

KNOWLEDGE GRAPHS AND SPATIOTEMPORAL DATA

SPATIAL DATA SCIENCE SYMPOSIUM, DEC. 2019

Gengchen Mai Ling Cai

STKO Lab, University of California Santa Barbara



KNOWLEDGE GRAPHS

A **knowledge graph (KG)** is a data repository that stores real-world knowledge under some schema, e.g., an ontology.

■ Directed multi-graphs

- Nodes: entities
- Edges: relationships between entities with relation types as labels
- Statements: (subject→predicate→object)

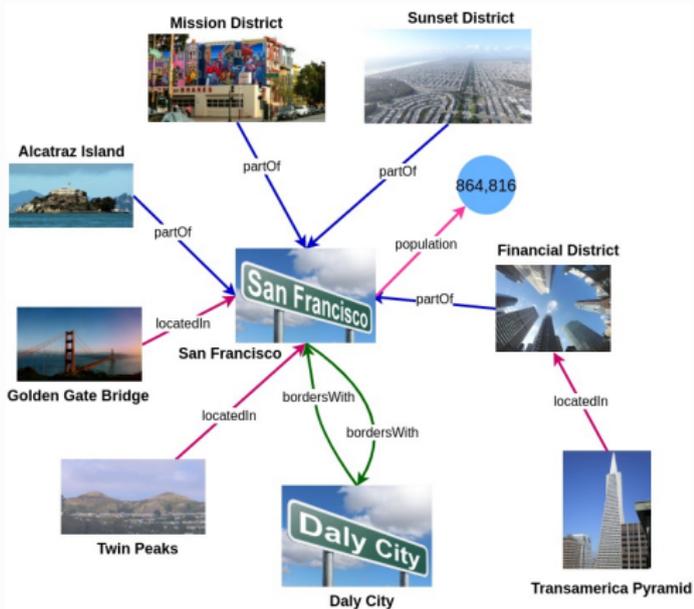


FIGURE 1: An Example of a KG

KNOWLEDGE GRAPHS

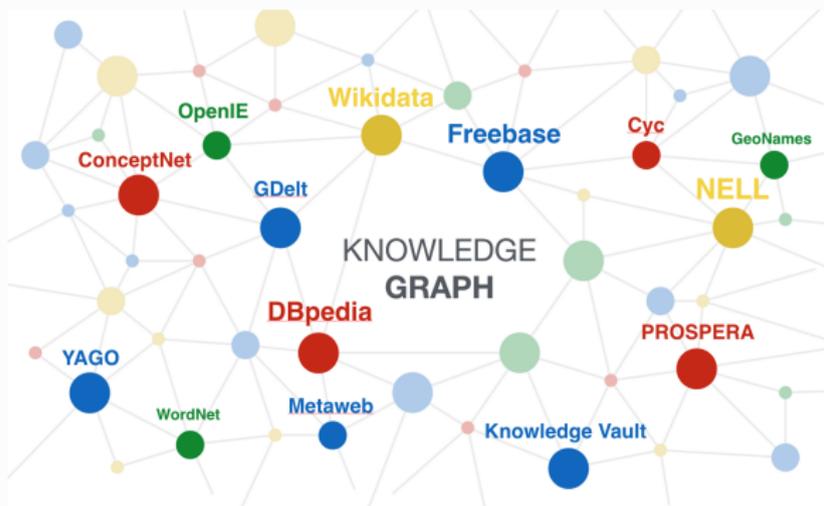


FIGURE 2: Figure From <https://medium.com/@sderymail/challenges-of-knowledge-graph-part-1-d9ffe9e35214>

- Knowledge graphs can be linked based on alignment techniques.
 - (dbr:Place, owl:**equivalentClass**, schma-org:Place)
 - (dbr:Santa_Barbara,_California, owl:**sameAs**, freebase:Santa_Barbara,_California)

APPLICATIONS OF KNOWLEDGE GRAPHS

■ Cross-domain Research

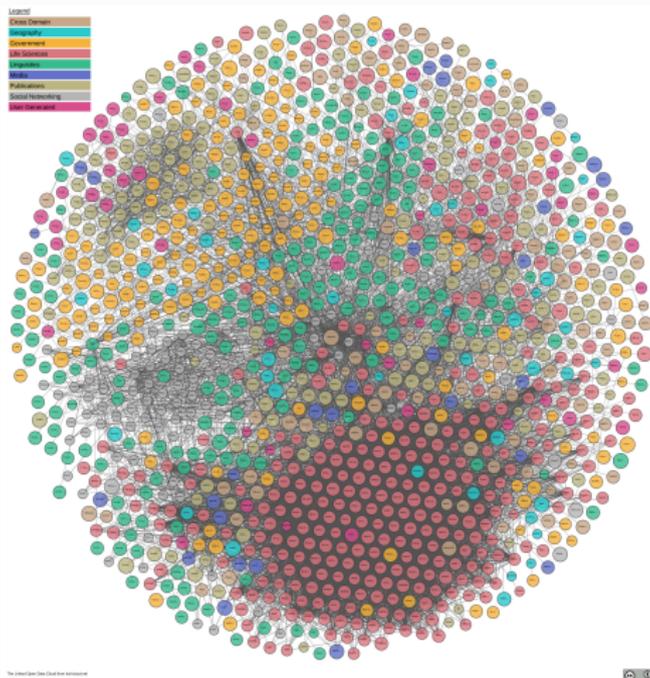


FIGURE 3: Linked Open Cloud

APPLICATIONS OF KNOWLEDGE GRAPHS

- Question Answering Systems, e.g., Apple Siri, Bing Search.



FIGURE 4: Siri

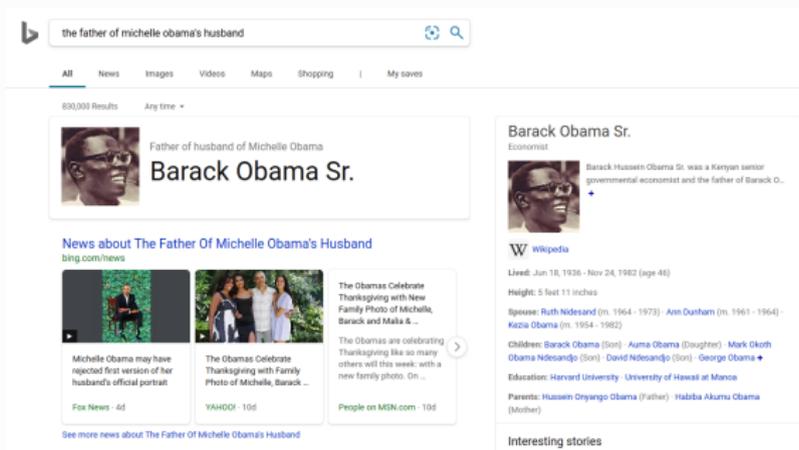


FIGURE 5: Bing search

APPLICATIONS OF KNOWLEDGE GRAPHS

■ Information Retrieval, e.g. Google Knowledge Graph

The image shows a Google search results page for the query "santa barbara". The search bar at the top shows the query and a magnifying glass icon. Below the search bar, there are navigation tabs for "All", "Flights", "Maps", "News", "Images", "More", "Settings", and "Tools". The main content area displays search results, including a "Santa Barbara, CA | Hotels, Restaurants, Events & Activities" card with a URL and a brief description. Below this is a "Visit Santa Barbara (@SantaBarbara) - Twitter" card with a URL and a tweet about Spanish food and romance. To the right, there is a "Santa Barbara, California - Wikipedia" card with a URL and a brief description. Below the Wikipedia card is a "Things to do in Santa Barbara" section with four cards: "Mission Santa Barbara", "Santa Barbara Zoo", "Santa Barbara Presidio", and "Santa Barbara County Courthouse". To the right of the main content area is a "Santa Barbara" knowledge panel with a map, a description of the city, weather information, and a "Plan a trip" section.

santa barbara

[All](#)
[Flights](#)
[Maps](#)
[News](#)
[Images](#)
[More](#)
[Settings](#)
[Tools](#)

About 1,250,000 results (0.60 seconds)

Santa Barbara, CA | Hotels, Restaurants, Events & Activities
<https://santabarbaraca.com/> •
 Welcome to Santa Barbara—The American Riviera! Plan your trip, find restaurants, things to do, wine tasting, shopping, outdoor activities and more.
[Santa Barbara Beauty in the ...](#) [Santa Barbara Bowl](#) [Santa Barbara County ...](#)

Visit Santa Barbara (@SantaBarbara) - Twitter
<https://twitter.com/SantaBarbara>

Delicious food, great cocktails and stunning views of the Steamers Wharf. What more could you ask for at [Blue Water Grill?](#)
[@AmericanRiviera](#) [@BogaloffBeauty](#)
<pic.twitter.com/MQFm6k...>
 2 days ago · [Twitter](#)

Modern chic meets Spanish romance. Introducing the new [@HiltonSBResort](#). [#AmericanRiviera](#)
<bit.ly/2K4a4Wx>
<pic.twitter.com/9x02ZxH...>
 2 days ago · [Twitter](#)

Winery-hopping around the Santa Barbara area just got a little easier with [UberWINE](#). [#TravelLettuce](#)
<https://t.me/20q4v1w>
[@TravelLettuce](#)
 2 days ago · [Twitter](#)

Santa Barbara, California - Wikipedia
https://en.wikipedia.org/wiki/Santa_Barbara,_California •
 Santa Barbara (Spanish for "Saint Barbara") is the county seat of Santa Barbara County in the U.S. state of California. Situated on a south-facing section of History of Santa Barbara - Santa Barbara County, California

Things to do in Santa Barbara

Mission Santa Barbara
 18th-century mission & Catholic church

Santa Barbara Zoo
 Compact zoo with hundreds of animals

Santa Barbara Presidio
 Green space featuring a 1782 fortress

Santa Barbara County Courthouse
 Spanish Colonial Revival courthouse

[Santa Barbara travel guide](#)

Santa Barbara
 City in California

Santa Barbara is a city on the central California coast, with the Santa Ynez Mountains as dramatic backdrop. Downtown, Mediterranean-style white stucco buildings with red-tile roofs reflect the city's Spanish colonial heritage. Upscale boutiques and restaurants offering local wines and occasional fare line State Street. On a nearby hill, Mission Santa Barbara, founded in 1786, houses Franciscan friars and a museum.

Weather: 62°F (17°C), Wind S at 6 mph (10 km/h), 67% Humidity
Population: 91,930 (2016)

Plan a trip

Santa Barbara travel guide

3-star hotel averaging \$230, 5-star averaging \$492

Upcoming Events

Colleges and Universities: University of California, Santa Barbara, MORE

People also search for

[California](#) [United States of America](#) [Los Angeles](#) [San Diego](#) [San Francisco](#)

[More about Santa Barbara](#)

FIGURE 6: Google Knowledge Graph

SPATIOTEMPORAL DATA IN KNOWLEDGE GRAPHS

Geographic Information

■ Geographic Information of Entities

■ Coordinate information

- (Santa Barbara -> coordinateLocation -> **(34°25'33"N, 119°42'51"W)**);

■ Topological relations

- (Santa Barbara -> **partOf** -> California) ;

■ Other Geospatial-Related Statements

- (France -> memberOf -> **European Union**);

- (Washington, D.C. -> hasPopulation -> 672,228);

- (Los Angeles -> twinnedAdministrativeBody -> **Berlin**);

SPATIOTEMPORAL DATA IN KNOWLEDGE GRAPHS

Temporal Information

■ Temporal Scope of a Statement

- (Poland → memberOf → Warsaw Pact, [1955, 1991]);
- (Washington, D.C. → hasPopulation → 672,228, 2015); ...

■ Time as Literals

- (Barack Obama → dateOfBirth → 4 August 1961);
- (Santa Barbara → inception → 1847); ...

■ Transaction Time

- (Fernando Torres → playFor → Chelsea, [2011,2015], [09/02/2017])

WHY DO SPATIOTEMPORAL DATA MATTER?

■ Examples:

- **Geographic question:** Find the cities in California which the longest river in California flowed through?
 - Find the longest river in California.
 - Spatial operations are imposed over the river and all the cities in California.
- **Temporal query:** (?Person) (?Person → workLocation → New York City) ∧ (?Person → positionHeld → President of the United States)
 - Find candidates that satisfy both statements.
 - Check the temporal scoping of the two statements.
- ...

KNOWLEDGE GRAPH EMBEDDINGS

- **Basic idea:** encode entities and relations as latent low-dimensional vectors, where each dimension represents one latent feature.
 - Take TransE as an example:
 - Given a statement (Santa Barbara \rightarrow partOf \rightarrow California), $|\mathbf{Santa\ Barbara} + \mathbf{partOf} - \mathbf{California}| = 0$

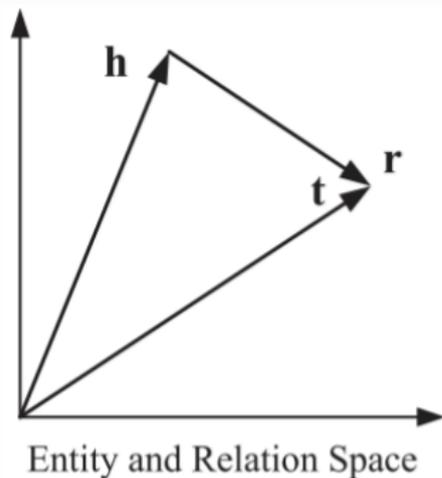
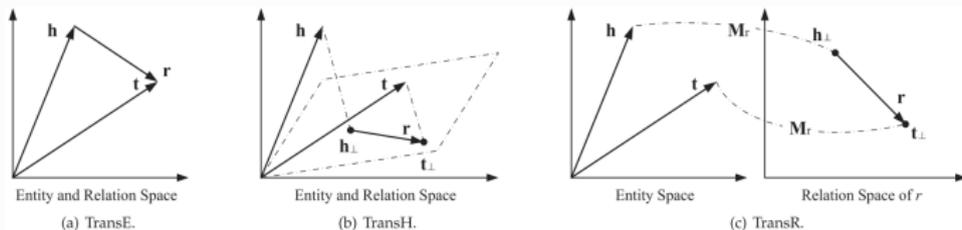


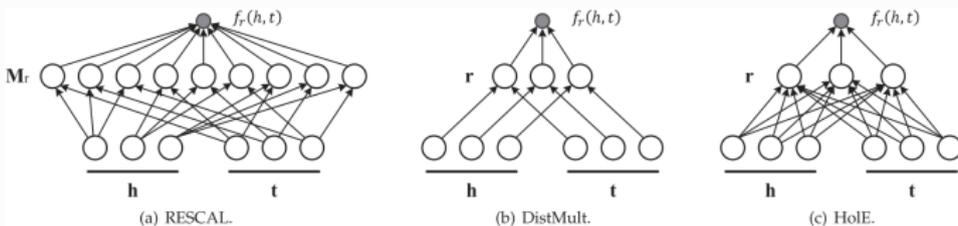
FIGURE 7: Knowledge Graph Embedding- TransE

KNOWLEDGE GRAPH EMBEDDINGS

Translation-based models (e.g. TransE, TransH, and TransR)



Semantic matching models (e.g. RESCAL, DisMult, and HolE)



SPATIALLY EXPLICIT MODEL

- **Spatially Explicit Model** (Goodchild et al., 2004): A model is said to be spatially explicit when it differentiates behaviors and predictions according to spatial location
- What makes a model spatially explicit? (Goodchild et al., 2001)
 - **The invariance test**: the results are **not invariant** under **relocation** of the studied phenomena
 - **The representation test**: contain **spatial representations** of the studied phenomena in their implementations (e.g., coordinates, spatial relations, place names, and so on)
 - **The formulation test**: use **spatial concepts** in their formulations, e.g. the notion of a neighborhood
 - **The outcome test**: the spatial structures/forms of inputs and outcomes of the model differ

SPATIALLY EXPLICIT MACHINE LEARNING MODEL

- **Spatially Explicit Machine Learning Model**: Improve the performance of current state-of-the-art machine learning models by using **spatial thinking and principles** such as:
 - **spatial variability**
 - **distance decay effect**
 - **map projection**
- Examples:
 - Geographic Question Answering
 - Geographic Knowledge Graph Summarization
 - Location Encoding

GEOGRAPHIC QUESTION ANSWERING

- Due to **missing information** and **logical inconsistency**, it is likely to receive **no answer** for questions given a knowledge graph.
- This challenge is commonly handled by **query relaxation/rewriting** based on **knowledge graph embedding**.
- Examples:
 - What is the weather like in **Montecito**? (**missing information**)
 - After **rewriting**: What is the weather like in **Santa Barbara**?
 - Which city spans Texas and Colorado? (**logical inconsistency**)
 - After **relaxation**: Which city locates in Texas?
- The relaxation of geo-queries should consider **spatial proximity** and **place hierarchy**.

WORKFLOW

Question



Query



Query relaxation / rewriting



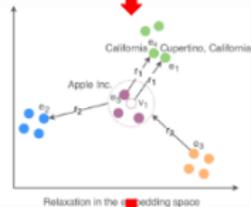
Relaxed query



Relaxed Question + Answer

Q: In which computer hardware company located in Cupertino is/was Steve Jobs a board member?

```
SELECT ?v
WHERE {
  ?v dbo:locationCity dbr:Cupertino, _California .
  ?v dbo:industry dbr:Computer_hardware .
  dbr:Steve_Jobs dbo:board ?v .
}
```

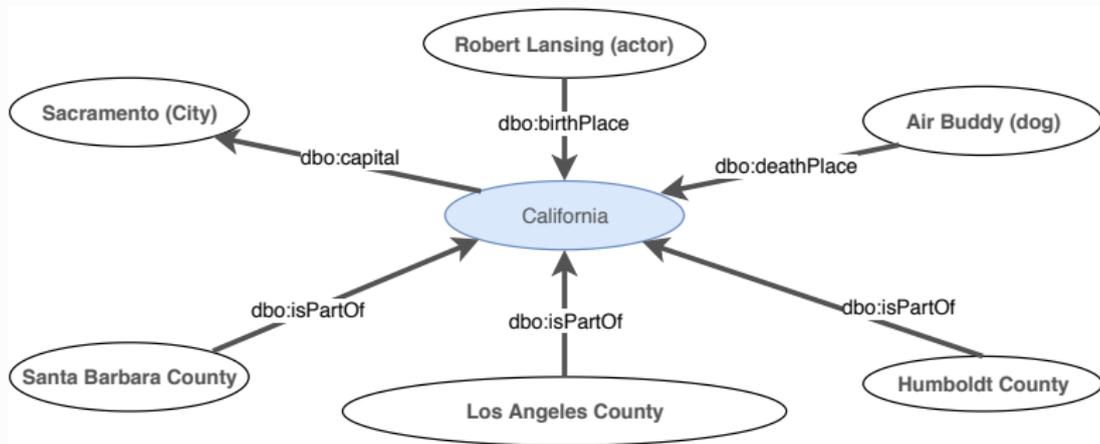


```
SELECT ?v
WHERE {
  ?v dbo:locationCity dbr:California .
  ?v dbo:industry dbr:Computer_hardware .
  dbr:Steve_Jobs dbo:board ?v .
}
```

Q: In which computer hardware company located in California is/was Steve Jobs a board member?
A: Apple Inc.

SPATIALLY EXPLICIT KNOWLEDGE GRAPH EMBEDDING

- **TransGeo**: to assign **larger weights** to geographical triples in an entity context, and these weights are modeled using a **distance decay function**



EVALUATION

- **Link prediction:** Given h, r , to predict the correct t
- **Answer prediction by relaxation/rewriting:** The rank of the correct answer in the queried answer ranking list

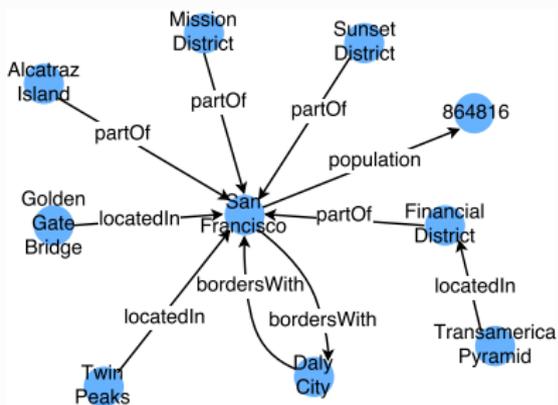
TABLE 1: Two evaluation tasks for different KG embedding models

	Link Prediction				Query Relaxation	
	MRR		HIT@10		MRR	HIT@10
	Raw	Filter	Raw	Filter		
<i>TransE</i> Model	0.122	0.149	30.00%	34.00%	0.008	5% (1 out of 20)
Wang et al. (2018)	0.113	0.154	27.20%	30.50%	0.000	0% (0 out of 20)
<i>TransGeo</i> _{regular}	0.094	0.129	28.50%	33.40%	0.098	25% (5 out of 20)
<i>TransGeo</i> _{unweighted}	0.108	0.152	30.80%	37.80%	0.043	15% (3 out of 20)
<i>TransGeo</i>	0.104	0.159	32.40%	42.10%	0.109	30% (6 out of 20)

GEO KNOWLEDGE GRAPH SUMMARIZATION

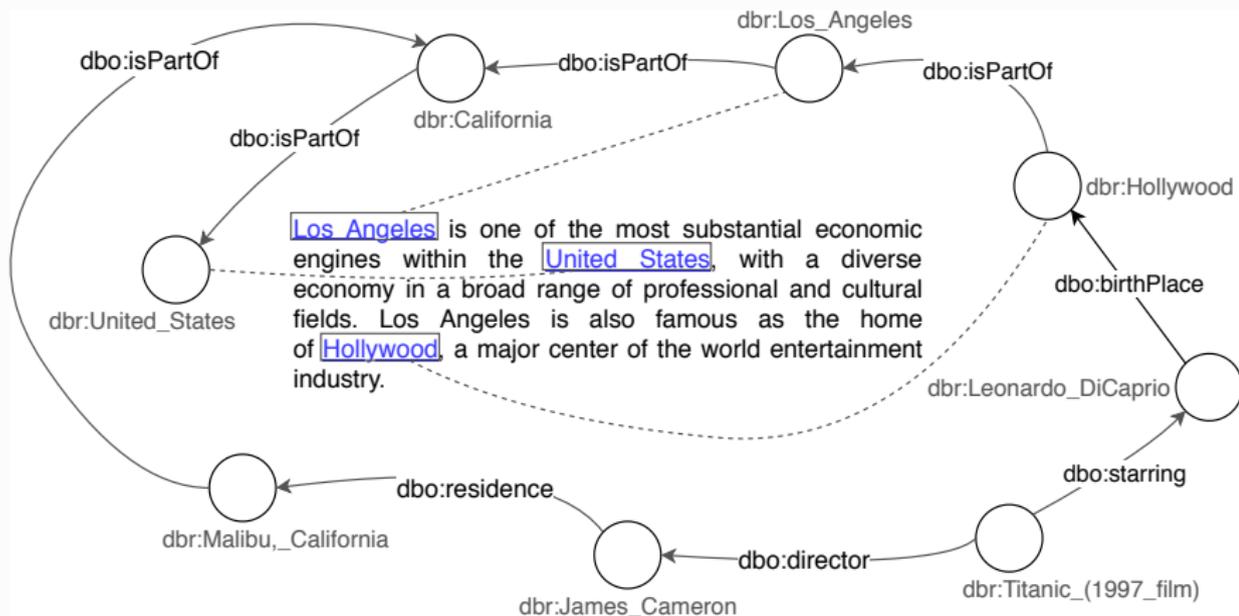
■ Summarization

- Identify the underlying structure and meaning of the original Geographic KG using a digest graph

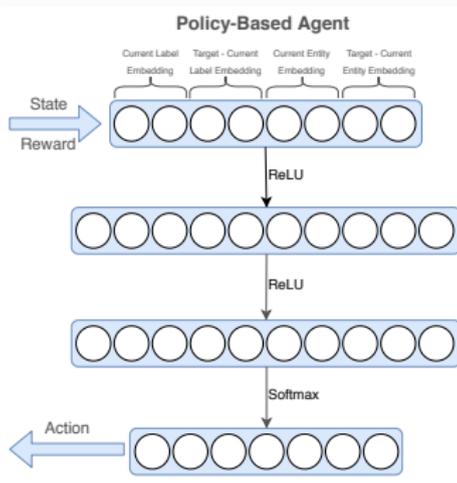
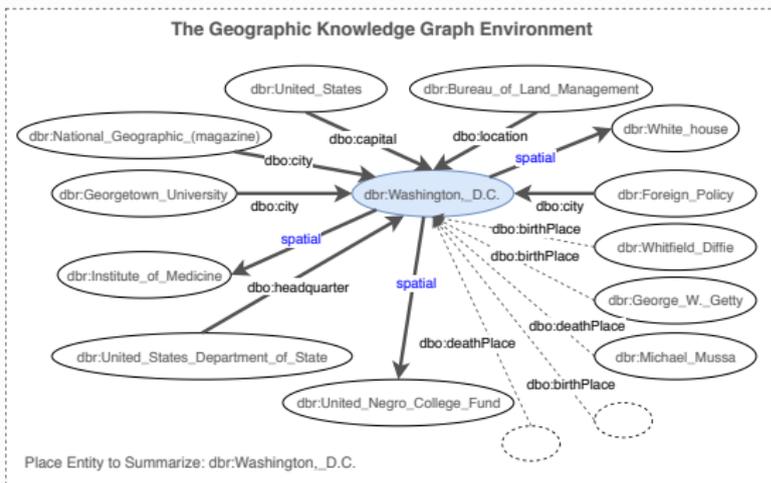


- **Question:** How can we leverage both **top-down** knowledge (e.g., considering **spatial component explicitly**) and **bottom-up** approaches (e.g., **machine learning**) to help summarize geo KGs by taking into account the balance between **commonality** and **variability**?

SUMMARIZATION EXAMPLE



REINFORCEMENT LEARNING FRAMEWORK

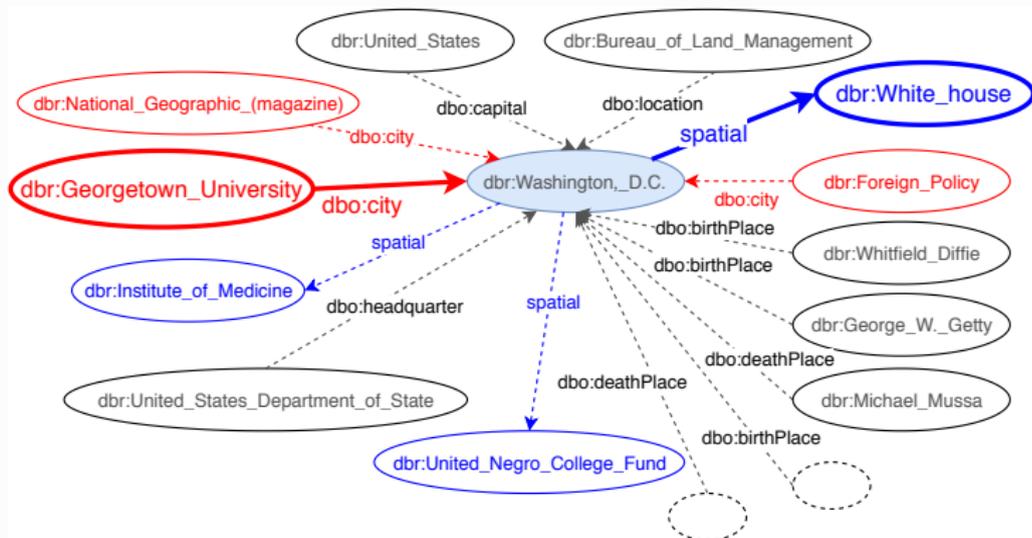


- The process starts with only **one node**
- The **agent** analyzes the **original graph structure** and the **Wikipedia summary**
- The agent iteratively adds **new relations and nodes** to the graph until the graph conveys information comparable to the Wikipedia summary

MARKOV DECISION PROCESS

■ Actions

■ 534 relations + 1 special **spatial** relation



RESULT

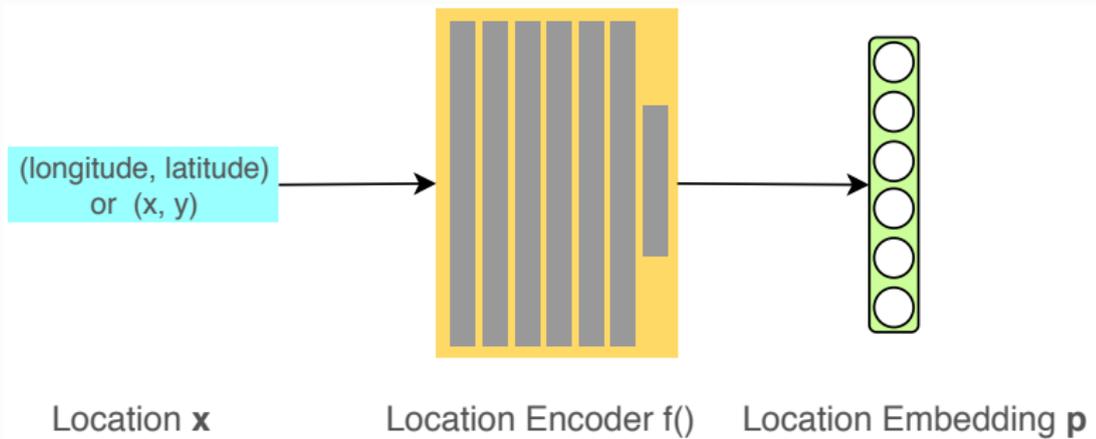
- RL-based models improve the cosine similarity (the summary graph is comparable to the Wikipedia abstract)

	RL (nonspatial-normal)	RL (spatial-normal)	RL (nonspatial-maxmin)	RL (spatial-maxmin)	RL (spatial-maxmin-pr)
Entity Embedding	0.0307	0.0496	0.0523	0.0732	0.0760
Word Embedding	0.1659	0.2527	0.2444	0.3025	0.3159

- **The spatially explicit model can perform twice as good as non-spatial models**

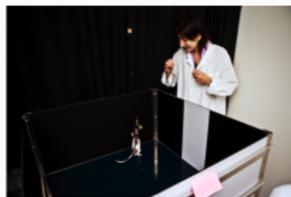
LOCATION ENCODING

- More direct approach?
- A **general-purpose representation model for space** is particular useful to design **spatially explicit models** for multiple tasks
- Advantage:
 - Preserve **spatial proximity** and **directions**
 - Easy to **generalize to unseen locations**
 - Avoid **explicit pairwise distance computation** which is unnecessarily expensive



GRID CELL

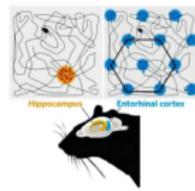
- Nobel winning Neuroscience research shows that **grid cells** in mammals provide a **multi-scale periodic representation** that functions as a metric for **coding space**.
- Grid cells are critical for integrating self-motion (path integration, or so-called dead-reckoning).



(a)



(b)



(c)



(d)

FIGURE 10: Figure from R. Gao et al., (2019)

GRID CELL

- Blair et al. (2007) show that the **multi-scale periodic representation of grid cells** can be simulated by summing three cosine grating functions oriented 60° apart.

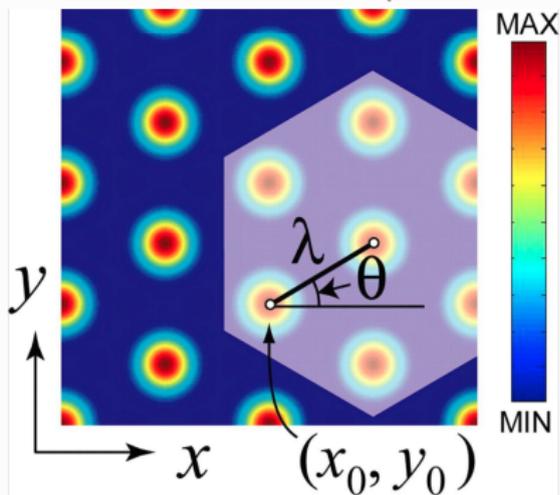


FIGURE 11: Figure from Blair et al. (2007)

- Encode locations with **multi-scale periodic representations** by using **3 sinusoidal functions**.

APPLICATIONS

- KG related tasks:
 - Geographic Question Answering
 - Geographic Knowledge Graph Summarization
- Other tasks:
 - Air Pollution Forecasting
 - **Location-Aware Image Classification**

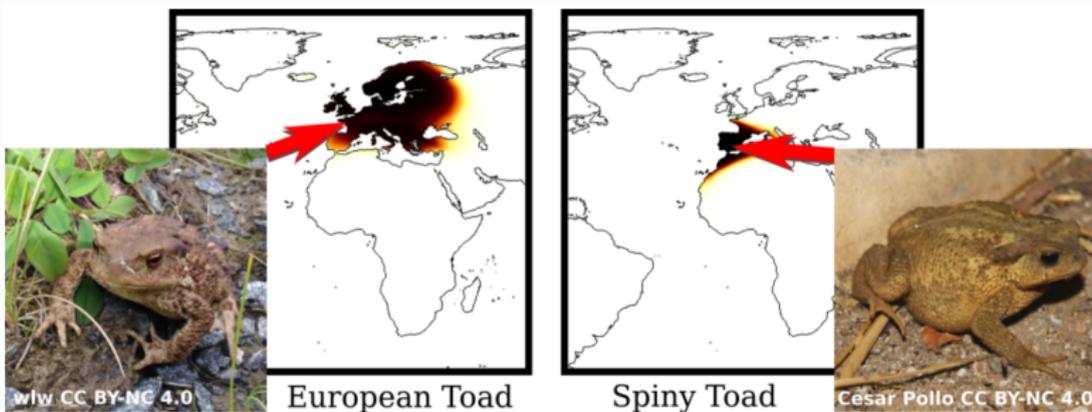


FIGURE 12: Figure from Mac Aodha et al. (2019)

LOCATION-AWARE IMAGE CLASSIFICATION

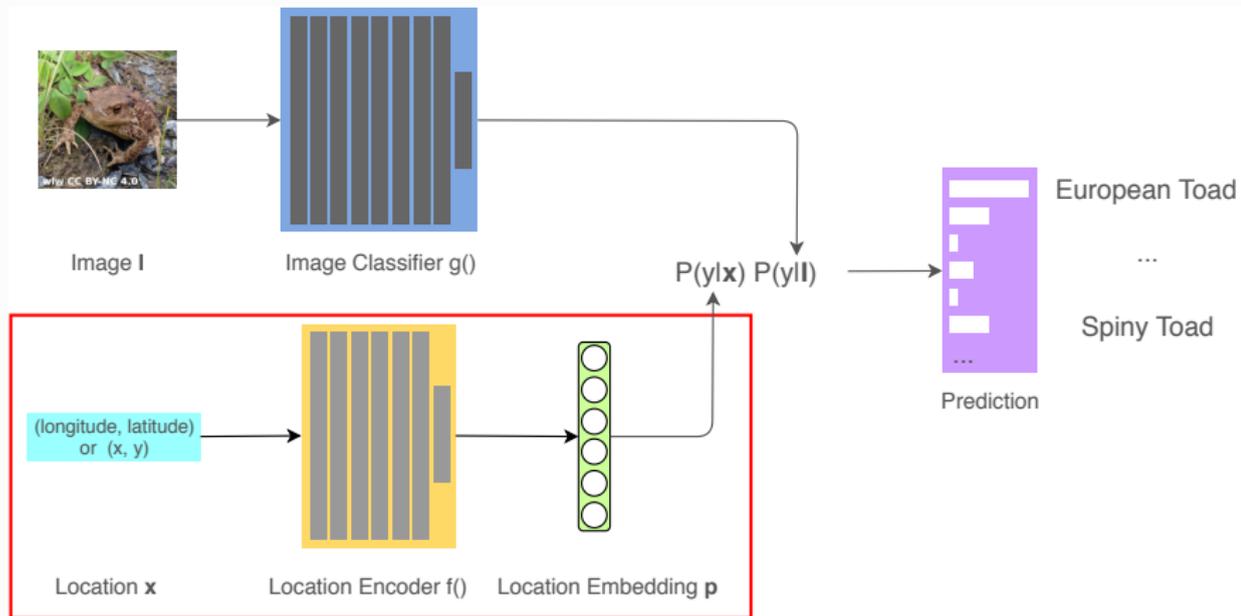


FIGURE 13: Location-Aware Image Classification

EVALUATION

	BirdSnap [†]	NABirds [†]
No Prior (i.e. uniform)	70.07	76.08
Nearest Neighbor (num)	77.76	79.99
Nearest Neighbor (spatial)	77.98	80.79
Adaptive Kernel (Berg et al., 2014)	78.65	81.11
<i>tile</i> (Tang et al., 2015) (location only)	77.19	79.58
<i>wrap</i> (Mac Aodha et al., 2019) (location only)	78.65	81.15
<i>grid</i> ($\lambda_{min}=0.0001$, $\lambda_{max}=360$, $S = 64$)	79.44	81.28
<i>theory</i> ($\lambda_{min}=0.0001$, $\lambda_{max}=360$, $S = 64$)	79.35	81.59

FIGURE 14: Evaluation Result for Location Aware Image Classification

SUMMARY

- Knowledge graphs play important roles in **data storage, data sharing, data synthesis, semantic search, cross-domain studies**, etc.
- **Spatiotemporal data** are abundant within and beyond knowledge graphs.
- **Spatially explicit models** are needed for the advancement of spatial data science.