NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2023 Track on Datasets and Benchmarks

UUKG: Unified Urban Knowledge Graph Dataset for Urban Spatiotemporal Prediction

Yansong NING[1], Hao LIU[1,2], Hao WANG[3], Zhenyu ZENG[3], Hui XIONG[1,2]

[1]The Hong Kong University of Science and Technology (Guangzhou)
[2]The Hong Kong University of Science and Technology
[3]Alibaba Cloud Intelligence Group

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Alibaba

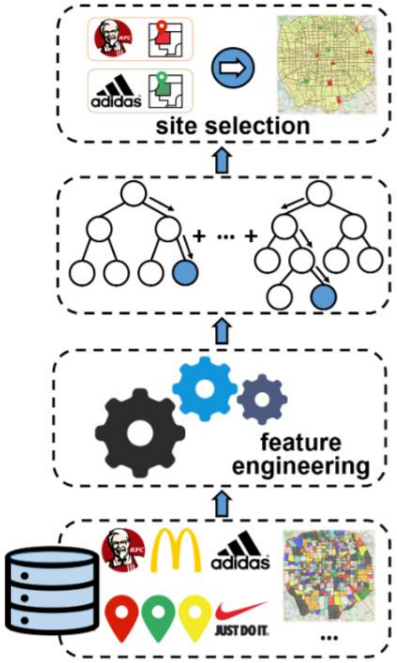
OUTLINE

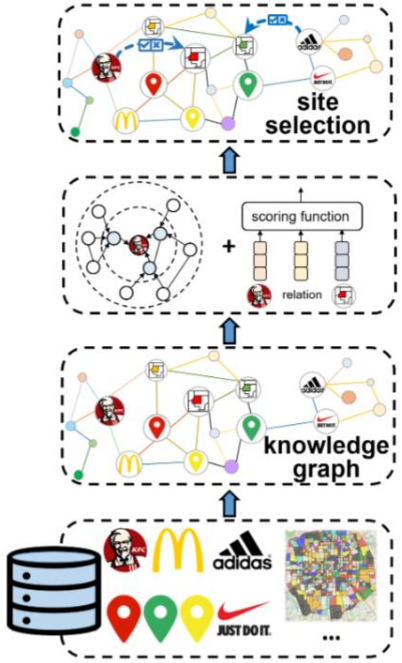
- **Motivation and Challenge**
- **Urban Knowledge Graph Construction**
- **Structure-aware Urban Knowledge Graph Embedding**
- **Knowledge-enhance Urban SpatioTemporal Prediction**
- **Conclusion and Future Work**

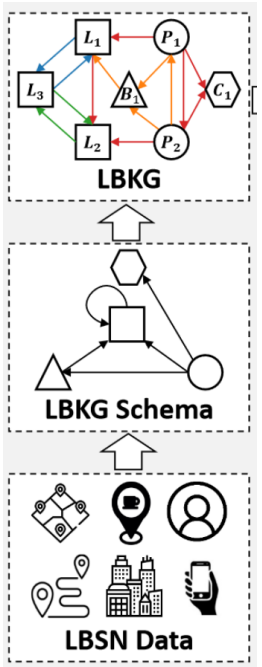
Motivation

Existing UrbanKGs

- Task-specific and publicly unavailable
- Limits the flourishing of knowledge-enhanced urban spatiotemporal prediction







[1] Liu, et al. Knowledge-driven site selection via urban knowledge graph. arXiv preprint arXiv:2111.00787 (2021). [2] https://www.iconexperience.com/
[3] Zhou, et al. Hierarchical knowledge graph learning enabled socioeconomic indicator prediction in location-based social network. Proceedings of the ACM Web Conference. 2023.

Challenge

How to Construct A Unified Urban Knowledge Graph

- The urban data describe the urban system from different aspects and granularities and they are usually disjoint
- It is appealing to extract and align heterogeneous urban knowledge in a unified graph organization

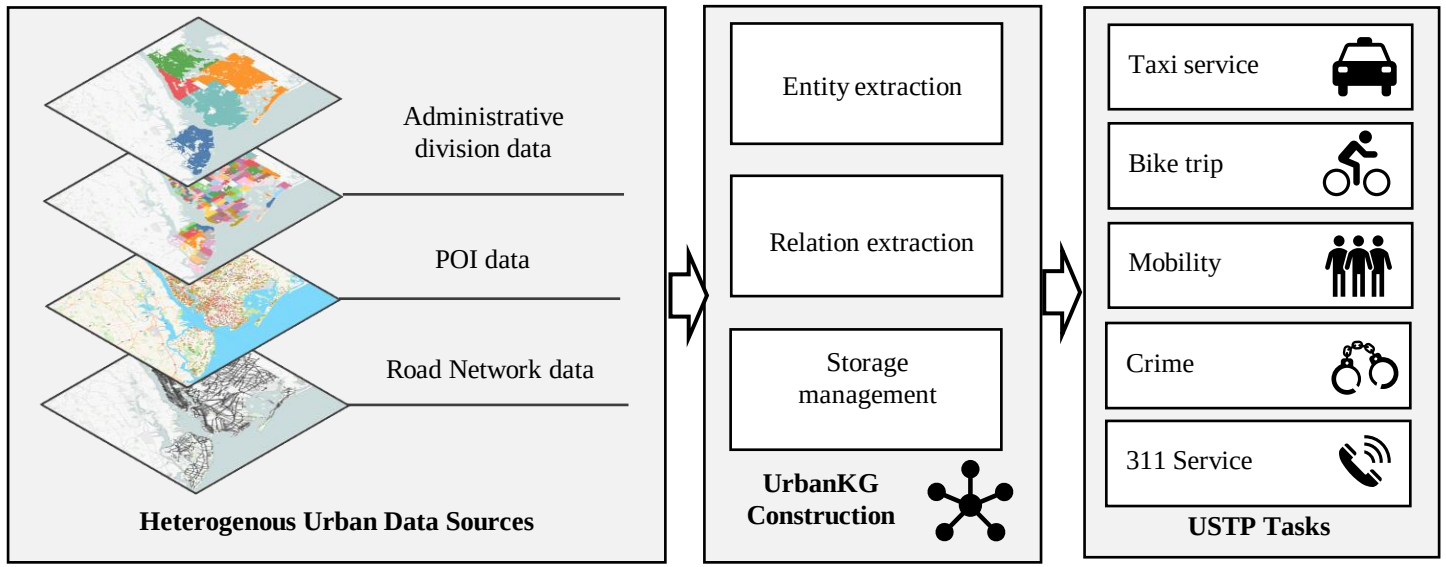
How to Preserve Complicated Structural Urban Knowledge

- Urban knowledge graph includes diverse structures such as hierarchy and cycle
- Capture such high-order semantic knowledge to empower downstream spatiotemporal prediction tasks

Unified Urban Knowledge Graph Dataset

Workflow

- Construct an urban knowledge graph from multi-sourced urban data
- Organize entities into a multi-relational heterogeneous graph
- Encode various high-order structural patterns in a unified configuration
- Facilitate joint processing for various downstream urban spatiotemporal prediction task



OUTLINE

Motivation and Challenge

Urban Knowledge Graph Construction

Structure-aware Urban Knowledge Graph Embedding

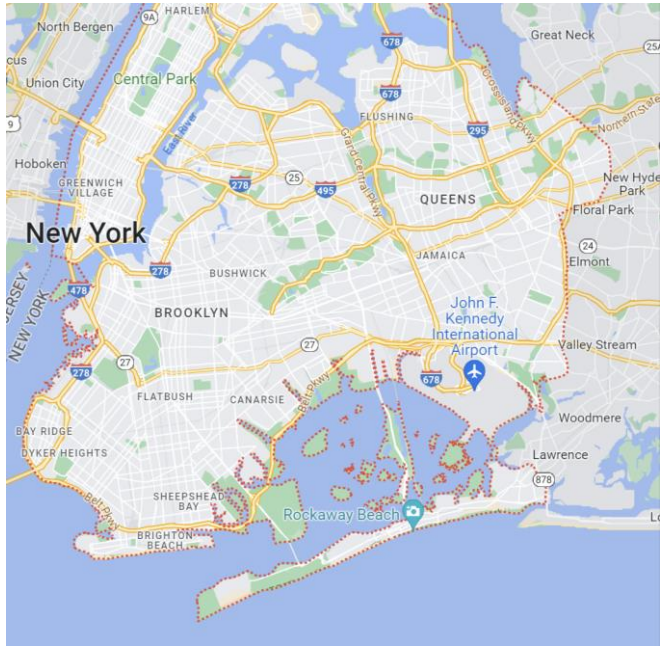
Knowledge-enhance Urban SpatioTemporal Prediction

Conclusion and Future Work

Urban Knowledge Graph (UrbanKG) Construction

UrbanKG Definition

- The UrbanKG is defined as a multi-relational graph $G = (E, R, F)$, where E, R, and F is the set of urban entities, relations and facts, respectively. In particular, facts are defined as set $\{\langle h, r, t\rangle | h, t \in E, r \in R\}$, where each triplet $\langle h, r, t\rangle$ describes that head entity h is connected with tail entity t via relation r



[1] https://www.google.com/maps [2] https://zola.planning.nyc.gov/

Urban Knowledge Graph (UrbanKG) Construction

Data Collection and Preprocessing

- Data sources: Open Street Map, NYC Gov and CHI Gov
- Data types:
  - ✓ Administrative division data: borough and area
  - ✓ Road Network data: segment and junction
  - ✓ Point-of-Interest (POI) data

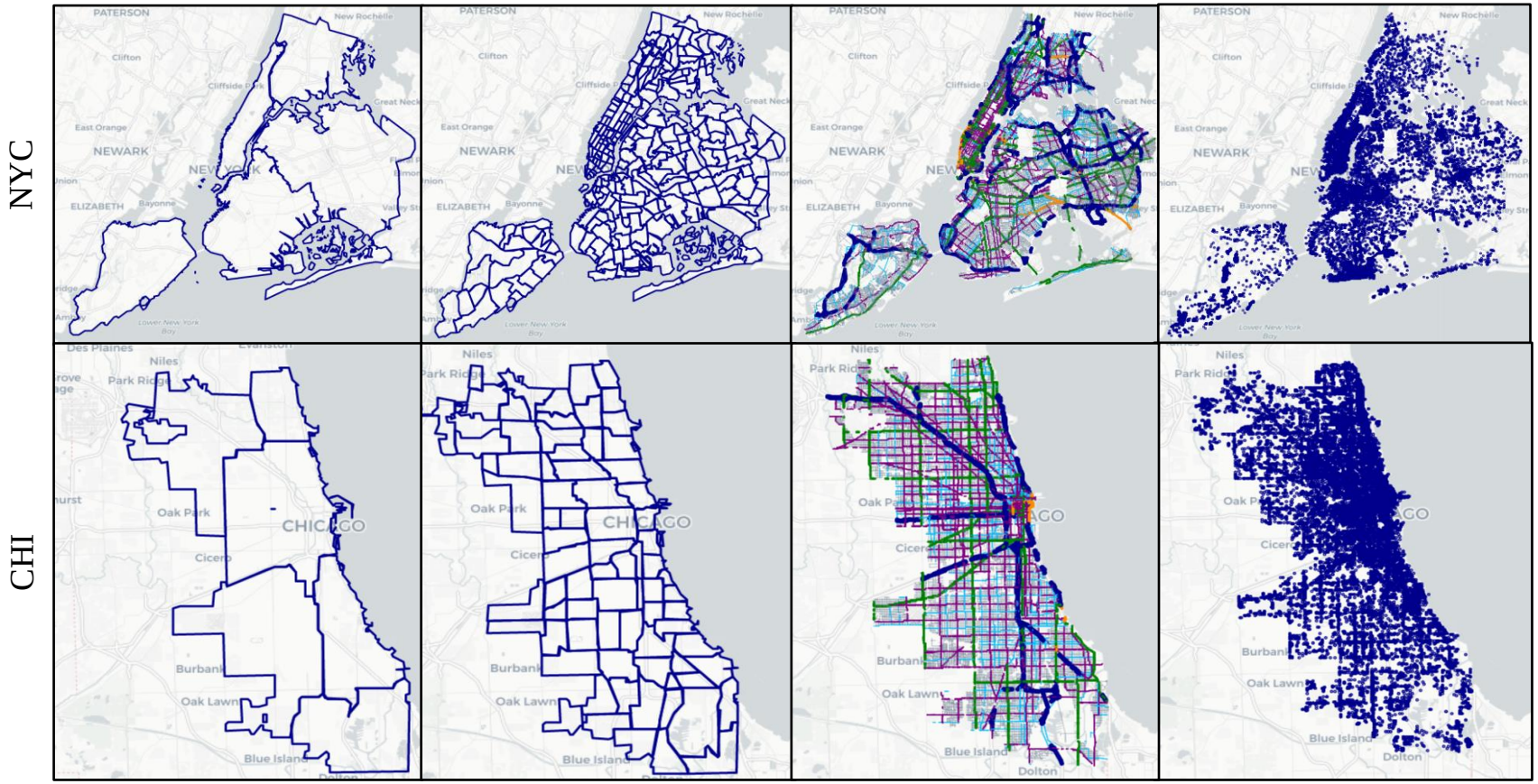
Dataset	Description	Sample format	Records New York	Records Chicago
Administrative division	Boundary	*["New Yorck", range(40.50, 40.91, -74.25, -73.70)]*	1	1
	# of Borough	*["Queens", polygon(40.54 -73.96, ...)]*	5	6
	# of Area	*["Jamaica", "Queens", polygon(40.69 -73.82, ...)]*	260	77
Road network	# of Segment	*[road id, road name, start junction, end junction, type, line range]*	110,919	71,578
	# of Category	*[191751, "Queens Boulevard", 59378, 4798, tertiary, line(40.73 -73.82, ...)]*	5	5
	# of Junction	*[junction id, junction type, coordinate]*	62,627	37,342
	# of Category	*[59378, crossing, coordinate(40.78 -73.98)]*	5	6
POI	# of POI	*[poi id, poi name, poi type, coordinate]*	62,450	31,573
	# of Category	*[34633854, "Empire State Building", corporation, coordinate(40.75, -73.99)]*	15	15

[1] https://www.openstreetmap.org/ [2] https://www.chicago.gov/ [3] https://www.nyc.gov/

Urban Knowledge Graph (UrbanKG) Construction

Entity Extraction

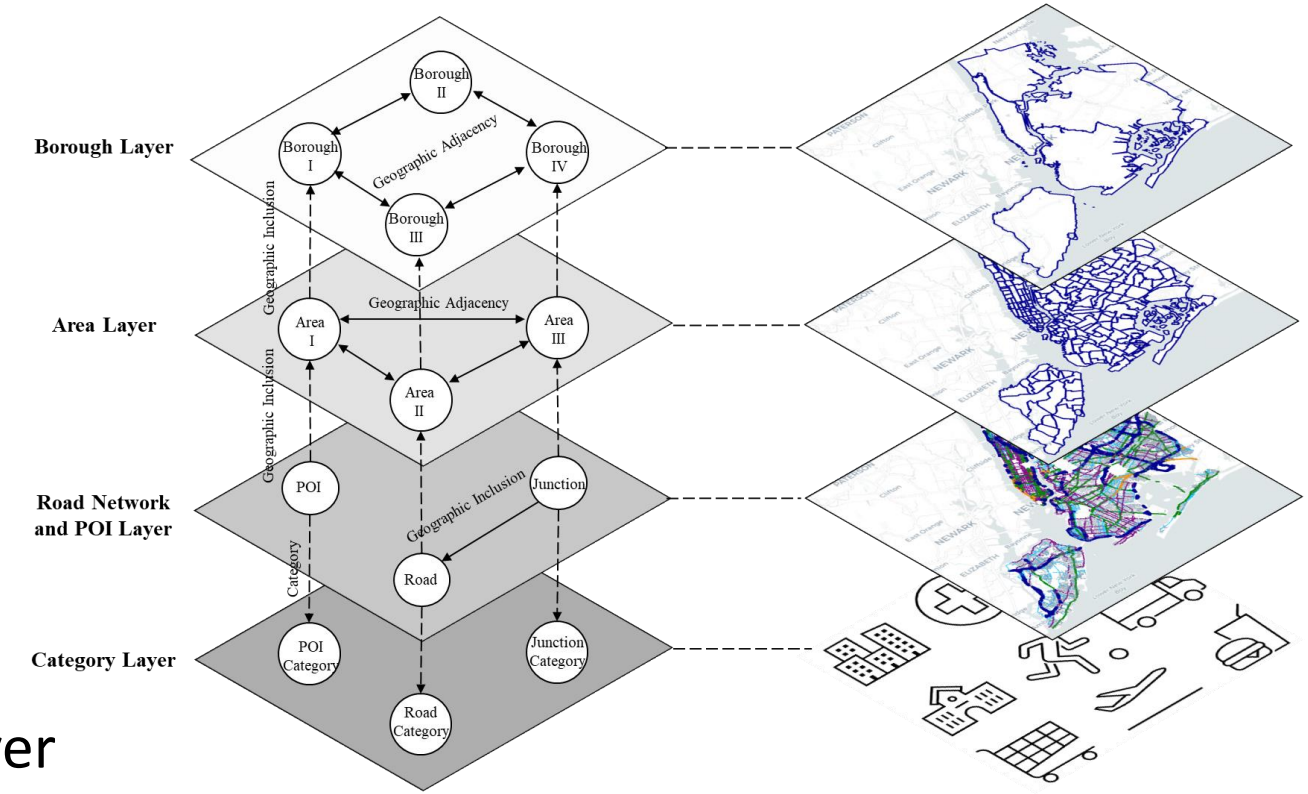
- Eight types of entities: borough, area, POI, road, junction, POI category, road category, junction category



Urban Knowledge Graph (UrbanKG) Construction

Relation Construction

- Three types of relations:
  - ✓ Geographic inclusion
  - ✓ Geographic adjacency
  - ✓ Category
- Four layers of UrbanKG:
  - ✓ Borough layer
  - ✓ Area layer
  - ✓ Road network and POI layer
  - ✓ Category layer



Relation	Symmetric	Head & Tail Entity	Abbrev.	Relation	Symmetric	Head & Tail Entity	Abbrev.
POI Locates at Area	×	(POI, Area)	PLA	Junction Belongs to Road	×	(Junction, Borough)	JBR
Road Locates at Area	×	(Road, Area)	RLA	Borough Nearby Borough	✓	(Borough, Borough)	BNB
Junction Locates at Area	×	(Junction, Area)	JLA	Area Nearby Area	✓	(Area, Area)	ANA
POI Belongs to Borough	×	(POI, Borough)	PBB	POI Has POI Category	×	(POI, PC)	PHPC
Road Belongs to Borough	×	(Road, Borough)	RBB	Road Has Road Category	×	(Road, RC)	RHRC
Junction Belongs to Borough	×	(Junction, Borough)	JBB	Junction Has Junction Category	×	(Junction, JC)	JHJC
Area Locates at Borough	×	(Area, Borough)	ALB	-	–	–	–

Urban Knowledge Graph (UrbanKG) Construction

Statistics of NYC and CHI UrbanKGs

- Two multi-level large-scale UrbanKGs
- Millions of triplets are preserved

Dataset	# Entity	# Relation	# Triplet	# Train	# Valid	# Test
NYC	236,287	13	930,240	837,216	46,512	46,512
CHI	140,602	13	564,400	507,960	28,220	28,220

[1] https://www.cntraveler.com/ [2] https://www.telegraph.co.uk/

OUTLINE

Motivation and Challenge

Urban Knowledge Graph Construction

Structure-aware Urban Knowledge Graph Embedding

Knowledge-enhance Urban SpatioTemporal Prediction

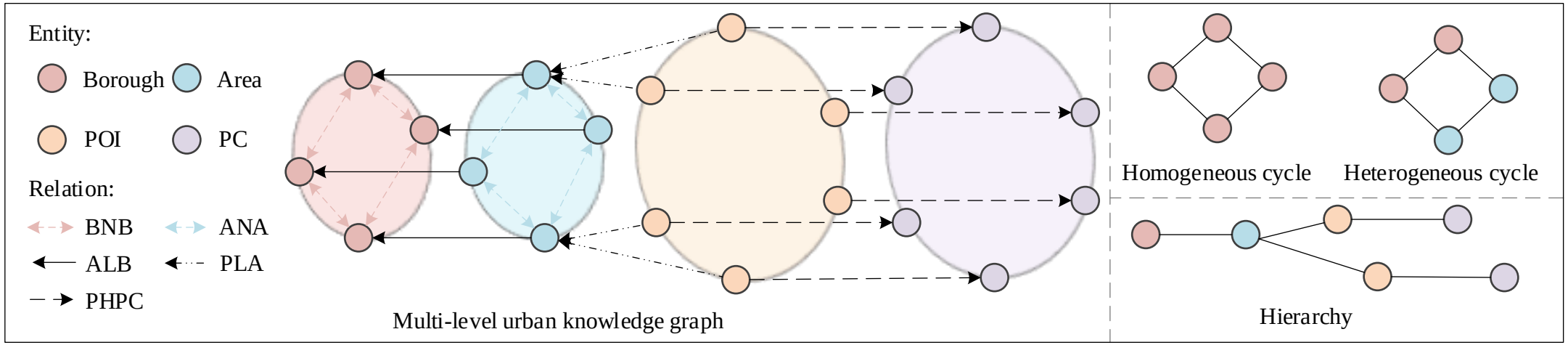
Conclusion and Future Work

Structure-aware Urban Knowledge Graph Embedding

Qualitative and Quantitative Structural Analysis

- High-order hierarchy: modeling such hierarchy empowers to uncover deeper semantics
- High-order cycle: geographic semantics could be induced by the topological cycles

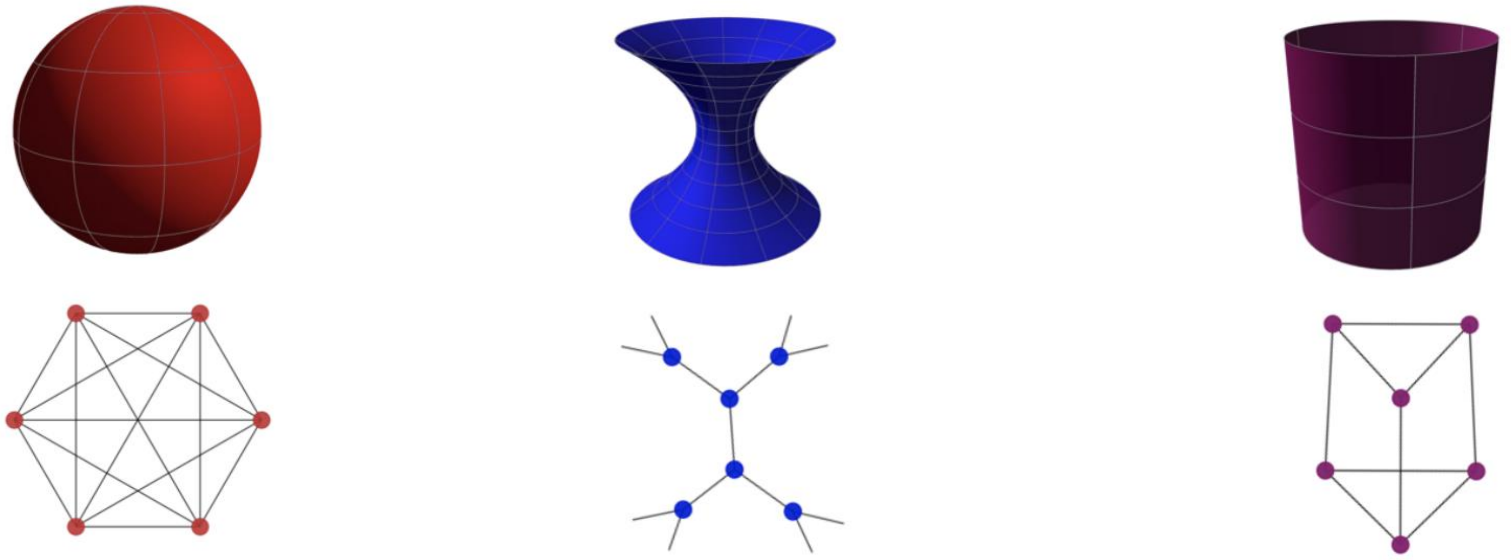
Dataset	# Entity	# Relation	# Triplet	# Train	# Valid	# Test	# Cycle	# Hyperbolicty
NYC	236,287	13	930,240	837,216	46,512	46,512	1,090,884	$\delta = 0$
CHI	140,602	13	564,400	507,960	28,220	28,220	532,108	$\delta = 0$



Structure-aware Urban Knowledge Graph Embedding

Derive Non-Euclidean Methods on UrbanKG Embedding

- Hierarchical structures are suitable to be represented in hyperbolic space
- Cyclic structures are suitable to be modeled in spherical space
- Model cycles and hierarchies in product space (i.e., the combination of hyperbolic and spherical space) simultaneously



spherical space　　hyperbolic space　　product space

[1] Giovanni, et al. Heterogeneous manifolds for curvature-aware graph embedding. ICLR 2022 Workshop on Geometric and Topological Representation Learning. 2022.

Structure-aware Urban Knowledge Graph Embedding

Experimental Setup

- Problem formulation: Learn low-dimensional embeddings of entities and relations. Evaluate UrbanKG embedding using the link prediction task, which aims to predict the missing head or tail entity for a triplet $\langle h, r, t \rangle$
- Baselines:
  - ✓ Euclidean models: TransE, DisMult, ComplEx, MuRE, TuckER, QuatE
  - ✓ Non-Euclidean models:
    - Hyperbolic space: MuRP, RotH, RefH, AttH, ConE
    - Spherical space: MuRS
    - Product space: M2GNN, GIE
- Evaluation metric:
  - ✓ MRR, the mean of the inverse of correct entity ranking
  - ✓ Hits@K, the percentage of correct entities in top-K ranked entities

Structure-aware Urban Knowledge Graph Embedding

Results

- Almost all non-Euclidean (i.e., hyperbolic and spherical space) embedding methods outperform Euclidean embedding method
- Product space obtains the dominant performance as it can capture hierarchies and cycles simultaneously

Type	Space	Model	NYC				CHI			
			MRR	Hits@10	Hits@3	Hits@1	MRR	Hits@10	Hits@3	Hits@1
Euclidean models	E	TransE	.507	.563	.528	.470	.485	.556	.501	.436
	E	DisMult	.401	.478	.433	.355	.395	.479	.469	.394
	E	MuRE	.516	.613	.545	.468	.493	.601	.536	.437
	E	TuckER	.513	.609	.541	.466	.488	.584	.525	.423
	C	RotatE	.274	.363	.309	.220	.306	.385	.336	.258
	C	ComplEx	.259	.357	.305	.195	.304	.385	.337	.253
	C	QuatE	.321	.388	.347	.282	.396	.490	.427	.345
Non-Euclidean models	S	MuRS	.528	.622	.552	.478	.501	.619	.536	.437
	H	MuRP	.545	.635	.570	.497	.519	.634	.555	.456
	H	RotH	.526	.601	.550	.472	.511	.626	.548	.447
	H	RefH	.524	.610	.549	.473	.509	.629	.542	.451
	H	ATTH	.539	.603	.556	.503	.514	.610	.542	.463
	H	ConE	.542	.629	.563	.485	.513	.617	.535	.465
	P	M2GNN	.561	.638	.578	.521	.540	.651	.571	.481
	P	GIE	.573	.665	.600	.523	.552	.660	.580	.498

OUTLINE

Motivation and Challenge

Urban Knowledge Graph Construction

Structure-aware Urban Knowledge Graph Embedding

Knowledge-enhance Urban SpatioTemporal Prediction

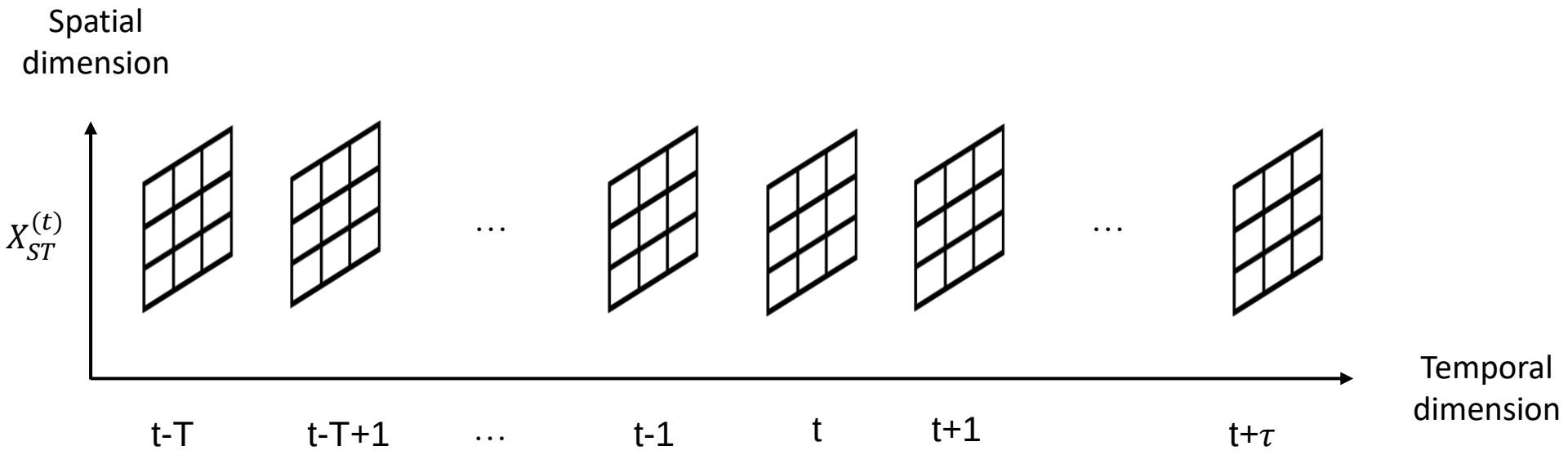
Conclusion and Future Work

Knowledge-enhance Urban SpatioTemporal Prediction

Problem Formulation

- Knowledge-enhanced urban spatiotemporal prediction (USTP) task aims to learn a multi-step prediction function f based on past observations $X_{ST}^{(t)} \in R^{N \times C}$ and the UrbanKG G:

$$\left(X_{ST}^{(t)}, X_{ST}^{(t+1)}, \ldots, X_{ST}^{(t+\tau)}\right) = f\left(\left(X_{ST}^{(t-T)}, X_{ST}^{(t-T+1)}, \ldots, X_{ST}^{(t-1)}\right), G\right)$$

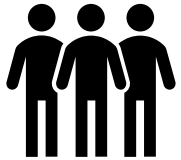


Knowledge-enhance Urban SpatioTemporal Prediction

Incorporating UrbanKG Embedding

- Five USTP tasks:
  - ✓ Taxi service: predict future area-level taxi inflow and outflow
  - ✓ Bike trip: forecast future road-level bike inflow and outflow
  - ✓ Human mobility: forecast future human inflow and outflow
  - ✓ Crime: binary classification
  - ✓ 311 service: binary classification
- Incorporate the UrbanKG by concatenating the entities' embeddings with the taxi flow features, e.g., $X_{ST}^{(t)} \in R^{N \times C_{taxi'}}$ where $C_{taxi'} = C_{taxi} \| e_{area}$, e_{area} is the embedding of area entity and $\|$ denotes the concatenation operation





Knowledge-enhance Urban SpatioTemporal Prediction

Statistics of Five USTP Datasets and Baselines

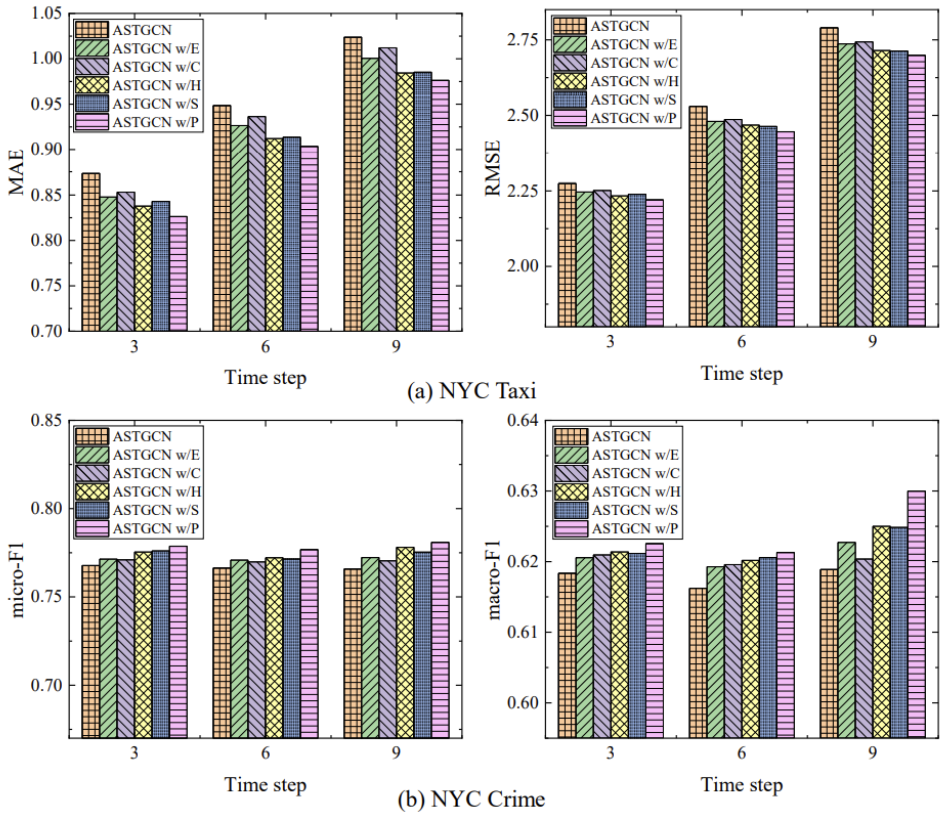
- Same dataset construction way for NYC and CHI
- Baselines:
  - ✓ Auto-Encoder, LSTM
  - ✓ STGCN, ASTGCN
  - ✓ MTGNN
  - ✓ AGCRN
  - ✓ TGCN, HGCN

Dataset	NYC					CHI				
	Taxi	Bike	Mobility	Crime	311 service	Taxi	Bike	Mobility	Crime	311 service
# of records	1,118,584	383,919	1,052,232	389,551	3,141,153	3,826,868	1,085,690	939,543	202,291	1,821,949
# of vertices	260	2,500	1,600	260	260	77	1,500	1,000	77	77
Time span	04/01/2020-06/31/2020			01/01/2021-31/12/2021		04/01/2019-06/31/2019			01/01/2021-31/12/2021	
Time interval	30 minutes			120 minutes		30 minutes			120 minutes	

Knowledge-enhance Urban SpatioTemporal Prediction

Comparison of UrbanKG Embedding Space Variants

- UrbanKG embedding could enhance USTP tasks regardless of which space they come from
- Hyperbolic space and spherical space embedding demonstrate greater performance improvement
- Embedding derived from the product space yields the most substantial benefit for both urban flow forecasting and urban event prediction



Knowledge-enhance Urban SpatioTemporal Prediction

Results

- UrbanKG embedding consistently improves the accuracy of five USTP task
- UrbanKG embedding consistently improves the accuracy of six existing USTP models

Model	NYC					CHI				
	Taxi MAE/RMSE	Bike MAE/RMSE	Mobility MAE/RMSE	Crime Micro/Macro-F1	311 service Micro/Macro-F1	Taxi MAE/RMSE	Bike MAE/RMSE	Mobility MAE/RMSE	Crime Micro/Macro-F1	311 service Micro/Macro-F1
GRU	1.802/3.587	0.920/1.115	1.003/1.318	-/-	-/-	3.372/8.895	1.431/2.619	1.157/1.800	-/-	-/-
LSTM	1.425/2.769	0.832/0.994	0.984/1.227	-/-	-/-	3.075/9.076	1.359/2.259	1.107/1.682	-/-	-/-
AE	1.399/2.484	0.701/0.898	0.956/1.233	56.42/52.12	65.68/59.64	3.313/11.64	1.416/2.706	1.148/1.721	57.49/57.27	54.67/55.34
STGCN	0.835/2.069	0.221/0.558	0.553/0.976	76.80/63.10	81.15/78.50	1.967/6.014	0.632/1.537	0.458/1.138	71.86/68.71	79.67/79.47
STGCN w/P	0.781/1.942	0.201/0.543	0.542/0.947	77.56/64.26	81.98/79.15	1.844/5.826	0.625/1.501	0.414/1.064	72.18/68.93	79.92/79.83
Improv %	6.47%/6.14%	9.05%/2.69%	1.99%/2.97%	0.99%/1.84%	1.02%/0.83%	6.25%/3.13%	1.11%/2.34%	9.61%/6.50%	0.45%/0.32%	0.31%/0.45%
MTGNN	1.289/2.275	0.967/1.041	0.963/1.163	72.88/58.61	79.32/75.75	2.167/5.965	1.153/1.723	1.081/1.504	68.48/66.46	76.72/76.32
MTGNN w/P	1.183/2.118	0.920/1.021	0.891/1.083	73.47/60.02	80.28/77.85	1.975/5.882	1.083/1.641	0.983/1.399	71.73/68.65	78.29/77.99
Improv %	8.22%/6.90%	4.86%/1.92%	7.48%/6.8%	0.81%/2.41%	1.21%/2.77%	8.86%/1.39%	6.07%/4.76%	9.07%/6.98%	4.75%/3.30%	2.05%/2.19%
AGCRN	1.315/2.391	0.958/1.038	0.945/1.148	65.98/60.62	75.68/69.64	2.403/7.277	1.187/1.768	0.213/1.281	64.14/63.35	71.18/67.07
AGCRN w/P	1.208/2.221	0.875/0.983	0.901/1.067	67.75/61.88	76.35/72.57	2.255/6.945	1.093/1.621	0.205/1.258	66.51/65.27	72.87/70.54
Improv %	8.14%/7.11%	8.66%/5.30%	4.66%/7.06%	2.68%/2.08%	0.89%/4.21%	6.16%/4.56%	7.92%/8.31%	3.76%/1.80%	3.71%/3.03%	2.37%/5.17%
ASTGCN	0.832/2.141	0.231/0.579	0.566/0.973	76.69/61.62	80.82/77.45	1.849/5.323	0.745/1.822	0.505/1.259	71.77/68.45	79.18/78.93
ASTGCN w/P	0.805/2.012	0.220/0.564	0.558/0.944	76.94/62.34	81.28/78.85	1.822/5.198	0.690/1.754	0.463/1.191	72.22/68.90	80.09/79.98
Improv %	3.25%/6.03%	4.76%/2.59%	1.41%/2.98%	0.33%/1.17%	0.57%/1.81%	1.46%/2.35%	7.38%/3.73%	8.32%/5.40%	0.63%/0.66%	1.15%/1.33%
TGCN	1.701/3.198	0.337/0.685	0.654/1.118	75.22/63.19	77.56/72.96	2.112/6.645	1.127/2.583	0.702/1.665	70.62/66.41	77.89/77.57
TGCN w/P	1.578/2.887	0.319/0.664	0.635/1.054	76.01/63.65	79.29/73.47	1.997/6.502	1.052/2.341	0.677/1.586	71.26/68.38	78.49/78.25
Improv %	7.23%/9.72%	5.34%/3.07%	2.91%/5.72%	1.05%/0.73%	2.23%/0.70%	5.45%/2.15%	6.65%/9.37%	3.56%/4.74%	0.91%/2.97%	0.77%/0.88%
HGCN	1.337/2.285	0.951/1.138	0.971/1.200	76.04/62.26	75.68/69.64	2.765/7.849	1.159/1.794	0.661/1.277	67.57/62.31	74.67/73.53
HGCN w/P	1.282/2.124	0.879/1.036	0.921/1.104	76.70/63.25	77.41/72.76	2.609/7.781	1.092/1.682	0.637/1.146	69.17/65.7	75.87/75.32
Improv %	4.11%/7.05%	7.57%/8.96%	5.15%/8.00%	0.87%/1.59%	2.29%/4.48%	5.64%/0.87%	5.78%/6.24%	3.63%/10.2%	2.37%/5.44%	1.61%/2.43%

OUTLINE

Conclusion and Future Work

Conclusion

- To the best of our knowledge, UUKG is the first open-source urban knowledge graph dataset compatible with various aligned USTP tasks
- Urban high-order structure modeling is important for urban knowledge representation and structure-aware KG embedding methods could capture them effectively

Limitation

- All experiments were conducted in two US metropolises and we only consider concatenation operation for embedding fusion

Future Work

- Derive extra multi-modal data (e.g., images, reviews) to enrich the UrbanKG and introduce more USTP tasks (e.g., trajectory prediction and site selection)

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Alibaba

Our Github repository will be continuously updated!
https://github.com/usail-hkust/UUKG

Thank You!
Q & A

Contact us if you have further questions.

✉ yning092@connect.hkust-gz.edu.cn, liuh@ust.hk