STEP: Semantics-Aware Sensor Placement for Monitoring Community-Scale Infrastructure

Andrew Chio[1], Jian Peng[2], Nalini Venkatasubramanian[1]

November 15-16, 2023, Istanbul, Turkiye

ACM BuildSys 2023

[1] University of California, Irvine, California, US

[2] Orange County Public Works, California, US

Built Utility Infrastructures are Critical Lifelines

Built Utility Infrastructure


Credit: Stormdrain, by Mississippi Watershed Management Organization, CC BY-NC

Stormwater


Credit: The interior of a drinking water treatment plant. DW Treatment Plant, CC0

Drinking Water


Credit: Oil and gas pipeline, by Forest Guardians, CC BY-SA

Gas Pipelines


Credit: Natrual Gas power Plant, by Jon Sullivan, CC BY-NC

Power Plants

• • •

support

Community Structures

Credit: Emergency Sign, CC0

Hospitals


Credit: UCI Donald Bren Hall, by David Eppstein, CC BY-SA

Schools


Credit: Drone Shot of Moving Automobiles on an Expressway, by Kendel Media

Roads/Bridges


Credit: High-rise Apartments in Yokohama, by ykanazawa1999, CC BY-NC-SA

High Rises

• • •

Built Infrastructures are Strained and Prone to Failure

Rising Urban Populations

- Infrastructure usage and impact from failures

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

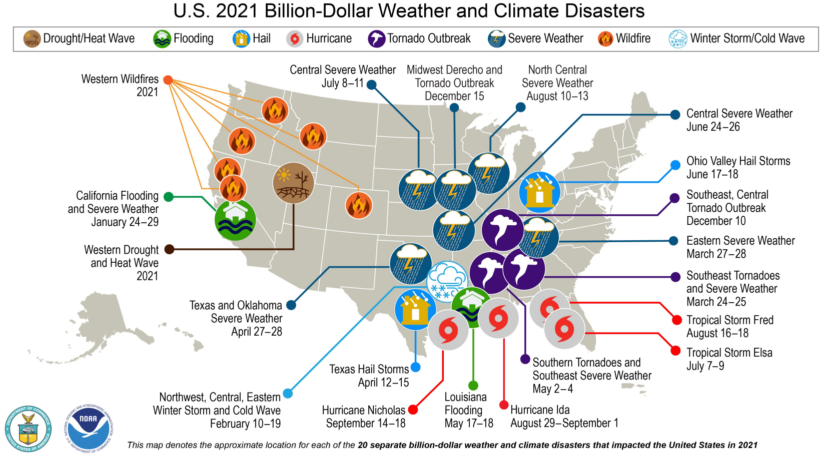
The World Bank - Urban Population Estimates

	2010	2020	2050
Urban population	249,849,720	272,364,755	334,298,000
Urban population (% of total population)	80.8	82.7	89.2

+22.5M

Worsening Climate Change

- Since 1980, **310** weather/climate disasters causing over **$2T*** in damages
- In past 5 years: **85** disasters, **$742.1B***



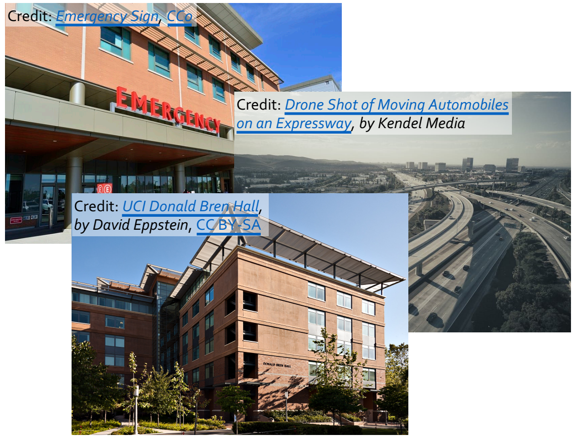
Aging & Low Investments in Modernization

- Large funding gaps for maintenance
- Systems exceeding or at end of life

Broken Pipes

Fallen Power Lines

Driving Use Case: Stormwater Infrastructure Networks



Cities and Communities

Rainfall *Stormwater*

Excess Irrigation

Oils and Greases *Unpermitted wastewater*

Pesticides *Chemicals*

Rivers, Bays, Ocean

Stormwater Network

Catch Basins

Network / Channels

Outfalls

Challenge: Addressing Pollutants

Current techniques are inadequate

- Manual inspections, citizen reports, site visits
- Water quality measured using testing kits and laboratory analysis
- 3-5 weeks to turnaround

Geo-distributed Infrastructure

- Large, regional coverage areas
- Scarce historical data
- 1000s of catch basins, outfalls as entry points

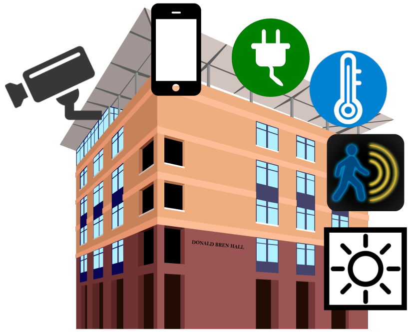
Nature of Pollutants

- *Transient phenomena*
  - e.g., bacterial decay, pollutant dilution
- *Spontaneous introduction*
  - e.g., Illegal dumping
- *Heterogeneous pollutants*

Enabling Smart Monitoring using the Internet-of-Things

Internet-of-Things (IoT)

Smart Campus[1]



Smart Healthcare[2]

Smart Agriculture[3]

High-Rise Buildings[4]

?

Stormwater Infrastructure

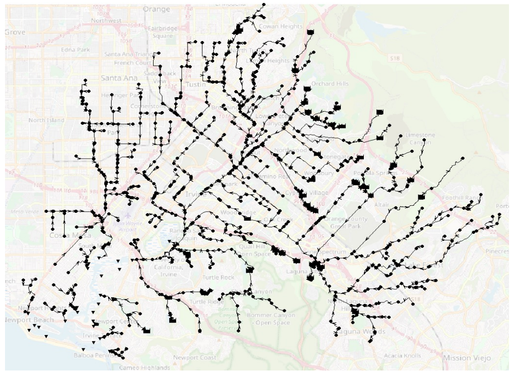
Where should IoT sensors be deployed to support reliable and effective monitoring solutions?

1 Lin, Y., Jiang, D., Yus, R., Bouloukakis, G., Chio, A., Mehrotra S., Venkatasubramanian, N. *Locater: Cleaning Wifi Connectivity Datasets for Semantic Localization*. In PVLDB 2021.
2 Catarinucci, L., De Donno, D., Mainetti, L., Palano, L., Patrono, L., Stefanizzi, M. L., Tarricone, L. *An IoT-aware architecture for smart healthcare systems*. In *IEEE IoT-J 2015*.
3 Haseeb, K., Ud Din, I., Almogren, A., Islam, N. *An energy efficient and secure IoT-based WSN framework: An application to smart agriculture*. *Sensors 2020*.
4 Liu, F., Baijnath-Rodino, J. A., Chang, T. C., Banerjee, T., Venkatasubramanian, N. *DOME: Drone-assisted Monitoring of Emergent Events For Wildland Fire Resilience*. In *ICCPS 2023*.

Gaining Insight into Sensor Deployments

Structural

Physical characteristics of network junctions and conduits



Junctions (Nodes):
- *location, elevation, depth*

Conduits (Edges):
- *length, cross-sectional area, roughness*

Gaining Insight into Sensor Deployments

Structural

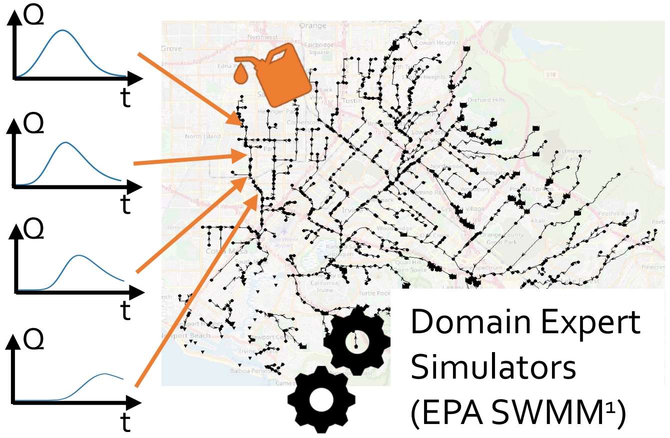
Physical characteristics of network junctions and conduits

Junctions (Nodes):
- *location, elevation, depth*

Conduits (Edges):
- *length, cross-sectional area, roughness*

Behavioral

Responses to various stimuli in the network, and their impact



Study effect and reach of anomalies on network through simulations

[1] EPA SWMM. https://www.epa.gov/water-research/storm-water-management-model-swmm

Gaining Insight into Sensor Deployments

Structural

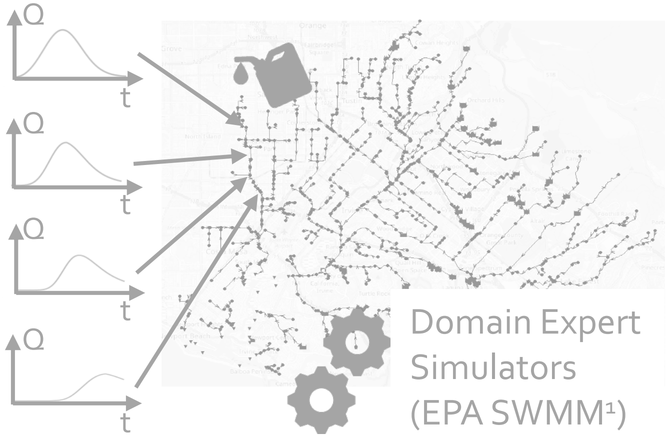
Physical characteristics of network junctions and conduits

Junctions (Nodes):
- *location, elevation, depth*

Conduits (Edges):
- *length, cross-sectional area, roughness*

Behavioral

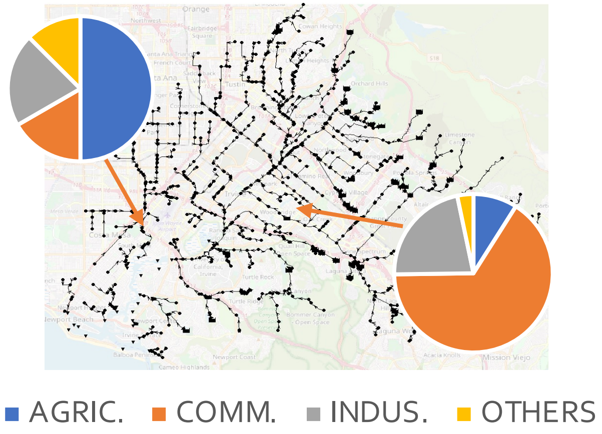
Responses to various stimuli in the network, and their impact



Study effect and reach of anomalies on network through simulations

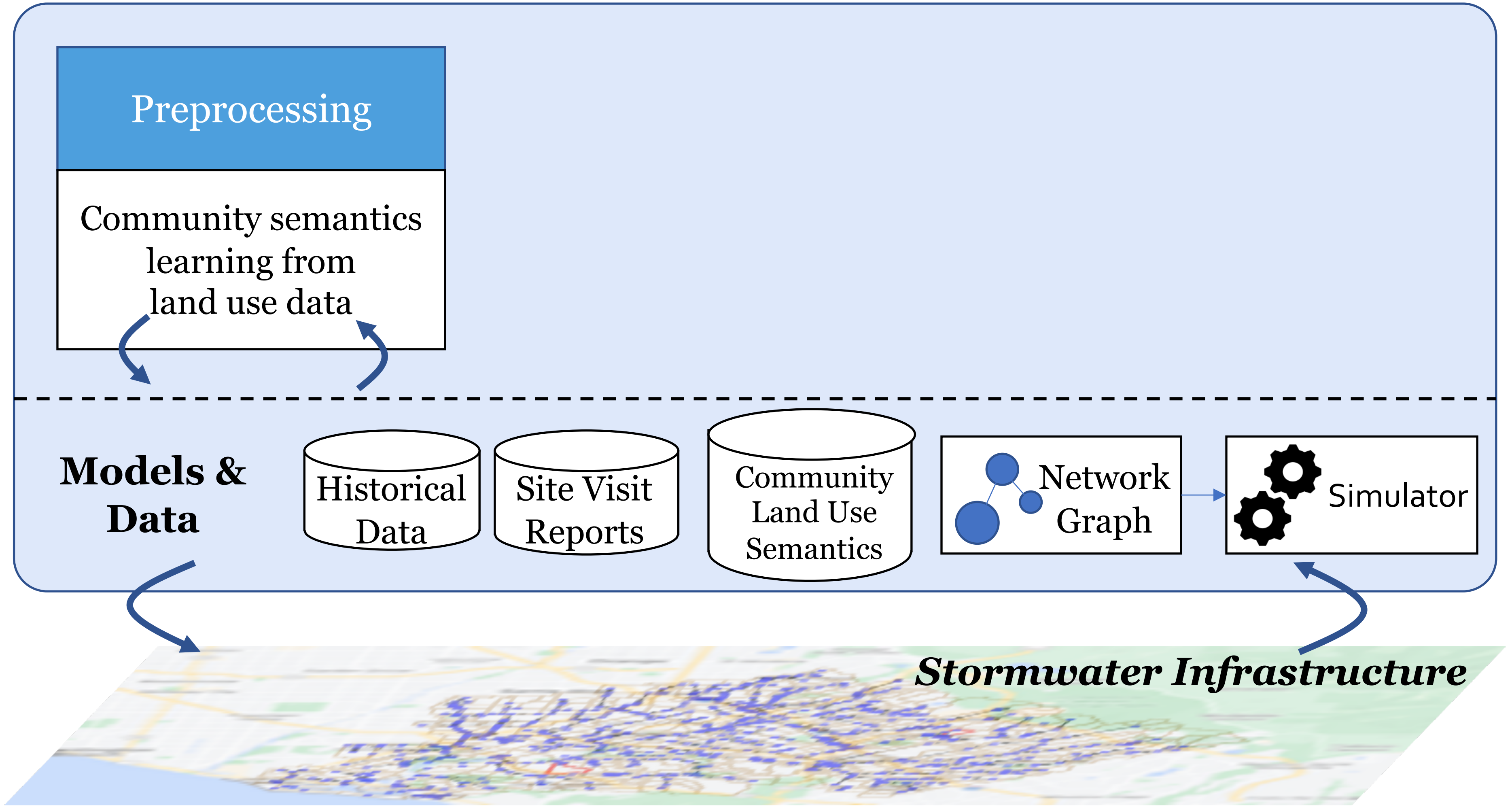
Semantic

Influences from specific land uses of a community on anomalies



Examine relationships between pollutants and potential sources

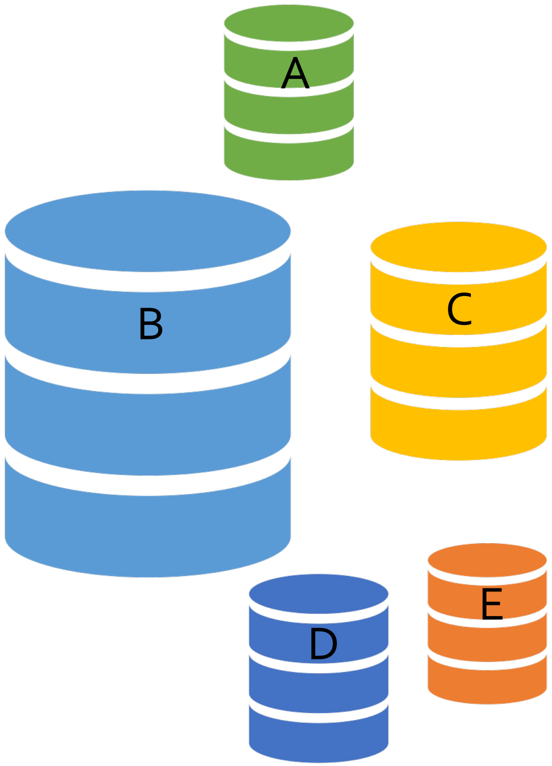
The STEP Approach



Realistic Anomalies to inform Sensor Placement

Historical Dataset D

Time t	Node v	Flow f	Obs. o
2017-05-23, 10:02	FVES@D05	0.03	Turbidity: 1.2
2017-06-30, 08:22	FVES@D05	0.14	Elec. Cond.: 906
2019-05-15, 08:00	COF07S01	0.28	Turbidity: 4.03
2019-07-24, 09:10	COF07S01	0.21	Temp. : 26.24
2019-08-07,08:28	GGC02S0172DS	0.09	Elec. Cond.: 6543
2019-07-17,08:25	IRVF06P06	0.11	Turbidity: 3.5
...	...	...	...



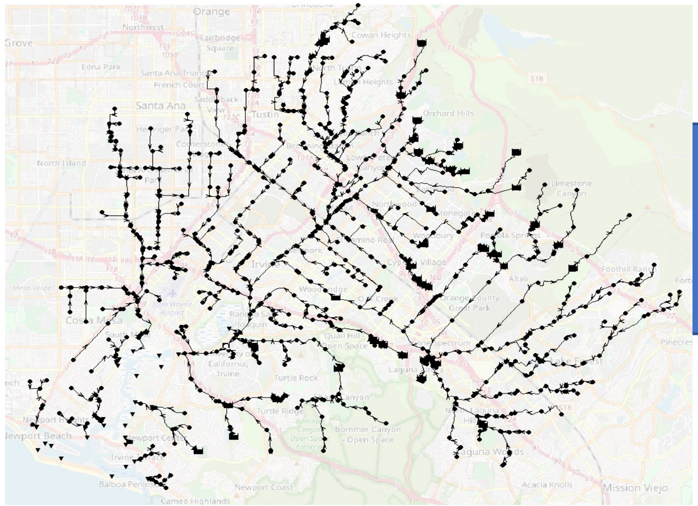
Sets of Realistic Anomalies

How can we learn from past instances of anomalies?

Extracting Anomalies from Water Quality Data

Learn different types of behaviors in the network

Stormwater Network



Simulated Anomalies

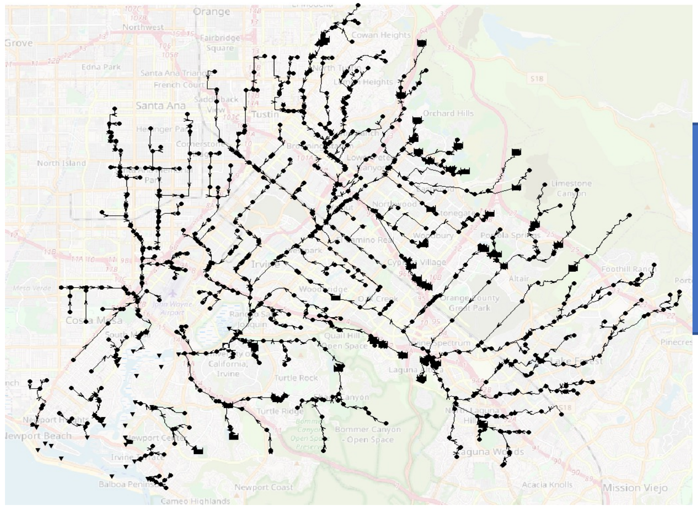
Generate uniformly distributed anomalies
Simulate and cache behavior in network

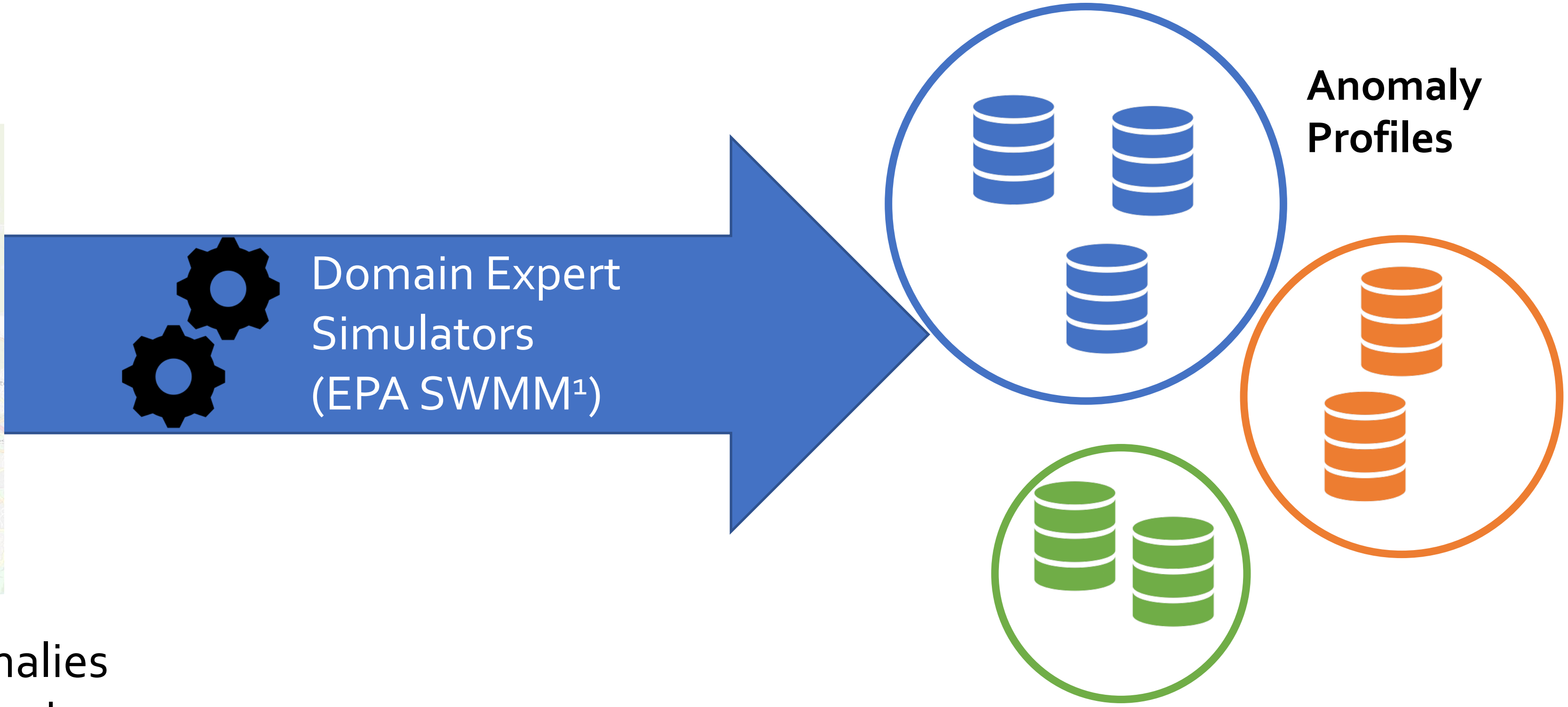
- *Reachability from an origin*
- *Duration of anomaly*
- *Phenomenon produced*
- *Nearby community-level semantics*

[1] EPA SWMM. https://www.epa.gov/water-research/storm-water-management-model-swmm

Extracting Anomalies from Water Quality Data

Stormwater Network





Generate uniformly distributed anomalies
Simulate and cache behavior in network

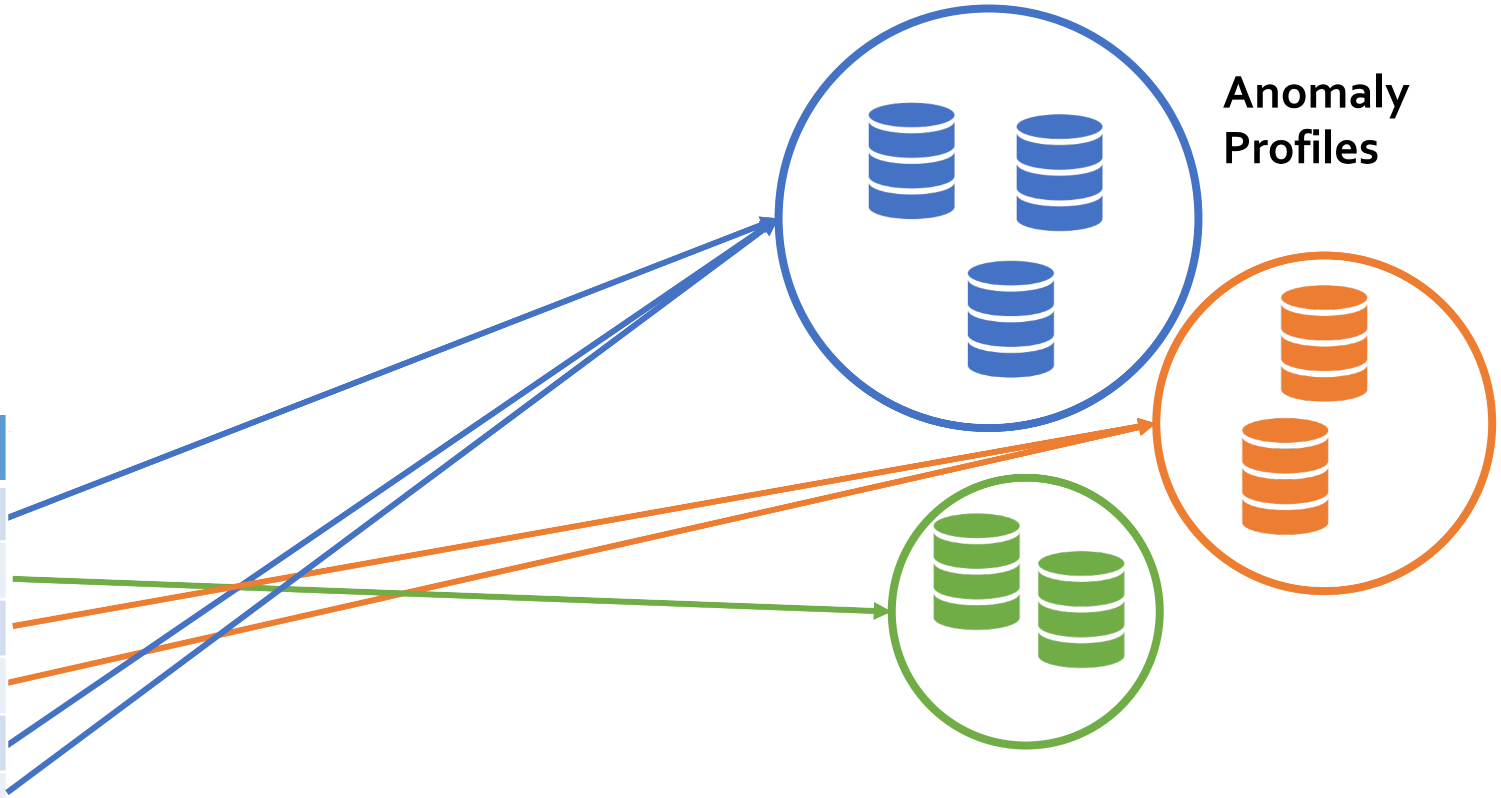
Apply *agglomerative clustering* to group anomalies into ***profiles***, based on similarity of impact (behavior) in network

[1] EPA SWMM. https://www.epa.gov/water-research/storm-water-management-model-swmm

Extracting Anomalies from Water Quality Data

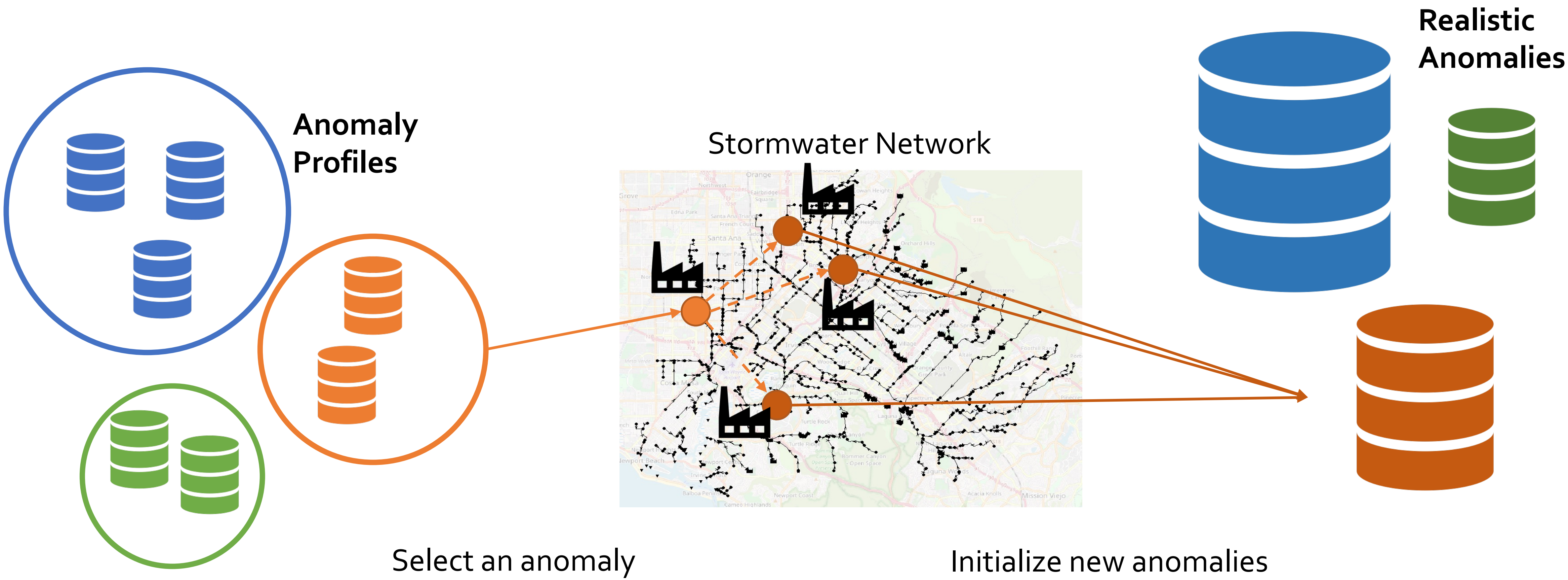
Historical Dataset D

Time t	Node v	Flow f	Obs. o
2017-05-23, 10:02	FVES@D05	0.03	Turbidity: 1.2
2017-06-30, 08:22	FVES@D05	0.14	Elec. Cond.: 906
2019-05-15, 08:00	COF07S01	0.28	Turbidity: 4.03
2019-07-24, 09:10	COF07S01	0.21	Temp. : 26.24
2019-08-07,08:28	GGC02S0172DS	0.09	Elec. Cond.: 6543
2019-07-17,08:25	IRVF06P06	0.11	Turbidity: 3.5
…	…	…	…



Match historical instances to anomaly profiles to weigh them by probability of occurrence

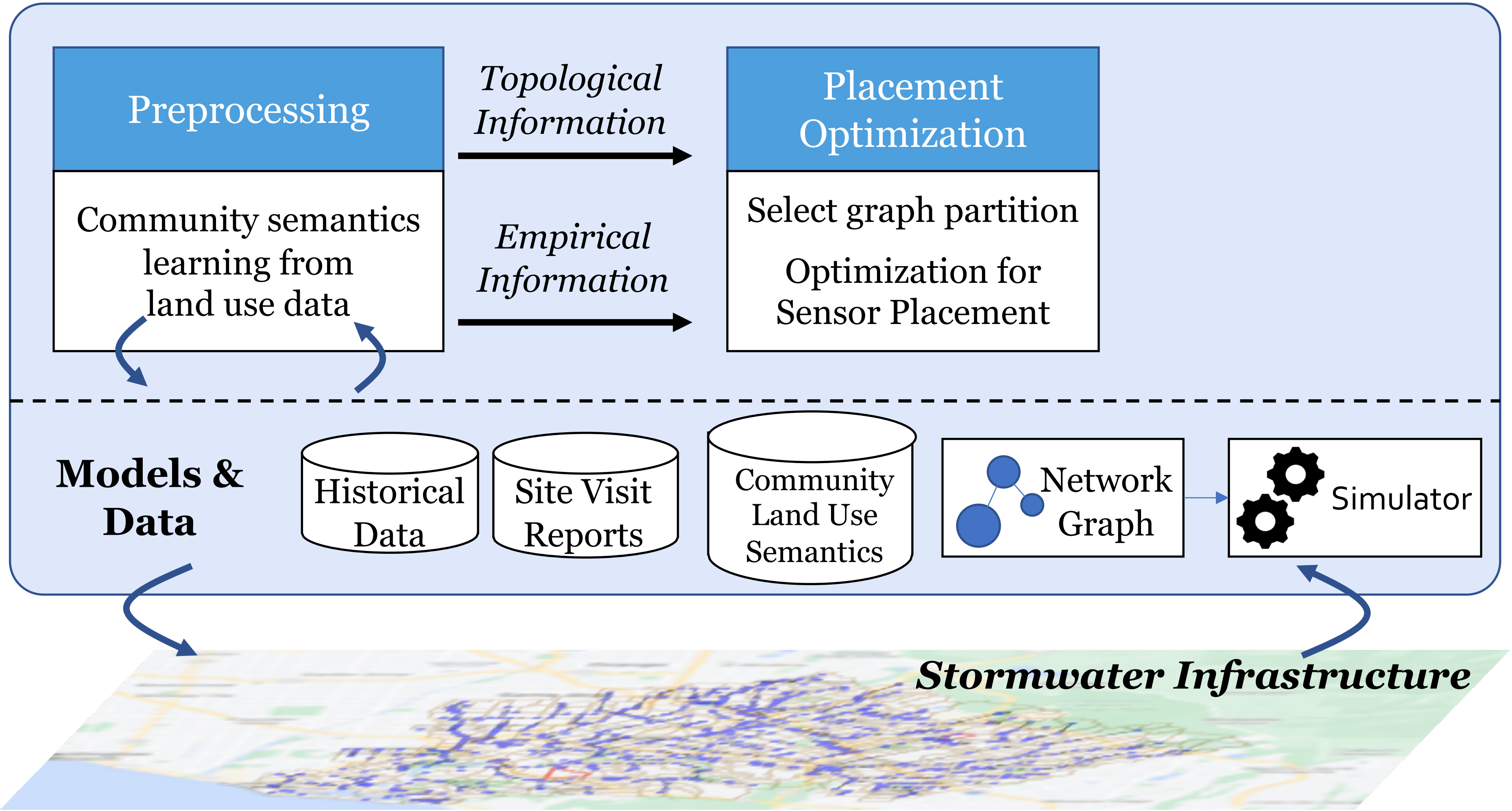
Generating Realistic Anomalies using Semantics



Select an anomaly (and corresponding semantic land use)

Initialize new anomalies at other locations with same semantic land use

The STEP Approach



Optimization Objectives

Objective: Coverage COV

Ability of a placement to capture and observe anomalies in the network

$$COV(\mathcal{X},\mathcal{A},\mathcal{G}) = \frac{1}{|\mathcal{V}|}\sum_{v_i\in\mathcal{V}} covered(v_i,\mathcal{X},\mathcal{A}(v_i))$$

$$covered(v_i,\mathcal{X},\mathcal{A}(v_i)) = \mathbb{1}\left[\sum_{\alpha_k\in\mathcal{A}(v_i)}\sum_{v_j\in\mathcal{V}}\sum_{s_l\in\mathcal{S}} x_{lj}OB(l,k)PT(k,j) \geq \rho\,|\mathcal{A}(v_i)|\right]$$

Objective: Traceability TR

Ability of a placement to use observations to track the origin of an anomaly

$$TR(\mathcal{X},\mathcal{A},\mathcal{G}) = \frac{1}{|\mathcal{A}|}\sum_{\alpha_k\in\mathcal{A}}\sum_{s_l\in\mathcal{S}(\mathcal{P}_k)}\sum_{v_j\in\mathcal{V}} \left|\mathcal{V}^{up}_{v_j,\alpha_k,\mathcal{X}}\right| \Big/ |\mathcal{V}|$$

Betweenness Centrality $\mathcal{BTN}$

- ***# of anomalies passing through a node***
- $\mathcal{BTN}(v) = \sum_{\alpha\in\mathcal{A}} 1[time(\alpha,v^*,v)\leq\tau]$

Branching Complexity $\mathcal{BC}$

- ***Degree of merging/splitting at upstream nodes***
- $\mathcal{BC}(v_j) = \begin{cases} 1 & if\ IsRoot(v_j) \\ \max\limits_{v_i\in pa(v_j)} \mathcal{BC}(v_i) + \sum_{v_i\in pa(v_j)} \frac{\mathcal{BC}(v_i)}{\max\limits_{v_i\in pa(v_j)} \mathcal{BC}(v_i)} & otherwise \end{cases}$

Semantic Entropy $\mathcal{SE}$

- ***Skewness in distribution of upstream semantic land uses***
- $\mathcal{SE}(\mathcal{U},\mathcal{G}) = \sum_{u\in\mathcal{U}} \lambda_m(-P(u_m)\log P(u_m))$

  where $P(u_m) = \sum_{v_i\in\mathcal{V}^{up}_{v_j}} (Area(v_i,u_m)/\sum_{u_m\in\mathcal{U}} Area(v_i,u_m))$

The STEP Sensor Deployment Optimization

Algorithm 2: Sensor Placement

Input: Graph $\mathcal{G}$, Sensors $\mathcal{S}$, Anomalies $\mathcal{A}$, int N_{part}, Budget B
Output: Placement $\mathcal{X}$

$placements \leftarrow \emptyset; G_{subs} \leftarrow \emptyset; best \leftarrow 0; v_{part} \leftarrow null$
for $i \leftarrow 1..N_{part}$ **do**
  for $v_j \in \mathcal{V}$, *for each subgraph* **do**
    $score \leftarrow w_{cov}\Delta\mathcal{BTN}(v_j) + \frac{w_{tr}}{2}\Delta\mathcal{BC}(v_j) + \frac{w_{tr}}{2}\Delta\mathcal{SE}(v_j)$
    if $score < best$ **then** $best \leftarrow (score, v_j)$;
    $G_{subs} \leftarrow G.AddSplit(v_{part})$
$nodes \leftarrow GetPartitionNodes(G_{subs})$
$placement \leftarrow placement \cup Sensorize(nodes)$
for $\mathcal{G}' \leftarrow \mathcal{G}_{subs}$ **do**
  $\mathcal{A}' \leftarrow \bigcup_{v \in \mathcal{V}'} \mathcal{A}(v)$
  $B' \leftarrow B \cdot \frac{|\mathcal{A}'|}{|\mathcal{A}|}$;
  $\mathcal{X}' \leftarrow$ Use MILP to solve Eqn. 7 with budget B'
  Add $\mathcal{X}'$ to *placements*
$\mathcal{X} \leftarrow AdjustPlacement(placement, hops = 5)$
return $\mathcal{X}$

Intuition:

Initialize variables for partitioning

Construct N_{part} graph partitions at key nodes

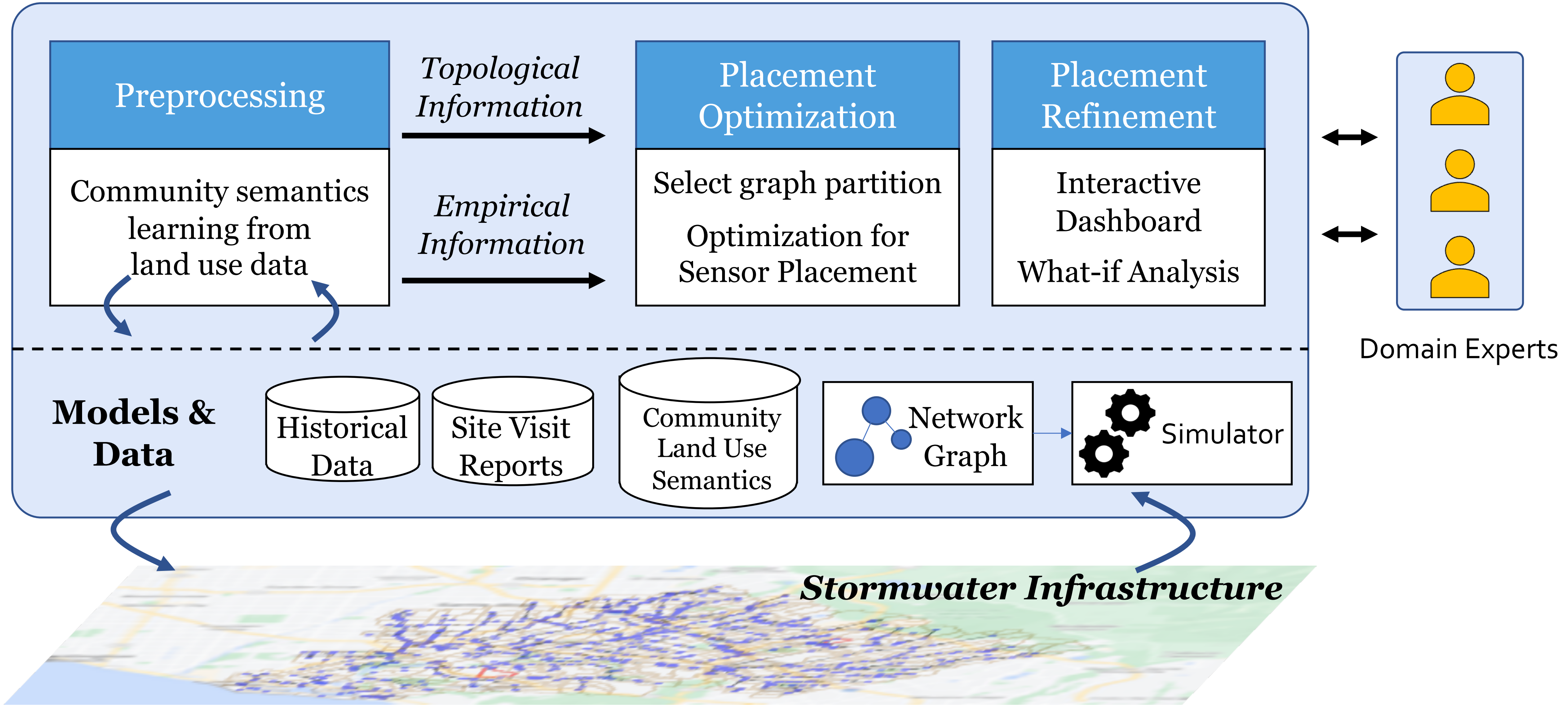
Score nodes based on their betweenness centrality (BTN), branching complexity (BC), and semantic entropy (SE)

Instrument key locations with sensors

Obtain ideal placements using MILP within each partitioned subgraph

Merge partial solutions, and adjust as needed

The STEP Approach



Placement Refinement

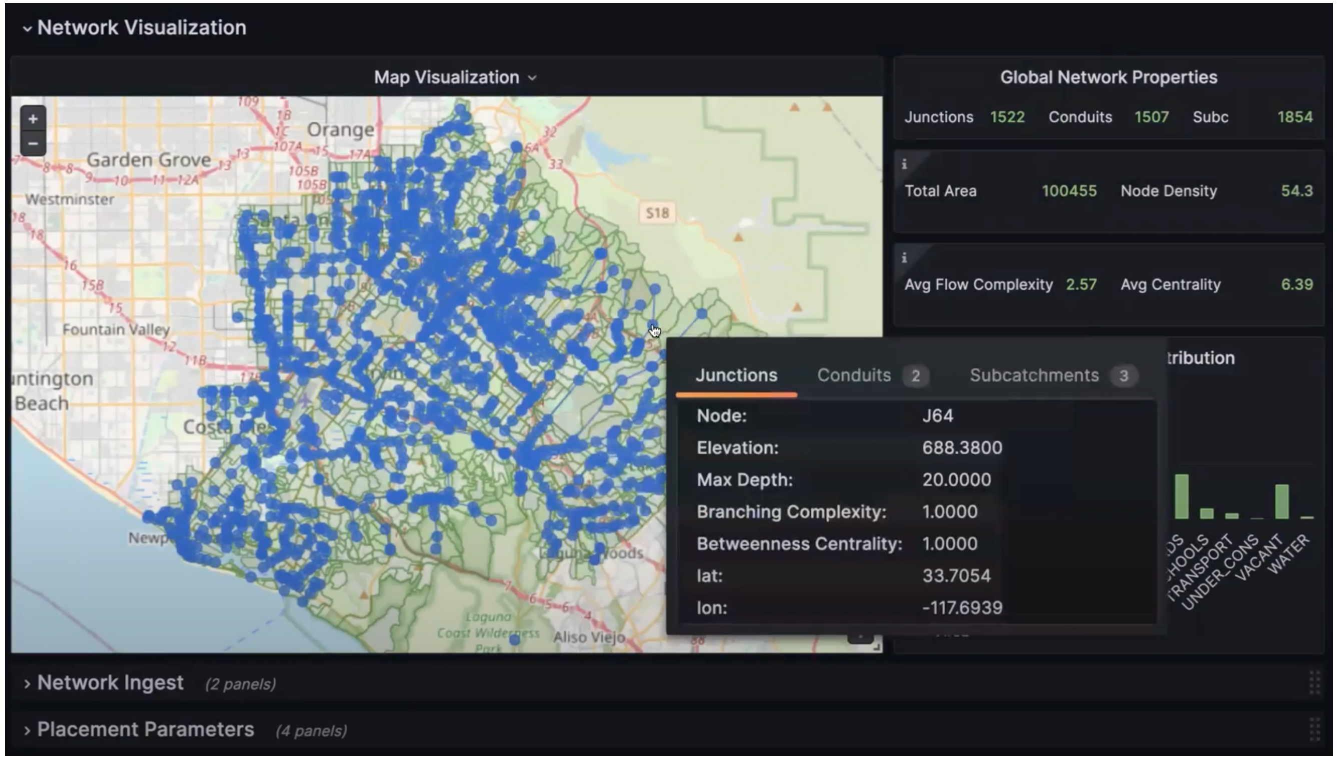
An ideal placement may not be practical!

Potential Vandalism

Location-specific communication issues

Physical barriers to easy human access

The STEP toolkit includes a dashboard for domain experts to refine a potential placement as needed

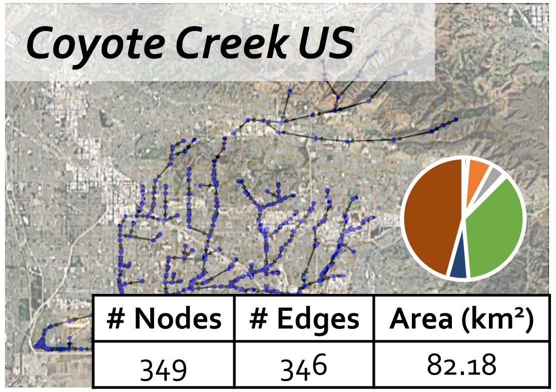


https://github.com/andrewgchio/STEP

Experiments: 6 Real-World Stormwater Networks

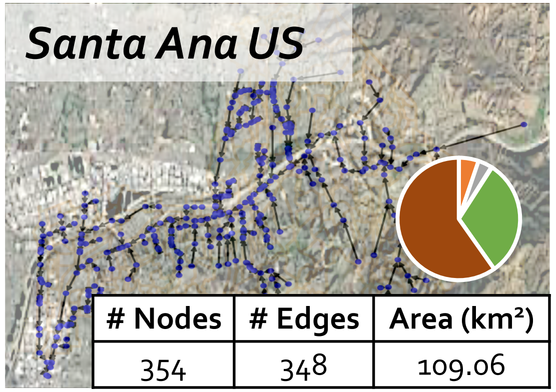
Real-World Networks

- 6 EPA SWMM[1] networks of stormwater systems in Southern California, USA of varying sizes
- Considered 7 different types of semantic land uses in networks
- *Provided by Orange County Public Works (OCPW)*



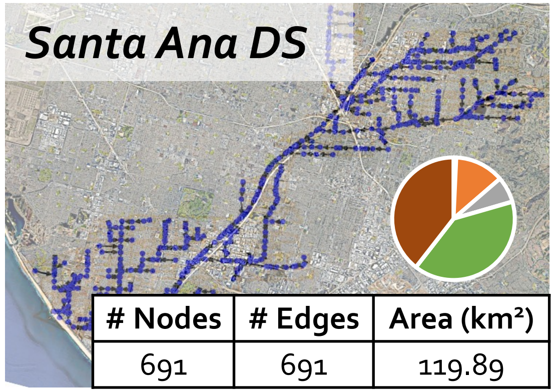
Coyote Creek US

# Nodes	# Edges	Area (km²)
349	346	82.18



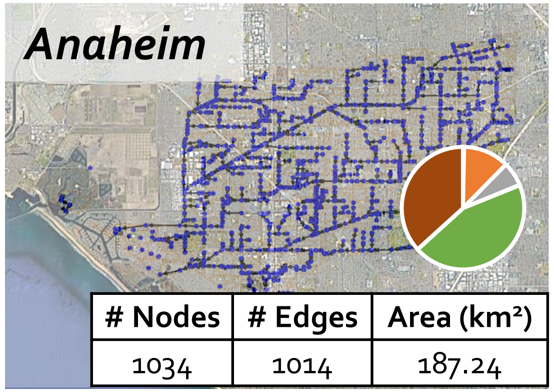
Santa Ana US

# Nodes	# Edges	Area (km²)
354	348	109.06



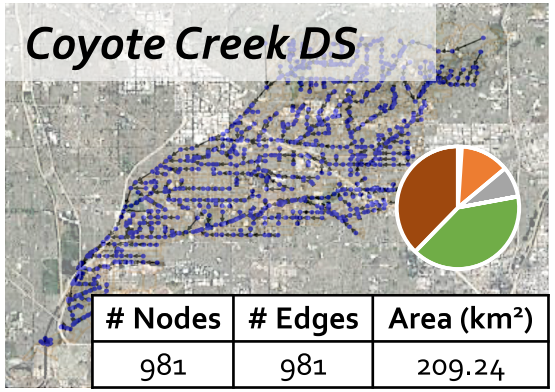
Santa Ana DS

# Nodes	# Edges	Area (km²)
691	691	119.89



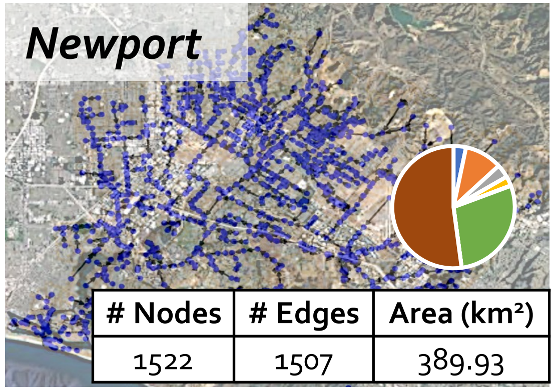
Anaheim

# Nodes	# Edges	Area (km²)
1034	1014	187.24



Coyote Creek DS

# Nodes	# Edges	Area (km²)
981	981	209.24



Newport

# Nodes	# Edges	Area (km²)
1522	1507	389.93

AGRICULTURE COMM_SERV INDUSTRIAL MIX_COMM
MIX_URBAN HI_DENSITY LO_DENSITY OTHER

OC Public Works

[1] EPA SWMM. https://www.epa.gov/water-research/storm-water-management-model-swmm

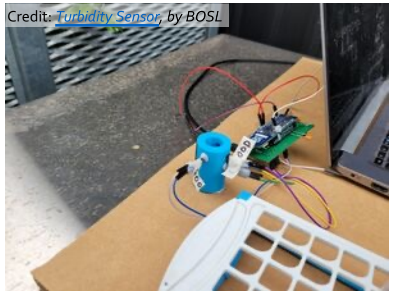
Sensors, Historical Data, and Anomalies

Sensors Considered [1,2,3]

- 5 real types of sensors considered

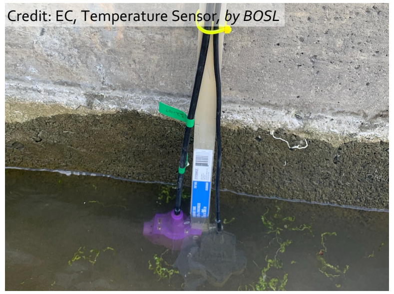
Table 1: Sensors considered in placement

Phenomenon	Accuracy	Hardware & Depl. Cost	Op. Cost
Turbidity	11.6%	\$100	\$300
Depth	1 mm	\$150	\$350
Temperature	$0.5°C$	\$200	\$300
Electric Cond.	10%	\$150	\$300
Velocity	5 mm/s	\$150	\$350



Turbidity Sensor

Depth Sensor



Electrical Conductivity, Temperature Sensor

Historical Data

OCPublicWorks

- 1292 historical grab samples from 30 locations
- Spans 16 years from 2006 to 2022
- *Provided by OCPW*

Anomalies

- Random anomalies defined uniformly across nodes in networks,
  - Random duration 30 ± 5 minutes
  - Random flow rate 0.2 ± 0.2 cfs,
  - Randomly sampled phenomena
- Realistic set of anomalies (derived from historical data) for evaluation

[1] S Catsamas et al., *Characterisation and development of a novel low-cost radar velocity and depth sensor*. In SPN 2022.
[2] B. Shi et al. *A low-cost water depth and electrical conductivity sensor for detecting inputs into urban stormwater networks*. In Sensors 2021.
[3] M Wang et al., *An Innovative Low-cost Turbidity Sensor for Long-term Turbidity Monitoring in the Urban Water System*. In ICUD 2021.

Baselines and Performance Metrics

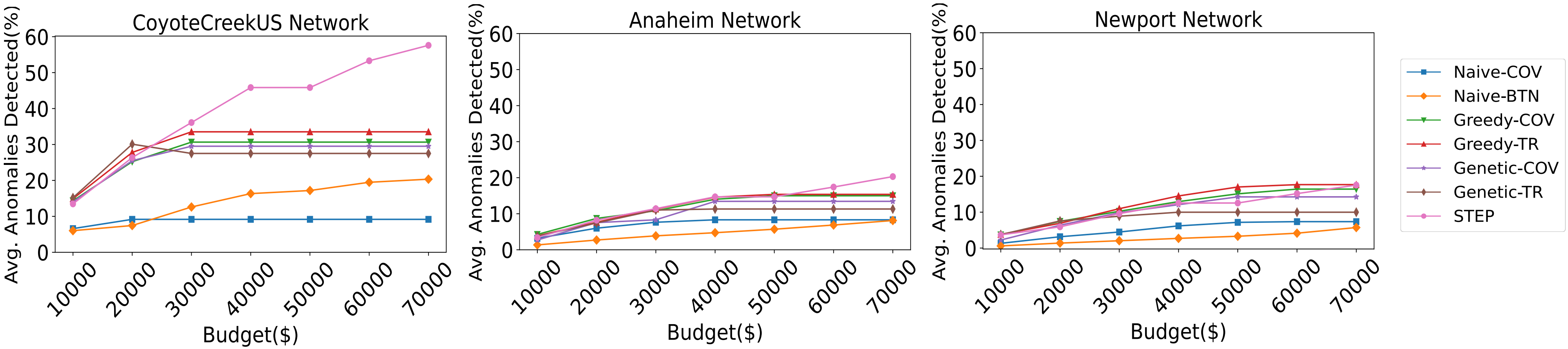
Baseline Comparisons

- Greedy Heuristics: select node/sensor to optimize heuristic
  - *Naïve-COV (radial), Naïve-BTN*
  - *Coverage, Traceability*
- Genetic Algorithm: simulates natural selection/evolution
  - *Coverage, Traceability*

Performance Measures

- Number of (realistic) *anomalies detected*
- *Traceability* of anomalies
- *Coverage* of nodes in network

Experiment 1: Number of Anomalies Detected



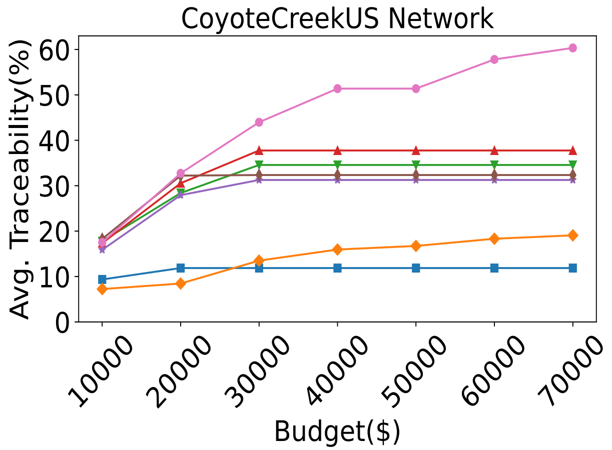
(a) Coyote Creek Upstream (Small 1) (d) Anaheim (Med 1) (f) Newport Beach (Large 1)

- On average, STEP detected **35-528%, 32-309% and 1-207% more anomalies** for the small, medium, and large networks than the baseline approaches

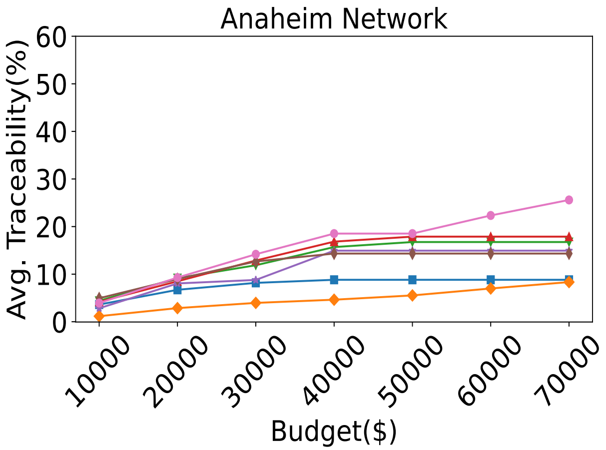
- Difference in efficacy of STEP is more pronounced in smaller networks

STEP proposes a heterogenous sensor placement that is effective at supporting detection of realistic anomalies

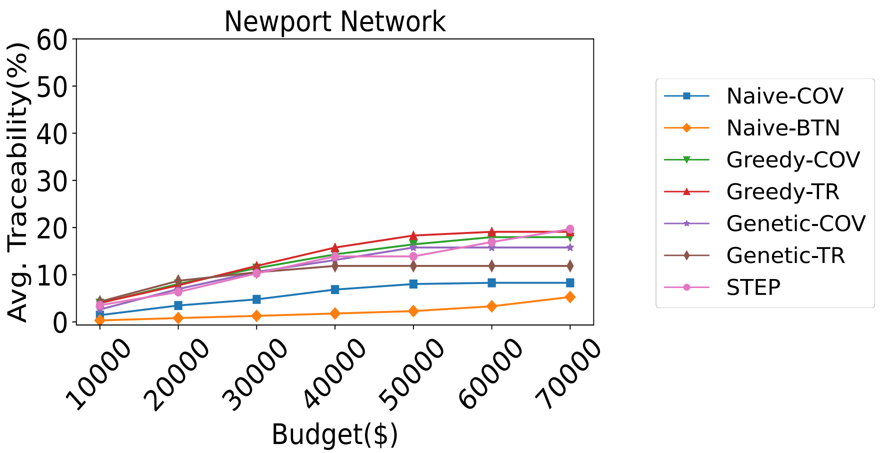
Experiment 2: Traceability



(a) Coyote Creek Upstream (Small 1)



(d) Anaheim Network (Med 1)



(f) Newport Beach (Large 1)

- On average, traceability of anomalies supported by STEP was **30-671%, 43-400% and 3-272% better** for the small, medium, and large networks than the baseline approaches

STEP proposes a heterogenous sensor placement that is effective at supporting potential traceability of anomalies after they occur

Key Takeaways and Future Directions

- We developed **STEP**: a system for sensor deployments that integrates *structural*, *behavioral*, and *semantic* aspects of an infrastructure
  - *A novel anomaly generator based on community-level semantics*
  - *A graph partitioning + optimization leveraging key network properties*
  - *A prototype system for deployment refinement*

- Our experiments show the efficacy of this approach on 6 real-world networks

- **Future Directions:**
  - Leverage proposed sensor deployment in a real stormwater network
  - Provide analysis support for pollutant source identification

- **Our code is publicly available on GitHub**: https://github.com/andrewgchio/STEP

- **Acknowledgements**:
  - NSF SWADE Project (https://www.sites.uci.edu/swade)
  - UC National Laboratory Fees Research Program – Los Alamos National Laboratory