A Spatially Explicit Reinforcement Learning Model for Geographic Knowledge Graph Summarization

Bo Yan, Krzysztof Janowicz, Gengchen Mai, and **Rui Zhu**

STKO Lab
Department of Geography
University of California, Santa Barbara

July, 2019

Geospatial Semantics

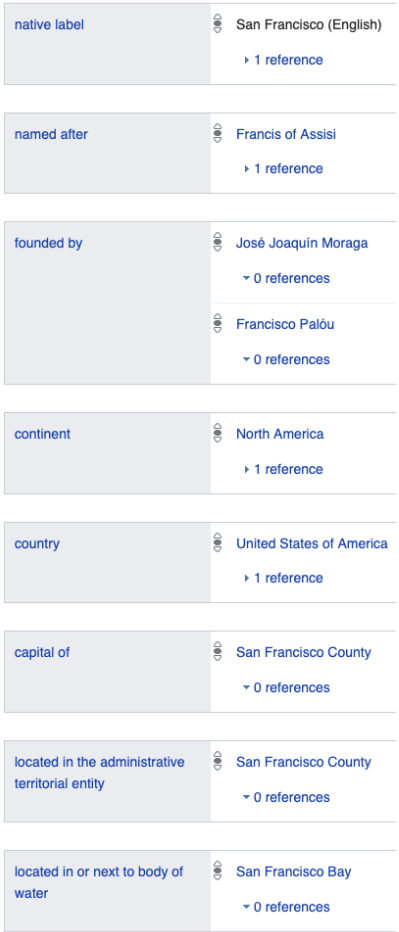
- Geospatial Semantics
  - Understanding the **meaning** of geographic concepts as well as their **cognitive** and **digital** counterparts
- Geographic Knowledge Graph
  - A nexus component in geospatial semantics to improve
    - **interoperability** (standardized and structured data)
    - **accessibility** (interface between human language and machine-readable data)
    - **conceptualization** (ontology)

Challenges in Utilizing Geographic KG

- Diversity
  - Because of the interconnected nature, KGs can cover **multiple domains** (e.g., life sciences, linguistics, media, and social networks)
  - **Heterogeneous** information networks (**multi**-relational, e.g., *isPartOf*, *birthPlace*, *headquarterOf*)
  - Objects can have **various data types**, e.g., population (numbers), multimedia (images)
- Quantity
  - Size of each KG dataset
    - **400M+ geo entities** belonging to **500 classes** in LinkedGeoData
  - Number of triples associated with each entity
    - wd:Q62 (San Francisco) has **685 outgoing** links and **22K+ incoming** links

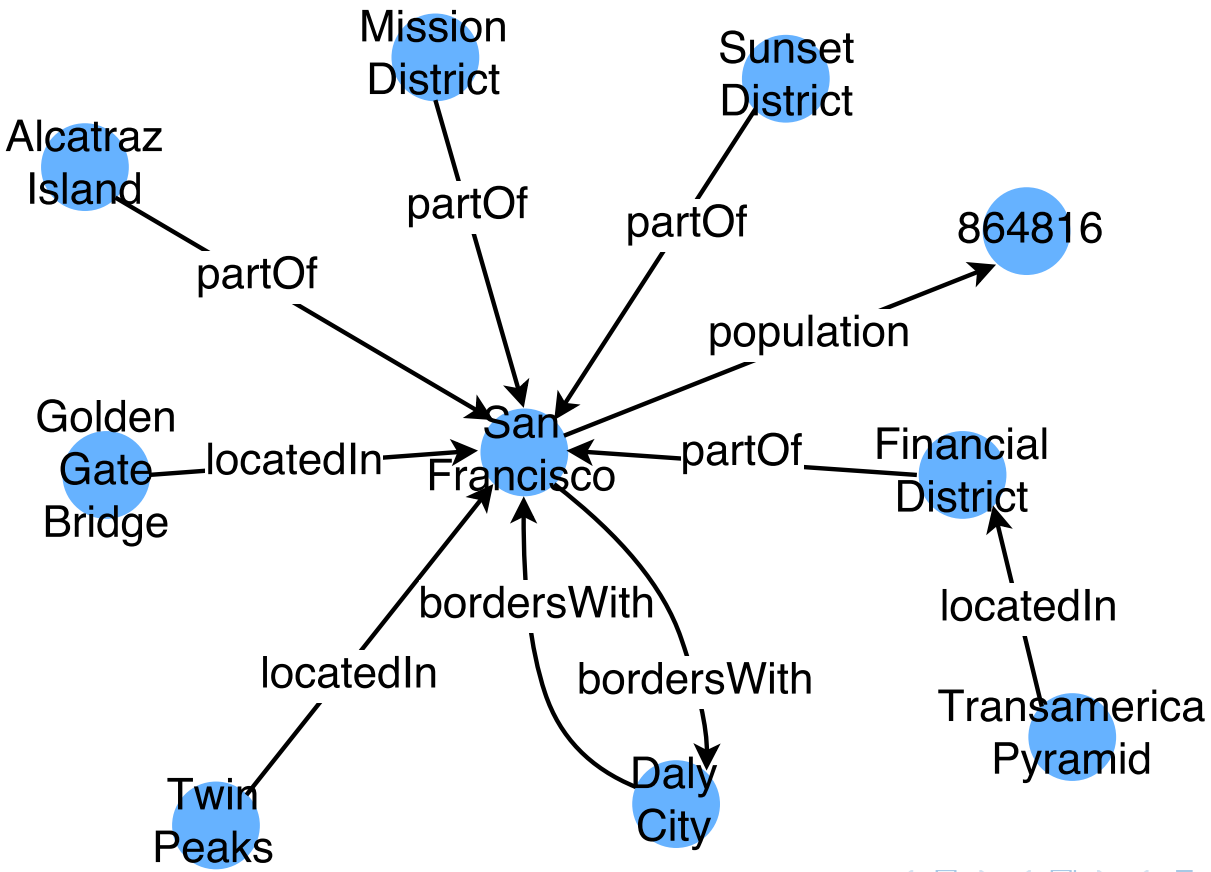
Information Overload

- Technical perspective
  - Noise
  - Interactive analysis
- Psychological perspective
  - Cognitive load



Less is More

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



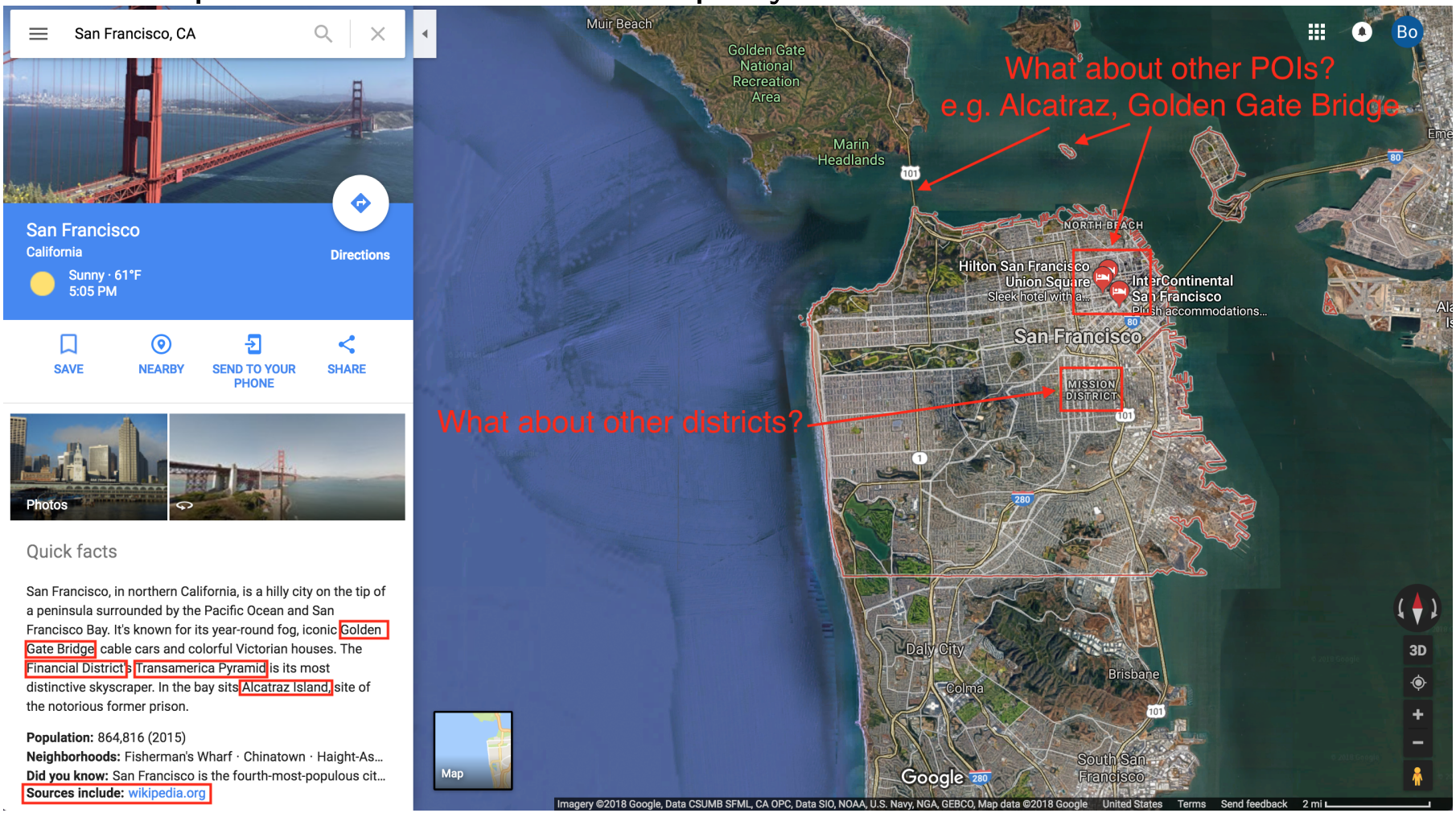
Geo KG Summarization

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**?

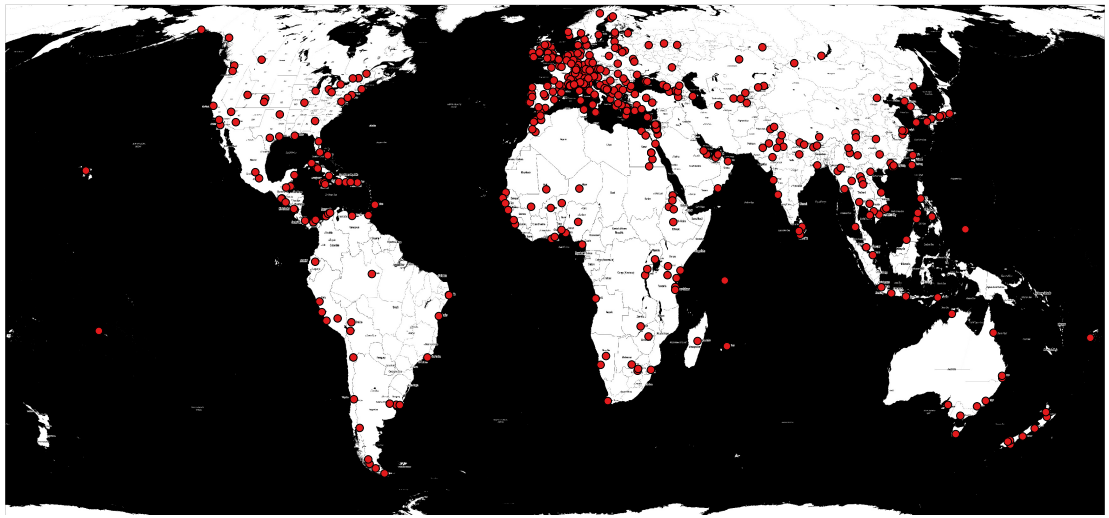
Summarization Guidance

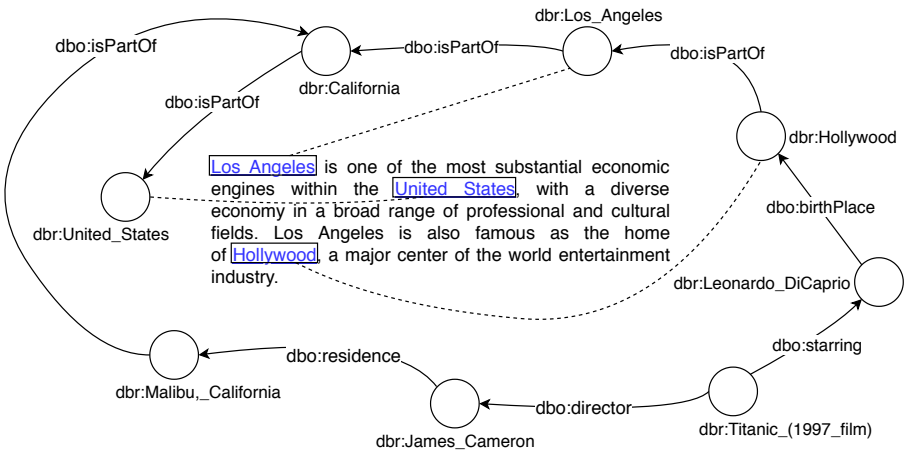
- Wikipedia abstracts are exemplary summaries



Dataset

- 369 famous places around the world
- Two parallel parts (Wikipedia summaries & DBpedia subgraphs)





Summarization as a Sequential Decision-making Process

- Intuition
  - 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

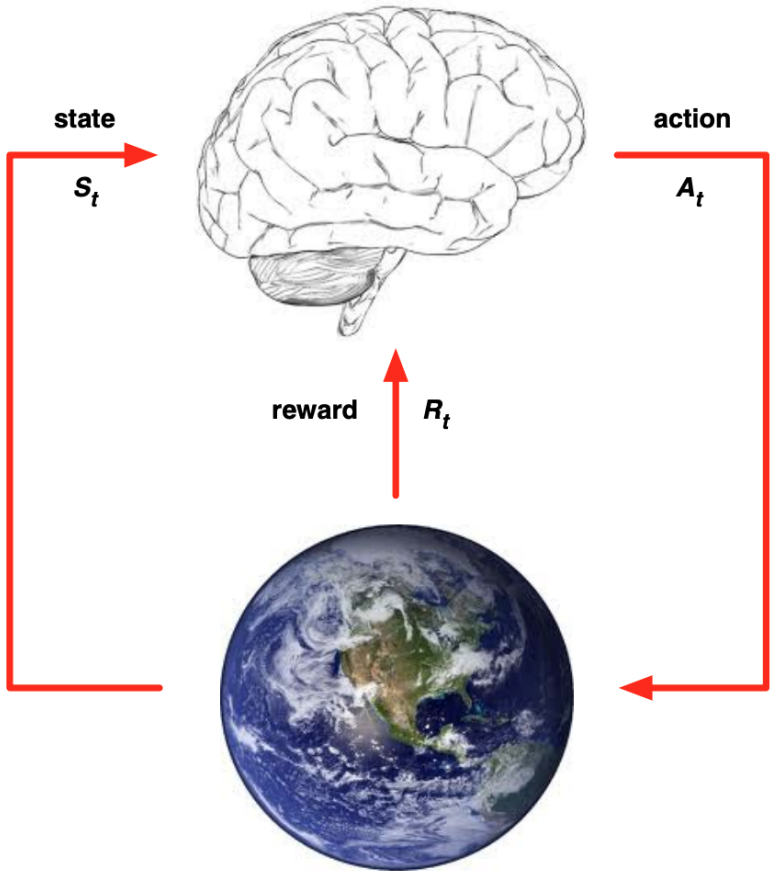
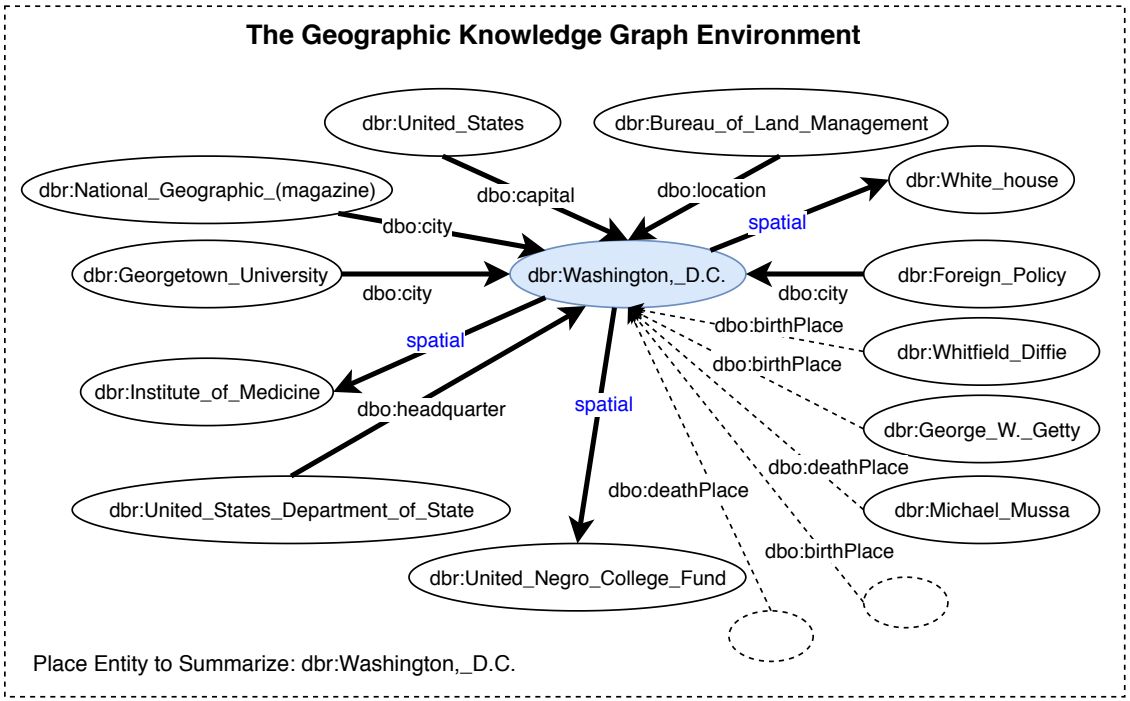
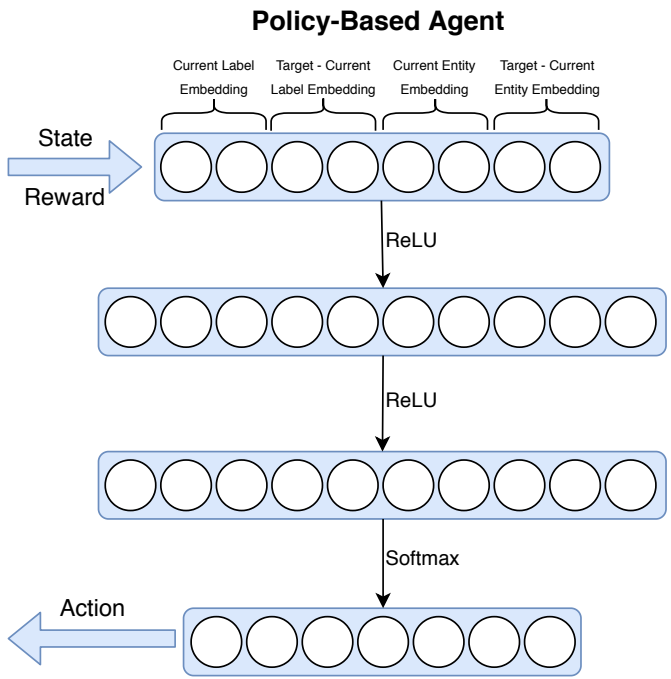


Figure from David Silver's slides

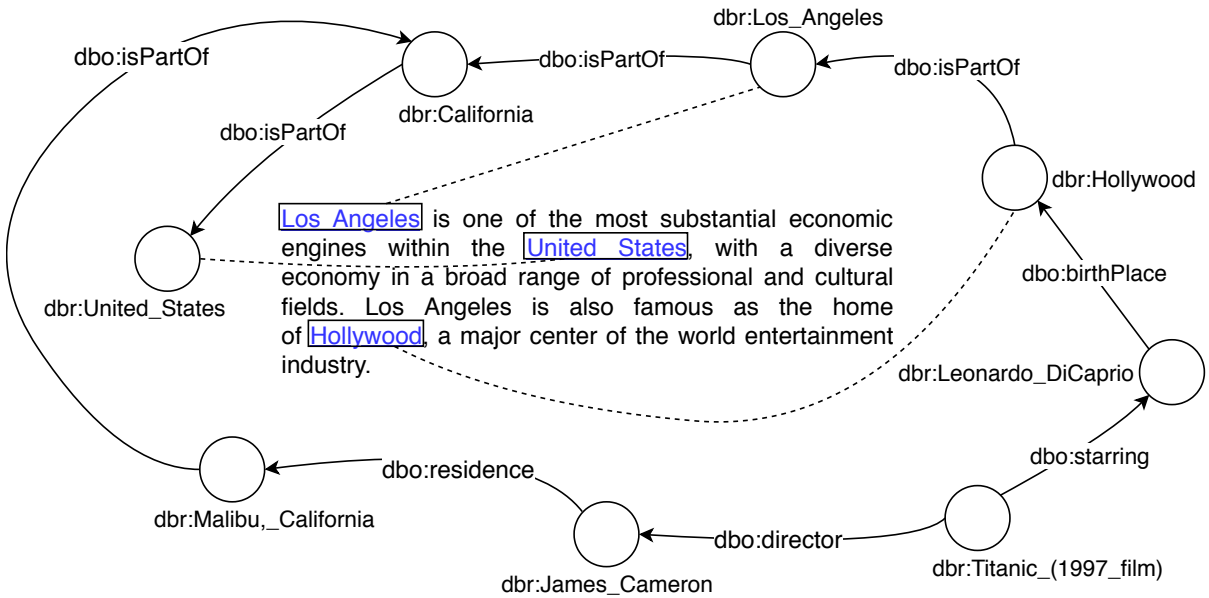
Reinforcement Learning Framework





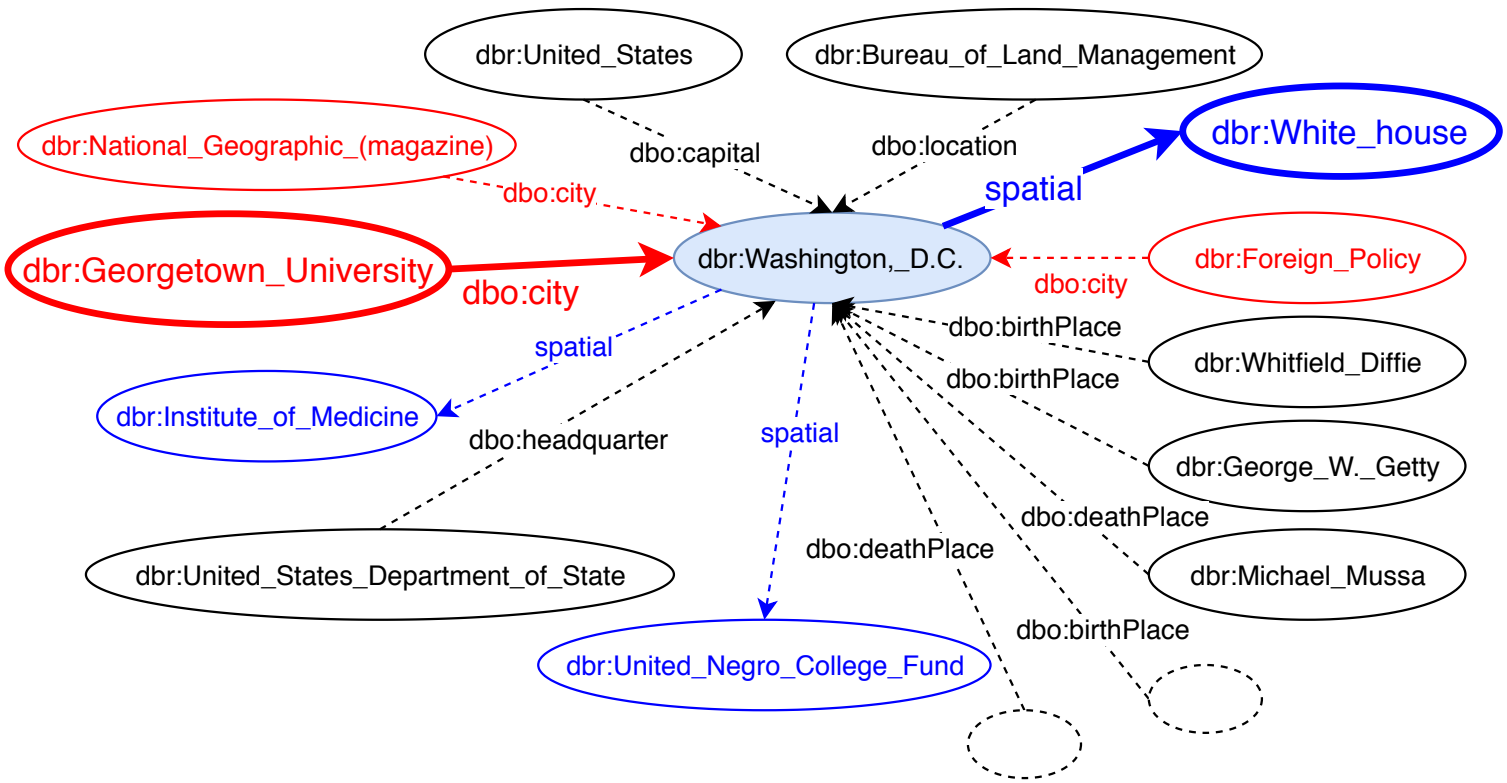
Markov Decision Process

- States
  - KG embeddings (TransE, Bordes et al. 2013)
  - Word embeddings (fastText, Bojanowski et al. 2017)



Markov Decision Process

- Actions
  - 534 relations + 1 special **spatial relation**



Markov Decision Process

- Rewards $R = r_{similarity} + r_{diversity} + r_{connection}$
  - Similarity $r_{similarity}$
    - Cosine similarity
  - Diversity $r_{divsersity}$
    - Balance between **commonality** and **variability**
  - Connection $r_{connection}$
    - Prefer nodes that are directly connected to the place to be summarized

Training Procedure

- Warm start the process using supervised policy
- REINFORCE (Monte Carlo Policy Gradient, Williams, 1992)
  - Optimize the policy $\pi_\theta(a|s)$ s.t. the total future expected reward J is maximized

$$\begin{aligned}\nabla_\theta J(\theta) &= \mathbb{E}_{s\sim \Pr(s), a\sim\pi_\theta(a|s)} Q(s,a)\nabla_\theta \log \pi_\theta(a|s) \\ &\approx \frac{1}{N}\sum_{i=0}^{N}\sum_{s,a\in eps_i} Q(s,a)\nabla_\theta \log \pi_\theta(a|s)\end{aligned} \tag{1}$$

  - Entropy-regularized (diversity)

$$H(\theta) = -\sum_{a\in A}\pi_\theta(a|s)\log \pi_\theta(a|s) \tag{2}$$

  - Loss function

$$\mathcal{L}_{REINFORCE} = -(J + \alpha H) \tag{3}$$

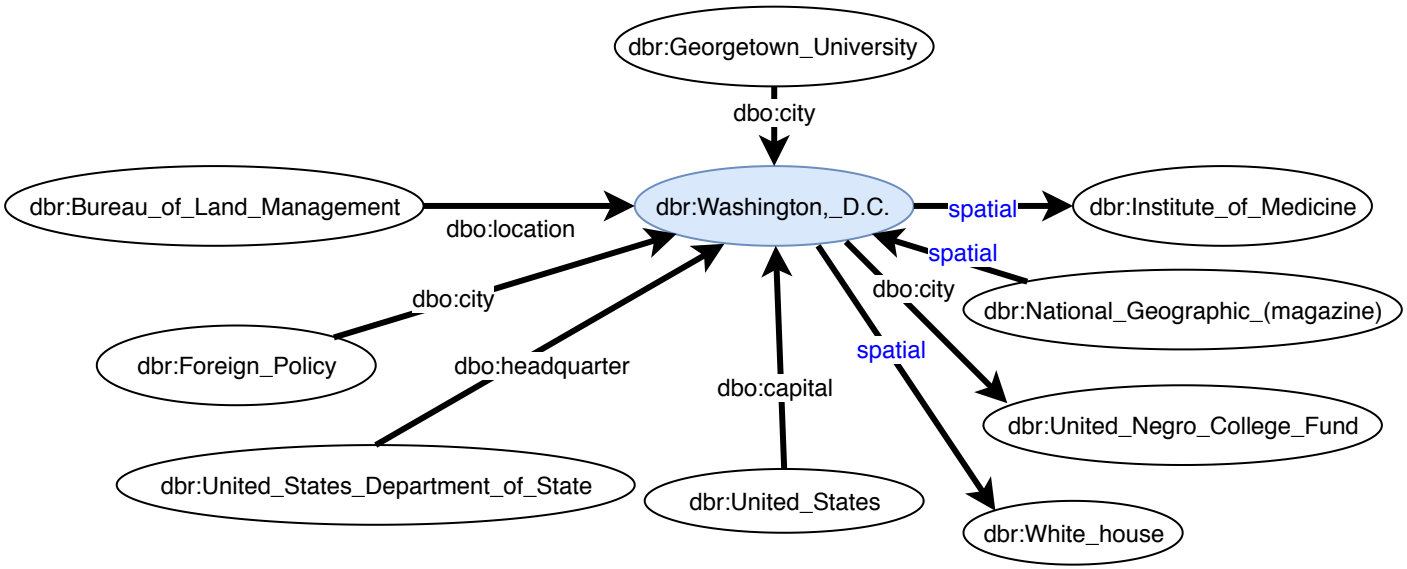
Result

- 35 places in the testing data
- Criteria: the improvement on the cosine similarity w.r.t. the baseline (the graph with the place entity itself)
- 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 Embdding	0.1659	0.2527	0.2444	0.3025	**0.3159**

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

Examples


dbr:Georgetown_University
dbo:city
dbr:Bureau_of_Land_Management
dbo:location
dbr:Washington,_D.C.
spatial
dbr:Institute_of_Medicine
spatial
dbo:city
dbr:National_Geographic_(magazine)
dbo:city
dbr:Foreign_Policy
dbo:headquarter
dbo:capital
spatial
dbr:United_Negro_College_Fund
dbr:United_States_Department_of_State
dbr:United_States
dbr:White_house

Conclusions

- Wikipedia summaries provide **guidance** to the **summarization** process
- **RL agent** can learn to summarize geo KG
- **Spatially explicit models** (the special spatial relation) yield better results