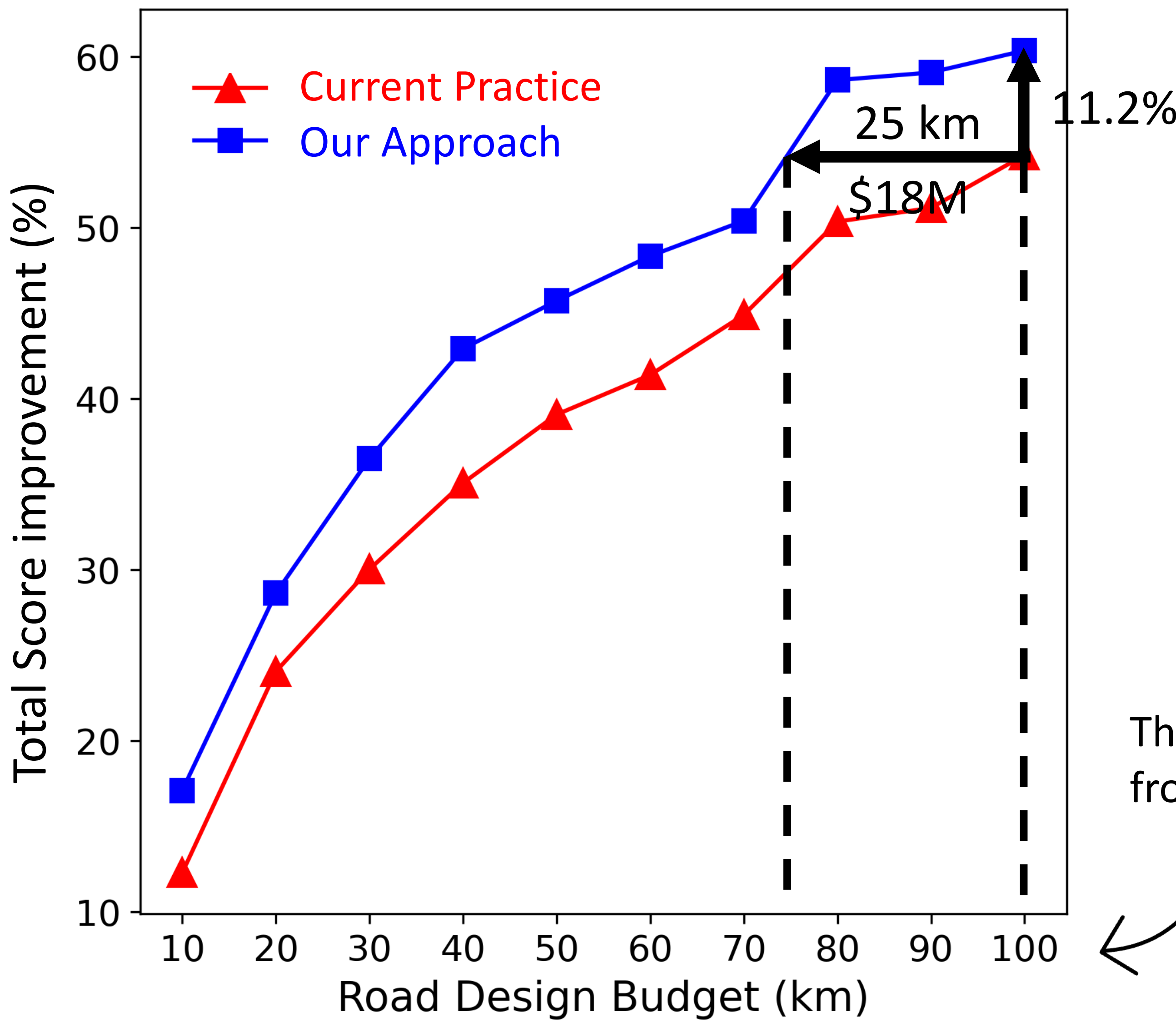


Synthetic Instances: Solution Quality

- 3,526 OD pairs, 1824 edges, 373 nodes



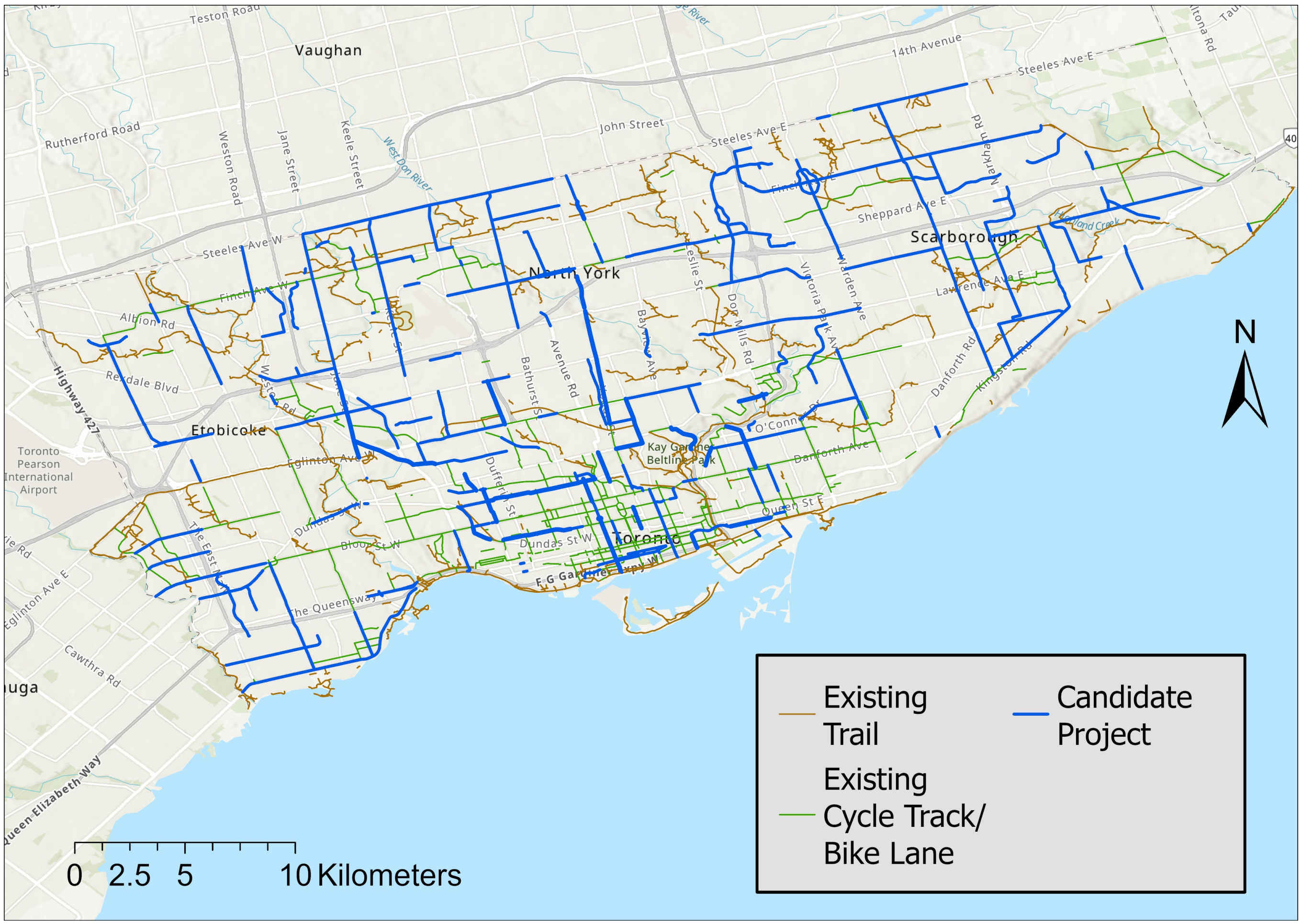
Strong performance: 2022 – 2024 planning horizon



The City's budget from 2022 – 2024

2025 – 2029 Network Expansion

- Candidate projects (339.3 km)



2025 – 2029 Network Expansion

- Two-phase budget structure
  - Phase I (2025—2027): 45 km
  - Phase II (2028—2029): 45 km

“Short-term” **“Long-term”**

Aligned?

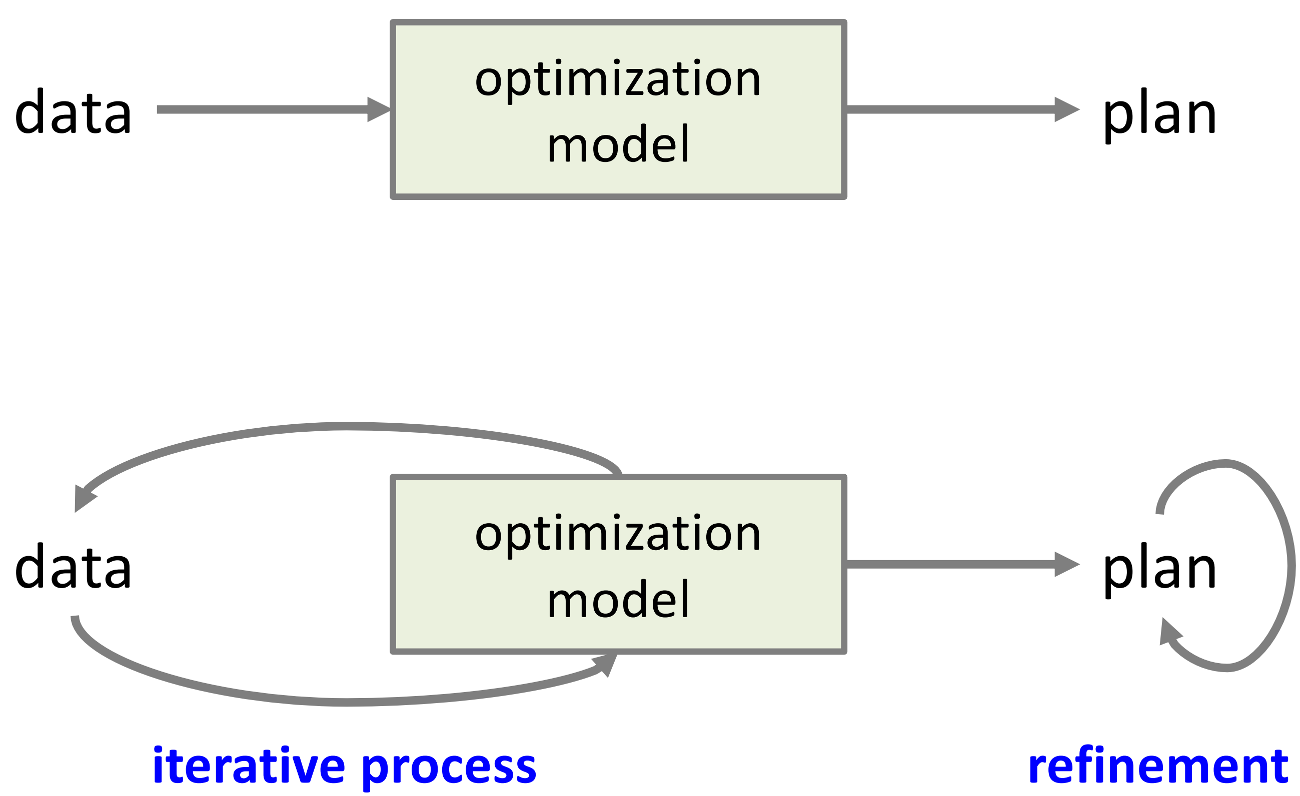
The short- and long-term objectives are largely aligned

Final implementation

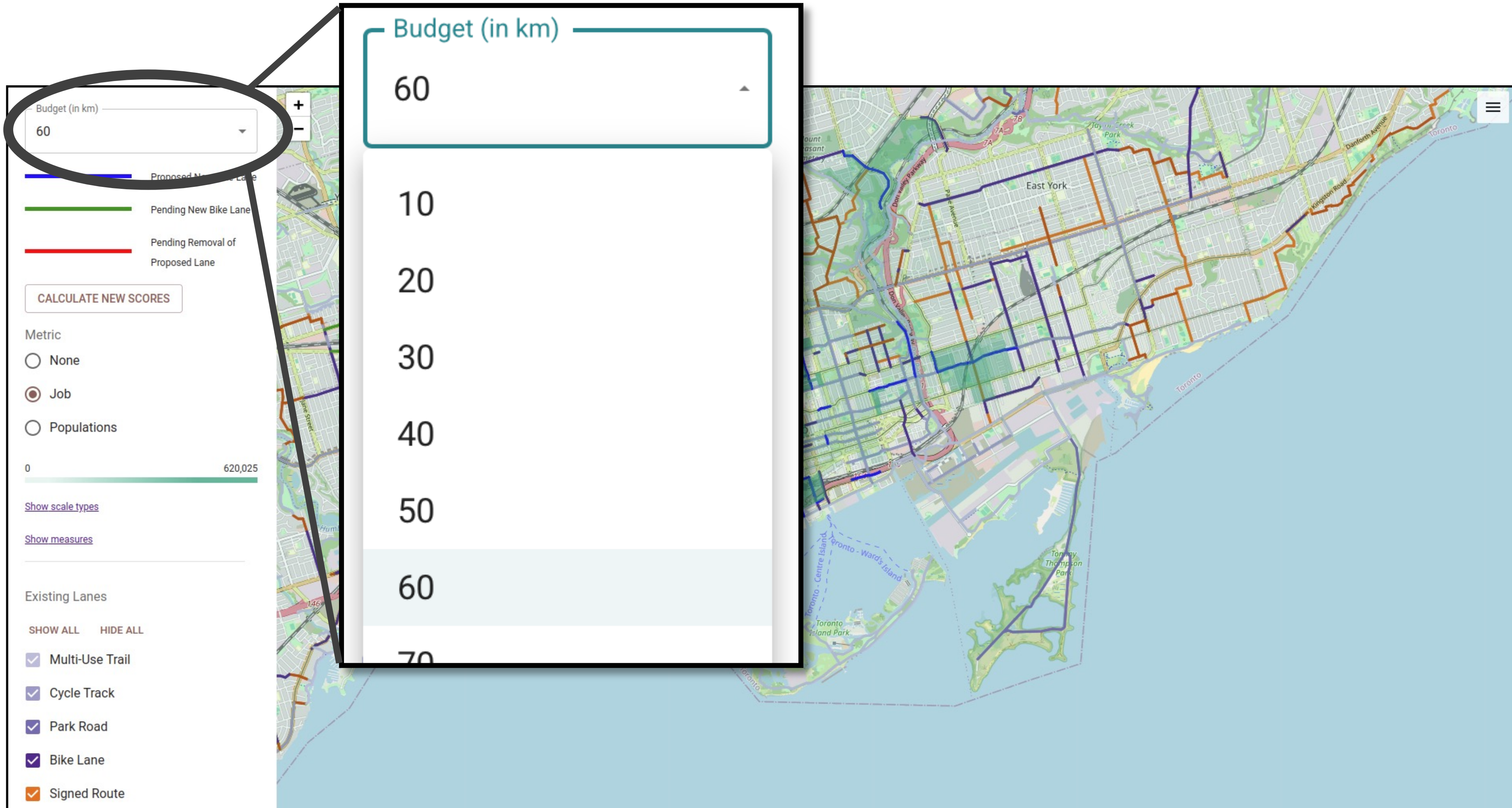
- Connectivity scores (5 points)
  - Appeared in both solutions (5 points)
  - Appeared in only one solution (4 points)
  - Others (0 point)
- Safety (5 points)
- Transit Access (5 points)
- …

29 km approved 2025 – 2027 **28 km** under study/design 2028 – 2030 **=** **$40M** Investment

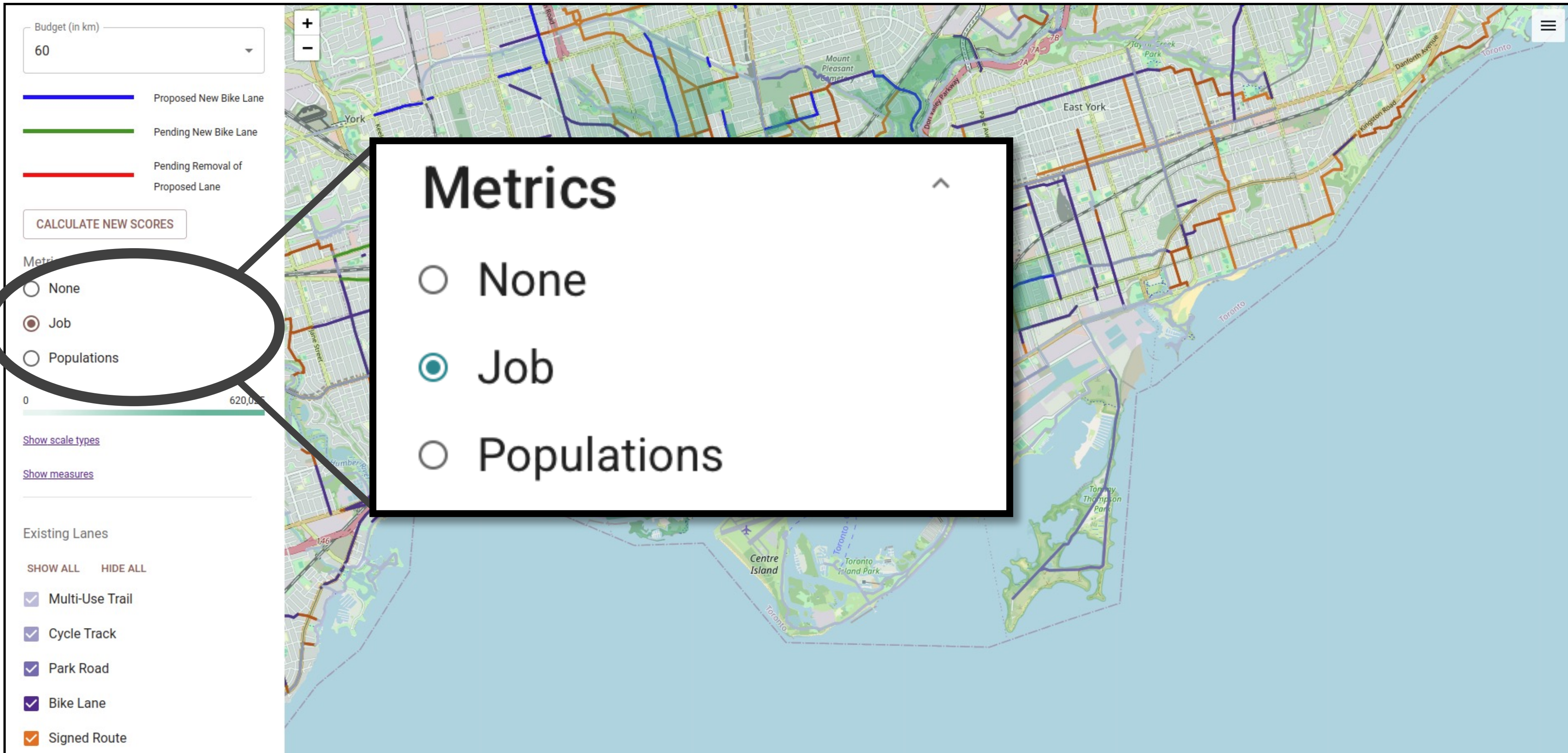
Real-world decision making isn't one-shot optimization



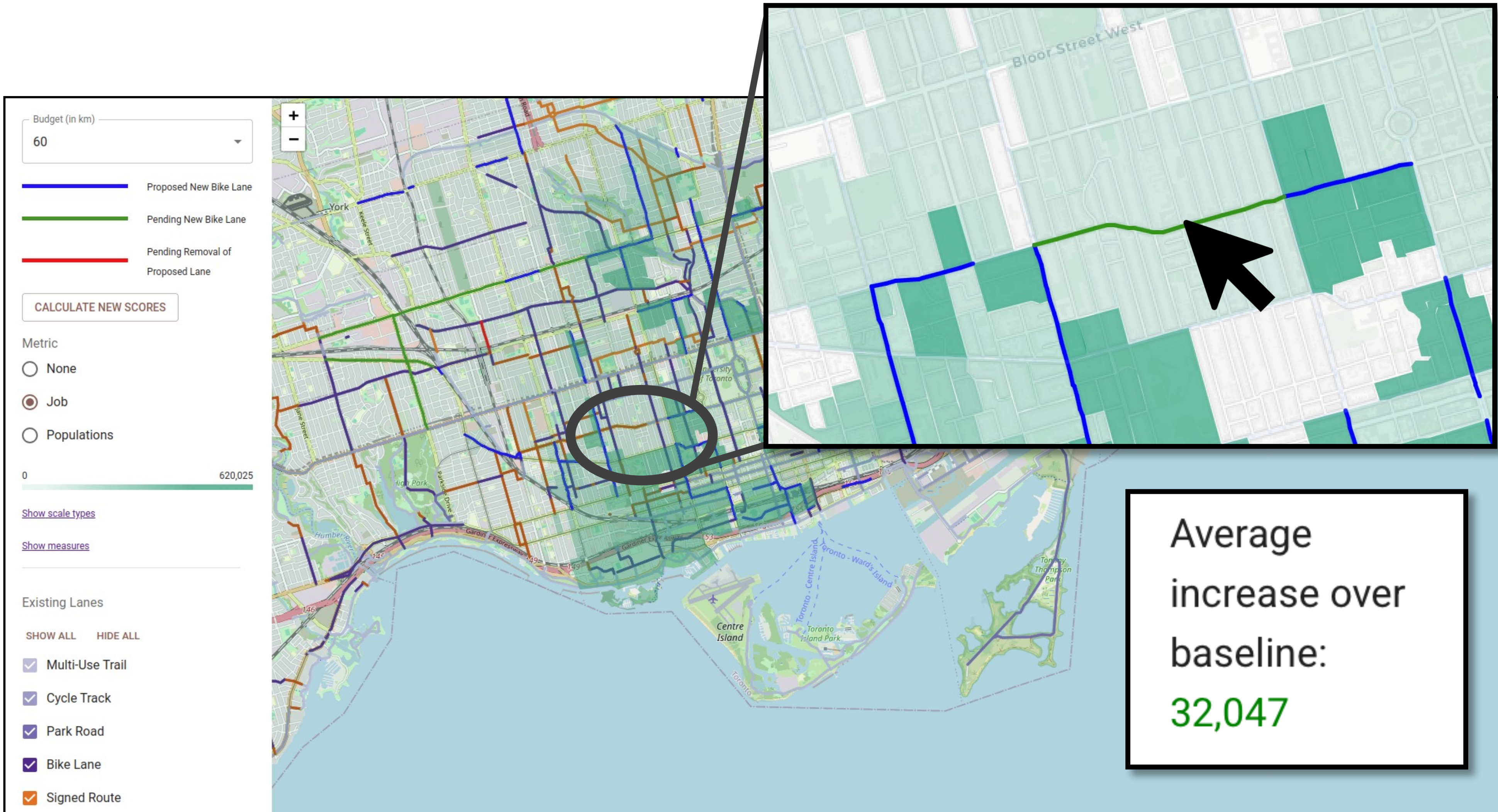
CycleLinx: optimizing with humans, not for them



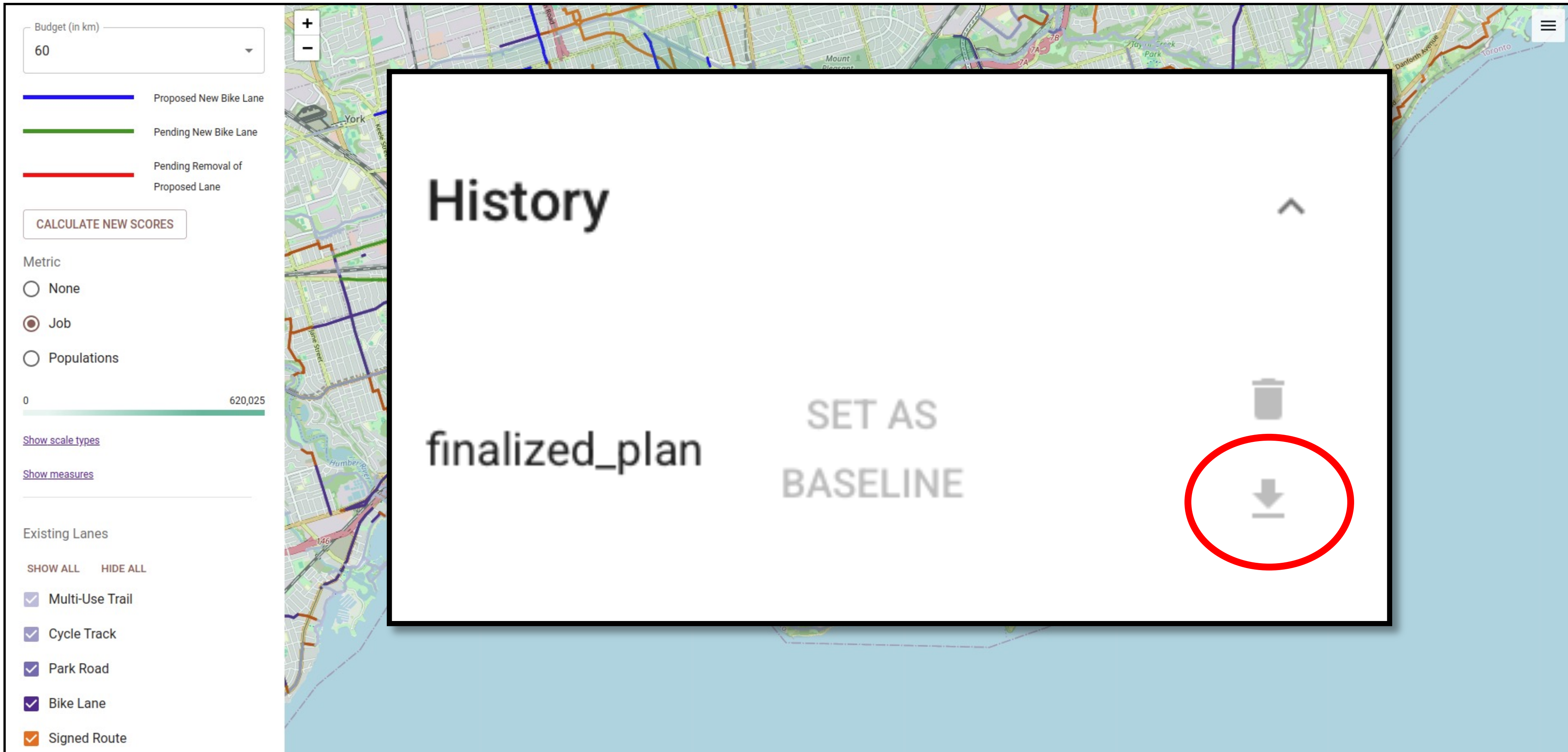
CycleLinx: optimizing with humans, not for them



CycleLinx: optimizing with humans, not for them



CycleLinx: optimizing with humans, not for them



An end-to-end, data-driven solution

40 km
Assessed

Widely cited

CBC news · U of T News · TechXplore · TORONTO STAR

Only research team presented at

Toronto City Budget Meeting
Legislative Assembly of Ontario

+11.2%	**-25%**	**$18M**	**29 km**	**28 km**
accessibility	**length**	**cost saving**	**approved**	**under study**
	2022 – 2024		**2025 – 2027**	**2028 – 2029**

Deployed in
Toronto

Rolling out in
Hamilton
& Mississauga

Assessment Framework

AutoLTS
Cycling Stress Assessment

→ **MaxANDP**
Cycling Network Expansion

→ **CycleLinx**
Solution Refinement & Visualization