

Analytics for Better Urban Cycling

Bo Lin

(soon) National University of Singapore

eMERGE Seminar, UC Berkeley

March 2nd, 2026

Cycling has become increasingly popular

3 GOOD HEALTH
AND WELL-BEING



11 SUSTAINABLE CITIES
AND COMMUNITIES



13 CLIMATE
ACTION



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



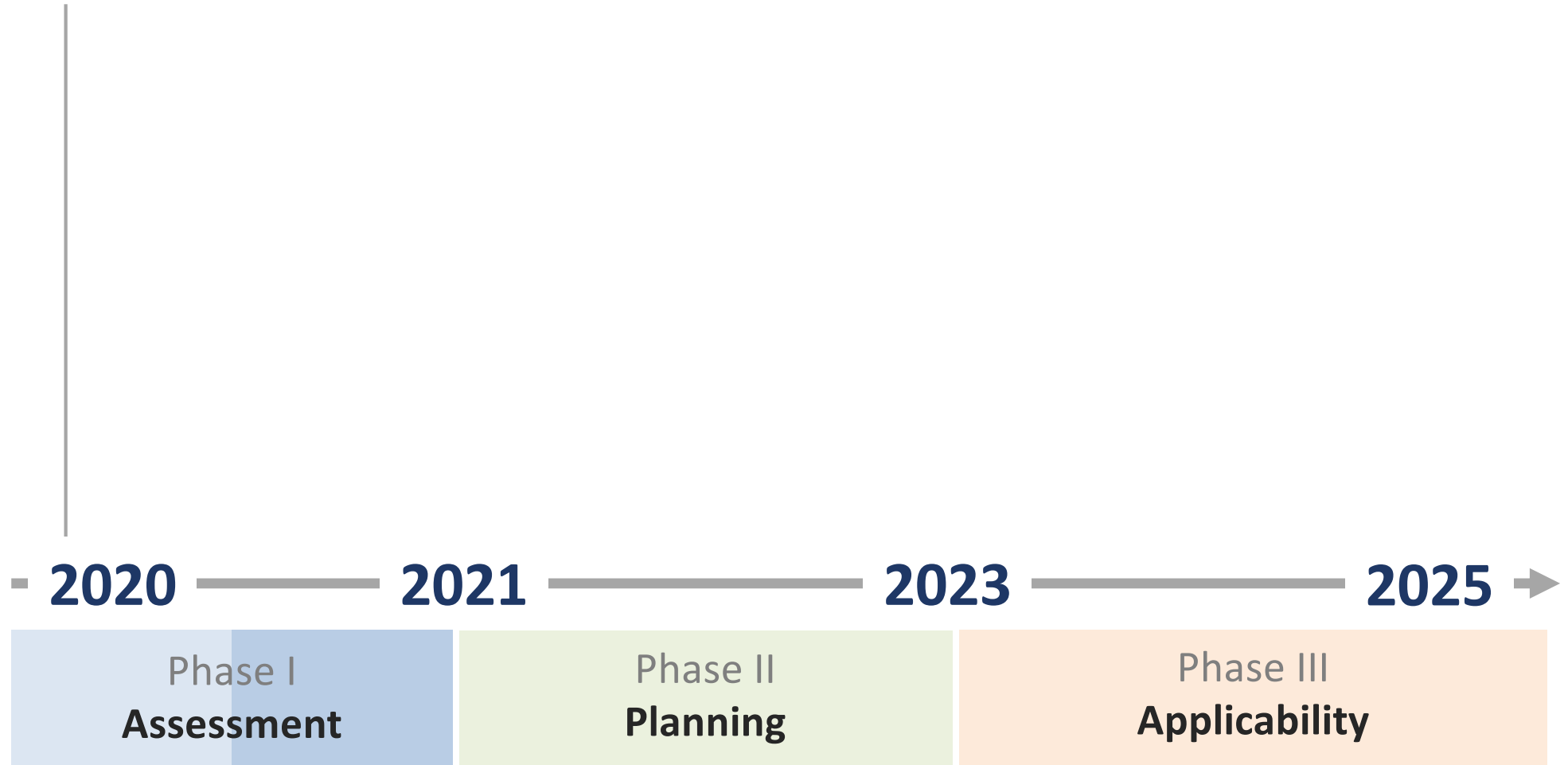
Laura Hanrahan | Dec 7 2021, 10:19 am



Analytics for better urban cycling

- 2020 — 2021 — 2023 — 2025 →

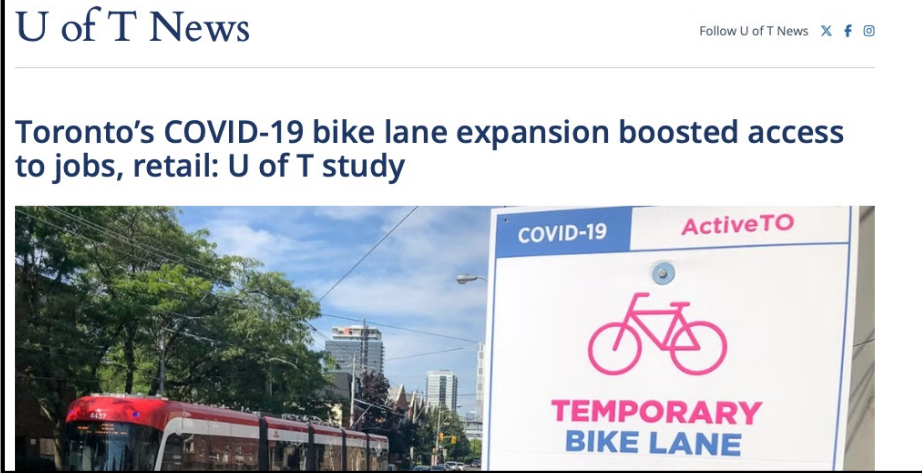
Analytics for better urban cycling



Analytics for better urban cycling

NSERC
Alliance
Grants

Assessment
framework



2020

2021

2023

2025 →

Phase I
Assessment

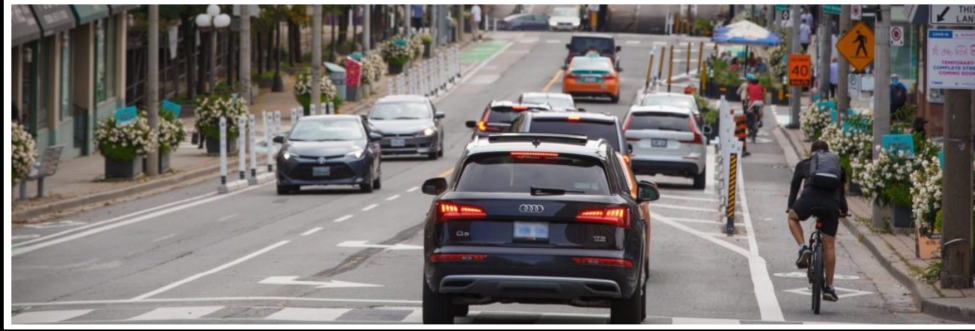
Analytics for better urban cycling

NSERC
Alliance
Grants

Assessment
framework

Yonge Street
(implemented)

ActiveTO Midtown Yonge Complete Street



2020

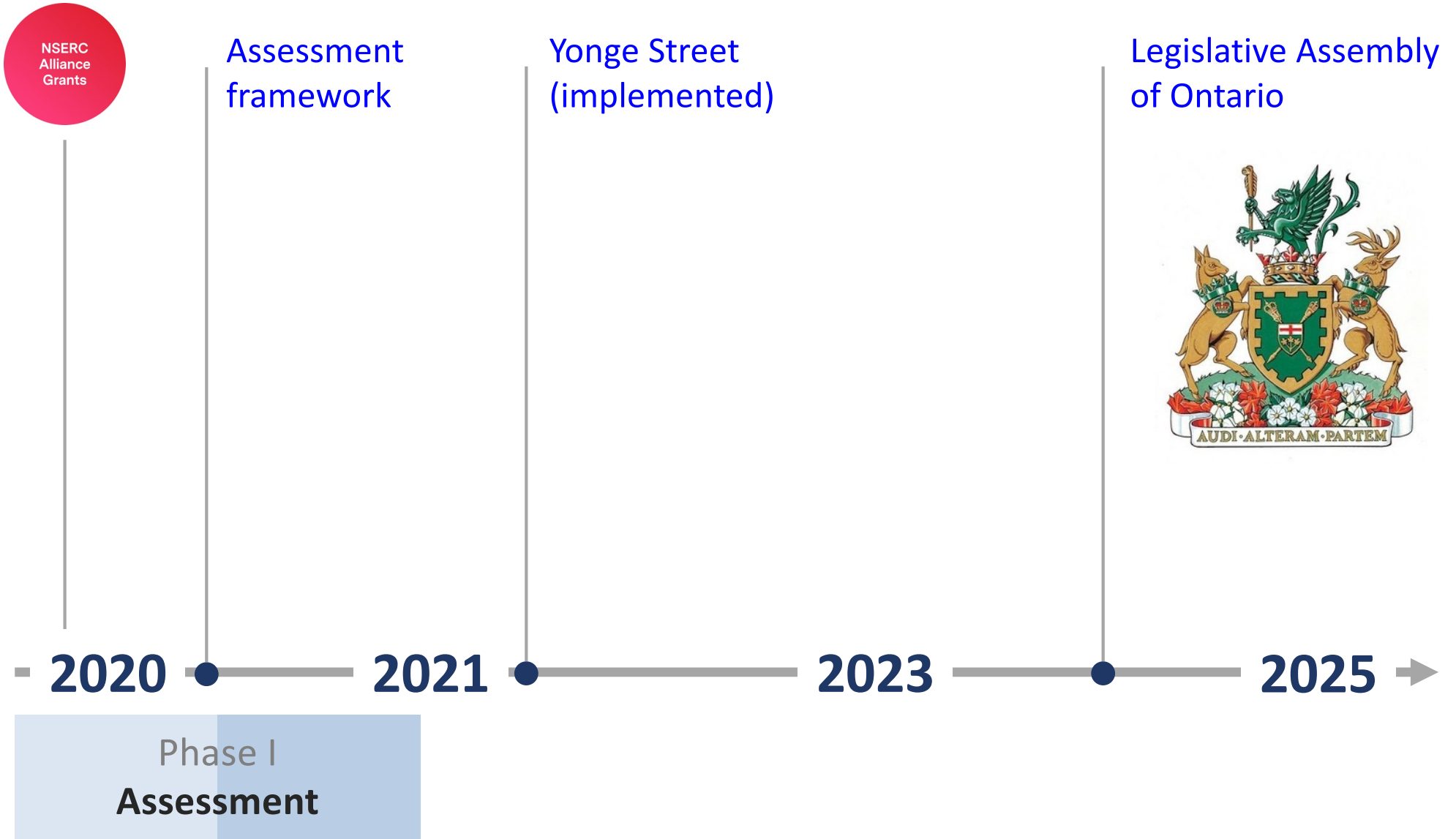
2021

2023

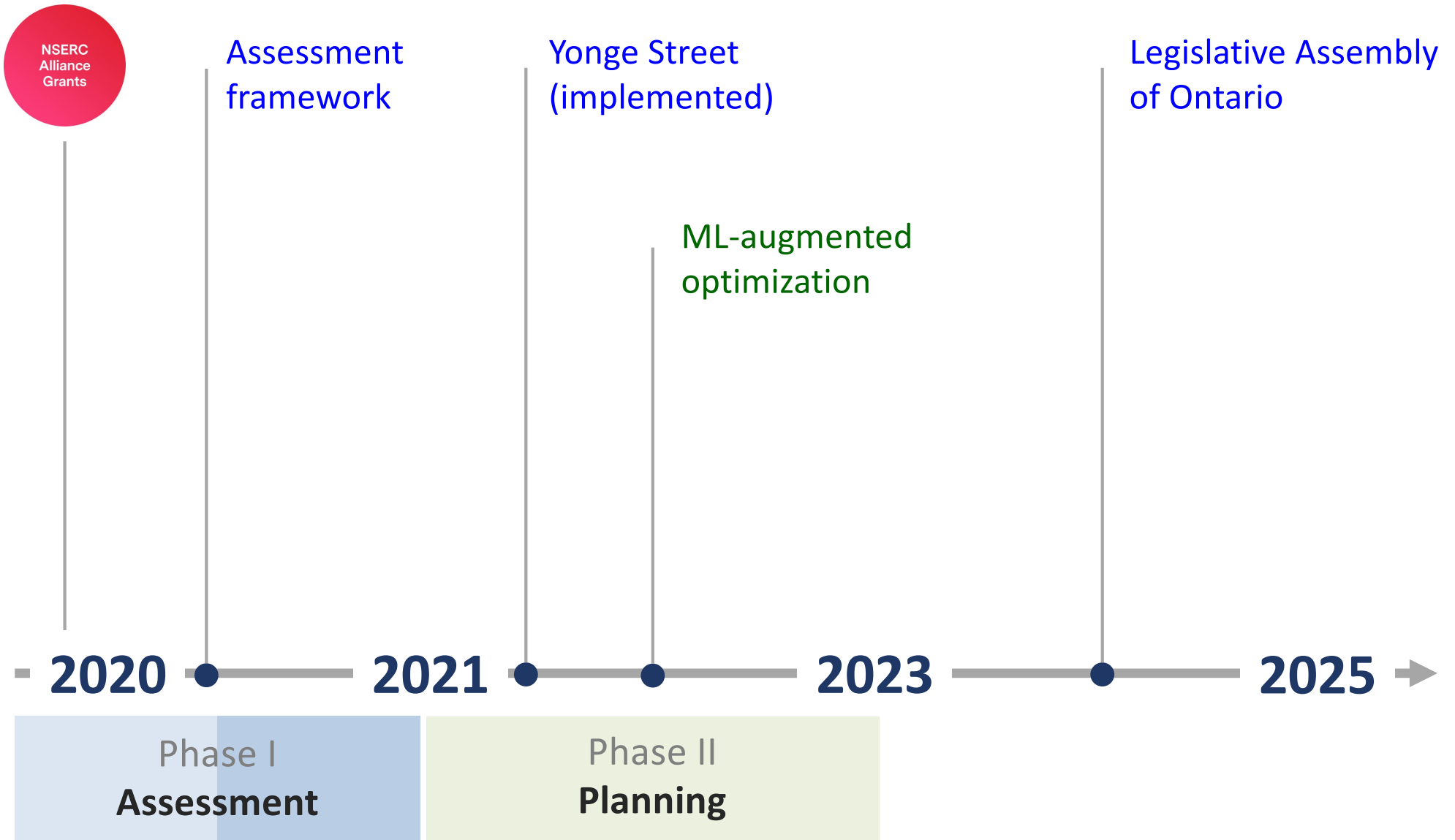
2025

Phase I
Assessment

Analytics for better urban cycling



Analytics for better urban cycling



Analytics for better urban cycling



Assessment framework

Yonge Street (implemented)

Legislative Assembly of Ontario

Cycling Network Plan

The Cycling Network 2025-2027 Implementation Program was adopted by Toronto City Council on June 26, 2024. Learn more about the analyses and recommended bikeway projects under the tabs below, and about the [public input](#).

Toronto's 2025—2029 plan

2020 2021 2023 2025 →

Phase I
Assessment

Phase II
Planning

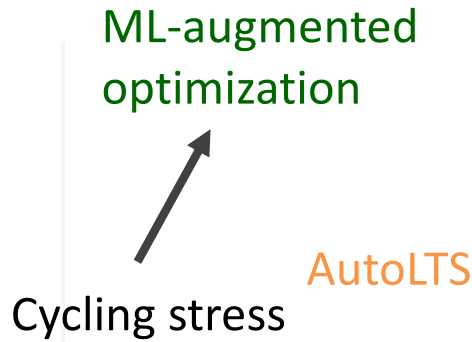
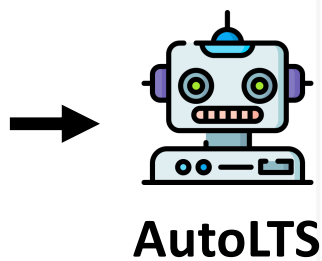
Analytics for better urban cycling



Assessment framework

Younge Street (implemented)

Legislative Assembly of Ontario



Toronto's 2025—2029 plan

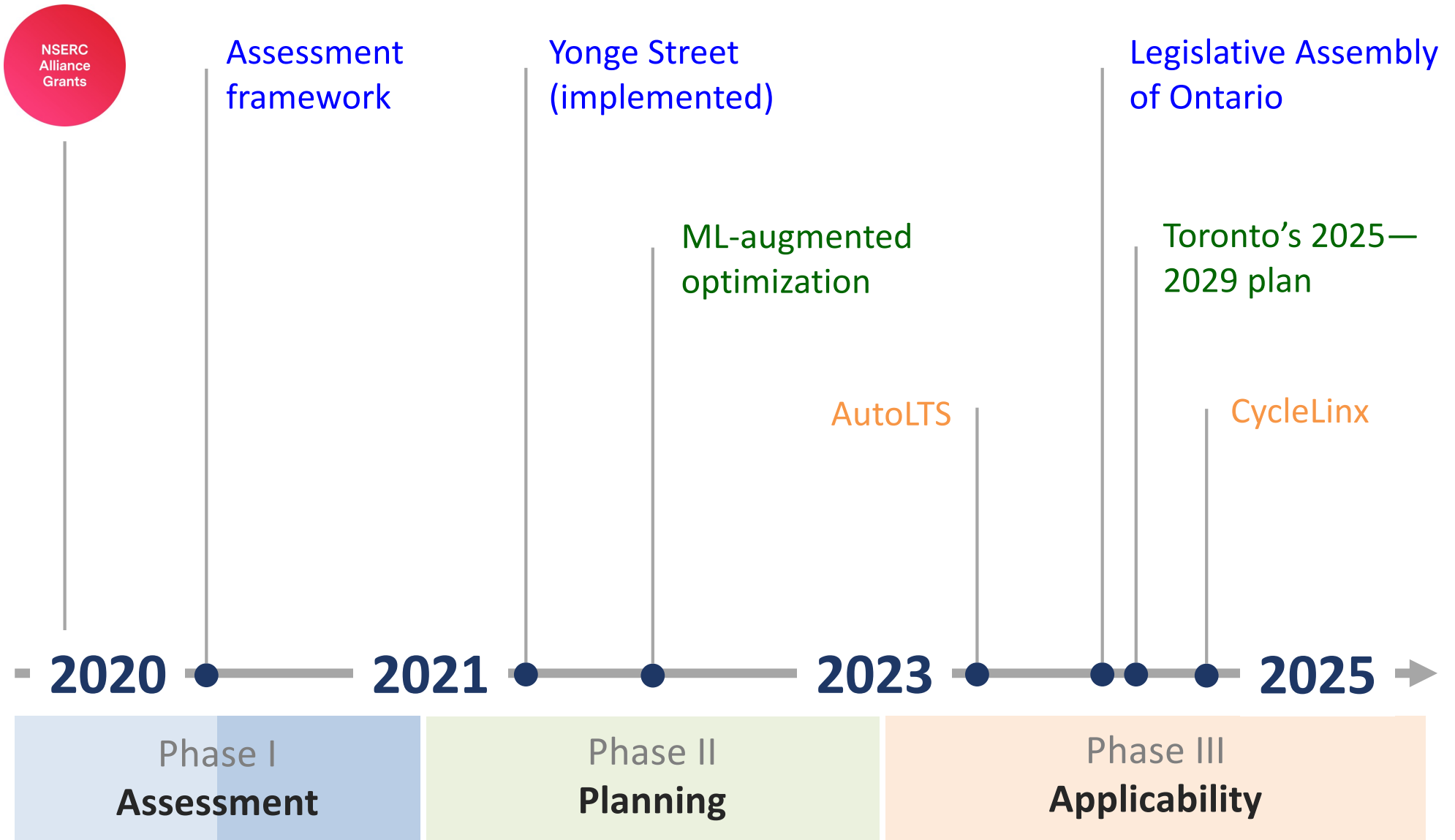


Phase I
Assessment

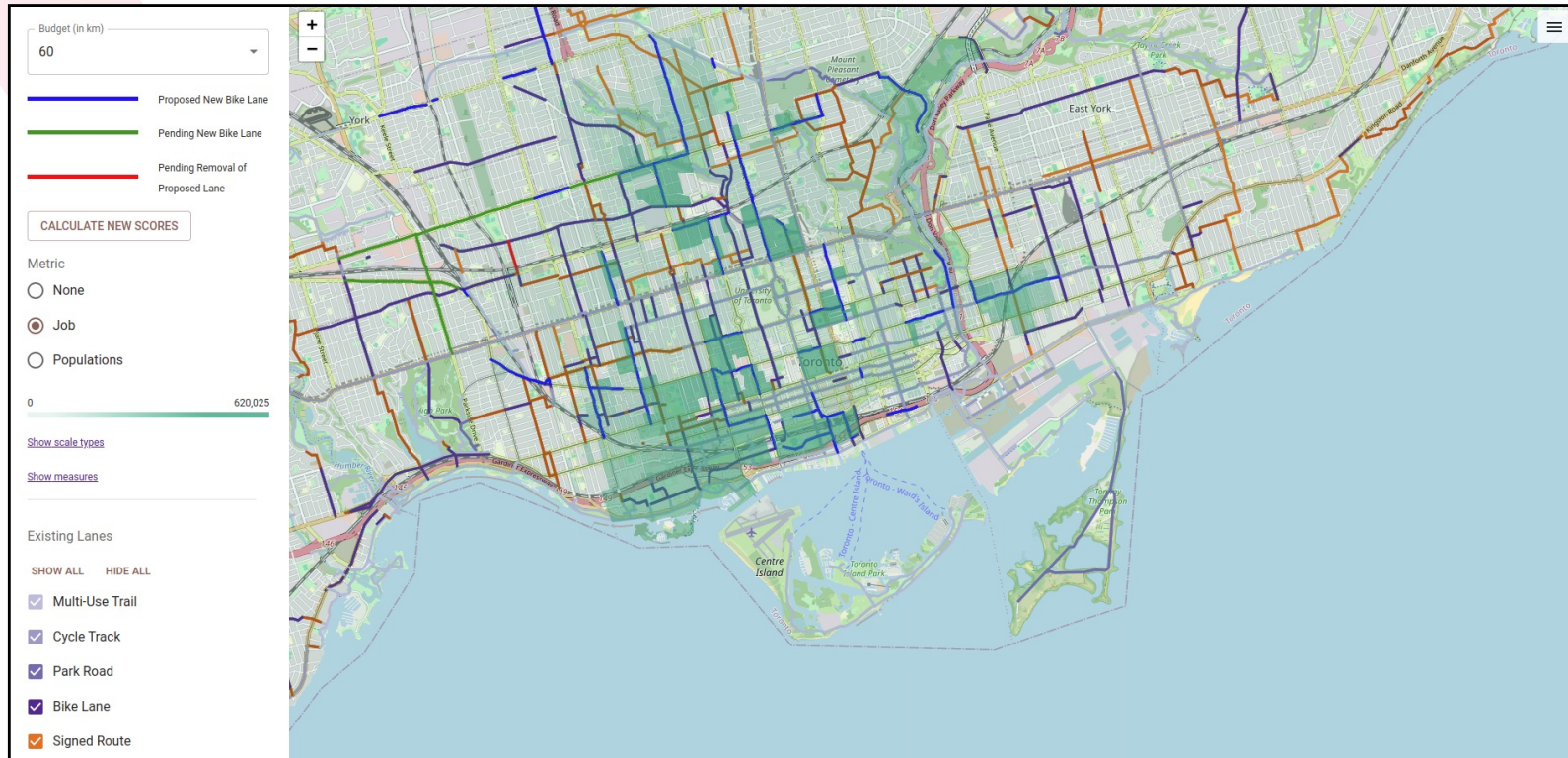
Phase II
Planning

Phase III
Applicability

Analytics for better urban cycling



Analytics for better urban cycling



Legislative Assembly
Ontario

CycleLinx

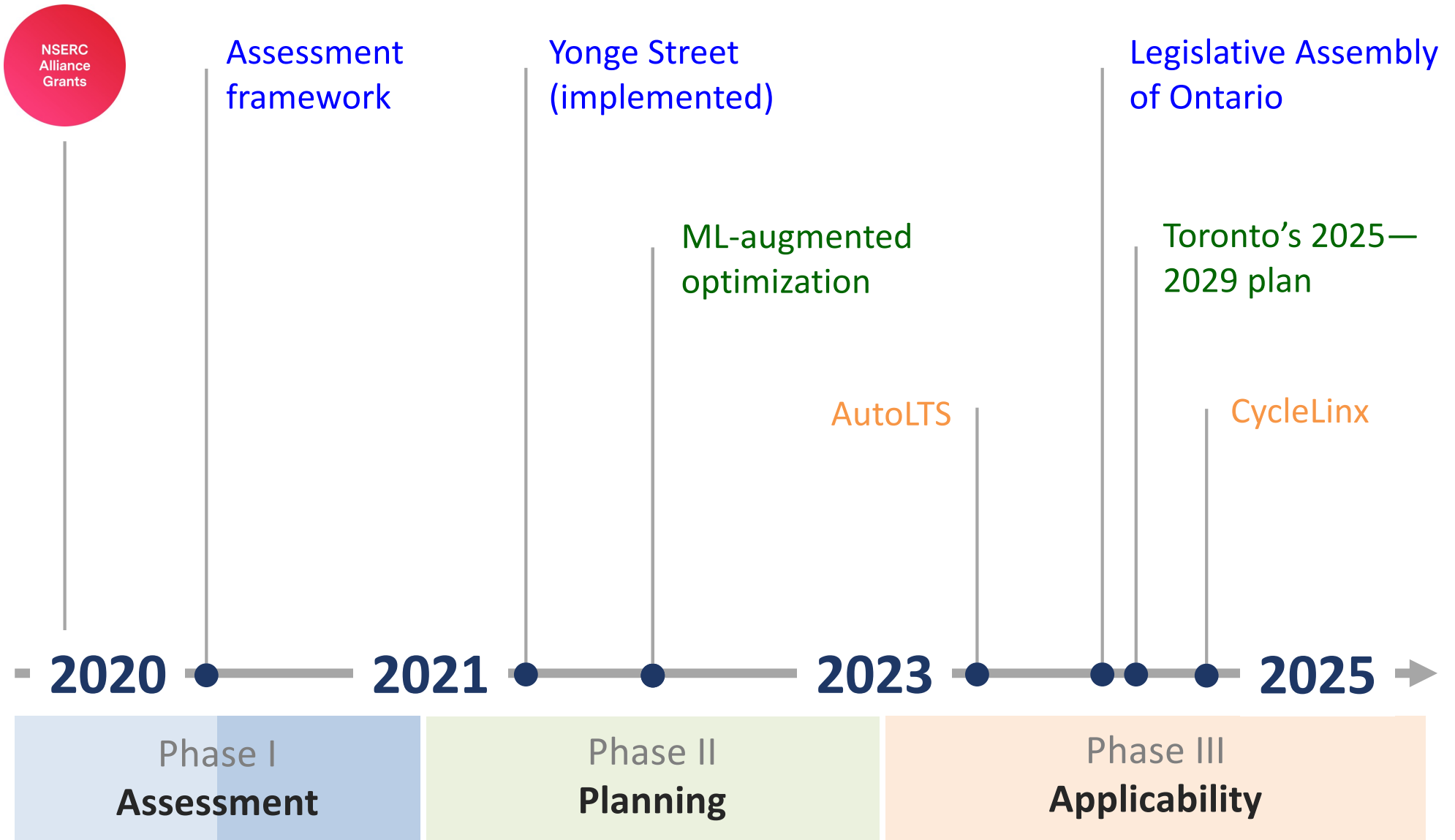


Phase I
Assessment

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

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

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

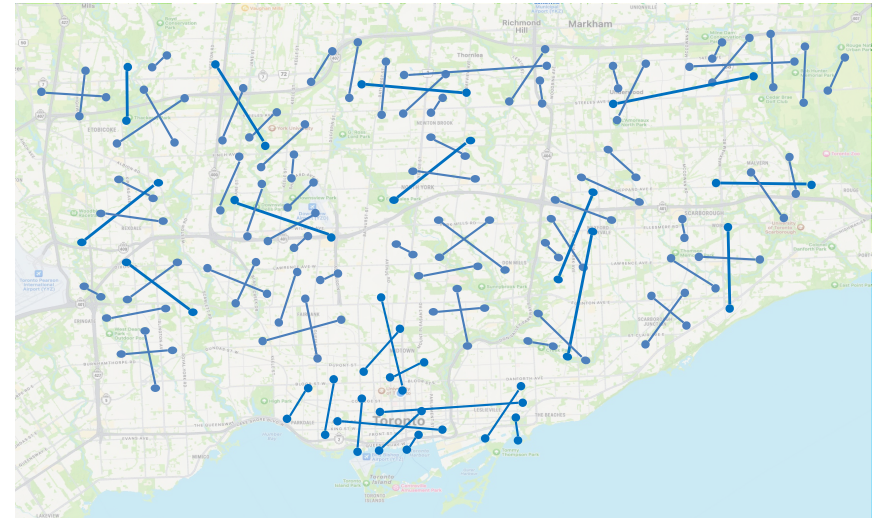
Low-stress shortest path (\mathbf{x})



follower

...

x 1 million!



**Bilevel
Program**

Leader

Followers

**Two-Stage
Stochastic Program**

1st-stage decision maker

2nd-stage decision makers
(in different scenarios)

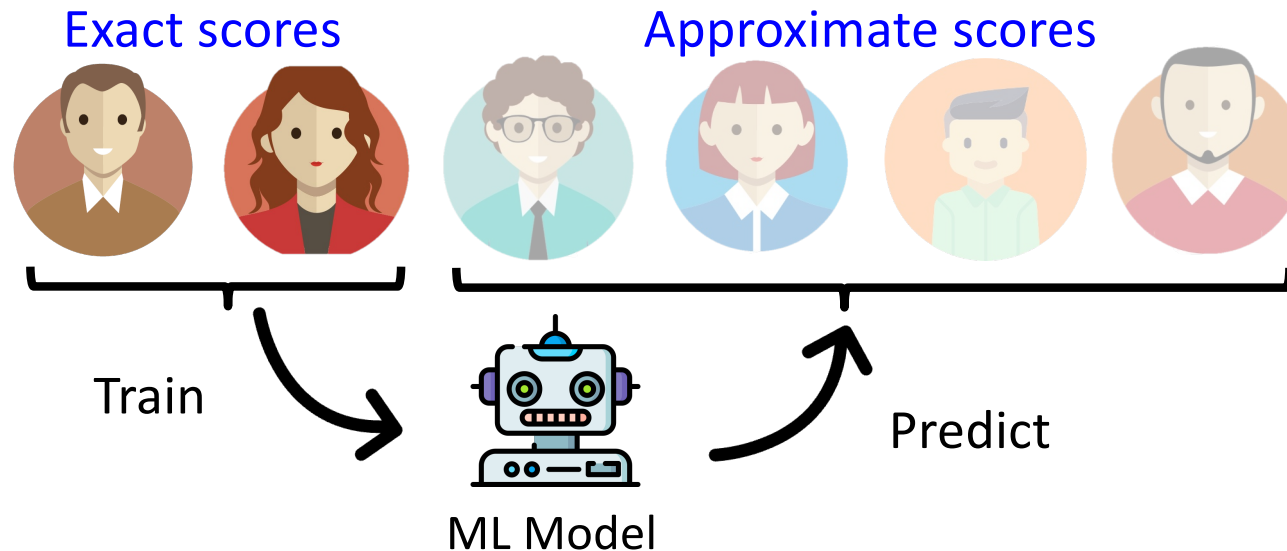


How to deal with one million followers?

Sampling!

Tractable, but lose sight of the “bigger picture”

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

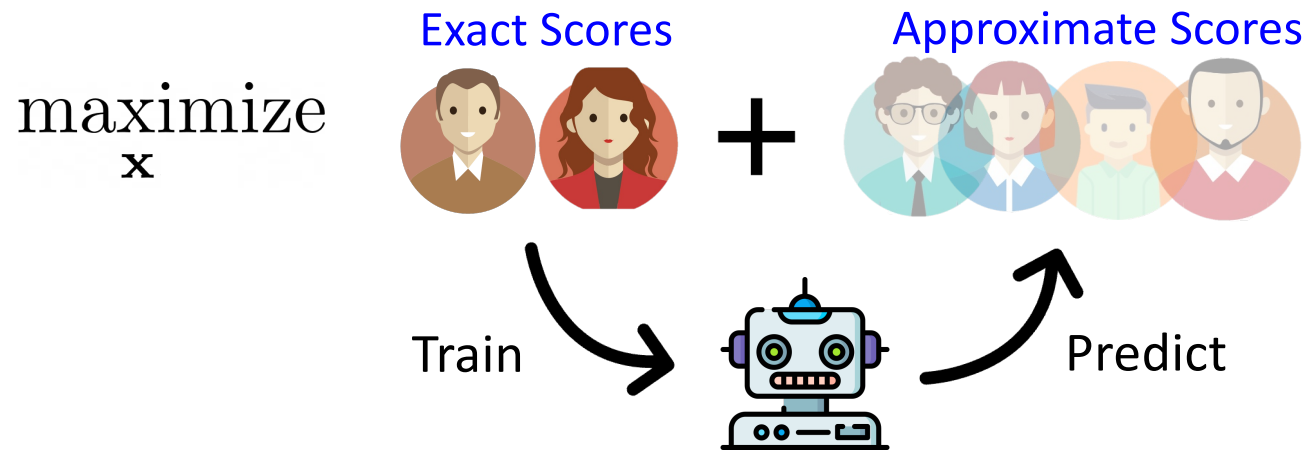


Key contribution: *ML-Augmented optimization*

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

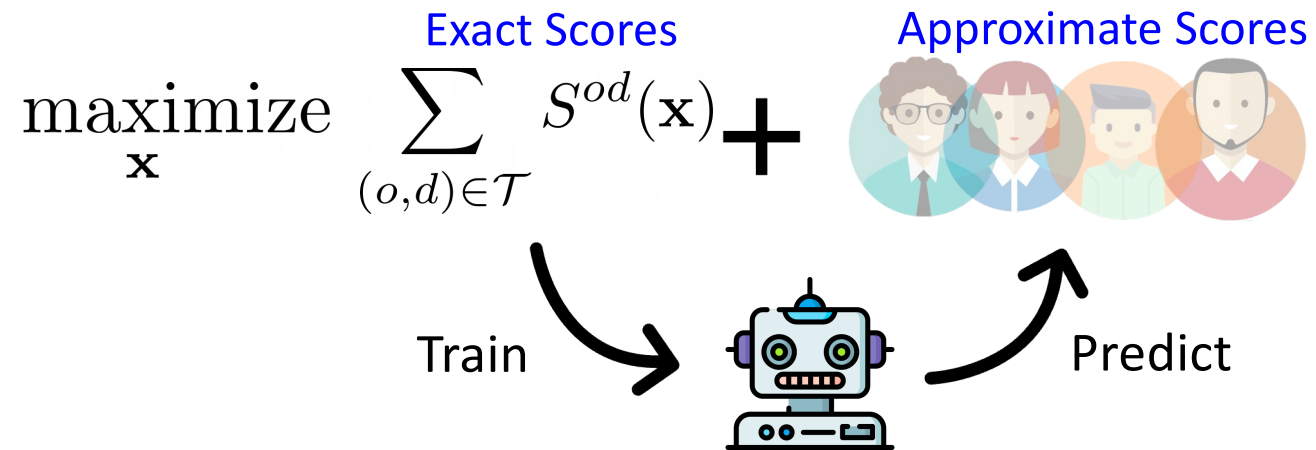
Our Approach

The ML-Augmented Model



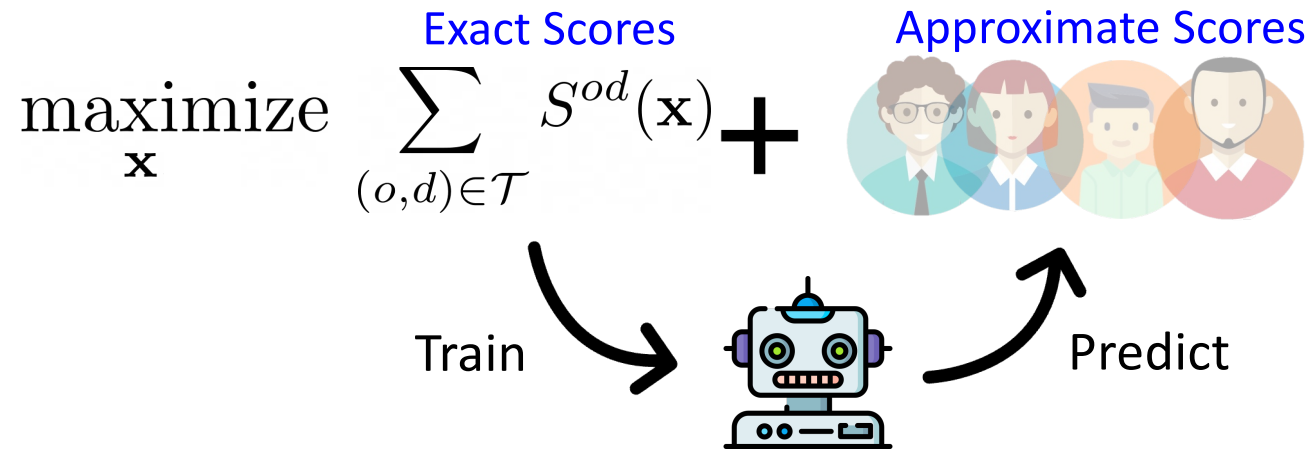
Our Approach

The ML-Augmented Model



Our Approach

The ML-Augmented Model



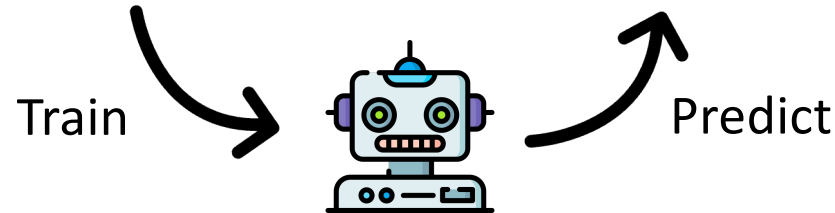
$$\underbrace{\mathbf{w}^\top}_{\text{ML model parameters}} \underbrace{\mathbf{f}^{od}}_{\text{features (OD locations, etc.)}} \longrightarrow S^{od}(\mathbf{x})$$

Our Approach

The ML-Augmented Model

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

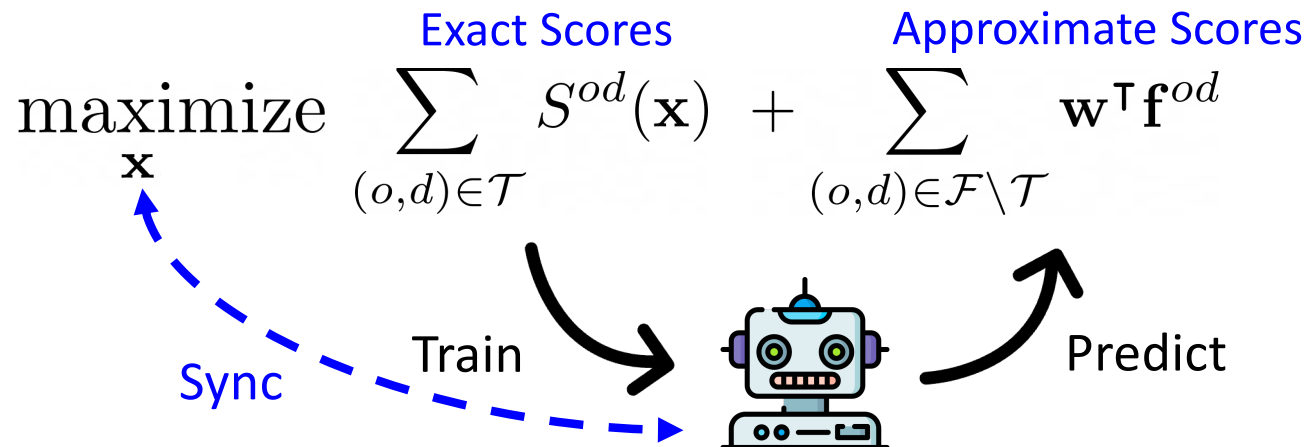
Exact Scores
Approximate Scores



$$\underbrace{\mathbf{w}^\top}_{\substack{\text{ML model} \\ \text{parameters}}} \underbrace{\mathbf{f}^{od}}_{\substack{\text{features} \\ \text{(OD locations, etc.)}}} \longrightarrow S^{od}(\mathbf{x})$$

Our Approach

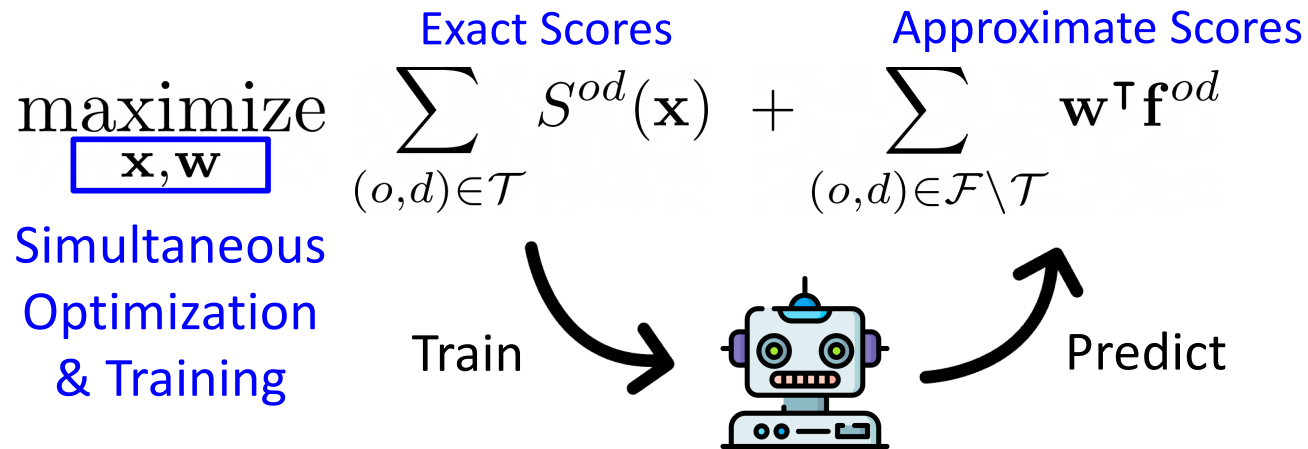
The ML-Augmented Model



$$\underbrace{\mathbf{w}^\top}_{\text{ML model parameters}} \underbrace{\mathbf{f}^{od}}_{\text{features (OD locations, etc.)}} \longrightarrow S^{od}(\mathbf{x})$$

Our Approach

The ML-Augmented Model



$$\underbrace{\mathbf{w}^\top}_{\text{ML model parameters}} \underbrace{\mathbf{f}^{od}}_{\text{features (OD locations, etc.)}} \longrightarrow S^{od}(\mathbf{x})$$

Our Approach

The ML-Augmented Model

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

Exact Scores
Approximate Scores

Simultaneous
Optimization
& Training

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

L1 Training loss is bounded

Our Approach

The ML-Augmented Model

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

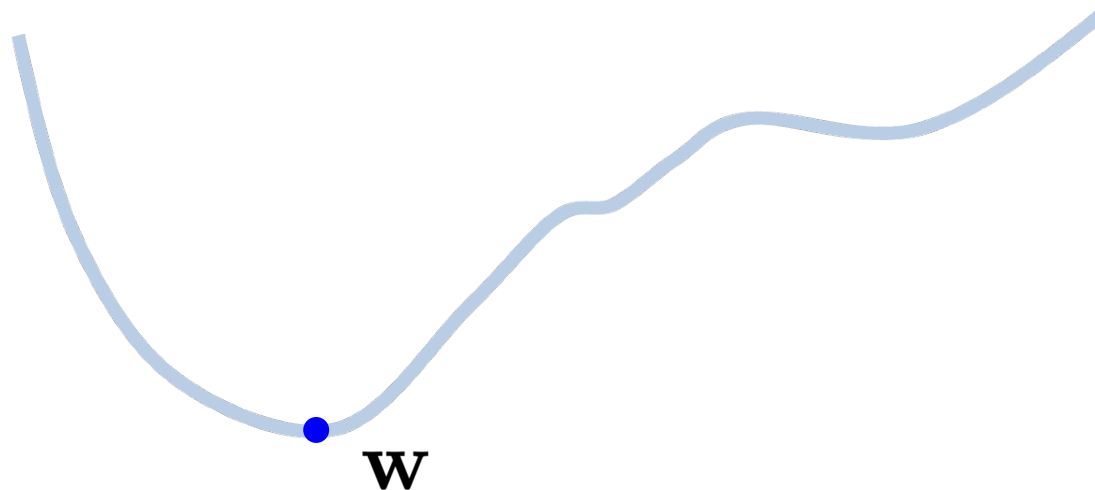
Exact Scores Approximate Scores

Simultaneous
Optimization
& Training

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

L1 Training loss is bounded

Training loss



Our Approach

The ML-Augmented Model

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

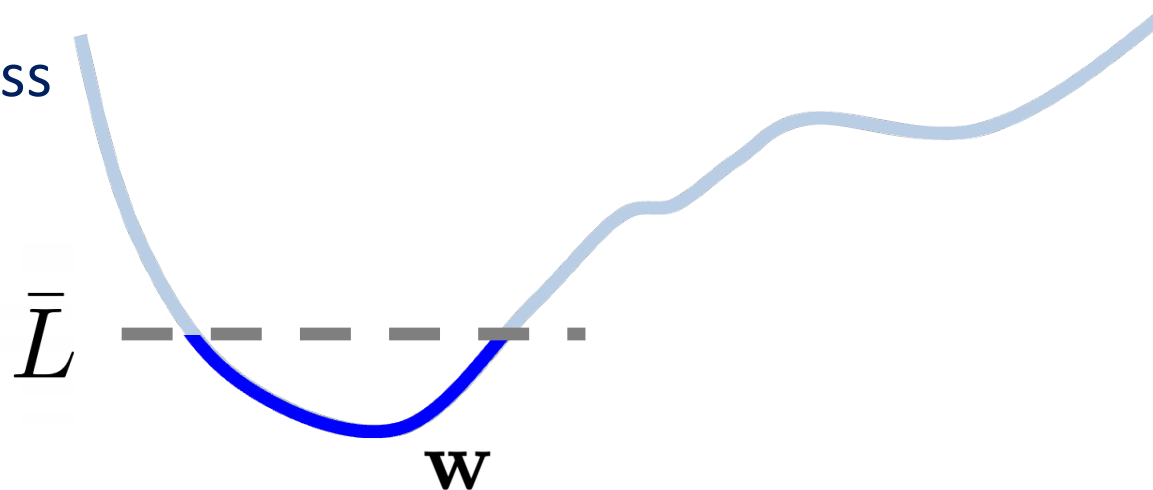
Exact Scores Approximate Scores

Simultaneous
Optimization
& Training

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

L1 Training loss is bounded

Training loss



Our Approach

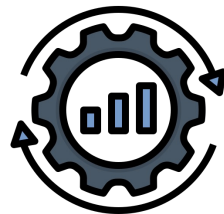
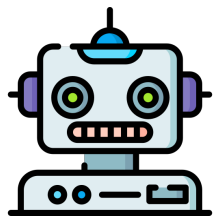
The ML-Augmented Model

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

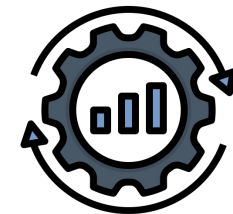
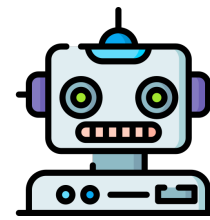
Simultaneous Optimization & Training

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

L1 Training loss is bounded



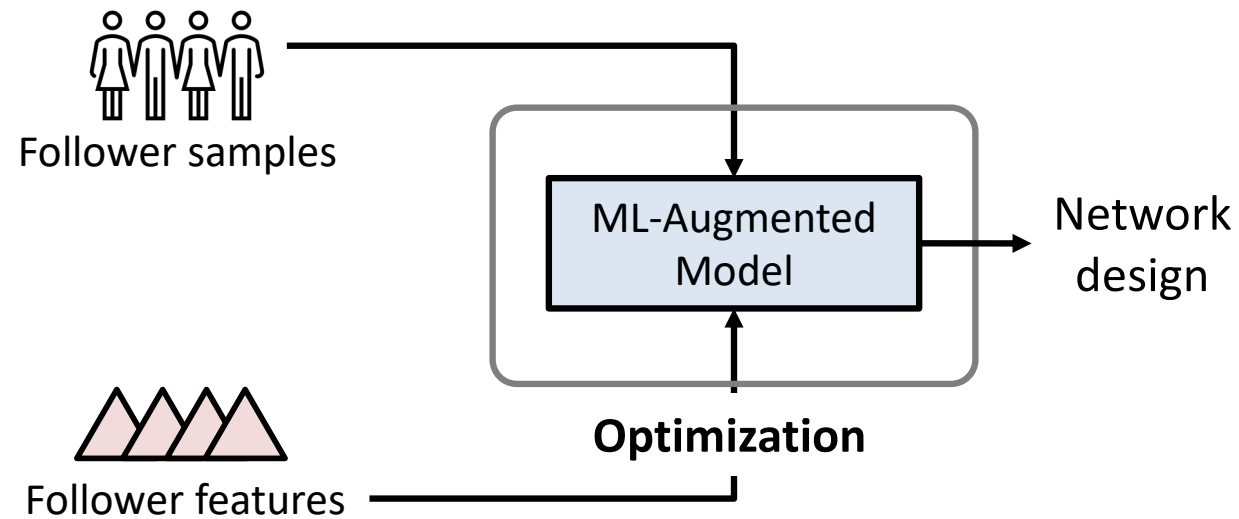
VS



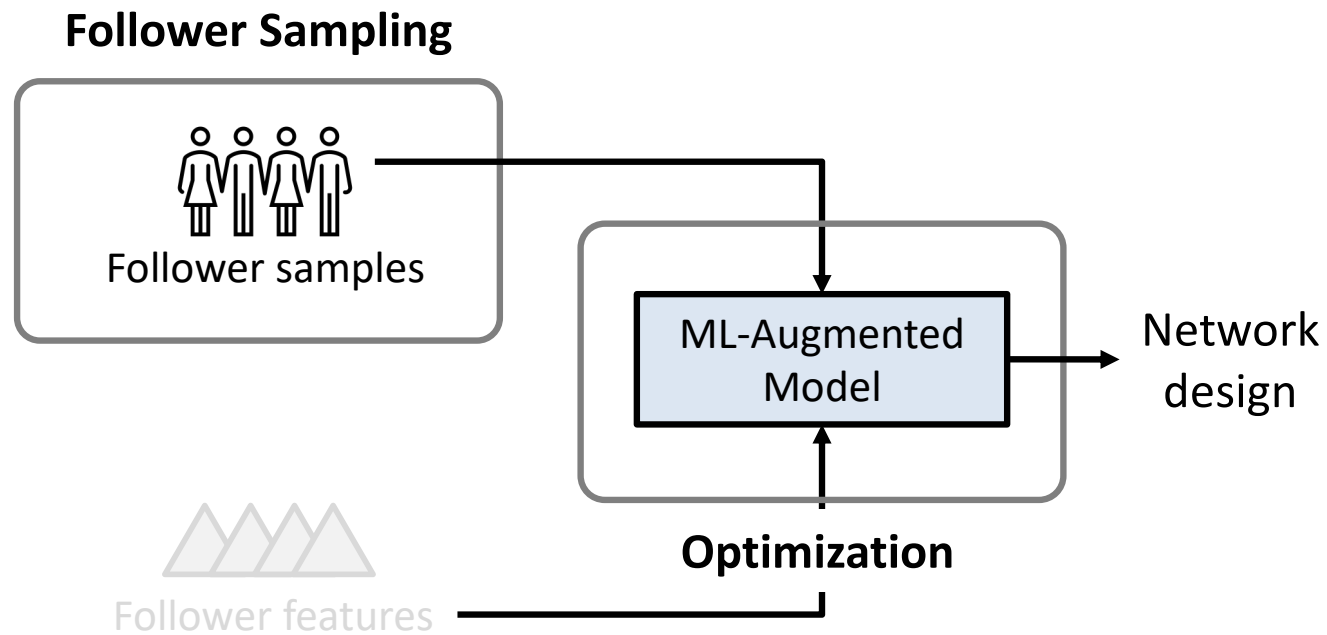
Predict, then optimize
(sequentially)

Predict and optimize
(simultaneously)

Roadmap

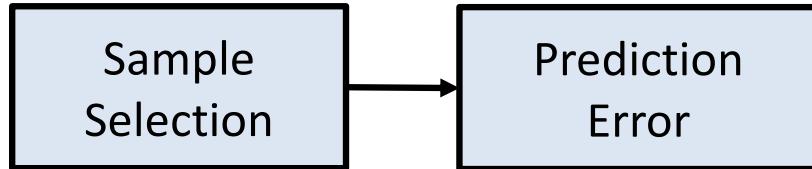


Roadmap

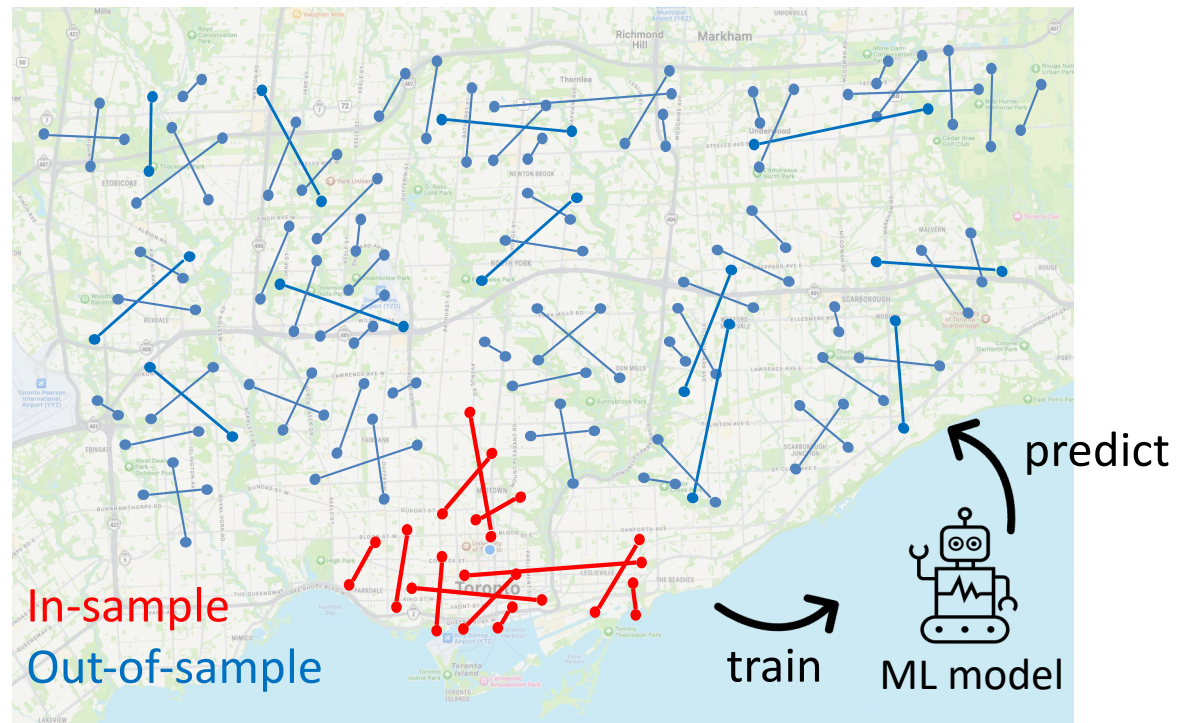
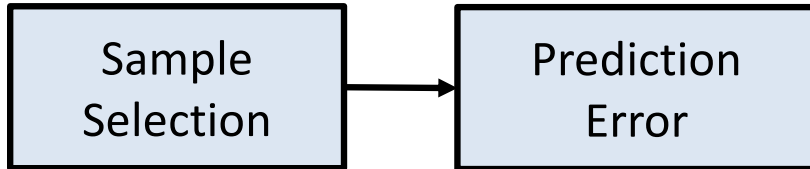


Why is sample selection so important?

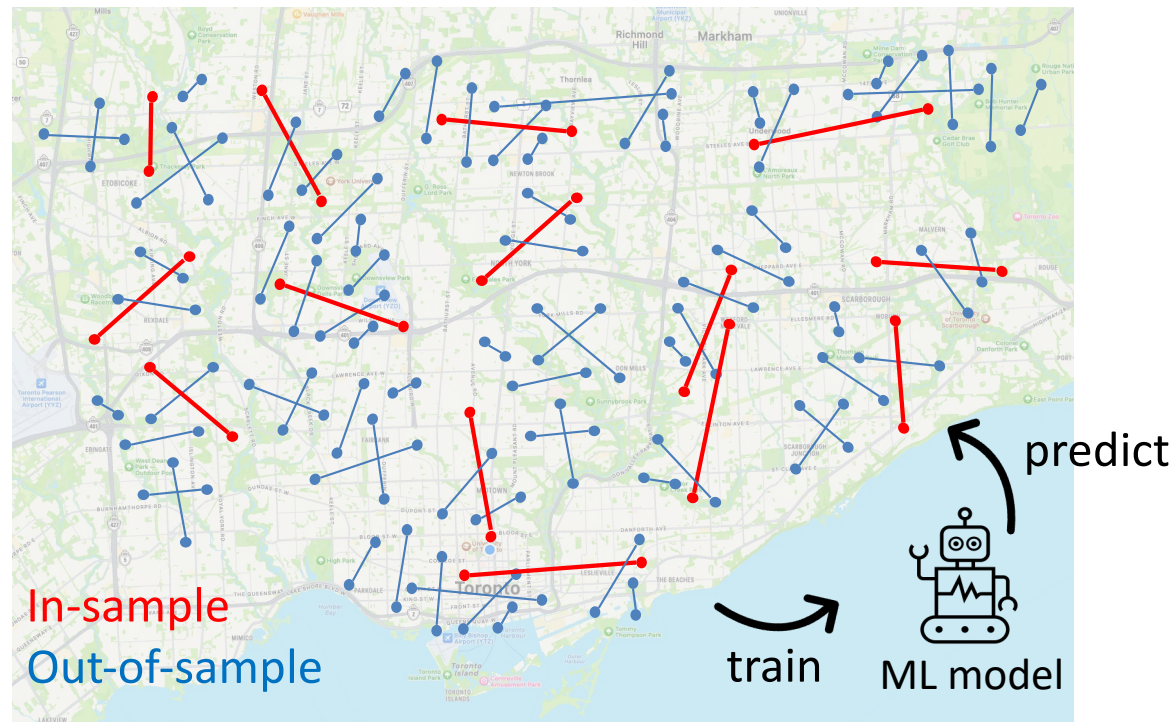
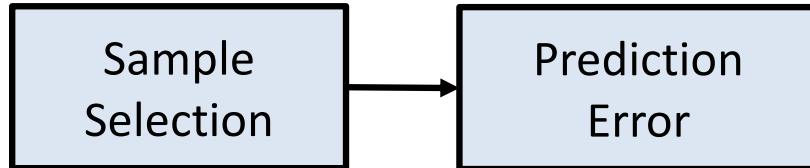
Why is sample selection so important?



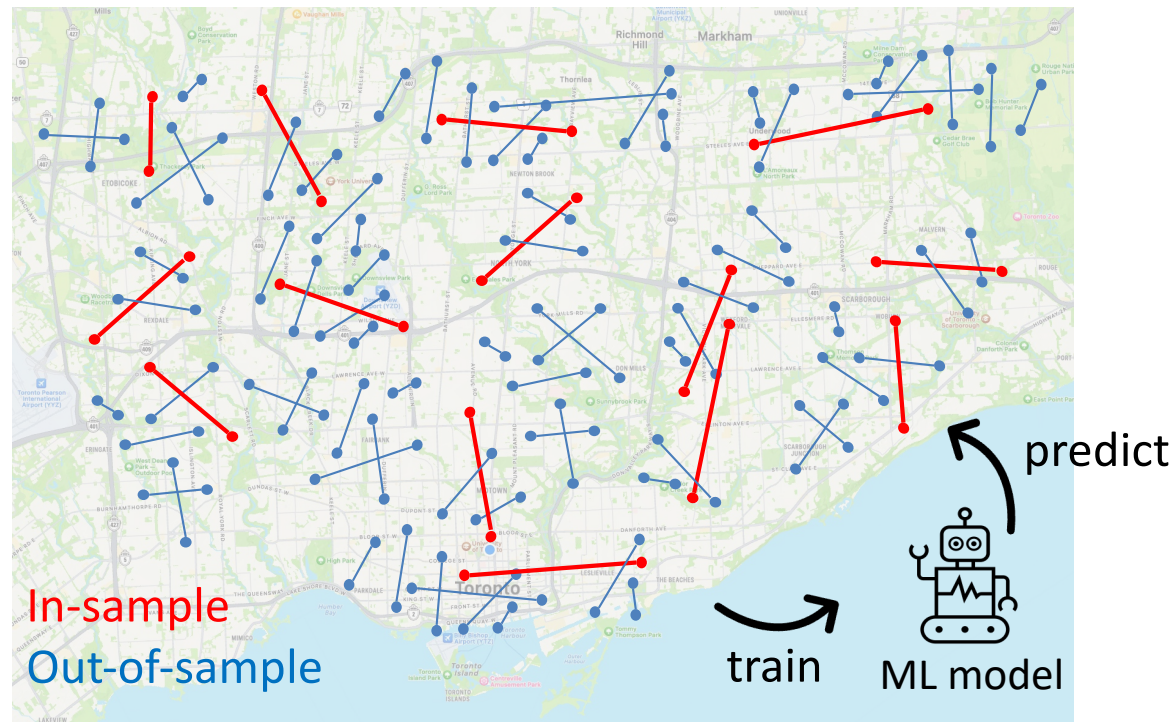
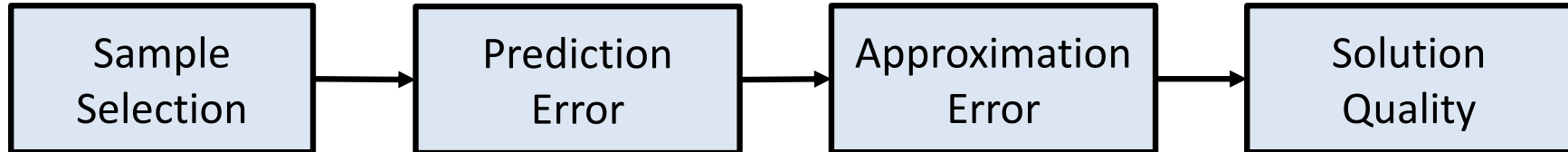
Why is sample selection so important?



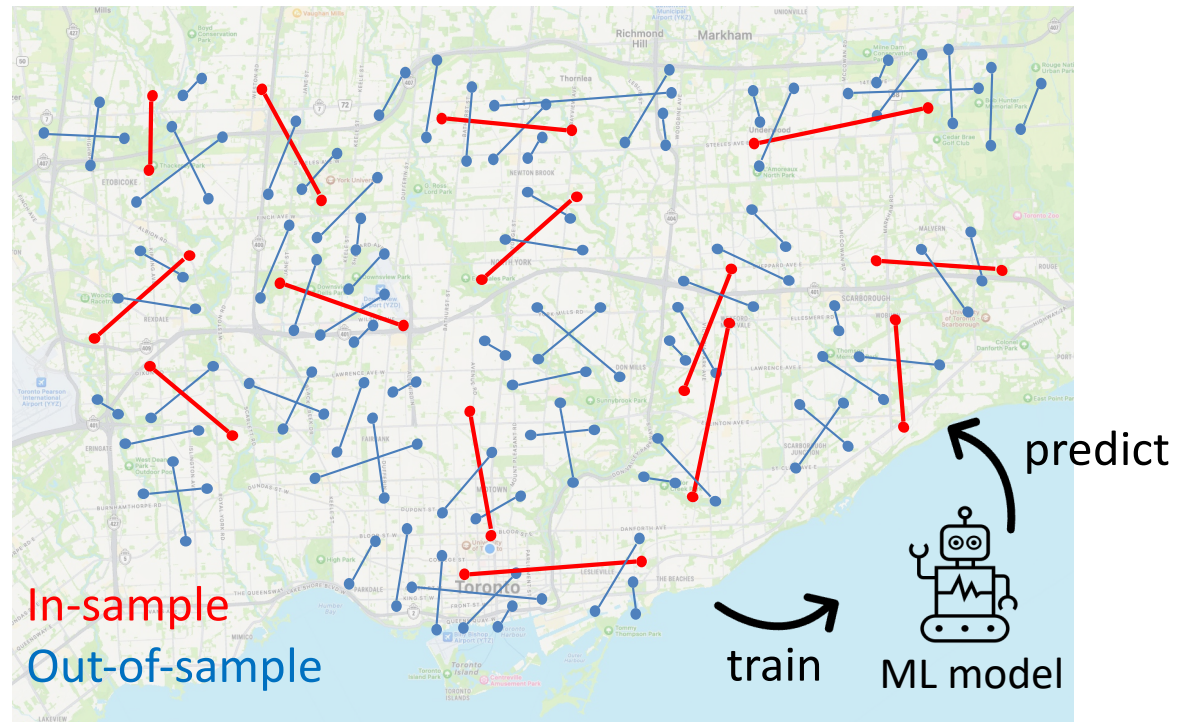
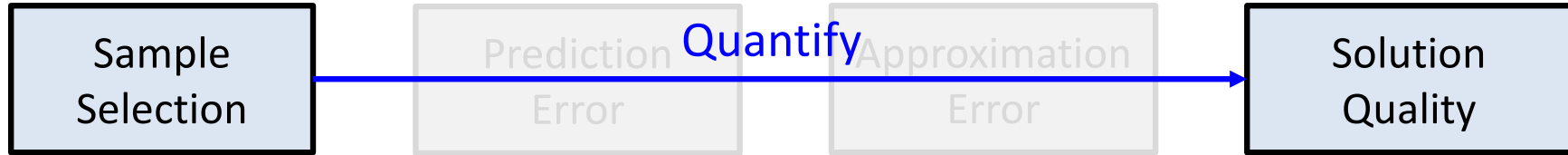
Why is sample selection so important?



Why is sample selection so important?



Why is sample selection so important?



How does follower sampling affect solution quality?

Theorem 2 (informal)

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

How does follower sampling affect solution quality?

Theorem 2 (informal)

optimality gap \leq

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

1st term

training loss

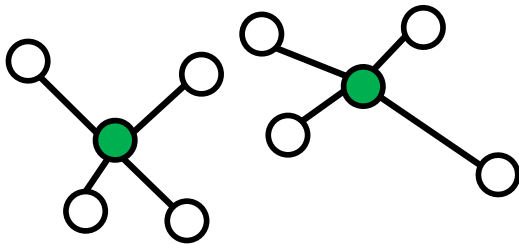
2nd term

**distance to the nearest
in-sample follower**

3rd term

**# of out-of-sample
followers assigned**

Feature Space



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

● in-sample follower (\mathcal{T})

Follower Sampling Problem

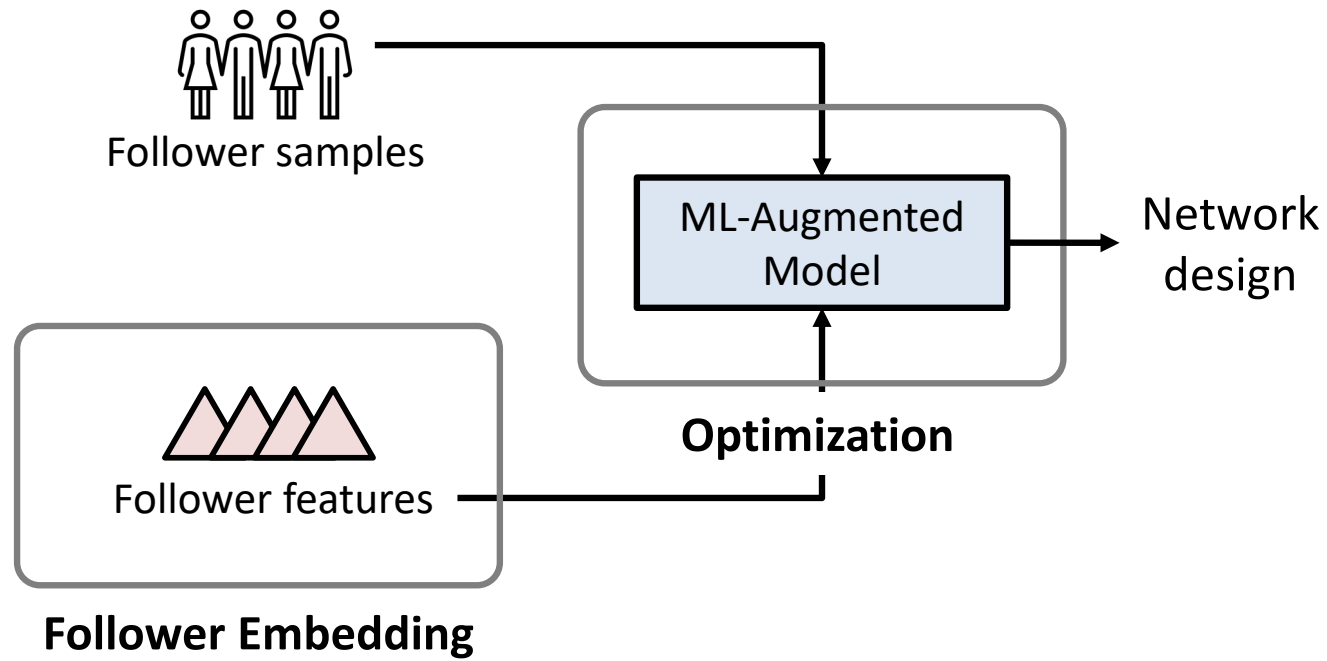
minimize total distance (2nd term)

subject to sample size $\leq p$

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

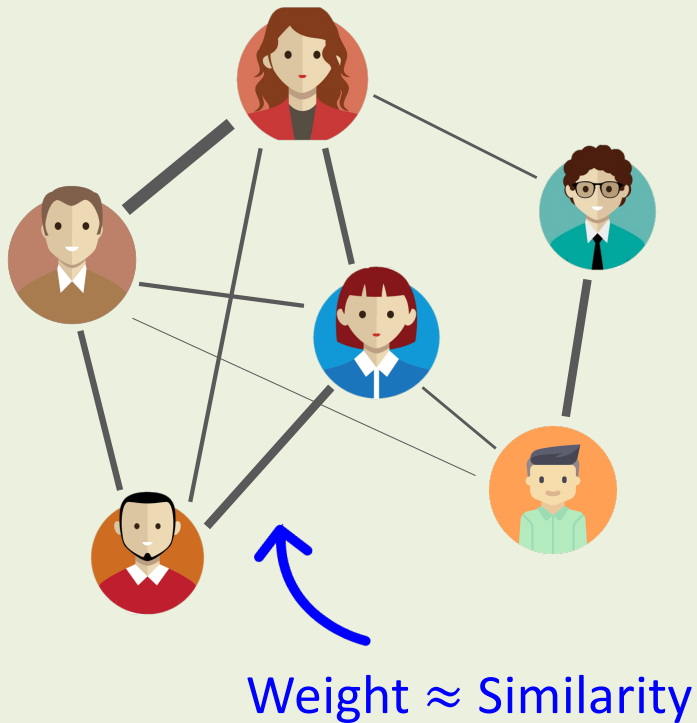
Balanced p-median problem

Roadmap

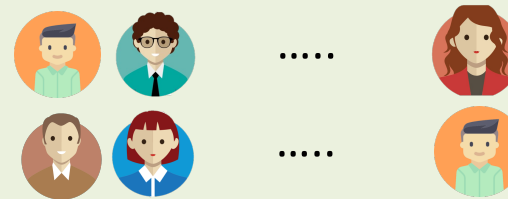


Follower embedding

1. Relationship Graph

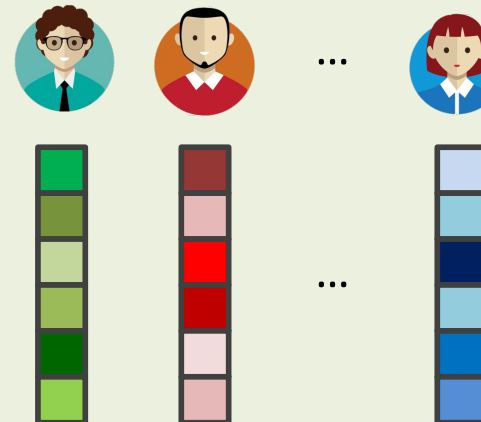


2. Random Walk

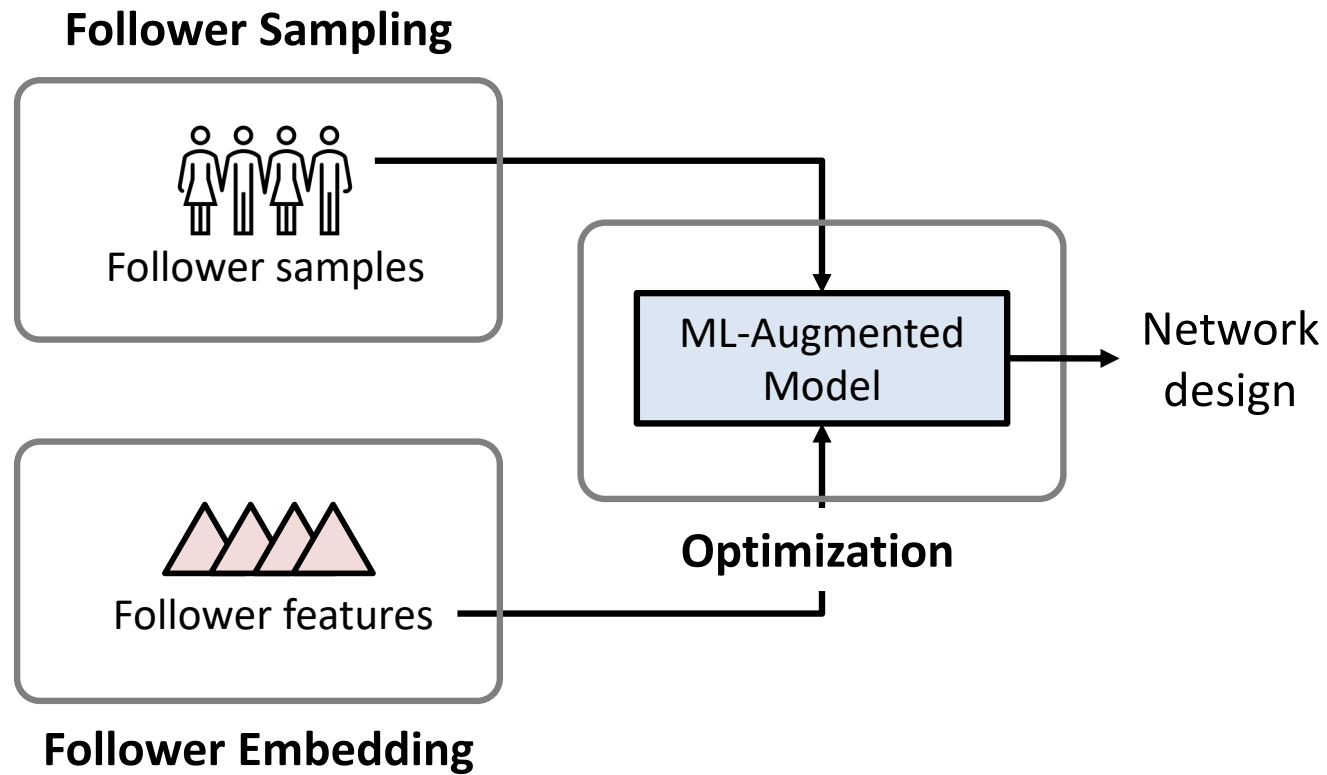


node = word walk = sentence

3. Word Embedding



Roadmap

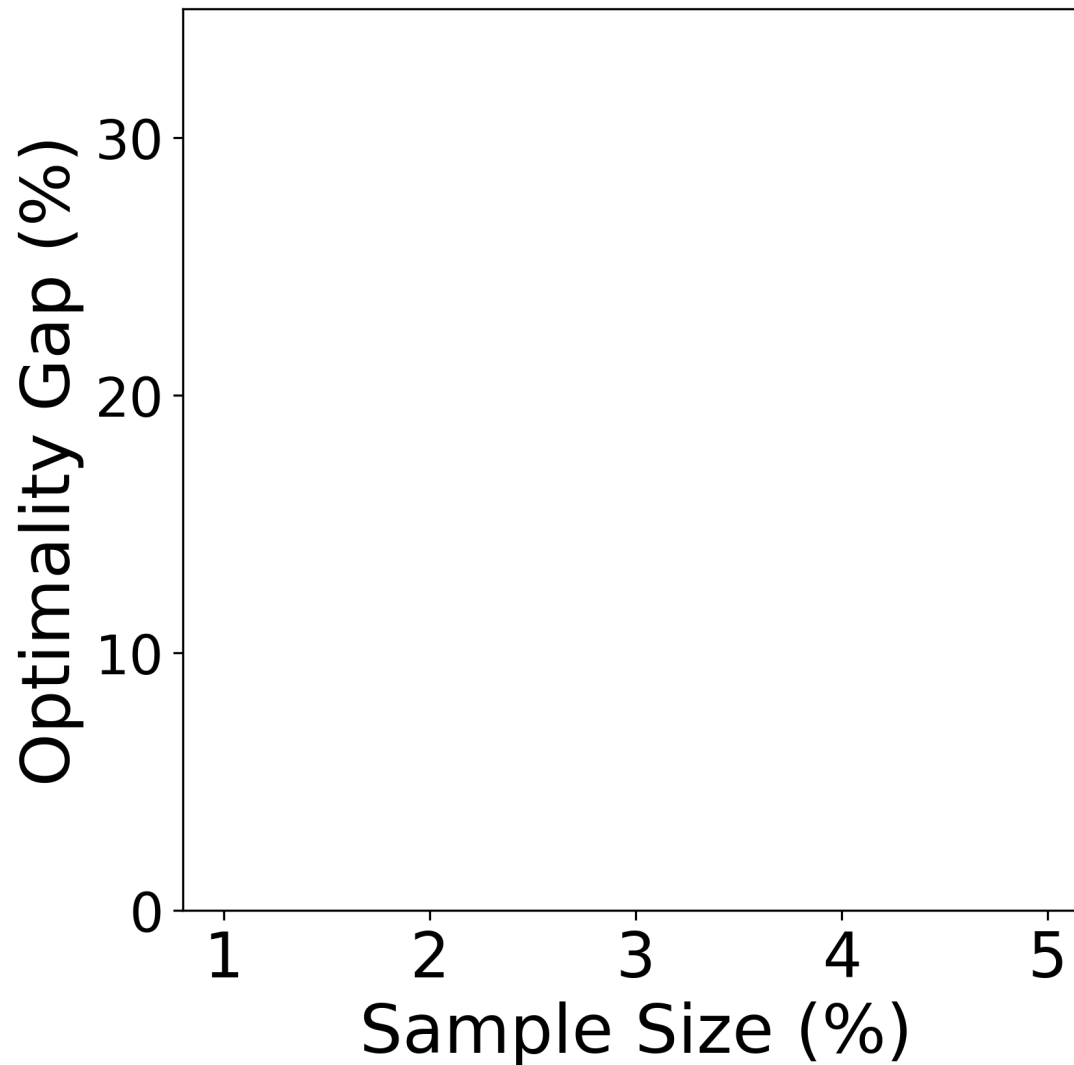


Synthetic Instances: Solution Quality

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

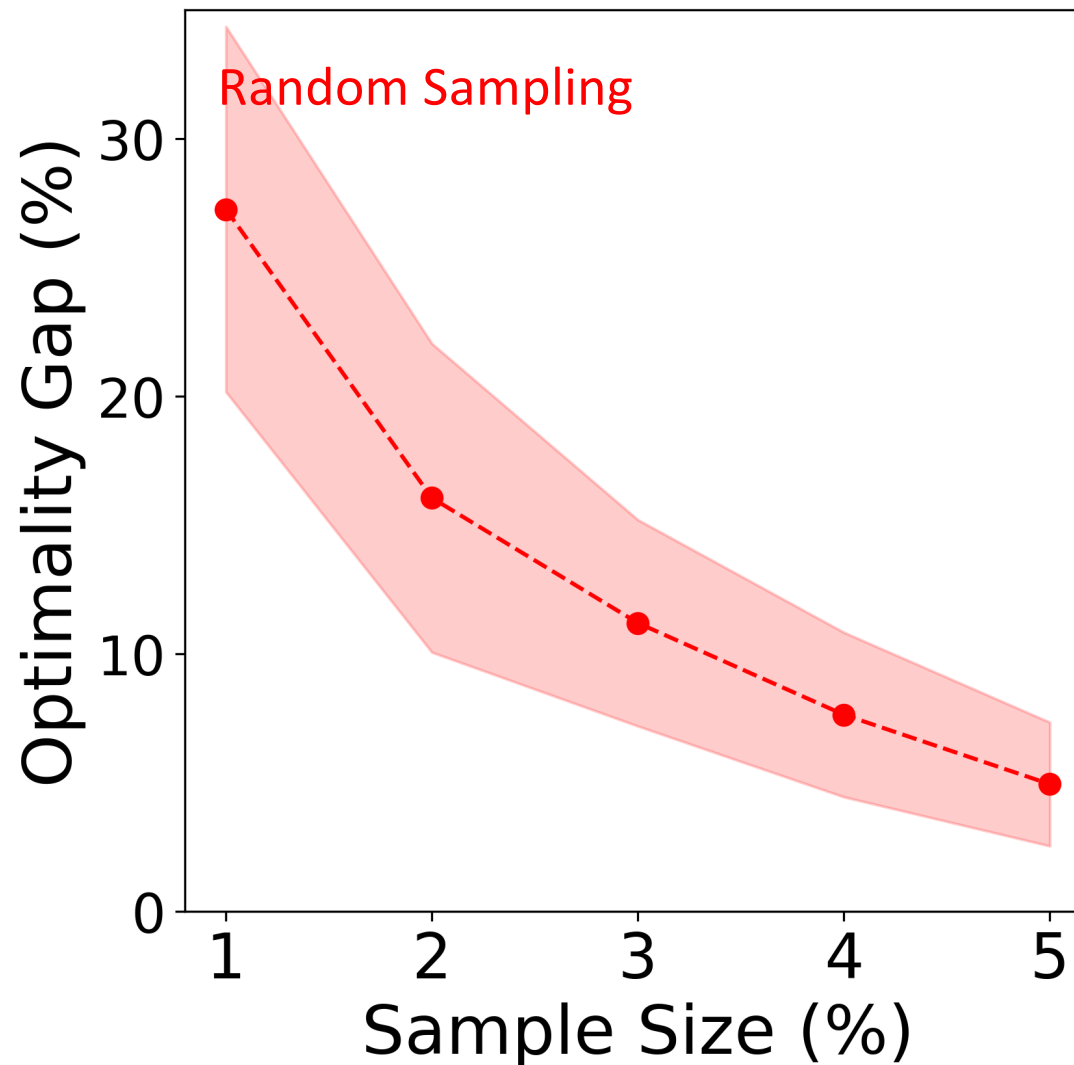
Synthetic Instances: Solution Quality

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



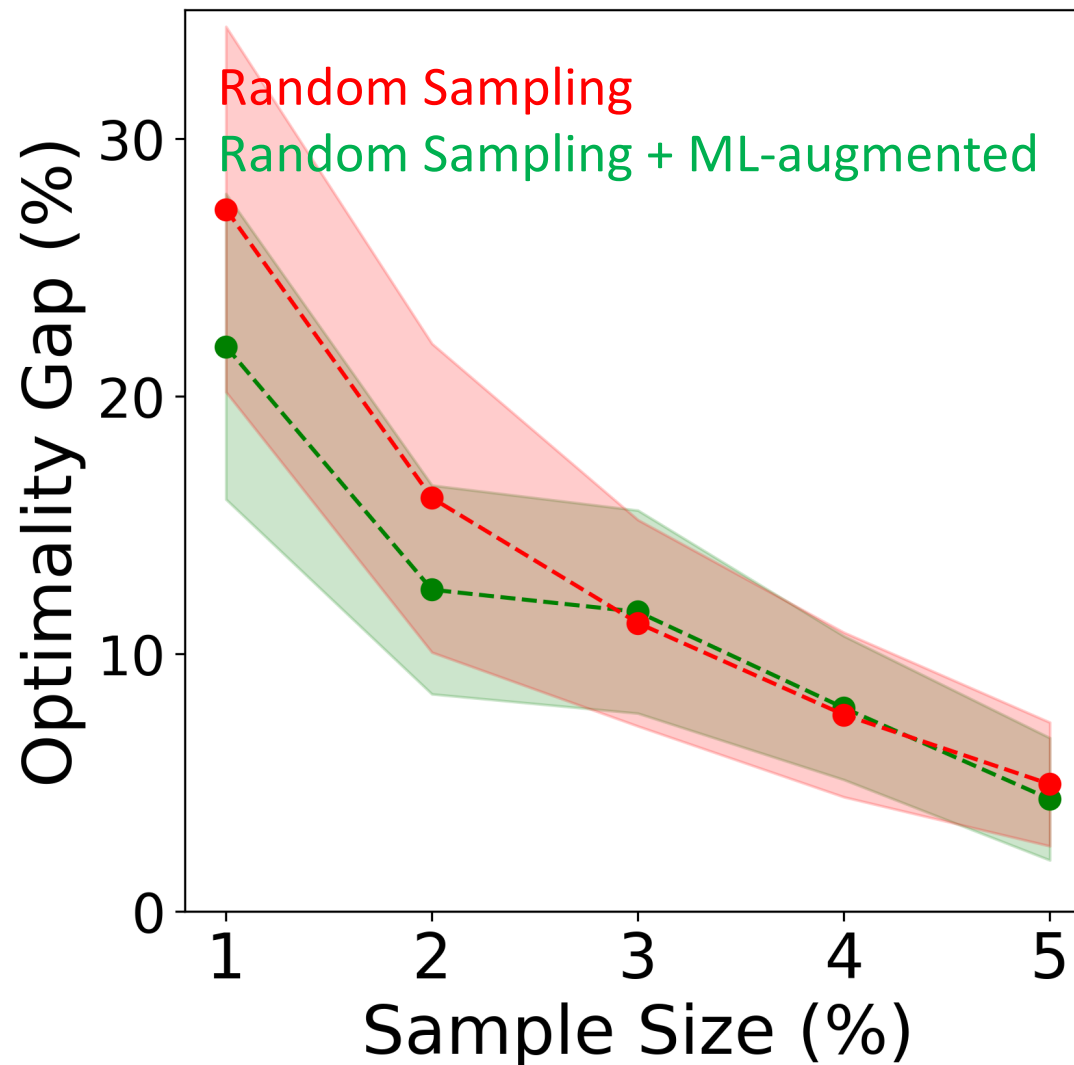
Synthetic Instances: Solution Quality

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



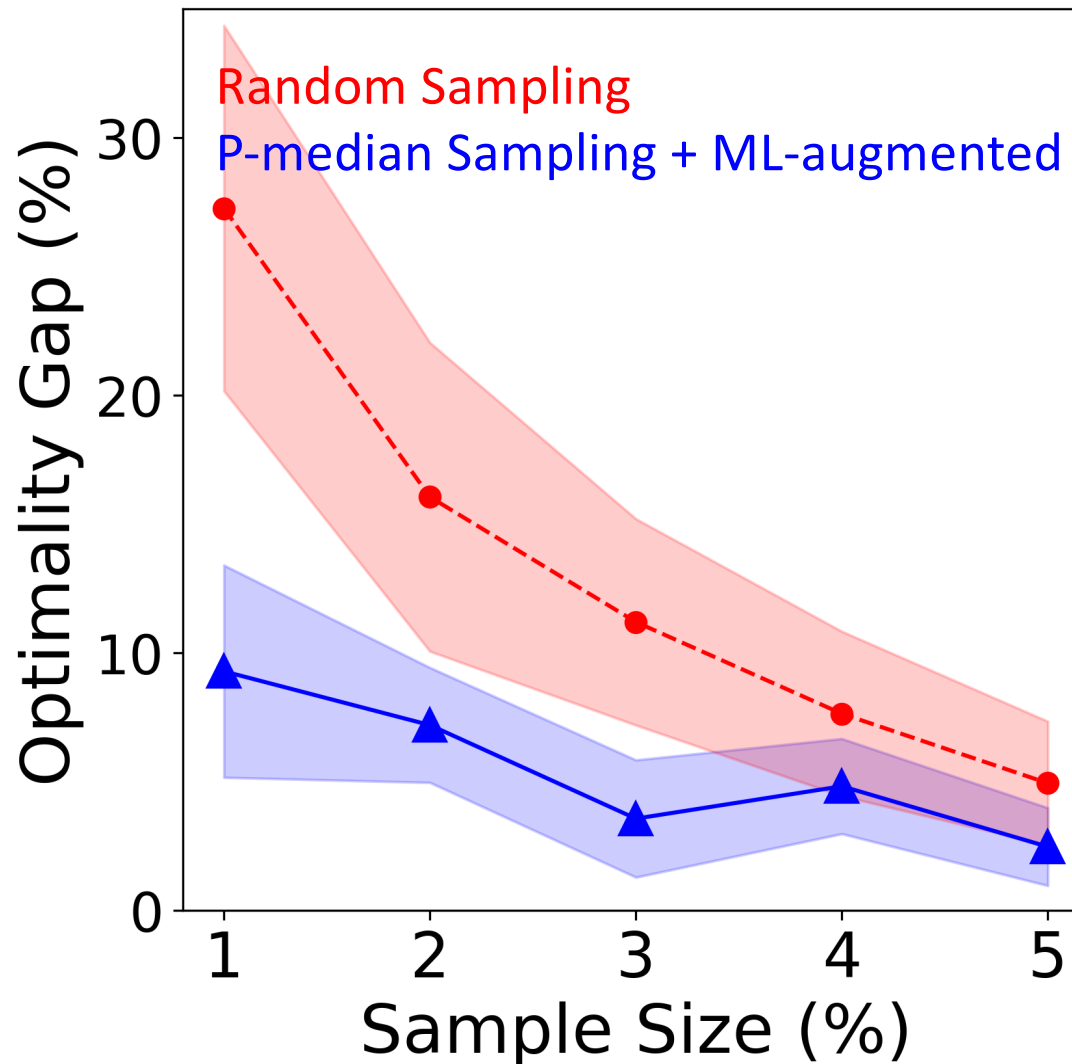
Synthetic Instances: Solution Quality

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



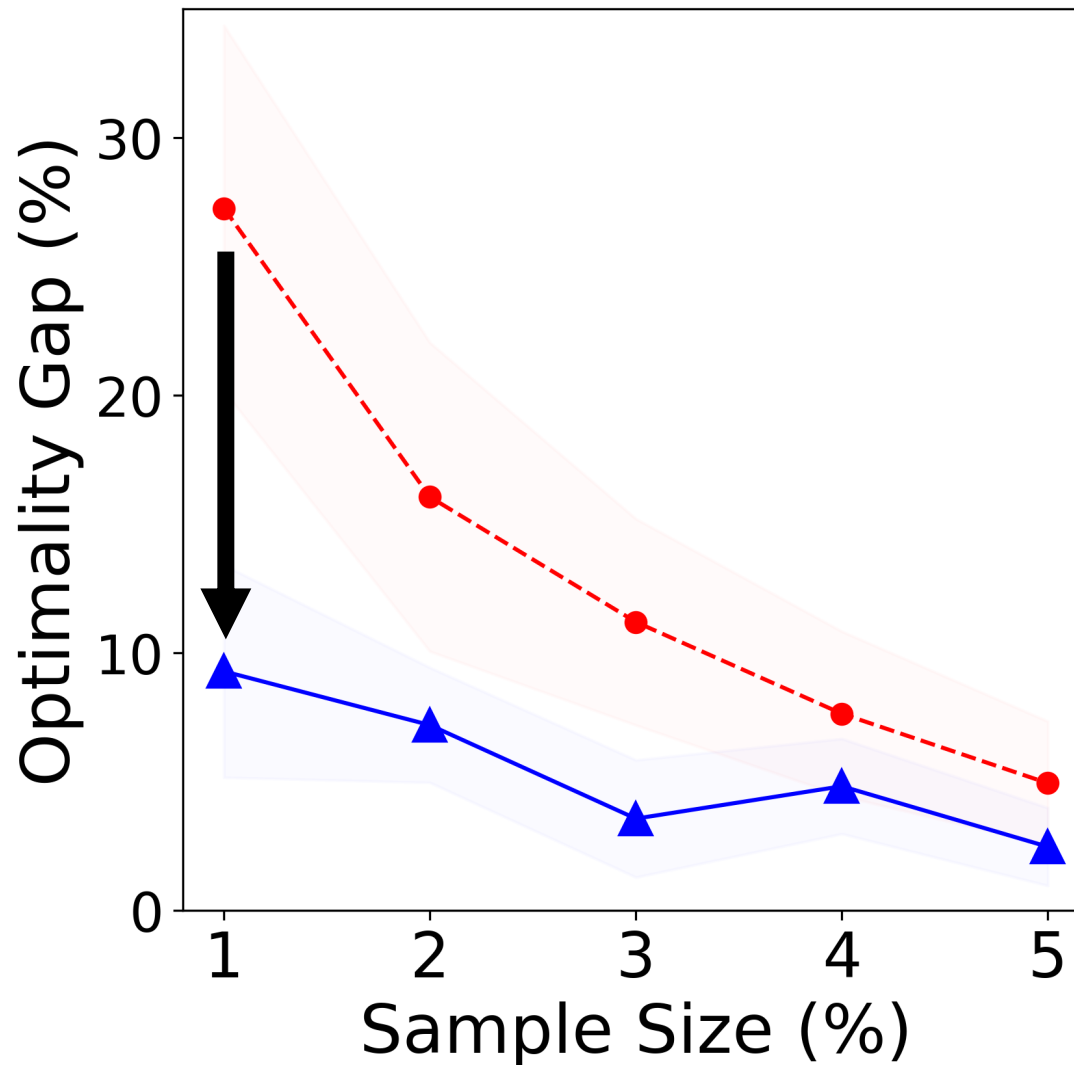
Synthetic Instances: Solution Quality

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

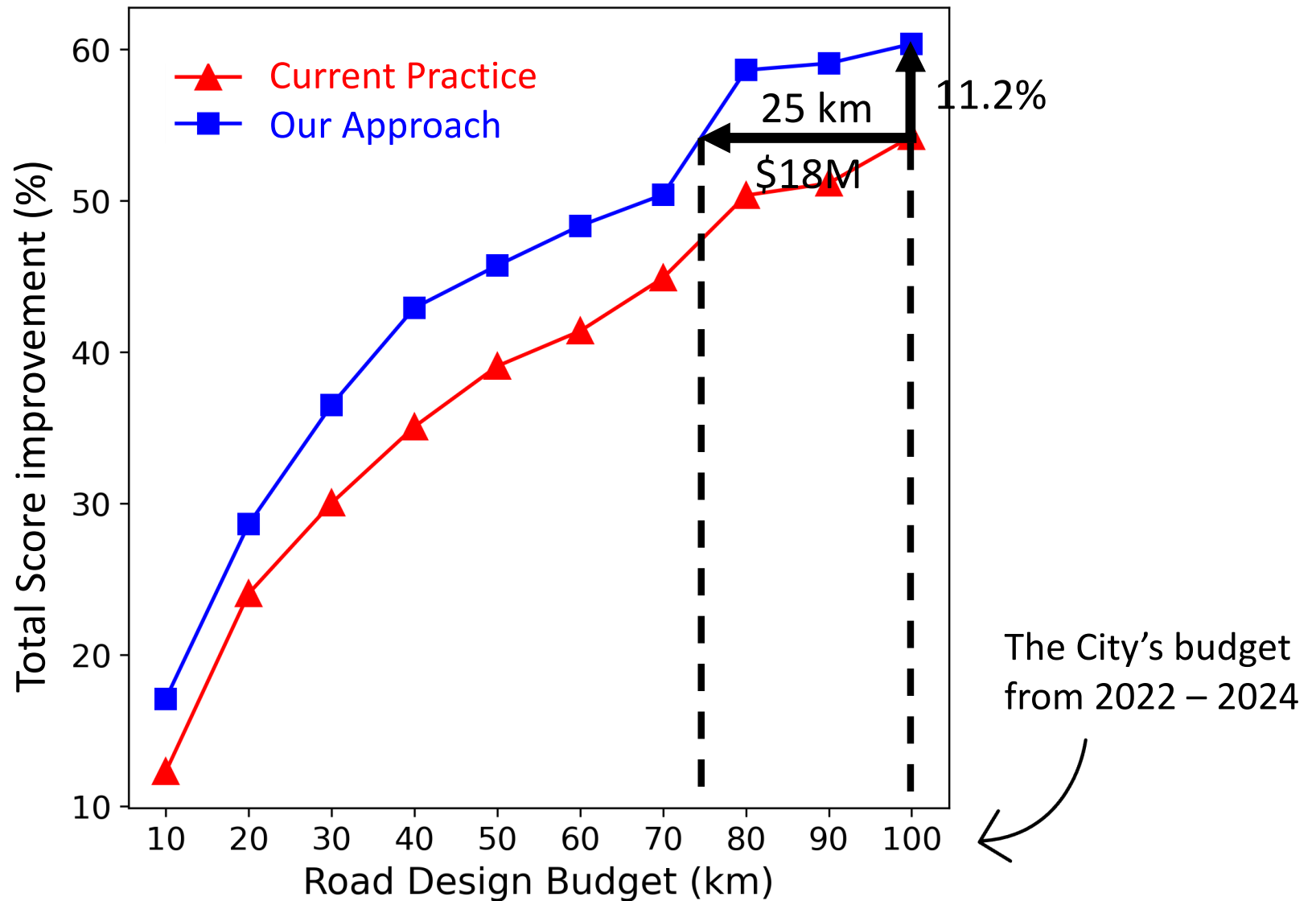


Synthetic Instances: Solution Quality

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

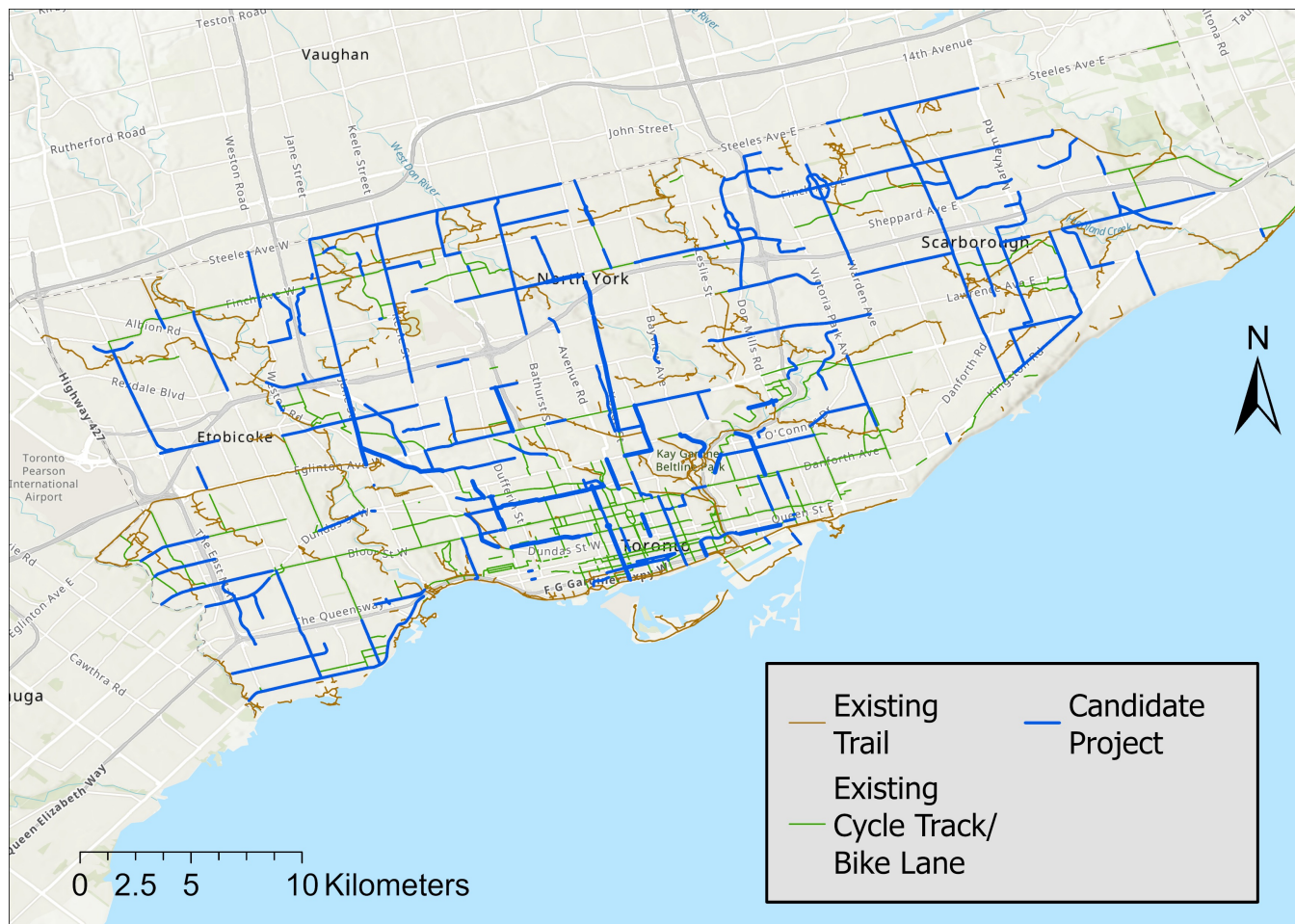


Strong performance: 2022 – 2024 planning horizon



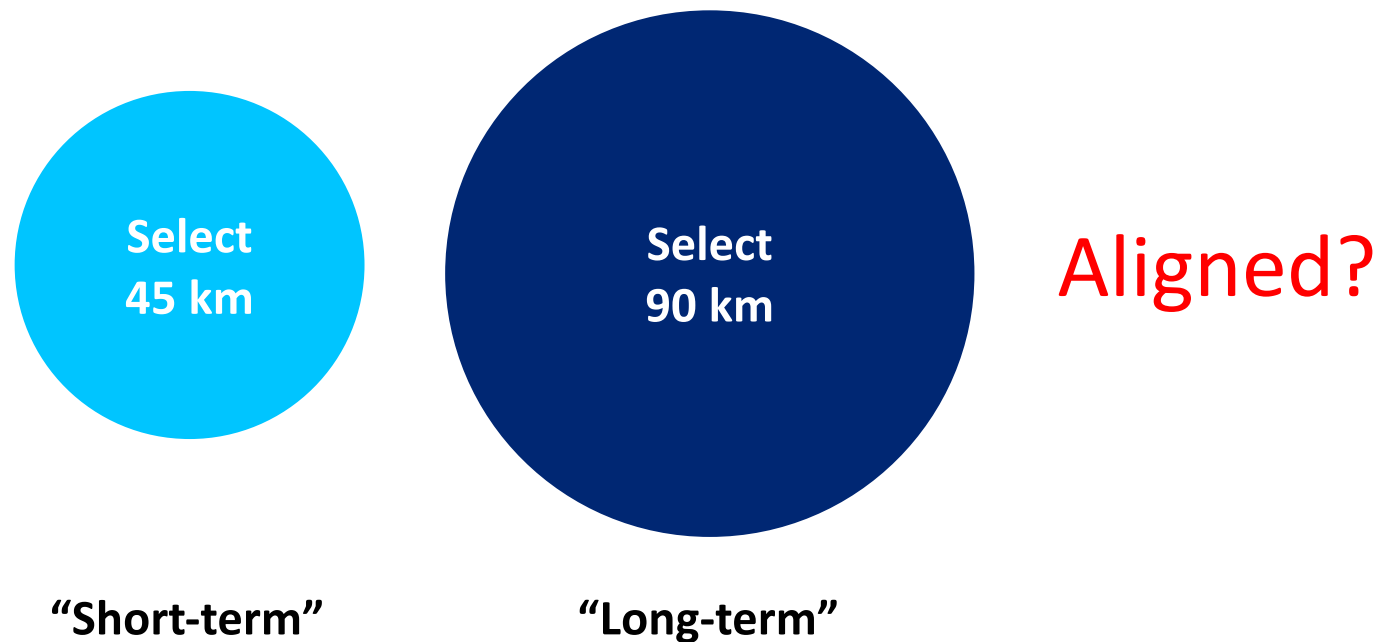
2025 – 2029 Network Expansion

- Candidate projects (339.3 km)

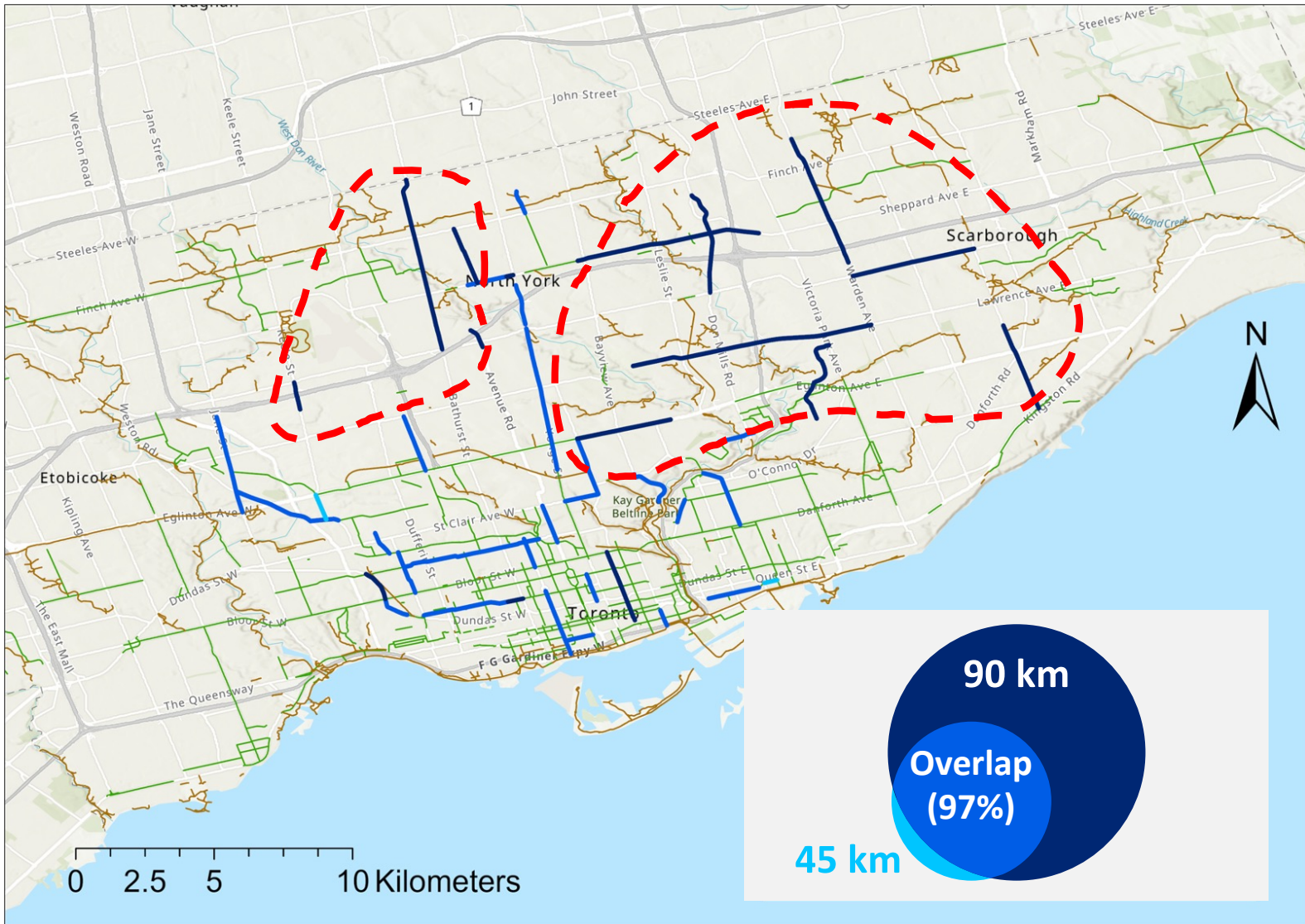


2025 – 2029 Network Expansion

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



The short- and long-term objectives are largely aligned

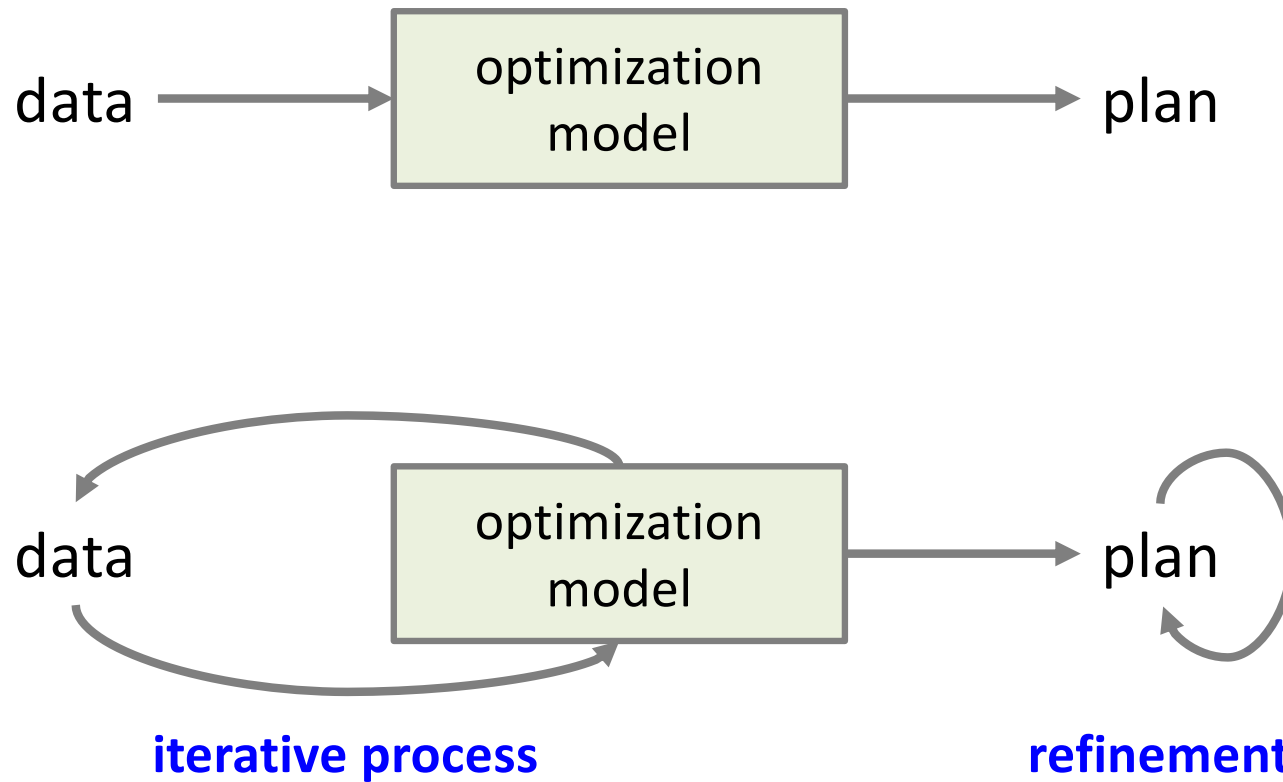


Final implementation

- 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



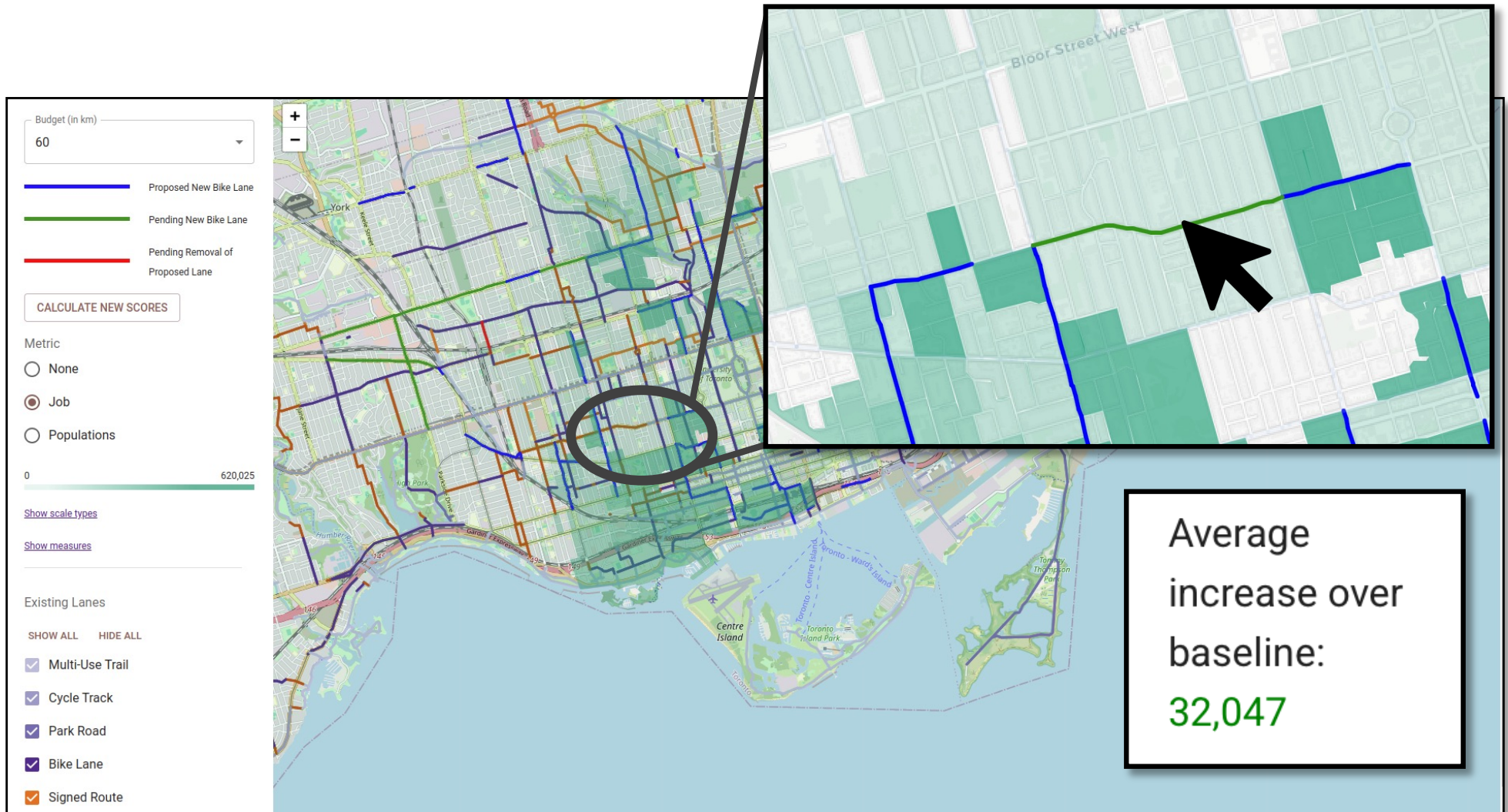
CycleLinx: optimizing with humans, not for them

The image displays the CycleLinx web application interface. On the left, a control panel includes a 'Budget (in km)' dropdown menu set to 60, a 'CALCULATE NEW SCORES' button, and a 'Metric' section with radio buttons for 'None', 'Job', and 'Populations'. Below this is a horizontal scale from 0 to 620,025 and links for 'Show scale types' and 'Show measures'. The 'Existing Lanes' section has 'SHOW ALL' and 'HIDE ALL' options, followed by checkboxes for 'Multi-Use Trail', 'Cycle Track', 'Park Road', 'Bike Lane', and 'Signed Route'. A legend identifies lane types: blue for 'Proposed New Lane', green for 'Pending New Bike Lane', and red for 'Pending Removal of Proposed Lane'. A large, semi-transparent slider is overlaid on the interface, showing a range from 10 to 70 km, with the current value set at 60. The background is a map of Toronto with various colored lines representing different lane types across the city grid.

CycleLinx: optimizing with humans, not for them

The screenshot displays the CycleLinx web application interface. On the left, a sidebar contains a 'Budget (in km)' dropdown set to 60, a legend for 'Proposed New Bike Lane' (blue), 'Pending New Bike Lane' (green), and 'Pending Removal of Proposed Lane' (red), and a 'CALCULATE NEW SCORES' button. Below this is a 'Metrics' section with three radio button options: 'None', 'Job' (which is selected and circled in black), and 'Populations'. A 'Show scale types' link is also present. At the bottom of the sidebar, an 'Existing Lanes' section has 'SHOW ALL' and 'HIDE ALL' buttons, followed by a list of lane types with checkboxes: Multi-Use Trail, Cycle Track, Park Road, Bike Lane, and Signed Route. The main area is a map of Toronto showing various colored bike lane proposals overlaid on a street grid. A 'Metrics' dialog box is open in the center, mirroring the 'Metrics' section of the sidebar, with 'Job' selected. The map includes labels for 'East York', 'Centre Island', and 'Toronto Island Park'.

CycleLinx: optimizing with humans, not for them



CycleLinx: optimizing with humans, not for them

The screenshot displays the CycleLinx web application interface. On the left, there is a control panel with a 'Budget (in km)' dropdown set to 60. Below this are three legend items: 'Proposed New Bike Lane' (blue line), 'Pending New Bike Lane' (green line), and 'Pending Removal of Proposed Lane' (red line). A 'CALCULATE NEW SCORES' button is present. The 'Metric' section has radio buttons for 'None', 'Job' (selected), and 'Populations'. A progress bar shows a value of 0 out of 620,025. There are links for 'Show scale types' and 'Show measures'. The 'Existing Lanes' section has 'SHOW ALL' and 'HIDE ALL' options, with checkboxes for 'Multi-Use Trail', 'Cycle Track', 'Park Road', 'Bike Lane', and 'Signed Route', all of which are checked.

The main area shows a map of a city with various colored lines representing different lane types. A large white panel is overlaid on the map, titled 'History'. It contains the text 'finalized_plan' and 'SET AS BASELINE'. On the right side of the panel, there are three icons: an upward arrow, a trash can, and a download arrow. The download arrow icon is circled in red.

An end-to-end, data-driven solution

40 km
Assessed

Widely cited



Only research team presented at



**Toronto City Budget Meeting
Legislative Assembly of Ontario**

+11.2%
accessibility

-25%
length

2022 – 2024

\$18M
cost saving

29 km
approved

2025 – 2027

28 km
under study

2028 – 2029

Deployed in
Toronto

Rolling out in
**Hamilton
& Mississauga**

