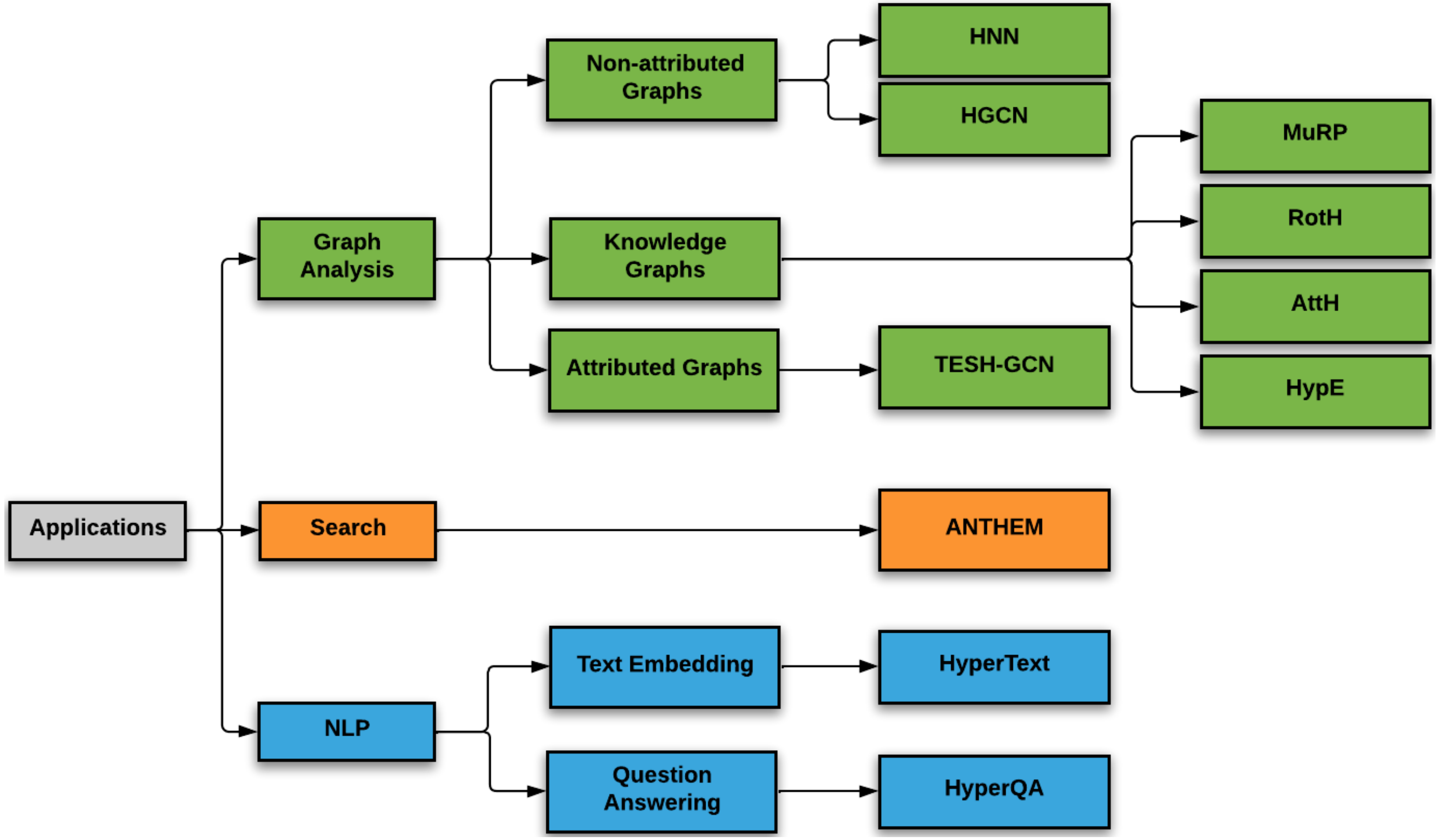
Part 4: Applications

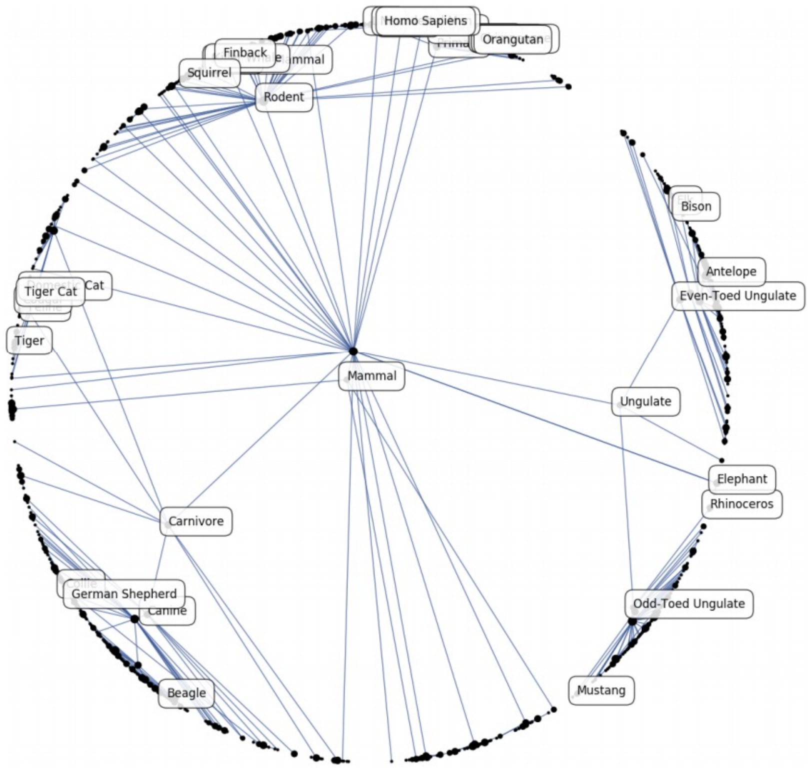


Overview of Applications

Graphs, Knowledge Graphs, Search, and NLP

Advantages of Hyperbolic Networks is primarily observed in cases when hierarchical relations exist in the underlying datasets.

1. Graph Analysis



Mammals in WordNet*

* Nickel, Maximillian, and Douwe Kiela. "Poincaré embeddings for learning hierarchical representations." *Advances in neural information processing systems* 30 (2017).

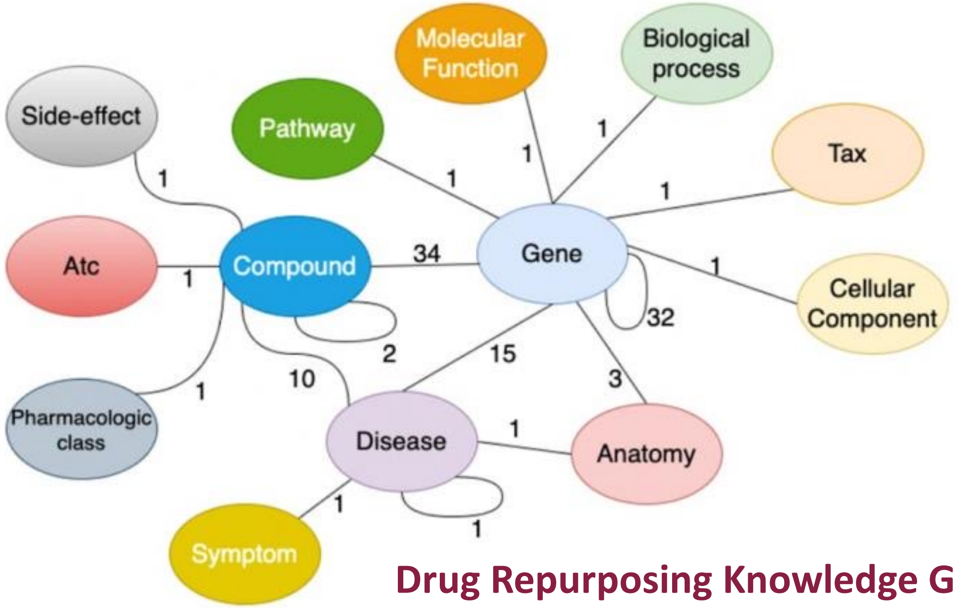
Overview of Applications

Graphs, Knowledge Graphs, Search, and NLP

Advantages of Hyperbolic Networks is primarily observed in cases when hierarchical relations exist in the underlying datasets.

While graph datasets are general candidates for hyperbolic networks, they have also been used to exploit the hierarchy in the following applications:

1. Graph Analysis
2. Knowledge Graphs



Drug Repurposing Knowledge Graph*

* Ioannidis, Vassilis N., et al. "Drkg-drug repurposing knowledge graph for covid-19." *arXiv preprint arXiv: 2010.09600* (2020).

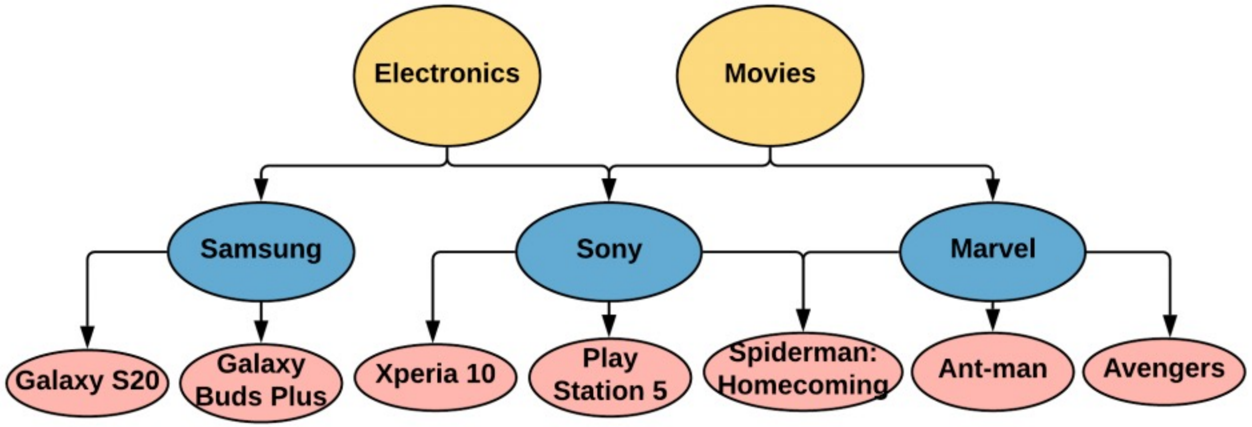
Overview of Applications

Graphs, Knowledge Graphs, Search, and NLP

Advantages of Hyperbolic Networks is primarily observed in cases when hierarchical relations exist in the underlying datasets.

While graph datasets are obvious candidates for hyperbolic networks, they have also been used to exploit the hierarchy in the following applications:

1. Graph Analysis
2. Knowledge Graphs
3. Search



Hierarchy in Product Catalogues*

*Choudhary, Nurendra, et al. "ANTHEM: Attentive Hyperbolic Entity Model for Product Search." WSDM 22 (2022): 21-25.

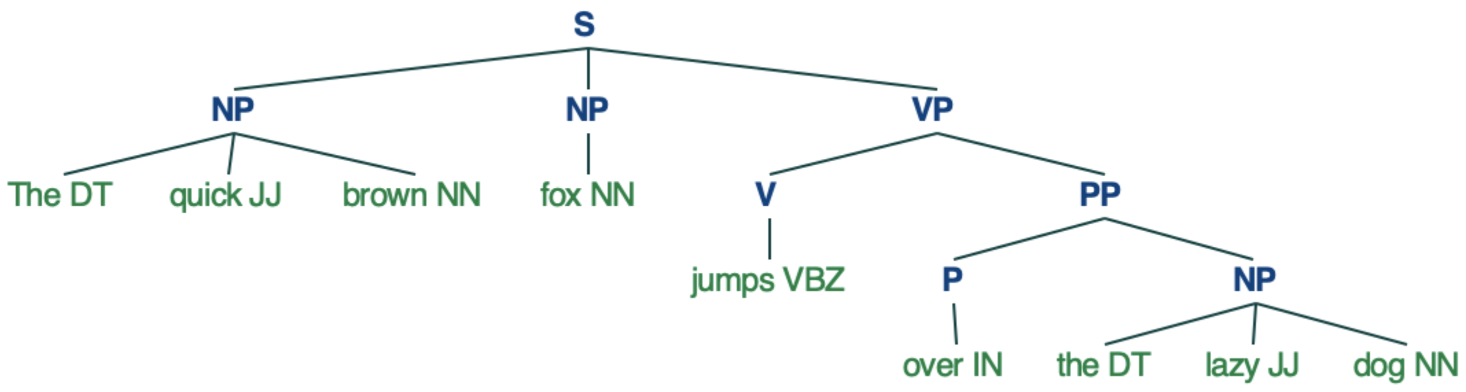
Overview of Applications

Graphs, Knowledge Graphs, Search, and NLP

Advantages of Hyperbolic Networks is primarily observed in cases when hierarchical relations exist in the underlying datasets.

While graph datasets are obvious candidates for hyperbolic networks, they have also been used to exploit the hierarchy in the following applications:

1. Graph Analysis
2. Knowledge Graphs
3. Search
4. Natural Language Processing



Parse trees of Sentences*

*https://www.geeksforgeeks.org/syntax-tree-natural-language-processing/

Overview of Applications

Graphs, Knowledge Graphs, Search and NLP

1. Graphs
   - ❑ Hyperbolic Neural Networks
   - ❑ Hyperbolic Graph Convolutional Neural Networks
   - ❑ Hyperbolic GRAph Meta Learner (H-GRAM)

2. Knowledge Graphs
   - ❑ Multi-relational Poincaré Embeddings
   - ❑ Low-Dimensional Hyperbolic Knowledge Graph Embeddings
   - ❑ Self-Supervised Hyperboloid Representations from Logical Queries over Knowledge Graphs

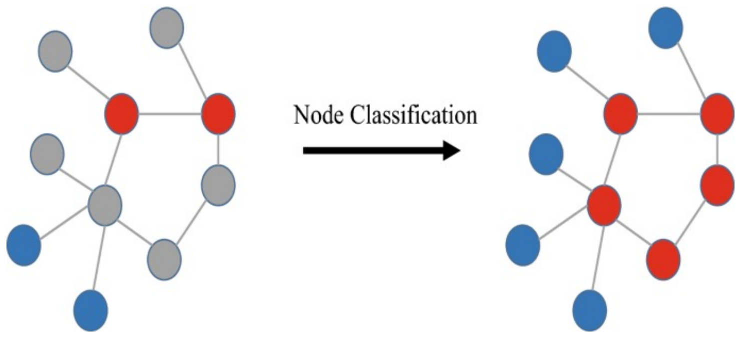
3. Search
   - ❑ ANTHEM: Attentive Hyperbolic Entity Model for Product Search

4. Natural Language Processing
   - ❑ Hyperbolic Representation Learning for Fast and Efficient Neural Question Answering
   - ❑ Text Enriched Sparse Hyperbolic Graph Convolution Network

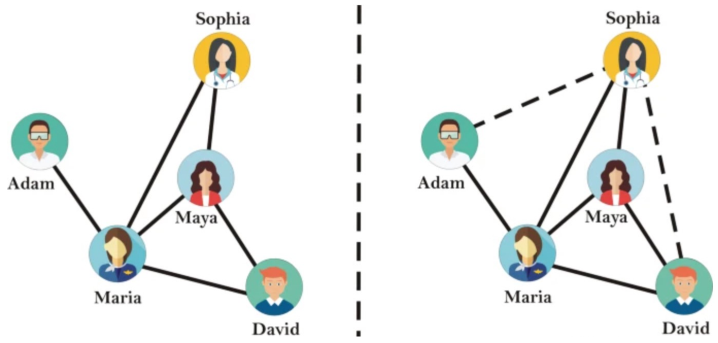
Applications
Graph Analysis: Introduction

Graphs are essential data structures that contain attributed (or) non-attributed nodes connected by edges.

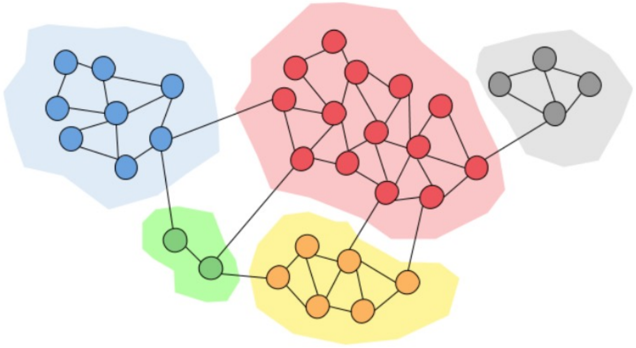
Several popular problems:
1. Node Classification
2. Link Prediction
3. Community Detection



Node Classification[1]



Link Prediction[2]



Community Detection[3]

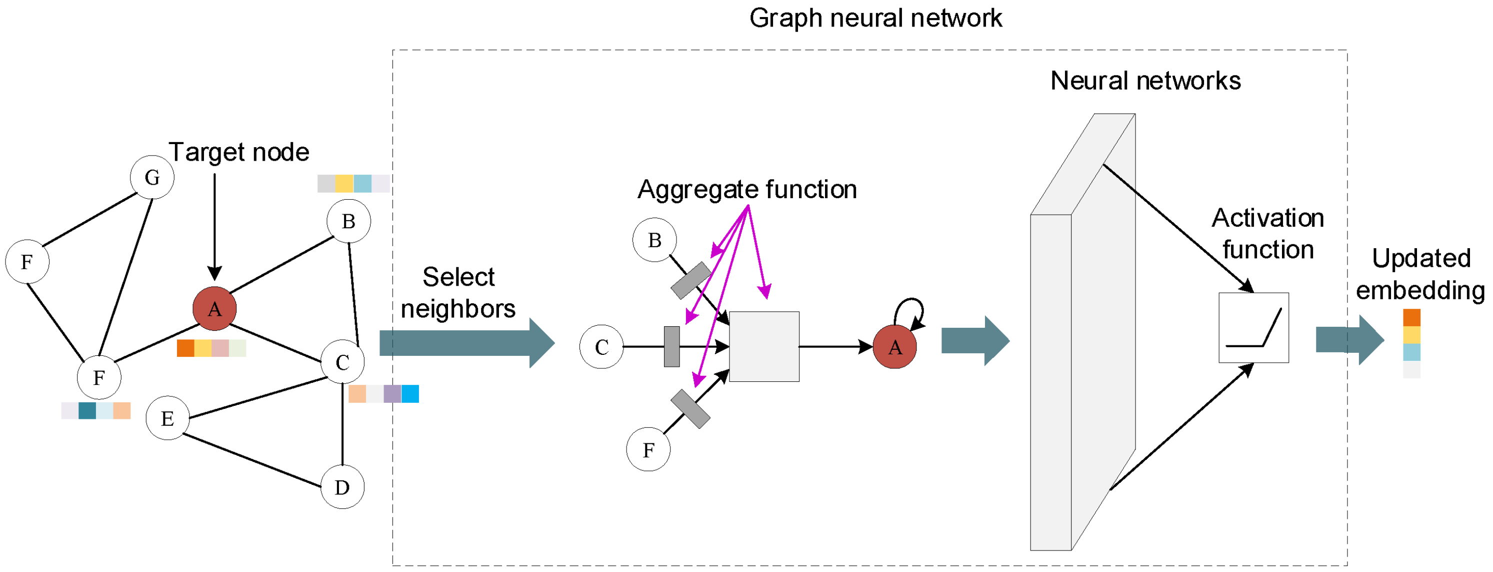
1. Chen, Linjun, Xingyi Liu, and Zexin Li. "Nonlinear Graph Learning-Convolutional Networks for Node Classification." Neural Processing Letters (2021): 1-10.
2. Ahmad, I., Akhtar, M. U., Noor, S., & Shahnaz, A. (2020). Missing link prediction using common neighbor and centrality based parameterized algorithm. Scientific reports.
3. Community Detection Algorithms, Thamindu Dilshan Jayawickrama, Towards Datascience.

Applications

Graph Analysis: Generic GNN

Generic Graph Neural Networks contains three components:

1. Neighborhood Selection:
   RandomWalk, DeepWalk, SkipWalk.
2. Message Passing:
   Linear, Convolution, Recurrent, Attention
3. Message Aggregation:
   Averaging, Max Pooling, MLP, Attention.

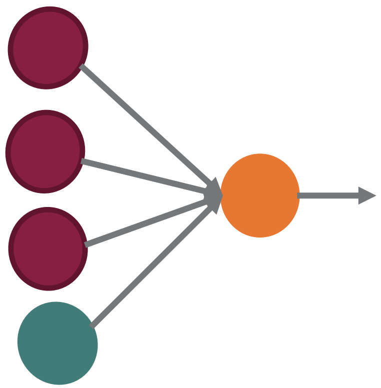


Generic Graph Neural Network*

* Zeng, Yufan, and Jiashan Tang. "Rlc-gnn: An improved deep architecture for spatial-based graph neural network with application to fraud detection." *Applied Sciences* 11.12 (2021): 5656.

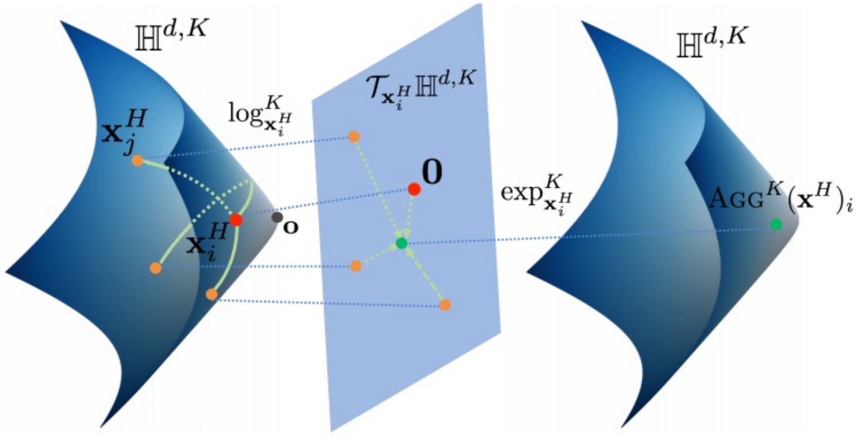
Applications

Graph Analysis: Hyperbolic Models (Previously discussed in Architectures)

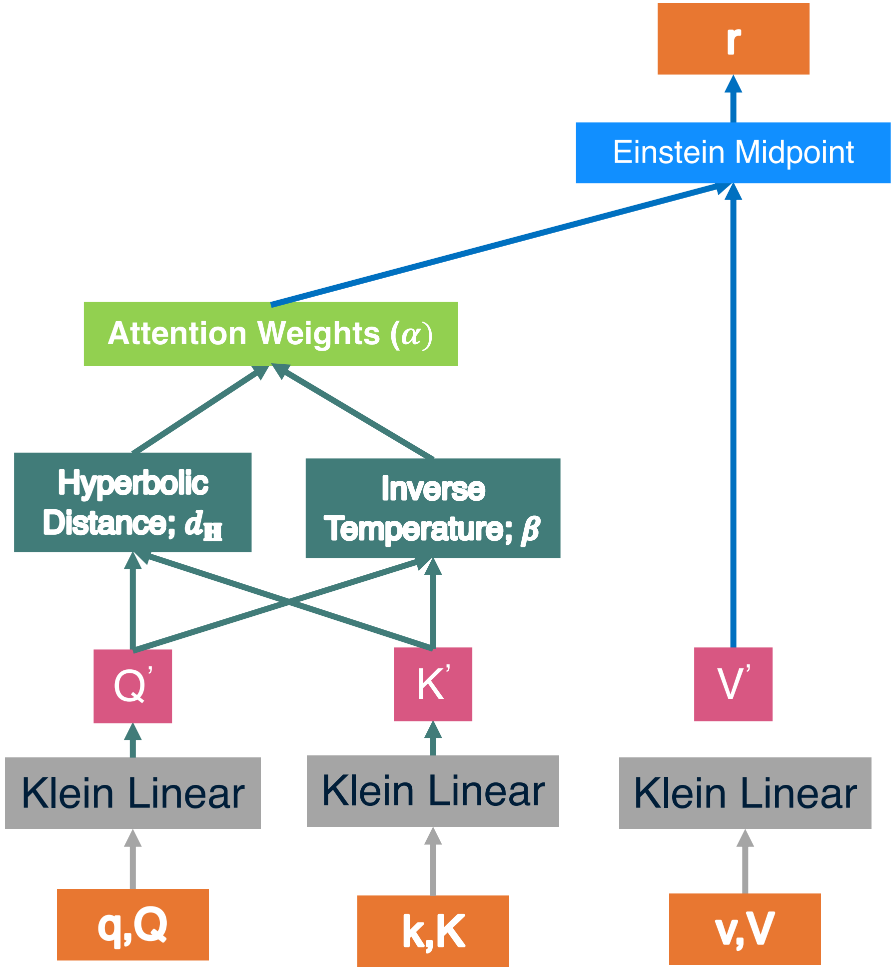


HNN

$w_1 \otimes_c x_1 \oplus_c w_2 \otimes_c x_2 \oplus_c w_3 \otimes_c x_3 \oplus_c \mathrm{b}$



HGCN



HGAT

Applications

Graph Analysis: Experimental Study

Evaluation Tasks:

Node Classification: Given a graph with nodes (attributed or non-attributed) and edges, estimate a model that predicts the class of unclassified existing (transductive) or new (inductive) nodes.

Link Prediction: Given a graph with nodes (attributed or non-attributed) and edges, estimate a model that predicts the probability of edges between existing (transductive) or new (inductive) node-pairs.

Datasets (δ is hyperbolicity, less δ implies more hierarchy in the dataset):

Area	Dataset	δ	Nodes	Edges	Labels	Description
Citation	CORA	11	Papers	Citations	Academic Subareas	Machine learning papers.
	PubMed	3.5	Papers	Citations	Academic Subareas	Medical papers.
Medicine	DISEASE	0	Testee	Interaction	Infected or not	SIR disease propagation model.
	PPI	1	Protein	Interaction	Stem cell growth rate	Union of ppi networks in human tissues.
Traffic	AIRPORT	1	Airports	Flight paths	Country population	Airline routes from OpenFlights.org.

Evaluation Metrics:

Node Classification: F1 score

Link Prediction: ROC-AUC score

Applications

Graph Analysis: Node Classification

Node Classification (F1 Score)										
		Shallow		Neural Nets		Graph Neural Nets				
Dataset	δ	EUC	HYP	MLP	HNN	GCN	GAT	SAGE	HGCN	HGAT
CORA	11	23.8 ± 0.7	22.0 ± 1.5	51.5 ± 1.0	54.6 ± 0.4	81.3 ± 0.3	83.0 ± 0.7	77.9 ± 2.4	79.9 ± 0.2	79.6 ± 0.3
PubMed	3.5	48.2 ± 0.7	68.5 ± 0.3	72.4 ± 0.2	69.8 ± 0.4	78.1 ± 0.2	79.0 ± 0.3	77.4 ± 2.2	80.3 ± 0.3	79.2 ± 0.3
DISEASE	0	32.5 ± 1.1	45.5 ± 3.3	28.8 ± 2.5	41.0 ± 1.8	69.7 ± 0.4	70.4 ± 0.4	69.1 ± 0.6	74.5 ± 0.9	73.4 ± 0.9
PPI	1	-	-	55.3+0.4	59.3 ± 0.4	69.7 ± 0.3	70.5 ± 0.4	69.1 ± 0.3	74.6 ± 0.3	73.6 ± 0.3
AIRPORT	1	60.9 ± 3.4	70.2 ± 0.1	68.6 ± 0.6	80.5 ± 0.5	81.4 ± 0.6	81.5 ± 0.3	82.1 ± 0.5	90.6 ± 0.2	89.4 ± 0.2

Applications

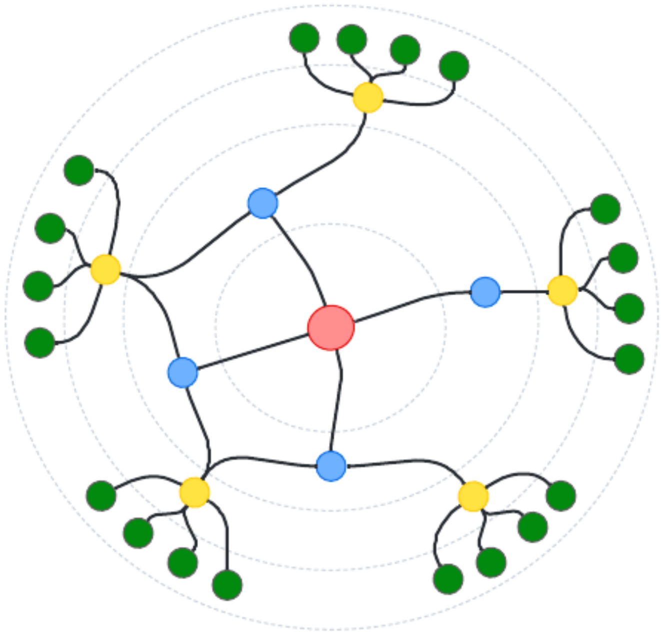
Graph Analysis: Link Prediction

Link Prediction (ROC AUC Score)										
		Shallow		Neural Nets		Graph Neural Nets				
Dataset	δ	EUC	HYP	MLP	HNN	GCN	GAT	SAGE	HGCN	HGAT
CORA	11	82.5 ± 0.3	87.6 ± 0.2	83.1 ± 0.5	89.0 ± 0.1	90.4 ± 0.2	93.7 ± 0.1	85.5 ± 0.6	92.9 ± 0.1	90.8 ± 0.2
PubMed	3.5	83.3 ± 0.1	87.5 ± 0.1	84.1 ± 0.9	94.9 ± 0.1	91.1 ± 0.5	91.2 ± 0.1	86.2 ± 1.0	96.3 ± 0.0	93.9 ± 0.2
DISEASE	0	59.8 ± 2.0	63.5 ± 0.6	72.6 ± 0.6	75.1 ± 0.3	64.7 ± 0.5	69.8 ± 0.3	65.9 ± 0.3	90.8 ± 0.3	88.6 ± 0.3
PPI	1	-	-	67.8 ± 0.2	72.9 ± 0.3	77.0 ± 0.5	76.8 ± 0.4	78.1 ± 0.6	84.5 ± 0.4	81.4 ± 0.3
AIRPORT	1	92.0 ± 0.0	94.5 ± 0.0	89.8 ± 0.5	90.8 ± 0.2	89.3 ± 0.4	90.5 ± 0.3	90.4 ± 0.5	96.4 ± 0.1	94.2 ± 0.2

Applications

Graph Analysis: Scalability in H-GRAM

- In Euclidean Graphs, we depend on local subgraph encodings to scale over large graph datasets.
- In Hyperbolic Graphs, we are not able to directly apply this because the representations are relative to an origin.

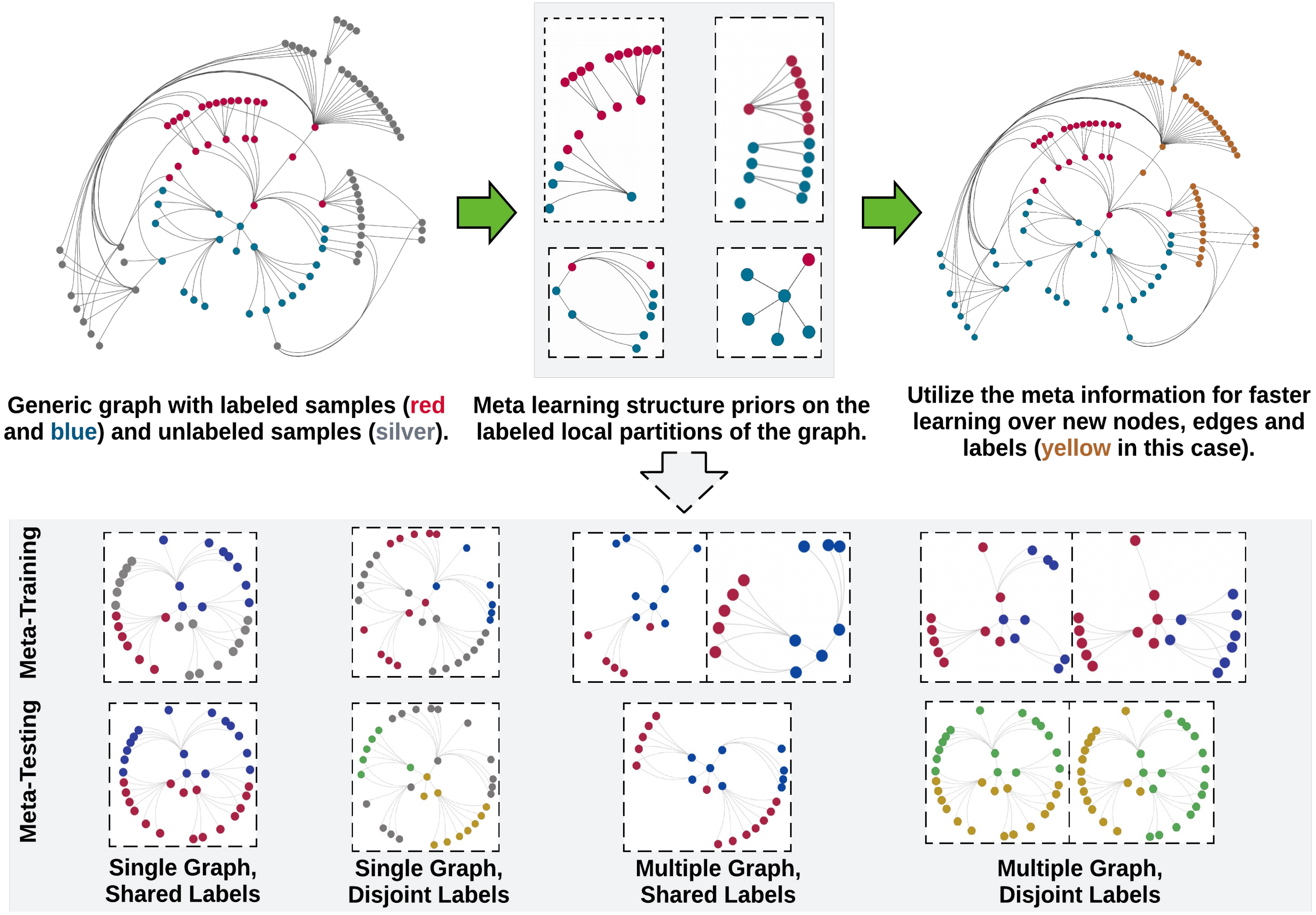


Hyperbolic Embeddings

Applications

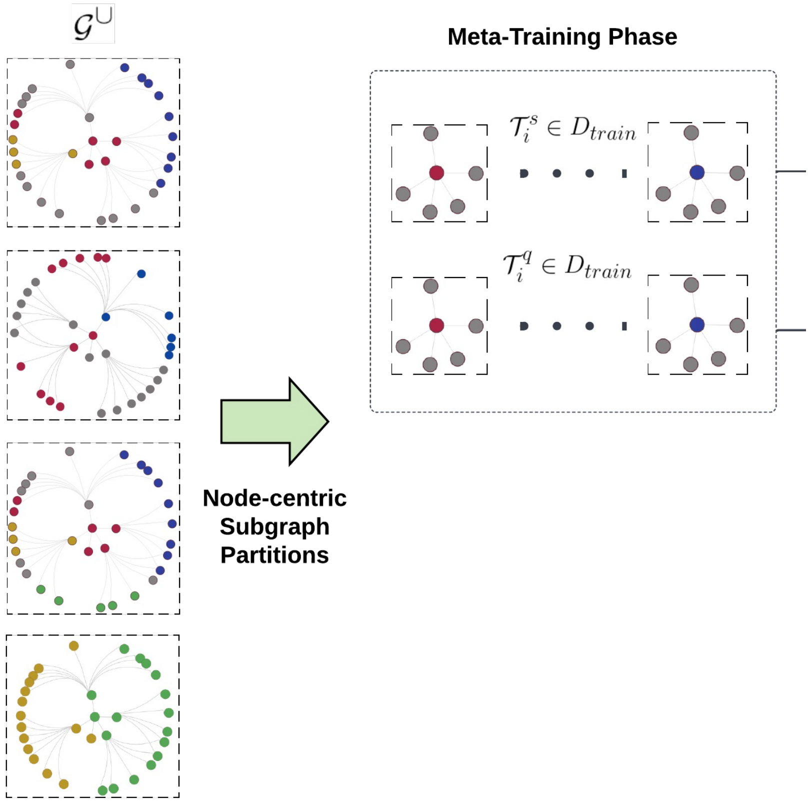
Graph Analysis: Scalability in H-GRAM

- It can be theoretically shown that one can move the origin to local subgraphs with a bounded information loss.
- Divide the graph into subgraphs and note four possible scenarios:
  - Single Graph, Shared Labels
  - Single Graph, Disjoint Labels
  - Multiple Graph, Shared Labels
  - Multiple Graph, Disjoint Labels



Applications

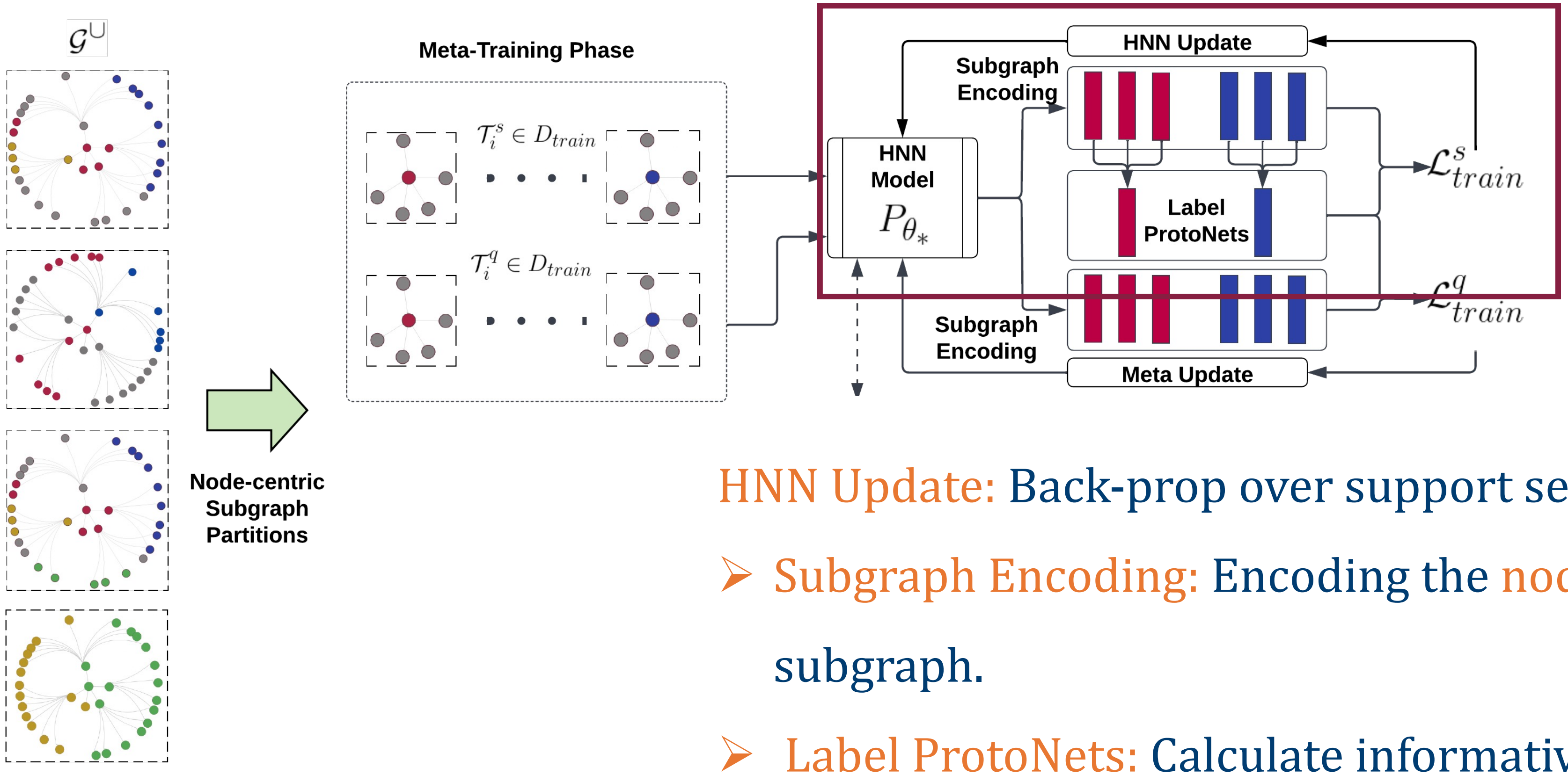
Graph Analysis: Scalability in H-GRAM



- In the case of Multiple Graphs or Disjoint Labels, we need to rely on Meta-learning for knowledge transfer between different subgraphs.
- In Meta-learning, we partition the problem into;
  - Meta-training: only training samples
  - Meta-testing: few training samples

Applications

Graph Analysis: Scalability in H-GRAM

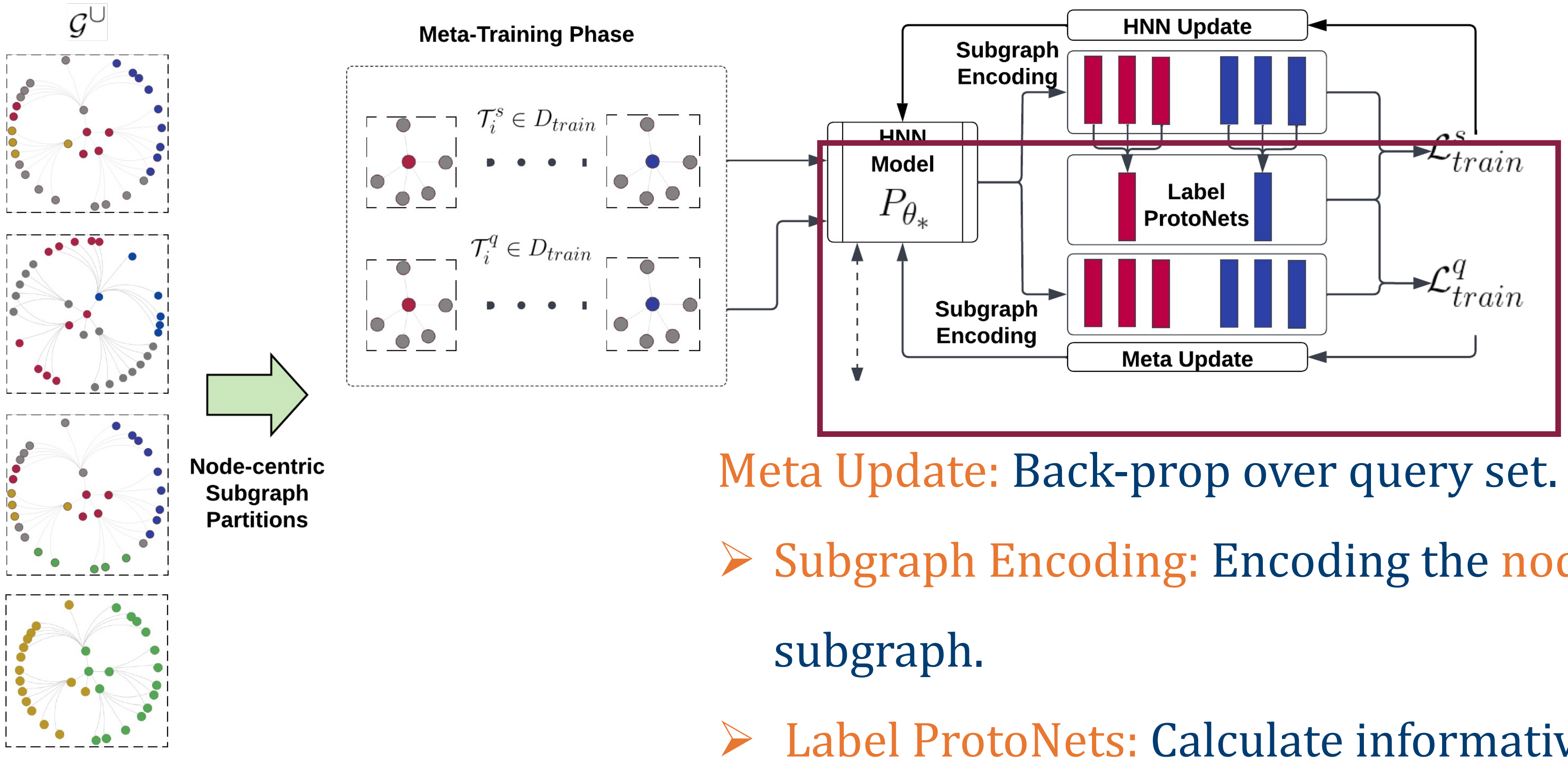


HNN Update: Back-prop over support set.

- Subgraph Encoding: Encoding the node-centric subgraph.
- Label ProtoNets: Calculate informative continuous label prototypes.

Applications

Graph Analysis: Scalability in H-GRAM

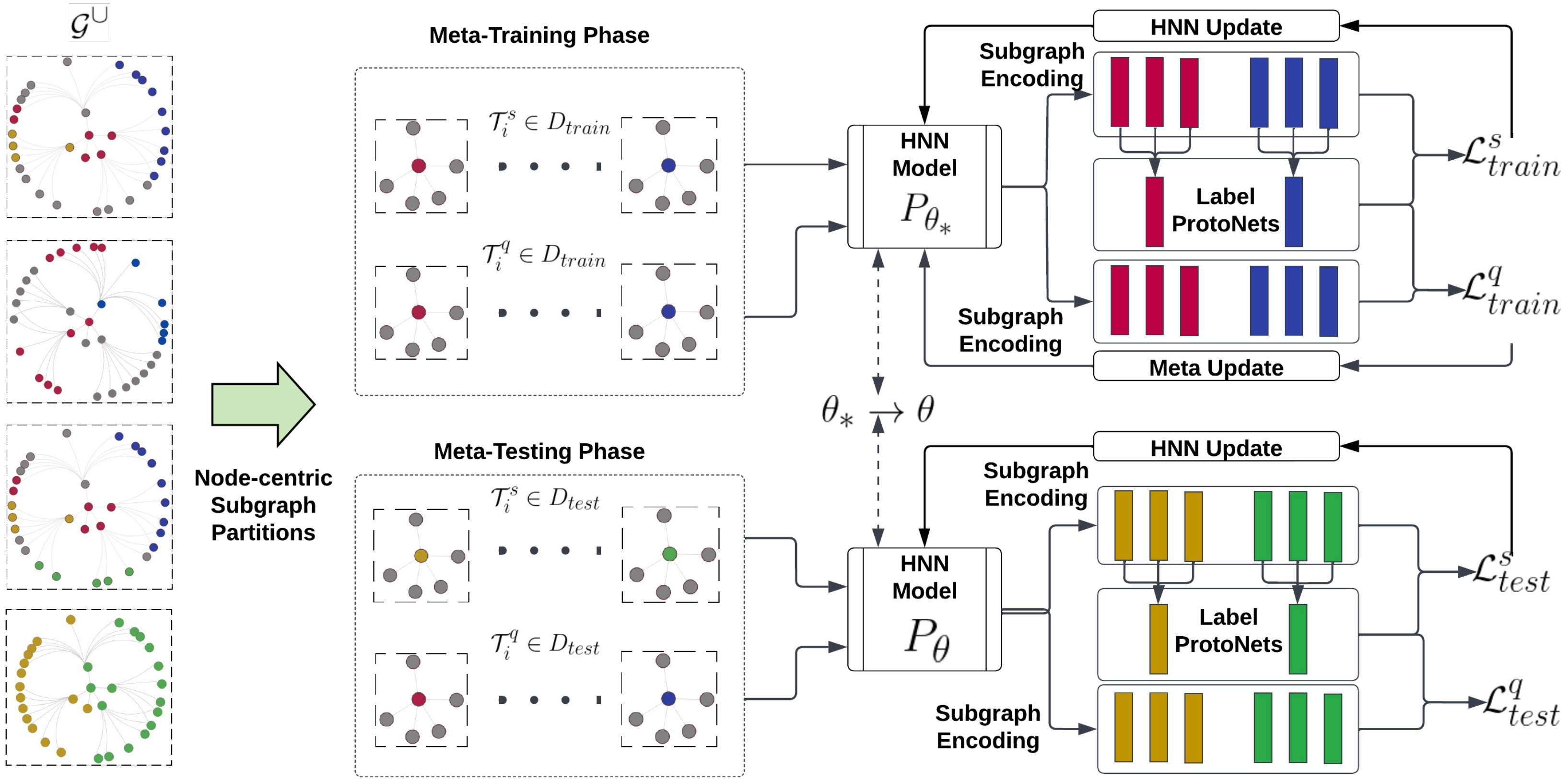


Meta Update: Back-prop over query set.

- Subgraph Encoding: Encoding the node-centric subgraph.
- Label ProtoNets: Calculate informative continuous label prototypes.
- Aggregate over a task and meta update.

Applications

Graph Analysis: Scalability in H-GRAM



Meta Testing:

- HNN Updates: Few-shot over support set of test data.

Prediction over query set of test data for final evaluation.

Applications

Graph Analysis: Scalability in H-GRAM

1. Performance of H-GRAM
2. Challenging Few-shot Settings
3. Time Comparison and Ablation Study

Applications

Graph Analysis: Scalability in H-GRAM

1. Datasets: Synthetic Cycle graph and Synthetic Barabási-Albert graph, ogbn-arxiv, Tissue-PPI, FirstMM-DB, Fold-PPI, Tree-of-Life, Cora, PubMed, and Citeseer.
2. Baselines: Meta-Graph, Meta-GNN, FS-GIN, FS-SGC, ProtoNet, MAML, HMLP, HGCN, and HAT.
3. Evaluation: Accuracy of Node Classification and Link Prediction

Applications

Graph Analysis: Scalability in H-GRAM

Task	Node Classification		Node Classification		Node Classification		Node Classification			Link Prediction	
Setup	SG,DL		MG,SL		MG,DL		SG,DL	MG,SL	MG,DL	MG,SL	MG,SL
Dataset	Syn. Cycle	Syn. BA	Syn. Cycle	Syn. BA	Syn. Cycle	Syn. BA	ogbn-arxiv	Tissue-PPI	Fold-PPI	FirstMM-DB	Tree-of-Life
Meta-Graph	-	-	-	-	-	-	-	-	-	0.719	0.705
Meta-GNN	0.72	0.694	-	-	-	-	0.273	-	-	-	-
FS-GIN	0.684	0.749	-	-	-	-	0.336	-	-	-	-
FS-SGC	0.574	0.715	-	-	-	-	0.347	-	-	-	-
ProtoNet	0.821	0.858	0.282	0.657	0.749	0.866	0.372	0.546	0.382	0.779	0.697
MAML	0.842	0.848	0.511	0.726	0.653	0.844	0.389	0.745	0.482	0.758	0.719
G-META	0.872	0.867	0.542	0.734	0.767	0.867	0.451	0.768	0.561	0.784	0.722
H-GRAM	0.883	0.873	0.555	0.746	0.779	0.888	0.472	0.786	0.584	0.804	0.742

Accuracy of H-GRAM compared to Euclidean baselines on Node Classification and Link Prediction

Applications

Graph Analysis: Scalability in H-GRAM

Task	Node Classification		Node Classification		Node Classification		Node Classification			Link Prediction	
Setup	SG,DL		MG,SL		MG,DL		SG,DL	MG,SL	MG,DL	MG,SL	MG,SL
Dataset	Syn. Cycle	Syn. BA	Syn. Cycle	Syn. BA	Syn. Cycle	Syn. BA	ogbn-arxiv	Tissue-PPI	Fold-PPI	FirstMM-DB	Tree-of-Life
Meta-Graph	-	-	-	-	-	-	-	-	-	0.719	0.705
Meta-GNN	0.720	0.694	-	-	-	-	0.273	-	-	-	-
FS-GIN	0.684	0.749	-	-	-	-	0.336	-	-	-	-
FS-SGC	0.574	0.715	-	-	-	-	0.347	-	-	-	-
ProtoNet	0.821	0.858	0.282	0.657	0.749	0.866	0.372	0.546	0.382	0.779	0.697
MAML	0.842	0.848	0.511	0.726	0.653	0.844	0.389	0.745	0.482	0.758	0.719
G-META	0.872	0.867	0.542	0.734	0.767	0.867	0.451	0.768	0.561	0.784	0.722
H-GRAM	0.883	0.873	0.555	0.746	0.779	0.888	0.472	0.786	0.584	0.804	0.742

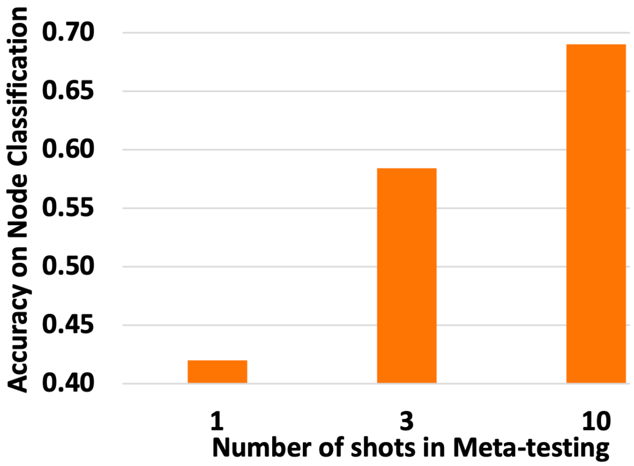
Accuracy of H-GRAM compared to Euclidean baselines on Node Classification and Link Prediction

Dataset	Cora		Pubmed		Citeseer	
Task	Node	Link	Node	Link	Node	Link
HMLP	0.754	0.765	0.657	0.848	0.879	0.877
HAT	0.796	0.792	0.681	0.908	0.939	0.922
HGCN	0.779	0.789	0.696	0.914	0.95	0.928
H-GRAM	0.827	0.790	0.716	0.896	0.924	0.936

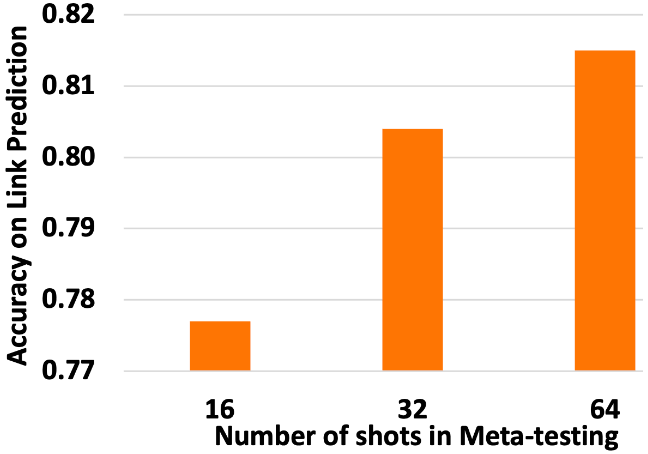
Accuracy of H-GRAM compared to Hyperbolic baselines on Node Classification and Link Prediction

Applications

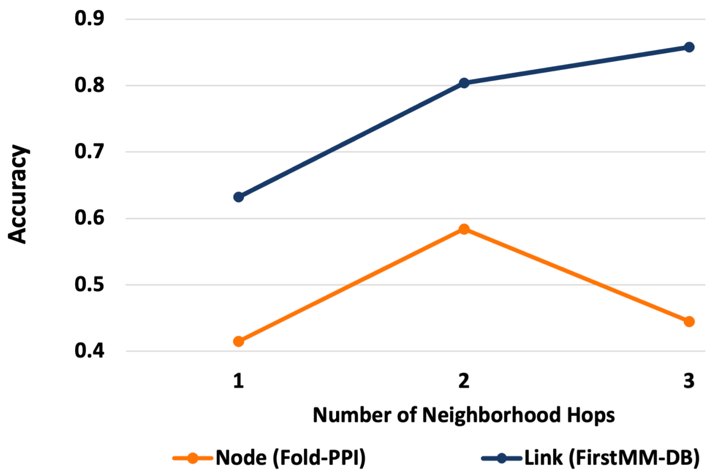
Graph Analysis: Scalability in H-GRAM



Number of shots (vs) Accuracy in Node Classification on Fold-PPI.



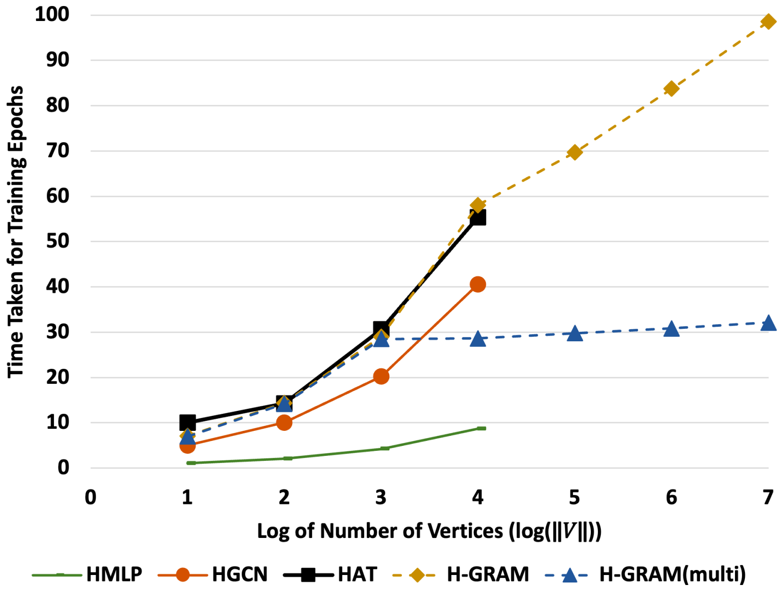
Number of shots (vs) Accuracy in Link Prediction on FirstMM-DB.



Number of neighborhood hops (vs) Accuracy on NC and LP.

Applications

Graph Analysis: Scalability in H-GRAM



Comparison of time taken by different HNNs with increasing number of vertices.

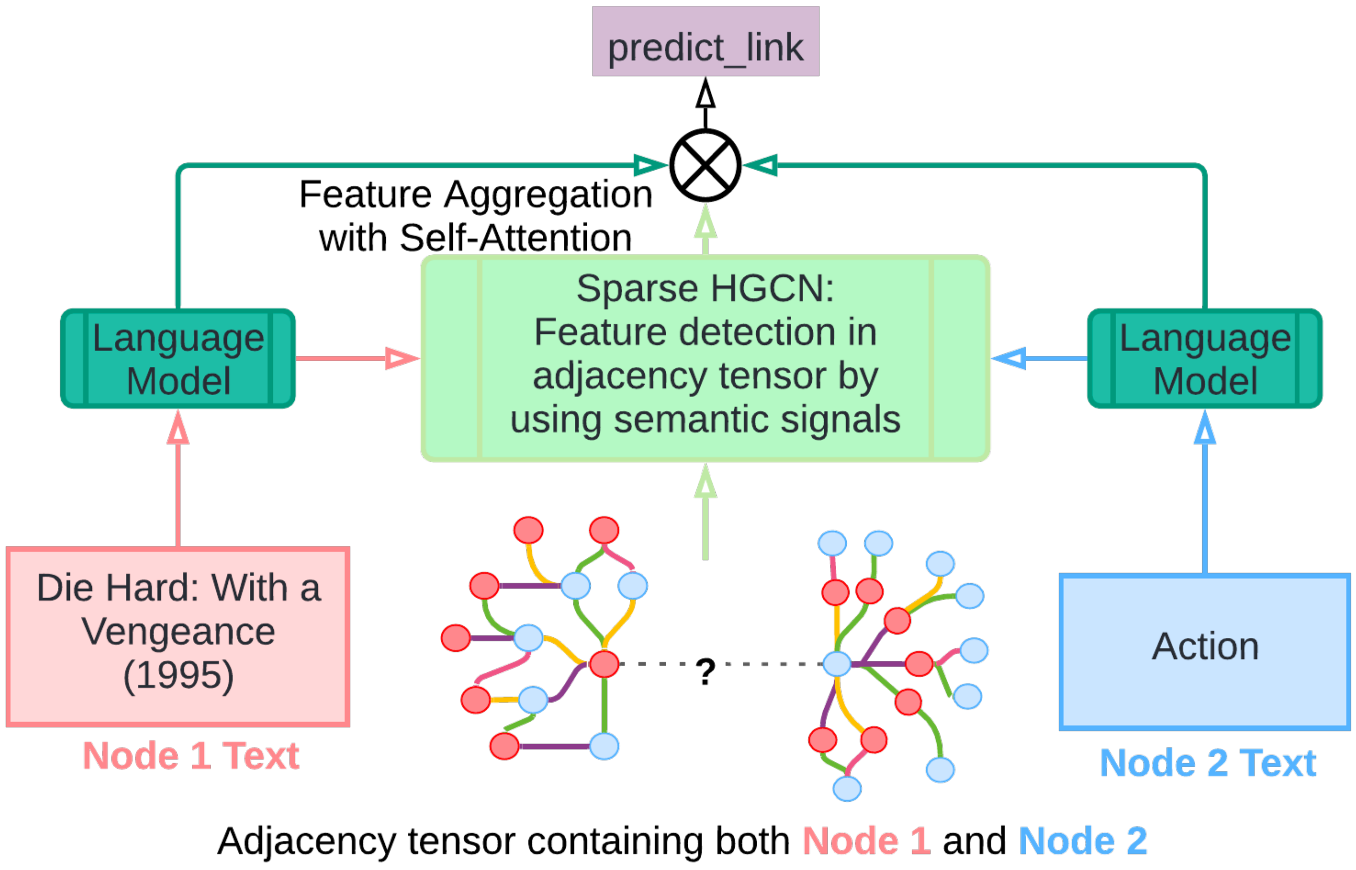
Task	Node Classification			Link Prediction	
Setup	SG,DL	MG,SL	MG,DL	MG,SL	MG,SL
Dataset	ogbn-arxiv	Tissue-PPI	Fold-PPI	FirstMM-DB	Tree-of-Life
H-ProtoNet	0.389	0.559	0.398	0.799	0.716
H-MAML	0.407	0.762	0.502	0.777	0.739
H-GRAM (HMLP)	0.37	0.537	0.372	0.772	0.688
H-GRAM (HAT)	0.462	0.777	0.573	0.794	0.732
H-GRAM (HGCN)	0.472	0.786	0.584	0.804	0.742

Ablation Study.

Applications

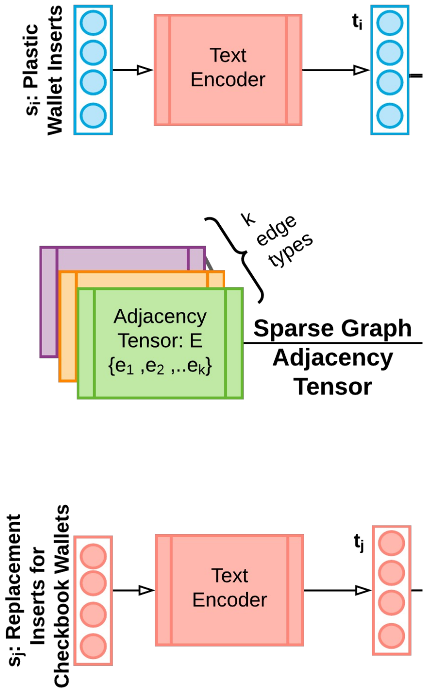
Multi-modal Graph Processing: TESH-GCN

In TESH-GCN, we try to utilize the inherent node information to extract both local and global graph features and integrate it with the nodes' information to finally solve a downstream task.



Applications

Multi-modal Graph Processing: TESH-GCN



We use Text and Graph-based Link Prediction as a running example in this work. But the method is extensible to any other modality and downstream task.

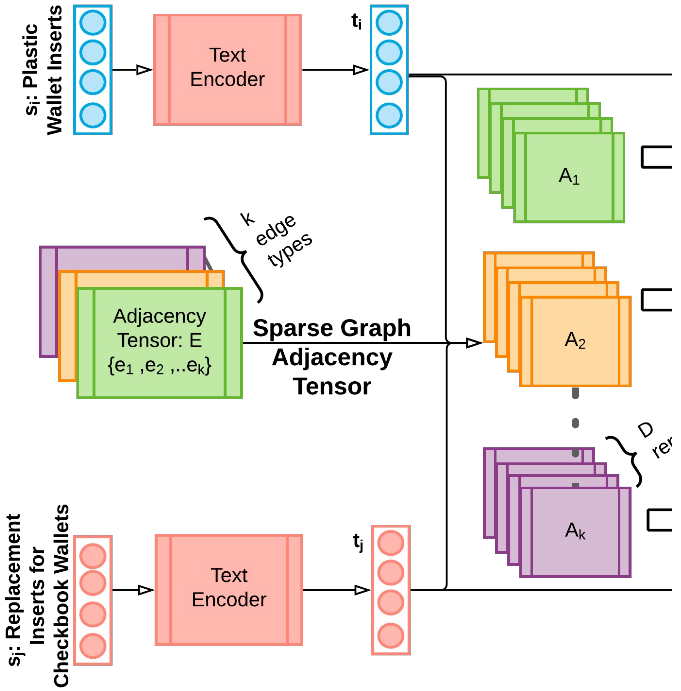
The inputs to TESH-GCN are:

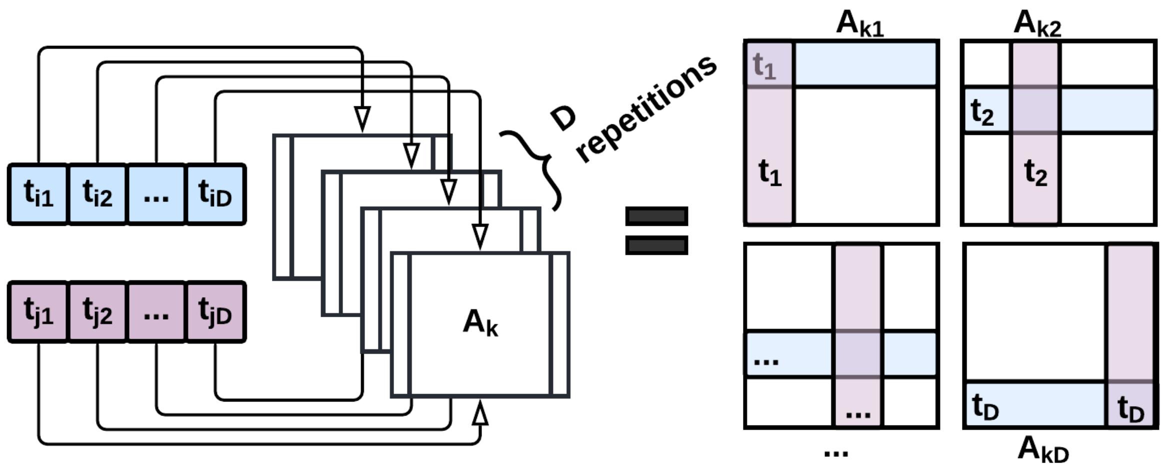
- Text from Node 1.
- Text from Node 2.
- Sparse Adjacency Tensor of the graph.
  - Graph is a tensor for multiple-relations and a matrix for the case of single-relation.

The text from Node 1 and Node 2 are encoded using a trainable language model.

Applications

Graph Analysis: Scalability with H-GRAM

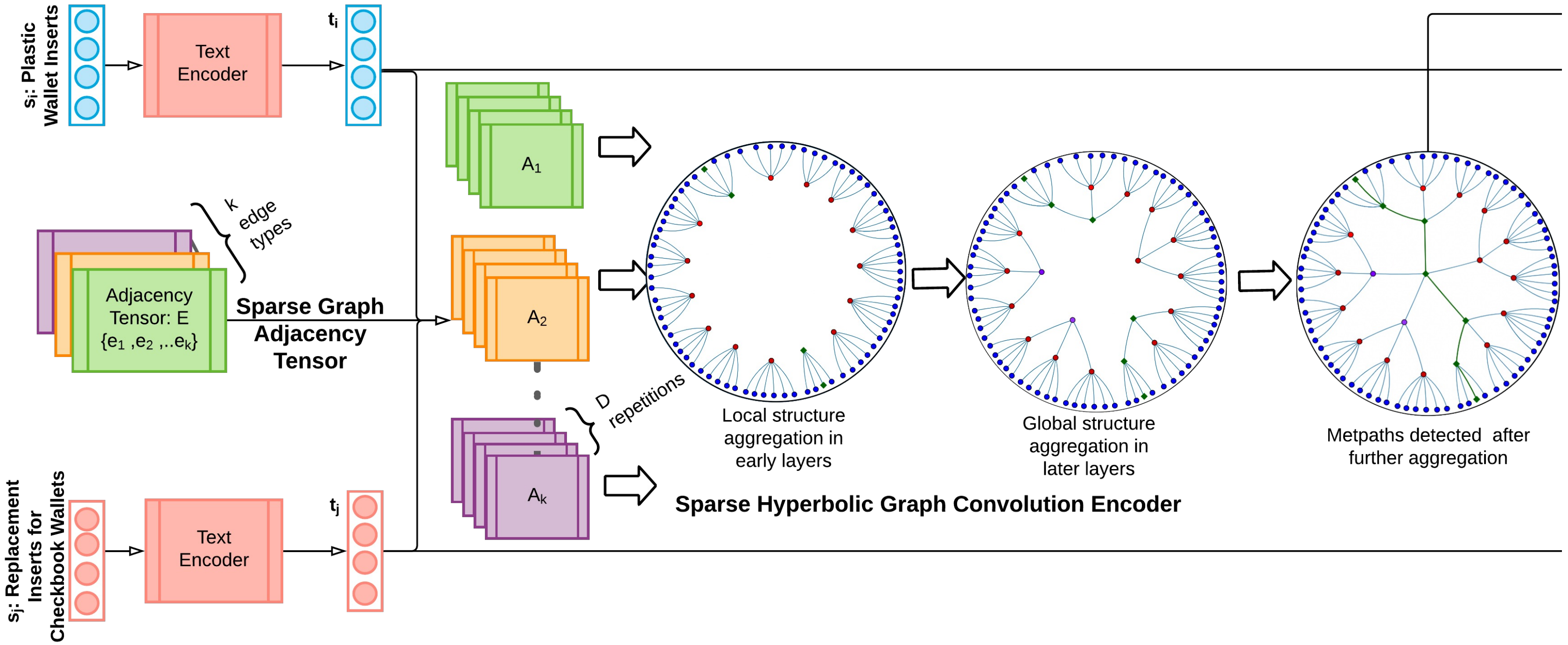




The text encoding from Node 1 and Node 2 are added to the Adjacency tensor to capture both semantic information and location of the nodes.

Applications

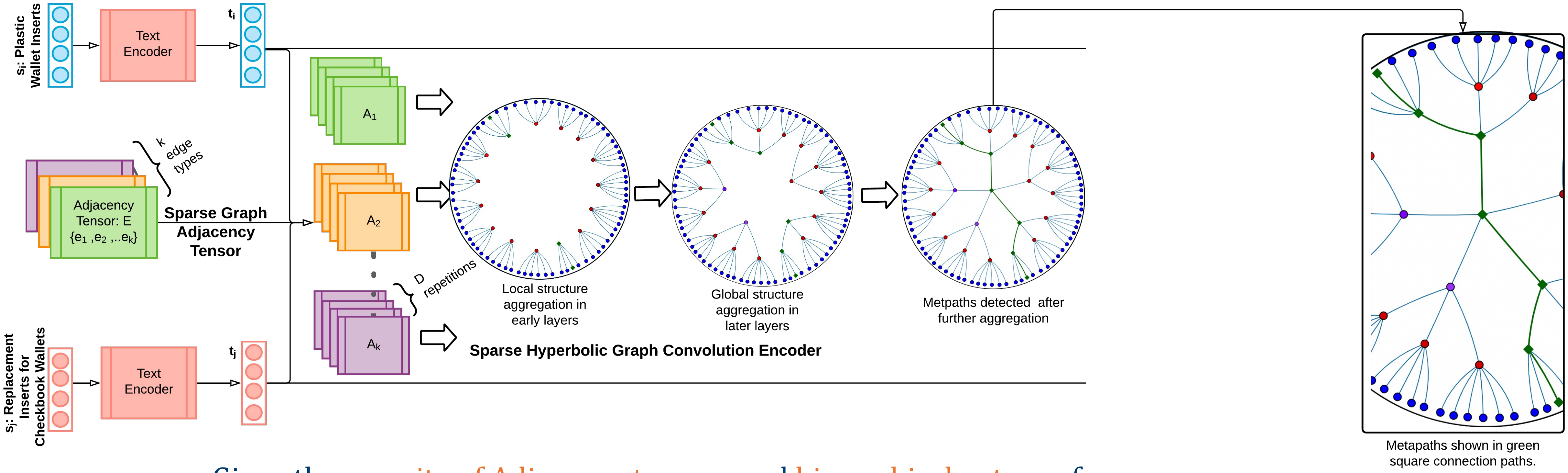
Multi-modal Graph Processing: TESH-GCN



Given the sparsity of Adjacency tensors and hierarchical nature of graph structure, we use an 8-layer Sparse HGCN encoder.
We notice that in most of the cases and 8-layer network is able to identify the metapath between the input nodes.

Applications

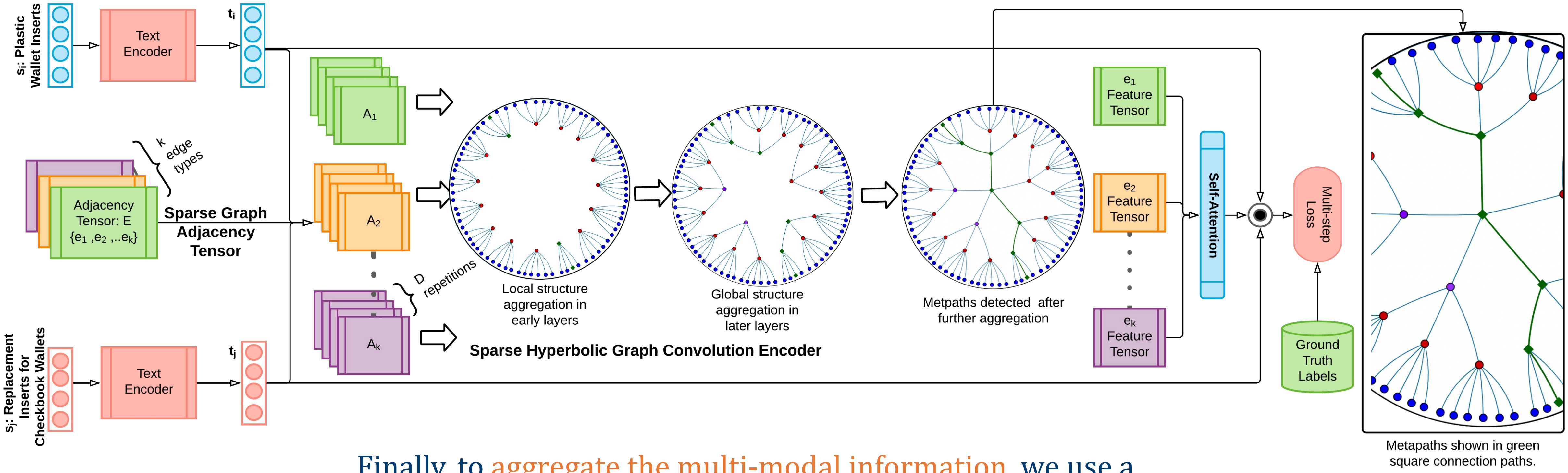
Multi-modal Graph Processing: TESH-GCN



Given the sparsity of Adjacency tensors and hierarchical nature of graph structure, we use an 8-layer Sparse HGCN encoder.
We notice that in most of the cases and 8-layer network is able to identify the metapath between the input nodes.

Applications

Multi-modal Graph Processing: TESH-GCN



Finally, to aggregate the multi-modal information, we use a self-attention network over the text and graph features.

Applications

Multi-modal Graph Processing: Evaluation

1. Performance on Link Prediction
2. Ablation Study
3. Time and Memory Complexity
4. Robustness
5. Example Metapaths

Applications

Multi-modal Graph Processing: Evaluation

- Task: Link Prediction
  - Dataset: Amazon, DBLP, Twitter, Cora, MovieLens
  - Baselines: Text-based (BERT), Graph-based (HGCN), Hybrid approach (TextGCN)
  - Evaluation Metrics: Accuracy, F1

Applications

Multi-modal Graph Processing: Evaluation

	Amazon		DBLP		Twitter		Cora		MovieLens	
Models	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
BERT	0.787	0.784	0.604	0.603	0.667	0.641	0.757	0.751	0.76	0.752
HGCN	0.71	0.703	0.547	0.533	0.608	0.598	0.929	0.923	0.685	0.677
TextGCN	0.817	0.809	0.624	0.616	0.671	0.669	0.862	0.856	0.789	0.78
TESH-GCN	**0.829**	**0.836**	**0.636**	**0.640**	**0.709**	**0.670**	**0.940**	**0.918**	**0.806**	**0.801**

Applications

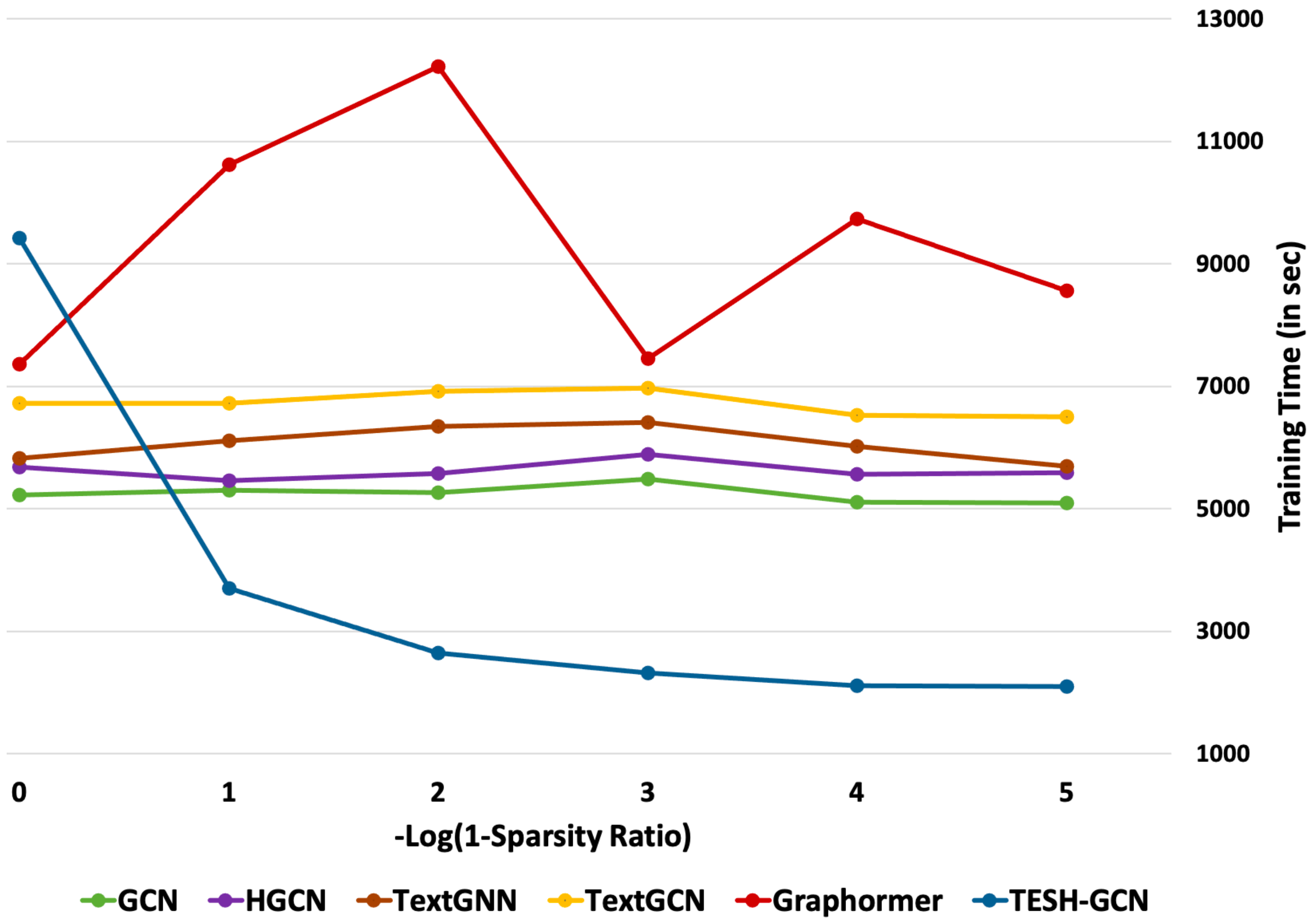
Multi-modal Graph Processing: Ablation Study

	Amazon		DBLP		Twitter		Cora		MovieLens	
Models	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
TESH-GCN	0.829	0.836	0.636	0.64	0.709	0.67	0.94	0.918	0.806	0.801
w/o Text	0.784	0.784	0.599	0.599	0.645	0.622	0.854	0.824	0.759	0.748
w/o Hyperbolic	0.677	0.678	0.522	0.516	0.577	0.585	0.787	0.757	0.655	0.66
w/o Residual	0.826	0.829	0.629	0.632	0.699	0.658	0.937	0.913	0.796	0.795

Applications

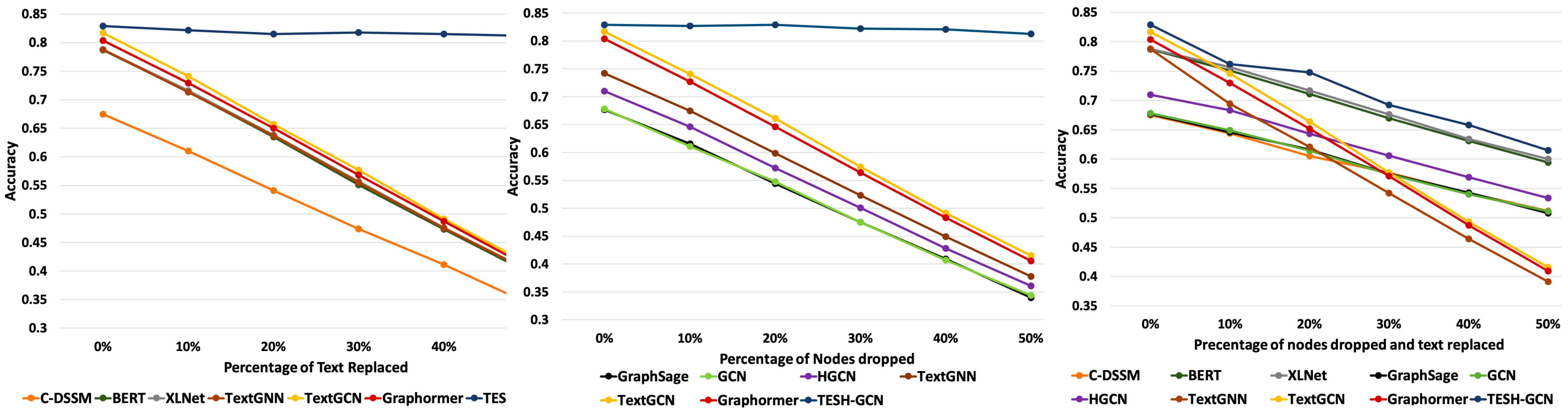
Multi-modal Graph Processing: Time and Memory Complexity

TESH-GCN's training time is more than other methods for dense graphs, but as the graph's get sparser, the training time reduces drastically.



Applications

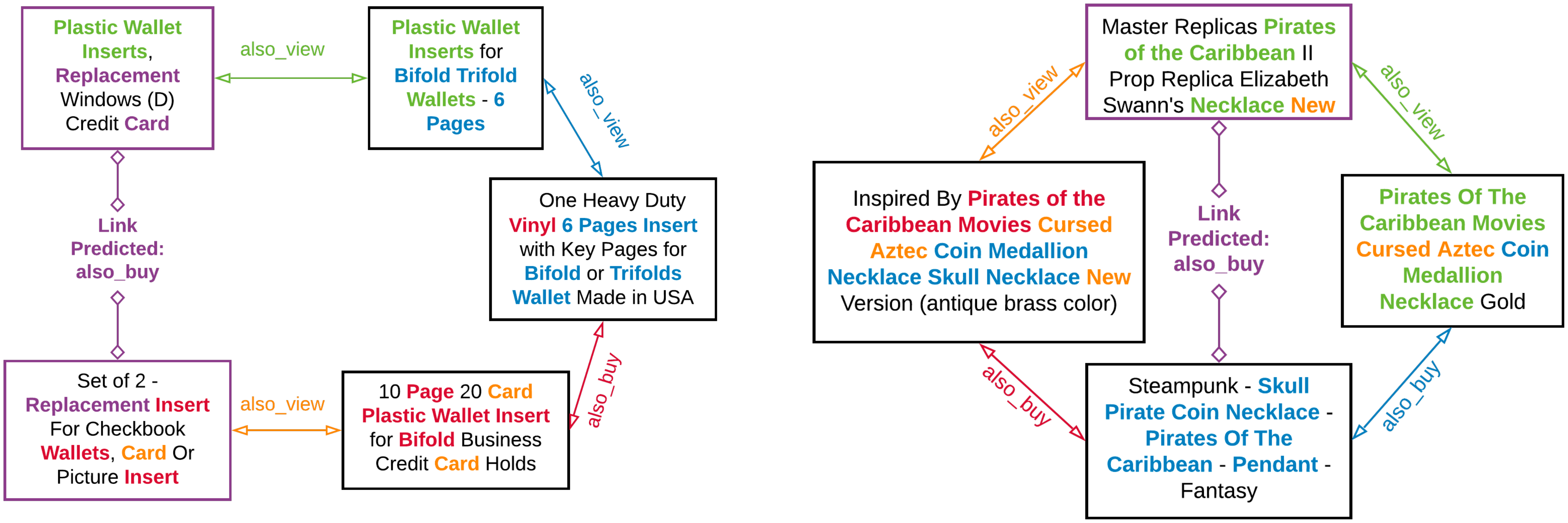
Multi-modal Graph Processing: Robustness



Robustness is tested by text replacement, node drops and a hybrid node drop with text replacement.

Applications

Multi-modal Graph Processing: Visualization of Metapaths



Applications

Graph Analysis: Learnings

- Hyperbolic space are consistently better at simultaneously capturing hierarchical structure from more tree-like datasets where the hyperbolicity is lower.
- Hyperbolic networks are able to perform better on both message aggregation (node classification) and message passing (link prediction)
- HGCN and HGAT are also able to reduce the error margins compared to their Euclidean counterparts.

Applications

Knowledge Graphs: Introduction

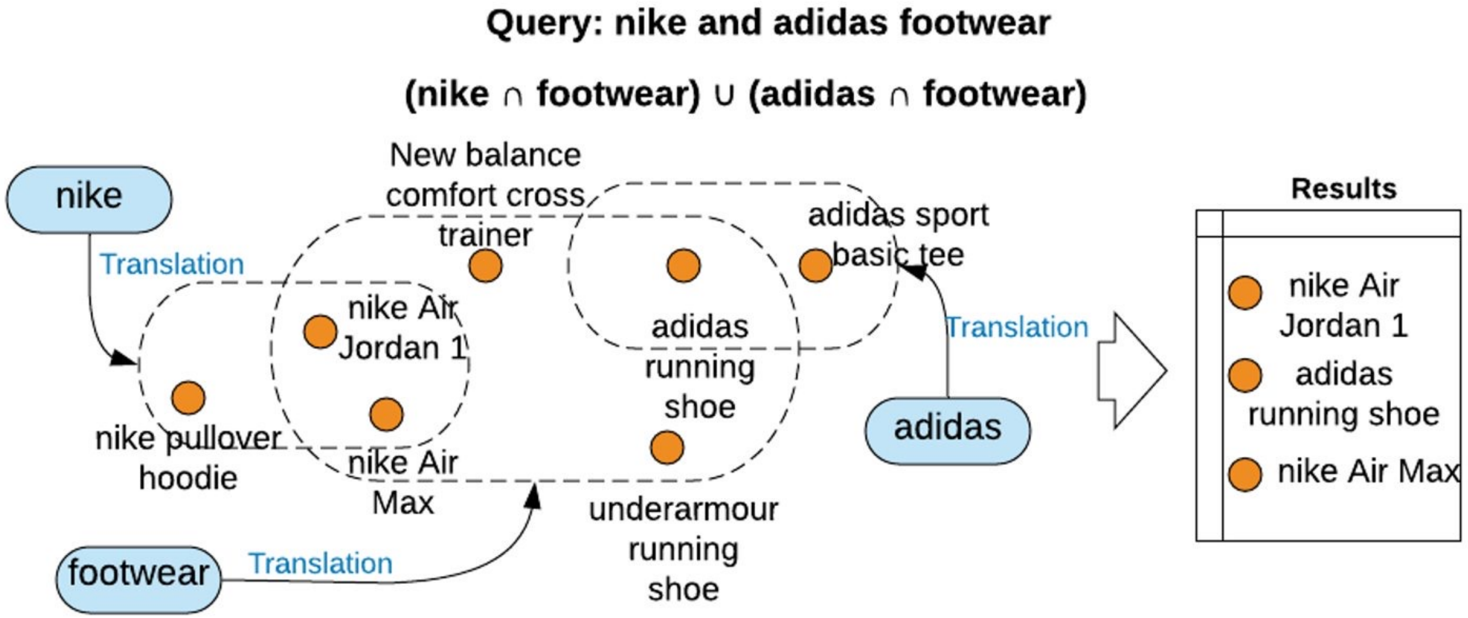
Knowledge Graphs are ubiquitous data structures.

KG querying is computationally expensive due to its size (≈10M nodes with trillions of relations).

Representation Learning can help!

Learn representations of entities and relations in a latent space.

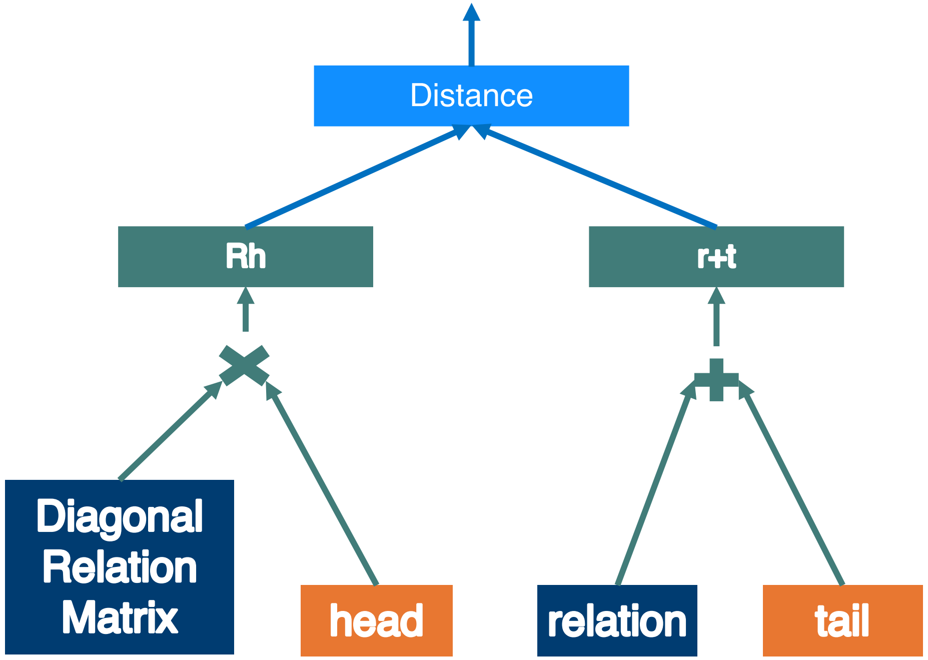
Apply logical operators to simulate querying behaviour.



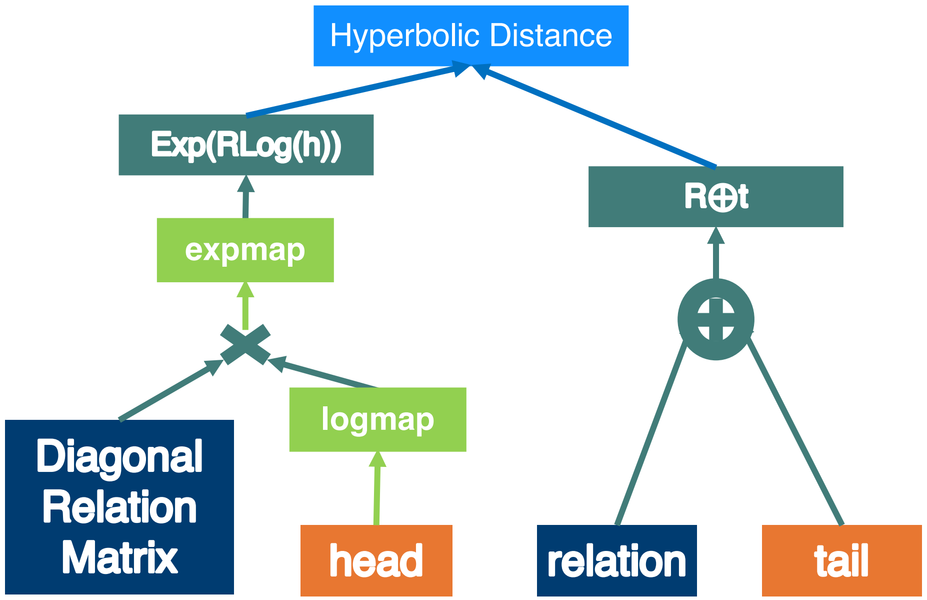
Applications

Knowledge Graphs: Poincaré Embeddings

Multi-relational Poincaré (MuRP) embedding model is an analogue to translational Euclidean representation model (MuRE).



MuRE



MuRP

Applications
MuRP: Experimental Study

Evaluation Tasks:

Relational Link Prediction: Given a graph with attributed nodes and attributed edges, estimate a model that predicts the relevance of other entities as tail answers for an input head entity and relation.

Datasets:

Dataset	# Entity	# Relation	Description
FB15k-237	14,541	237	Collection of real world facts from Freebase network.
WN18RR	40,943	11	Hierarchical collection of relations between words.

Evaluation Metrics:

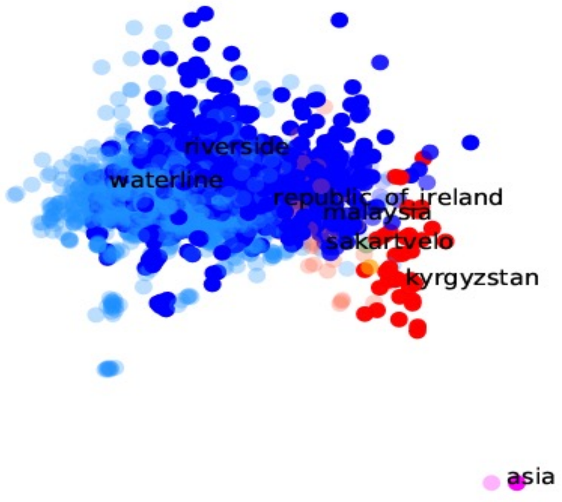
Relational Link Prediction : HITS@1,3,10 and Mean Reciprocal Rank (MRR).

Applications

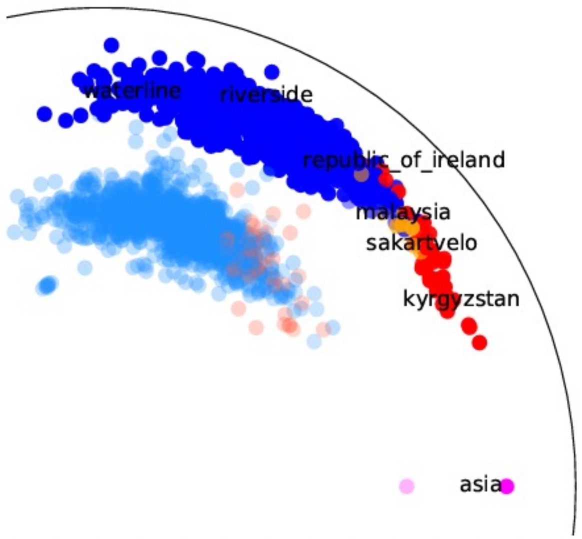
MuRP: Relational Link Prediction

Relational Link Prediction								
Dataset	WN18RR				FB15k-237			
Models	MRR	HITS@10	HITS@3	HITS@1	MRR	HITS@10	HITS@3	HITS@1
TransE	0.226	0.501	-	-	0.294	0.465	-	-
DistMult	0.43	0.49	0.44	0.39	0.241	0.419	0.263	0.155
ComplEx	0.44	0.51	0.46	0.41	0.247	0.428	0.275	0.158
Neural LP	-	-	-	-	0.25	0.408	-	-
MINERVA	-	-	-	-	-	0.456	-	-
ConvE	0.43	0.52	0.44	0.4	0.325	0.501	0.356	0.237
M-Walk	0.437	-	0.445	0.414	-	-	-	-
RotatE	-	-	-	-	0.297	0.48	0.328	0.205
MuRP	0.481	566	0.495	0.44	0.335	0.518	0.367	0.243

Applications
MuRP: Visualization



MuRE



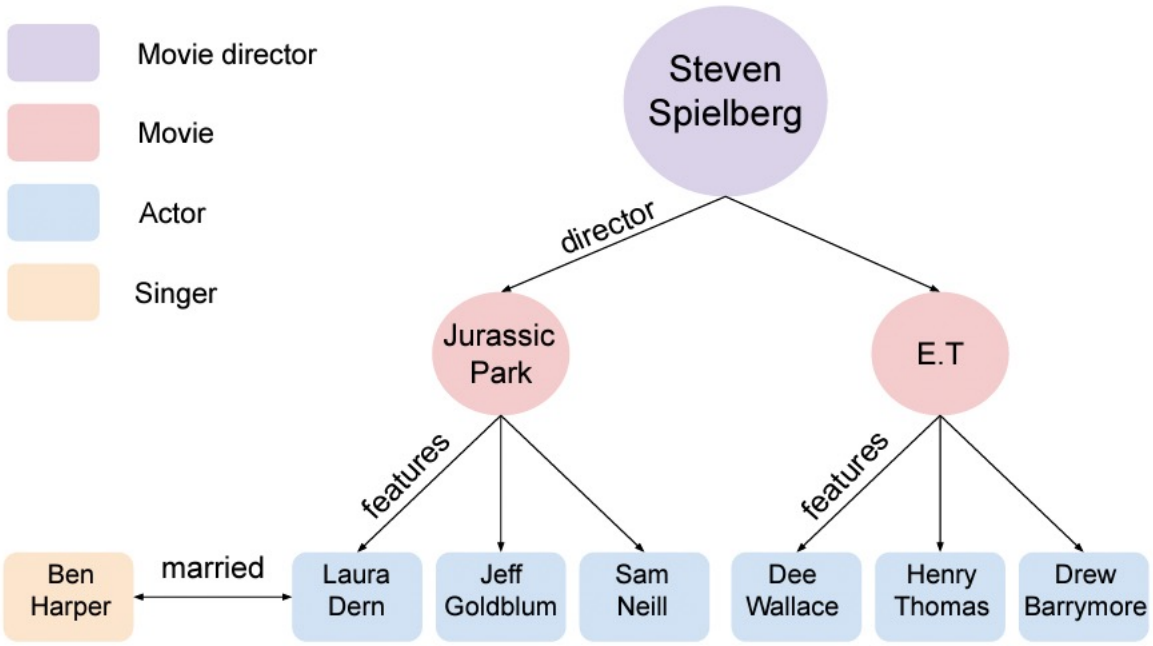
MuRP

Learned 40-dimensional MuRP and MuRE embeddings for WN18RR relation has_part.
Blue indicates **true positives** and **red** indicates **true negatives**. The lighter shades of the color indicate the entity before the relational translation.
Notice that the points are more separable in the Poincaré space when compared to the Euclidean space, where they are clustered together.

Applications

Knowledge Graphs: Rotation, Reflection and Attention

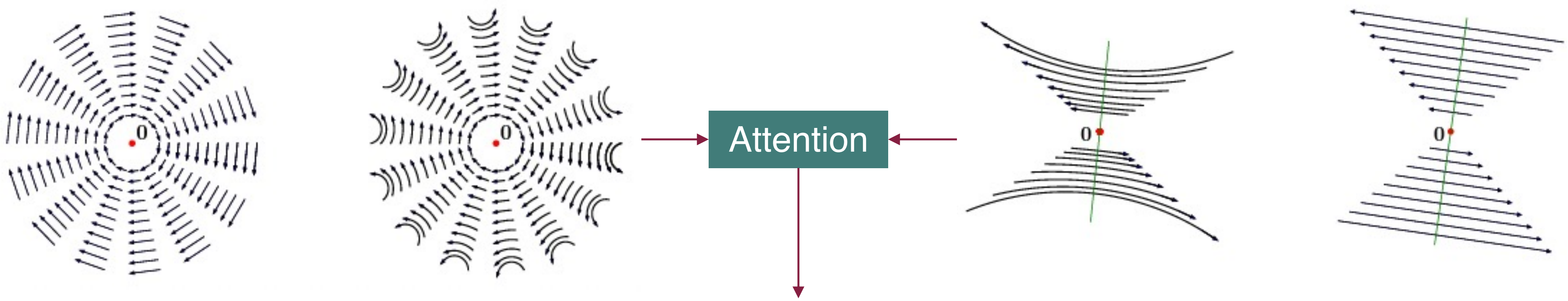
Knowledge Graphs have mixed topology, i.e., some relations are symmetric (*married*) while others are not (*director, features*).



Applications

Knowledge Graphs: Rotation, Reflection and Attention

Hence, the method proposes two sets of embeddings; rotational (RotH) and reflectional (RefH), which are finally combined with Attention (AttH).



Rotation
(Euclidean vs Hyperbolic)

AttH embeddings

Reflection
(Euclidean vs Hyperbolic)

Applications

RotH,RefH, AttH : Experimental Study

Evaluation Tasks:

Relational Link Prediction: Given a graph with attributed nodes and attributed edges, estimate a model that predicts the relevance of other entities as tail answers for an input head entity and relation.

Datasets:

Dataset	# Entity	# Relation	Description
FB15k-237	14,541	237	Collection of real world facts from Freebase network.
WN18RR	40,943	11	Hierarchical collection of relations between words.
YAGO3-10	123,182	37	Subset of YAGO3, a large semantic knowledge base, derived from Wikipedia, WordNet, WikiData, GeoNames, and other data sources.

Evaluation Metrics:

Relational Link Prediction : HITS@1,3,10 and Mean Reciprocal Rank (MRR).

Applications

RotH,RefH, AttH: Relational Link Prediction

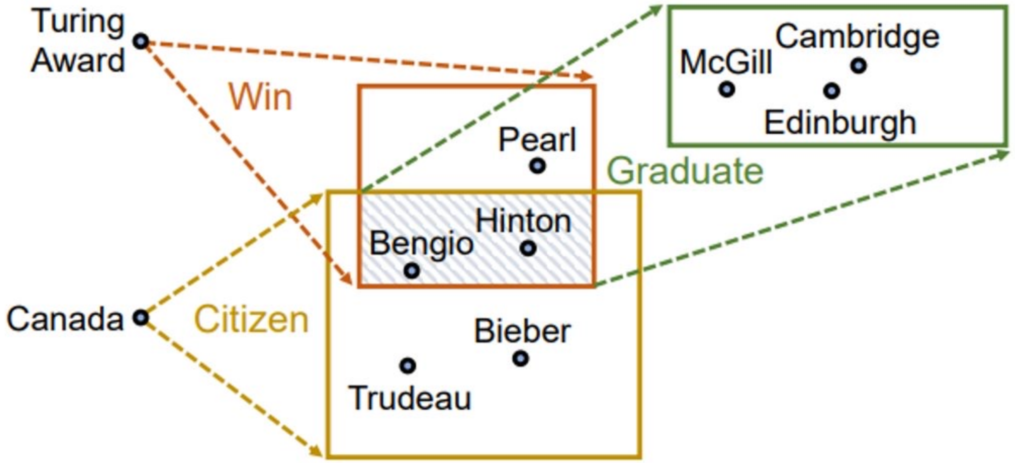
Relational Link Prediction													
	Dataset	WN18RR				FB15k-237				YAGO3-10			
Space	Models	MRR	H@10	H@3	H@1	MRR	H@10	H@3	H@1	MRR	H@10	H@3	H@1
Euclidean	RotatE	0.387	0.33	0.417	0.491	0.29	0.208	0.316	0.458	-	-	-	-
	MuRE	0.458	0.421	0.471	0.525	0.313	0.226	0.34	0.489	0.283	0.187	0.317	0.478
	RefE	0.42	0.39	0.42	0.46	0.294	0.211	0.322	0.463	0.336	0.259	0.367	0.484
	RotE	0.465	0.42	0.484	0.544	0.323	0.235	0.353	0.501	0.23	0.15	0.247	0.392
	AttE	0.455	0.419	0.47	0.521	0.302	0.216	0.33	0.474	0.37	0.289	0.403	0.527
Hyperbolic	MuRP	0.463	0.426	0.477	0.529	0.307	0.22	0.337	0.482	0.381	0.295	0.417	0.548
	RefH	0.456	0.419	0.471	0.526	0.311	0.223	0.339	0.488	0.374	0.29	0.41	0.537
	RotH	0.447	0.408	0.464	0.518	0.312	0.224	0.342	0.489	0.381	0.302	0.415	0.53
	AttH	0.472	0.428	0.49	0.553	0.314	0.223	0.346	0.497	0.393	0.307	0.435	0.559

Applications

Knowledge Graphs: Euclidean Methods

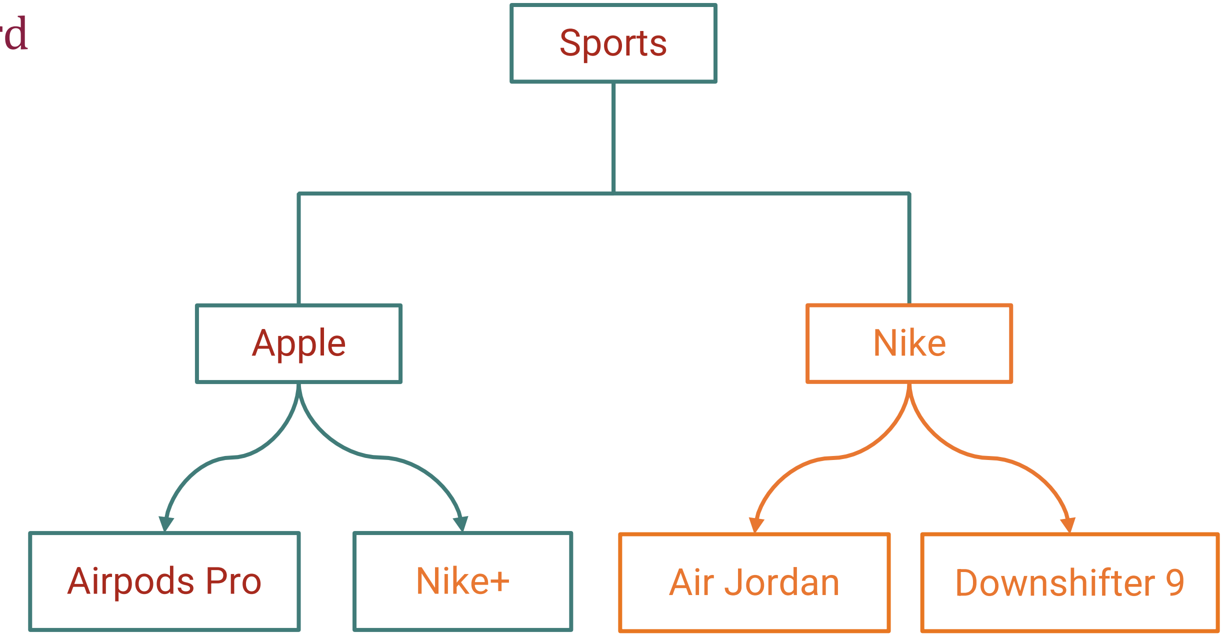
Spatial Representations are better for modelling Knowledge Graphs.

From which universities did the Canadian Turing Award winners graduate?



Box Embeddings in Euclidean Space

Knowledge Graphs possess inherent hierarchy.



Hierarchical Product Graph

Applications

Knowledge Graphs: Hyperboloid Embeddings

Capture spatial features of entities and relations.

Capture hierarchical features of entities and relations.

HypE

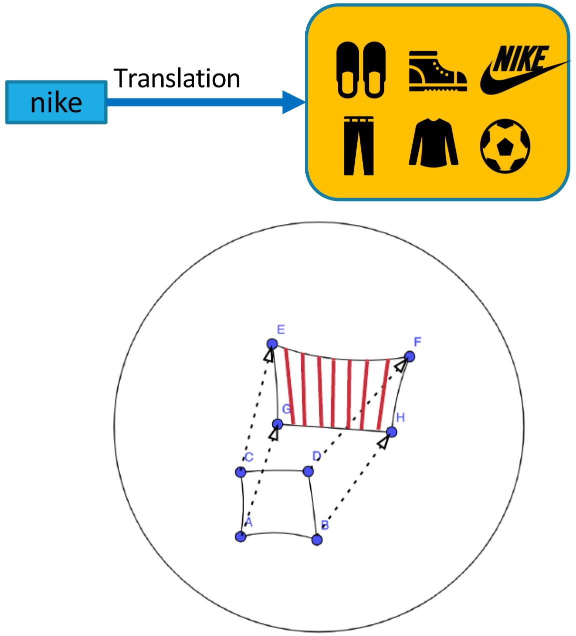
Handle different query types simultaneously

No labelled data

(self-supervised models)

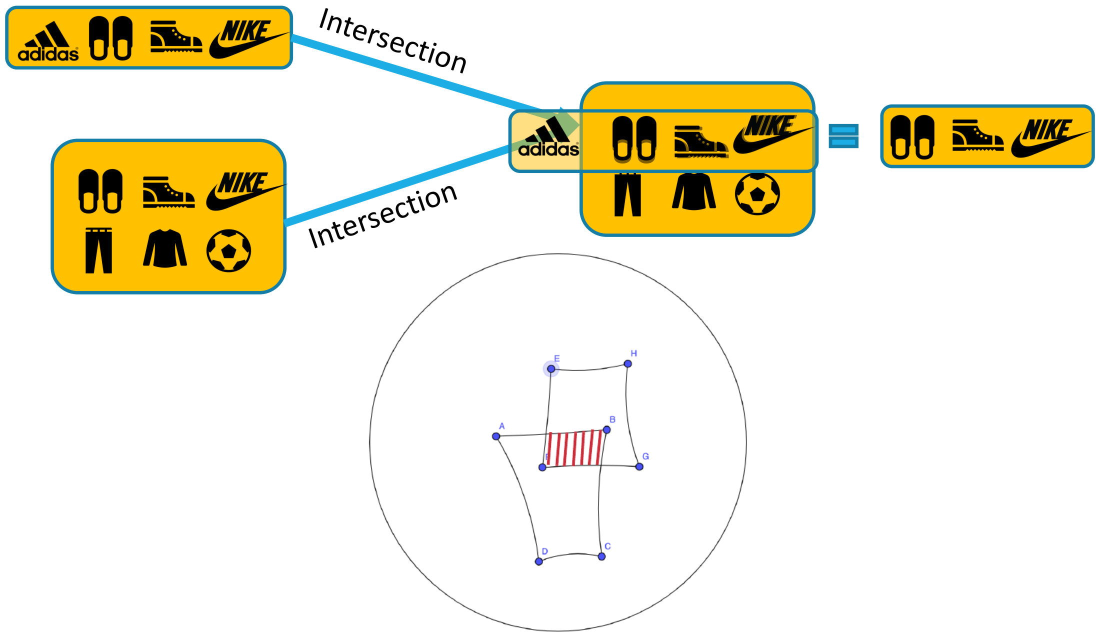
Applications
Knowledge Graphs: Reasoning Operations

Translation: Gives all children of a query
Q: nike



Hyperboloid Translation

Intersection: Gives intersection for two queries.
Q: nike footwear = nike ∩ footwear

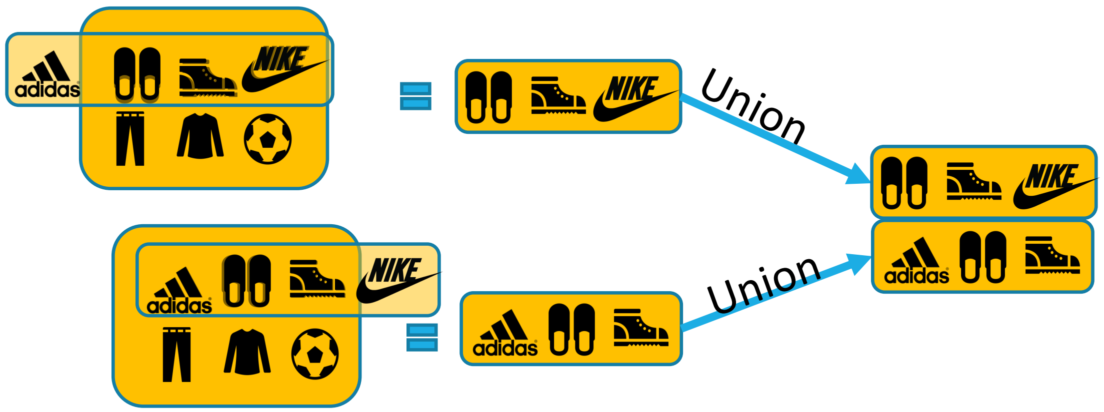


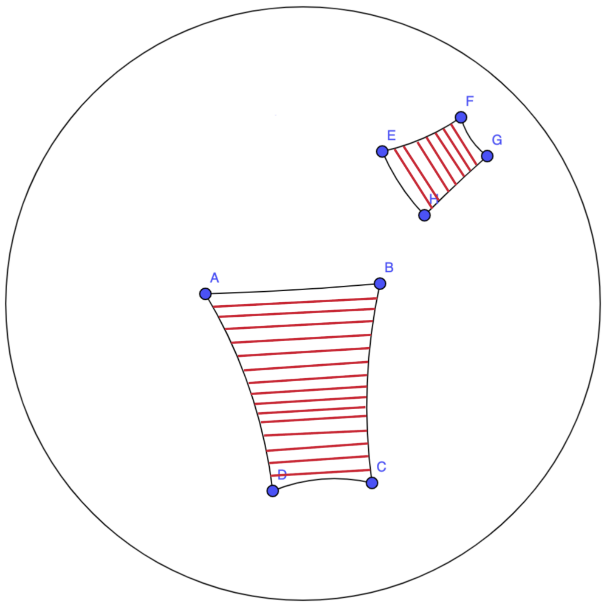
Hyperboloid Intersection

Applications
Knowledge Graphs: Hyperboloid Embeddings

Union Queries: Gives union of two queries.

Q: (nike ∪ adidas) ∩ footwear = (nike ∩ footwear) ∪ (adidas ∩ footwear)



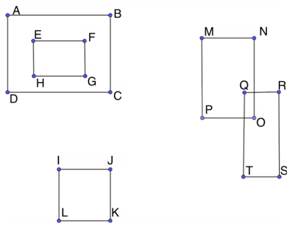


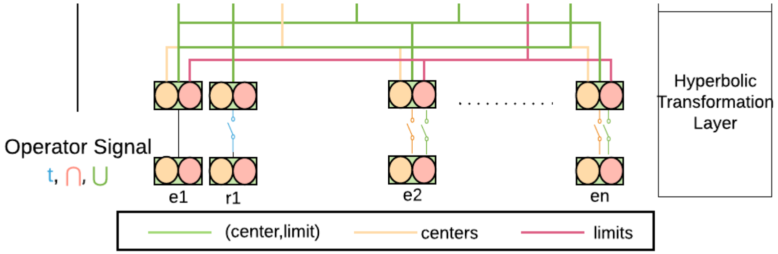
Hyperboloid Union

Applications

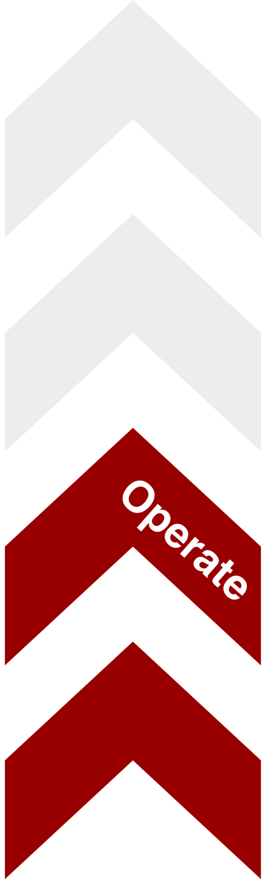
Knowledge Graphs: Hyperboloid Embeddings

Initialize entities and relations with Random Euclidean Boxes

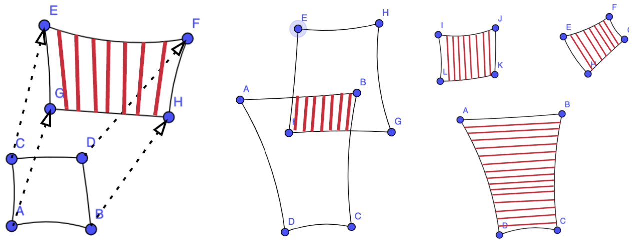


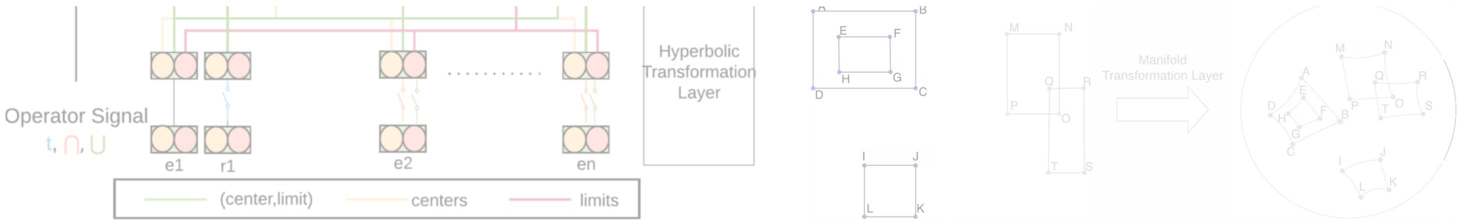


Applications
Knowledge Graphs: Hyperboloid Embeddings



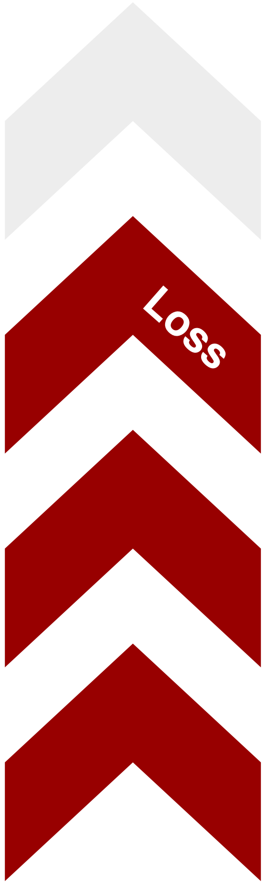
Operate according to signal



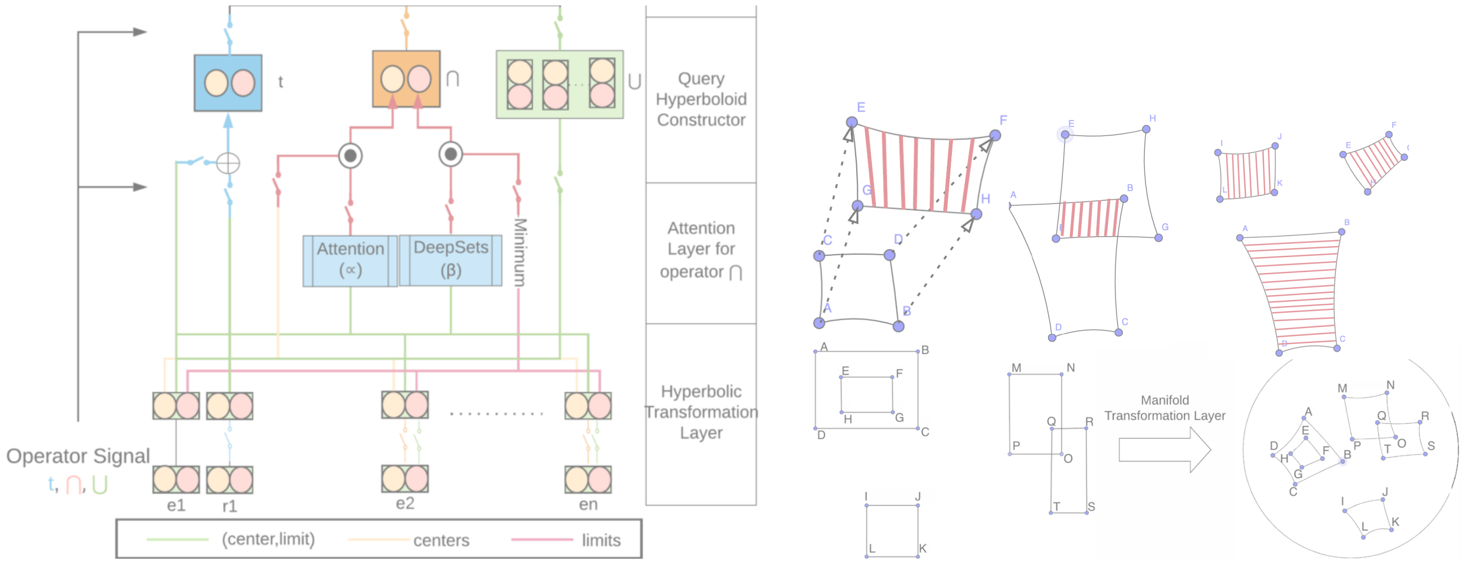


Applications

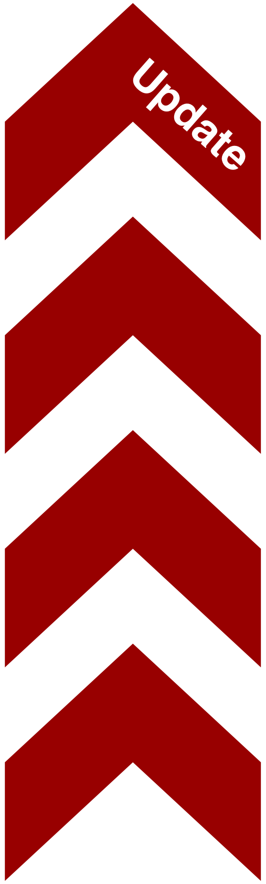
Knowledge Graphs: Hyperboloid Embeddings

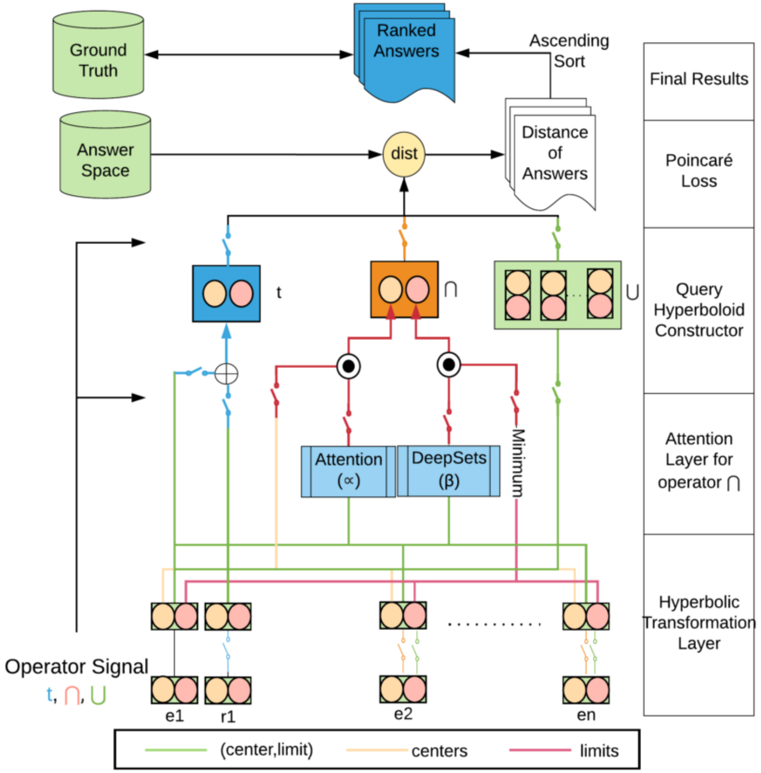


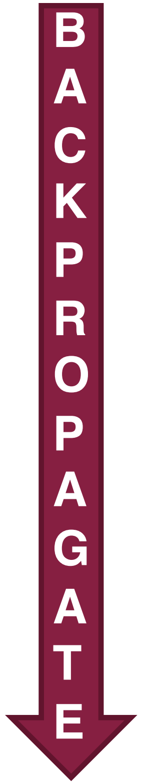
Calculate Loss based on distance of samples from the query space.



Applications
Knowledge Graphs: Hyperboloid Embeddings





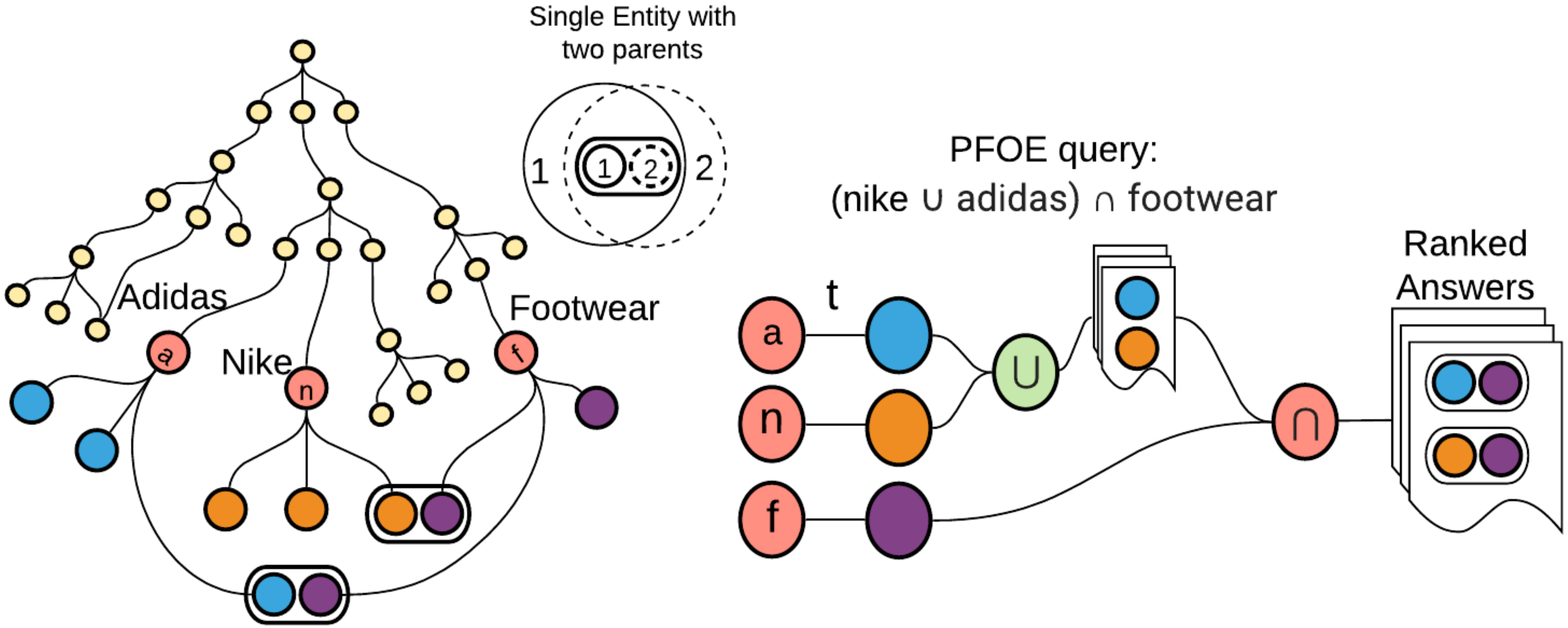


Update entity and relation representations based on the loss.

Applications

Knowledge Graphs: Self-supervised Learning

Generate pseudo-queries by using the structure/relations in the training graph.



Generation of Pseudo-query from a Product Graph

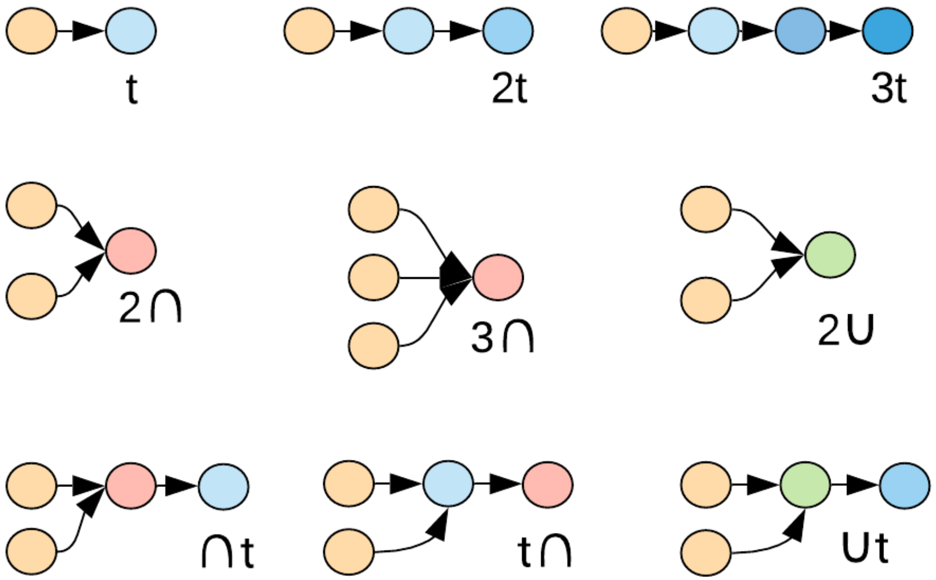
Applications

Knowledge Graphs: Evaluation

1. Efficacy of the Query Search Space
2. Anomaly Detection
3. Visualization

Applications

Knowledge Graphs: Efficacy of the Query Space



Translation (t):

1t: "nike", "shoes", "adidas"

2t: "women shoes" ("shoes" → "women")

3t: "furniture" ("furniture"→"chair","table","dining", etc→"ikea","wayfair", etc)

Intersection (∩):

2∩: "nike shoes" ("nike" AND "shoes")

3∩: "nike jordan laces" ("nike" AND "jordan" AND "laces")

∩t: "nike shoes" ("nike" AND "shoes" →products in the space)

t∩: "furniture ikea" ("furniture"→"chair","table","dining", etc AND "ikea")

Union (∪):

2∪: "nike and adidas" ("nike" OR "adidas")

∪t: "nike and adidas shoes" ("nike" OR "adidas" →products in the space)

Applications

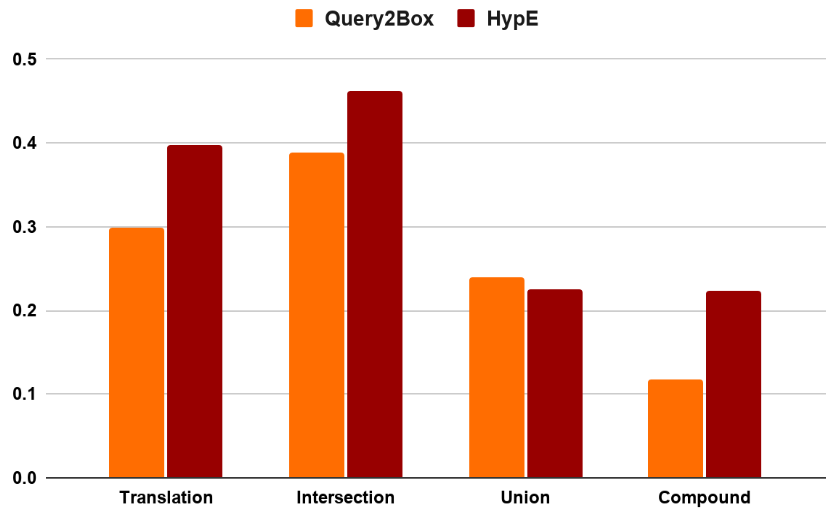
Knowledge Graphs: Efficacy of the Query Space

- Logical Query Reasoning
  - Dataset: FB15K, FB15K-237, NELL995, DBPedia, E-commerce Product Graph
  - Primary Baseline: Query2Box (ICLR 2020)
  - Evaluation Metrics: HITS@3, Mean Reciprocal Rank

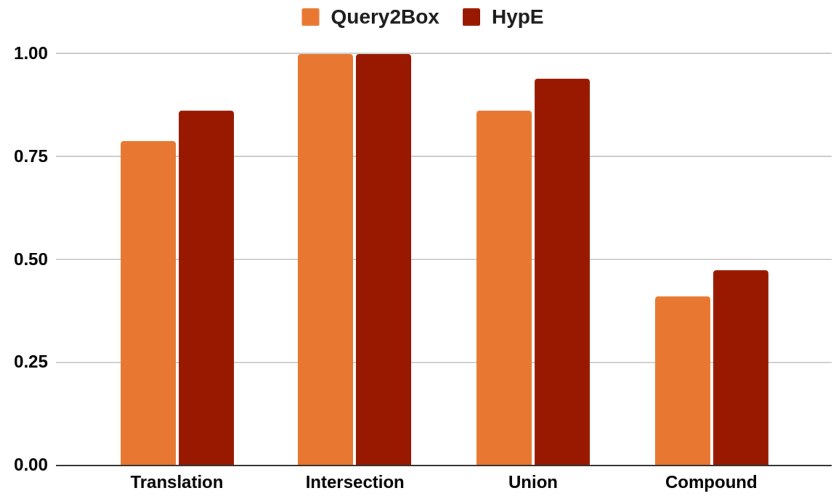
Applications

Knowledge Graphs: Efficacy of the Query Space

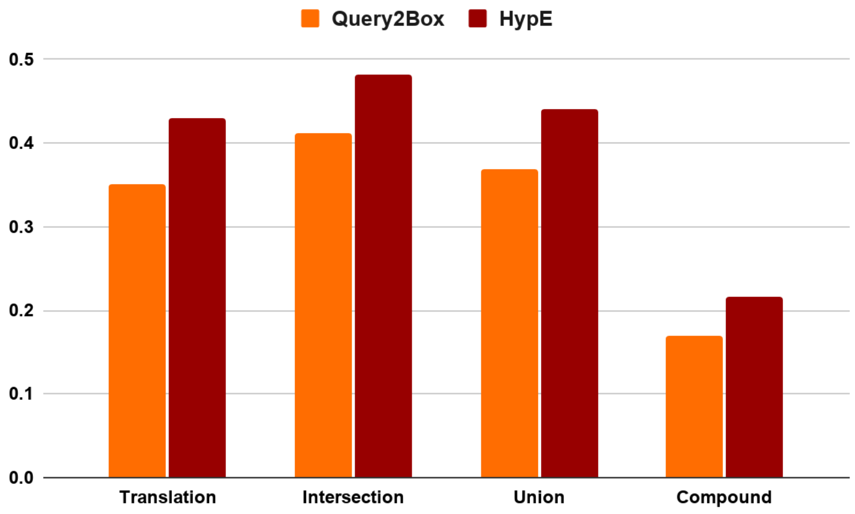
FB15K-237



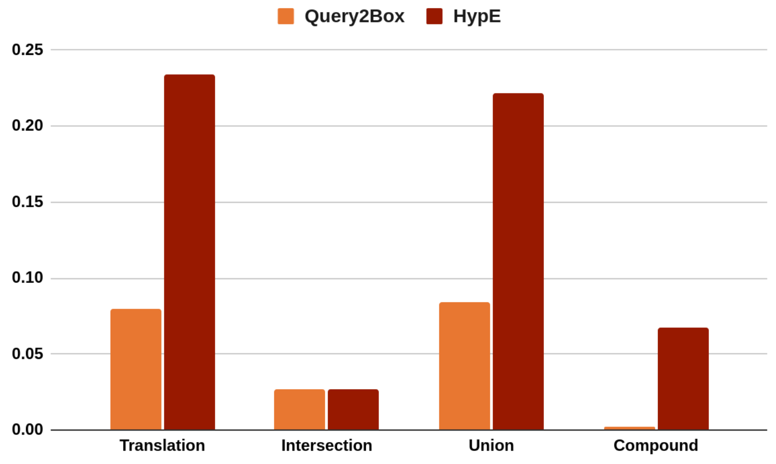
DBPedia



NELL995

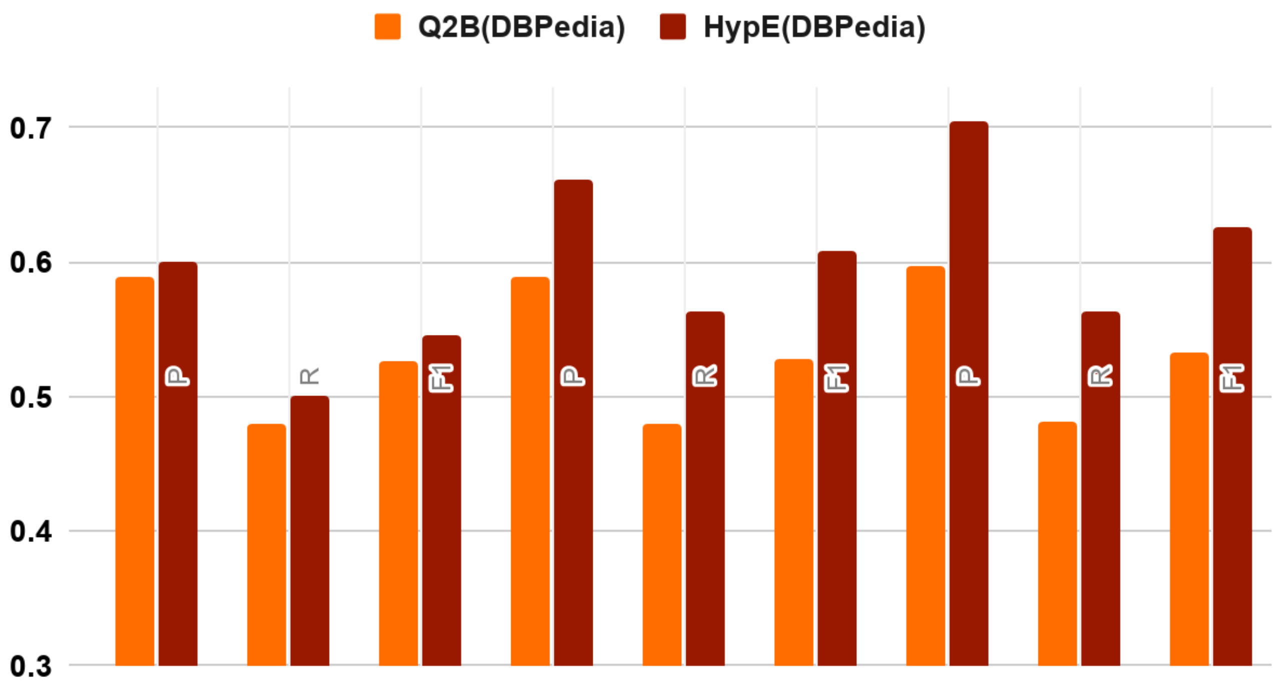


E-commerce (Relative)



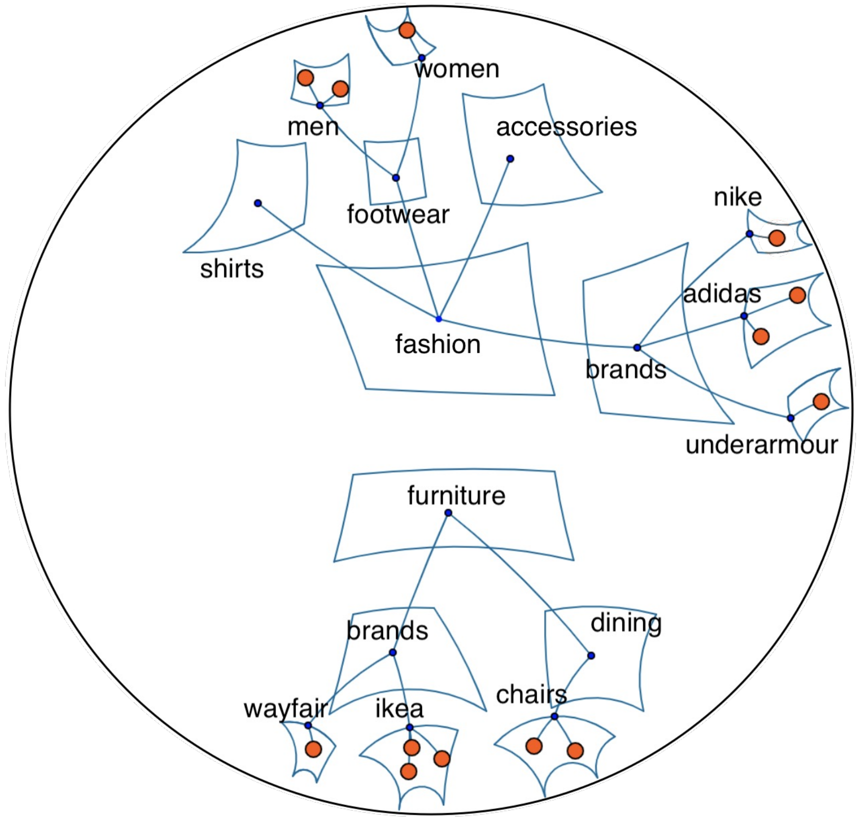
Applications

Knowledge Graphs: Anomaly Detection (Downstream task)

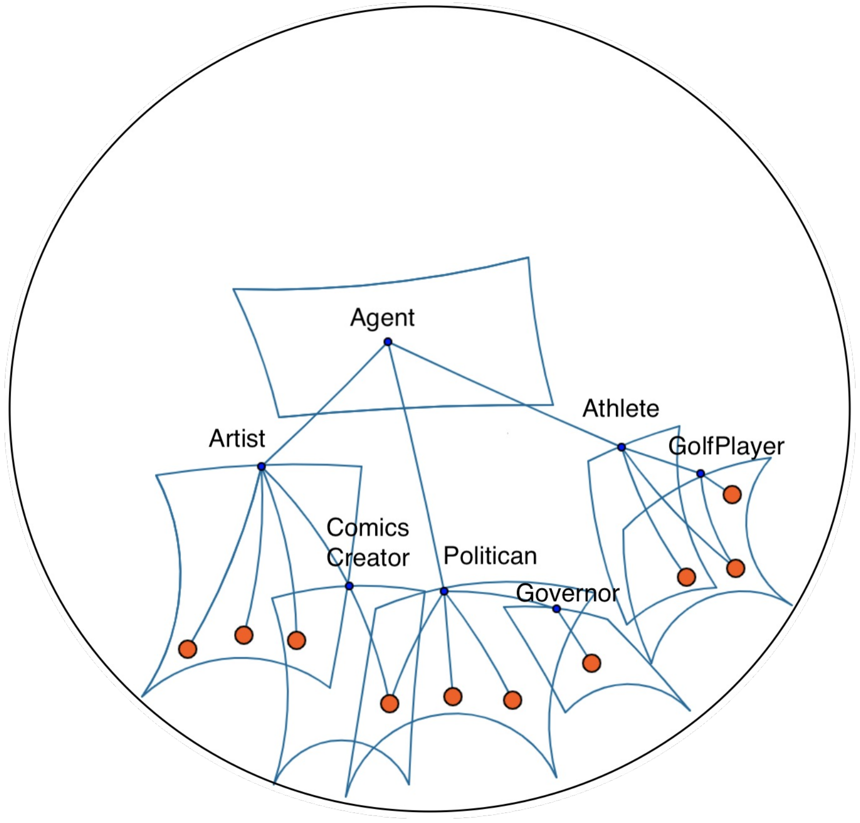


- Miscategorized Nodes detection
- Evaluation Metrics: Precision, Recall, F1-score
- Different levels of parents considered: Parent, Grandparent and Great-grandparent to the child nodes.

Applications
Knowledge Graphs: Visualization



Hyperboloids of the E-commerce Graph



Hyperboloids of DBPedia

More results can be found in the paper: Choudhary, N., Rao, N., Katariya, S., Subbian, K., & Reddy, C. K. (2021, April). Self-Supervised Hyperboloid Representations from Logical Queries over Knowledge Graphs. (WWW 2021)

Applications

Knowledge Graphs: Learnings

- Hyperbolic space is better at simultaneously capturing spatial and hierarchical structure information by pseudo-querying the knowledge graphs.

- The ablation study shows the clear importance of using relatively complex queries such as intersection and union in enhancing HypE's performance.

- HypE's representation, in congruence with/without additional information such as semantics, can also be used for downstream tasks (anomaly detection).

- The hyperboloids can also be visualized in a Poincaré ball for better human comprehension.

Applications

Extending Logical Reasoning to Product Search

Challenges:

- Search queries are presented in Natural Language and not logical queries.
- Reasoning needs logical steps.
- The products lie in a hierarchical catalogue.

Can we learn the logical reasoning steps from the natural language queries to retrieve products from a hierarchical catalogue?

Applications

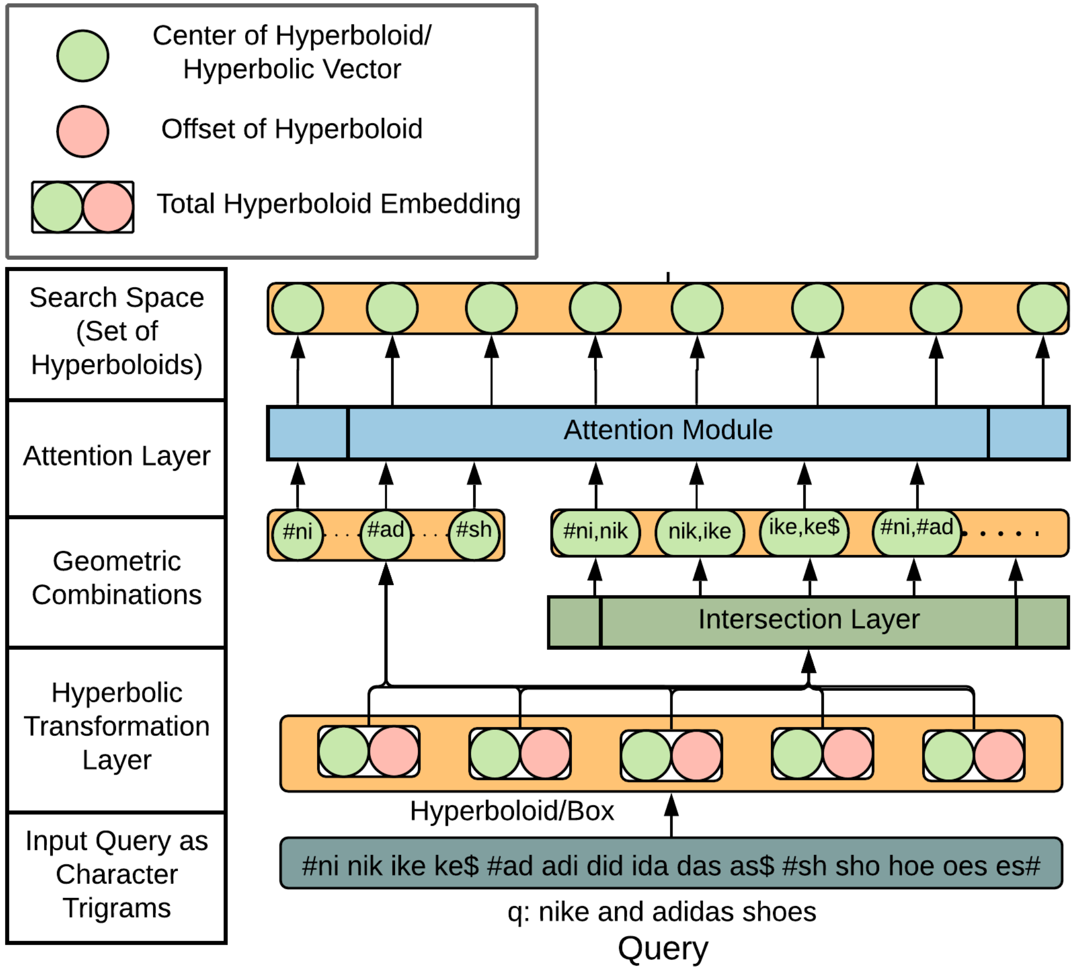
Product Search: ANTHEM

ANTHEM

Learn query representations in the Euclidean space.	Learn query operations through Attentive aggregation.
Encode hierarchically catalogued products in hyperbolic space.	Use a joint-learning architecture to match queries and products.

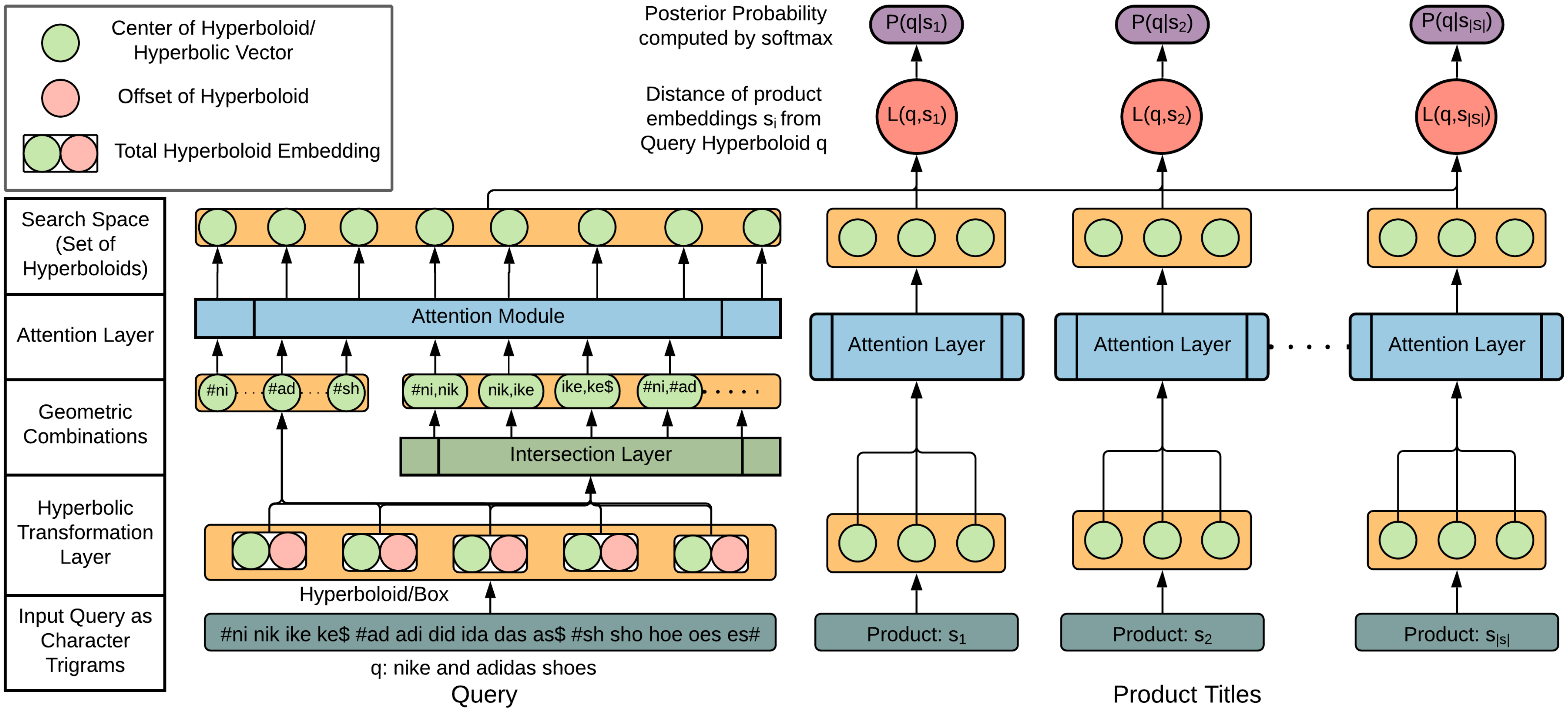
Applications

Product Search: ANTHEM

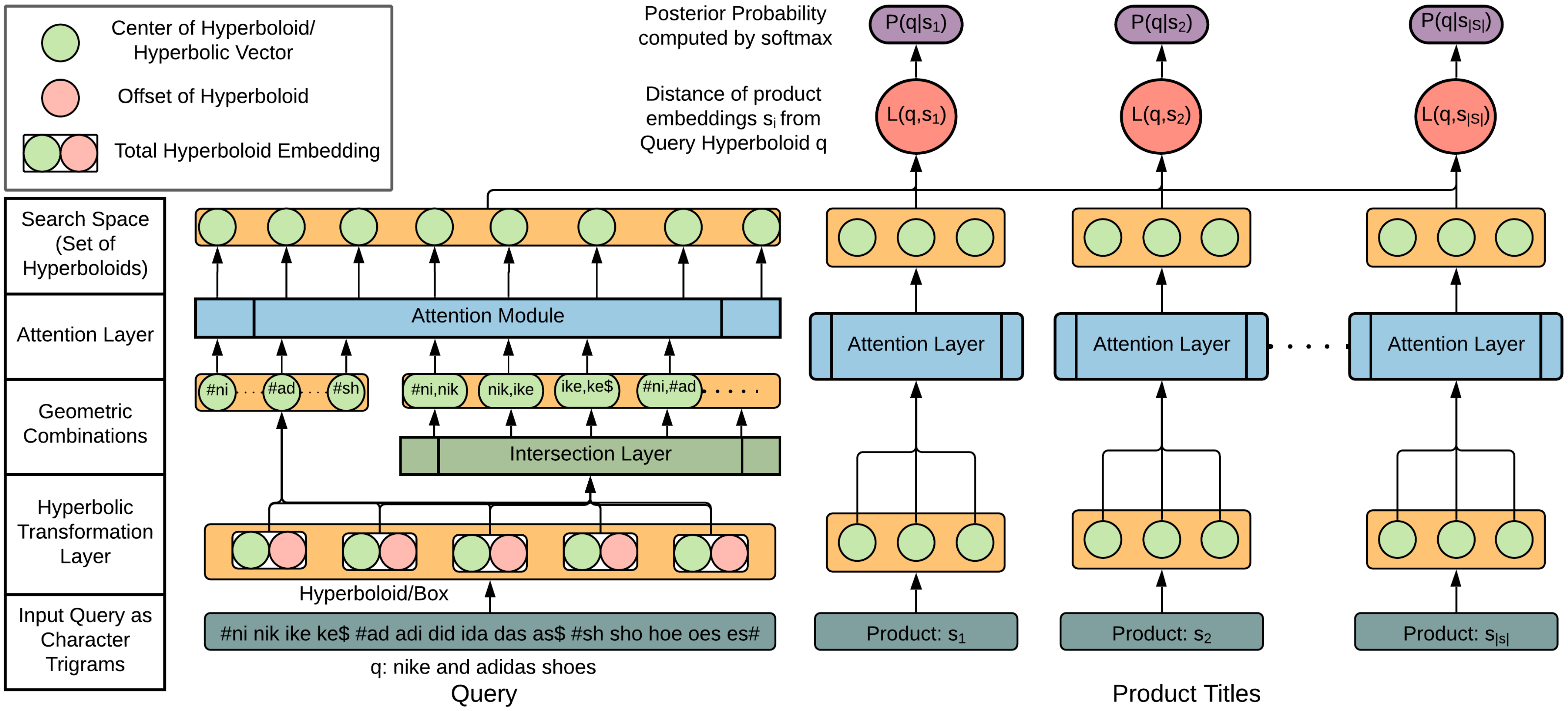


Applications

Product Search: ANTHEM



Applications
Product Search: ANTHEM



Applications

Product Search: Evaluation

1. Performance on Product Search
2. Performance on Query Matching
3. Explainability Study

Applications

Product Search: Performance on Product Search

- Dataset for Product Search: E-commerce (Amazon) Product Search, Public E-commerce Search Relevance
- Dataset for Query Matching: E-commerce (Amazon) ESCI Query Matching
- Baselines: ARC-II, KNRM, DUET, DRMM, aNMM, MatchPyramid, C-DSSM, MV-LSTM, BERT.
- Evaluation Metrics: NDCG@3, NDCG@5, NDCG@10, Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR)

Applications

Product Search: Performance on Product Search

Datasets	E-commerce Product Search (in %)					Public E-commerce Search Relevance (in %)				
Models	NDCG@3	NDCG@5	NDCG@10	MAP	MRR	NDCG@3	NDCG@5	NDCG@10	MAP	MRR
ARC-II	0	0	0	0	0	59.2	58.1	54.4	58.2	48.5
KNRM	12.5	12.8	15.0	12.8	16.7	66.6	65.5	62.6	65.6	56.6
Duet	13.1	13.3	15.4	13.4	14.7	66.9	65.8	62.8	66.0	55.6
DRMM	20.5	22.8	24.4	20.3	21.7	71.3	71.3	67.7	70.0	59.0
aNMM	21.0	23.9	26.8	20.3	22.9	71.6	72.0	69.0	70.0	59.6
MatchPyramid	25.6	25.6	26.8	25.0	34.7	74.3	72.9	69.0	72.8	65.3
C-DSSM	31.3	29.4	44.7	32.6	27.8	77.7	75.2	78.7	77.1	61.9
MV-LSTM	34.7	33.3	55.7	34.3	37.7	79.7	77.5	84.7	78.2	66.8
BERT	38.6	37.2	65.9	40.1	51.0	82.1	79.7	90.2	81.5	73.2
E-ANTHEM	49.4	46.7	66.7	51.2	62.9	88.5	85.2	90.7	88.0	79.0
ANTHEM	**51.1**	**48.9**	**80.9**	**53.5**	**65.4**	**89.5**	**86.5**	**98.4**	**89.3**	**80.2**

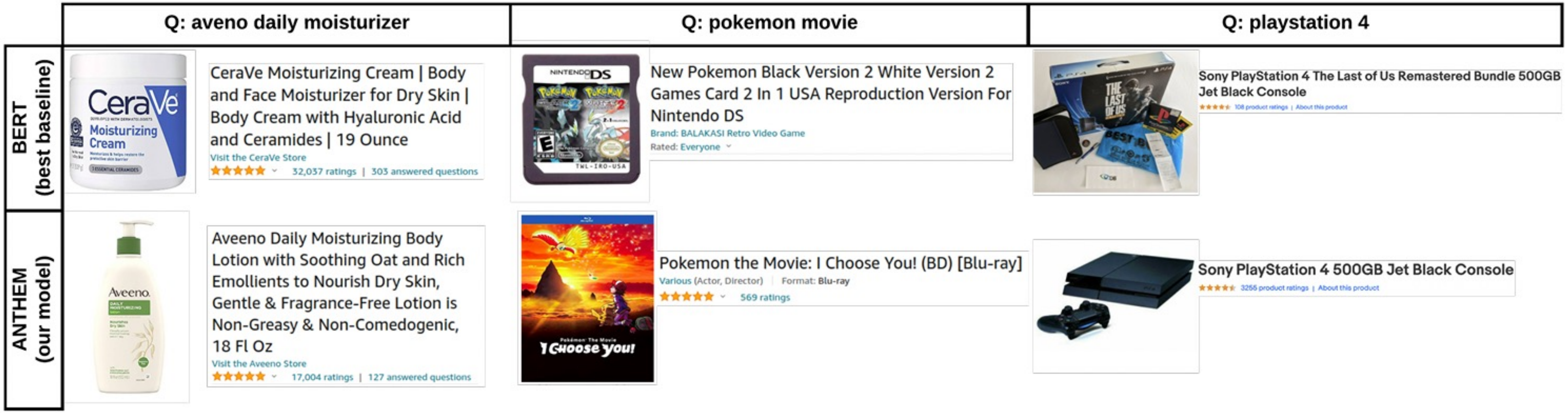
Applications

Product Search: Performance on Query Matching

Models	Accuracy (in %)	F-score (in %)	AUC (in %)
ARC-II	0.0	0.0	0.0
KNRM	-4.1	-24.9	-19.1
Duet	-2.3	-4.7	0.9
DRMM	25.1	15.4	33.1
aNMM	-1.3	-8.6	4.0
MatchPyramid	-14.5	-17.8	-9.7
C-DSSM	21.2	21.7	30.1
MV-LSTM	41.1	21.2	48.9
BERT	40.1	33.5	54.7
E-ANTHEM	43.2	40.3	61.4
ANTHEM	**43.9**	**40.8**	**62.6**

Applications

Product Search: Explainability Study



ANTHEM detects misspellings due to the use of char trigrams.

ANTHEM gives less false positives due to hierarchical information.

ANTHEM results have higher diversity due to information gain from different product hierarchies.

More results can be found in the paper: Choudhary, N., Rao, N., Katariya, S., Subbian, K., & Reddy, C. K. (2022, February). ANTHEM: Attentive Hyperbolic Entity Model for Product Search. (WSDM 2022)

Applications

Product Search: Learnings

- ➔ Hyperbolic space is better at simultaneously capturing spatial and hierarchical structure information by pseudo-querying the knowledge graphs.
- ➔ The ablation study shows the clear importance of using attention mechanism to capture intersection and union operations in product search queries.
- ➔ HypE's product representation, in congruence with matching architecture can be utilized for downstream tasks (product search and query matching).
- ➔ The attention weights in ANTHEM can also be visualized for better human comprehension.

Applications

Natural Language Processing: Text Classification

- Natural Language Inference: Given two sentences, a premise (e.g. "Little kids A. and B. are playing soccer.") and a hypothesis (e.g. "Two children are playing outdoors."), the binary classification task is to predict whether the second sentence can be inferred from the first one.
  - Standard SNLI dataset; 570K training, 10K validation and 10K test sentence pairs.
- PREFIX: Given two sentences, model has to decide if the second sentence is a noisy prefix of the first, or a random sentence.
  - PREFIX-Z% (Z=10, 30 or 50): For each random first sentence and one random prefix of it, a second positive sentence is generated by randomly replacing Z% of the words of the prefix, and a second negative sentence of same length is randomly generated. 500K training, 10K validation and 10K test pairs.

Applications

Text Classification: Evaluation

Metric: Accuracy	SNLI	PREFIX-10%	PREFIX-30%	PREFIX-50%
FULLY EUCLIDEAN RNN	79.34	89.62	81.71	72.10
HYPERBOLIC RNN+FFNN, EUCL. MLR	79.18	96.36	87.83	76.50
FULLY HYPERBOLIC RNN	78.21	96.91	87.25	62.94
FULLY EUCLIDEAN GRU	**81.52**	95.96	86.47	75.04
HYPERBOLIC GRU+FFNN, EUCL. MLR	79.76	**97.36**	**88.47**	**76.87**
FULLY HYPERBOLIC GRU	81.19	97.14	88.26	76.44

Applications
Text Classification: Evaluation

- Hyperbolic architectures, on average, perform better than the Euclidean variants.
- The improvement is better on hierarchical task; PREFIX
- No apparent benefit of using Hyperbolic MLR.

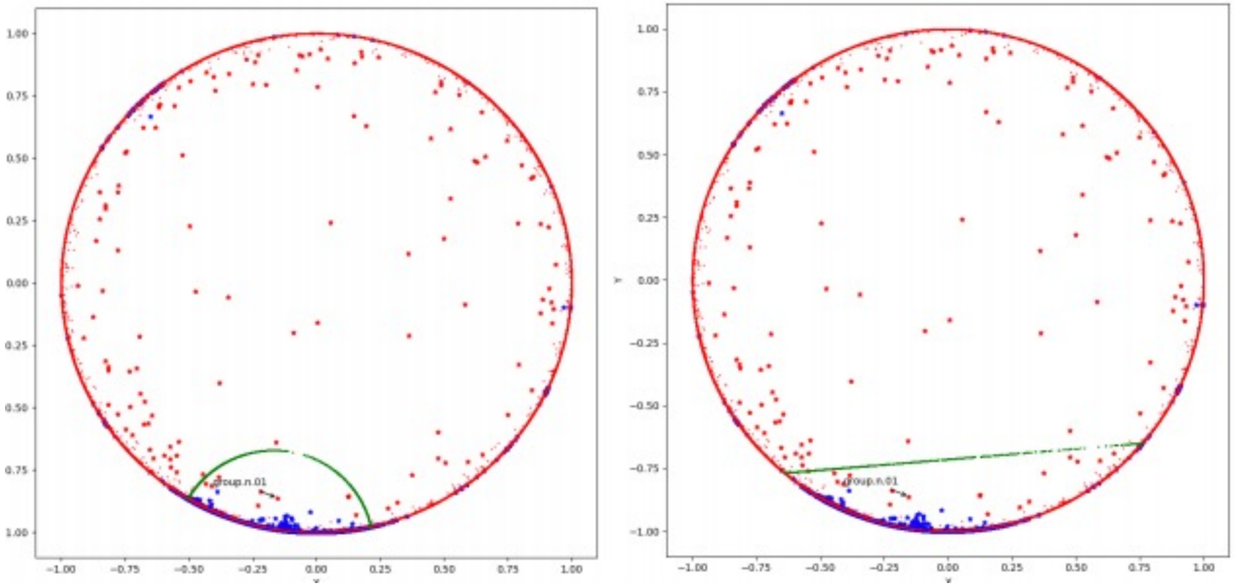


Image Source: Ganea, Octavian, Gary Bécigneul, and Thomas Hofmann. "Hyperbolic neural networks." Advances in neural information processing systems 31 (2018).

Q&A

Please provide your feedback at:

https://bit.ly/feedback-hyperbolic-tutorial