

Deep Learning on Graphs for Natural Language Processing

Lingfei Wu, Yu Chen, Heng Ji,
Yunyao Li, and Bang Liu

AAAI-2022 Tutorial
February 23th, 2022



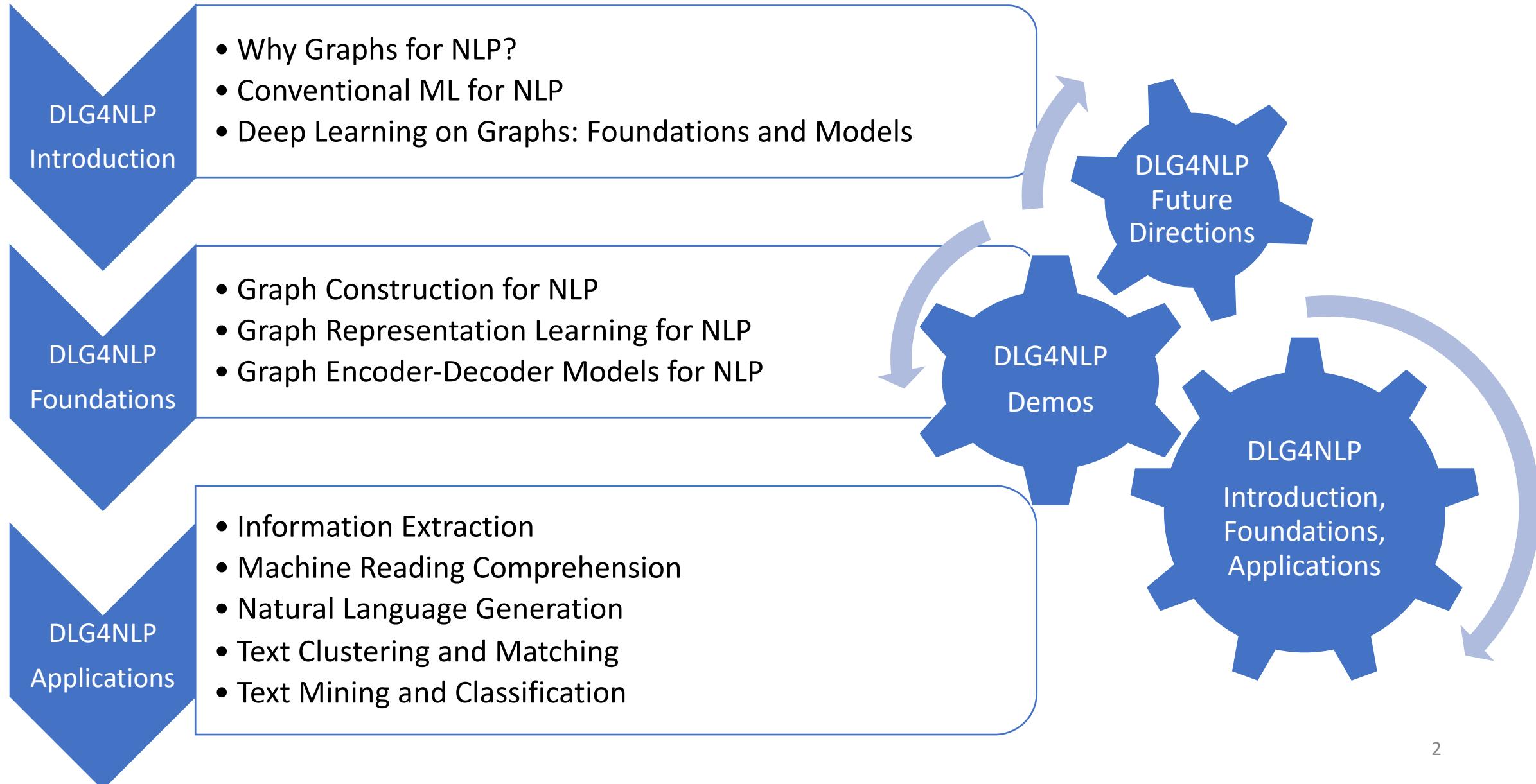
JD.COM  Meta



Université de Montréal 



Outline

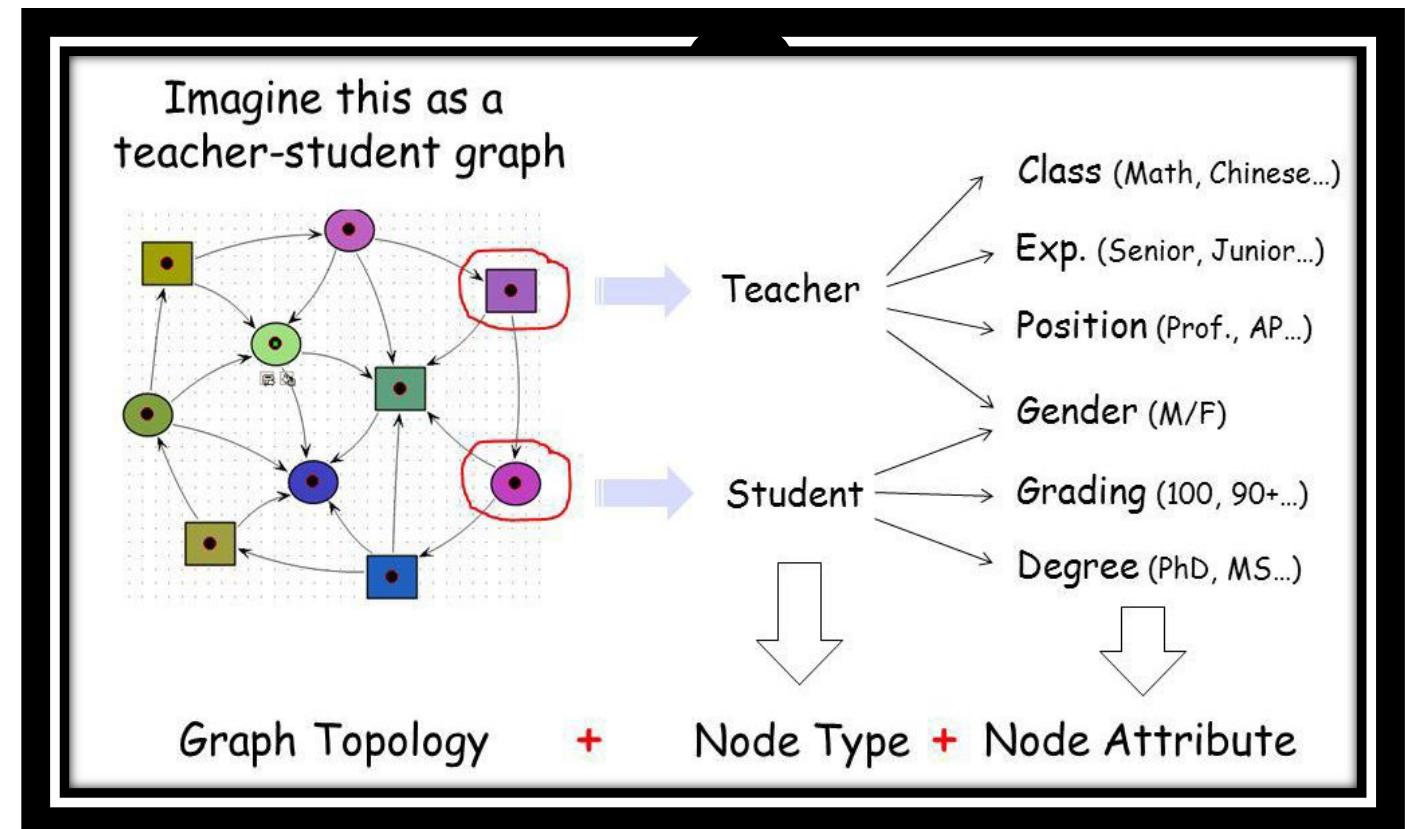


DLG4NLP

Introduction

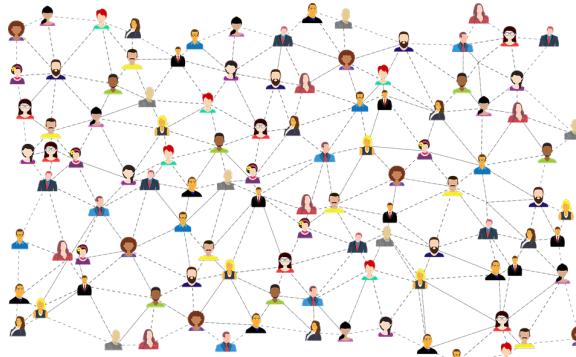
Why graphs?

- Graphs are a general language for describing and modeling complex systems



Graph!

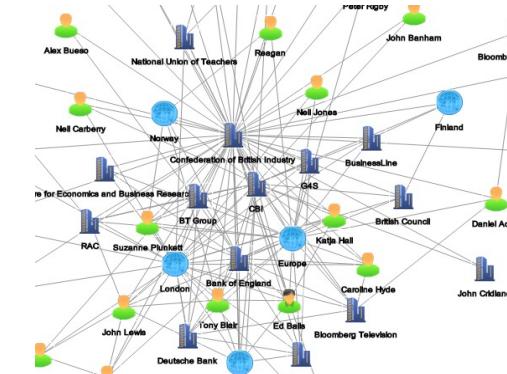
Graph-structured data are ubiquitous



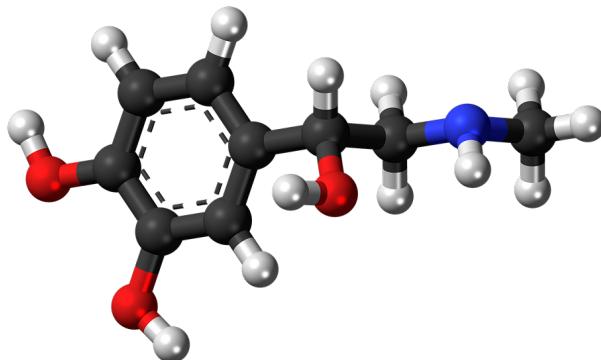
Social networks



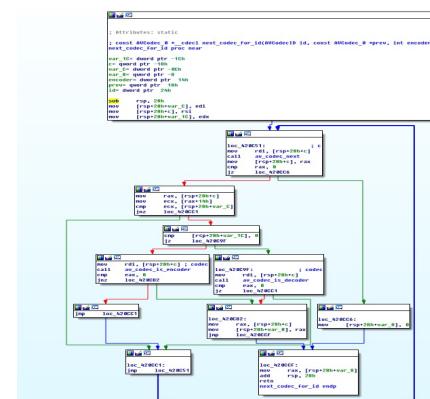
Information retrieval



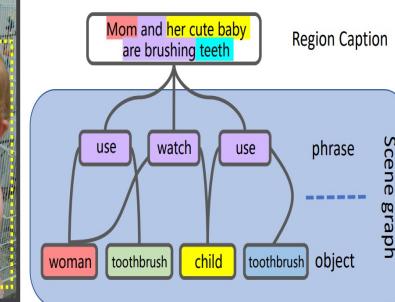
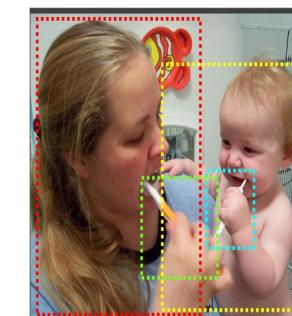
Financial transactions



Biomedical graphs

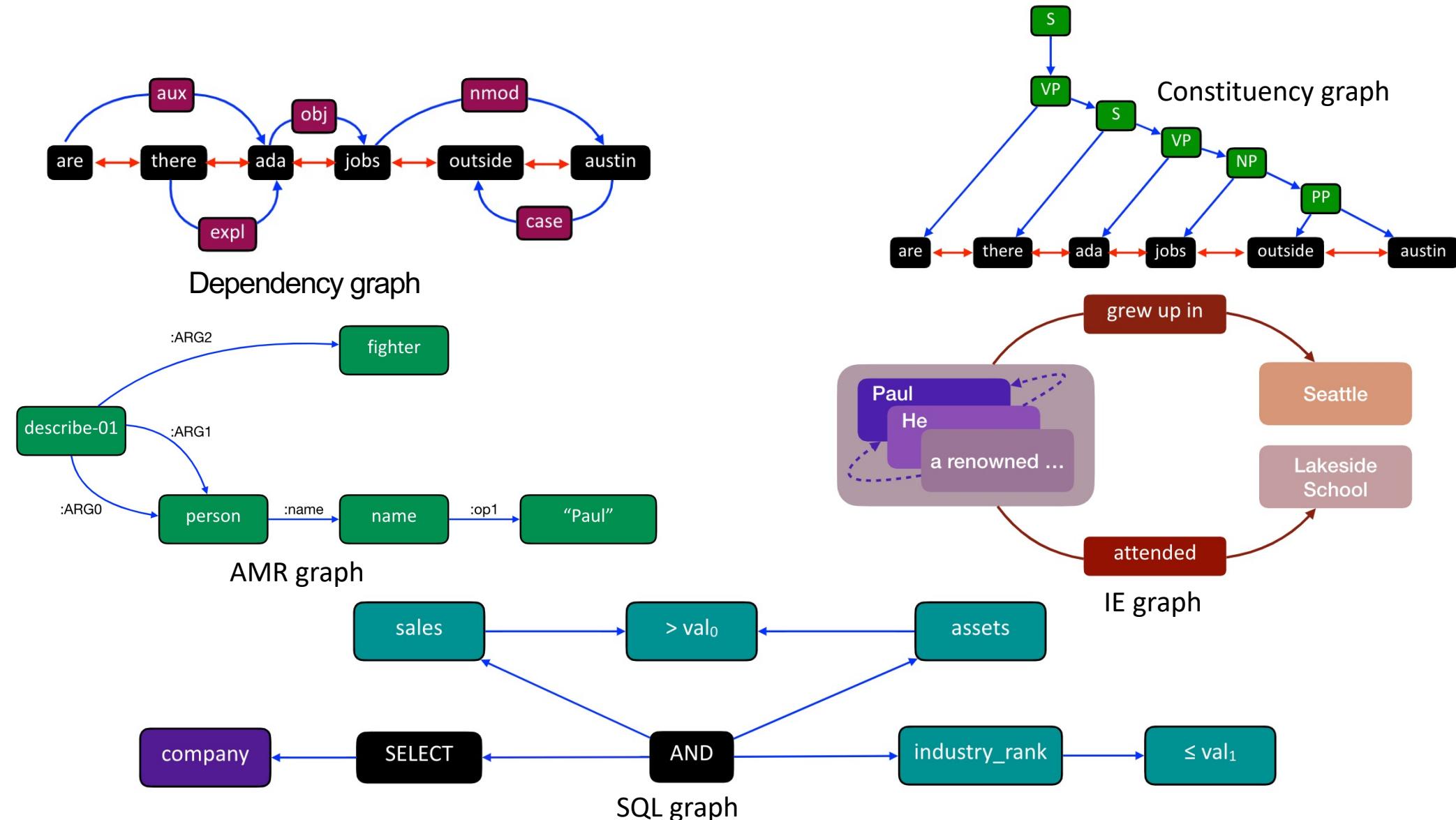


Program graphs



Scene graphs

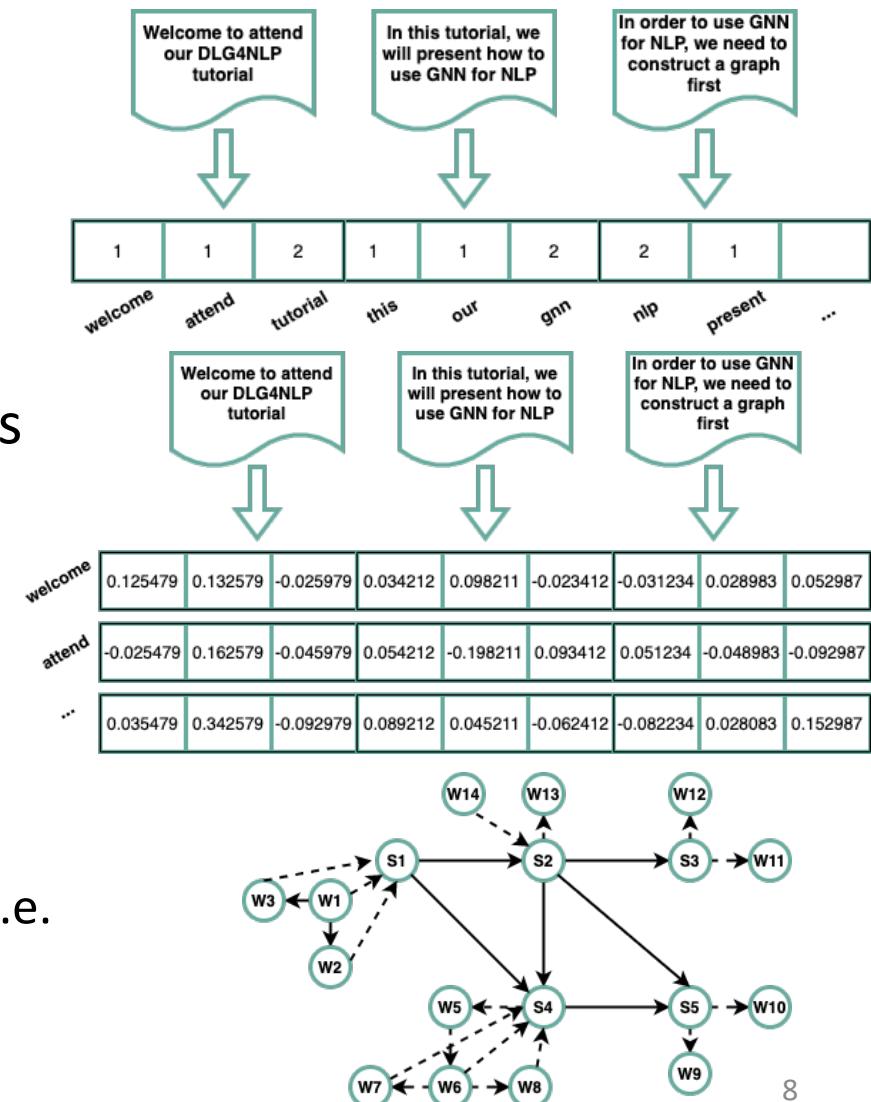
Graphs are ubiquitous in NLP As Well



Machine Learning on Graphs for NLP

Natural Language Processing: A Graph Perspective

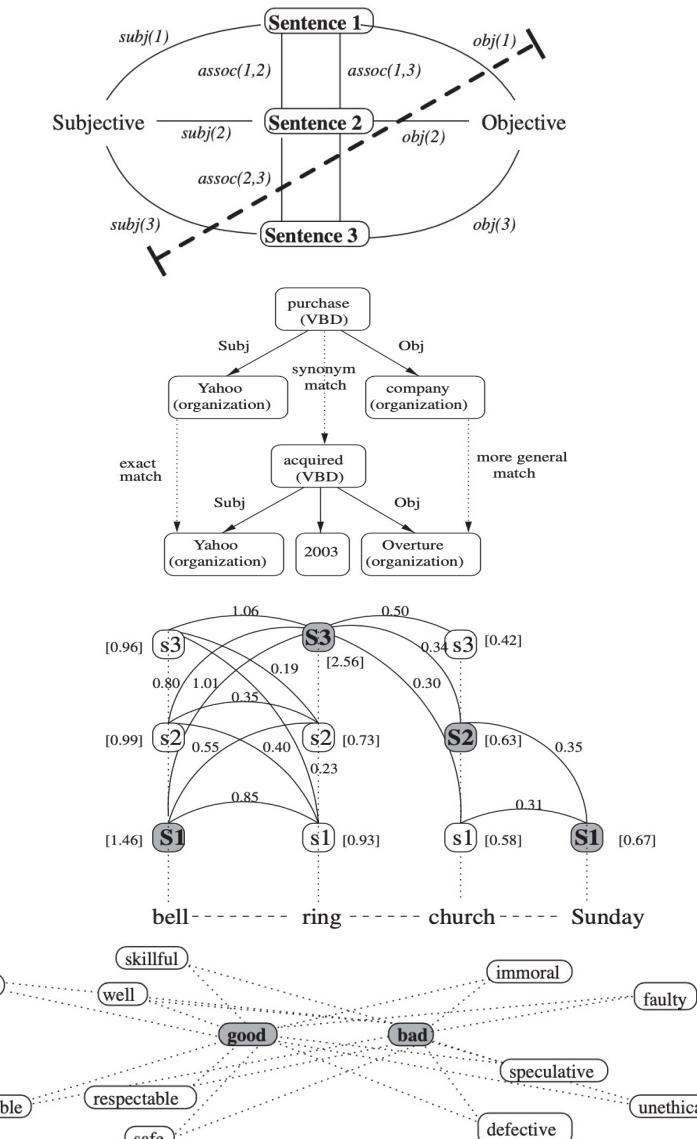
- Represent natural language as a bag of tokens
 - BOW, TF-IDF
 - Topic Modeling: text as a mixture of topics
- Represent natural language as a sequence of tokens
 - Linear-chain CRF
 - Word2vec, Glove
- **Represent natural language as a graph**
 - Dependency graphs, constituency graphs, AMR graphs, IE graphs, and knowledge graphs
 - Text graph containing multiple hierarchies of elements, i.e. document, sentence and word



Graph Based Methods for NLP

- Random Walk Algorithms
 - Generate random paths, one can obtain a stationary distribution over all the nodes in a graph
 - Applications: semantic similarity of texts, name disambiguation
- Graph Clustering Algorithms
 - Spectral clustering, random walk clustering and min-cut clustering for text clustering
- Graph Matching Algorithms
 - Compute the similarity between two graphs for textual entailment task
- Label Propagation Algorithms
 - Propagate labels from labeled data points to previously unlabeled data points
 - Applications: word-sense disambiguation, sentiment analysis

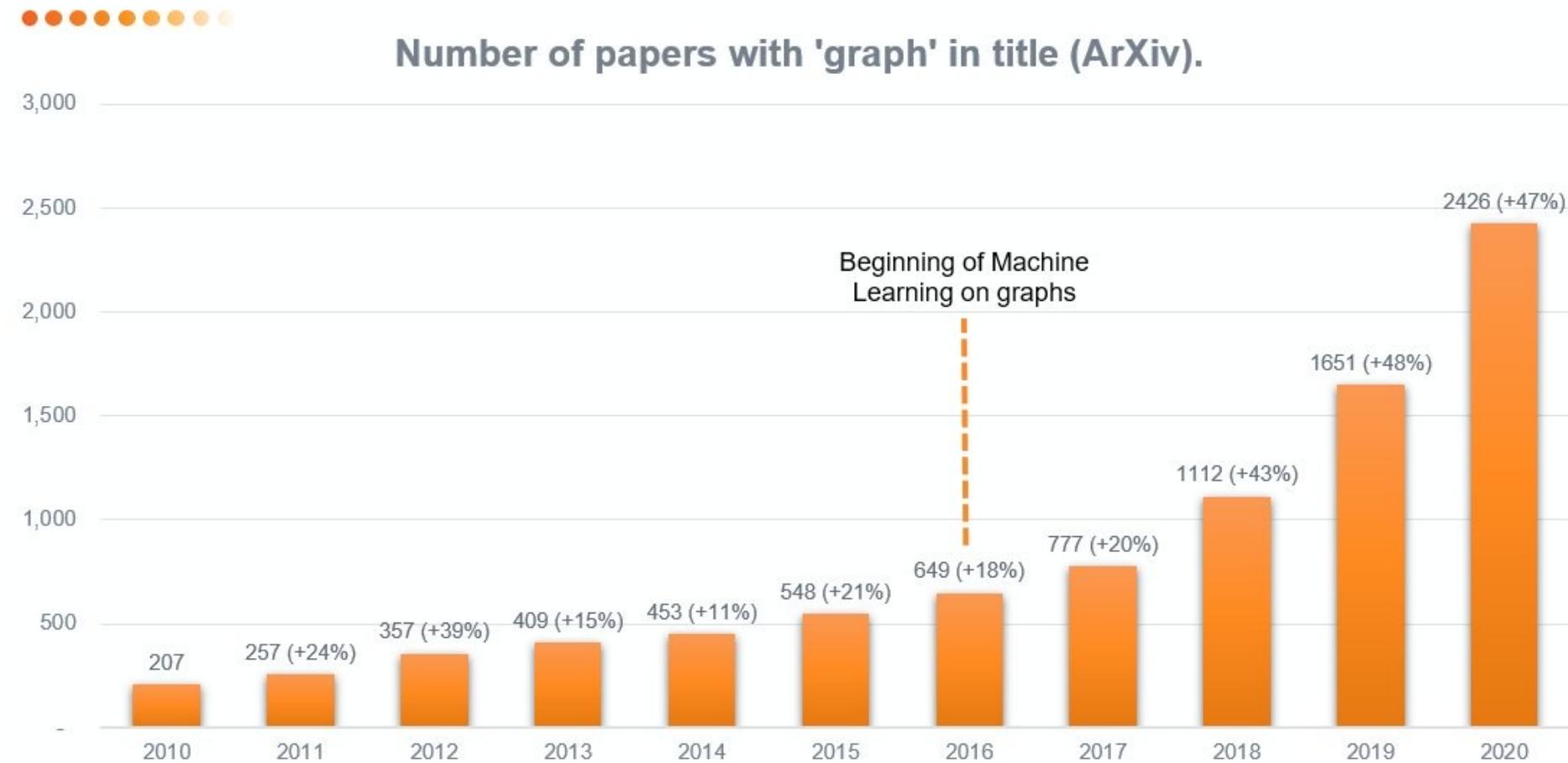
[Mihalcea and Radev, 2011]



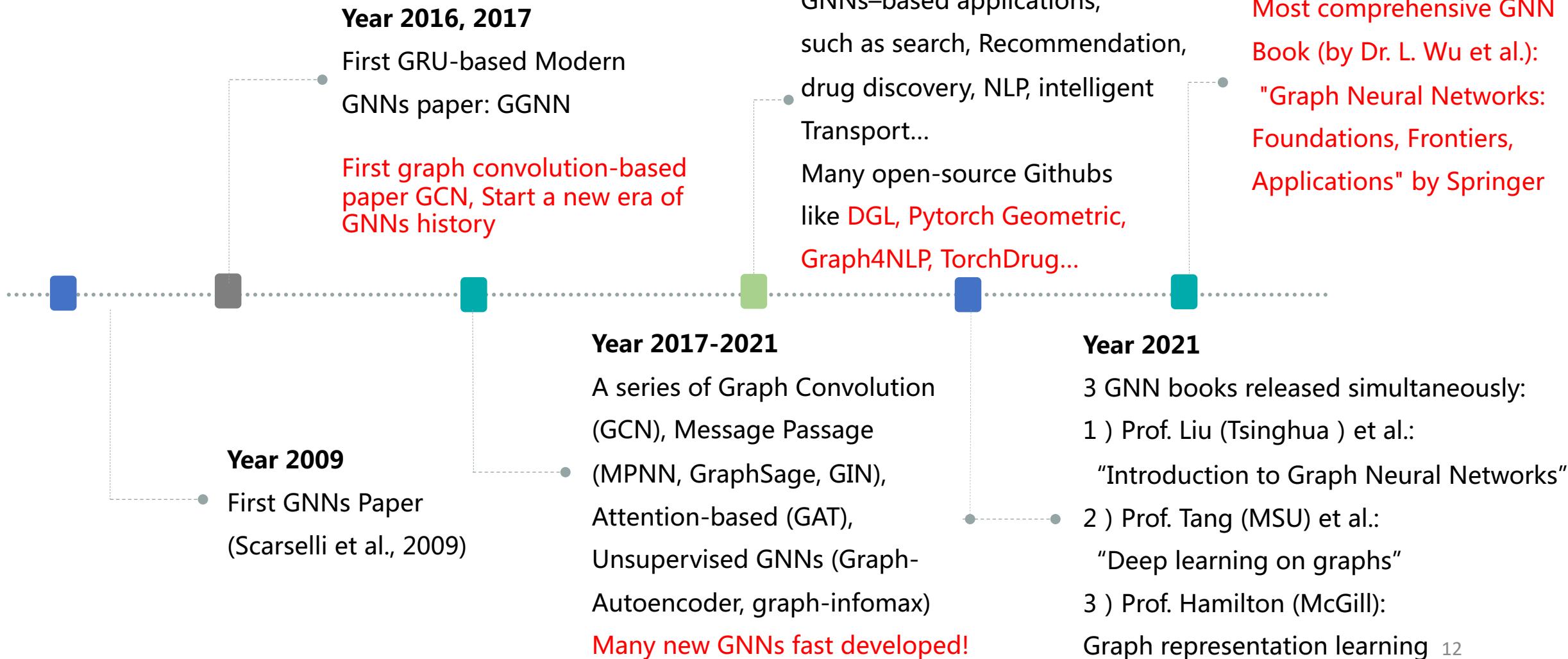
Deep Learning on Graphs: Foundations and Models

Graph Machine Learning: Recent Trending

Graph Machine Learning is on fire 🔥

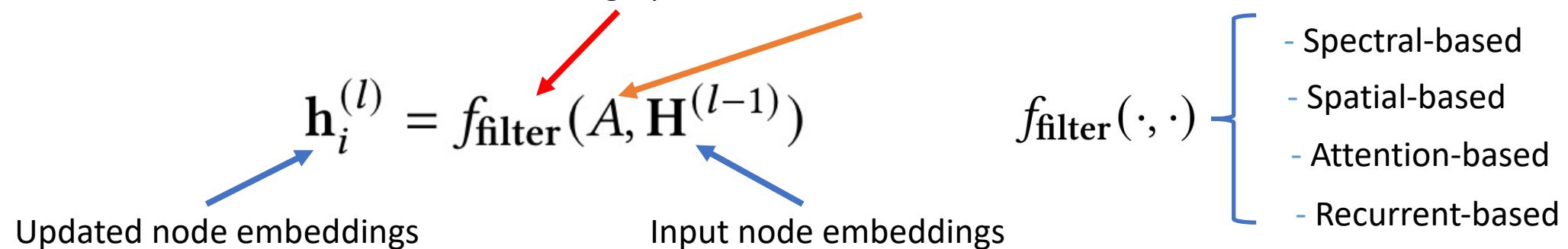


GNNs: A Brief History

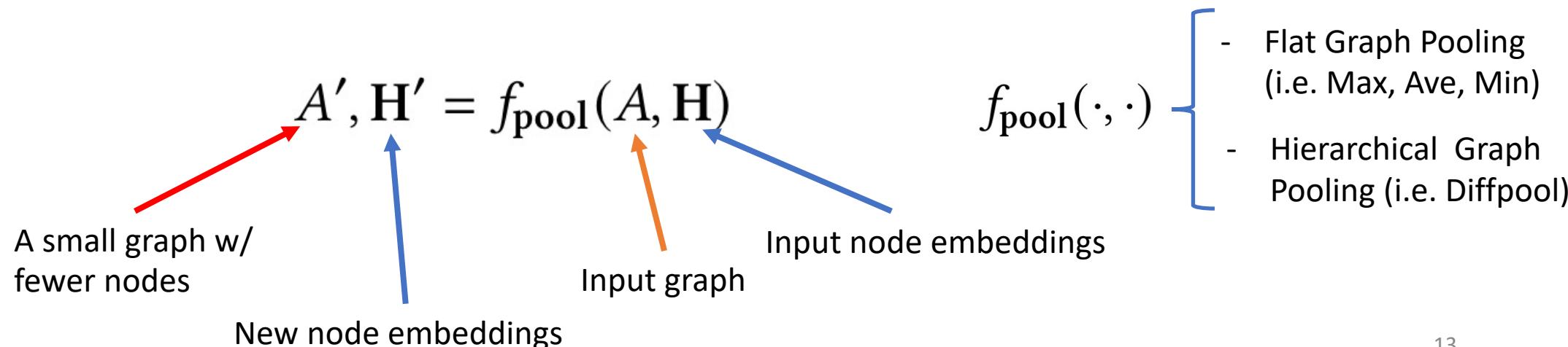


Graph Neural Networks: Foundations

- Learning node embeddings: A graph filter adjacency matrix

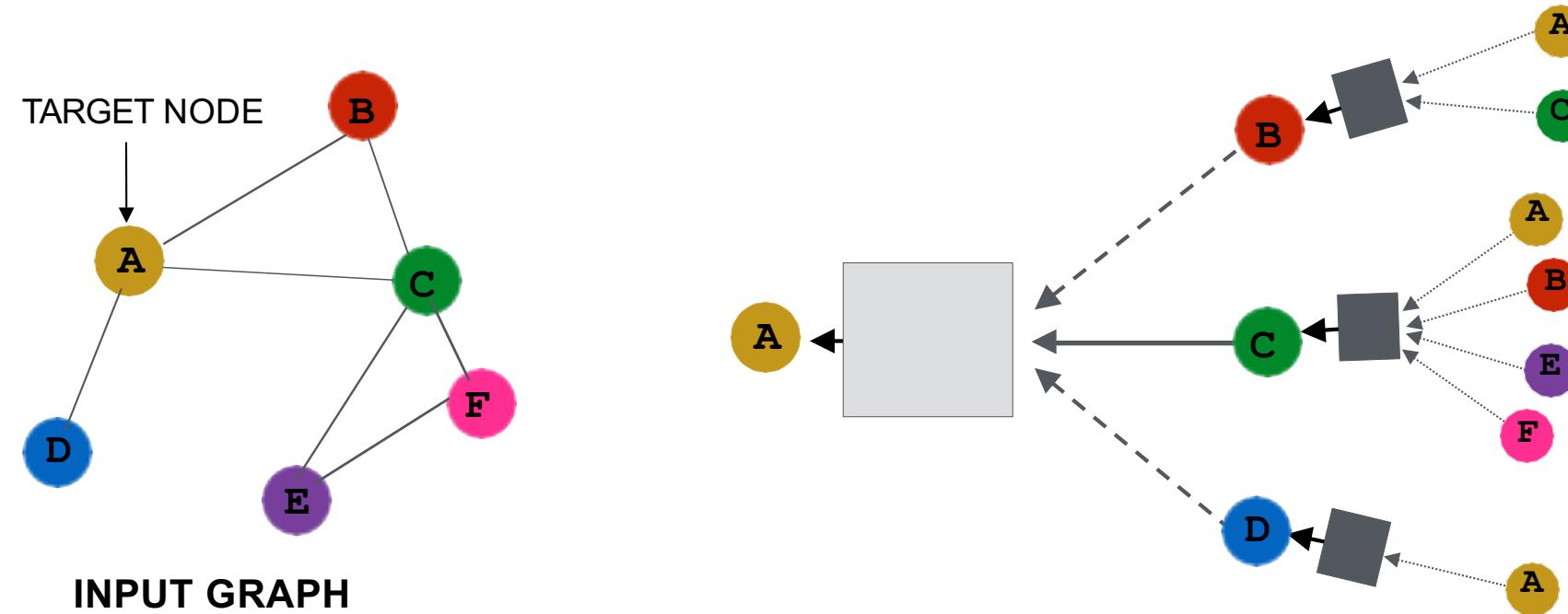


- Learning graph-level embeddings:



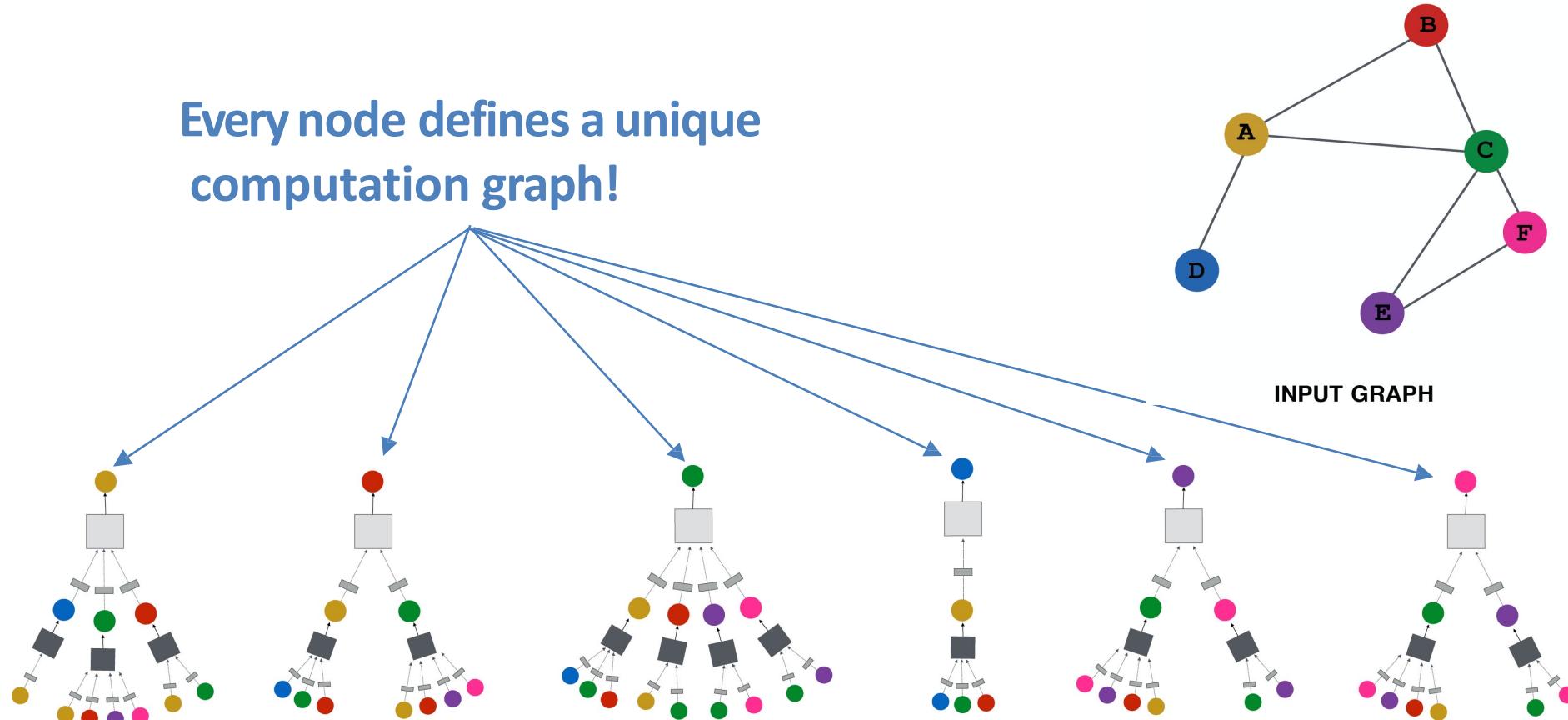
Graph Neural Networks: Basic Model

- **Key idea:** Generate node embeddings based on local neighborhoods.



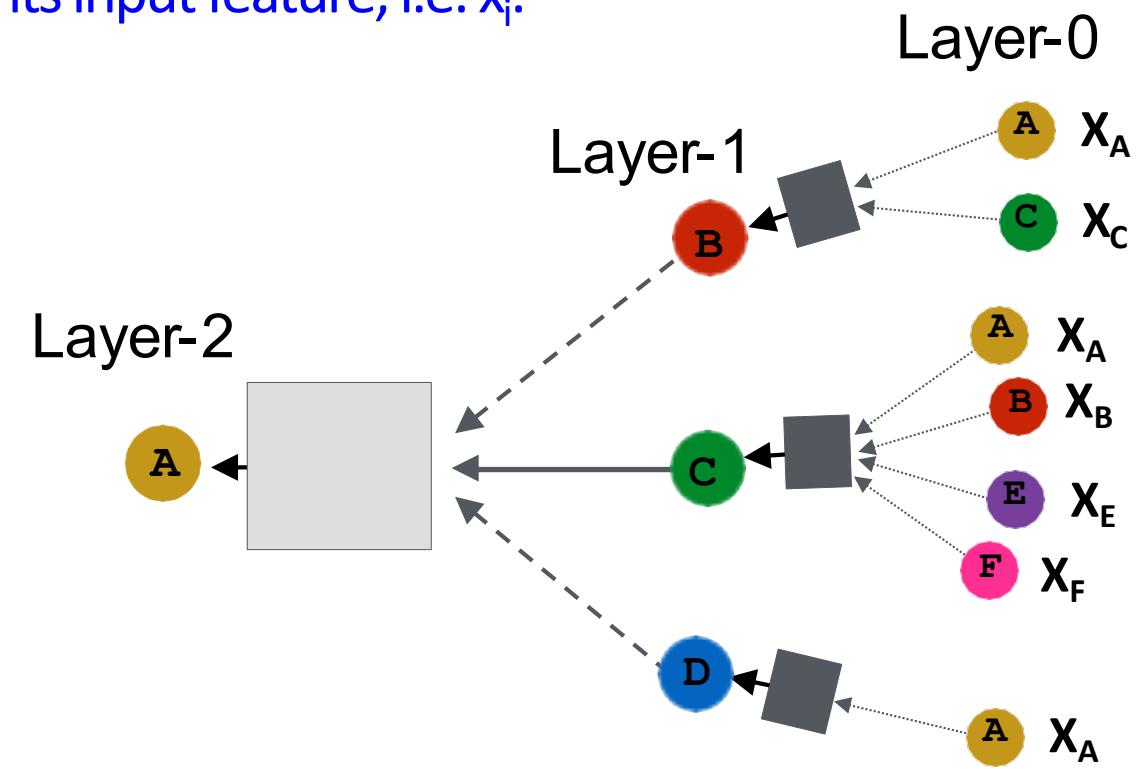
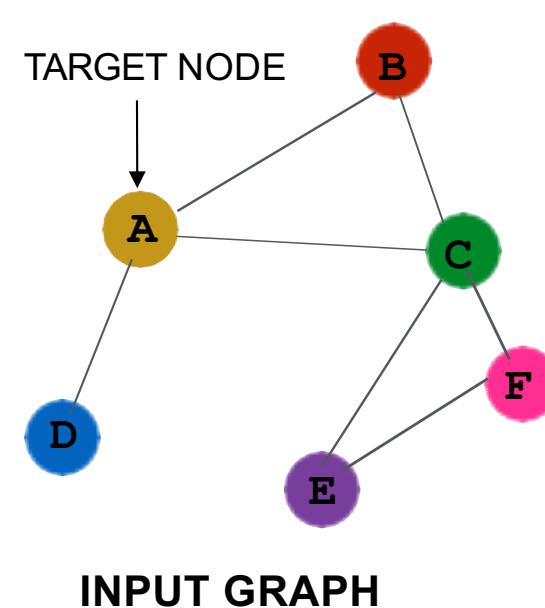
Neighborhood Aggregation

- **Intuition:** Network neighborhood defines a computation graph



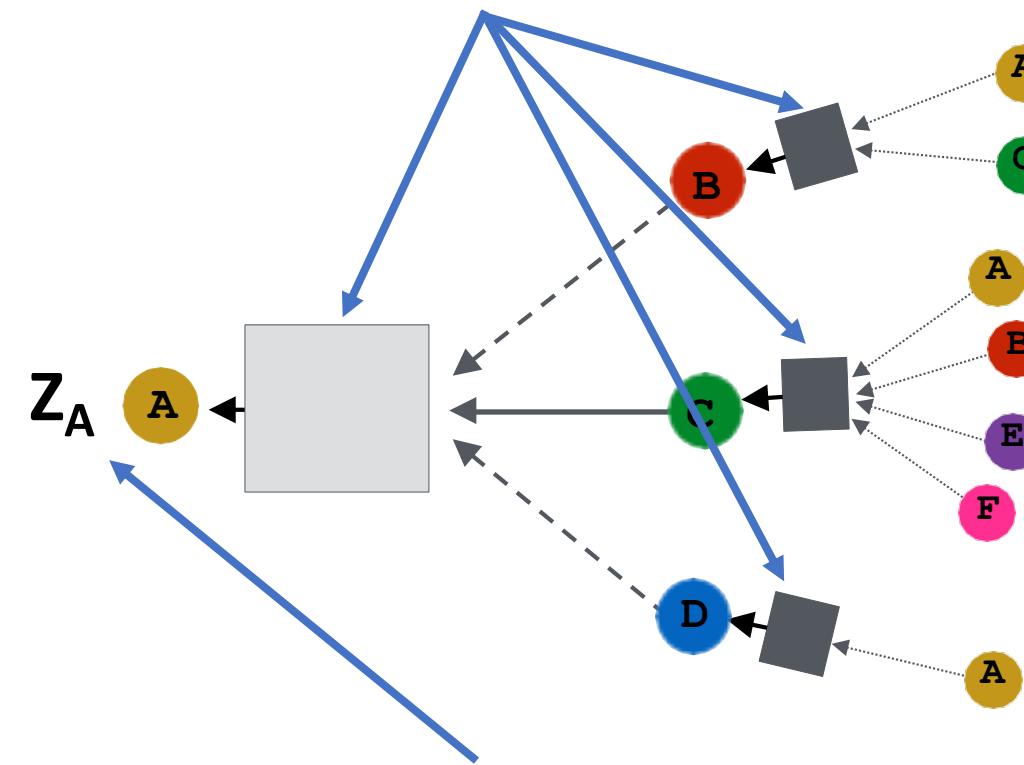
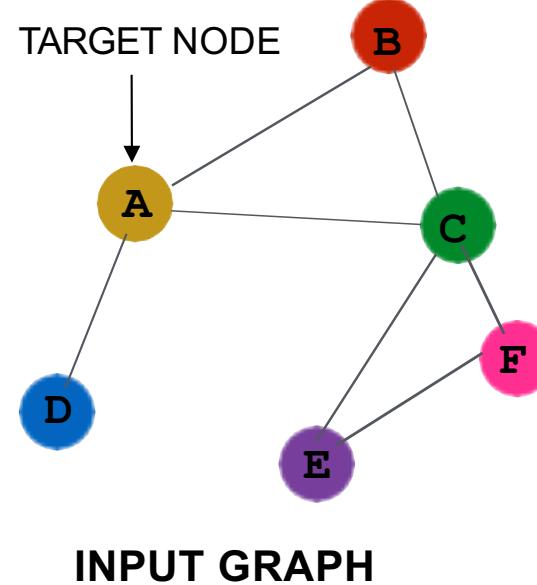
Neighborhood Aggregation

- Nodes have embeddings at each layer.
- Model can be arbitrary depth.
- “layer-0” embedding of node i is its input feature, i.e. x_i .



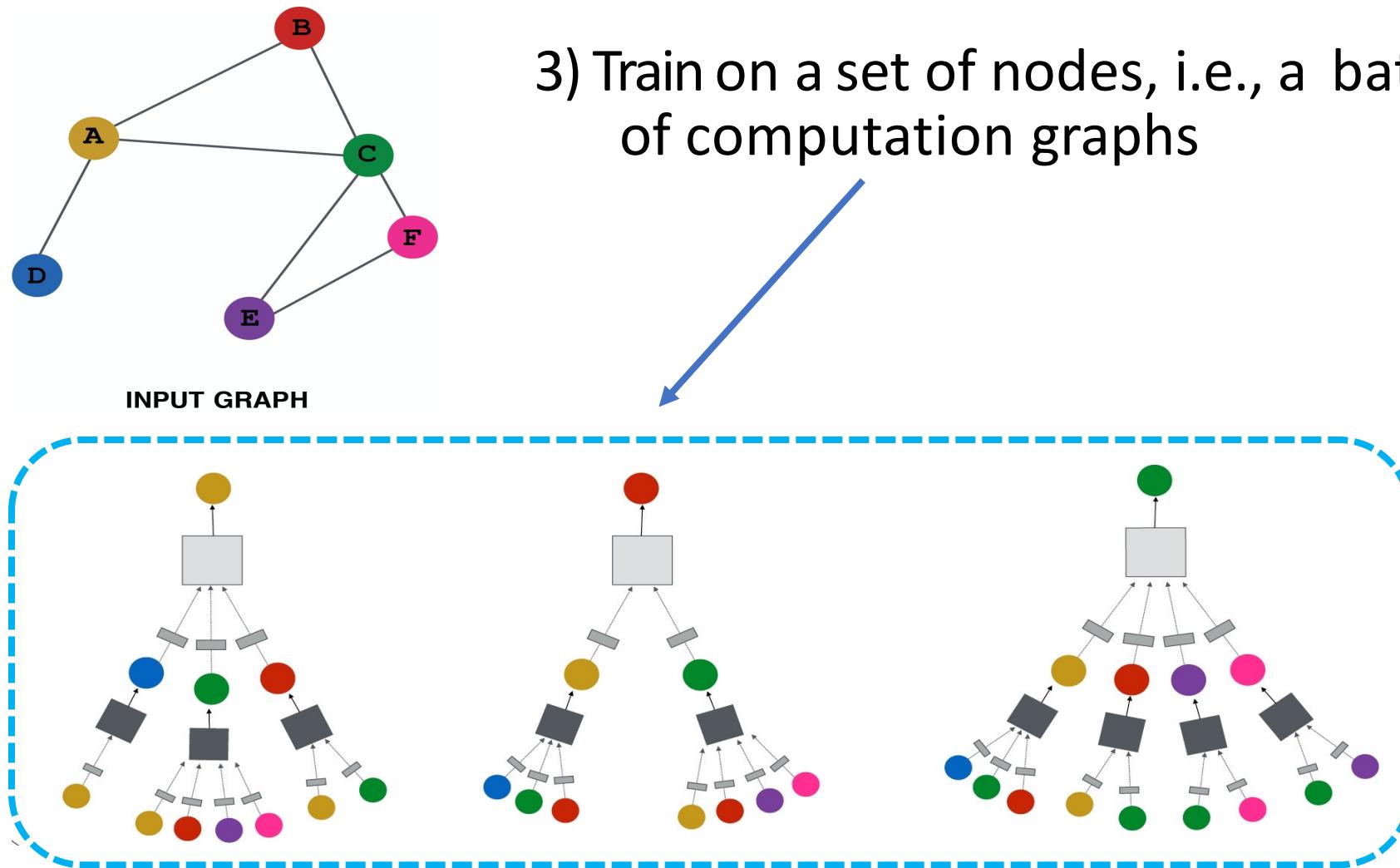
Overview of GNN Model

1) Define a neighborhood aggregation function

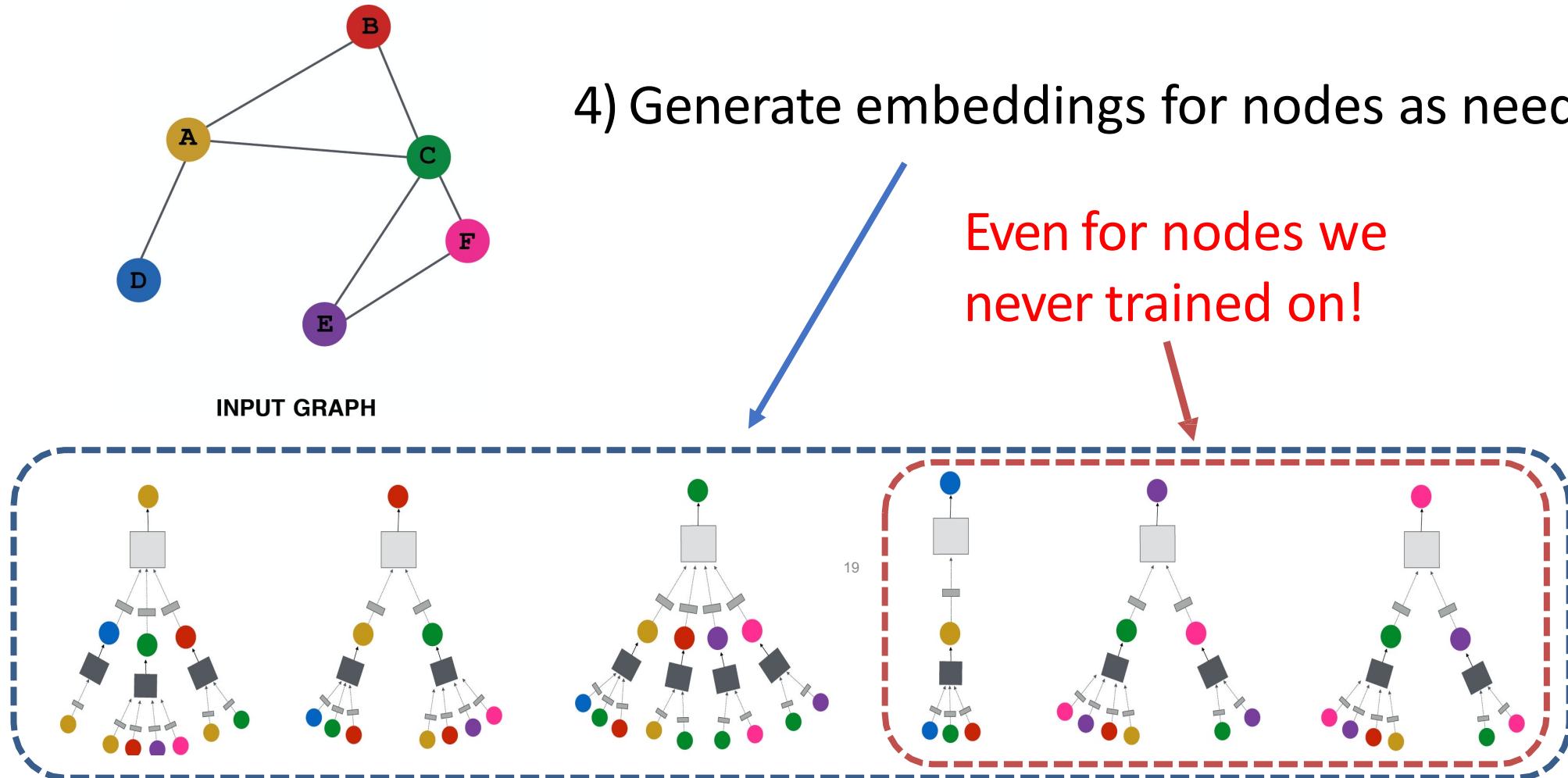


2) Define a loss function on the embeddings, $L(z_v)$

Overview of GNN Model

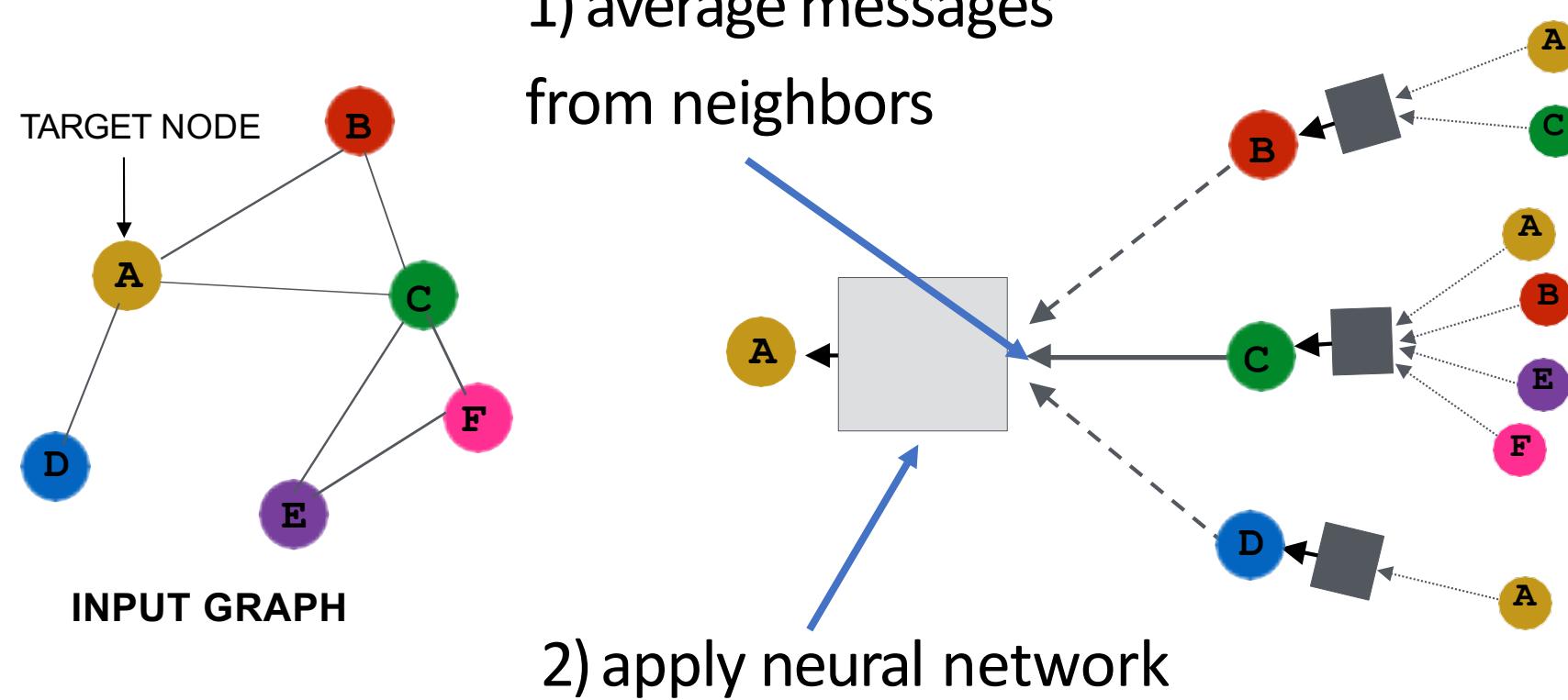


Overview of GNN Model



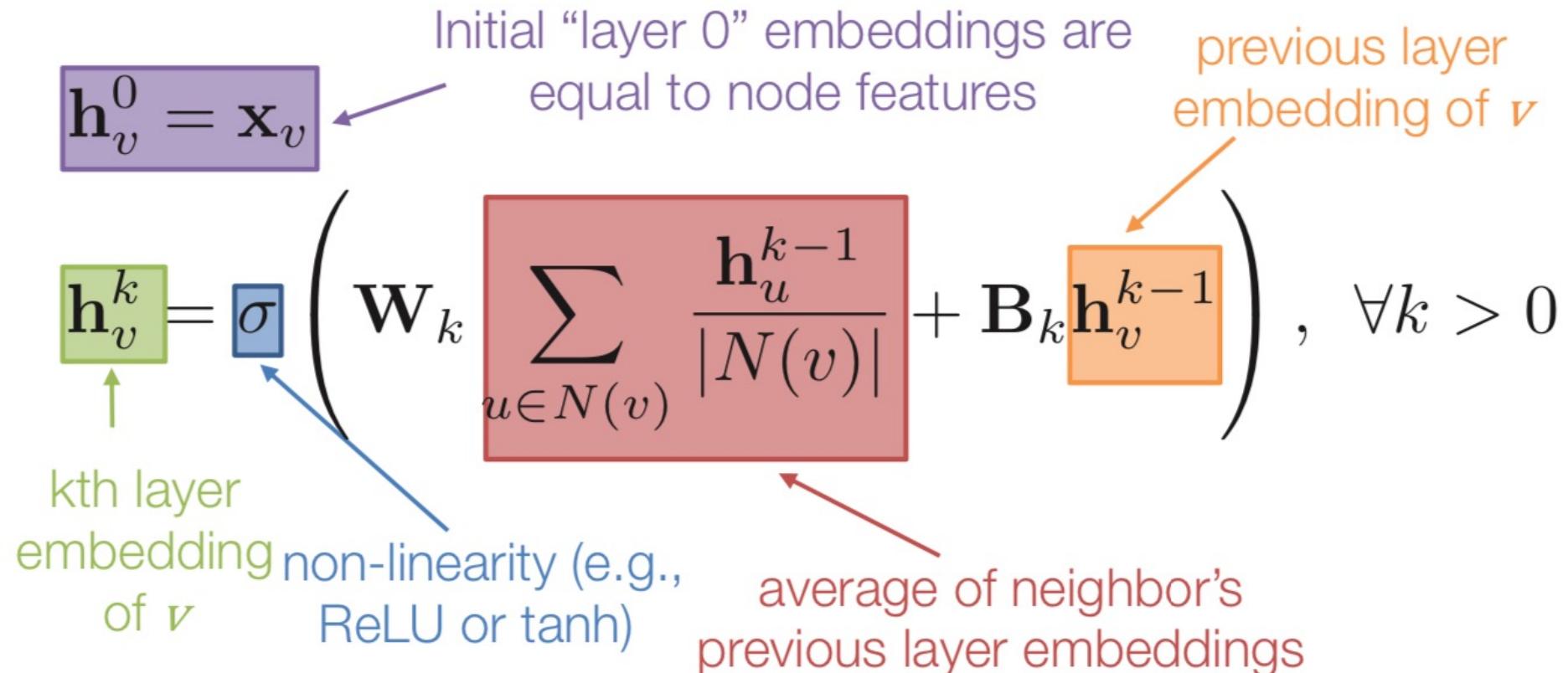
GNN Model: A Case Study

- **Basic approach:** Average neighbor information and apply a neural network



GNN Model: A Case Study

- **Basic approach:** Average neighbor information and apply a neural network.



GNN Model: Quick Summary

- Key idea: generate node embeddings by aggregating neighborhood information.
 - Allows for parameter sharing in the encoder
 - Allows for inductive learning

Graph Neural Networks: Popular Models

- Spectral-based Graph Filters
 - GCN (Kipf & Welling, ICLR 2017), Chebyshev-GNN (Defferrard et al. NIPS 2016)
- Spatial-based Graph Filters
 - MPNN (Gilmer et al. ICML 2017), GraphSage (Hamilton et al. NIPS 2017)
 - GIN (Xu et al. ICLR 2019)
- Attention-based Graph Filters
 - GAT (Velickovic et al. ICLR 2018)
- Recurrent-based Graph Filters
 - GGNN (Li et al. ICLR 2016)

Graph Convolution Networks (GCN)

Key idea: spectral convolution on graphs

Eigen-decomposition
is **expensive**

Chebyshev polynomials
accelerates but still not
powerful

First-order approxima-
tion fast and powerful

Renormalization trick
stabilizes the numerical
computation

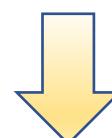
$$f_{\text{filter}} * \mathbf{x}_i = \mathbf{U} f(\Lambda) \mathbf{U}^T \mathbf{x}_i$$



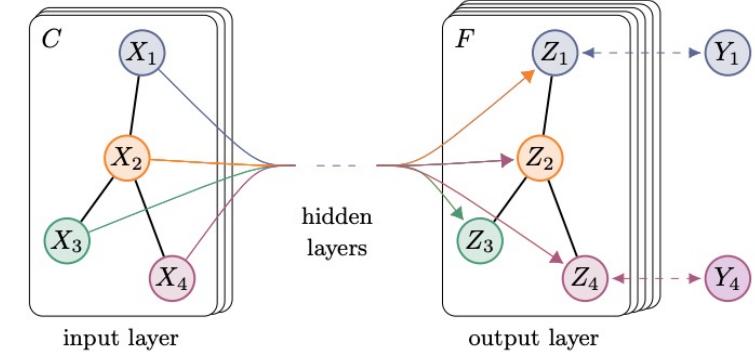
$$f'_{\text{filter}} * \mathbf{x}_i \approx \sum_{p=0}^P \theta'_p T_p(\tilde{\mathbf{L}}) \mathbf{x}_i$$



$$f_{\text{filter}} * \mathbf{h}_i^{(l)} \approx \theta(I_n + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) \mathbf{h}_i^{(l)}$$



$$\mathbf{H}^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}^{(l)})$$



GCN in NLP Tasks:

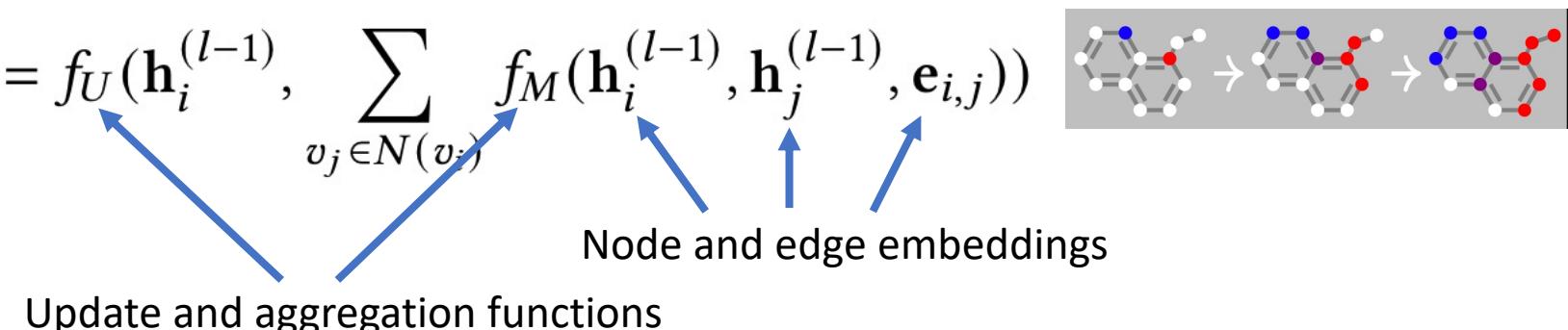
- Text classification
- Question Answering
- Text Matching
- Topic Modeling
- Information Extraction

Message Passing Neural Network (MPNN)

Key idea: graph convolutions as a message passing process

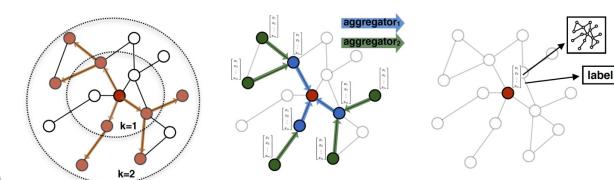
$$\text{MPNN: } \mathbf{h}_i^{(l)} = f_{\text{filter}}(A, \mathbf{H}^{(l-1)}) = f_U(\mathbf{h}_i^{(l-1)}, \sum_{v_j \in N(v_i)} f_M(\mathbf{h}_i^{(l-1)}, \mathbf{h}_j^{(l-1)}, \mathbf{e}_{i,j}))$$

expensive if
the number
of nodes are
large



$$\text{GraphSage: } f_{\text{filter}}(A, \mathbf{H}^{(l-1)}) = \sigma(\mathbf{W}^{(l)} \cdot f_M(\mathbf{h}_i^{(l-1)}, \{\mathbf{h}_j^{(l-1)}, \forall v_j \in N(v_i)\}))$$

sampling to
obtain a fixed
number of
neighbors



- MPNN and GraphSage in NLP Tasks:**
- Knowledge graph
 - Information extraction
 - Semantic parsing
- 25

Graph Attention Network (GAT)

Key idea: dynamically learn the weights (attention scores) on the edges when performing message passing

Weighted sum of node embeddings

$$\mathbf{h}_i^{(l)} = f_{\text{filter}}(A, \mathbf{H}^{(l-1)}) = \sigma \left(\sum_{v_j \in N(v_i)} \alpha_{ij} \mathbf{W}^{(l)} \mathbf{h}_j^{(l-1)} \right)$$

Learned local weights with self-attention

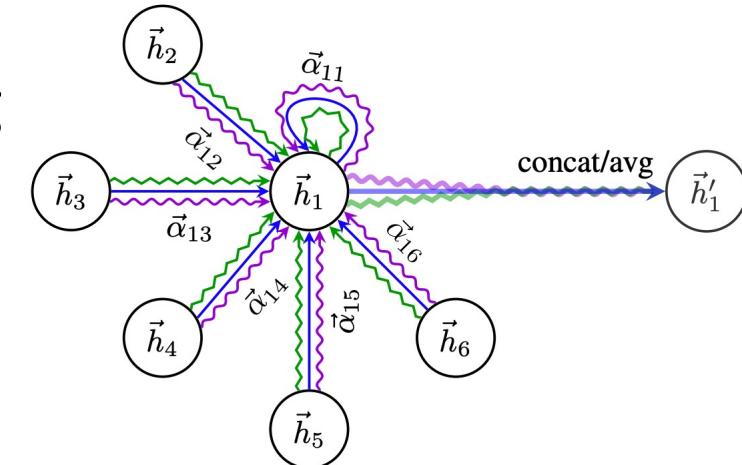
$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{u}^{(l)T} [\mathbf{W}^{(l)} \mathbf{h}_i^{(l-1)} || \mathbf{W}^{(l)} \mathbf{h}_j^{(l-1)}]))}{\sum_{v_k \in N(v_i)} \exp(\text{LeakyReLU}(\mathbf{u}^{(l)T} [\mathbf{W}^{(l)} \mathbf{h}_i^{(l-1)} || \mathbf{W}^{(l)} \mathbf{h}_k^{(l-1)}]))}$$

Intermediate node embeddings

$$f_{\text{filter}}(A, \mathbf{H}^{(l-1)}) = \parallel_{k=1}^K \sigma \left(\sum_{v_j \in N(v_i)} \alpha_{ij}^k \mathbf{W}_k^{(l)} \mathbf{h}_j^{(l-1)} \right)$$

Final node embeddings

$$f_{\text{filter}}(A, \mathbf{H}^{(L-1)}) = \sigma \left(\frac{1}{K} \sum_{k=1}^K \sum_{v_j \in N(v_i)} \alpha_{ij}^k \mathbf{W}_k^{(L)} \mathbf{h}_j^{(L-1)} \right)$$



GAT in NLP Tasks:

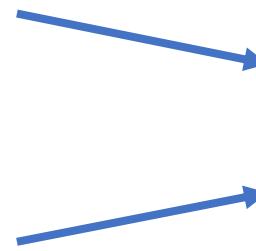
- Text classification
- Question Answering
- Knowledge graph
- Information extraction
- Semantic parsing

Gated Graph Neural Networks (GGNN)

Key idea: the use of Gated Recurrent Units while taking into account edge type and directions

Zero-padding
input node
embeddings

Incoming &
outcoming
edges for
node v_i



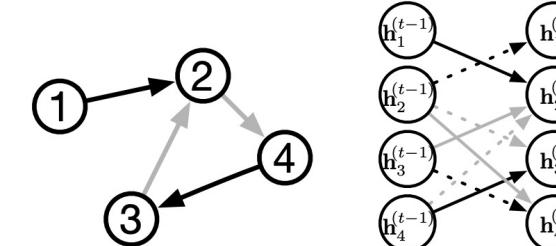
$$\mathbf{h}_i^{(0)} = [\mathbf{x}_i^T, \mathbf{0}]^T$$

$$\mathbf{a}_i^{(l)} = A_{i:}^T [\mathbf{h}_1^{(l-1)} \dots \mathbf{h}_n^{(l-1)}]^T$$

$$\mathbf{h}_i^{(l)} = \text{GRU}(\mathbf{a}_i^{(l)}, \mathbf{h}_i^{(l-1)})$$



GRU for fusing node embeddings

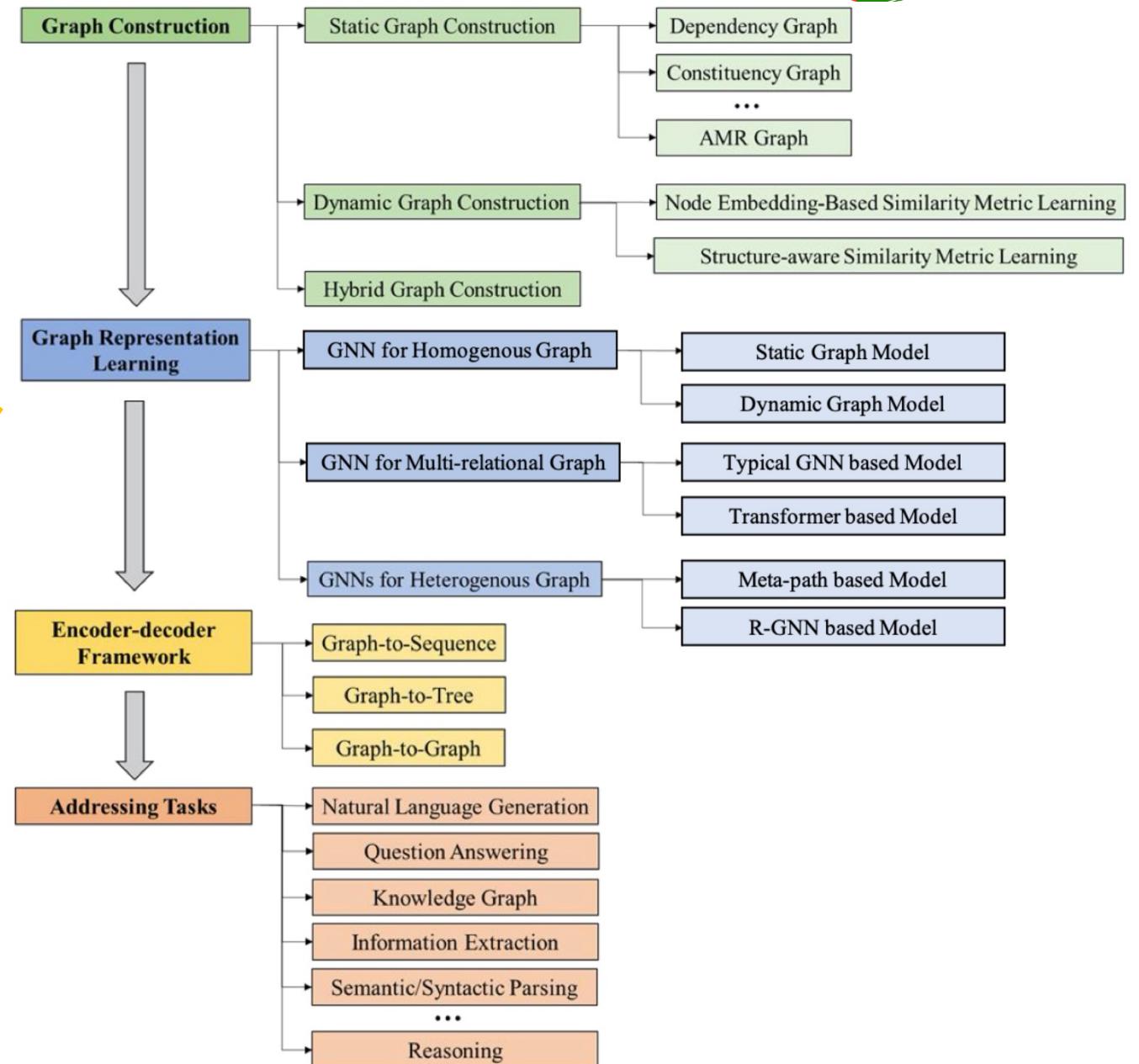
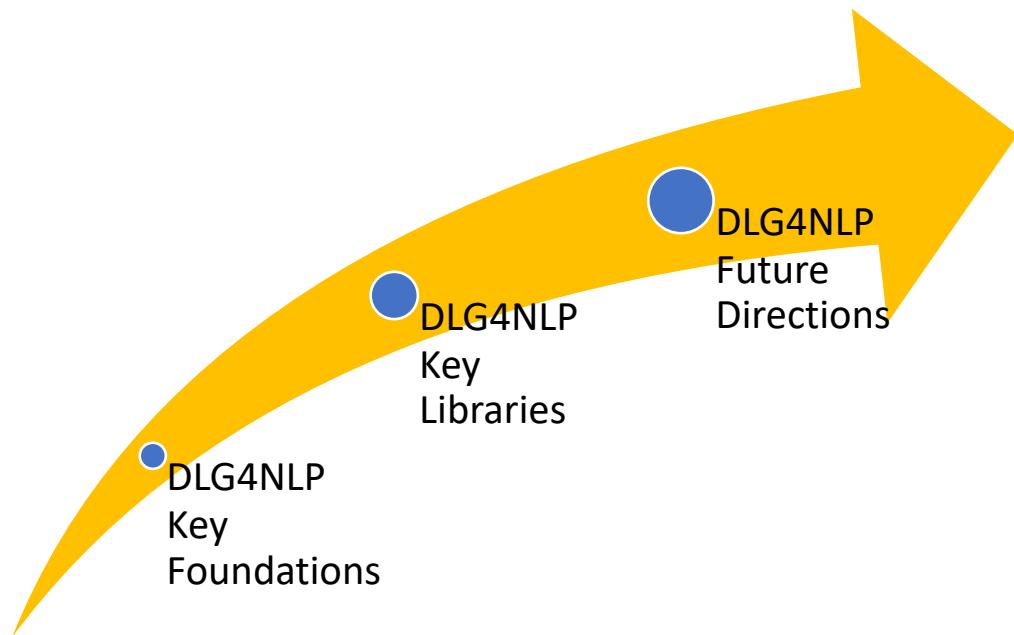


	Outgoing Edges				Incoming Edges			
	1	2	3	4	1	2	3	4
1	B							
2		C	B'			C'		
3	C						B'	
4		B		C'				

GGNN in NLP Tasks:

- Semantic parsing
- Machine translation

DLG4NLP: A Roadmap

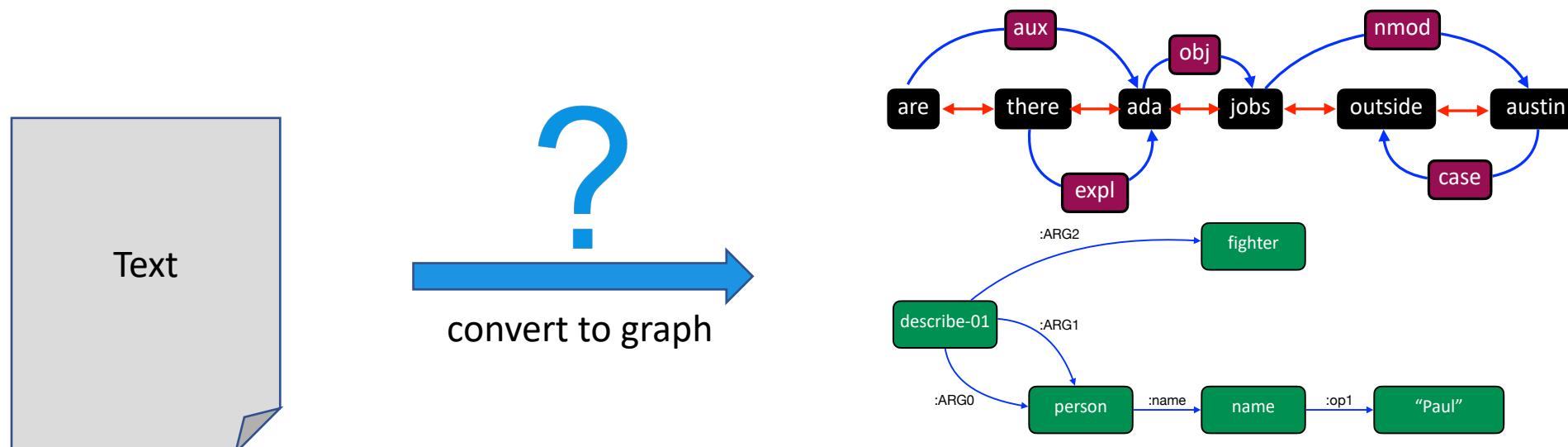


DLG4NLP Foundations

Graph Construction for NLP

Why Graph Construction for NLP?

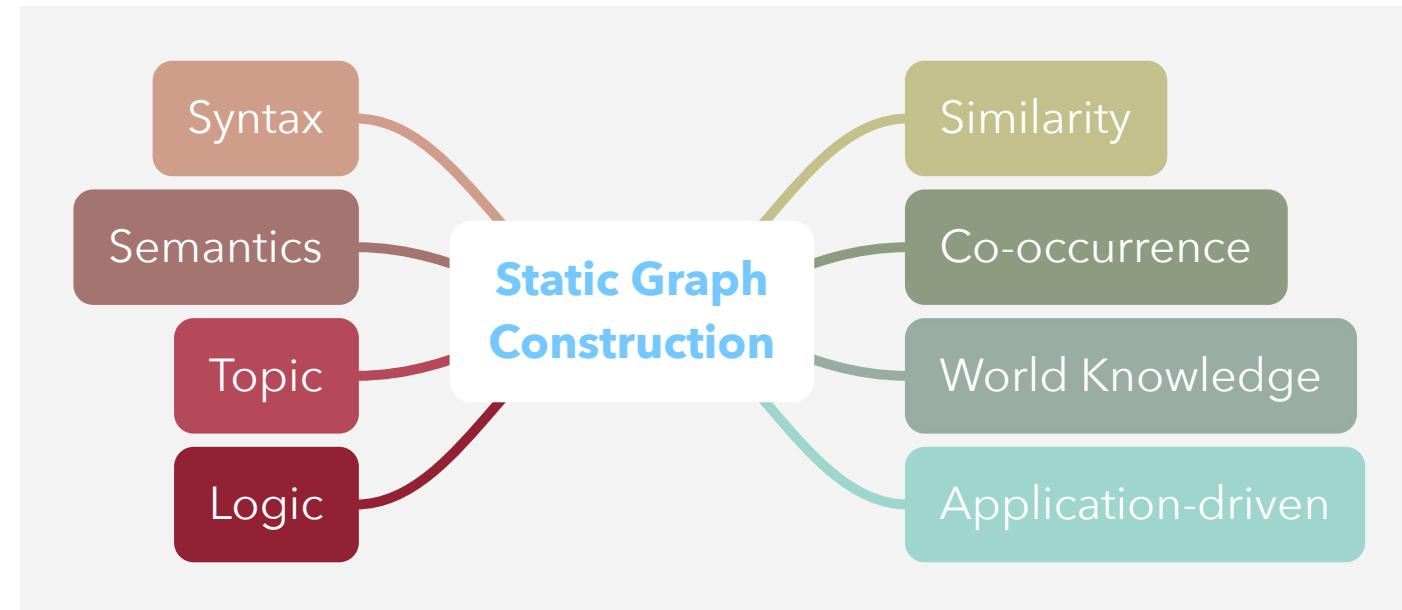
- Representation power: **graph** > sequence > bag
- Different NLP tasks require **different aspects** of text , e.g., syntax, semantics.
- Different graphs capture different aspects of the text
- Two categories: static vs dynamic graph construction
- Goal: good downstream task performance



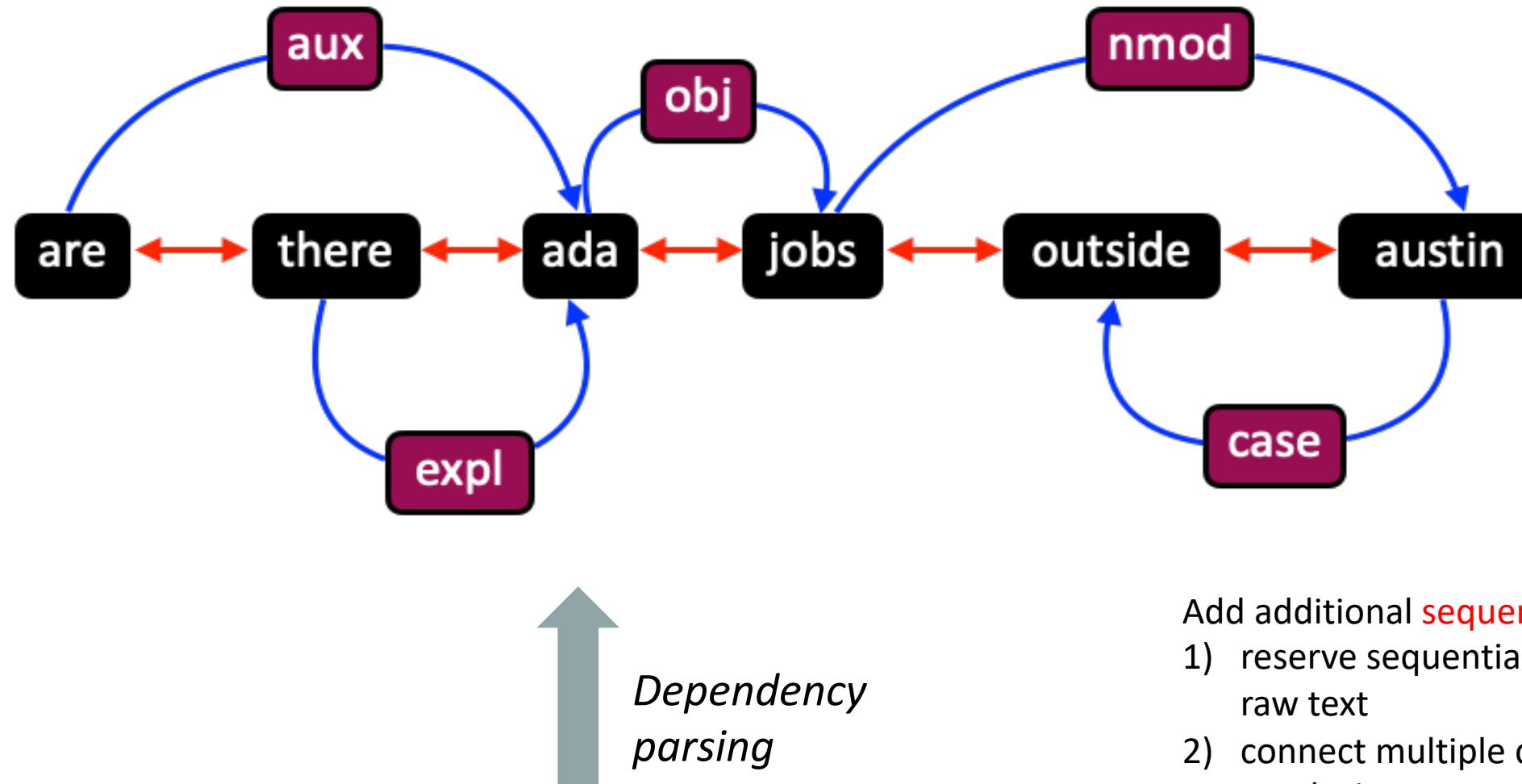
many more graph options...

Static Graph Construction

- Problem setting:
 - **Input:** raw text (e.g., sentence, paragraph, document, corpus)
 - **Output:** graph
- Conducted during **preprocessing** by augmenting text with **domain knowledge**



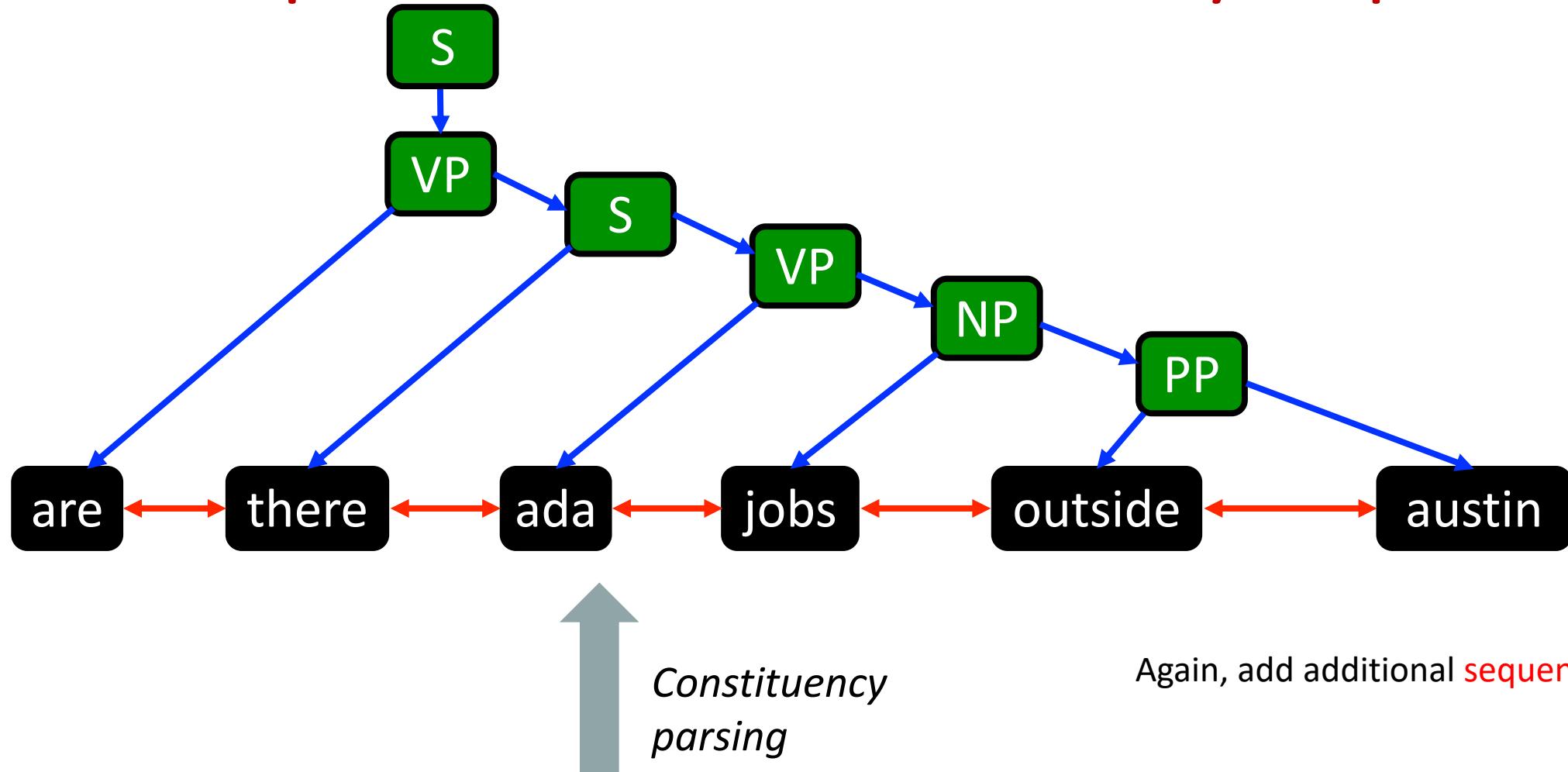
Static Graph Construction: Dependency Graph



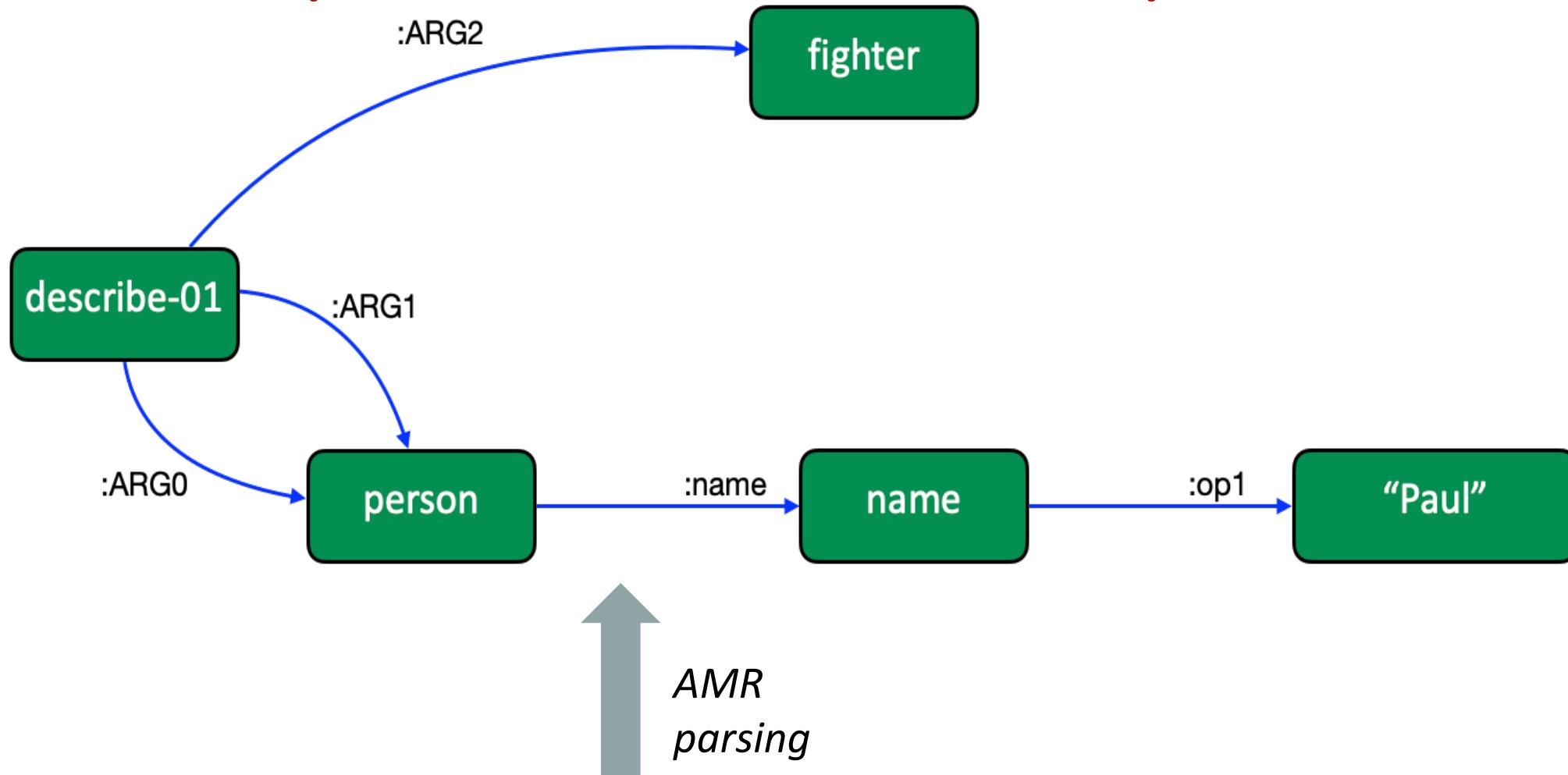
Text input: are there ada jobs outside austin

- Add additional **sequential edges** to
- 1) reserve sequential information in raw text
 - 2) connect multiple dependency graphs in a paragraph

Static Graph Construction: Constituency Graph

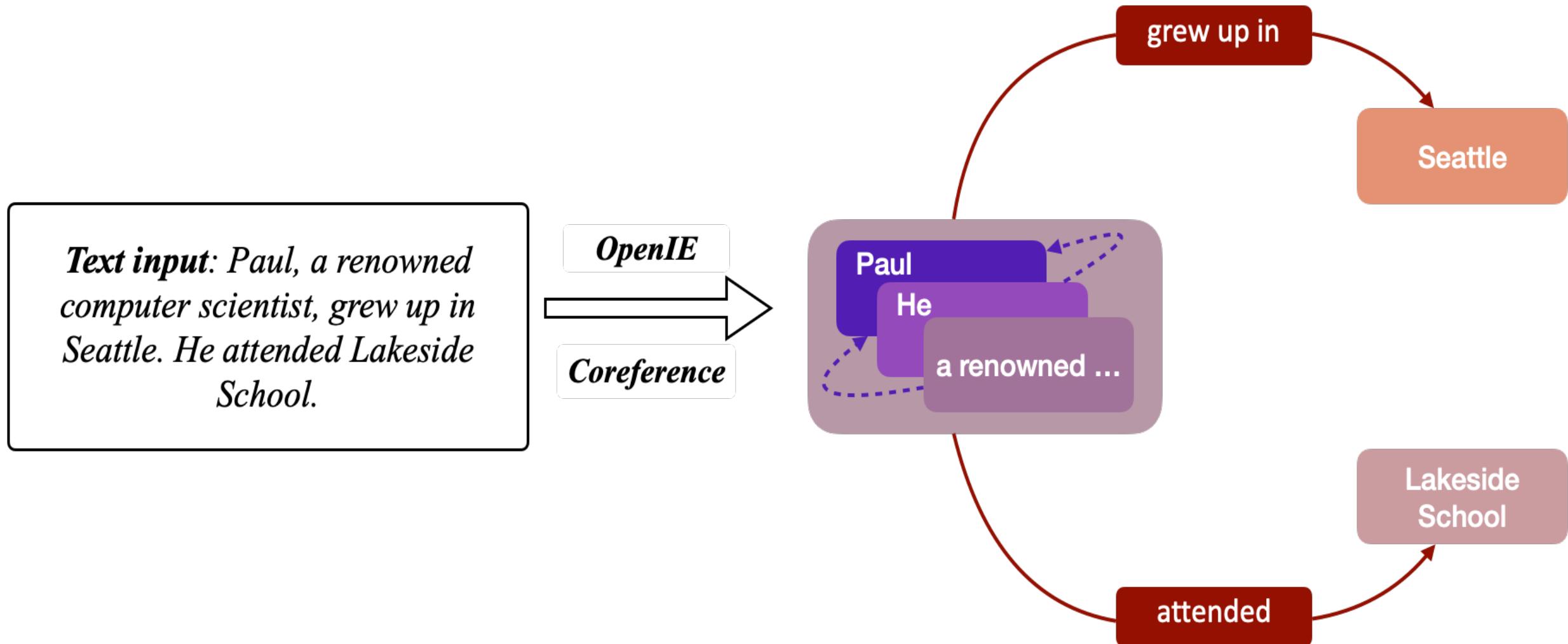


Static Graph Construction: AMR Graph

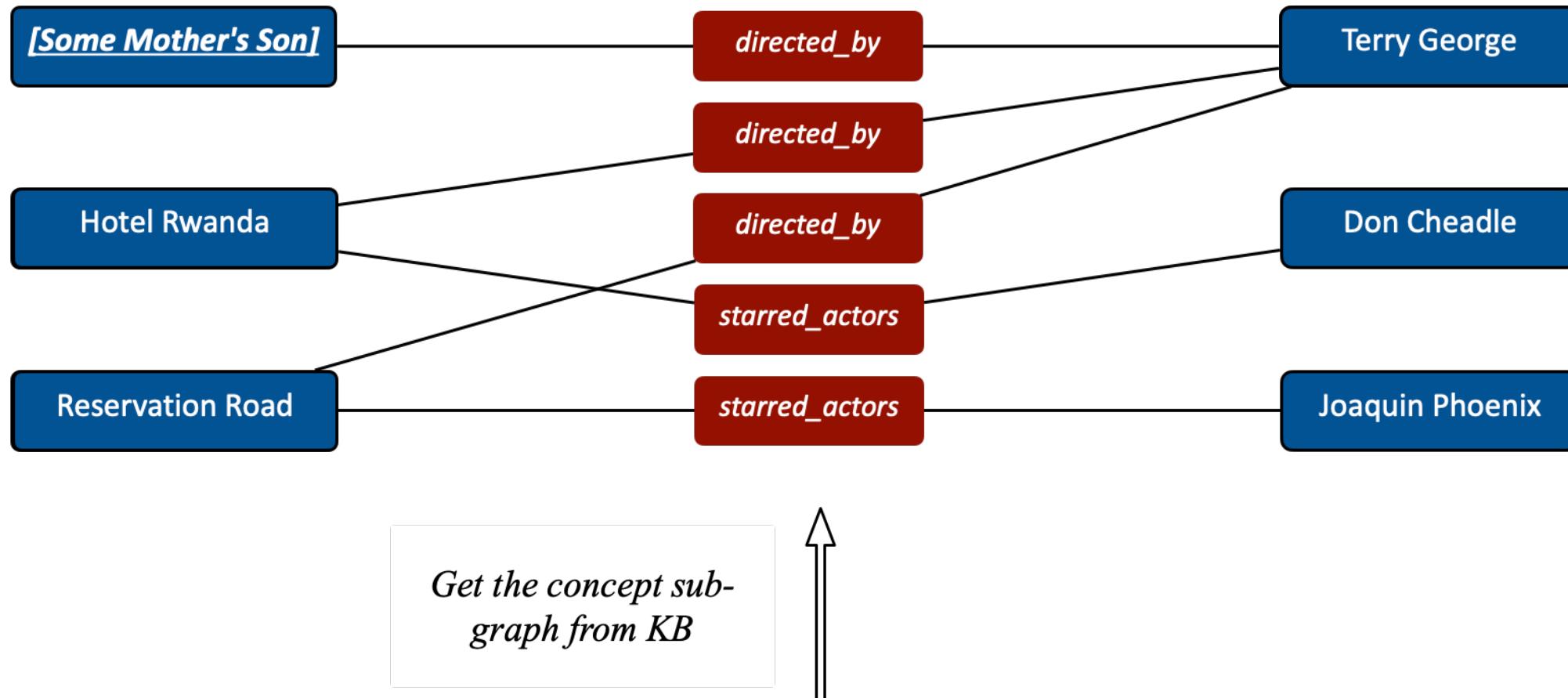


Text input: Paul's description of himself: a fighter

Static Graph Construction: IE Graph



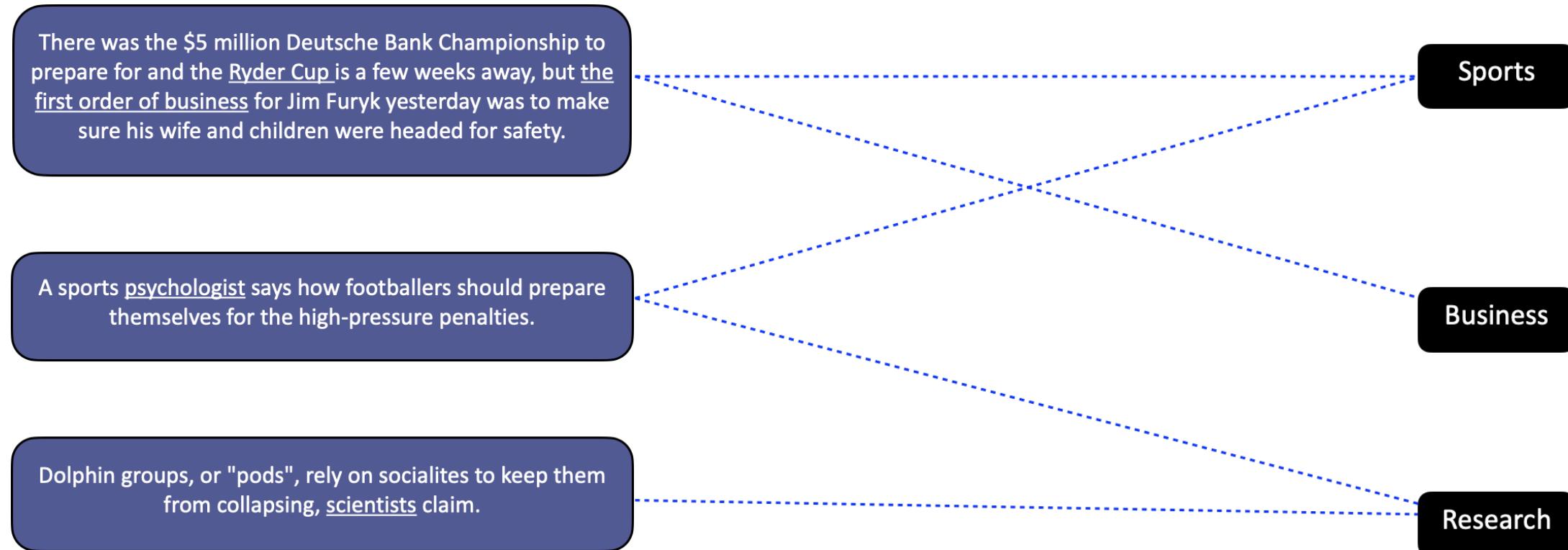
Static Graph Construction: Knowledge Graph



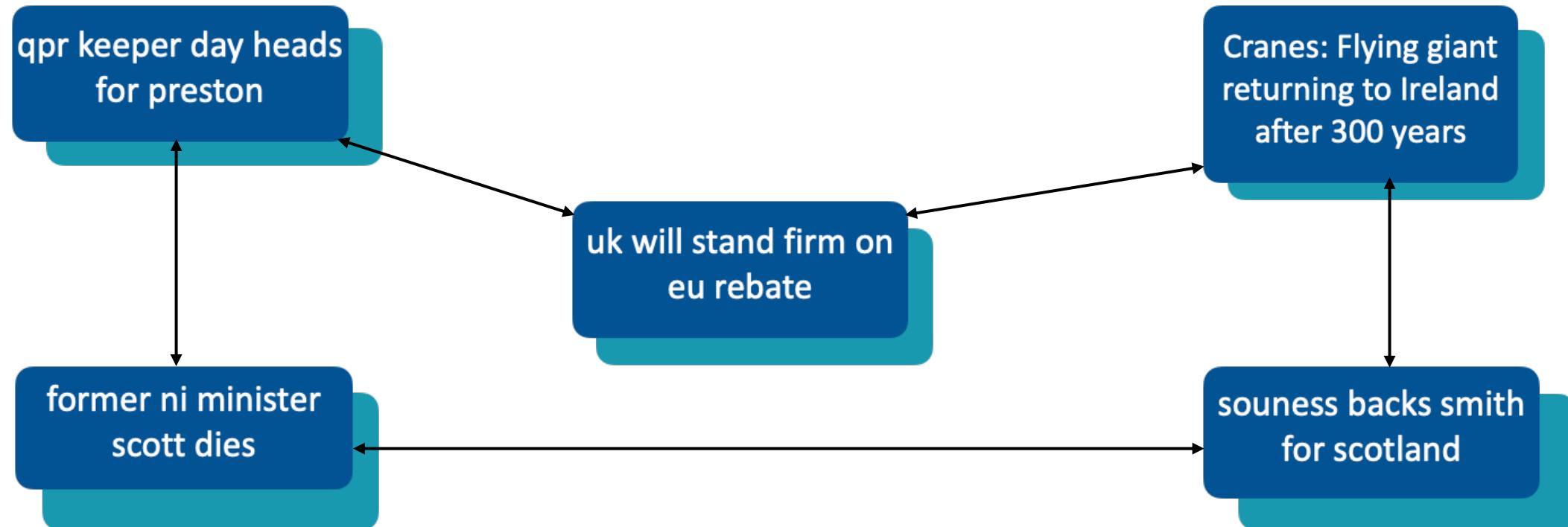
Question: who acted in the movies directed by the director of **[Some Mother's Son]**

Answer: Don Cheadle, Joaquin Phoenix

Static Graph Construction: Topic Graph



Static Graph Construction: Similarity Graph

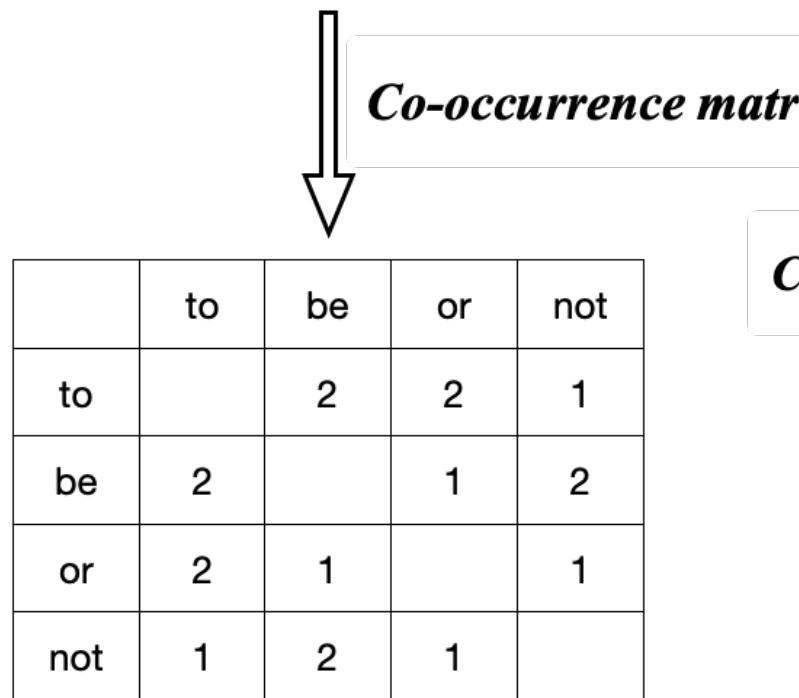


Sentence

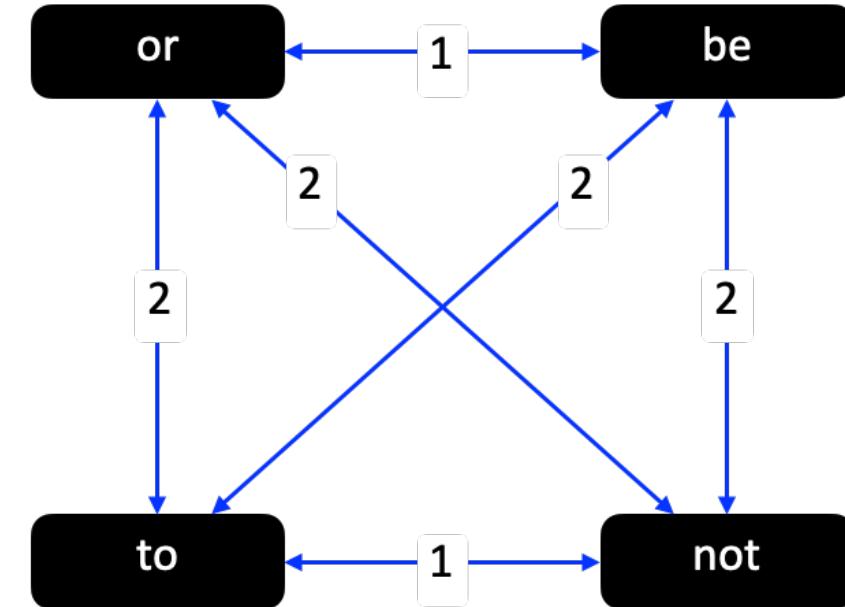
Sentence
TF-IDF vector

Static Graph Construction: Co-occurrence Graph

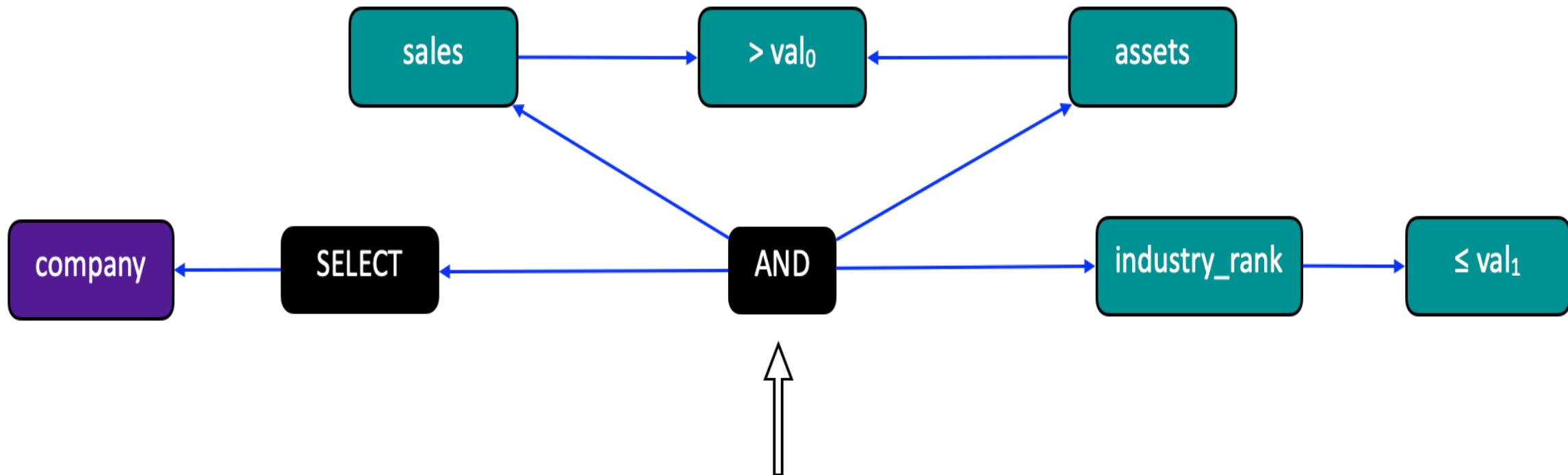
Text input: To be, or not to be: ...



Co-occurrence graph



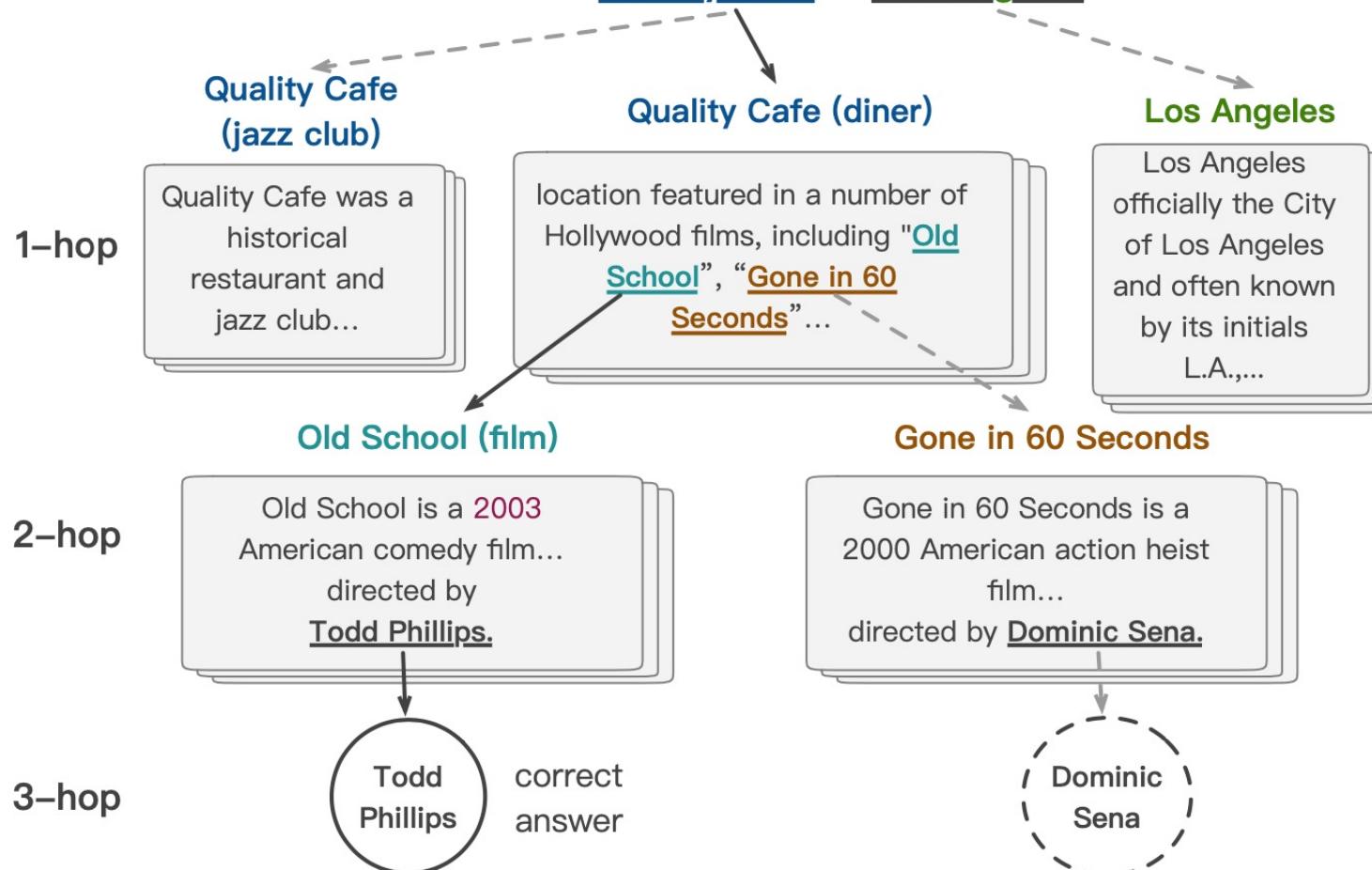
Static Graph Construction: SQL Graph



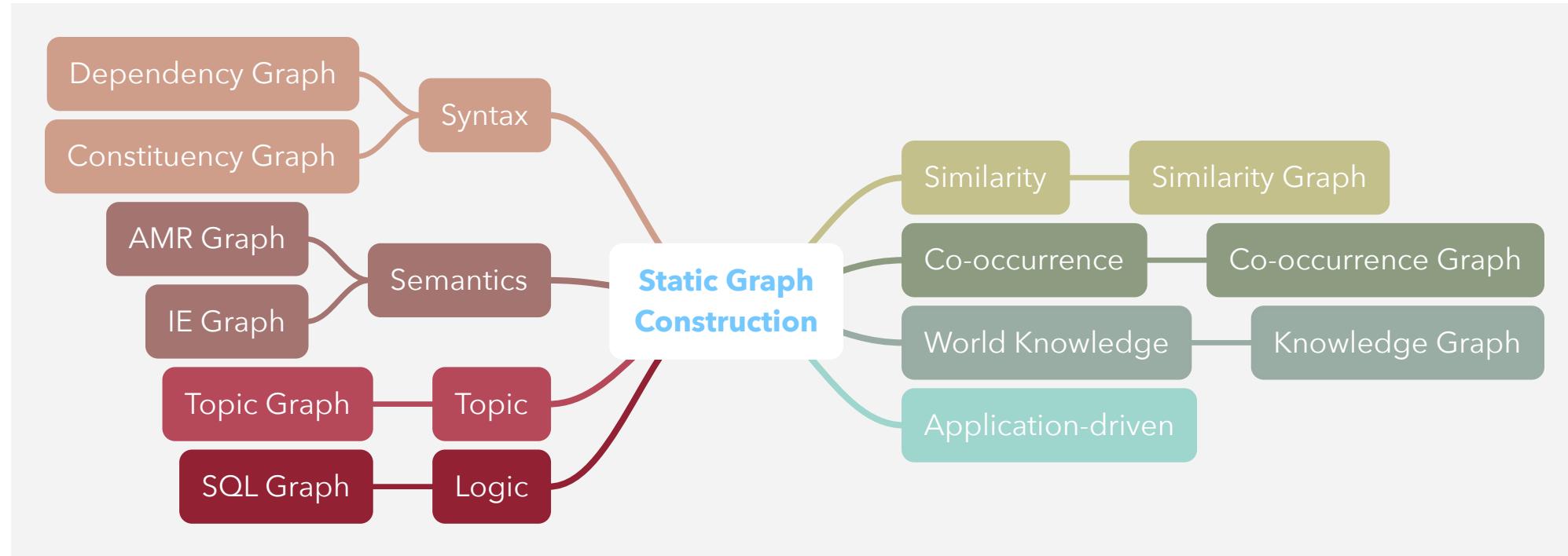
*SQL query input: **SELECT** company **WHERE** assets > val₀ **AND** sales > val₀ **AND** industry_rank ≤ val₁*

Static Graph Construction: Application-driven Graph

Question: Who is the director of the **2003** film which has scenes
in it filmed at the **Quality Cafe** in **Los Angeles**?



Static Graph Construction: Summary

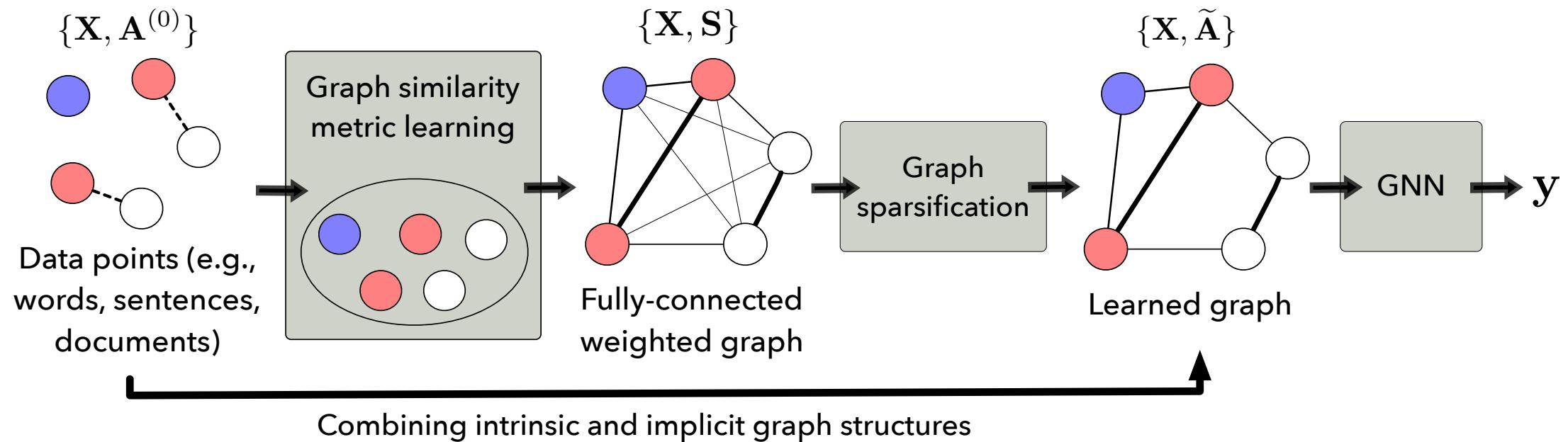


Widely used in various NLP applications such as NLG, MRC, semantic parsing, etc.

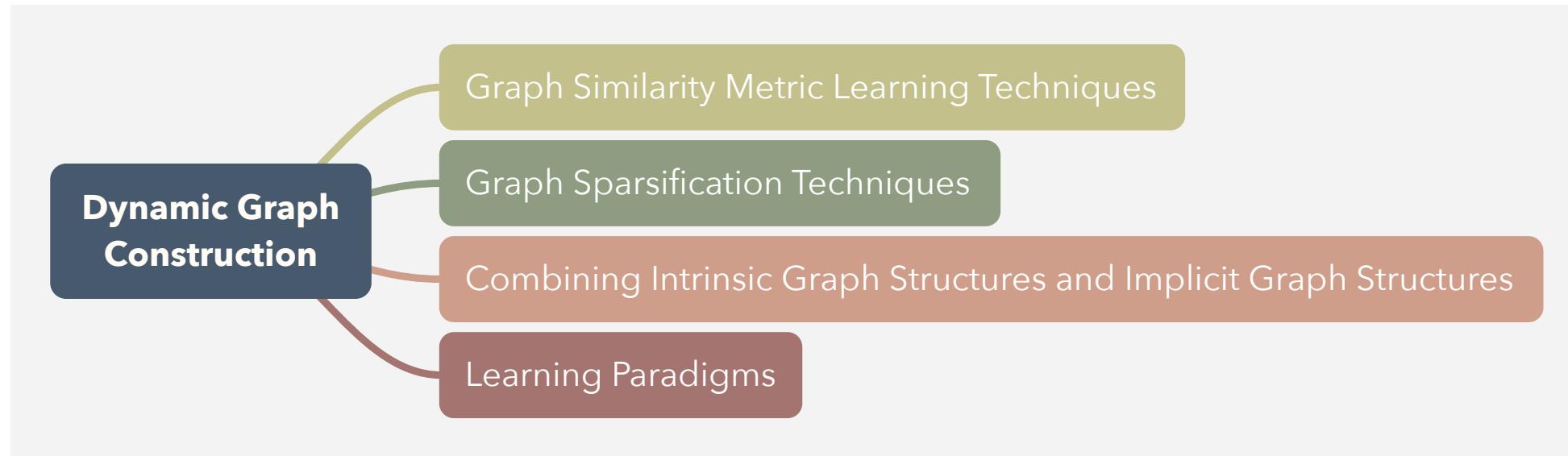
Dynamic Graph Construction

- Problem setting:
 - **Input:** raw text (e.g., sentence, paragraph, document, corpus)
 - **Output:** graph
- Graph structure (adjacency matrix) learning **on the fly**, joint with graph representation learning

Dynamic Graph Construction: Overview

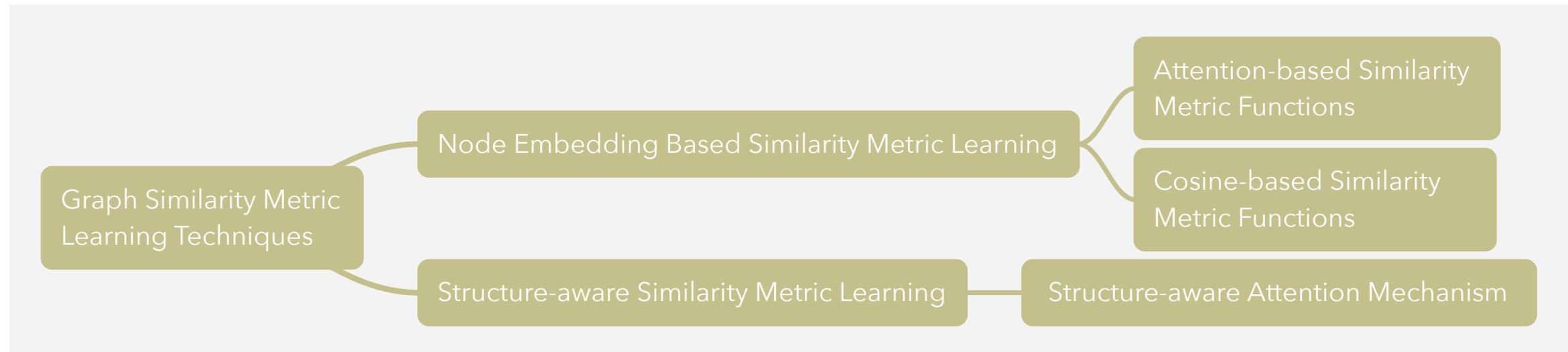


Dynamic Graph Construction Outline



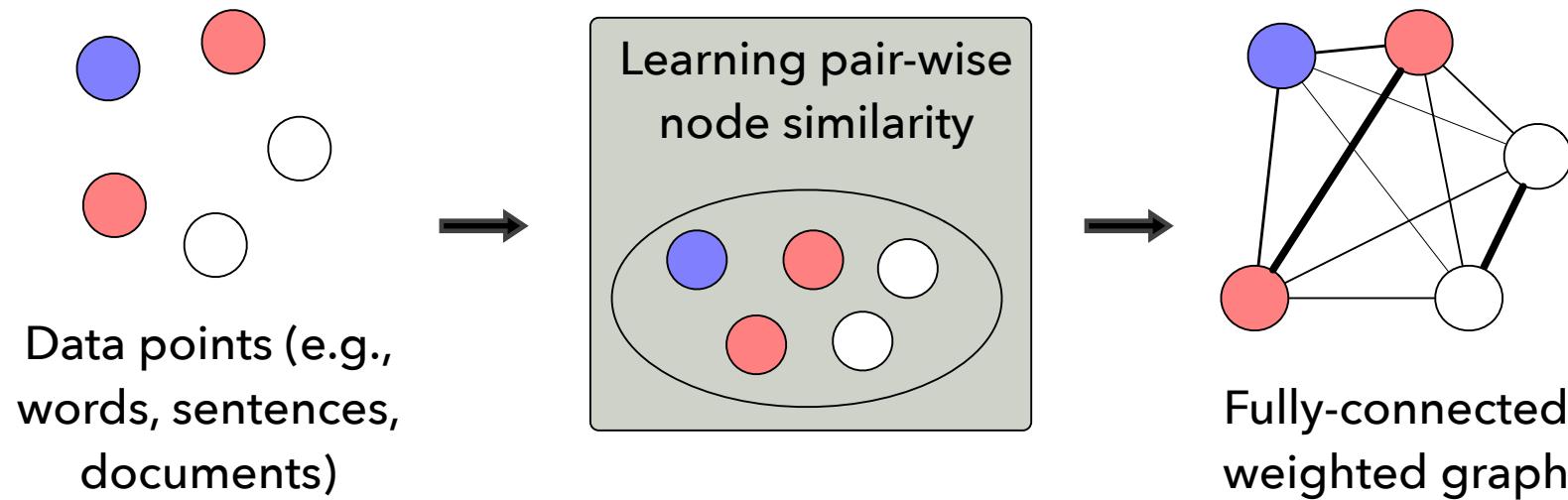
Graph Similarity Metric Learning Techniques

- Graph structure learning as **similarity metric learning** (in the node embedding space)
- Enabling **inductive learning**
- Various metric functions



Node Embedding Based Similarity Metric Learning

- Learning a weighted adjacency matrix by computing the **pair-wise node similarity** in the embedding space
- Common metrics functions
 - Attention-based similarity metric functions
 - Cosine-based similarity metric functions

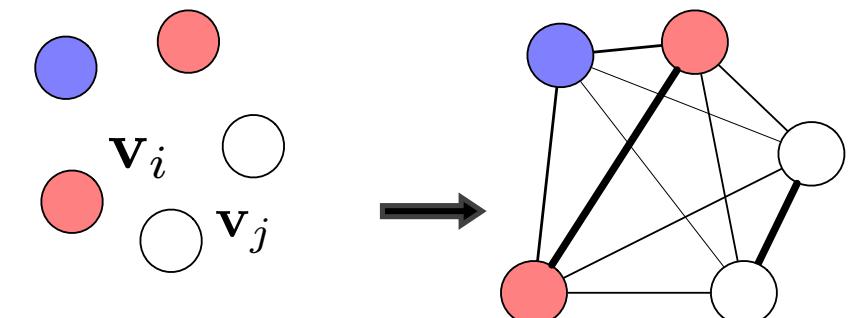


Attention-based Similarity Metric Functions

Variant 1)

$$S_{i,j} = (\mathbf{v}_i \odot \mathbf{u})^T \mathbf{v}_j$$

Node feature vector
 Non-negative learnable weight vector



Variant 2)

$$S_{i,j} = \text{ReLU}(\mathbf{W}\mathbf{v}_i)^T \text{ReLU}(\mathbf{W}\mathbf{v}_j)$$

Learnable weight matrix

Data points (e.g., words, sentences, documents)
 Fully-connected weighted graph

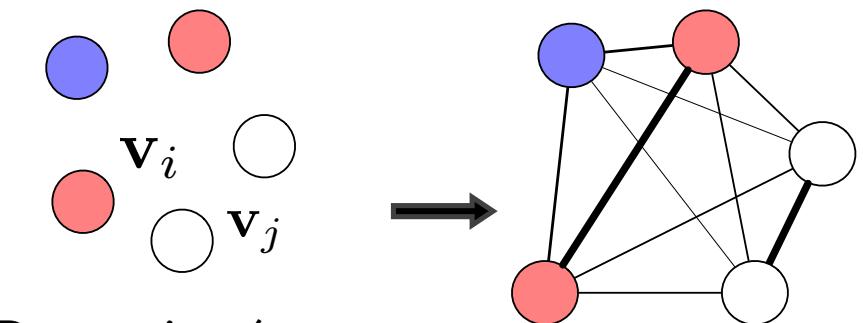
Cosine-based Similarity Metric Functions

$$S_{i,j}^p = \cos(\mathbf{w}_p \odot \mathbf{v}_i, \mathbf{w}_p \odot \mathbf{v}_j)$$

Learnable weight vector

$$S_{i,j} = \frac{1}{m} \sum_{p=1}^m S_{ij}^p$$

Multi-head similarity scores

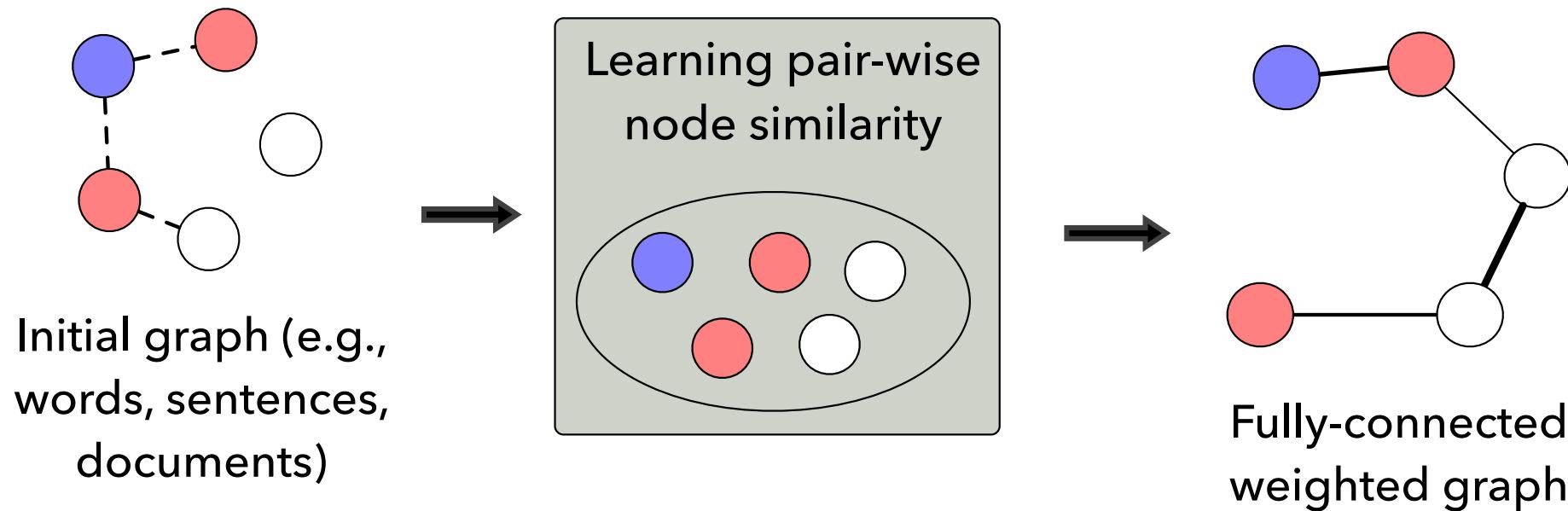


Data points (e.g., words, sentences, documents)

Fully-connected weighted graph

Structure-aware Similarity Metric Learning

- Learning a weighted adjacency matrix by computing the **pair-wise node similarity** in the embedding space
- Considering **existing edge information** of the intrinsic graph in addition to the node information

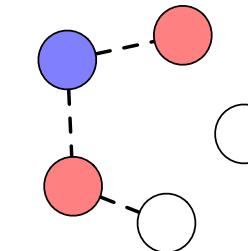


Attention-based Similarity Metric Functions

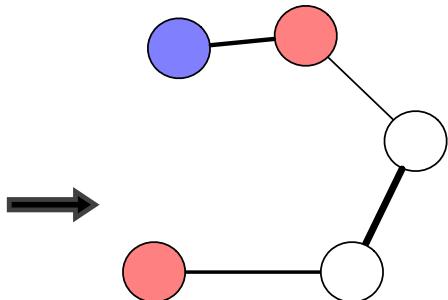
Variant 1)

$$S_{i,j}^l = \text{softmax}(\mathbf{u}^T \tanh(\mathbf{W}[\mathbf{h}_i^l, \mathbf{h}_j^l, \mathbf{v}_i, \mathbf{v}_j, \mathbf{e}_{i,j}]))$$

Edge embeddings



Initial graph (e.g., words, sentences, documents)



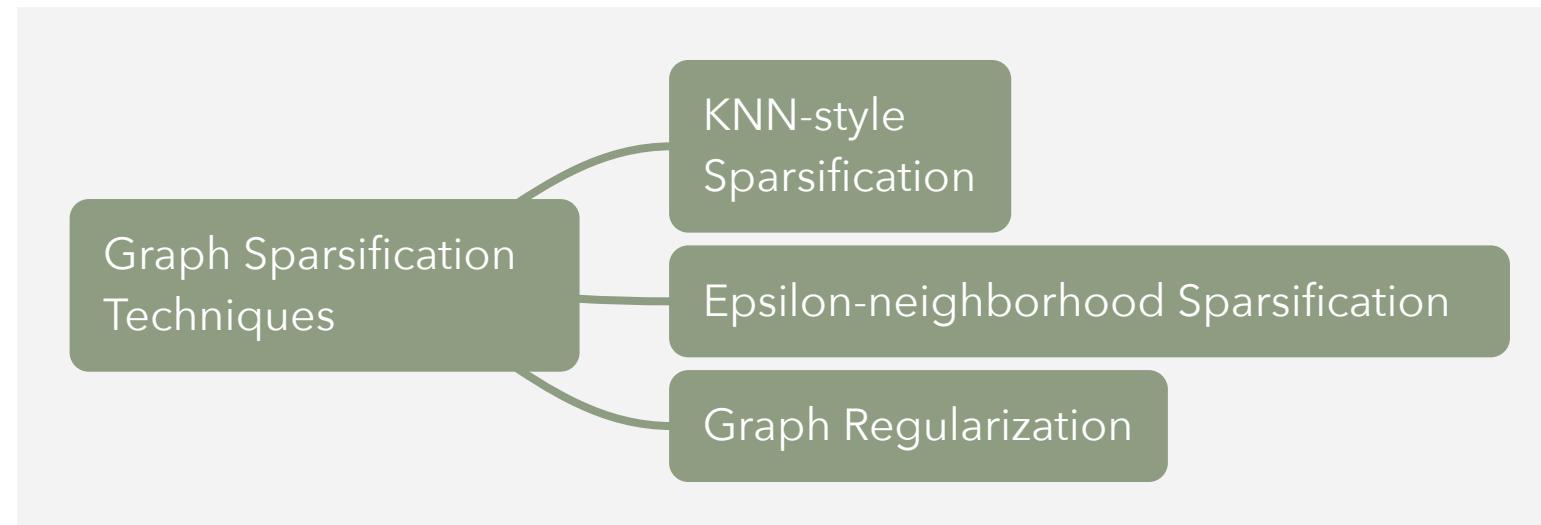
Fully-connected weighted graph

Variant 2)

$$S_{i,j} = \frac{\text{ReLU}(\mathbf{W}^Q \mathbf{v}_i)^T (\text{ReLU}(\mathbf{W}^K \mathbf{v}_i) + \text{ReLU}(\mathbf{W}^R \mathbf{e}_{i,j}))}{\sqrt{d}}$$

Graph Sparsification Techniques

- Similarity metric functions learn a fully-connected graph
- Fully-connected graph is **computationally expensive** and might introduce **noise**
- Enforcing sparsity to the learned graph structure
- Various techniques



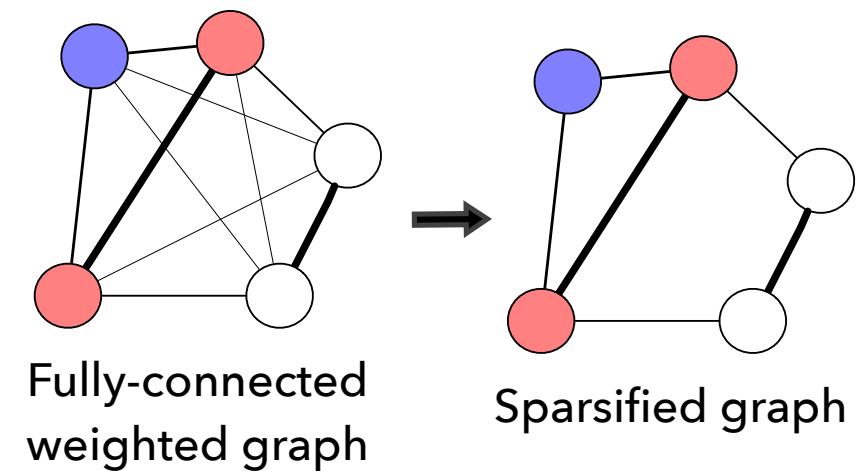
Common Graph Sparsification Options

Option 1) KNN-style Sparsification

$$\mathbf{A}_{i,:} = \text{topk}(\mathbf{S}_{i,:})$$

Option 2) epsilon-neighborhood Sparsification

$$A_{i,j} = \begin{cases} S_{i,j} & S_{i,j} > \varepsilon \\ 0 & \text{otherwise} \end{cases}$$



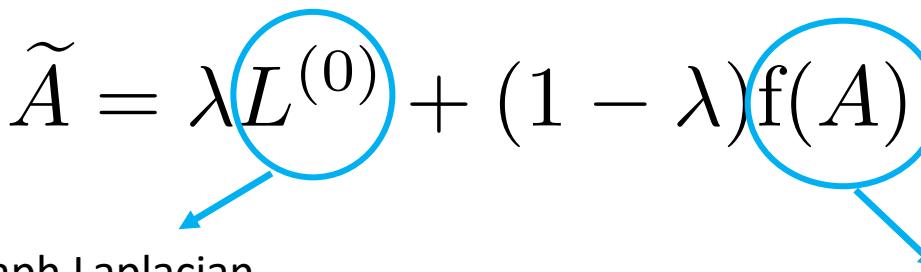
Option 3) graph Regularization

$$\frac{1}{n^2} ||A||_F^2$$

Combining Intrinsic and Implicit Graph Structures

- Intrinsic graph typically still carries rich and useful information
- Learned implicit graph is potentially a “shift” (e.g., substructures) from the intrinsic graph structure

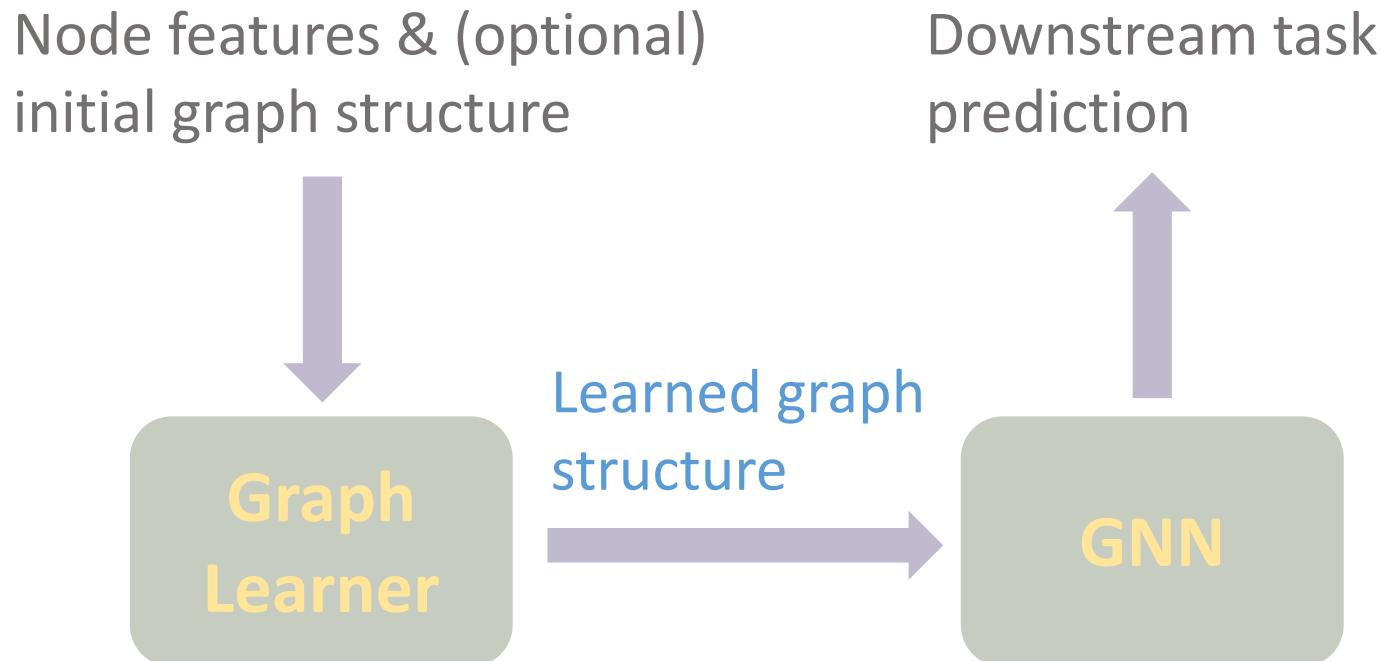
$$\tilde{A} = \lambda L^{(0)} + (1 - \lambda)f(A)$$



Normalized graph Laplacian

$f(A)$ can be arbitrary operation, e.g., graph Laplacian, row-normalization

Learning Paradigms: Joint Learning



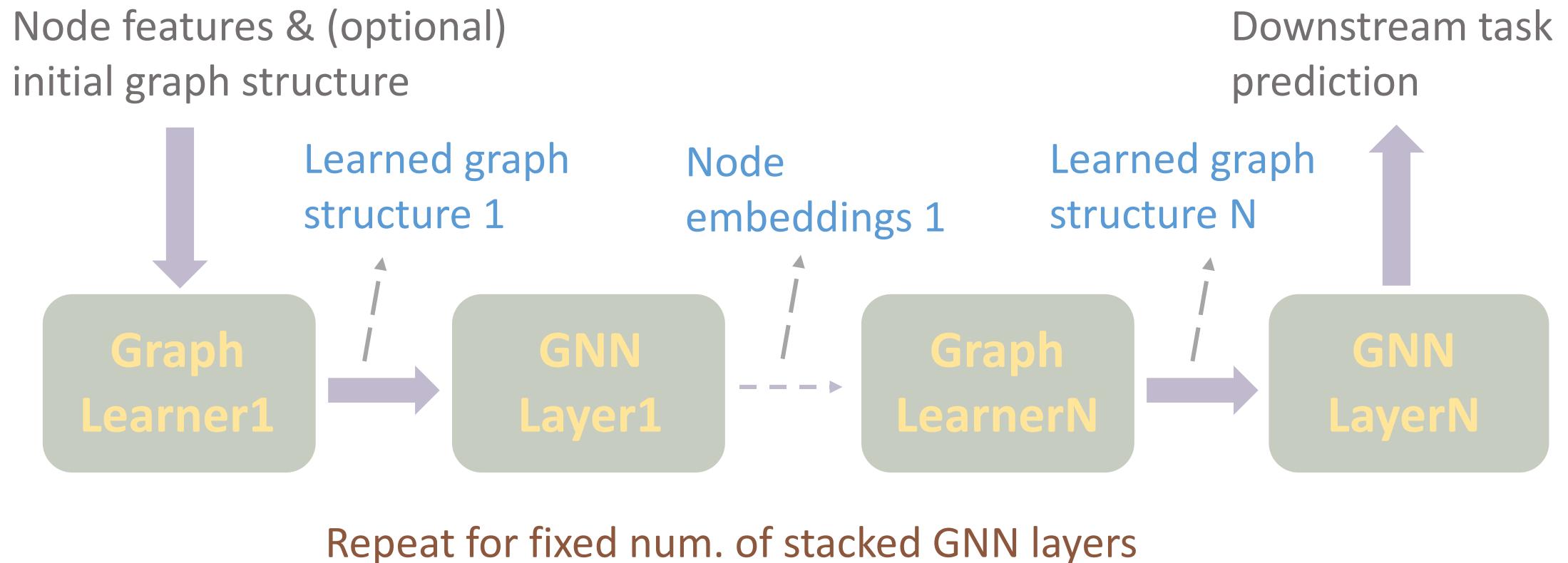
Chen et al. "GraphFlow: Exploiting Conversation Flow with Graph Neural Networks for Conversational Machine Comprehension". IJCAI 2020.

Chen et al. "Reinforcement Learning Based Graph-to-Sequence Model for Natural Question Generation". ICLR 2020.

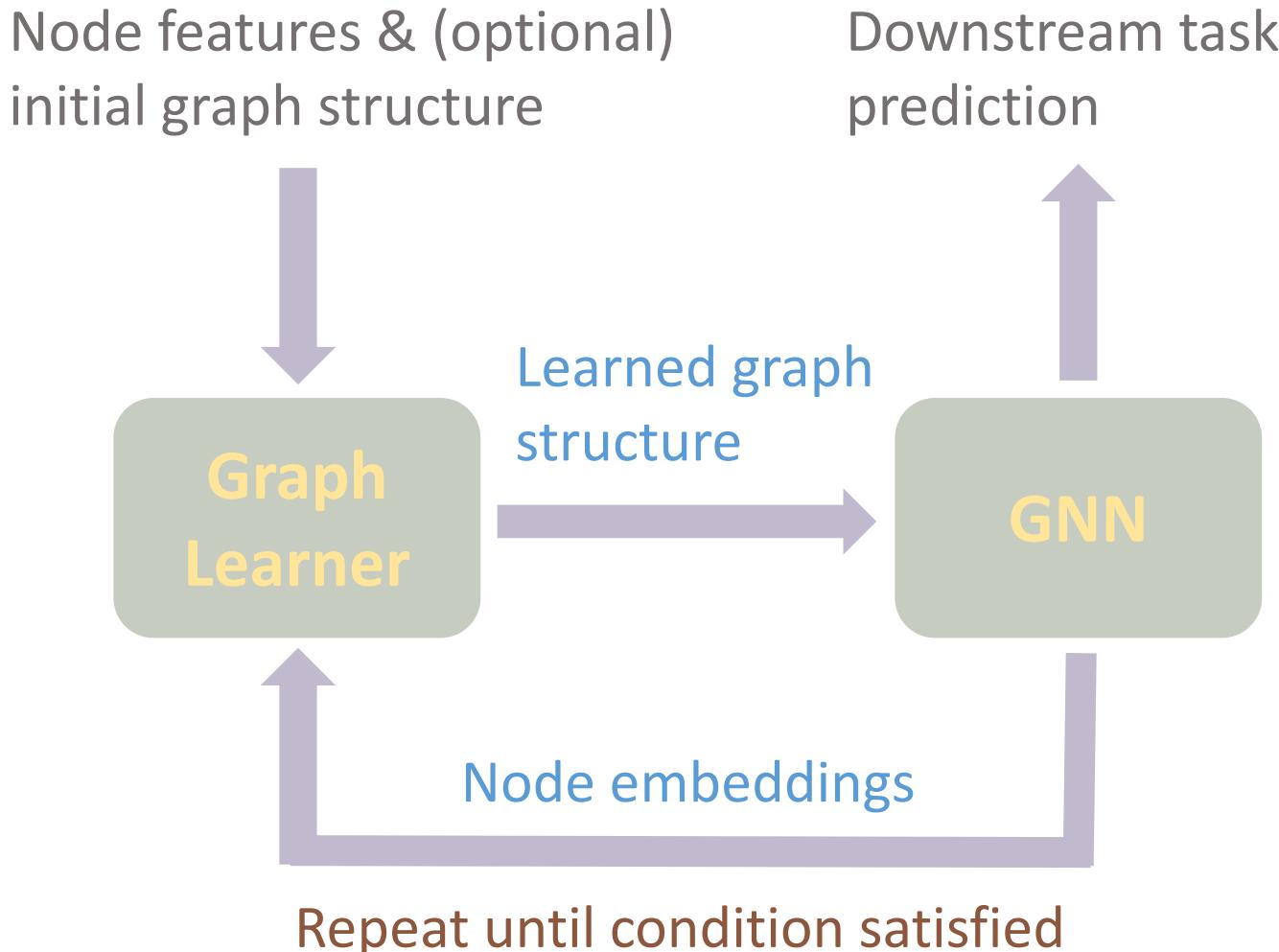
Liu et al. "Contextualized Non-local Neural Networks for Sequence Learning". AAAI 2019.

Liu et al. "Retrieval-Augmented Generation for Code Summarization via Hybrid GNN". ICLR 2021.

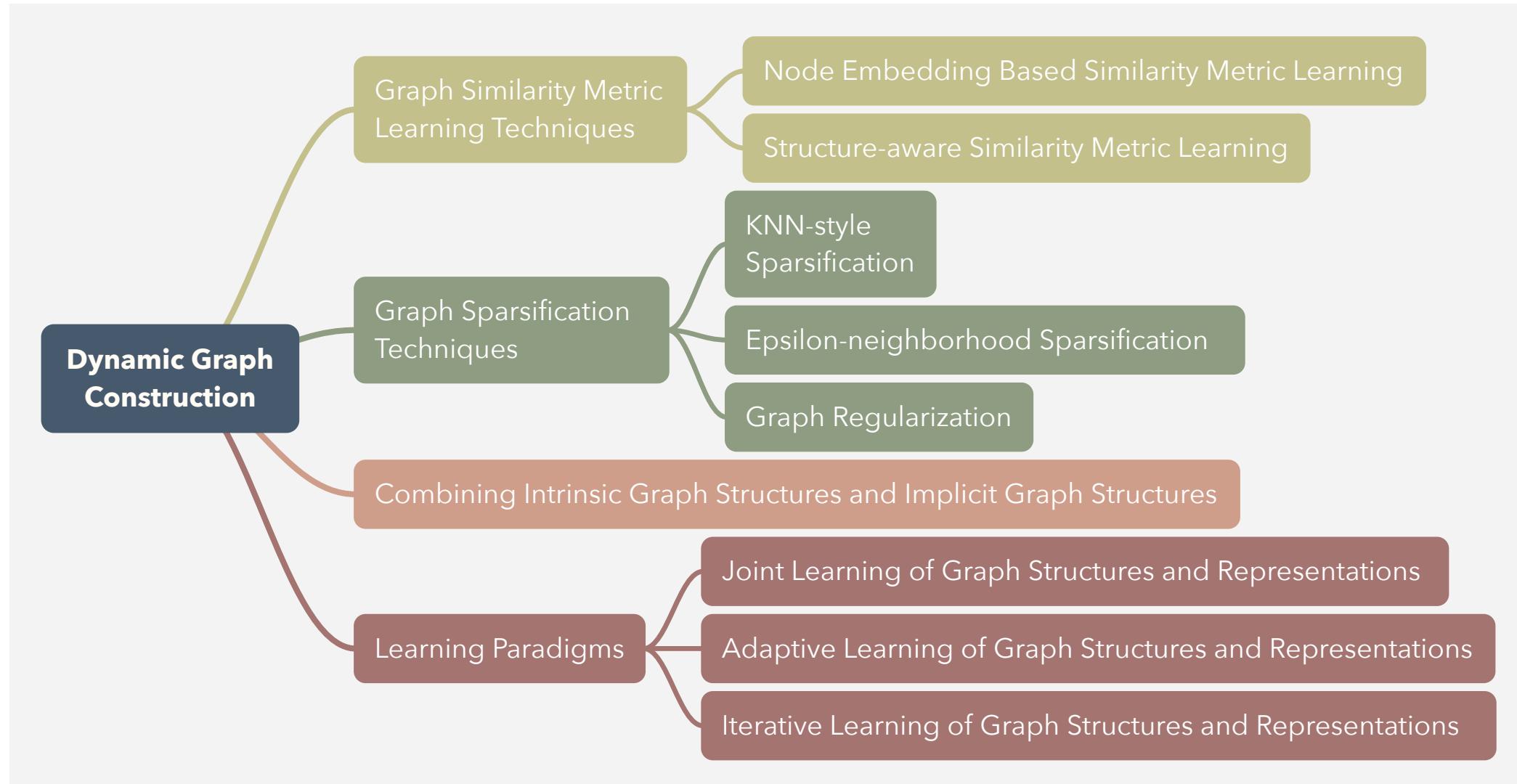
Learning Paradigms: Adaptive Learning



Learning Paradigms: Iterative Learning



Dynamic Graph Construction Summary



Static vs. Dynamic Graph Construction

New topic in DLG4NLP!

Static graph construction	Dynamic graph construction
Pros	Pros
prior knowledge	no domain expertise
	joint graph structure & representation learning
Cons	Cons
extensive domain expertise	scalability
<ul style="list-style-type: none"> • error-prone (e.g., noisy, incomplete) • sub-optimal 	explainability
<ul style="list-style-type: none"> • disjoint graph structure & representation learning • error accumulation 	

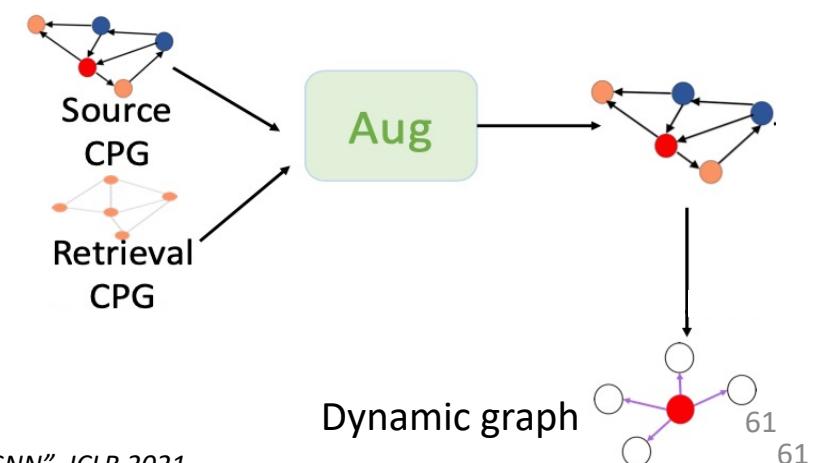
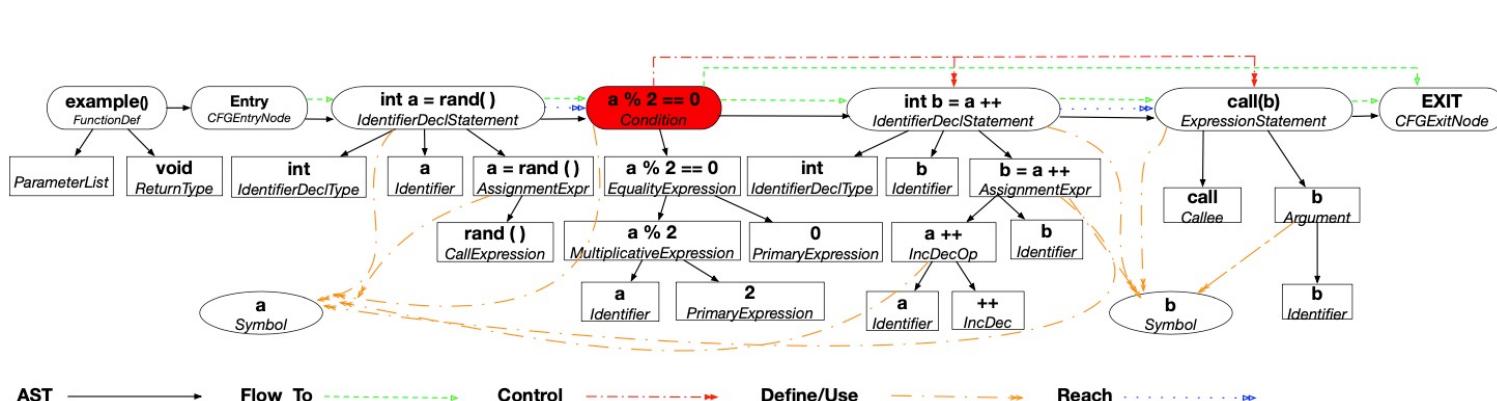
Static vs. Dynamic Graph Construction (cont)

When to use static graph construction

- Domain knowledge which fits the task and can be presented as a graph

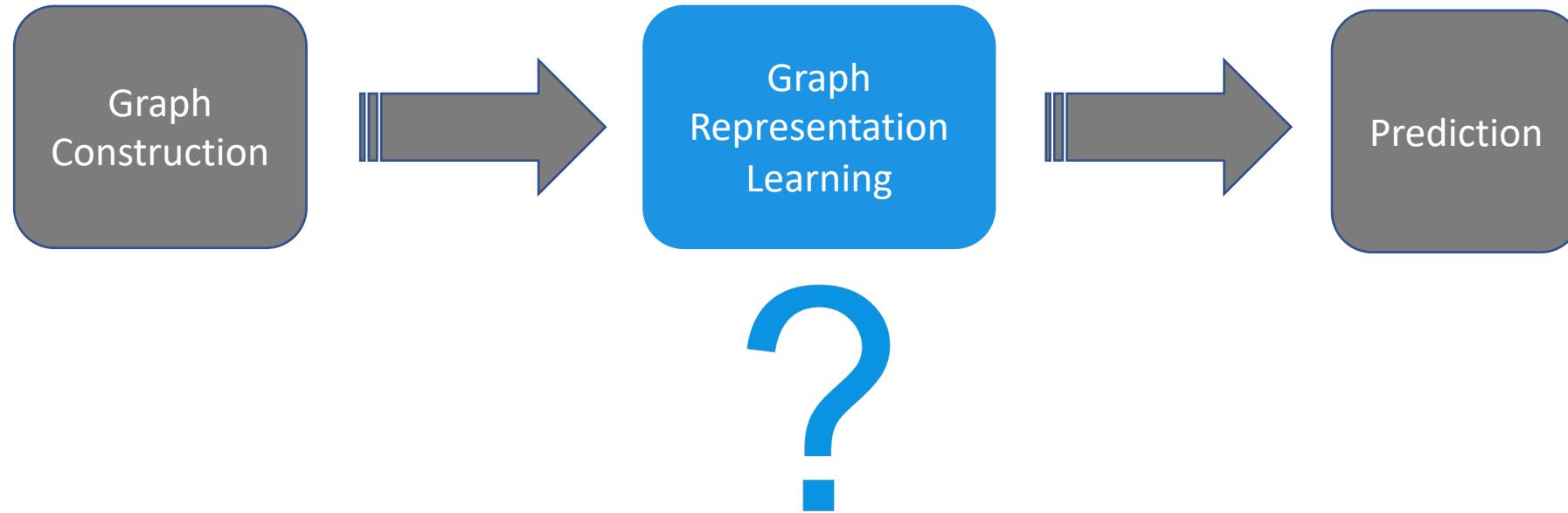
When to use dynamic graph construction

- Lack of domain knowledge which fits the task or can be presented as a graph
- Domain knowledge is incomplete or might contain noise
- To learn implicit graph which augments the static graph



Graph Representation Learning for NLP

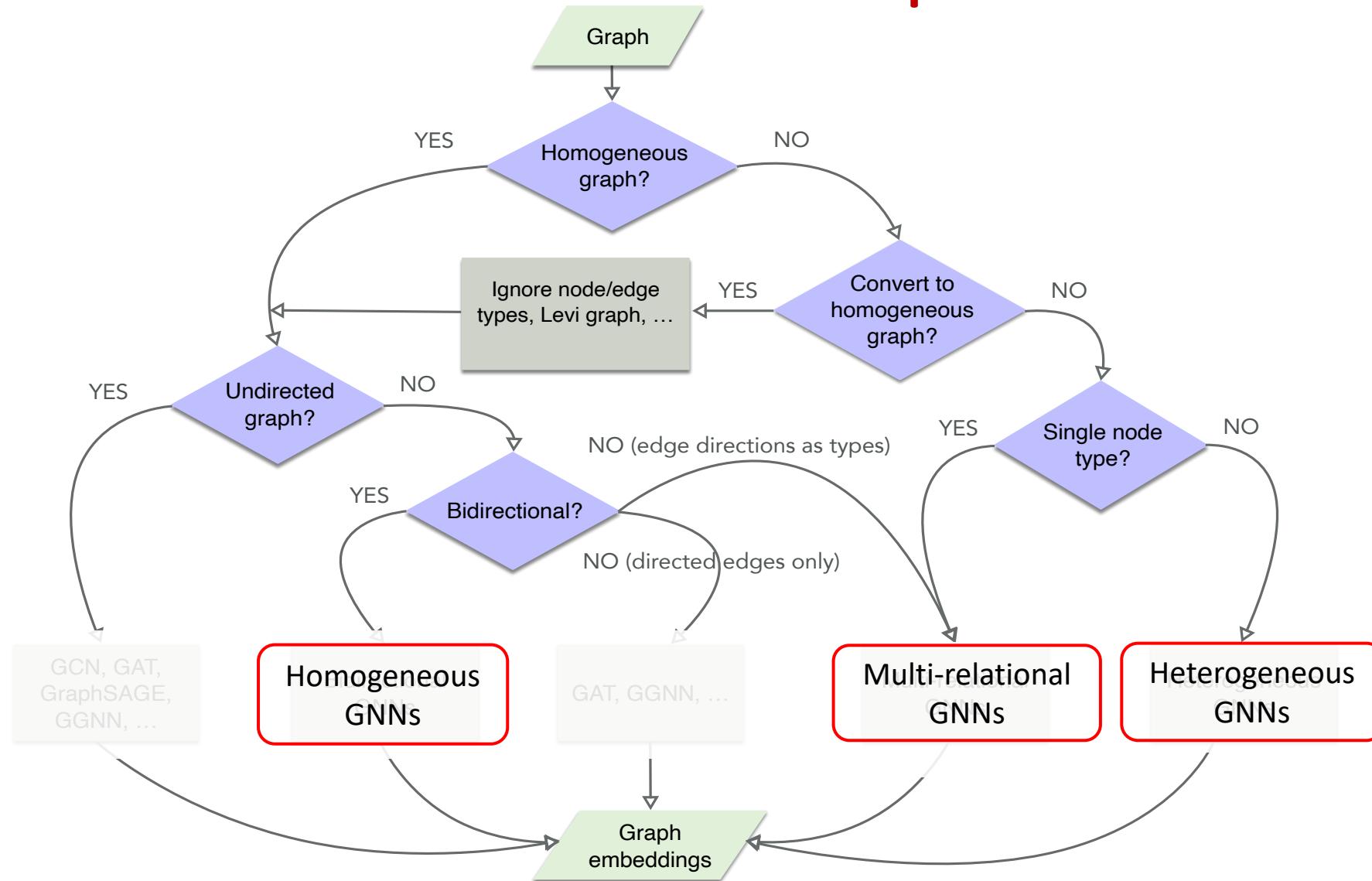
GNNs for Graph Representation Learning



Homogeneous vs Multi-relational vs Heterogeneous Graphs

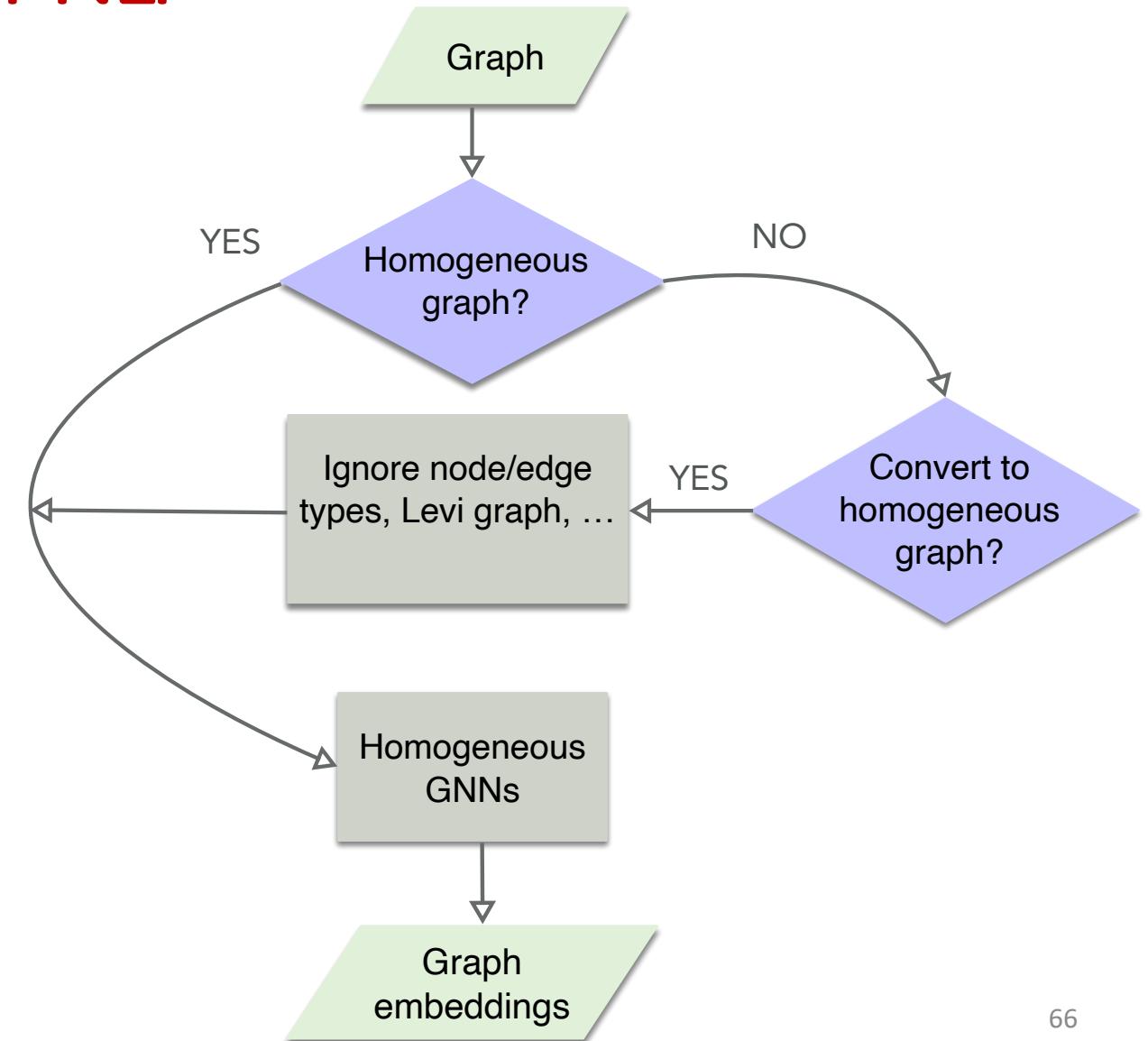
Graph types	Homogeneous	Multi-relational	Heterogeneous
# of node types	1	1	> 1
# of edge types	1	> 1	≥ 1

Which GNNs to Use Given a Graph?

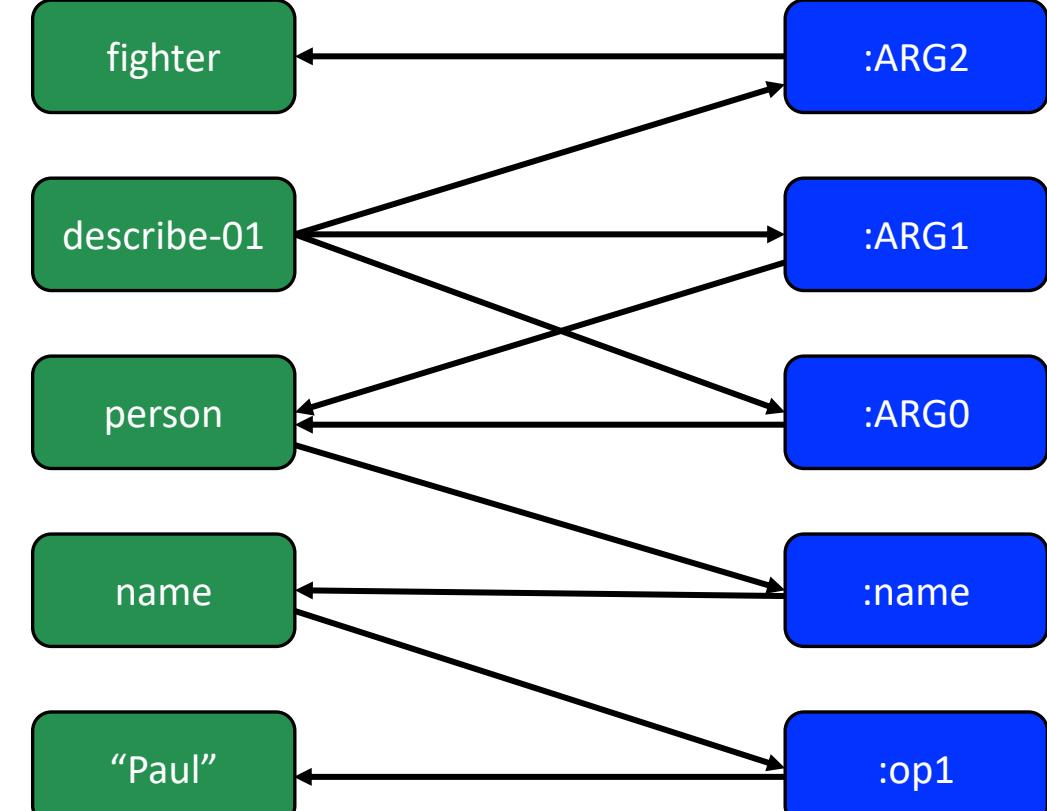
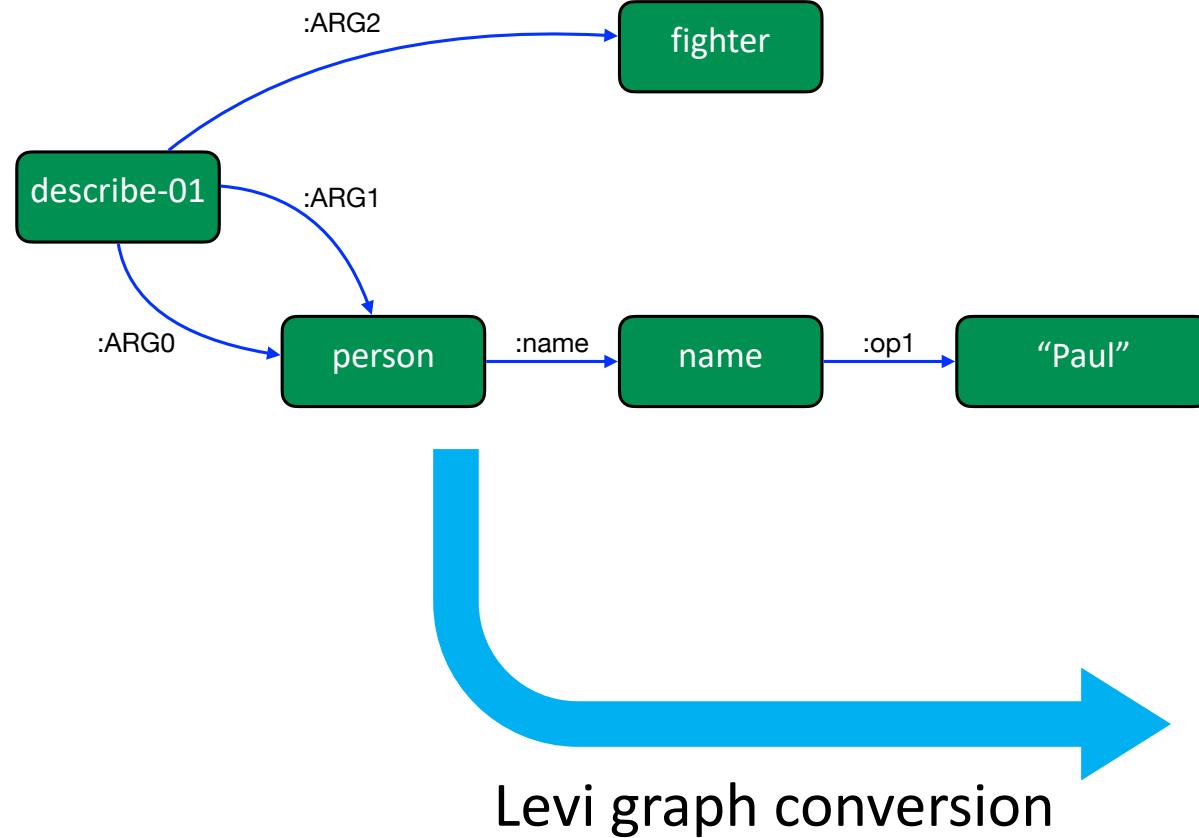


Homogeneous GNNs for NLP

- When to use homogeneous GNNs?
- Homogeneous GNNs
 - GCN
 - GAT
 - GraphSAGE
 - GGNN
 - ...



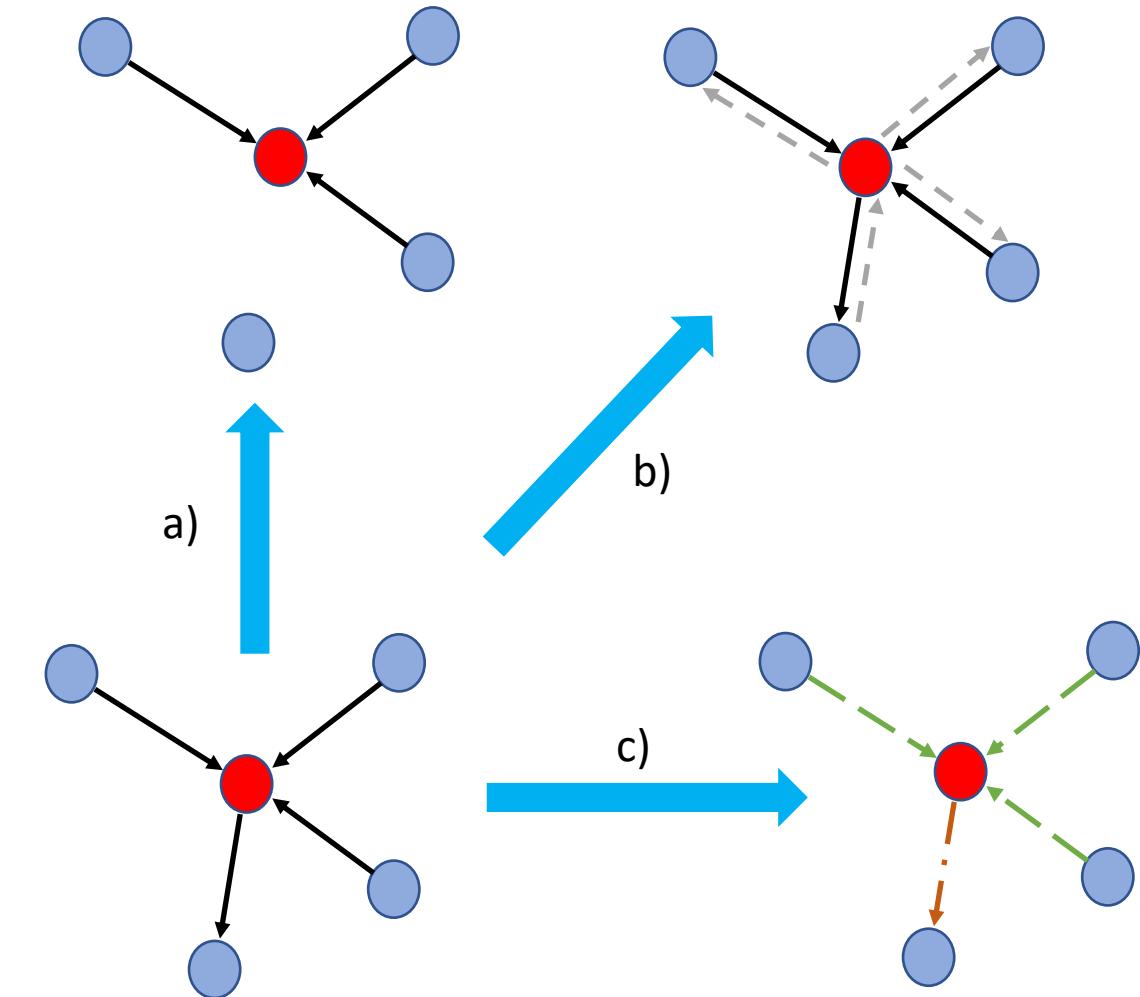
Non-homogeneous to Homogeneous Conversion via Levi Graph



Levi graph: edges as new nodes

How to Handle Edge Direction Information?

- Edge direction is important (think about BiLSTM, BERT)
- Common strategies for handling directed graphs
 - a) Message passing only along directed edges (e.g., GAT, GGNN)
 - b) Regarding edge directions as edge types (i.e., adding “reverse” edges)
 - c) Bidirectional GNNs



Edge Directions as Edge Types

- Regarding edge directions as edge types, resulting in a multi-relational graph

$$dir_{i,j} = \begin{cases} default, & e_{i,j} \text{ is originally existing in the graph} \\ inverse, & e_{i,j} \text{ is the inverse edge} \\ self, & i = j \end{cases}$$

Then we can apply multi-relational GNNs

Bidirectional GNNs for Directed Graphs

Bi-Sep GNNs formulation:

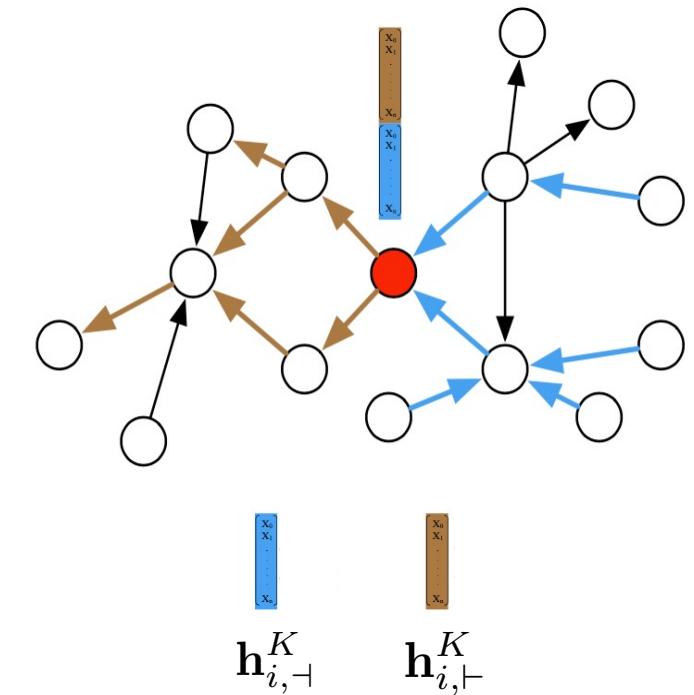
Run multi-hop backward/forward GNN on the graph

$$\mathbf{h}_{i,\dashv}^k = GNN(\mathbf{h}_{i,\dashv}^{k-1}, \{\mathbf{h}_{j,\dashv}^{k-1} : \forall v_j \in \mathcal{N}_{\dashv}(v_i)\})$$

$$\mathbf{h}_{i,\vdash}^k = GNN(\mathbf{h}_{i,\vdash}^{k-1}, \{\mathbf{h}_{j,\vdash}^{k-1} : \forall v_j \in \mathcal{N}_{\vdash}(v_i)\})$$

Concatenate backward/forward node embeddings at last hop

$$\mathbf{h}_i^K = \mathbf{h}_{i,\dashv}^K || \mathbf{h}_{i,\vdash}^K$$



Bidirectional GNNs for Directed Graphs (cont)

Bi-Fuse GNNs formulation:

Run one-hop backward/forward node aggregation

$$\mathbf{h}_{\mathcal{N}_-(v_i)}^k = AGG(\mathbf{h}_i^{k-1}, \{\mathbf{h}_j^{k-1} : \forall v_j \in \mathcal{N}_-(v_i)\})$$

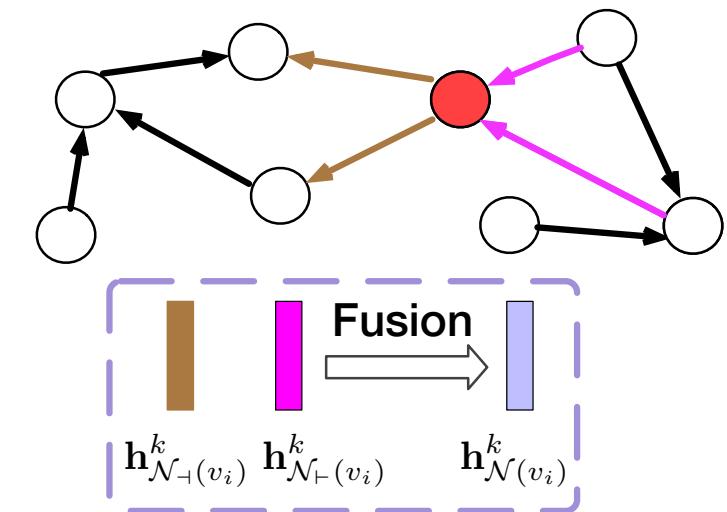
$$\mathbf{h}_{\mathcal{N}_+(v_i)}^k = AGG(\mathbf{h}_i^{k-1}, \{\mathbf{h}_j^{k-1} : \forall v_j \in \mathcal{N}_+(v_i)\})$$

Fuse backward/forward aggregation vectors at each hop

$$\mathbf{h}_{\mathcal{N}(v_i)}^k = Fuse(\mathbf{h}_{\mathcal{N}_-(v_i)}^k, \mathbf{h}_{\mathcal{N}_+(v_i)}^k)$$

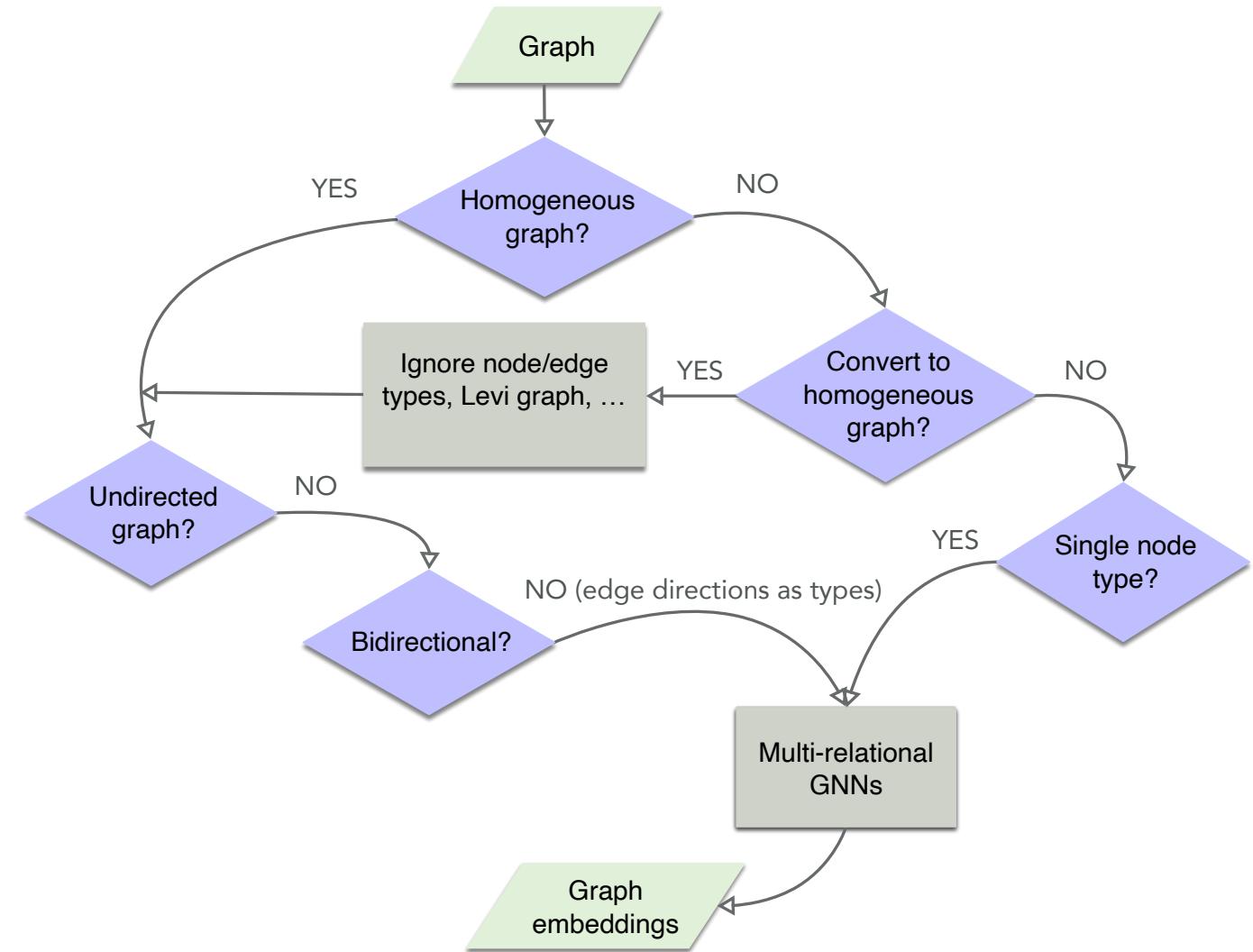
Update node embeddings with fused aggregation vectors at each hop

$$\mathbf{h}_i^k = \sigma(\mathbf{h}_i^{k-1}, \mathbf{h}_{\mathcal{N}(v_i)}^k)$$



Multi-relational GNNs for NLP

- When to use multi-relational GNNs?
- Multi-relational GNNs
 - Including relation-specific transformation parameters in GNN
 - Including edge embeddings in GNN
 - Multi-relational Graph Transformers



R-GNN: Overview

$$\mathbf{h}_i^k = \sigma(\mathbf{h}_i^{k-1}, \sum_{v_j \in \mathcal{N}(v_i)} AGG(\mathbf{h}_j^{k-1}, \theta^k))$$

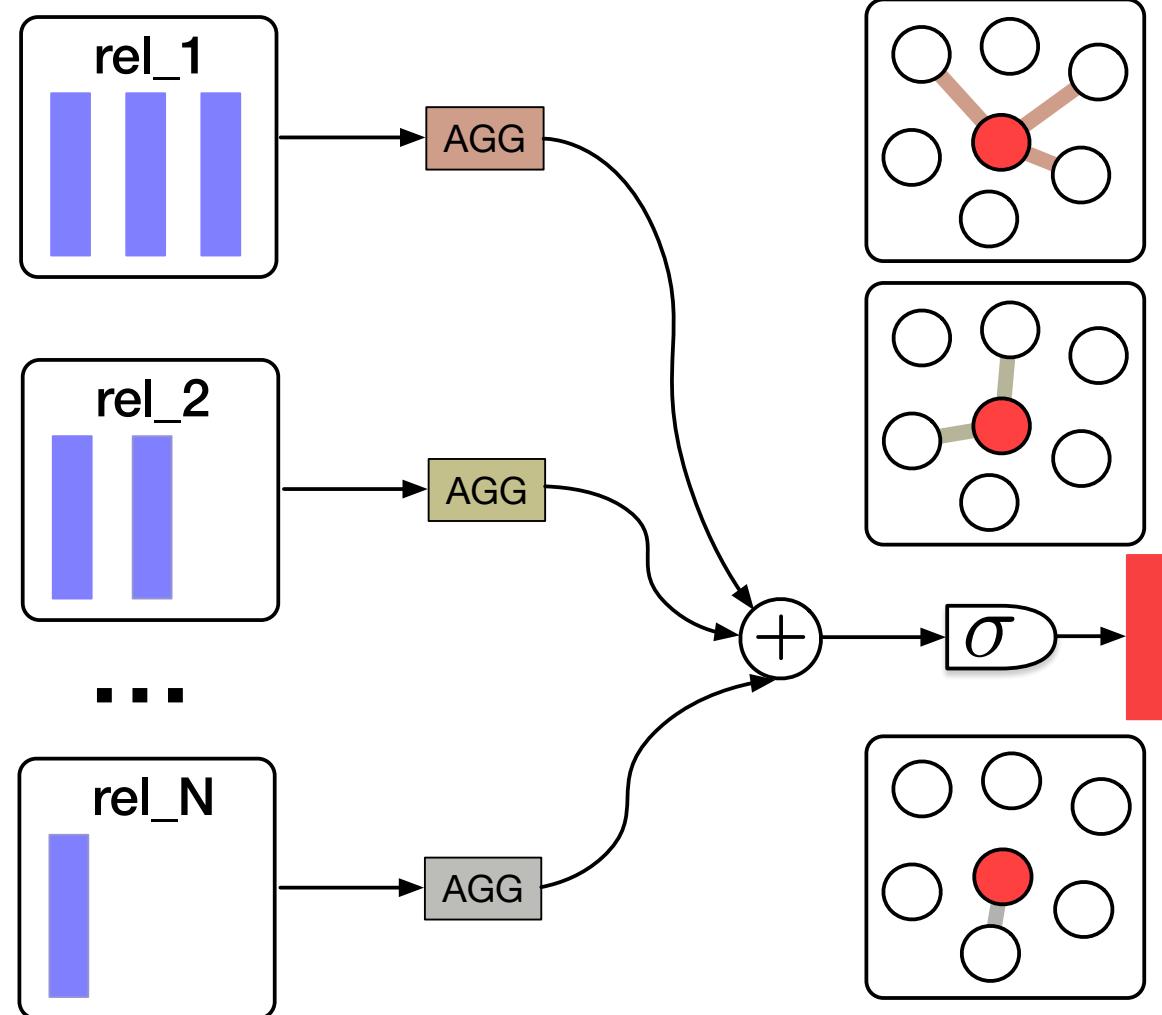
GNN

R-GNN

1) relation-specific transformation,
e.g., node feature transformation,
attention weight ...

$$\mathbf{h}_i^k = \sigma(\mathbf{h}_i^{k-1}, \sum_{r \in \mathcal{E}} \sum_{v_j \in \mathcal{N}_r(v_i)} AGG(\mathbf{h}_j^{k-1}, \theta_r^k))$$

2) aggregation per relation-specific subgraph



R-GNN Variant: R-GCN

- Relation-specific node feature transformation during neighborhood aggregation

$$\mathbf{h}_i^k = \sigma \left(\sum_{r \in \mathcal{E}} \sum_{v_j \in \mathcal{N}_r(v_i)} \frac{1}{c_{i,r}} \mathbf{W}_r^k \mathbf{h}_j^{k-1} + \mathbf{W}_0^k \mathbf{h}_i^{k-1} \right), \quad c_{i,r} = |\mathcal{N}_r(v_i)|$$



 Relation-specific $d \times d$ learnable weight matrix

R-GNN: Avoiding Over-parameterization

Learning $d \times d$ transformation weight matrix for each relation is expensive!

$O(Rd^2)$ parameters every GNN layer where R is the num of relation types

How to avoid over-parameterization?

Option 1) basis decomposition - linear hypothesis

$$\theta_r^k = \sum_{b=1}^B a_{rb}^k \mathbf{V}_b^{(k)}, \quad \mathbf{V}_b^{(k)} \in \mathbb{R}^{d \times d} \quad O(RB + Bd^2) \text{ parameters}$$

 Basis matrices

Option 2) block-diagonal decomposition - sparsity hypothesis

$$\theta_r^k = \bigoplus_{b=1}^B \mathbf{Q}_{br}^k = \text{diag}(\mathbf{Q}_{1r}^k, \mathbf{Q}_{2r}^k, \dots, \mathbf{Q}_{Br}^k), \quad \mathbf{Q}_{br}^{(k)} \in \mathbb{R}^{d/B \times d/B} \quad O(Rd^2/B) \text{ parameters}$$

 Submatrices

Including Edge Embeddings in GNNs

Variant 1) Include edge embeddings in message passing

$$\mathbf{h}_i^k = \sigma(\mathbf{h}_i^{k-1}, \sum_{v_j \in \mathcal{N}(v_i)} AGG(\mathbf{h}_j^{k-1}, \mathbf{e}_{i,j}, \theta^k))$$

Edge embeddings

Variant 2) Update edge embedding in message passing

$$\mathbf{h}_i^k = \sigma(\mathbf{h}_i^{k-1}, \sum_{v_j \in \mathcal{N}(v_i)} AGG(\mathbf{h}_j^{k-1}, \mathbf{e}_{i,j}^{k-1}, \theta^k)), \quad \mathbf{e}_{i,j}^k = f(\mathbf{e}_{i,j}^{k-1}, \theta_{rel}^k)$$

Update edge embeddings

Multi-relational Graph Transformers

- Transformers as a special class of GNNs which
 - jointly learn and encode a **fully-connected graph** via self-attention
 - share many similarities with GAT
 - fail to effectively handle **arbitrary graph-structured data**
 - e.g., position embeddings for sequential data, removing position embeddings for set
- Multi-relational graph transformers
 - employed with **structure-aware self-attention**
 - respect **various relation types**

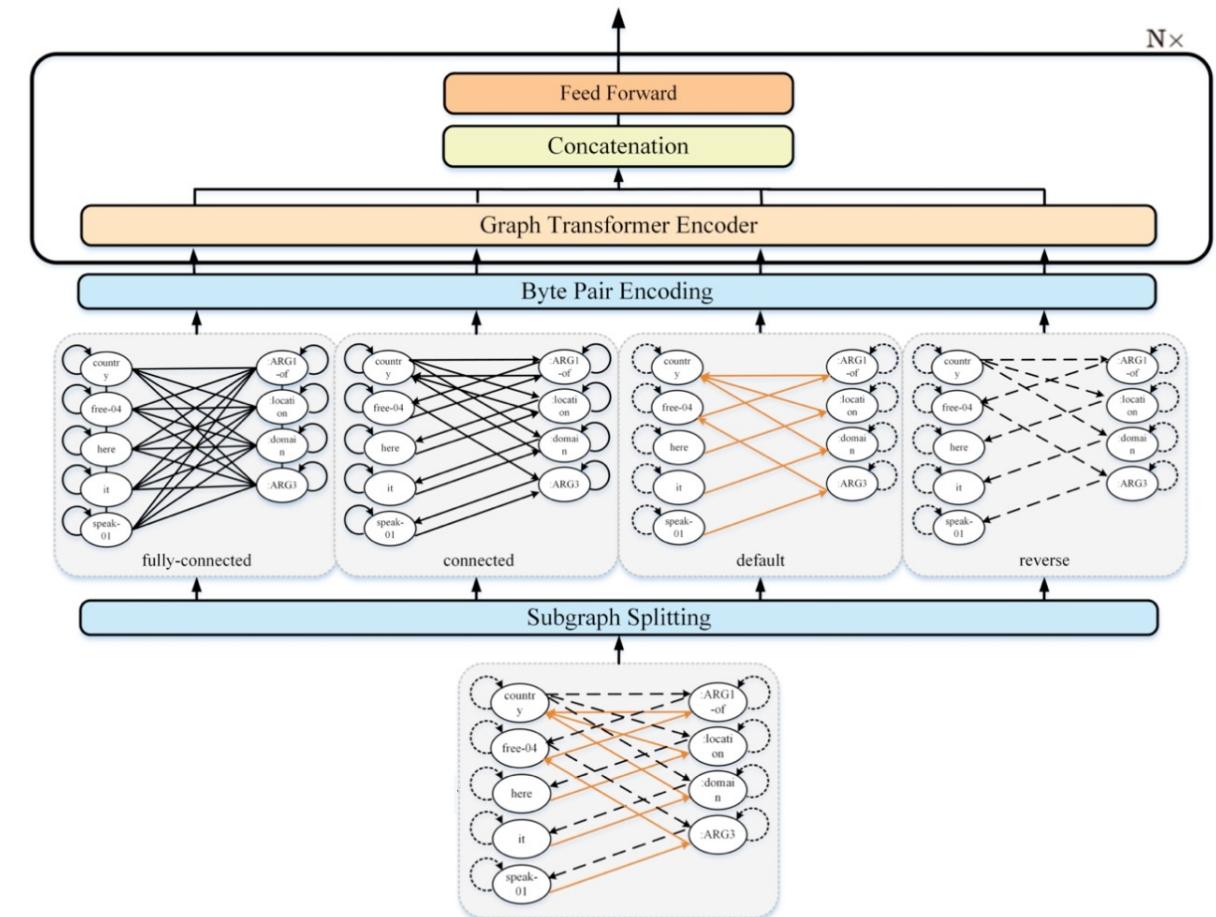
R-GAT based Graph Transformers

GAT-like masked attention

$$\mathbf{z}_i^{r,k} = \sum_{v_j \in \mathcal{N}_r(v_i)} \alpha_{i,j}^k \mathbf{W}_V^k \mathbf{h}_j^{k-1}, r \in \mathcal{E}$$

$$\mathbf{h}_i^k = \text{FFN}^k(\mathbf{W}_O^k [\mathbf{z}_i^{R_1,k}, \dots, \mathbf{z}_i^{R_m,k}])$$

Relation-specific learnable weight matrix



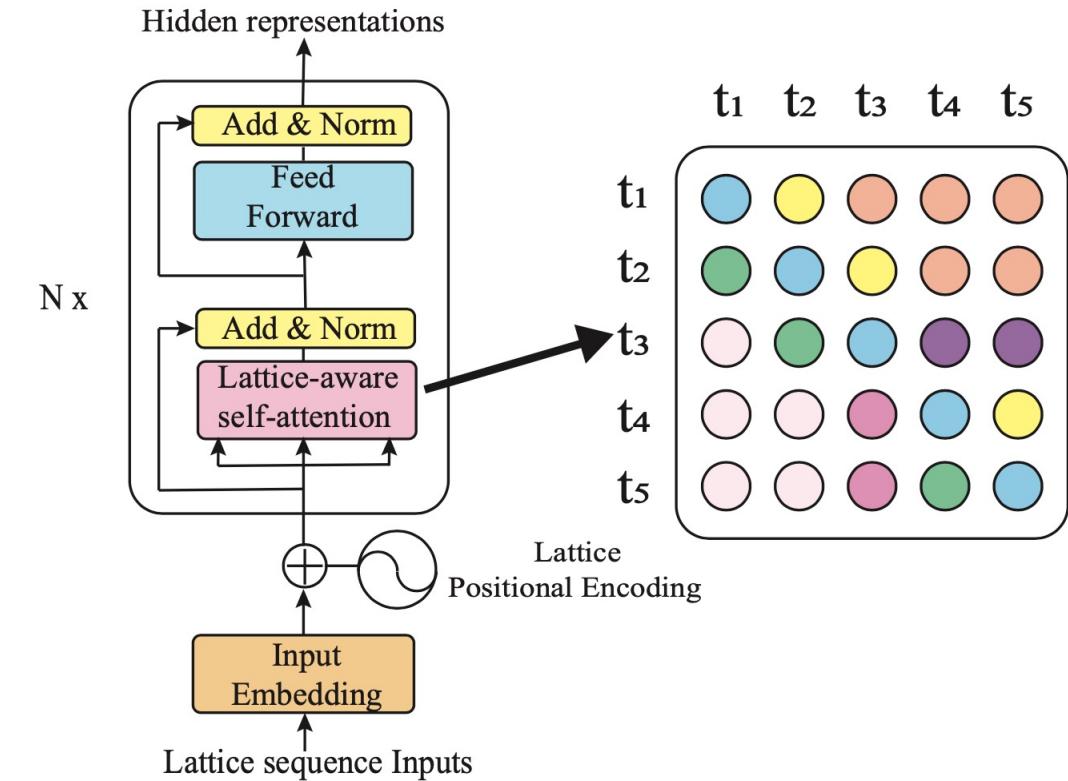
Structure-aware Self-attention based Graph Transformers

$$\mathbf{h}_i^k = \sum_j \alpha_{i,j}^k (\mathbf{W}_V^k \mathbf{h}_j^{k-1} + \mathbf{W}_F^k \mathbf{e}_{i,j})$$

Edge embeddings

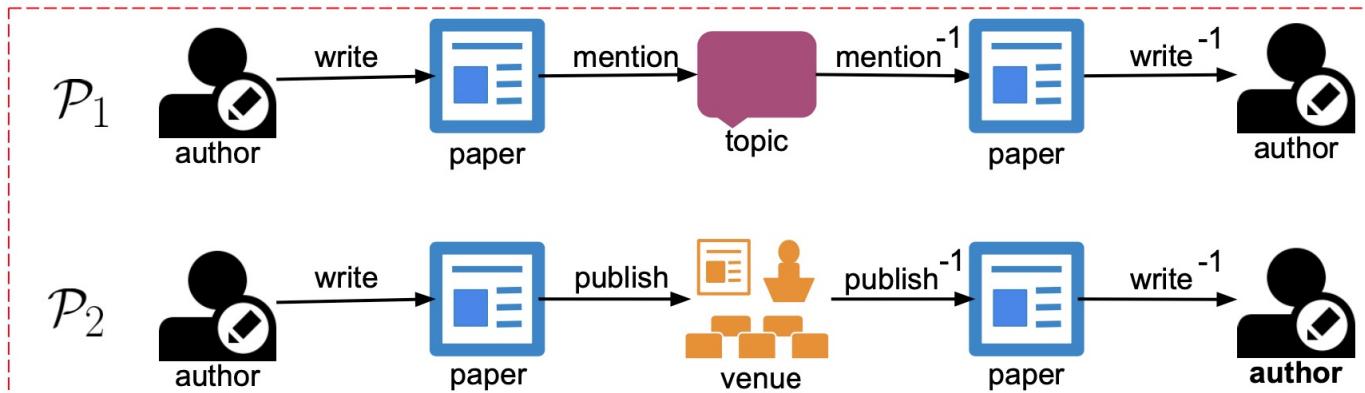
$$\alpha_{i,j}^k = \text{softmax}(u_{i,j}^k)$$

$$u_{i,j}^k = \frac{(\mathbf{W}_Q^k \mathbf{h}_i^{k-1})^T (\mathbf{W}_K^k \mathbf{h}_j^{k-1} + \mathbf{W}_R^k \mathbf{e}_{i,j})}{\sqrt{d}}$$

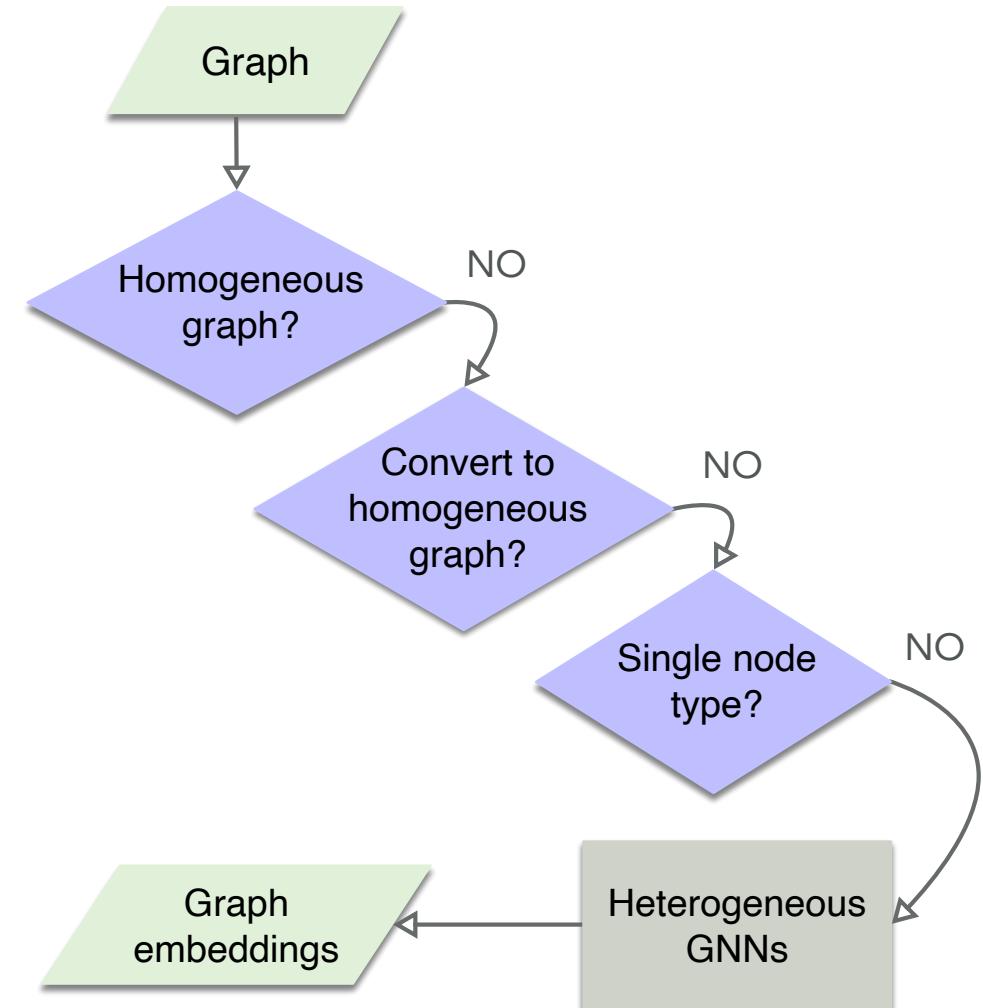


Heterogeneous GNNs

- When to use Heterogeneous GNNs?
- Heterogeneous GNNs
 - a) Meta-path based Heterogeneous GNNs



Meta paths among author nodes



Meta-path based Heterogeneous GNN example: HAN

Step 1) type-specific node feature transformation

$$\mathbf{h}_i = \mathbf{W}_{\tau(v_i)} \mathbf{v}_i$$

Node-type specific learnable weight matrix

Step 2) node-level aggregation along each meta path

$$\mathbf{z}_{i,\Phi_k} = \sigma \left(\sum_{v_j \in \mathcal{N}_{\Phi_k}(v_i)} \alpha_{i,j}^{\Phi_k} \mathbf{h}_j \right)$$

Aggregate over neighboring nodes
in k-length meta path

Step 3) meta-path level aggregation

$$\mathbf{z}_i = \sum_{k=1}^p \beta_{\Phi_k} \mathbf{z}_{i,\Phi_k}$$

Attention weights over meta paths

Graph Encoder-Decoder Models for NLP

Seq2Seq: Applications and Challenges

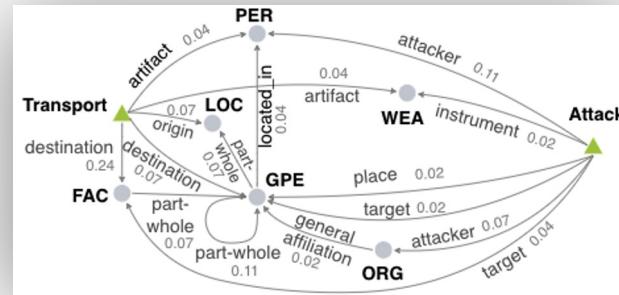
- Applications

- Machine translation
- Natural language generation
- Logic form translation
- Information extraction

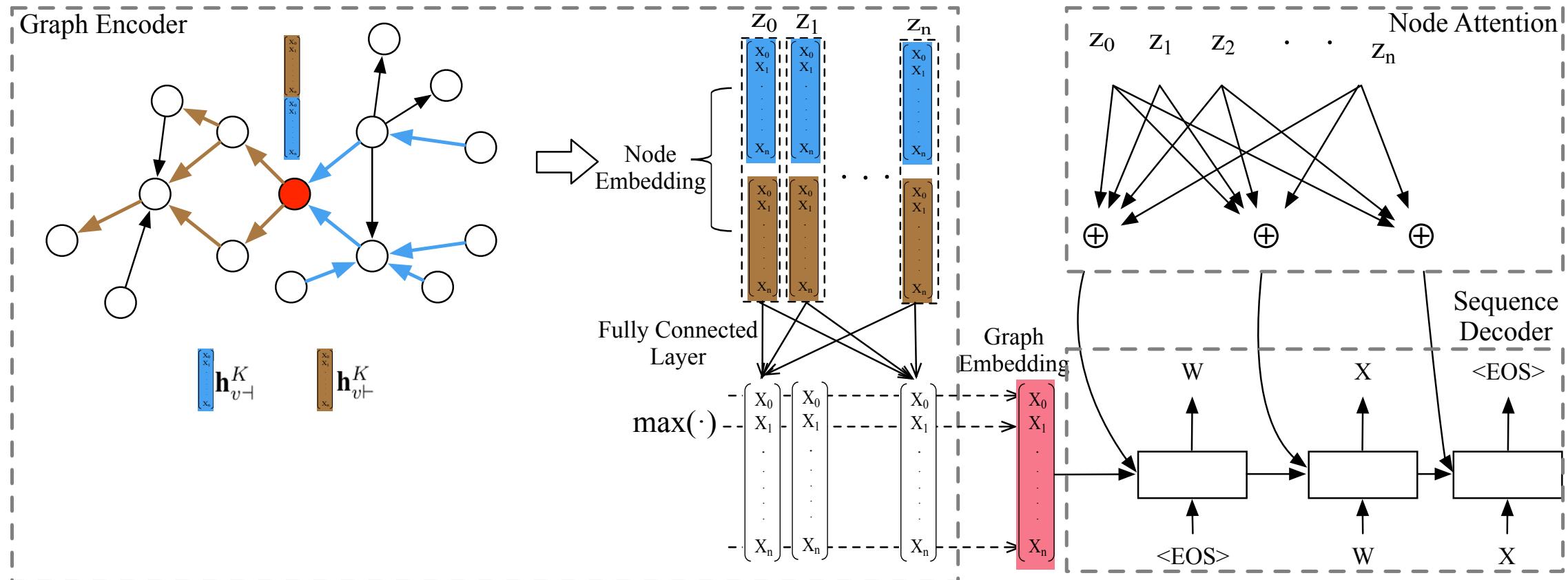


- Challenges

- Only applied to problems whose inputs are represented as sequences
- Cannot handle more complex structure such as graphs
- Converting graph inputs into sequences inputs lose information
- Augmenting original sequence inputs with additional structural information enhances word sequence feature



Graph-to-Sequence Model



- [1] Kun Xu*, Lingfei Wu*, Zhiguo Wang, Yansong Feng, Michael Witbrock, and Vadim Sheinin (Equally Contributed), "Graph2Seq: Graph to Sequence Learning with Attention-based Neural Networks", arXiv 2018.
- [2] Yu Chen, Lingfei Wu** and Mohammed J. Zaki (**Corresponding Author), "Reinforcement Learning Based Graph-to-Sequence Model for Natural Question Generation", ICLR'20.

Bidirectional GNNs for Directed Graphs

Bi-Sep GNNs formulation:

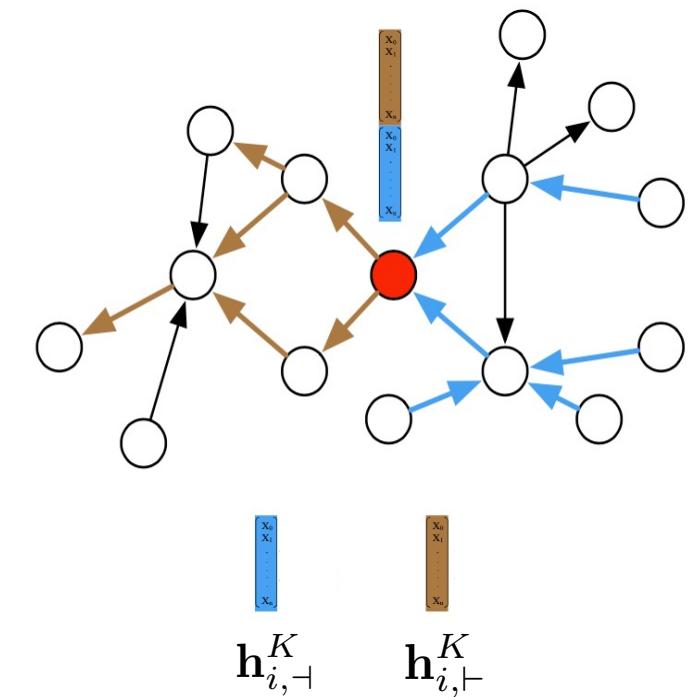
Run multi-hop backward/forward GNN on the graph

$$\mathbf{h}_{i,\dashv}^k = GNN(\mathbf{h}_{i,\dashv}^{k-1}, \{\mathbf{h}_{j,\dashv}^{k-1} : \forall v_j \in \mathcal{N}_{\dashv}(v_i)\})$$

$$\mathbf{h}_{i,\vdash}^k = GNN(\mathbf{h}_{i,\vdash}^{k-1}, \{\mathbf{h}_{j,\vdash}^{k-1} : \forall v_j \in \mathcal{N}_{\vdash}(v_i)\})$$

Concatenate backward/forward node embeddings at last hop

$$\mathbf{h}_i^K = \mathbf{h}_{i,\dashv}^K || \mathbf{h}_{i,\vdash}^K$$



Bidirectional GNNs for Directed Graphs (cont)

Bi-Fuse GNNs formulation:

Run one-hop backward/forward node aggregation

$$\mathbf{h}_{\mathcal{N}_-(v_i)}^k = AGG(\mathbf{h}_i^{k-1}, \{\mathbf{h}_j^{k-1} : \forall v_j \in \mathcal{N}_-(v_i)\})$$

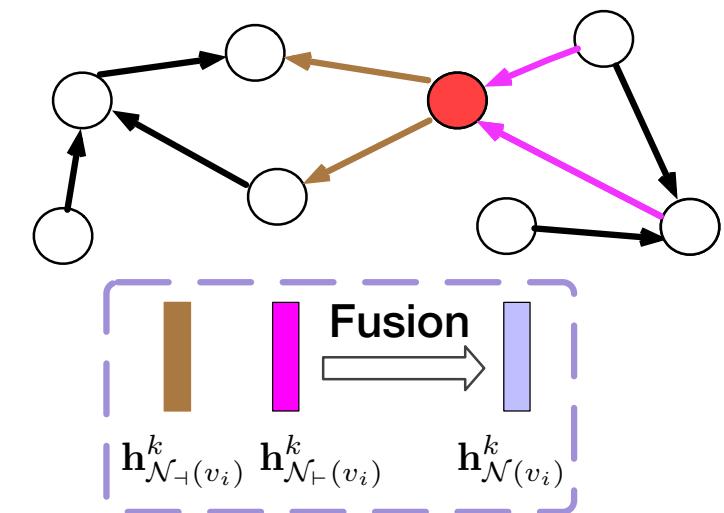
$$\mathbf{h}_{\mathcal{N}_+(v_i)}^k = AGG(\mathbf{h}_i^{k-1}, \{\mathbf{h}_j^{k-1} : \forall v_j \in \mathcal{N}_+(v_i)\})$$

Fuse backward/forward aggregation vectors at each hop

$$\mathbf{h}_{\mathcal{N}(v_i)}^k = Fuse(\mathbf{h}_{\mathcal{N}_-(v_i)}^k, \mathbf{h}_{\mathcal{N}_+(v_i)}^k)$$

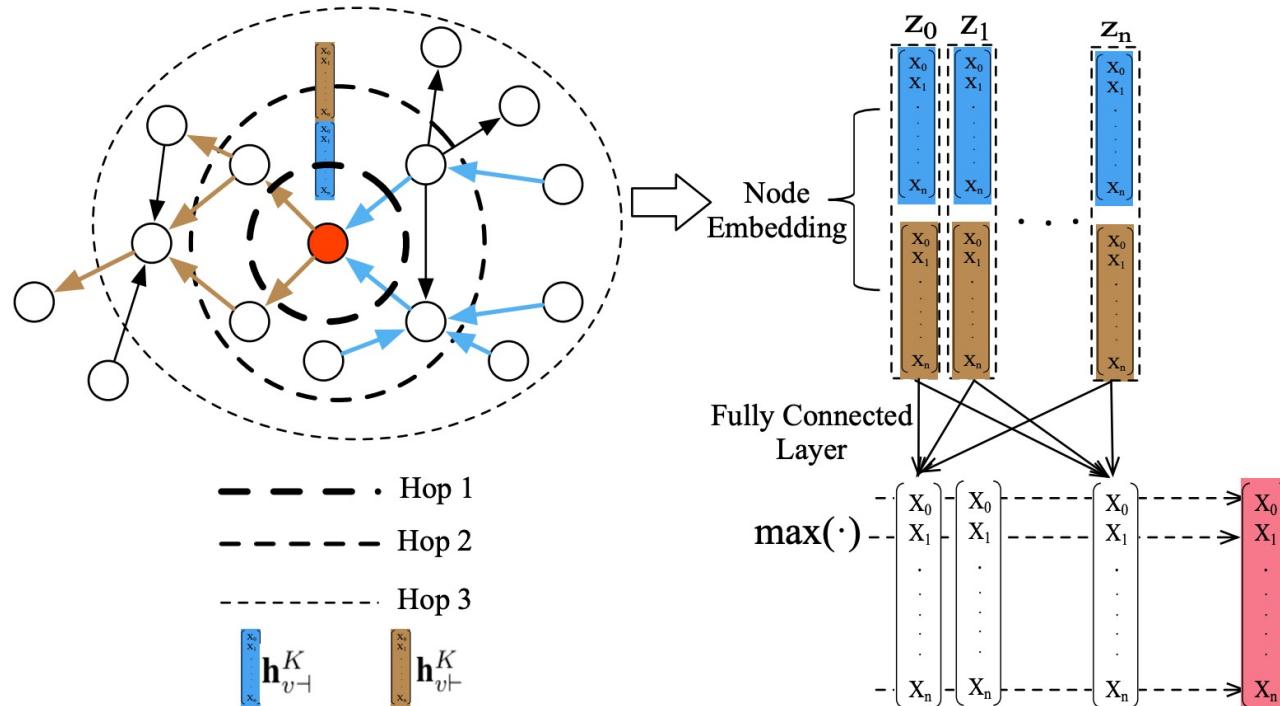
Update node embeddings with fused aggregation vectors at each hop

$$\mathbf{h}_i^k = \sigma(\mathbf{h}_i^{k-1}, \mathbf{h}_{\mathcal{N}(v_i)}^k)$$



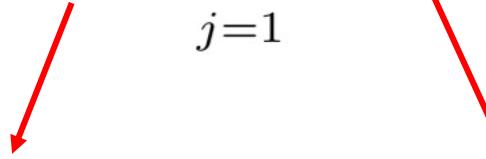
Graph Encoding

- Graph embedding
 - Pooling based graph embedding (*max, min and average pooling*)
 - Node based graph embedding
 - Add one super node which is connected to all other nodes in the graph
 - The embedding of this super node is treated as graph embedding



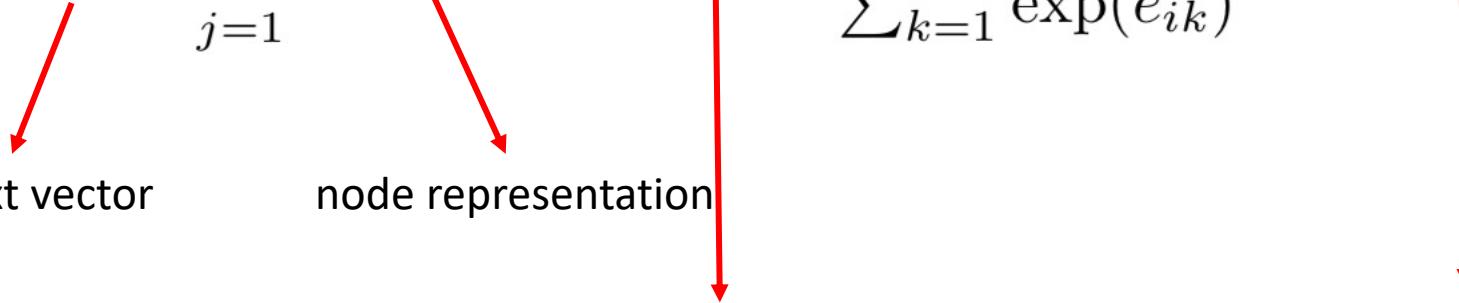
Attention Based Sequence Decoding

$$c_i = \sum_{j=1}^{\mathcal{V}} \alpha_{ij} h_j, \text{ where } \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{\mathcal{V}} \exp(e_{ik})}, e_{ij} = a(s_{i-1}, h_j)$$


context vector node representation

Attention Based Sequence Decoding

$$c_i = \sum_{j=1}^{\mathcal{V}} \alpha_{ij} h_j, \text{ where } \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{\mathcal{V}} \exp(e_{ik})}, e_{ij} = a(s_{i-1}, h_j)$$


 context vector → $\sum_{j=1}^{\mathcal{V}} \alpha_{ij} h_j$
 node representation → α_{ij}
 alignment model → $a(s_{i-1}, h_j)$
 attention weights → $\sum_{k=1}^{\mathcal{V}} \exp(e_{ik})$

Attention Based Sequence Decoding

$$c_i = \sum_{j=1}^{\mathcal{V}} \alpha_{ij} h_j, \text{ where } \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{\mathcal{V}} \exp(e_{ik})}, e_{ij} = a(s_{i-1}, h_j)$$

↓ context vector node representation ↓
↓ attention weights ↓ alignment model

- Objective Function

$$\theta^* = \arg \max_{\theta} \sum_{n=1}^N \sum_{t=1}^{T_n} \log p(y_t^n | y_{<t}^n, x^n)$$

Text Reasoning and Shortest Path

garden (A) bathroom (B) bedroom (C)
hallway (D) office (E) kitchen (F)

- 1 The **garden** is west of the **bathroom**.
- 2 The **bedroom** is north of the **hallway**.
- 3 The **office** is south of the **hallway**.
- 4 The **bathroom** is north of the **bedroom**.
- 5 The **kitchen** is east of the **bedroom**.

Transform

A	west	B
B	north	D
E	south	D
B	north	C
F	east	C

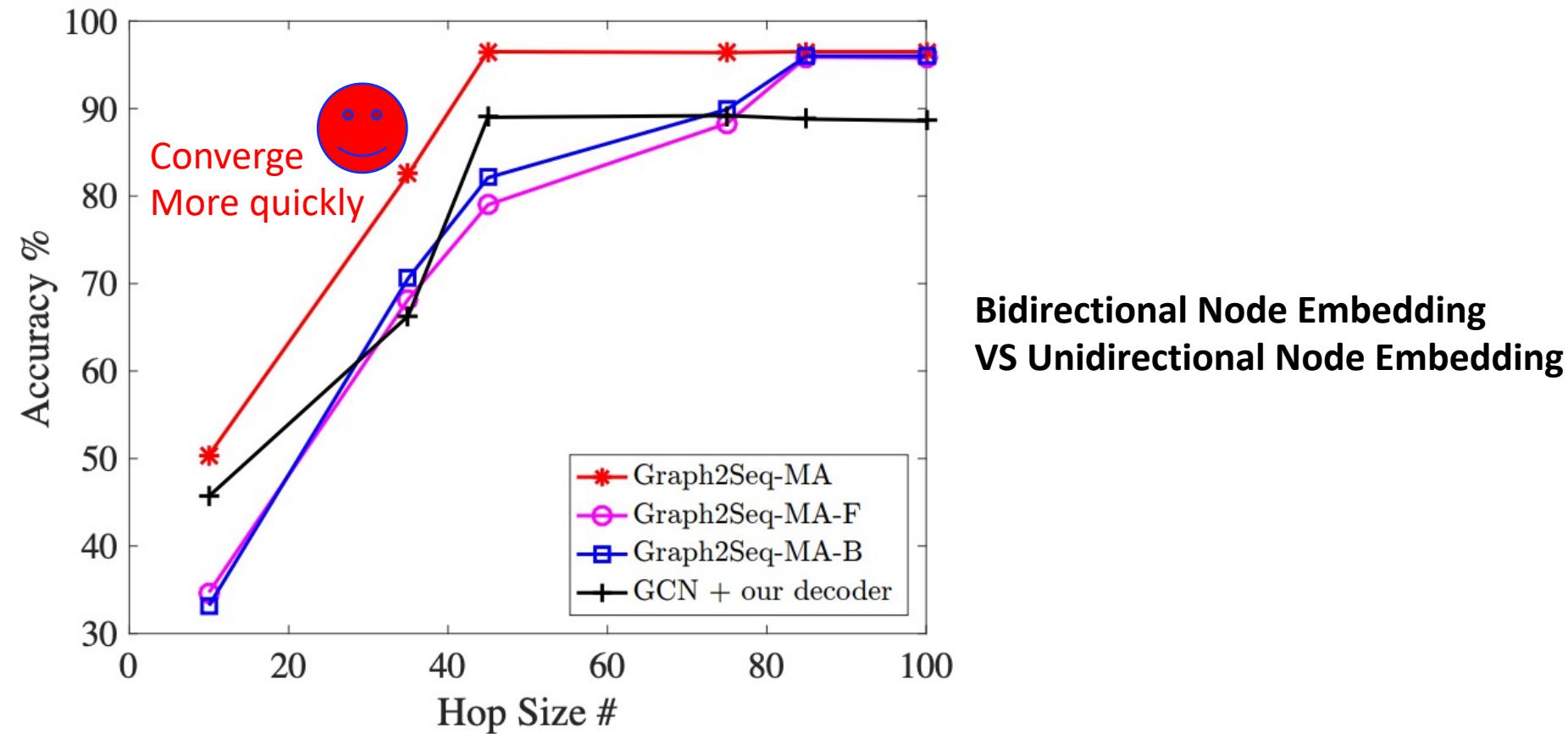
Q: How do you go from the **bathroom** to
the **hallway**

Transform

Q:path(B, D)

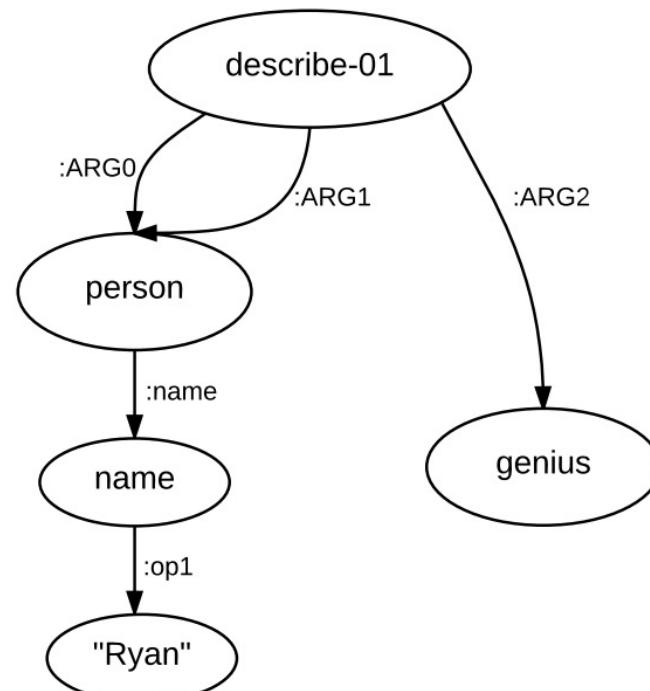
	bAbI T19	SP-S	SP-L
LSTM	25.2%	8.1%	2.2%
GGS-NN	98.1%	100.0%	95.2%
GCN	97.4%	100.0%	96.5%
Graph2Seq	99.9%	100.0%	99.3%

Effect of Bidirectional Node Embedding



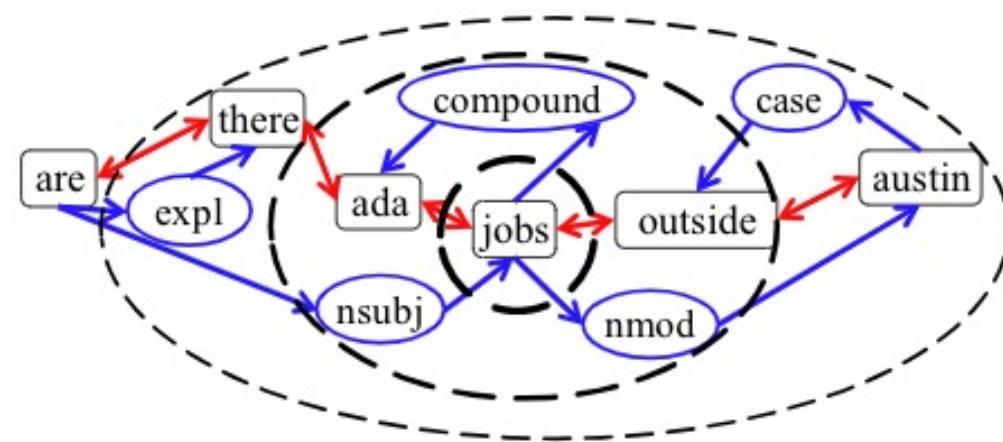
When Shall We Use Graph2Seq?

- Case I: the inputs are naturally or best represented in graph



“Ryan’s description of himself: a genius.”

- Case II: Hybrid Graph with sequence and its hidden structural information



Augmenting “are there ada jobs outside Austin” with its dependency parsing tree results

Learning Structured Input-Output Translation

- To bridge the semantic gap between the human-readable words and machine-understandable logics.
- Semantic parsing is important for question answering, text understanding
- Automatically solving of MWP is a growing interest.

SP	<p>Text Input: what jobs are there for web developer who know 'c++' ?</p> <p>Structured output: answer(A , (job (A) , title (A , W) , const (W , 'Web Developer') , language (A , C) , const (C , 'c++')))</p>
MWP	<p>Text input. 0.5 of the cows are grazing grass . 0.25 of the cows are sleeping and 9 cows are drinking water from the pond . find the total number of cows .</p> <p>Structured output: $((0.5 * x) + (0.25 * x)) + 9.0 = x$</p>

Graph and Tree Constructions

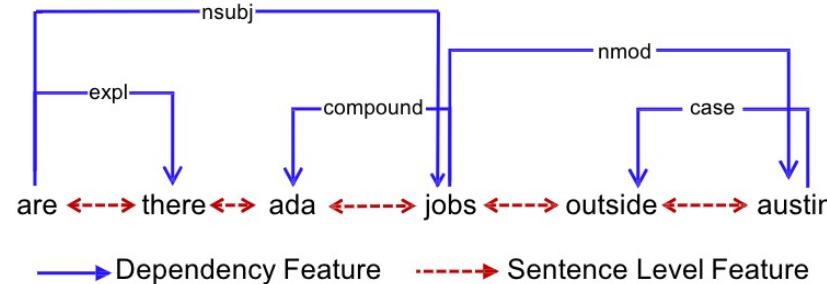


Figure 1: Dependency tree augmented text graph

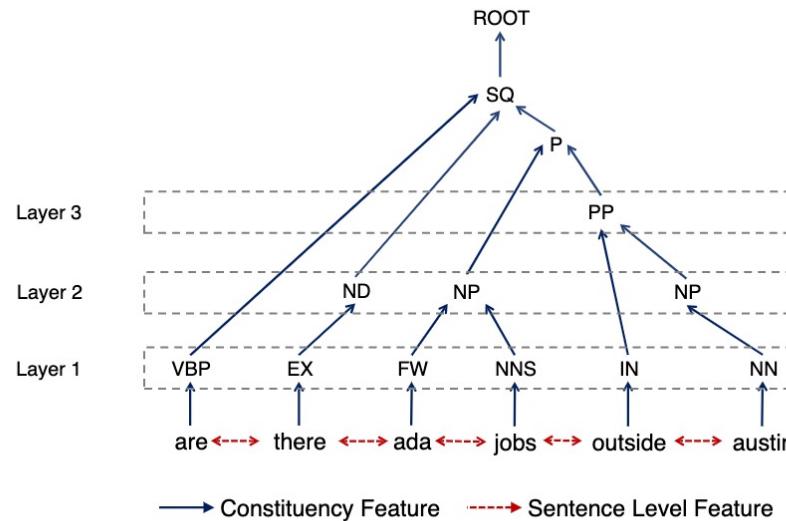


Figure 2: Constituency tree augmented text graph

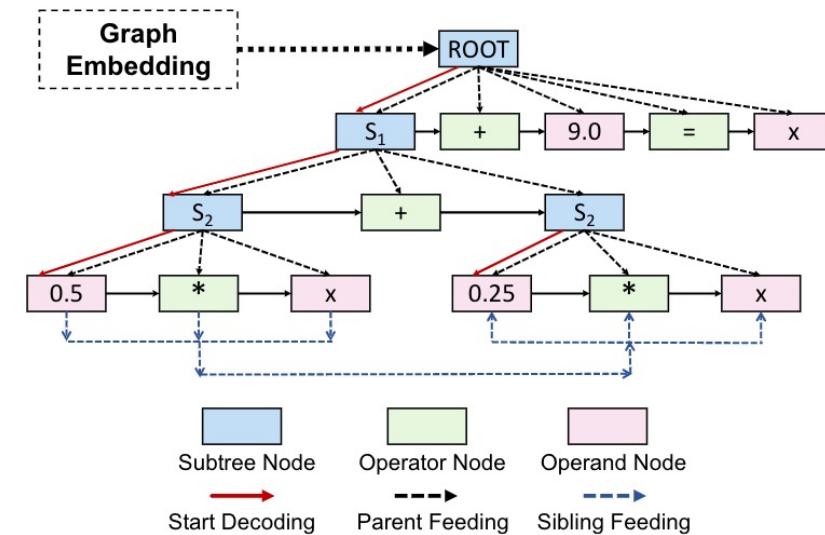
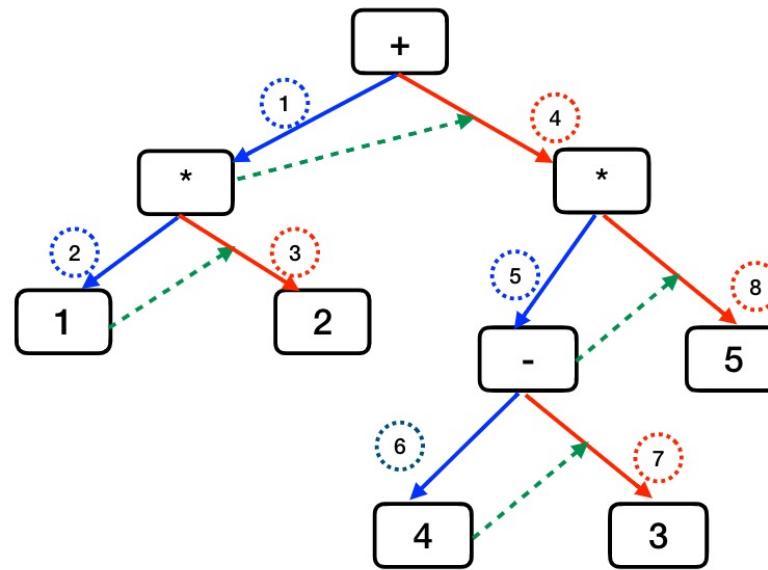


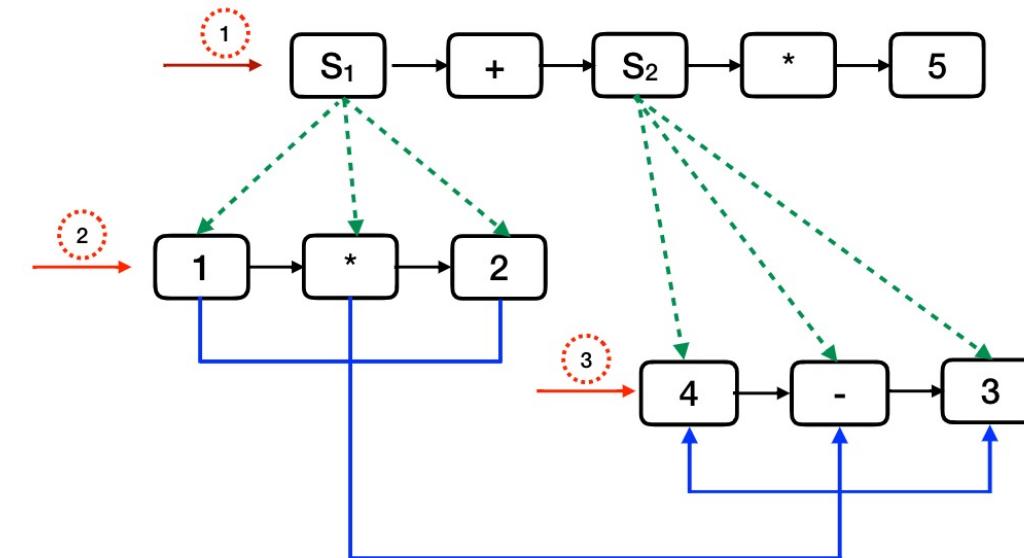
Figure 3: A sample tree output in our decoding process from expression $((0.5 * x) + (0.25 * x)) + 9.0 = x$

Tree Decoding



- Right node generation
- Left node generation
- Left sub-tree embedding

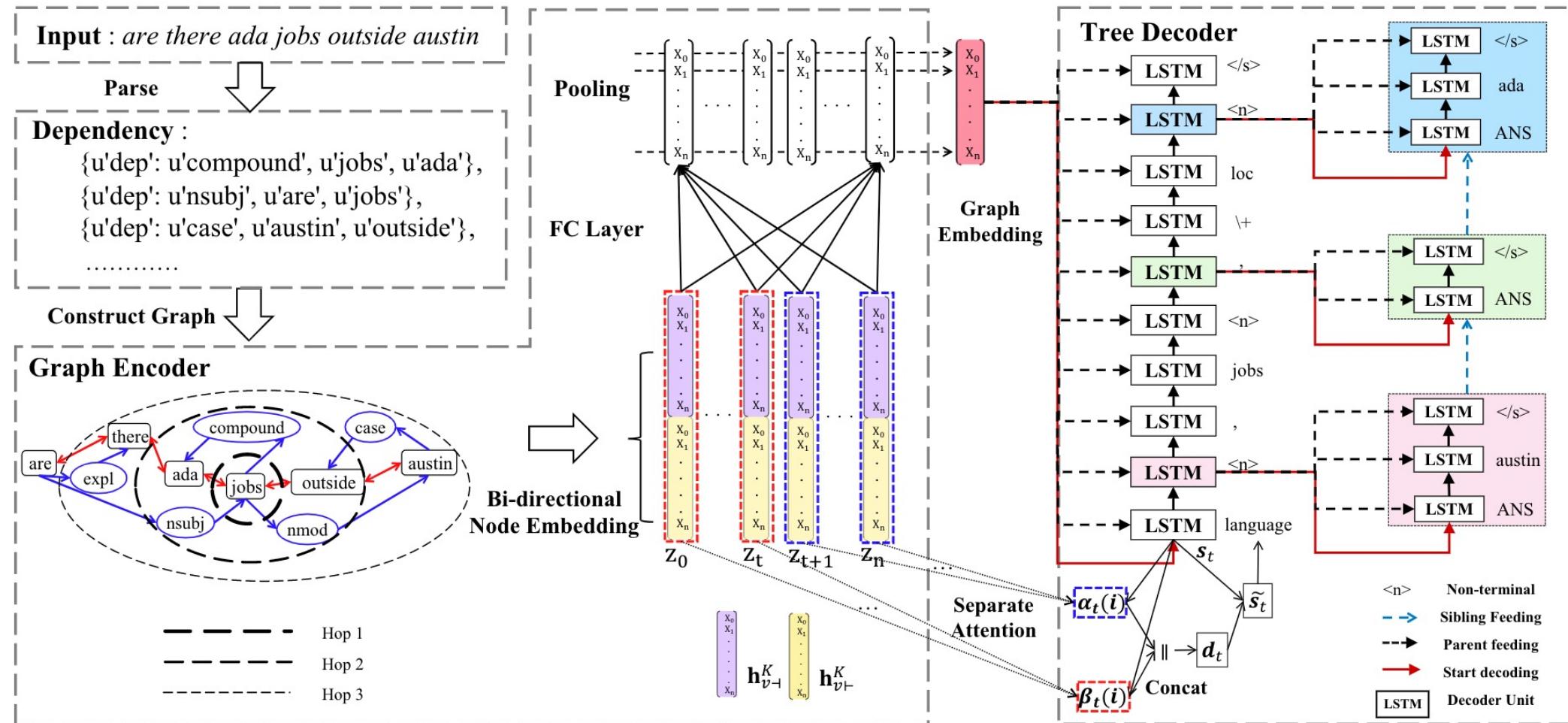
DFS-based tree decoder



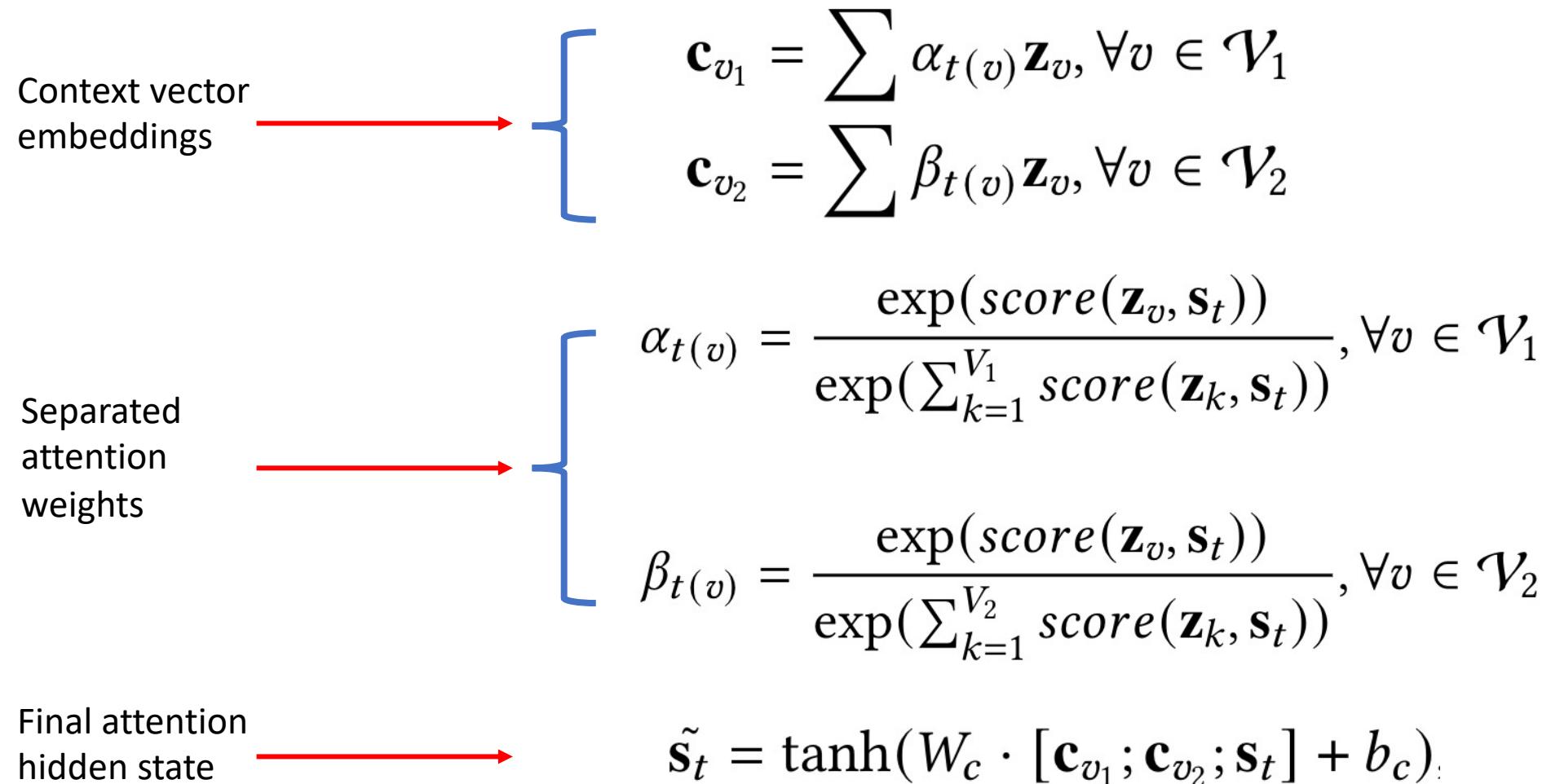
- Begin a new branch decoding
- Sequential decoding
- Sibling feeding
- Parent feeding

BFS-based tree decoder

Graph-to-Tree Model



Separated Attention Based Tree Decoding



Math Word Problem

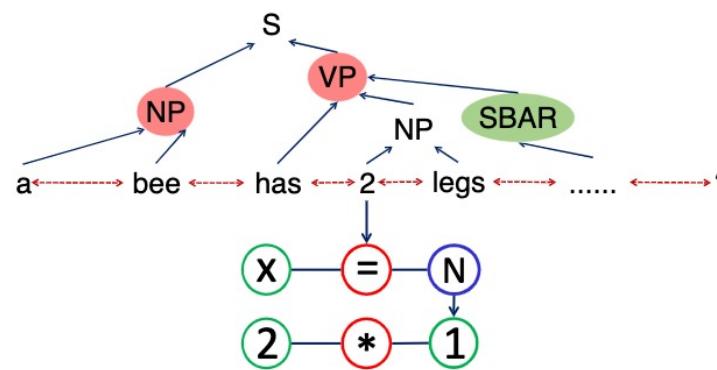
Methods		MAWPS
Oracle		84.8
Retrieval	Jaccard	45.6
	Cosine	38.8
Classification	BiLSTM	62.8
	Self-attention	60.4
Seq2seq	LSTM	25.6
	CNN	44.0
Seq2Tree		65.2
Graph2Seq		70.4
MathDQN		60.25
T-RNN	Full model	66.8
	W/o equation normalization	63.9
	W/o self-attention	66.3
Group-Att		76.1
Graph2Tree	with constituency graph	78.8
	with dependency graph	76.8

Table 5: Solution accuracy comparison on *MAWPS*

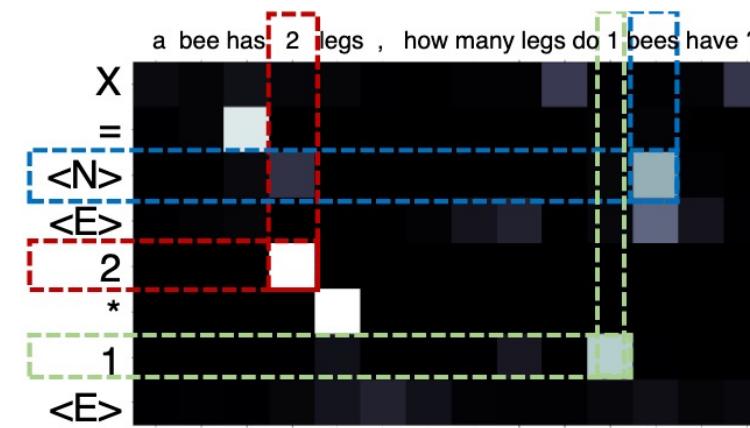
Methods	MATHQA
Seq2Prog	51.9
Seq2Prog+Cat	54.2
TP-N2F	55.95
Seq2seq	58.36
Seq2Tree	64.15
Graph2Seq	65.36
Graph2Tree	with constituency graph
	69.65
	with dependency graph
	65.66

Table 6: Solution accuracy comparison on *MATHQA*

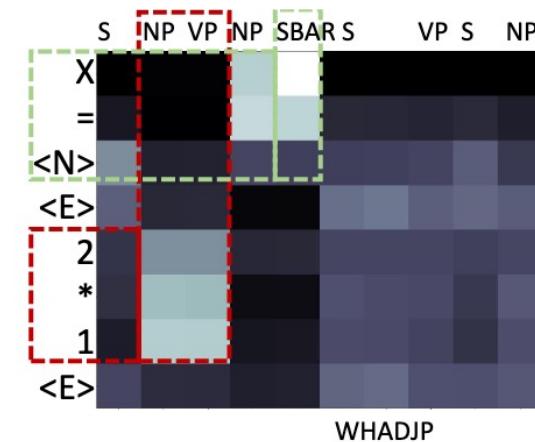
Visualization of Separated Attentions



(a) A graph-to-tree translation example



(b) Attention for word nodes



(c) Attention for structure nodes

Figure 5: Effect visualization of our separated attentions on both word and structure nodes in a graph.

10 Minutes Break

Want to prepare for our demo session?

- 1) git clone https://github.com/graph4ai/graph4nlp_demo
- 2) follow Get Started instructions in README

References:

- Website: <https://dlg4nlp.github.io/index.html>
- Survey: <https://arxiv.org/abs/2106.06090>
- Library: <https://github.com/graph4ai/graph4nlp>
- Demo: https://github.com/graph4ai/graph4nlp_demo

DLG4NLP Applications

Information Extraction

Outline

➤ Semantic Graph Parsing for Event Extraction



Improve Quality

• Cross-media Structured Common Space for Multimedia Event Extraction



Improve Portability

• Graph Schema-guided Event Extraction and Prediction



• Cross-media Knowledge Graph based Question Answering and Timeline Generation

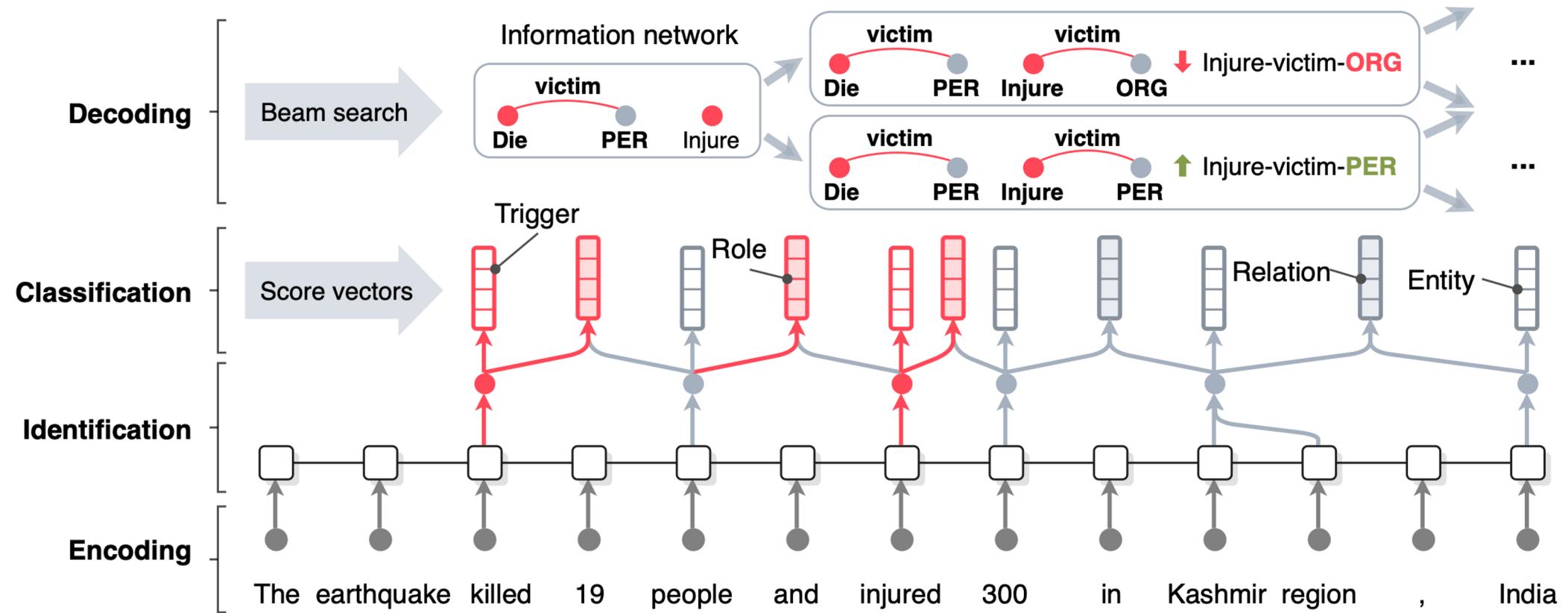
Extend Scope

• Cross-media Knowledge Graph based Misinformation Detection



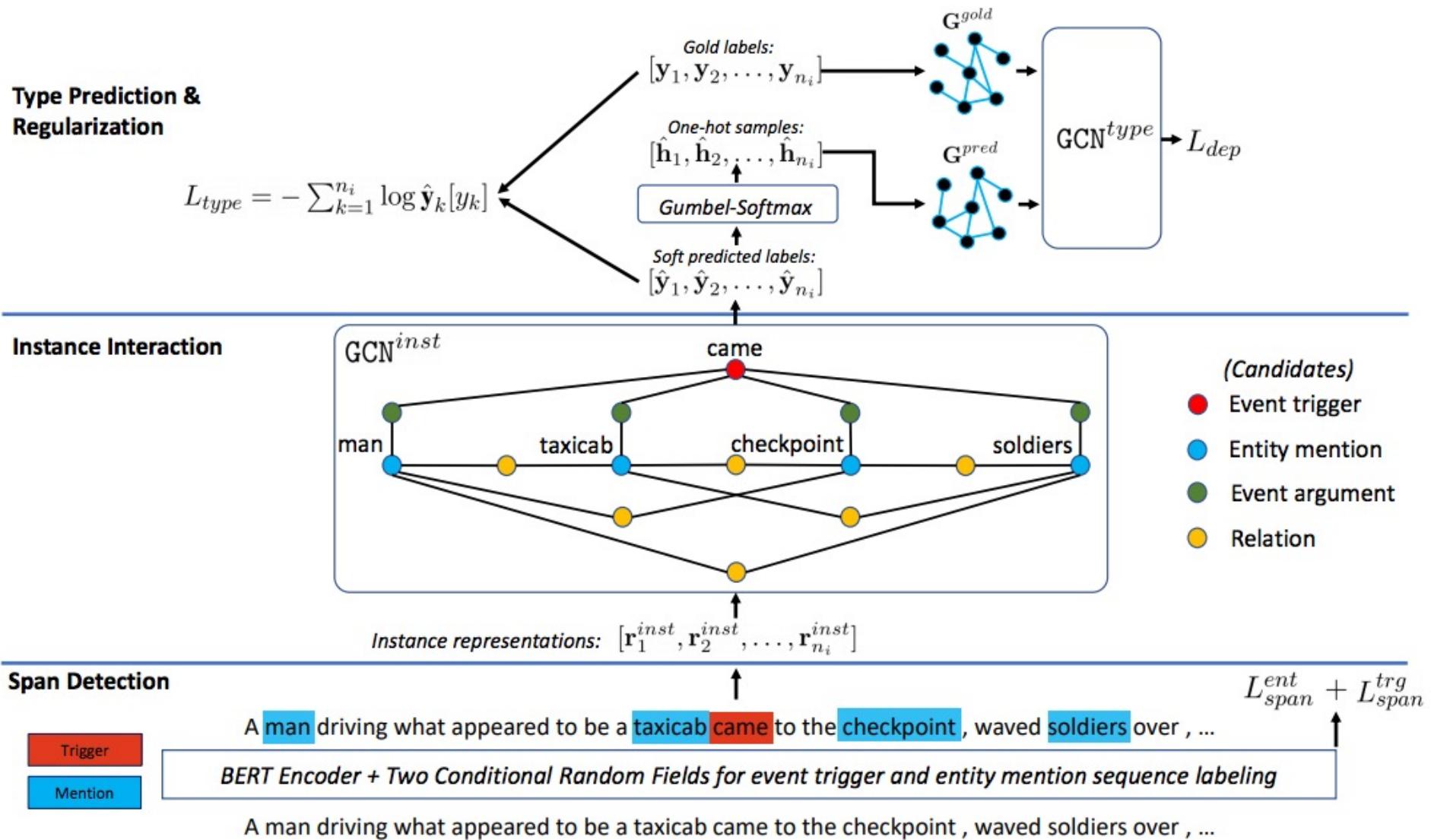
New Application

Information Extraction: a Sequence-to-Graph Task



- OneIE [Lin et al., ACL2020] framework extracts the information graph from a given sentence in four steps: encoding, identification, classification, and decoding

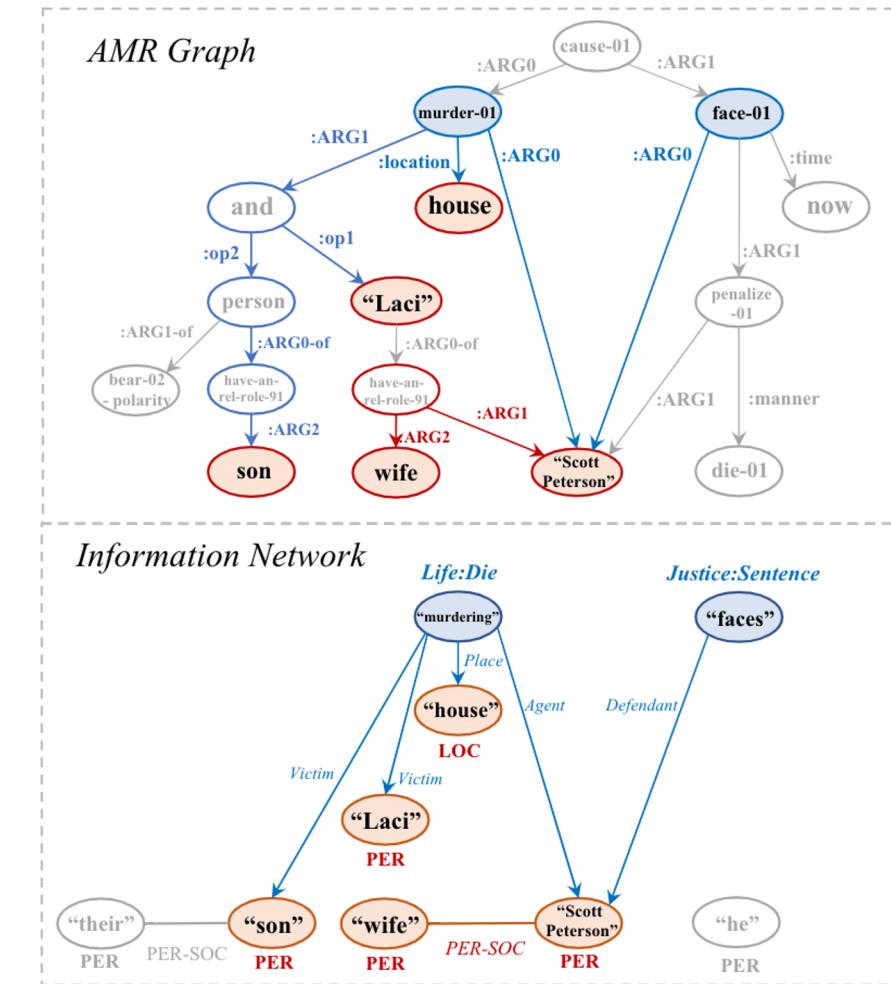
Extending to Graph-to-Graph Task



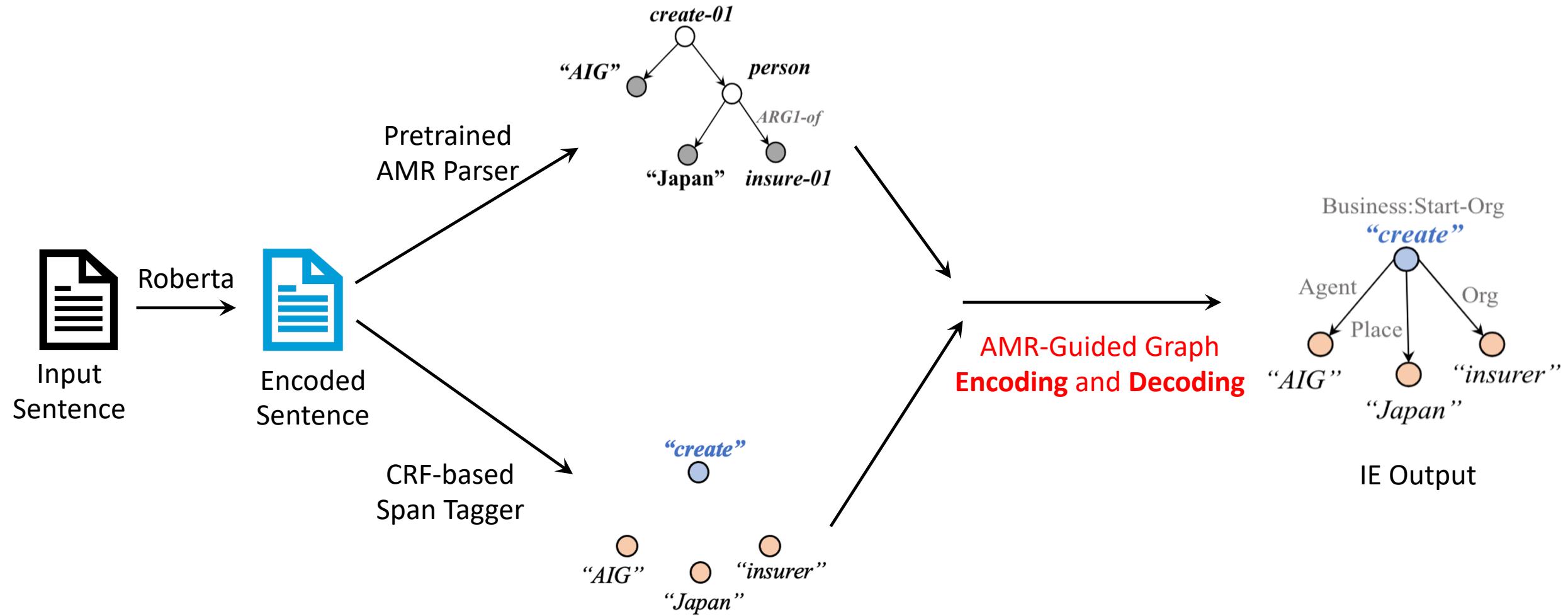
- [Nyuyen et al., NAACL2021]

Moving from Seq-to-Graph to Graph-to-Graph

- [Zhang and Ji, NAACL2021]
- Abstract Meaning Representation (AMR):
 - A kind of **rich semantic parsing**
 - Converts input sentence into a **directed** and **acyclic** graph structure with **fine-grained** node and edge type labels
- AMR parsing shares inherent similarities with information network (IE output)
 - Similar node and edge semantics
 - Similar graph topology
- Semantic graphs can better capture **non-local context** in a sentence
- **Exploit the similarity between AMR and IE to help on joint information extraction**

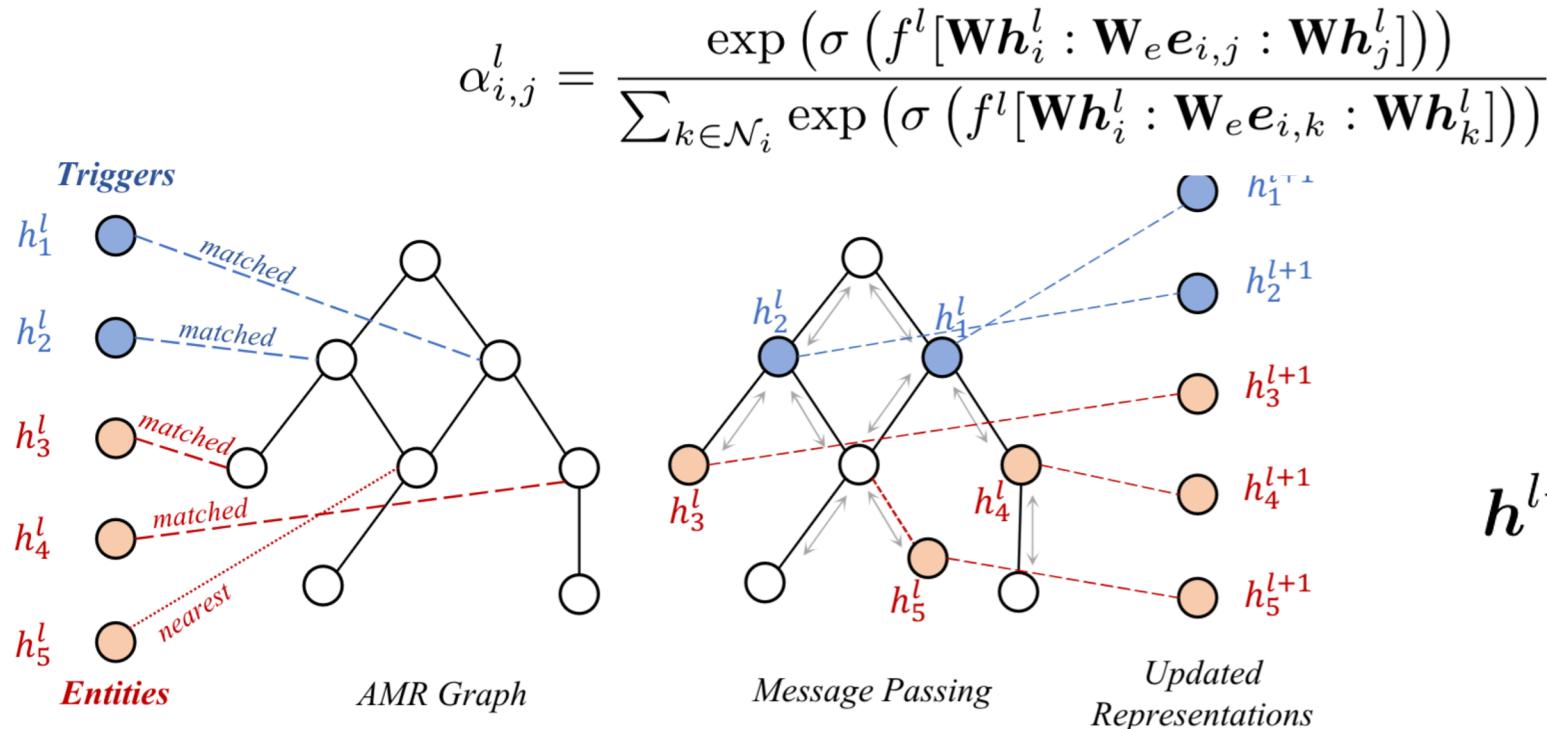


AMR-IE: An AMR-guided encoding and decoding framework for IE



AMR Guided Graph Encoding: Using an Edge-Conditioned GAT

- Map each candidate entity and event to AMR nodes.
- Update entity and event representations using an edge-conditioned GAT to incorporate information from AMR neighbors.



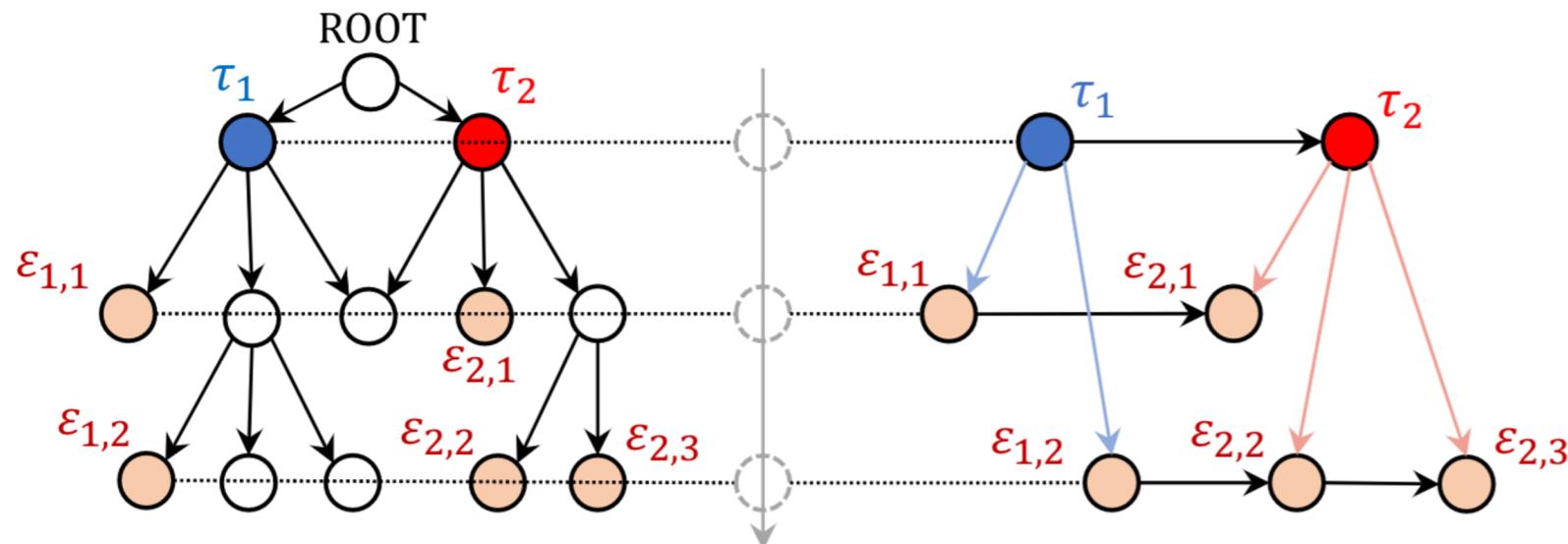
$$\mathbf{h}^* = \sum_{j \in \mathcal{N}_i} \alpha_{i,j}^l \mathbf{h}_i^l$$

$$\mathbf{h}^{l+1} = \mathbf{h}^l + \gamma \cdot \mathbf{W}^* \mathbf{h}^*$$

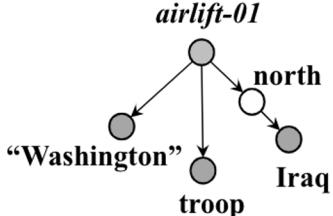
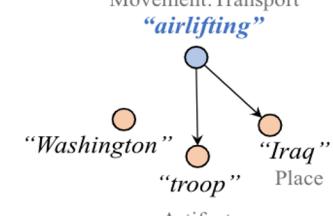
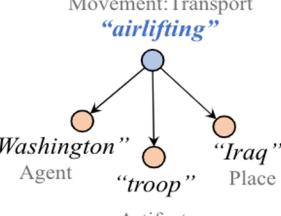
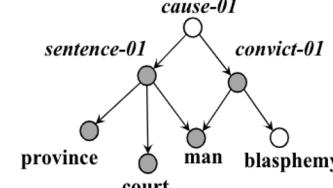
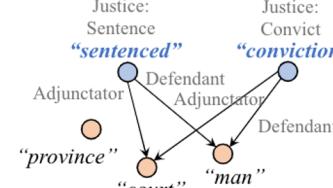
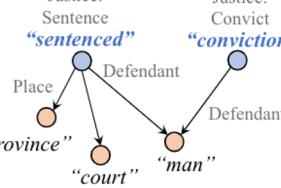
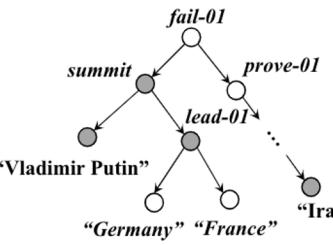
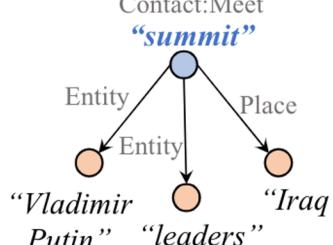
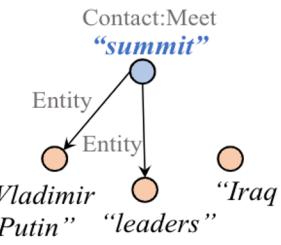
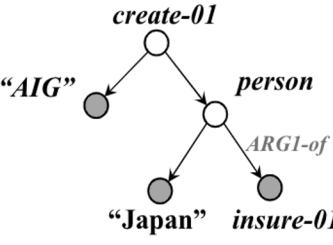
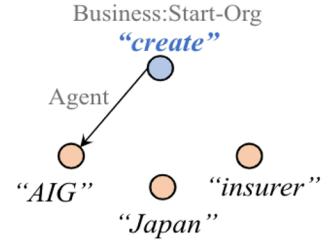
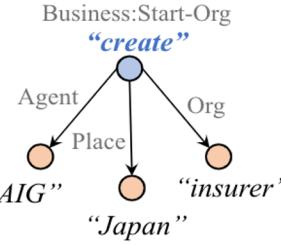
AMR Guided Graph Decoding: Ordered decoding guided by AMR

- Beam search based decoding as in *OneIE* (Lin et al. 2020).
- The decoding order of candidate nodes are determined by the hierarchy in AMR in a **top-to-down manner**.
- For example, the correct ordered decoding in the following graph is:

$$\tau_1, \tau_2, \varepsilon_{1,1}, \varepsilon_{2,1}, \varepsilon_{1,2}, \varepsilon_{2,2}, \varepsilon_{2,3}$$



Examples on how AMR graphs help

Sentence	AMR Parsing	<i>OneIE outputs</i>	<i>AMR-IE outputs</i>
If the resolution is not passed, Washington would likely want to use the airspace for strikes against Iraq and for airlifting troops to northern Iraq .	<p><i>airlift-01</i></p>  <pre> graph TD Washington((“Washington”)) -- north --> troop((troop)) troop --> Iraq((Iraq)) </pre>	<p>Movement:Transport “airlifting”</p>  <pre> graph TD Washington((“Washington”)) --> troop((troop)) troop --> Iraq((Iraq)) style Washington fill:#f0f0f0 style troop fill:#f0f0f0 style Iraq fill:#f0f0f0 </pre>	<p>Movement:Transport “airlifting”</p>  <pre> graph TD Washington((“Washington”)) --> troop((troop)) troop --> Iraq((Iraq)) style Washington fill:#f0f0f0 style troop fill:#f0f0f0 style Iraq fill:#f0f0f0 </pre>
A Pakistani court in central Punjab province has sentenced a Christian man to life imprisonment for a blasphemy conviction , police said Sunday.	<p><i>cause-01</i></p>  <pre> graph TD court((court)) -- sentence-01 --> man((man)) man -- convict-01 --> blasphemy((blasphemy)) style court fill:#f0f0f0 style man fill:#f0f0f0 </pre>	<p>Justice: Sentence “sentenced”</p> <p>Justice: Convict “conviction”</p>  <pre> graph TD court((court)) --> man((man)) man --> blasphemy((blasphemy)) style court fill:#f0f0f0 style man fill:#f0f0f0 style blasphemy fill:#f0f0f0 </pre>	<p>Justice: Sentence “sentenced”</p> <p>Justice: Convict “conviction”</p>  <pre> graph TD court((court)) --> man((man)) man --> blasphemy((blasphemy)) style court fill:#f0f0f0 style man fill:#f0f0f0 style blasphemy fill:#f0f0f0 </pre>
Russian President Vladimir Putin 's summit with the leaders of Germany and France may have been a failure that proves there can be no long-term "peace camp" alliance following the end of war in Iraq .	<p><i>fail-01</i></p>  <pre> graph TD VladimirPutin((“Vladimir Putin”)) -- summit --> Germany((“Germany”)) VladimirPutin -- summit --> France((“France”)) Germany -- lead-01 --> Iraq((Iraq)) style VladimirPutin fill:#f0f0f0 style Germany fill:#f0f0f0 style France fill:#f0f0f0 </pre>	<p>Contact:Meet “summit”</p>  <pre> graph TD VladimirPutin((“Vladimir Putin”)) --> leaders((“leaders”)) leaders --> Iraq((Iraq)) style VladimirPutin fill:#f0f0f0 style leaders fill:#f0f0f0 </pre>	<p>Contact:Meet “summit”</p>  <pre> graph TD VladimirPutin((“Vladimir Putin”)) --> leaders((“leaders”)) leaders --> Iraq((Iraq)) style VladimirPutin fill:#f0f0f0 style leaders fill:#f0f0f0 </pre>
Major US insurance group AIG is in the final stage of talks to take over General Electric's Japanese life insurance arm in a deal to create Japan's sixth largest life insurer , reports said Wednesday.	<p><i>create-01</i></p>  <pre> graph TD AIG((“AIG”)) -- person --> Japan((“Japan”)) Japan -- “ARG1-of” --> insure01((“insure-01”)) </pre>	<p>Business:Start-Org “create”</p>  <pre> graph TD AIG((“AIG”)) --> Japan((“Japan”)) Japan --> insure01((“insure-01”)) style AIG fill:#f0f0f0 style Japan fill:#f0f0f0 </pre>	<p>Business:Start-Org “create”</p>  <pre> graph TD AIG((“AIG”)) --> Japan((“Japan”)) Japan --> insure01((“insure-01”)) style AIG fill:#f0f0f0 style Japan fill:#f0f0f0 </pre>

Outline

- Semantic Graph Parsing for Event Extraction
- Cross-media Structured Common Space for Multimedia Event Extraction
- Graph Schema-guided Event Extraction and Prediction
- Cross-media Knowledge Graph based Question Answering and Timeline Generation
- Cross-media Knowledge Graph based Misinformation Detection

Multimedia Event Extraction (M^2E^2)

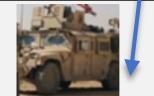
[Li et al., ACL2020]

Last week , U.S . Secretary of State Rex Tillerson visited Ankara, the first senior administration official to visit Turkey, to try to seal a deal about the battle for Raqqa and to overcome President Recep Tayyip Erdogan's strong objections to Washington's backing of the Kurdish Democratic Union Party (PYD) militias. Turkish forces have attacked SDF forces in the past around Manbij, west of Raqqa, forcing the **United States** to **deploy** dozens of **soldiers** on the **outskirts** of the town in a mission to prevent a repeat of clashes, which risk derailing an assault on Raqqa.



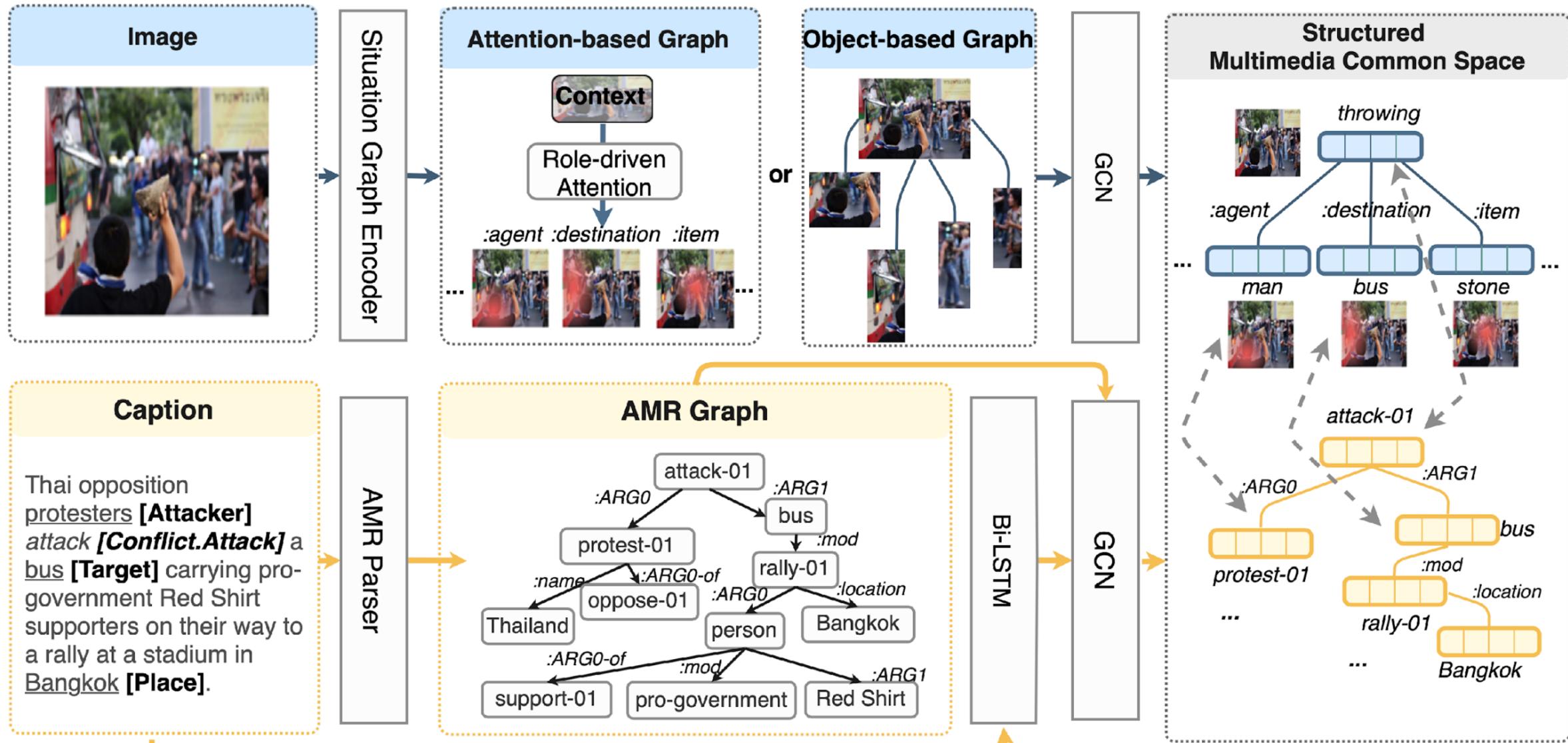
Output: Multimedia Events & Argument Roles

Event Type	Movement.Transport	
Event	Text Trigger	deploy
	Image	

Arguments	Agent	United States
	Destination	outskirts
	Artifact	soldiers
	Vehicle	
	Vehicle	

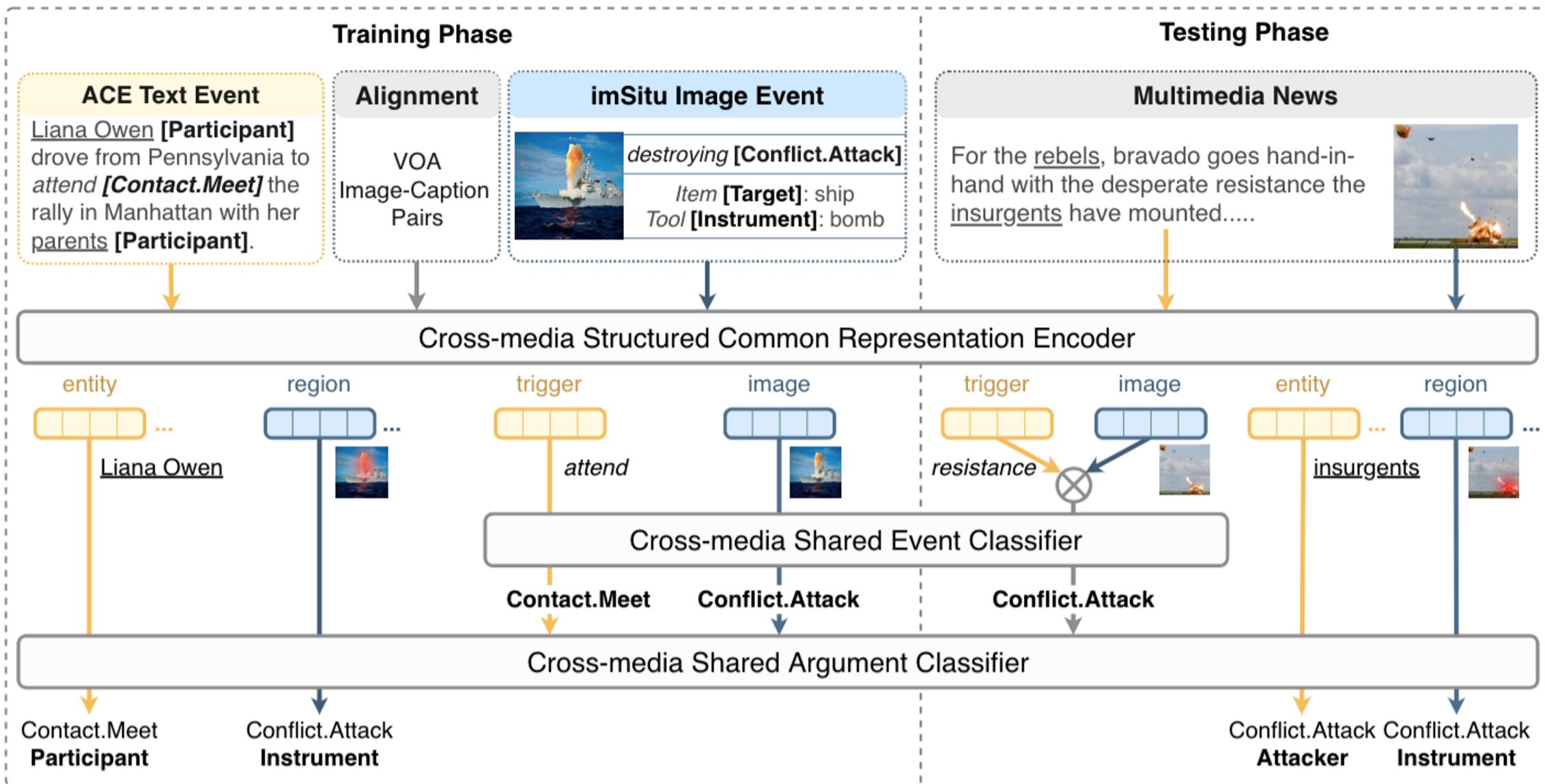
Weakly Aligned Structured Embedding

-- Training Phase (Common Space Construction)



Weakly Aligned Structured Embedding

-- Training and Test Phase (Cross-media shared classifiers)



Compare to Single Data Modality Extraction

- Surrounding sentence helps visual event extraction.
- Image helps textual event extraction.

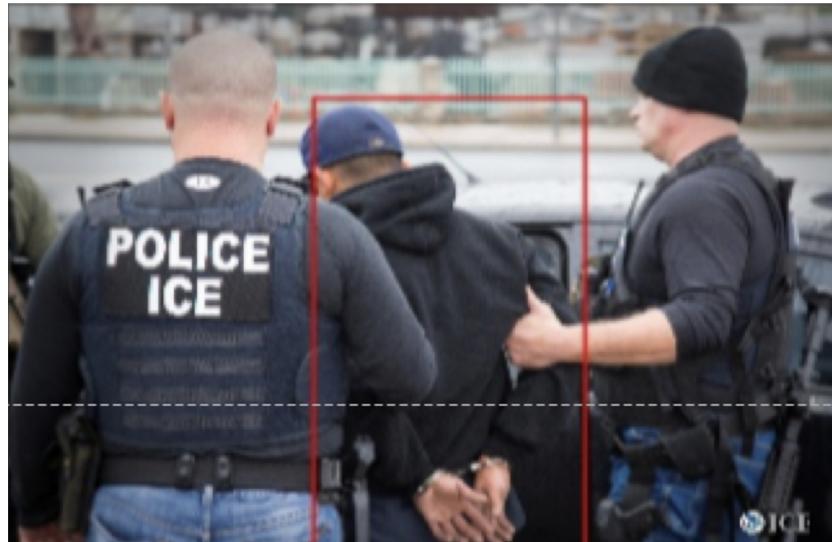


People celebrate Supreme Court ruling on Same Sex Marriage in front of the Supreme Court in Washington.



Iraqi security forces search [**Justice.Arrest**] a civilian in the city of Mosul.

Compare to Cross-media Flat Representation



Model	Event Type	Argument Role	
Flat	Justice.ArrestJail	Agent =	man
Ours	Justice.ArrestJail	Entity =	man

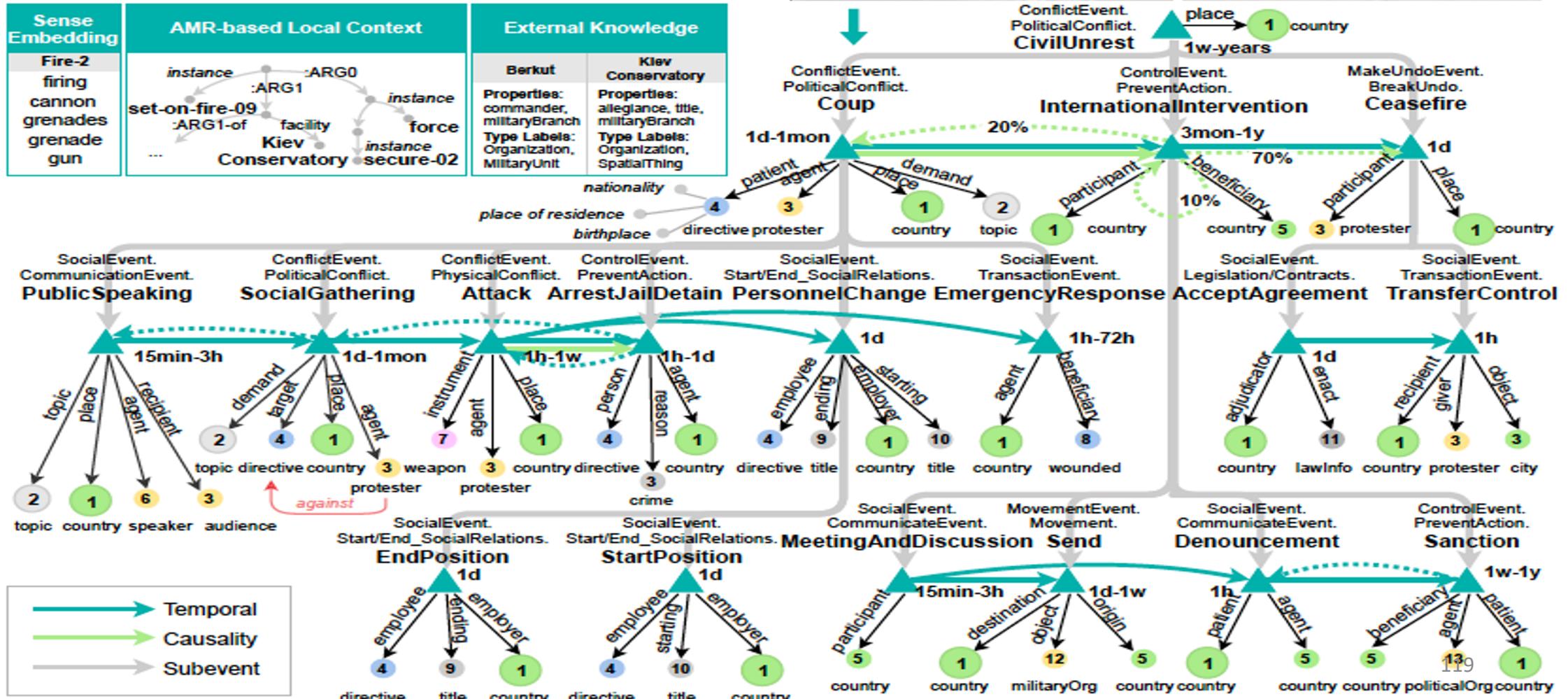
Model	Event Type	Argument Role	
Flat	Movement.Transport	Artifact =	none
Ours	Movement.Transport	Artifact =	man

Outline

- Semantic Graph Parsing for Event Extraction
- Cross-media Structured Common Space for Multimedia Event Extraction
- Graph Schema-guided Event Extraction and Prediction
 - Cross-media Knowledge Graph based Question Answering and Timeline Generation
 - Cross-media Knowledge Graph based Misinformation Detection

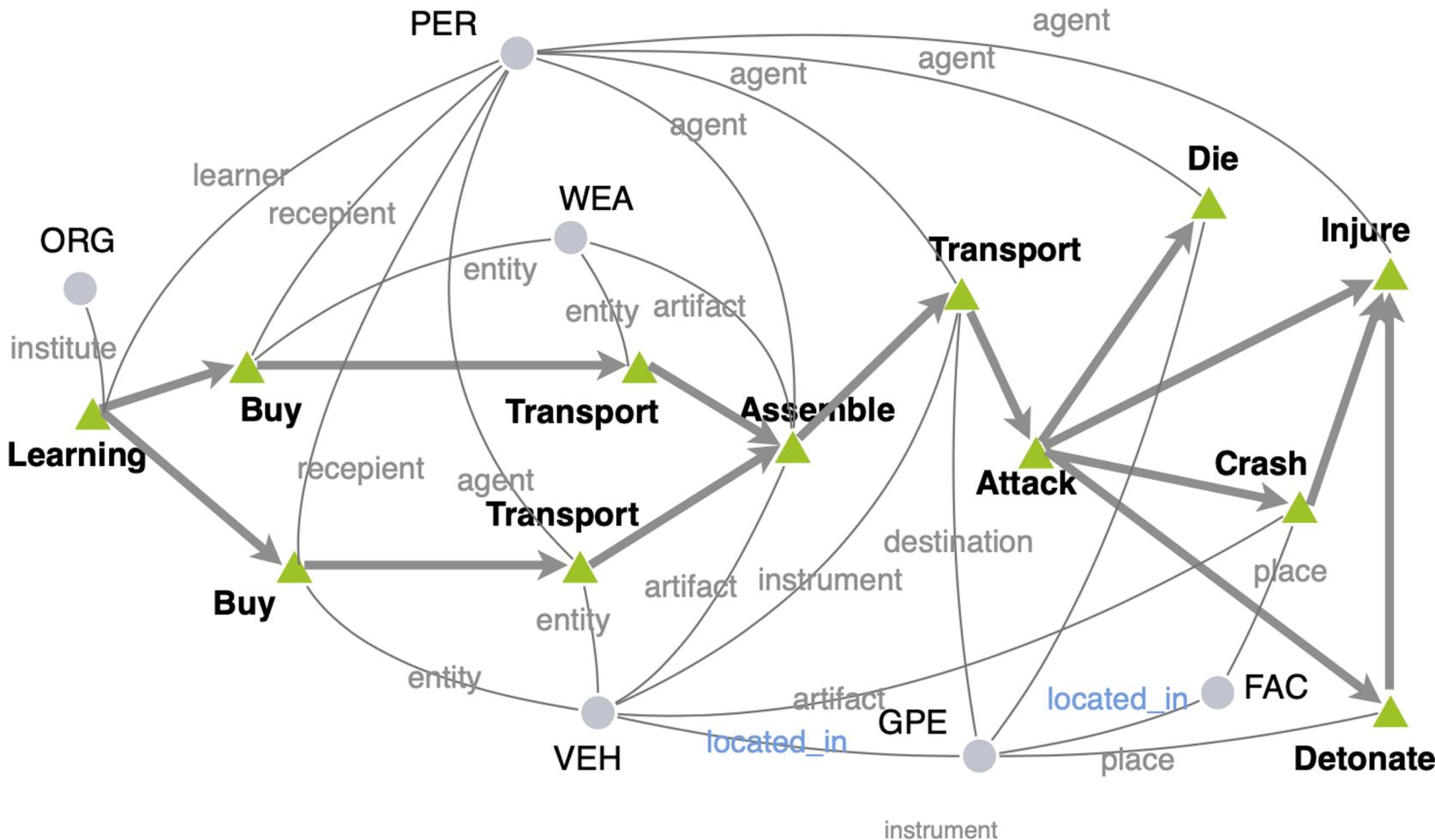
Global Event Knowledge Acquisition via Schema Induction

2014 Thai coup d'état: Однако протесты и блокада делятся уже почти 3 месяца, а военные так и не перешли к действиям
2013 Egyptian coup d'état: ... General Abdel Fattah el-Sisi announced that he there would be calling new presidential and Shura Council elections.
Ukrainian crisis: At 09:25, protesters pushed the Berkut back to the October Palace after security forces tried to set fire to Kiev Conservatory, which was being used as a field hospital for wounded protesters.



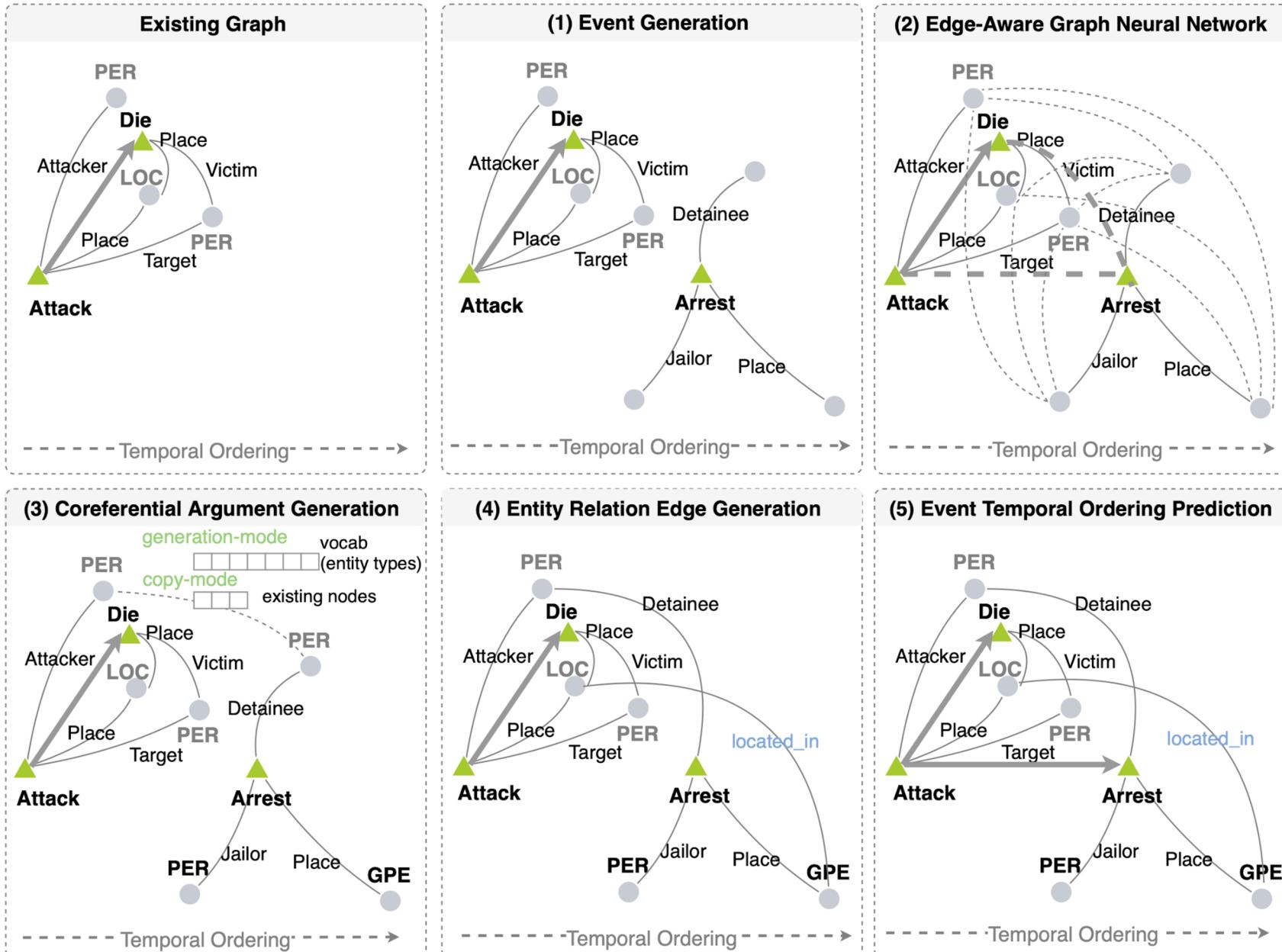
Temporal Complex Event Schema Composition [Li et al., 2021a]

- Graph Structure Aware:
 - Encode entity coreference and entity relation
 - Capture the interdependency of events and entities (sequences can not)
- Scenario guided:
 - Train one model based on instance graphs of the same scenario
- Probabilistic:
 - Support downstream tasks, such as event prediction



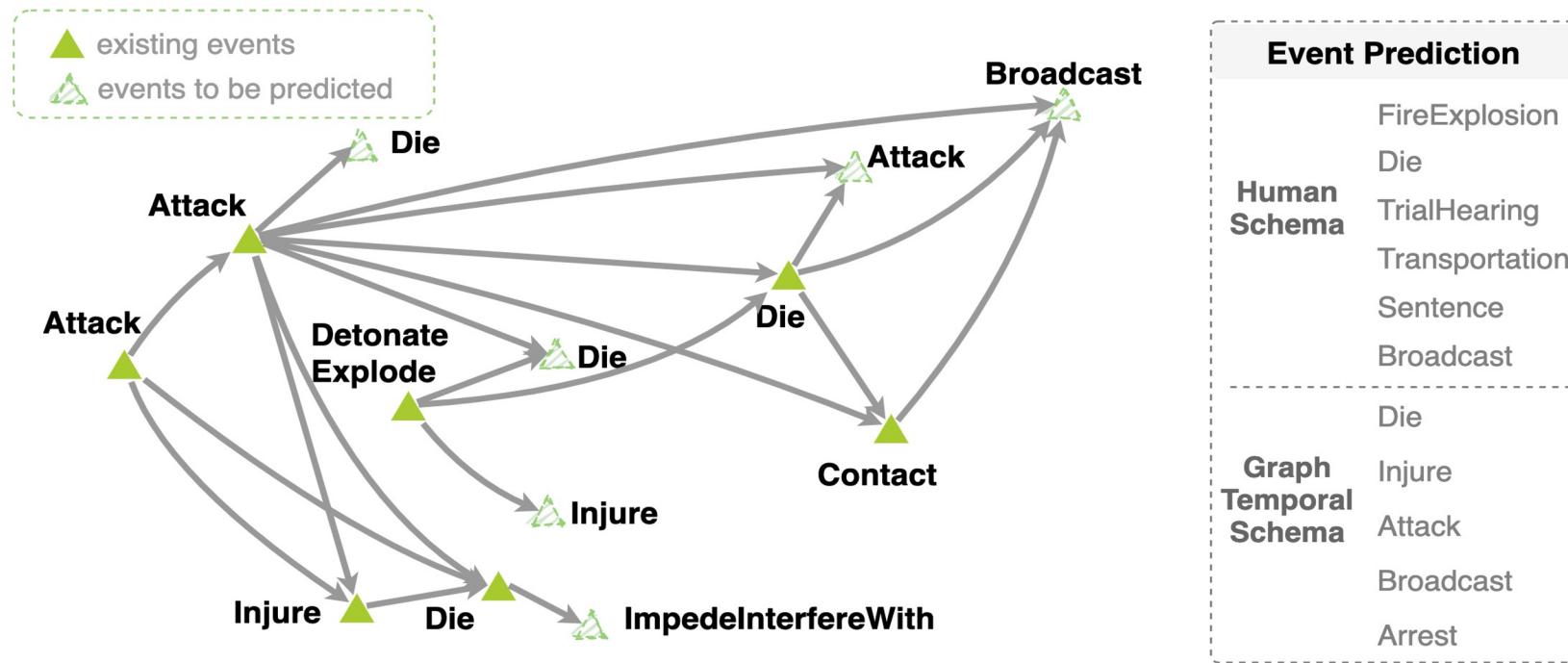
Generative Event Graph Model

- Schemas are the hidden knowledge to control instance graph generation
- Step 1. Event Node Generation
- Step 2. Message Passing
- Step 3. Argument Node Generation
- Step 4. Relation Edge Generation
- Step 5. Temporal Edge Generation



Schema-guided Event Prediction

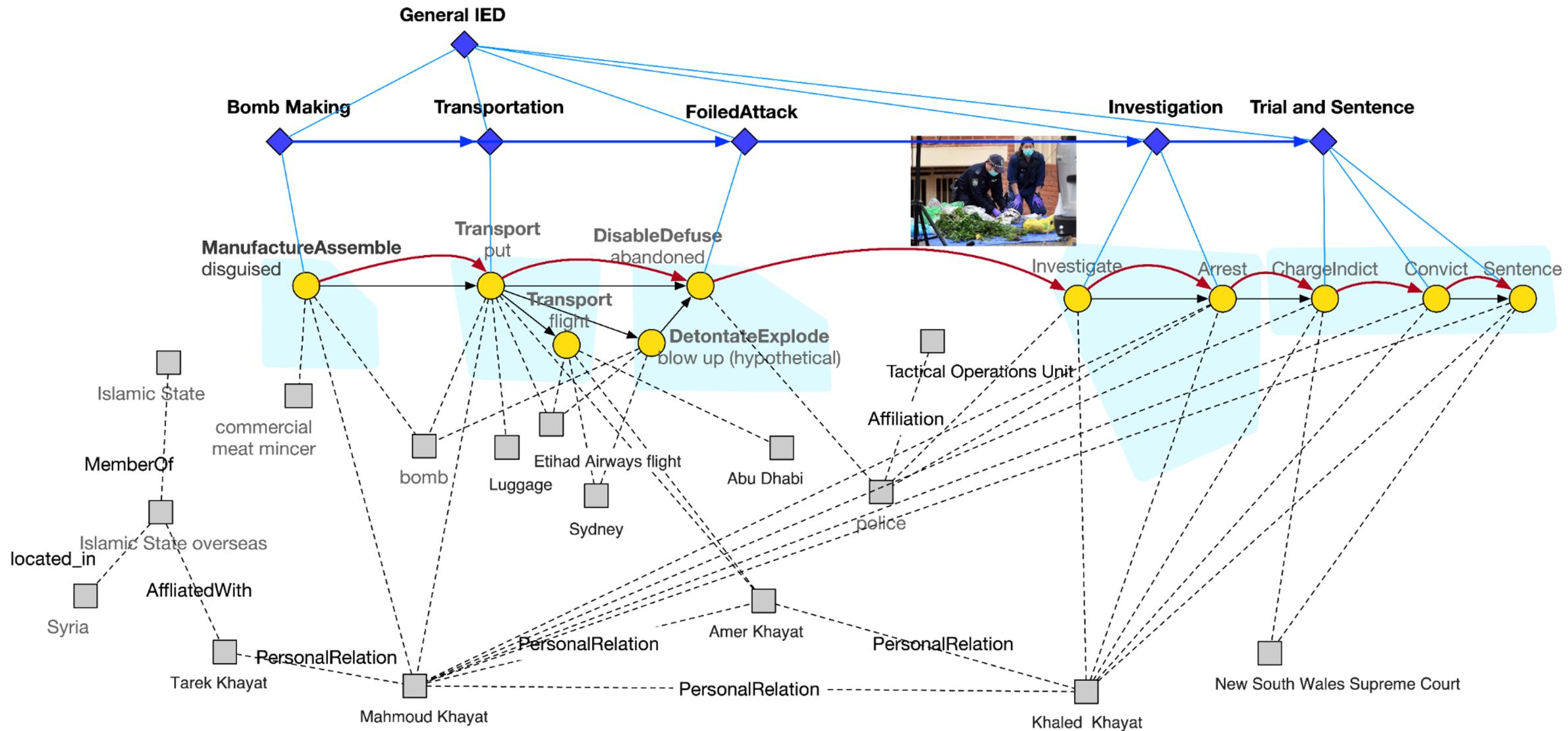
- **Schema-guided Event Prediction:** The task aims to predict ending events of each graph.
 - Considering that there can be multiple ending events in one instance graph, we rank event type prediction scores and adopt MRR and HITS@1 as evaluation metrics.



Dataset	Models	MRR	HITS@1
General	Human Schema	0.173	0.205
	Event Graph Model	0.457	0.591

Dataset	Models	MRR	HITS@1
IED	Human Schema	0.072	0.222
	Event Graph Model	0.203	0.426

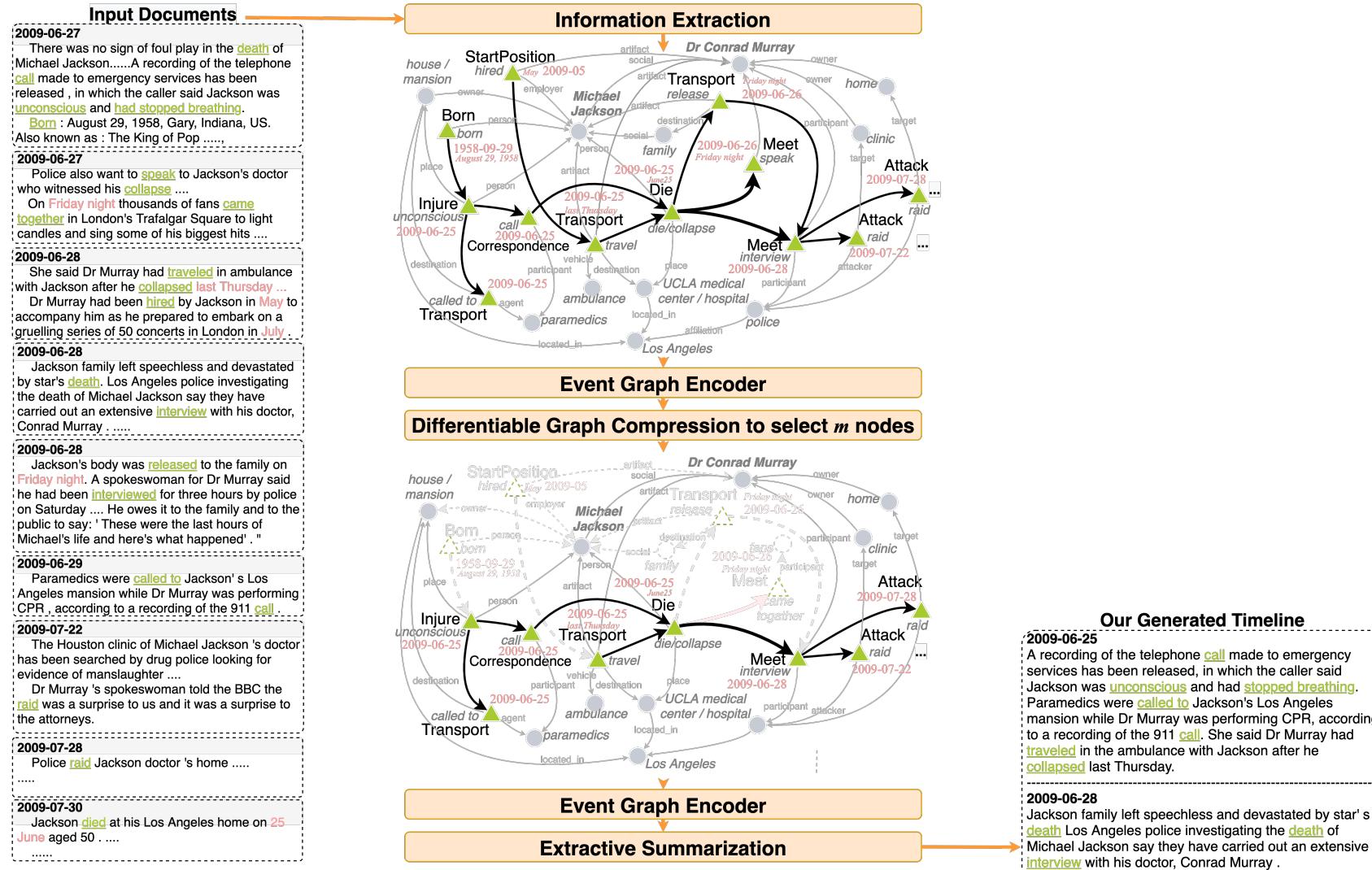
Instance-level Event Prediction



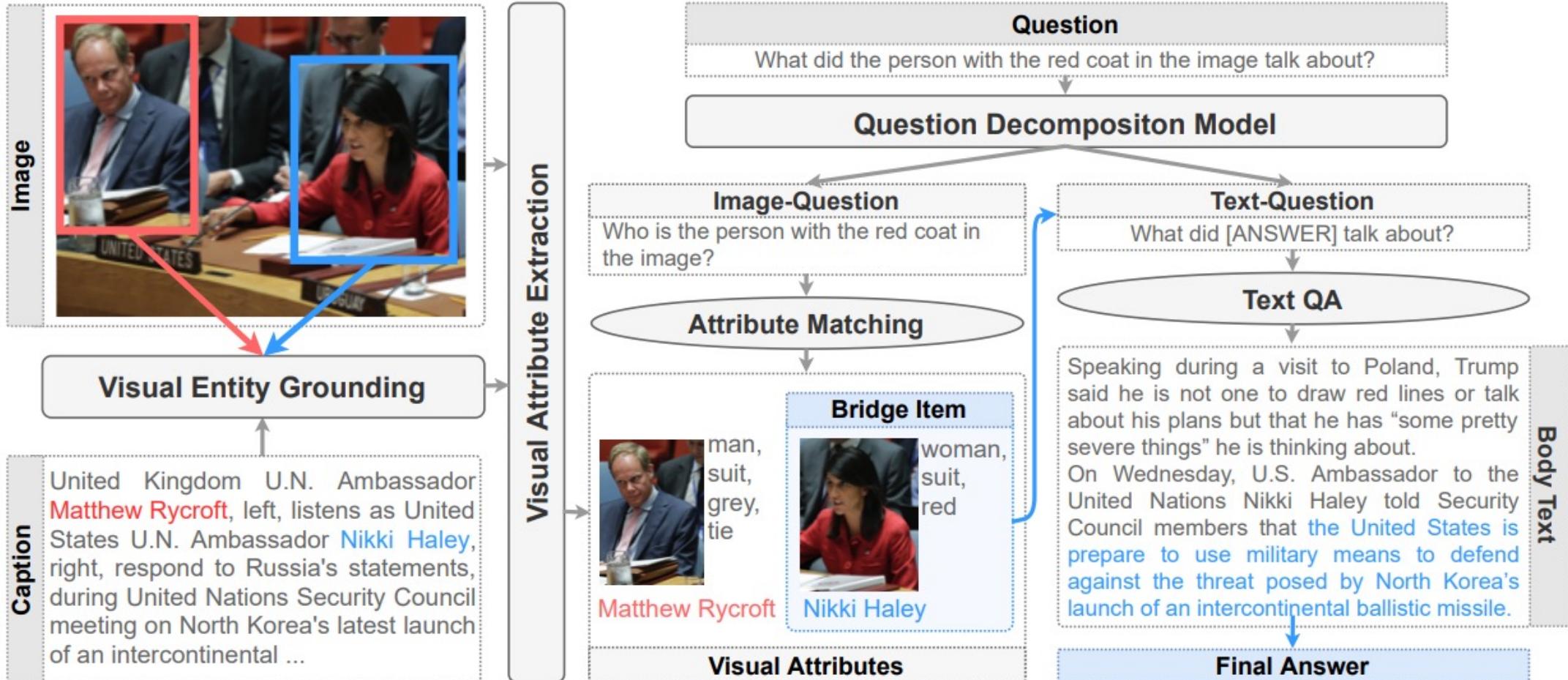
Outline

- Semantic Graph Parsing for Event Extraction
- Cross-media Structured Common Space for Multimedia Event Extraction
- Graph Schema-guided Event Extraction and Prediction
- Cross-media Knowledge Graph based Question Answering and Timeline Generation
- Cross-media Knowledge Graph based Misinformation Detection

Knowledge Graph based Event Timeline Generation [Li et al., EMNLP2021b]



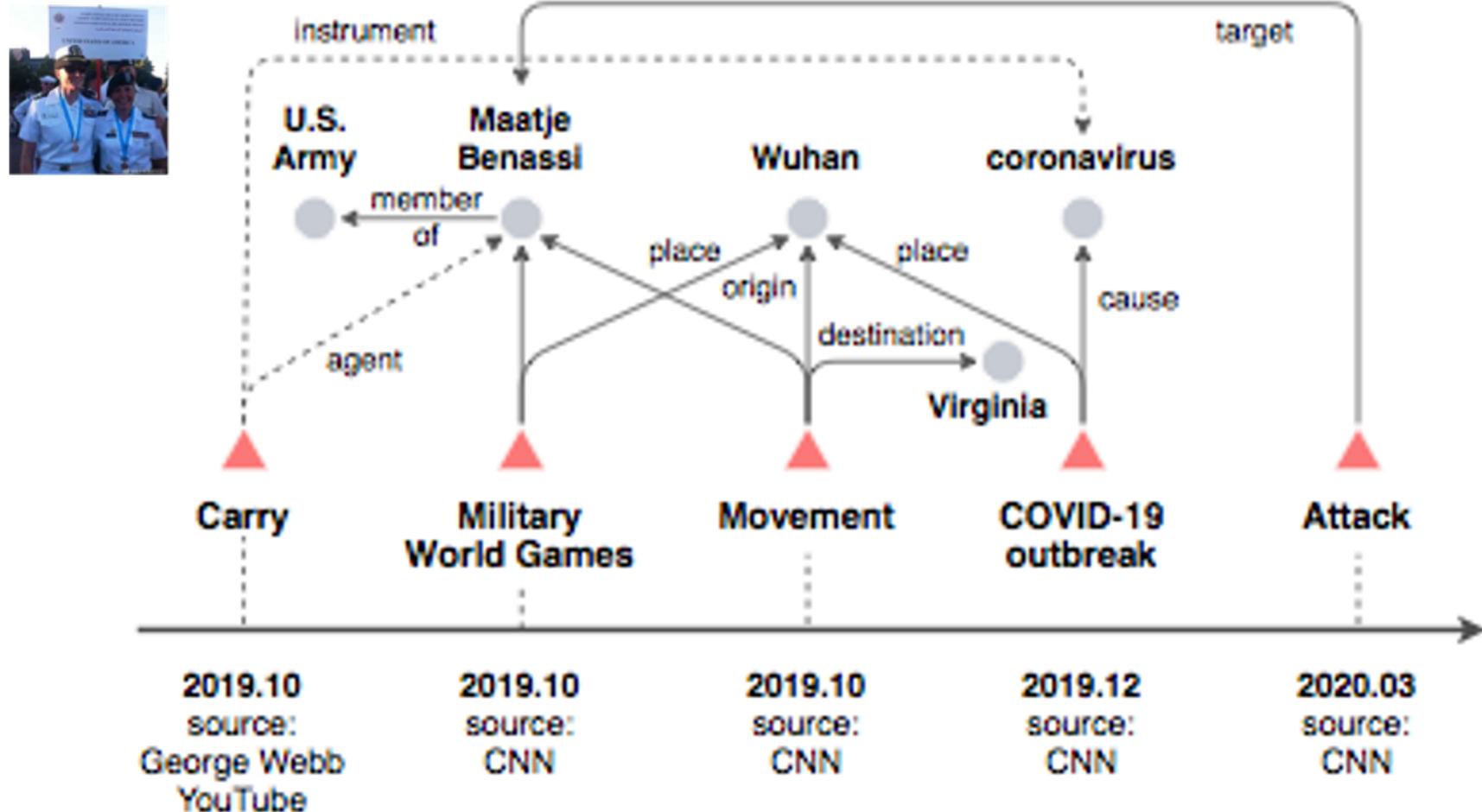
Multimedia Knowledge Graph based Question Answering [Reddy et al., AAAI2022]



Outline

- Semantic Graph Parsing for Event Extraction
 - Cross-media Structured Common Space for Multimedia Event Extraction
 - Graph Schema-guided Event Extraction and Prediction
 - Cross-media Knowledge Graph based Question Answering and Timeline Generation
- Cross-media Knowledge Graph based Misinformation Detection

Information Pollution



- Why would anyone ever believe these rumors?
- Because humans are very good at connecting dots
- And perhaps too good →

Quiz Time! Which one is Fake News?

Burma's once-outlawed National League for Democracy is holding its first party congress since the opposition group was founded 25 years ago. Delegates in Rangoon will draw up a policy framework and elect a central committee during the three-day meeting that began Friday. Democracy icon Aung San Suu Kyi is also expected to be reappointed as head of the party. The Nobel laureate helped the NLD to a strong showing in historic April by-elections, which saw the party win 43 of the 45 contested seats. But the NLD is setting its sights on 2015, when it hopes to take power during national elections. But the party faces several challenges as it attempts to fashion itself into a viable political alternative to the military, which still dominates parliament and other government institutions. One of the most pressing issues is electing younger leaders to replace the party's elderly founding members, many of whom are in their 80s or 90s and in poor health.



Congress delegates prepare to pose for photographs as they arrive to attend the National League for Democracy party's (NLD) congress in Rangoon, March 8, 2013.

Delegates from the NLD gather in Rangoon for the party's annual congress. The NLD is headed by Nobel Peace Prize winner Aung San Suu Kyi. **The party is expected to win a majority of seats in the parliament.**

This year's NLD Congress is the first time the party has been able to elect its own leadership. Nyan Win, a member of NLD's executive committee, told VOA that the party is looking forward to the new generation of leaders.

The party has come a long way since the military seized power in 1962. **The NLD was founded by a Briton.** Since then, Burma has been ruled by a quasi-civilian government. However, the military has still maintained tight control over the country's political institutions. **Phil Robertson**, Asia director for Human Rights Watch, **said he hopes the party will push forward with reforms that will allow the army to step down and allow the civilian government to take over.**



Quiz Time! Which Caption is Fake?

On 24 May 2017 the Philippines **militants left** their barrack in the outskirts of southern Marawi city to reinforce fellow troops who had been under siege by Islamic **troops**.



Philippine **troops arrive** at their barracks to reinforce fellow troops following the siege by Muslim **militants**, on the outskirts of Marawi city in the southern Philippines, May 24, 2017.

Quiz Time! Which Caption is Fake?

Anis Amri (L), the Tunisian suspect of the [Berlin](#) Christmas market [attack](#), is seen in this photo taken from security cameras at the [Milan](#) Central Train Station in downtown [Milan](#), Italy December 23, 2016.



Anis Amri, a Tunisian suspected of [defending](#) the Christmas market in [Milan](#), was seen in this photo given from a security camera at the Central Train Station of downtown [Berlin](#) on 23 December 2016 .



Knowledge Element-Level Misinformation Detection

[Fung et al., ACL2021]

Motivation: misinformative parts of a fake news article lie along the fine-grained details

Current Issues:

- Fake news detection approaches tend to focus on checking facts, semantic inconsistencies, style or bias, lacking a *unified framework*.
- The document-level detection task lacks *precision* and *explainability*

Ex of Grover-Generated Fake News - News Spoofing

Hong Kong declared Independence from China Yesterday

- February 19, 2021

In a historic decision made yesterday, Hong Kong declared its independence from mainland China. The Senate of Hong Kong, the local government's legislative body, passed the inaugural Resolution of Independence after members of all

races, sects and ages gathered in the senate chambers...

"As the Chief Executive Council today endorsed the proposal of the Chief Executive Council to confirm the first proposed Resolution of Independence, Hong Kong is

determined to complete the path of self-rule," said London-

factually incorrect → based broadcaster CNN yesterday... awkward linguistic

"We look forward to the motion being made by the style

Legislative Council and to firmly reaffirming our

commitment to a prosperous and stable life of our people,

while working together with China," Hong Kong's Chief

Executive, Carrie Lam, said in a statement, according to AFP.

Compare with Previous Work

- ❖ Motivation: misinformative parts of a fake news article lie along the fine-grained details
 - Existing approaches lack a *unified framework* in checking facts, semantic inconsistencies, text features and bias

	Text Features	Structured Knowledge	Source Bias	Multimedia	Knowledge Element Level Detection
Perez-Rosas et al. (2018)	✓				
Pan et al. (2017)		✓			
Baly et al. (2018)	✓		✓		
Zellers et al. (2019)	✓		✓		
Tan et al. (2020)	✓			✓	
<i>InfoSurgeon (Ours)</i>	✓	✓	✓	✓	✓

Comparison with related work on fake news detection

Knowledge Element-Level Misinformation Detection

- ❖ Combine *local* and *global* features
- ❖ Leverage external knowledge to help pinpoint misinformation

bbc.com

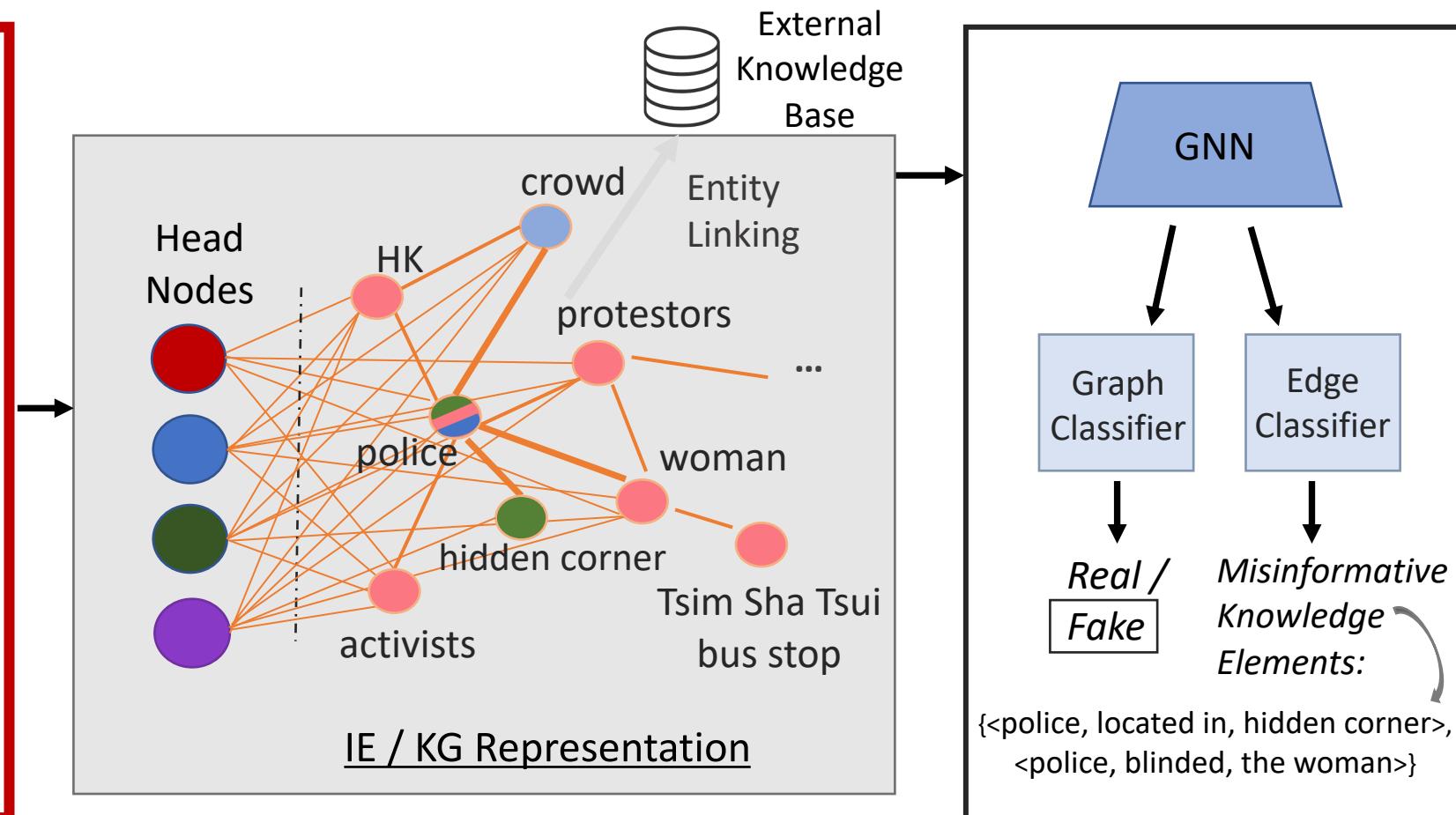
Police Brutality in HK at new Extreme Levels

Aug 11, 2019 | Lisa Lu

Police brutality has risen to a new, extreme level in HK this past weekend. HK police started shooting at protestors on the streets, including the unarmed, peaceful protestors. One notable incidence involved a woman at the Tsim Sha Tsui bus stop being shot in the eye by a policeman hiding behind corners. No warning was issued beforehand, and the woman was permanently blinded. Local activists are avidly calling for international attention on the HK police brutality.



HK police shoot cold bullets at protestors from hidden corners.

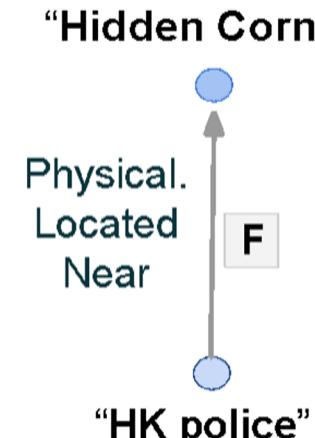


Knowledge Element-Level Misinformation Detection

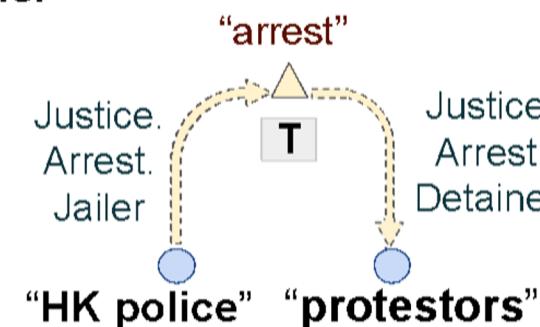
We also propose a **new task** in addition to document-level fake news detection that is more challenging but interesting.

Label each triplet connecting two entities as True/False

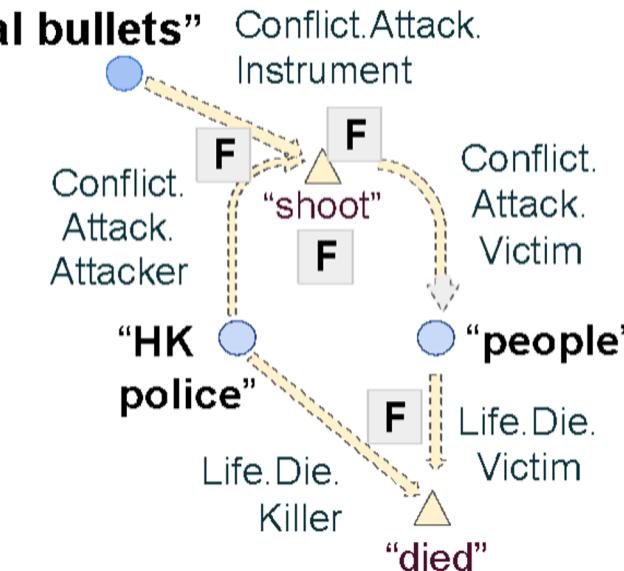
Relation:



Event:



Events/Relations:



T = True, F = False, ● = entity

Multimedia Knowledge Graph Construction

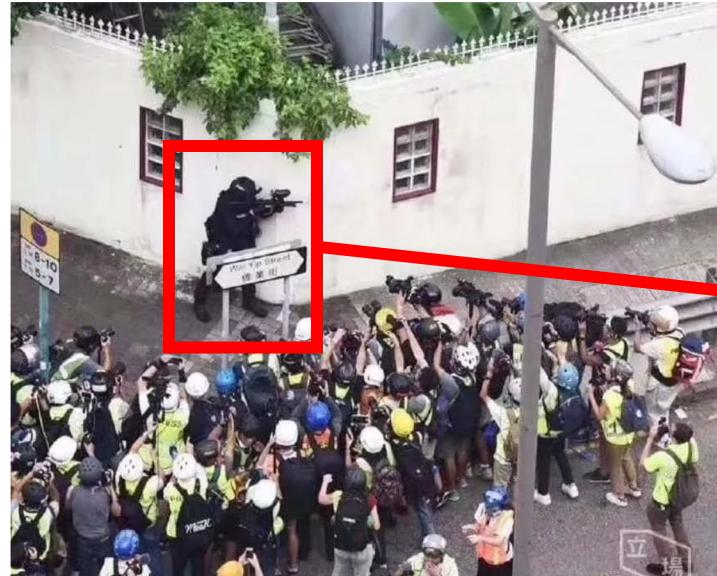
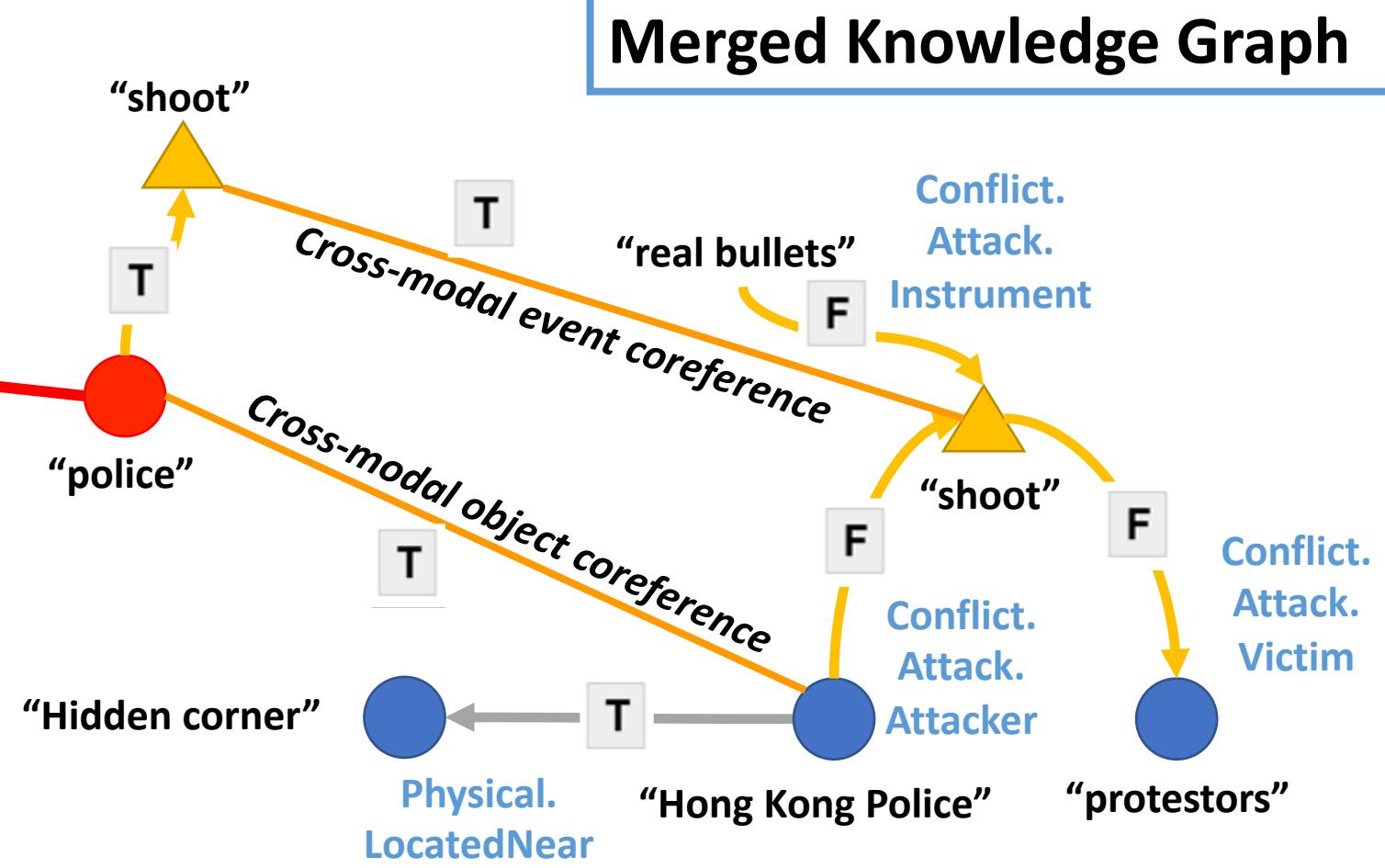


Image Caption

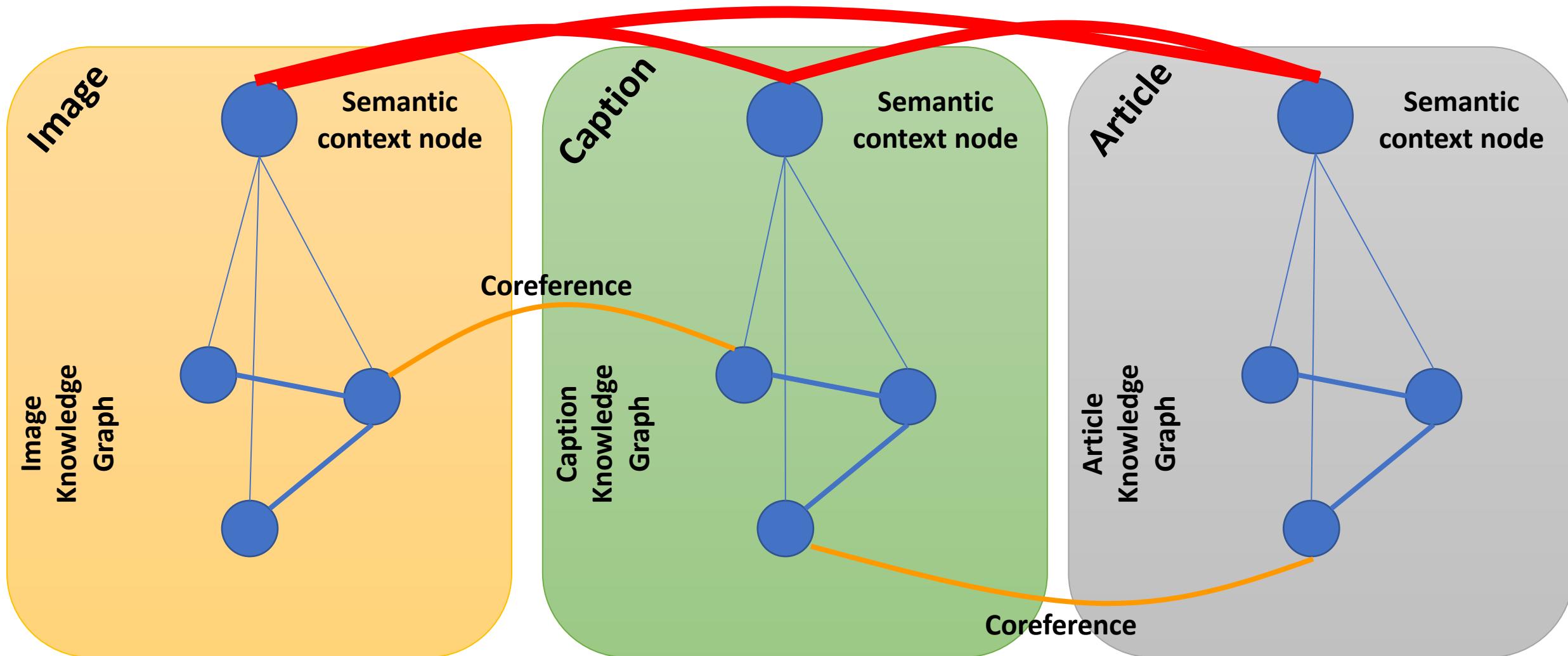
Hong Kong police shoot real bullets at protestors from hidden corners.

T = True, **F** = False, **●** = entity

▲ = event trigger



Graph Propagation



KG-Conditional Fake News Generation

- Annotating specifically which elements in a KG are fake is time-intensive / difficult
- We propose a solution to automatically obtain knowledge-element labeled knowledge graphs for free
- Given a set of real news articles, we extract KGs from the real articles
- Train a text generator model that learns to recreate an article from its KG
- To generate fake data, perform manipulation operations on the KG (editing knowledge relations, events, entities, etc.) to produce KG'
- Generate a fake article from KG'
- Key insight – We now know specifically which elements in KG' were manipulated!

Manipulated KG-to-Article Synthesis

- We perform the following manipulations on KGs:
- **Entity swapping** – Swapping entity that has same type and similar embedding (so they are harder to tell apart)
- **Addition of new relation or event** – Randomly select relation / event argument roles and append a new entity to the relation / event
- **Subgraph replacement** – Select a subgraph of the news article from an entity and replace it with a subgraph from another news article

Generating Text from Structured Representations

Article Image



Manipulated KG

```
<team, Manufacture.Artifact_Artifact.Artifact_Manufacturer,  
Zambia>  
<Zambia, Manufacture.Artifact_Manufacturer, men> ...  
<bicycle> <truck> <hospital>
```

Generated Article

The Zambia-based Bamboo **truck** is the first of its kind in the world, and it's designed to be a cheaper and lighter alternative to the traditional, heavy-duty, imported, metal ambulance. The Zambian-designed **truck** is built of bamboo, which is an environmentally friendly material...

- We manipulate knowledge graphs to synthesize fake news which contain known types of inconsistencies.

- This example is trivially detectable due to a inconsistency with the image

With Cross-media Manipulation Constraints

Article Image



Manipulated KG

```
<team, Manufacture.Artifact_Artifact.Artifact_Manufacturer,  
Zambia Fiji>  
<Zambia Fiji, Manufacture.Artif  
act_Manufacturer, men> ...  
<bicycle> <hospital> ...
```

Generated Article

A team of two Californians living in **Fiji** is trying to build the world's smallest and most affordable bicycle. They are using bamboo as the frame for their bicycles. The team is made up of 25 young men who met at a university in the Pacific island nation of **Fiji**. They're using their...

- By imposing cross-media knowledge graph manipulation constraints, we **prevent generating text with obvious inconsistencies.**

- Enables generating more realistic / challenging data for training detector

Caption Manipulation – AMR-to-Caption Synthesis

- Use text parser to get AMR graphs (Banarescu et al., 2013) from captions
- Use AMR since they capture fine-grained relations expressing who does what to whom
- Manipulations:
 - **Role switching** – Swapping entity positions in AMR graph
 - **Predicate negation** – Replace triggers / verbs with antonyms from WordNet
- Use off-the-shelf model for AMR to text synthesis (Ribeiro et al, 2020)



True Caption:

In Afghanistan, the Taliban released to the **media** this picture, which it said shows the suicide bombers who **attacked** the **army** base in Mazar-i-Sharif, April 21, 2017

Fake caption:

On 21 April 2017 the Taliban released this picture to the **army** in Afghanistan which they said was a suicide bomber **hiding** at a **media** base in the city of Mazar-i-Sharif

- **Ethical Statement: we are not going to share our generator, but sharing our detector!**

Knowledge Element-Level Misinformation Dataset

- To address the lack of data for the detection task, we further contribute a **KG2txt fake news generation** approach, which allows for control over knowledge element manipulation and creating silver standard annotation data.

	Overall	Real Documents	Fake Documents
Human Detection Accuracy	61.3%	80.4%	42.3%

The Turing Test results above show that our automatically generated fake documents are also very hard for humans to detect.

Knowledge Element-Level Misinformation Detection

Experimental result on traditional document-level detection:

	NYTimes Neural News Dataset	VOA Manipulated KG2Txt Dataset
Grover	56.0%	86.4%
DIDAN	77.6%	88.3%
InfoSurgeon (Our Model)	94.5%	92.1%

Experimental result on the novel task, knowledge element level misinformation detection:

	VOA Manipulated KG2Txt Dataset
Random (baseline)	27%
InfoSurgeon (Our Model)	37%

A Successful Example

- ❖ Example of fake news article in which baseline misses, but *InfoSurgeon* successfully detects

Image	Caption	Body Text	Misinformative KEs
	<p><i>Aerial view of Fort McHenry.</i></p>	<p><i>The battle of Fort McHenry, which took place in September of 1814, was a pivotal moment in the U.S. War of Independence...When the British finally left, they left behind a trail of destruction, including the destruction of the twin towers of the World Trade Center ...</i></p>	<p><British, Conflict.Attack, twin towers></p>

Demo 1: Multimedia Event Recommendation

Event Search

Query: Target: Януковича (Yanukovych) Number of Events: 2

Automated Summary:
Source Document Translation from Ukrainian/Russian Show Visual Knowledge Elements Hide Visual Knowledge Elements

Столкновения 20 февраля стали одним из ключевых факторов, вынудивших Президента Украины Виктора Януковича пойти на подписание Соглашения об урегулировании политического кризиса на Украине, потеряв доверия к самому Януковичу и к переформатированию парламентского большинства (Translation: The clashes on February 20 became one of the key factors that forced President of Ukraine Viktor Yanukovich to sign the Agreement on the settlement of the political crisis in Ukraine, the loss of confidence in Yanukovich himself and the reformatting of the parliamentary majority, which issued a resolution on the evening of February 20 banning the use of force), что согласно всем имеющимся уликам те милиционеры и демонстранты, что стали жертвами снайперского огня, застрелены одними и теми же снайперами (Translation: that according to all available evidence, those policemen and demonstrators shot by the same snipers)

Event Summary

Source Doc & Text Extraction Result
Not Secure | 159.89.180.81/demo/video_recom...
HC000T6CP, 2011-01-19

Visual Entity Linking

Visual Entity Extraction
Andriy Parubiy: 79%
CO.M.Equipment: 87%
COM: 42%

Event Arguments
Date: 201402 Location: Unknown Attackers: Unknown Target: Януковича (Yanukovych) Instrument: Unknown Type of Attack: Conflict.Attack

Event Type

Knowledge Elements based Ranking Incorporating User Feedback
Similar Events Dissimilar Events

Recommended Events
Date: 201402
Location: Unknown
Attackers: Unknown
Target: Unknown
Instrument: Unknown
Type of Attack: Conflict.Attack

Source Doc Translation
Not Secure | 159.89.180.81/demo/video_recom...
1 The case of Kiev snipers
2 Red Cross Volunteers of Ukraine provide first aid to a wounded man on Institutska at the beginning of the eleventh hour on February 20, 2014
3 Self-defenders carry out comrade on Institutskaya Street to the rear at the end of eleventh hour on February 20
4 Mark 13 on a pierced bullet on Institutskaya Street, pasted by criminologists near the side opposite the Maidan
5 The case of Kiev snipers question about the organizers and perpetrators of sniper Euromaidan participants and at the same time law enforcement officers in Kiev on 20, 2014, which killed 53 people (49 protesters and 4 law enforcement officers)

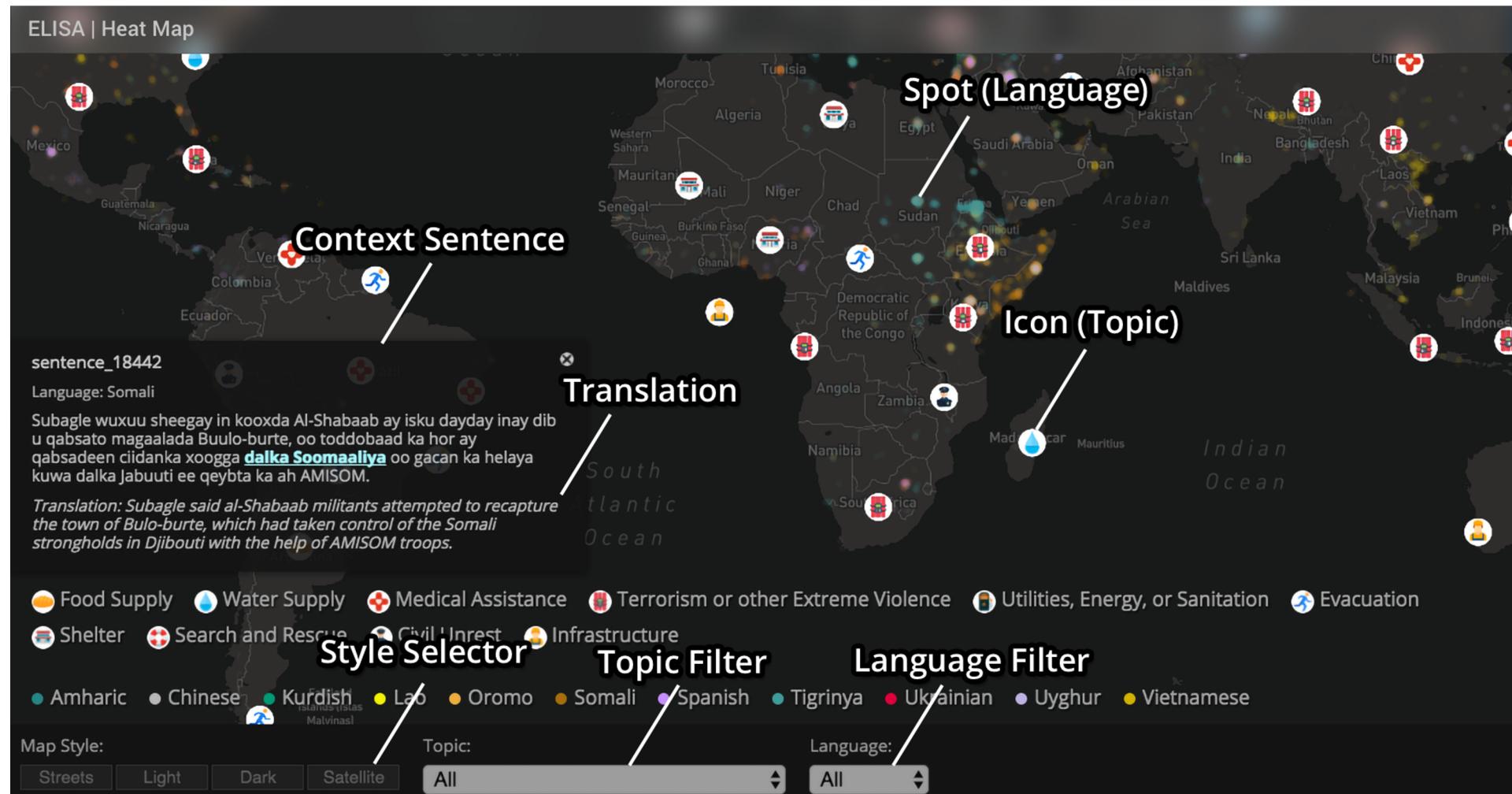
(Li et al., ACL2020 Best Demo Paper Award)

GitHub: <https://github.com/GAIA-IE/gaia>

DockerHub: <https://hub.docker.com/orgs/blenderNLP/repositories>

Demo: http://159.89.180.81/demo/video_recommendation/index_attack_dark.html

Demo 2: Event Heatmap for Disaster Relief



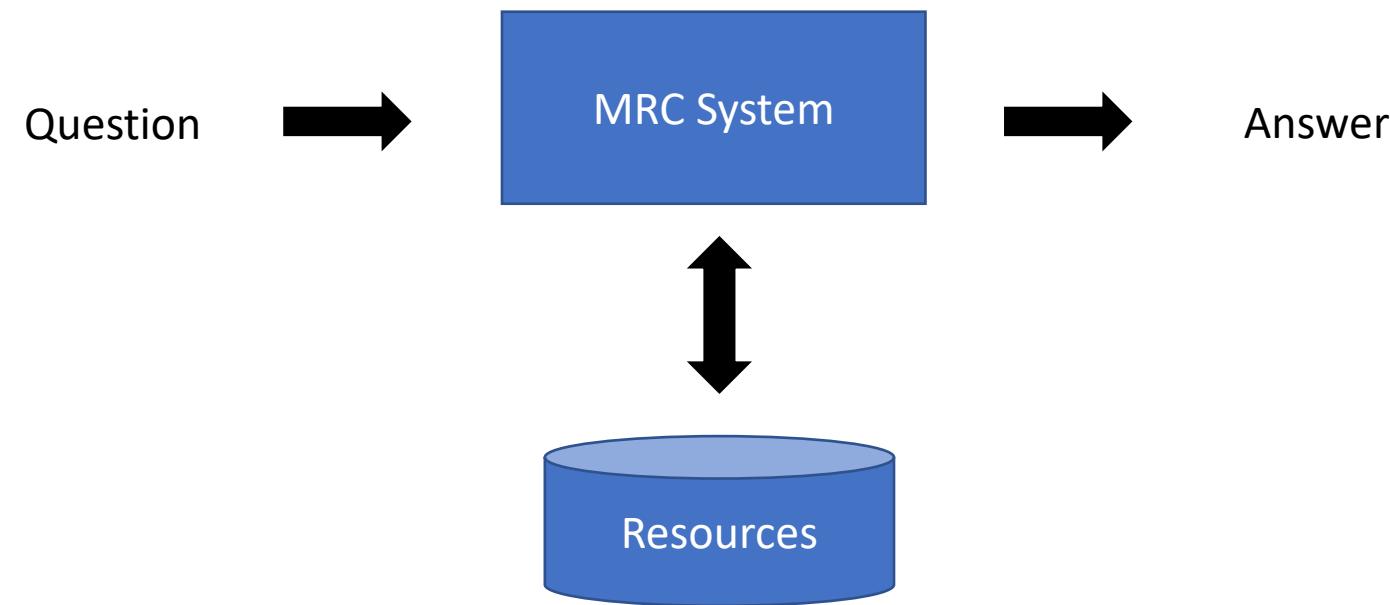
- Re-trainable Systems: http://159.89.180.81:3300/elisa_ie/api
- Demos: http://159.89.180.81:3300/elisa_ie
- Heat map: <http://159.89.180.81:8080/>

Software and Resources

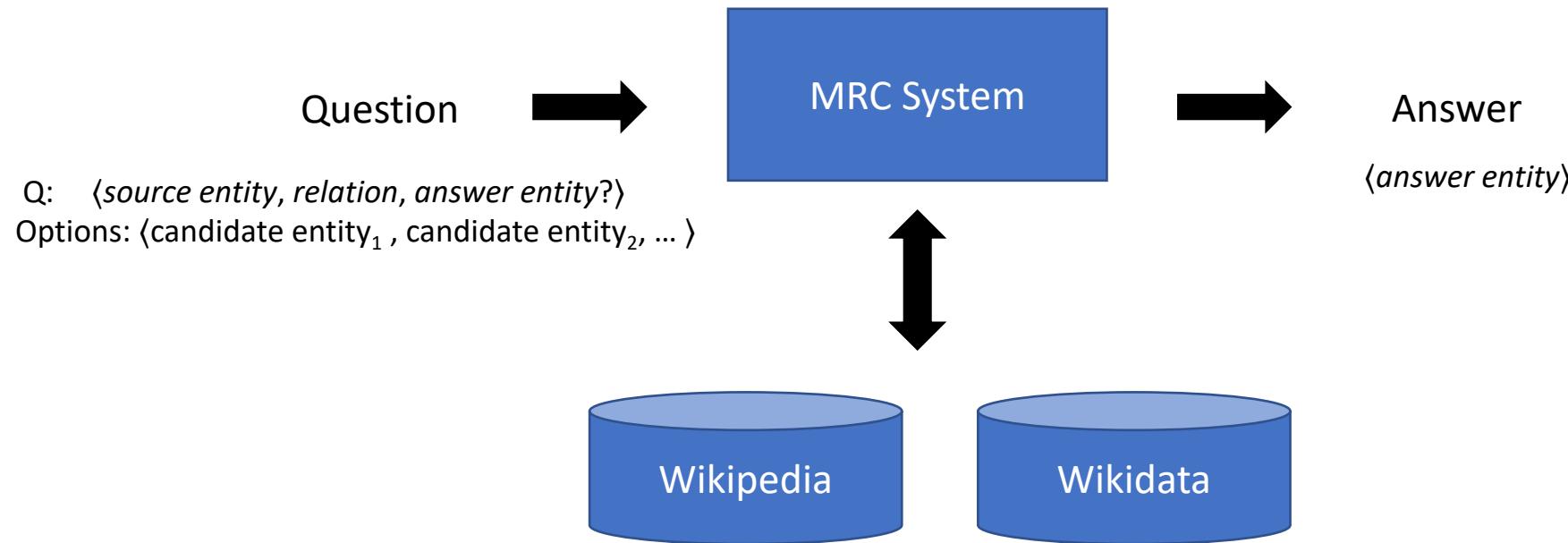
- KAIROS RESIN Cross-document Cross-lingual Cross-media Information Extraction system (Wen et al., NAACL2021 demo)
 - <https://github.com/RESIN-KAIROS/RESIN-pipeline-public>
- Joint Neural Information Extraction system (Lin et al., ACL2020)
 - <http://blender.cs.illinois.edu/software/oneie/>
- GAIA Multimedia Event Extraction system and new benchmark with annotated data set (Li et al., ACL2020 demo)
 - GitHub: https://github.com/GAIA-AIDA/uiuc_ie_pipeline_fine_grained
 - Text IE DockerHub: <https://hub.docker.com/orgs/blendernlp/>
 - Visual IE repositories: <https://hub.docker.com/u/dannapierskitoptal>

Machine Reading Comprehension

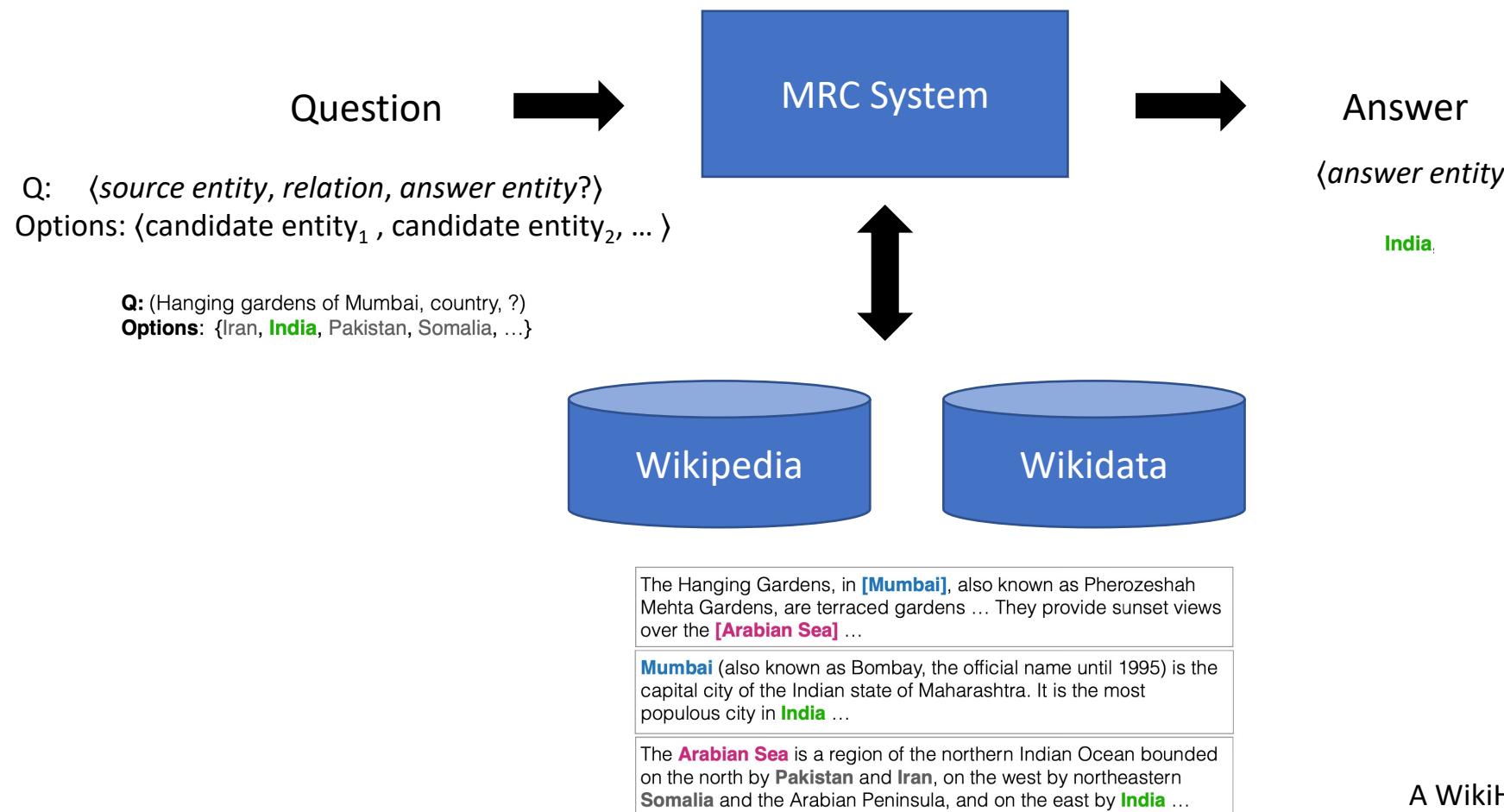
Machine Reading Comprehension (MRC)



Multi-Hop Reading Comprehension: WikiHop



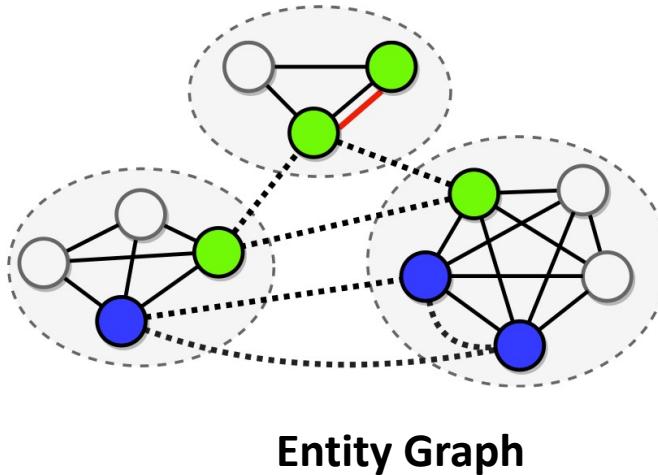
Multi-Hop Reading Comprehension: WikiHop



Multi-hop Reading Comprehension across Multiple Documents

[Cao et al. NAACL'19]

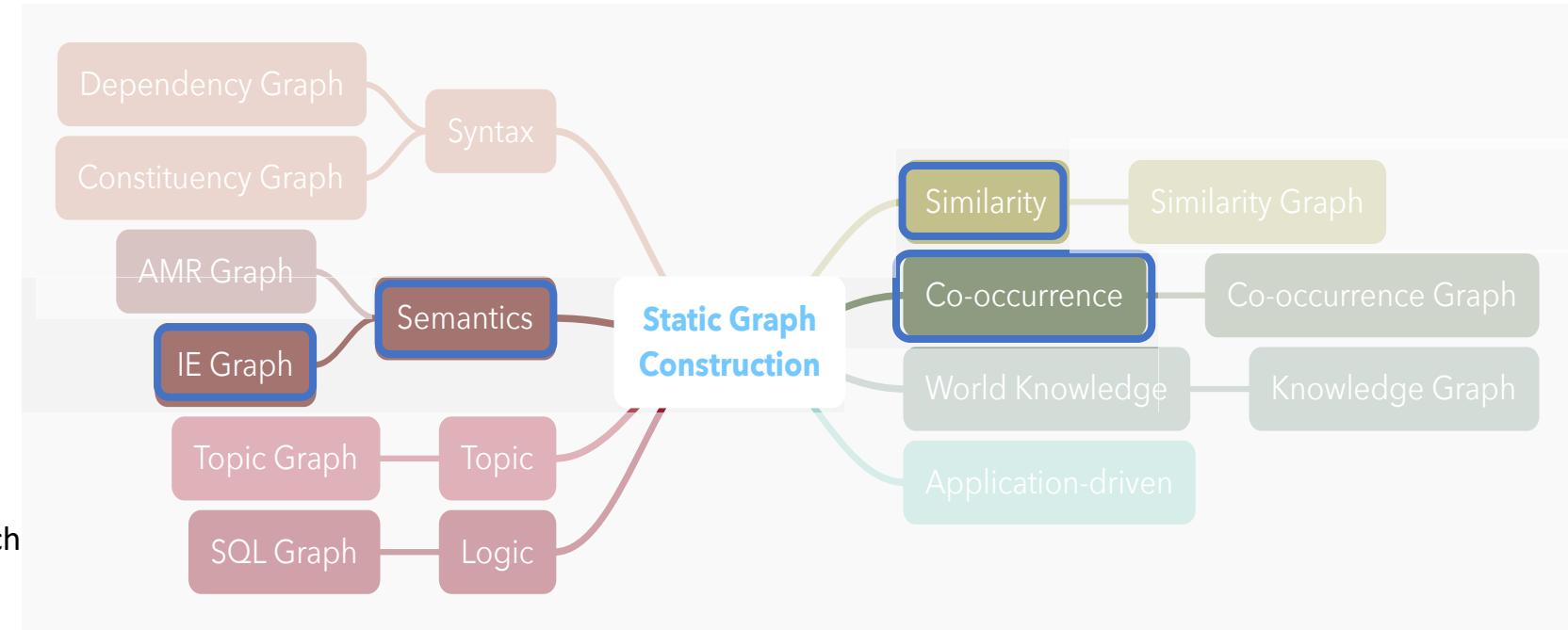
One of the first paper leveraging GNN for multi-hop reading comprehension



Node: Unambiguous mentions identified via exact match coreference

Edges

- **DOC-BASED:** co-occurrence in the same document
- **MATCH:** exact match (*most reliable but sparser*)
- **COREF:** co-reference (*less reliable*)
- **COMPLEMENT:** for nodes not connected otherwise



Multi-hop Reading Comprehension across Multiple Documents

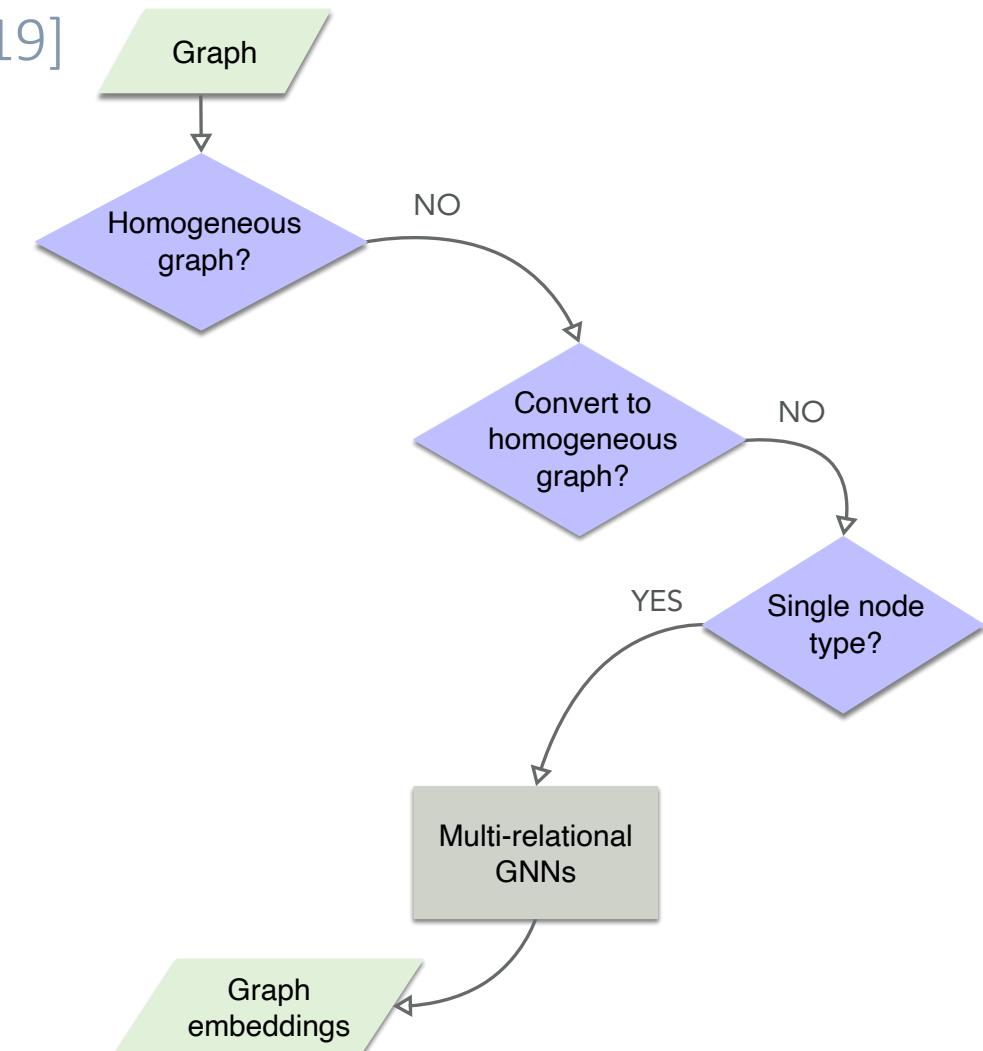
[Cao et al. NAACL'19]

$$\mathbf{u}_i^{(\ell)} = f_s(\mathbf{h}_i^{(\ell)}) + \frac{1}{|N_i|} \sum_{j \in N_i} \sum_{r \in R_{ij}} f_r(\mathbf{h}_j^{(\ell)}) \quad \left. \right\} \text{The update vector } (\mathbf{u}) \text{ of a node is a function of its neighbours } (N) \text{ conditioned on the relations between them}$$

$$\mathbf{a}_i^{(\ell)} = \sigma \left(f_a \left([\mathbf{u}_i^{(\ell)}, \mathbf{h}_i^{(\ell)}] \right) \right) \quad \left. \right\} \text{attention gate}$$

$$\mathbf{h}_i^{(\ell+1)} = \phi(\mathbf{u}_i^{(\ell)}) \odot \mathbf{a}_i^{(\ell)} + \mathbf{h}_i^{(\ell)} \odot (1 - \mathbf{a}_i^{(\ell)}) \quad \left. \right\} \text{the new node embedding}$$

Entity Relational Graph
Convolutional Network
(Gated RGCN)



Multi-hop Reading Comprehension across Multiple Documents

[Cao et al. NAACL'19]

$$P(c | q, C_q, S_q) \propto \exp \left(\max_{i \in M_c} f_o([\mathbf{q}, \mathbf{h}_i^{(L)}]) \right)$$

↑
final node embedding

question representation

Candidates scoring
2-layers MLP

use the final node embeddings and the question representation to predict a distribution over candidates.

Multi-hop Reading Comprehension across Multiple Documents

[Cao et al. NAACL'19]

Model	
Full (ensemble)	68.5
Full (single)	65.1 ± 0.11
GloVe w/o R-GCN	51.2
GloVe w/ R-GCN	59.2
No relation type	62.7
No DOC-BASED	62.9
No MATCH	64.3
No COREF	64.8
No COMPLEMENT	64.1
Induced edges	61.5

Figure 2. Accuracy on WikiHop validation set.

R-GCN is useful

Relations are important

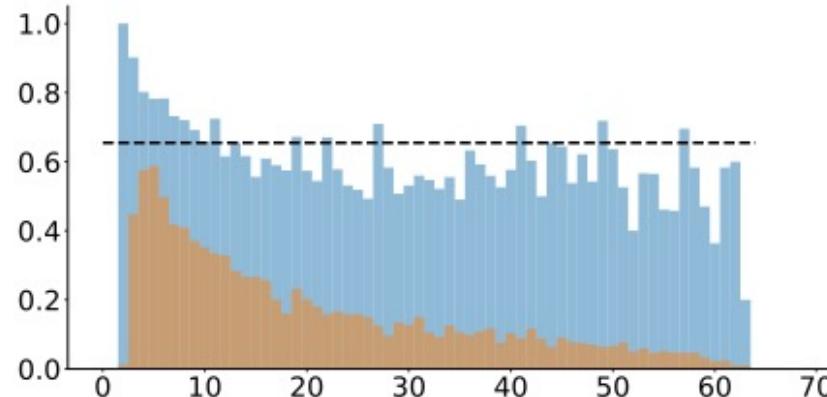
DOC-BASED relations are the most significant

Learning to predict edges is hard and an open problem

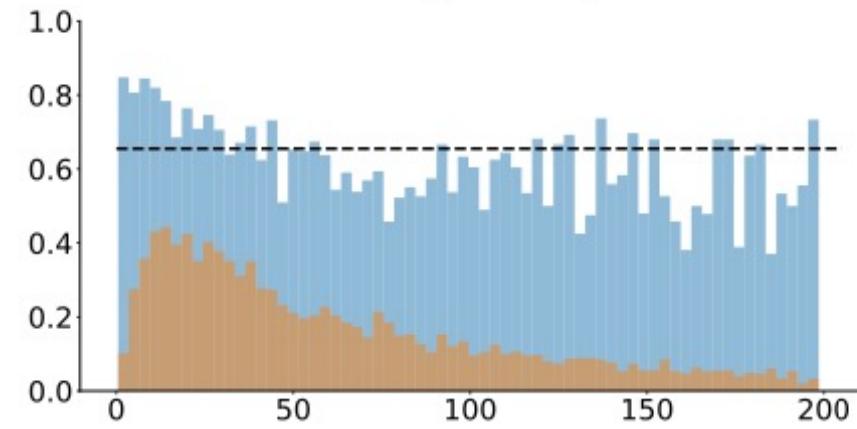
Multi-hop Reading Comprehension across Multiple Documents

[Cao et al. NAACL'19]

Performance gradually \downarrow as # candidates or # nodes \uparrow



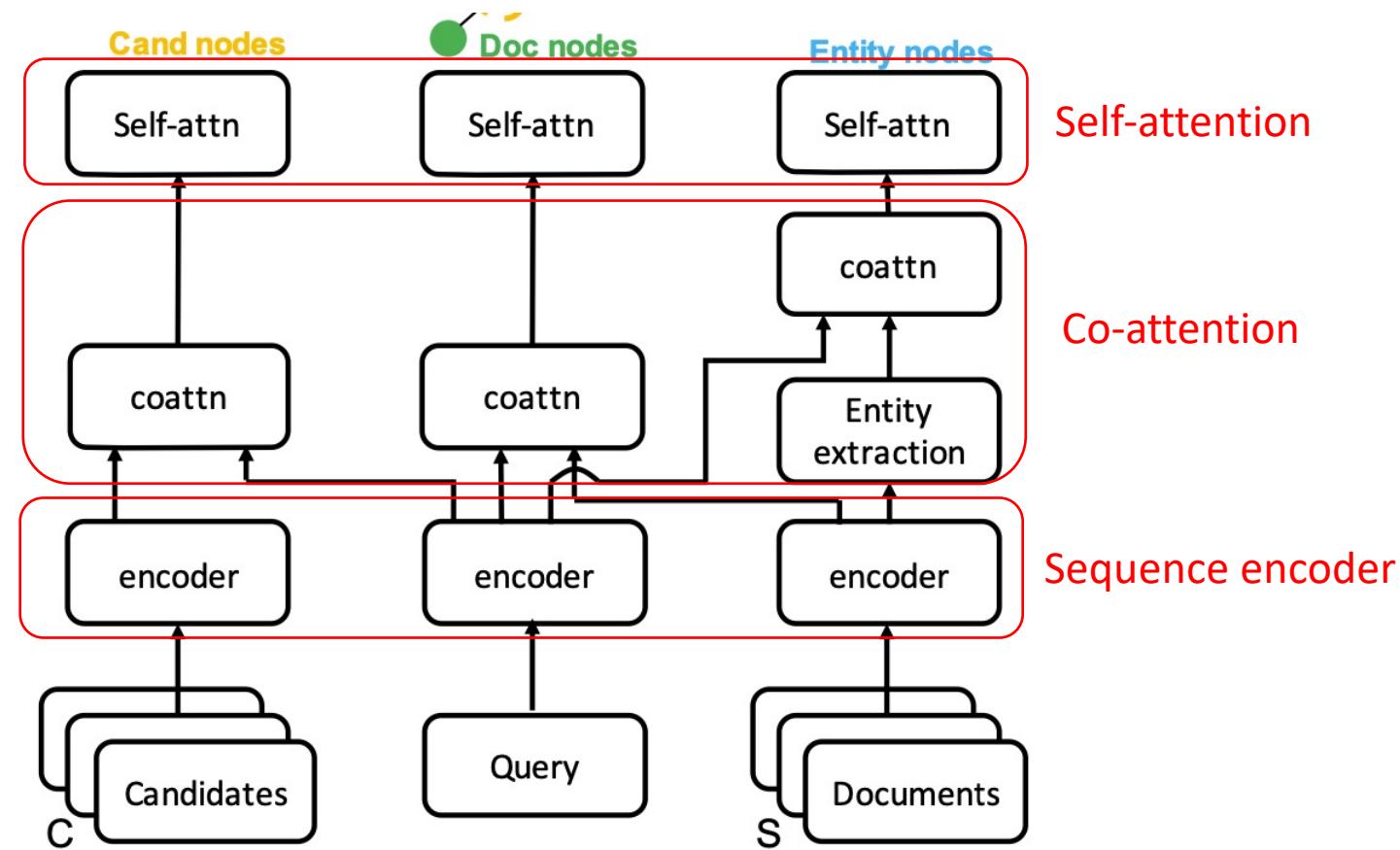
(a) Candidates set size (x-axis) and accuracy (y-axis). Pearson's correlation of -0.687 ($p < 10^{-7}$).



(b) Nodes set size (x-axis) and accuracy (y-axis). Pearson's correlation of -0.385 ($p < 10^{-7}$).

Multi-hop Reading Comprehension across Multiple Documents

[Tu et al. ACL'19]



Tu et al. Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs. ACL'19.

Multi-hop Reading Comprehension across Multiple Documents

[Tu et al. ACL'19]

Document & Candidate: if the candidate appear in the document

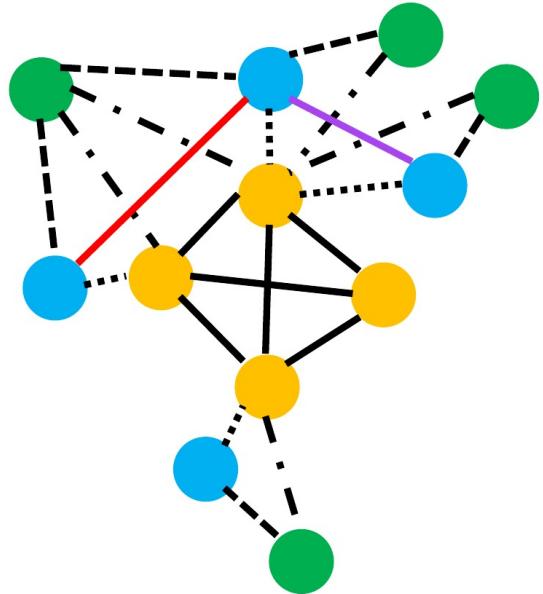
Document & Entity: If the entity if extracted from the document

Candidate & Entity: If the entity if a mention of the candidate

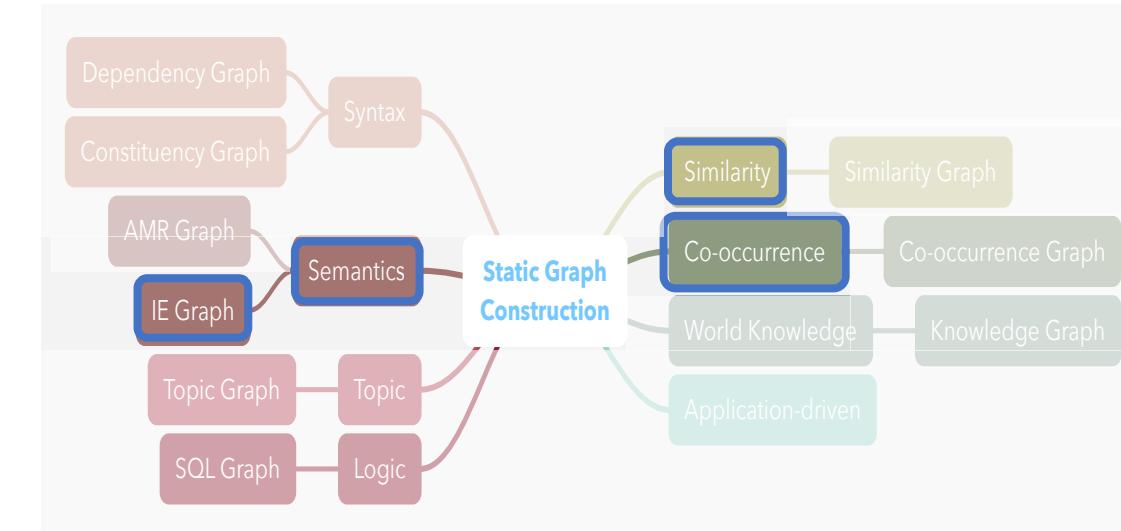
Entity & Entity: If they are extracted from the same document

Entity & Entity: If they are extracted mentions of the same candidate or query subjects from different documents

Entity & Entity: other pairs of entities



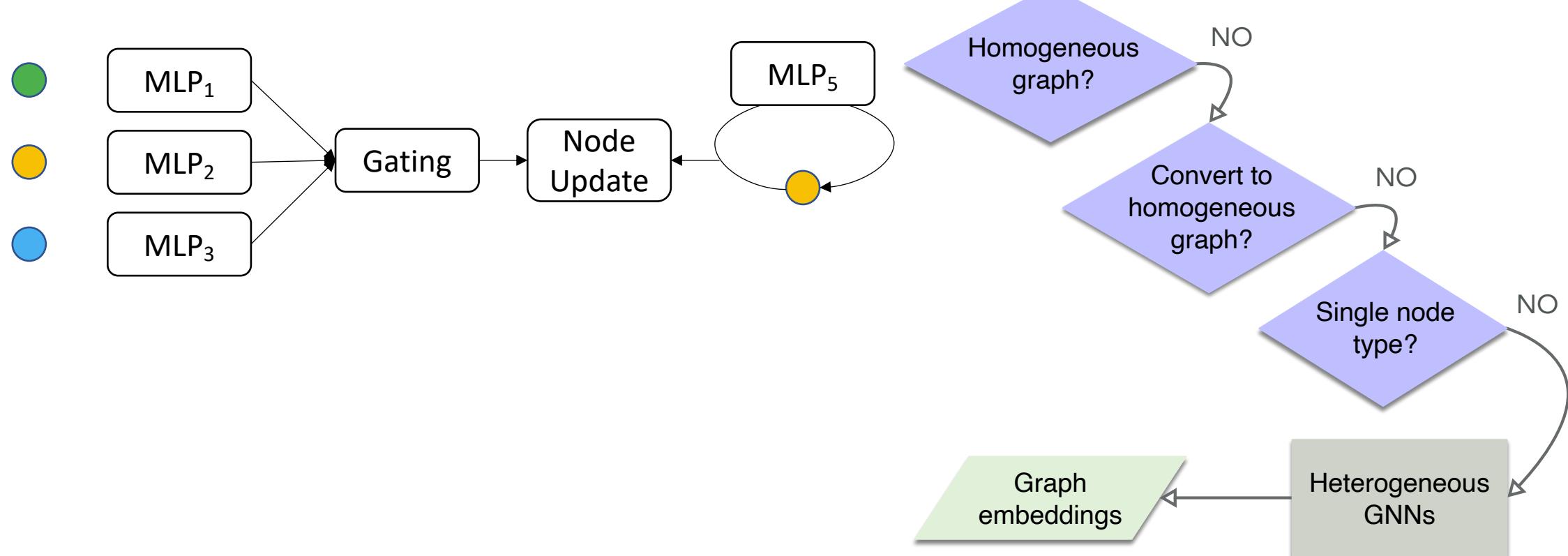
Heterogeneous Document-Entity Graph



Tu et al. Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs. ACL'19.

Multi-hop Reading Comprehension across Multiple Documents

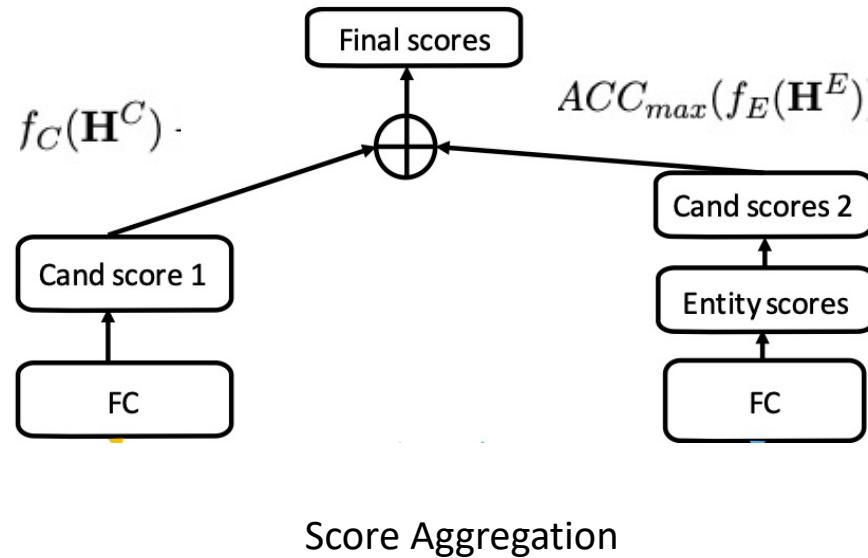
[Tu et al. ACL'19]



Tu et al. Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs. ACL'19.

Multi-hop Reading Comprehension across Multiple Documents

[Tu et al. ACL'19]



takes the maximum over scores of entities
that belong to the same candidate.

Multi-hop Reading Comprehension across Multiple Documents

Model	Accuracy (%)	
	Dev	Δ
Full model	68.1	-
- HDE graph	65.5	2.6
- different edge types	66.7	1.4
- candidate nodes scores	67.1	1.0
- entity nodes scores	66.6	1.5
- candidate nodes	66.2	1.9
- document nodes	67.6	0.5
- entity nodes	63.6	4.5

Table 2: Ablation results on the WIKIHOP dev set.

Multi-hop Reading Comprehension across Multiple Documents

[Tu et al. ACL'19]

Performance ↓ as # support documents or # candidates ↑

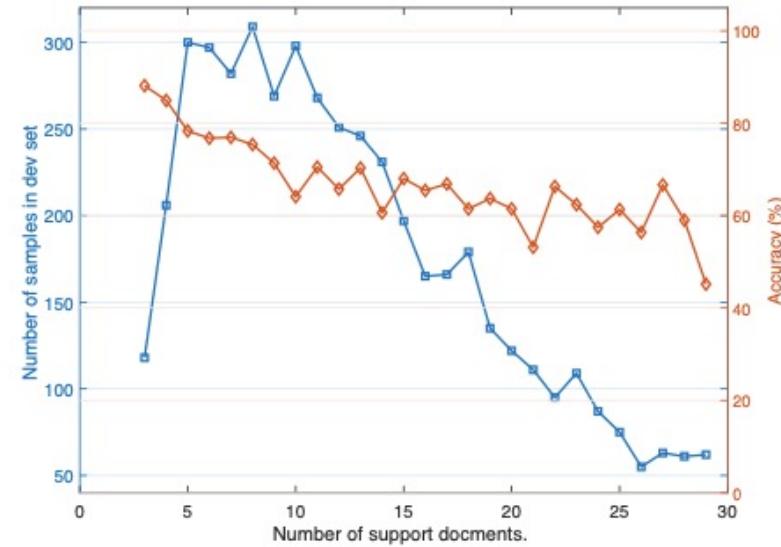


Figure 4: Plots between number of support documents (x-axis) and number of examples (left y-axis), and between number of support documents and accuracy (right y-axis).

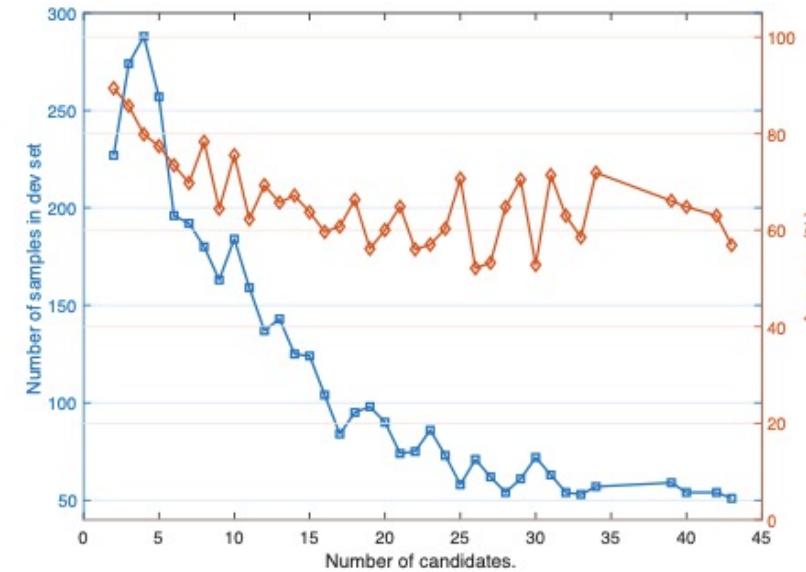
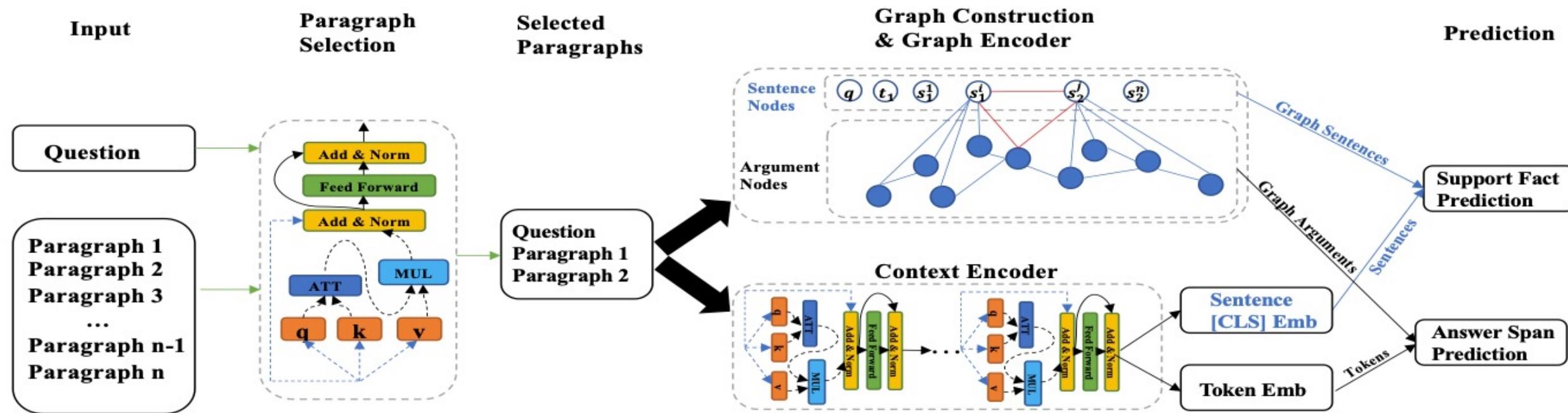


Figure 5: Plots between number of candidates (x-axis) and number of examples (left y-axis), and between number of candidates and accuracy (right y-axis).

Tu et al. Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs. ACL'19.

Multi-hop Reading Comprehension across Multiple Documents

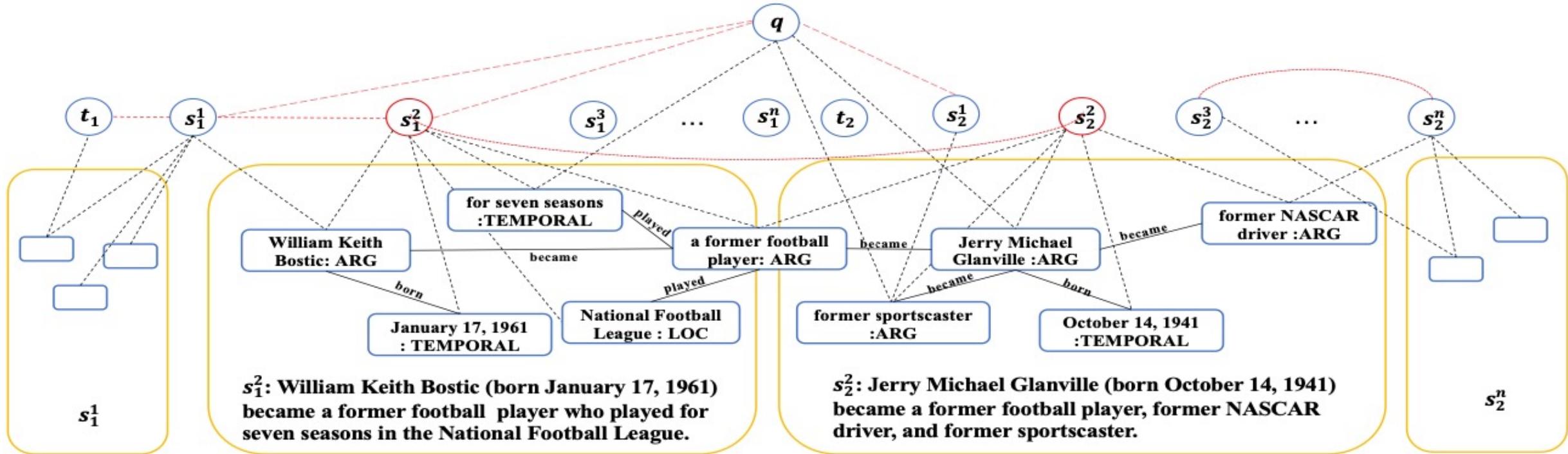
[Zheng et al. EMNLP'19]



Zheng and Kordjamshidi. **SRLGRN: Semantic Role Labeling Graph Reasoning Network**. EMNLP'20.

Multi-hop Reading Comprehension across Multiple Documents

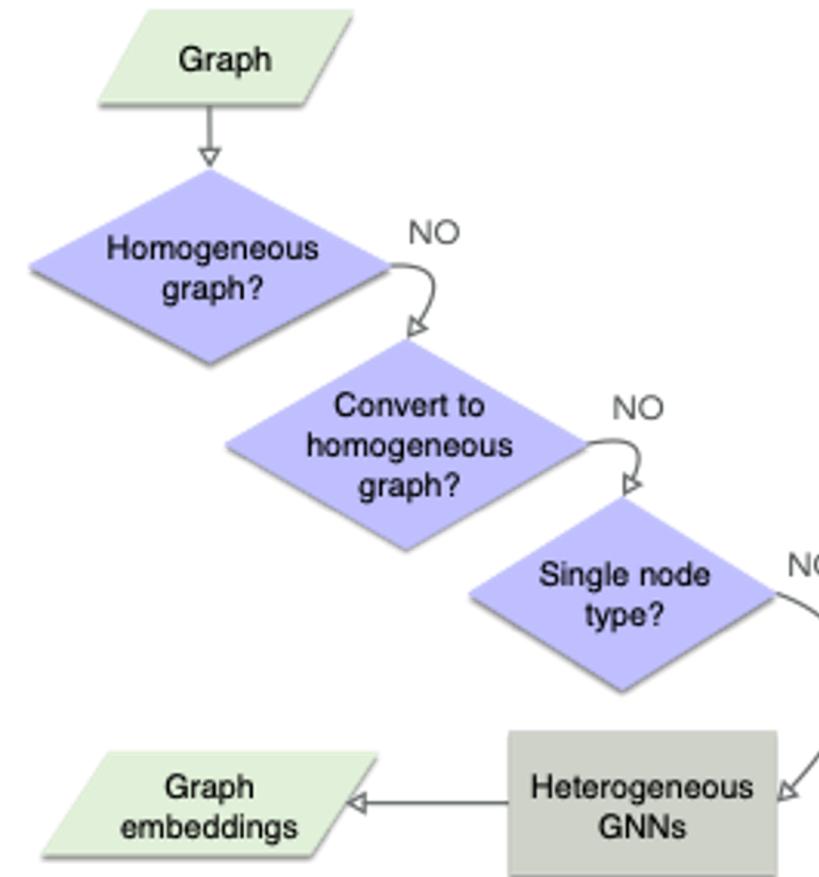
[Zheng et al. EMNLP'19]



- Argument nodes (with argument phrase and type)
- Document-level nodes (questions, titles and other sentences)
- SRL argument-predicate sub-graphs for sentences
- Predicate edges
- - - Edges connecting sentence nodes and argument nodes
- - - - Edges connecting sentences with shared arguments (exact match)

Multi-hop Reading Comprehension across Multiple Documents

[Zheng et al. EMNLP'19]



Zheng and Kordjamshidi. **SRLGRN: Semantic Role Labeling Graph Reasoning Network**. EMNLP'20.

Multi-hop Reading Comprehension across Multiple Documents

[Zheng et al. EMNLP'19]

HotpotQA Result on Distractor setting.

Model	Ans(%)		Sup(%)		Joint(%)	
	EM	F1	EM	F1	EM	F1
Baseline Model (Yang et al., 2018)	45.60	59.02	20.32	64.49	10.83	40.16
KGNN (Ye et al., 2019)	50.81	65.75	38.74	76.79	22.40	52.82
QFE (Nishida et al., 2019)	53.86	68.06	57.75	84.49	34.63	59.61
DecompRC (Min et al., 2019)	55.20	69.63	-	-	-	-
DFGN (Xiao et al., 2019)	56.31	69.69	51.50	81.62	33.62	59.82
TAP	58.63	71.48	46.84	82.98	32.03	61.90
SAE-base (Tu et al., 2019)	60.36	73.58	56.93	84.63	38.81	64.96
ChainEx (Chen et al., 2019)	61.20	74.11	-	-	-	-
HGN-base (Fang et al., 2019)	-	74.76	-	86.61	-	66.90
SRLGRN-base	62.65	76.14	57.30	85.83	39.41	66.37

SRL graph improves the completeness of the graph network over NER graph

Zheng and Kordjamshidi. **SRLGRN: Semantic Role Labeling Graph Reasoning Network**. EMNLP'20.

Multi-hop Reading Comprehension across Multiple Documents

[Zheng et al. EMNLP'19]

Ablation	Model	Ans(%)	
		EM	F1
Graph	w/o graph	53.06	67.68
	w/o Argument type and Semantic edge	60.10	73.24

SRL graph is important

Argument type and Predicate edges are important

Multi-hop Reading Comprehension across Multiple Documents

[Zheng et al. EMNLP'19]

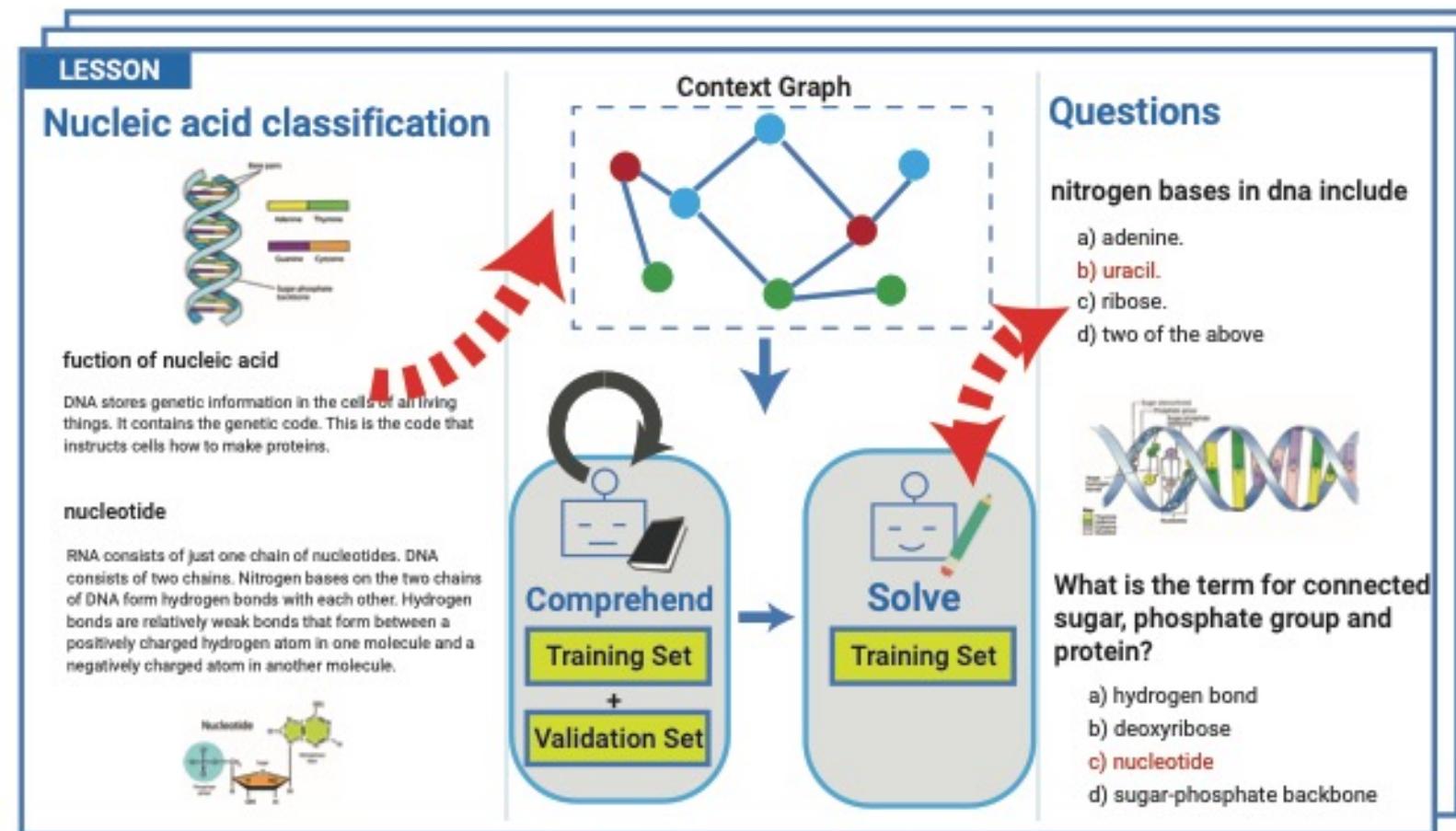
Error Types

Error Type	Model Prediction	Label
Synonyms	washington dc sars ey writer	district of columbia severe acute respiratory syndrome ernst young author
MLV	australian hessian mcdonald's, co	australia hessians mcdonalds
Month-Year	1946 25, november, 2015 10, july, 1873	1945 3, december 1, september, 1864
Number	11 fourth 2402	10 4 5922
External Knowledge	Coker	NCAA I FBS football
Other	taylor, swift film fourteenth	usher documentary 500th episode

Zheng and Kordjamshidi. **SRLGRN: Semantic Role Labeling Graph Reasoning Network**. EMNLP'20.

TQA: Textbook Question Answering

[Kim et al., ACL'19]



Kim et al. Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension ACL'19

Textbook Question Answering

[Kim et al., ACL'19]

- Key challenges:
 - Multi-modality
 - Long context
 - Difficulty to solve unseen problems

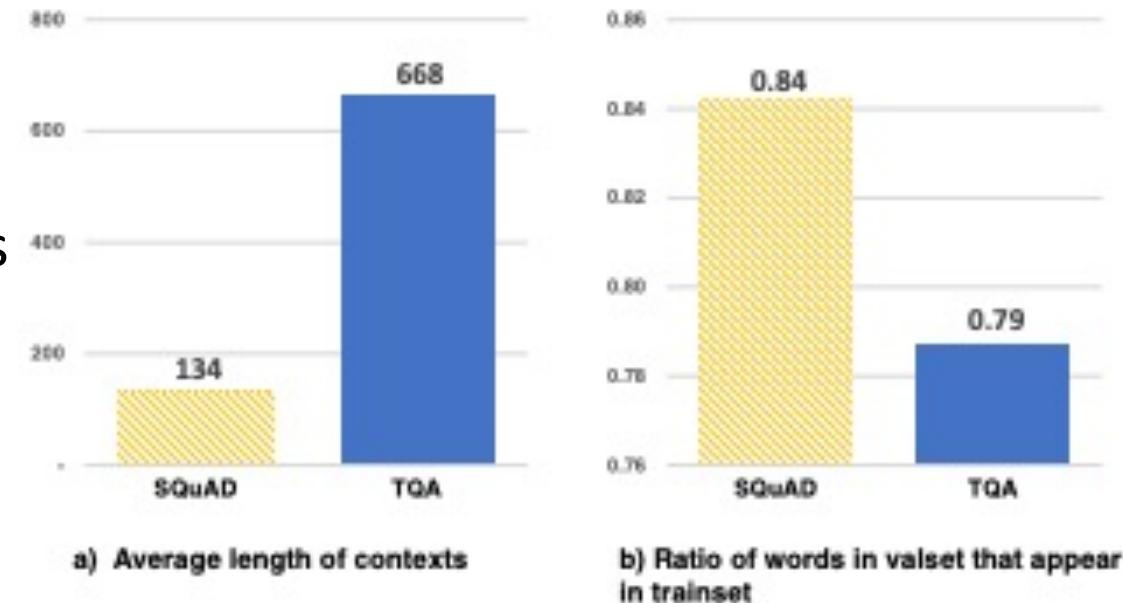
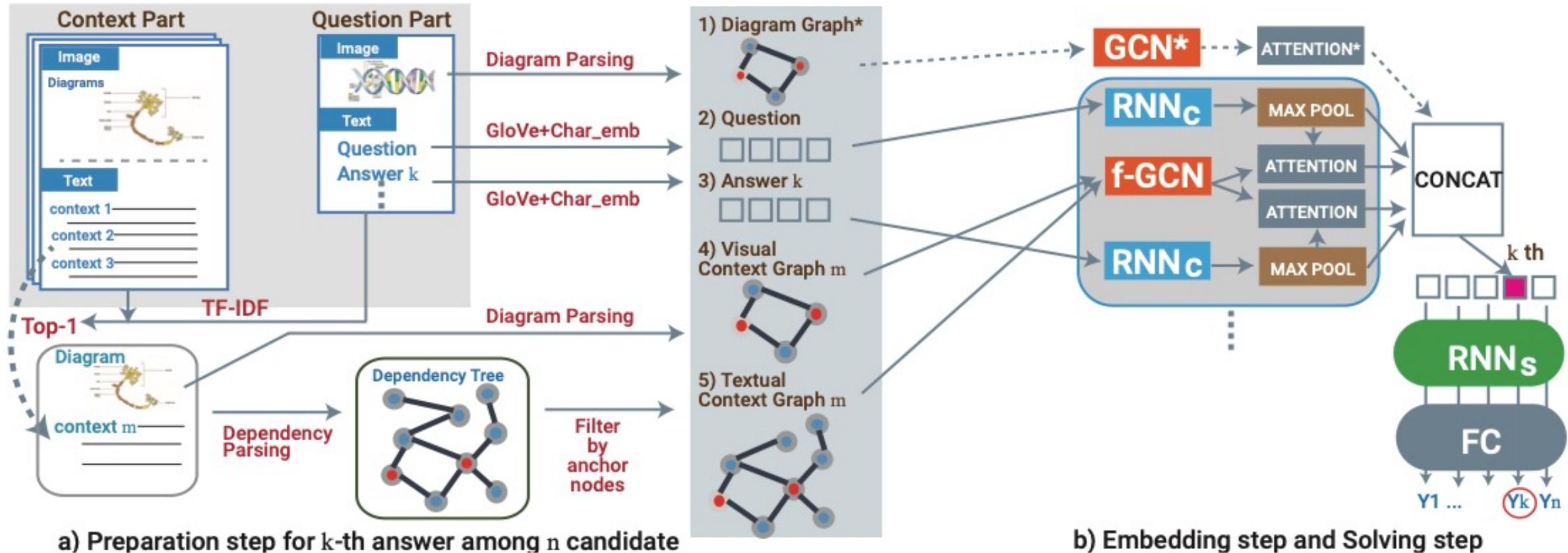


Figure 2: Analysis of contexts in TQA and SQuAD datasets.

Textbook Question Answering

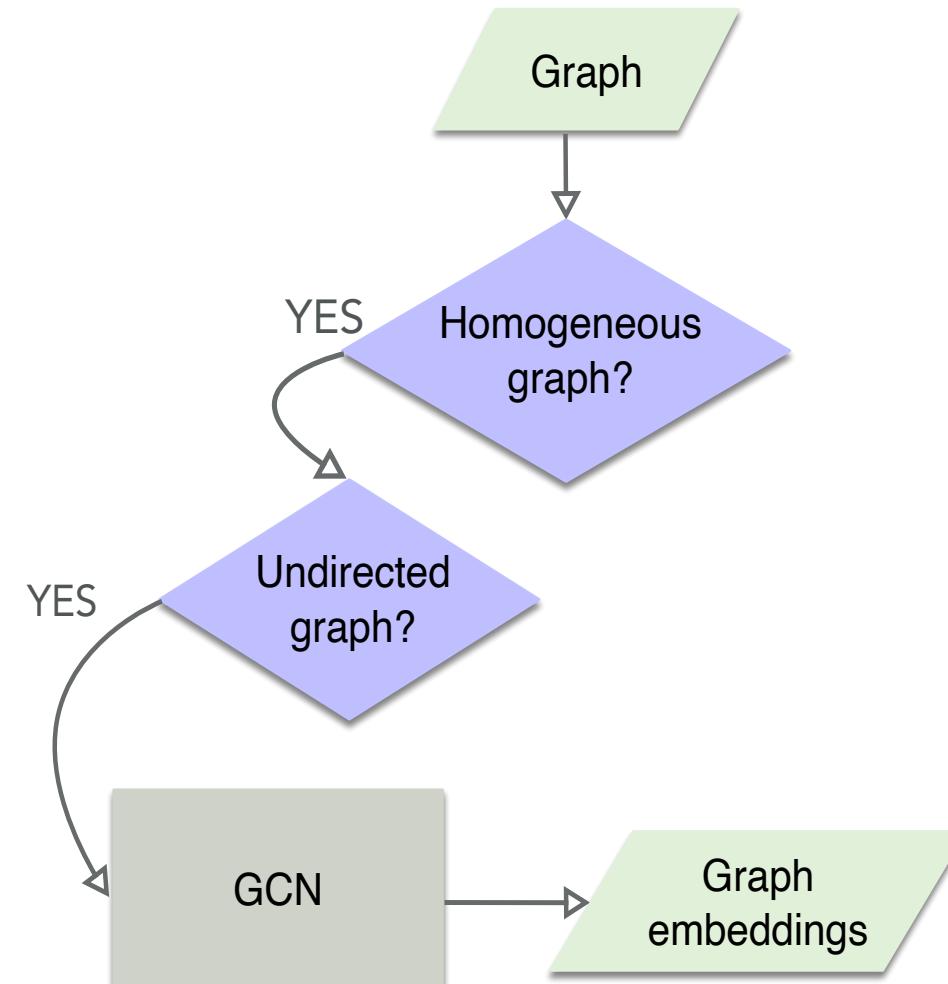
[Kim et al., ACL'19]



Kim et al. **Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension** ACL'19

TQA: Textbook Question Answering

[Kim et al., ACL'19]

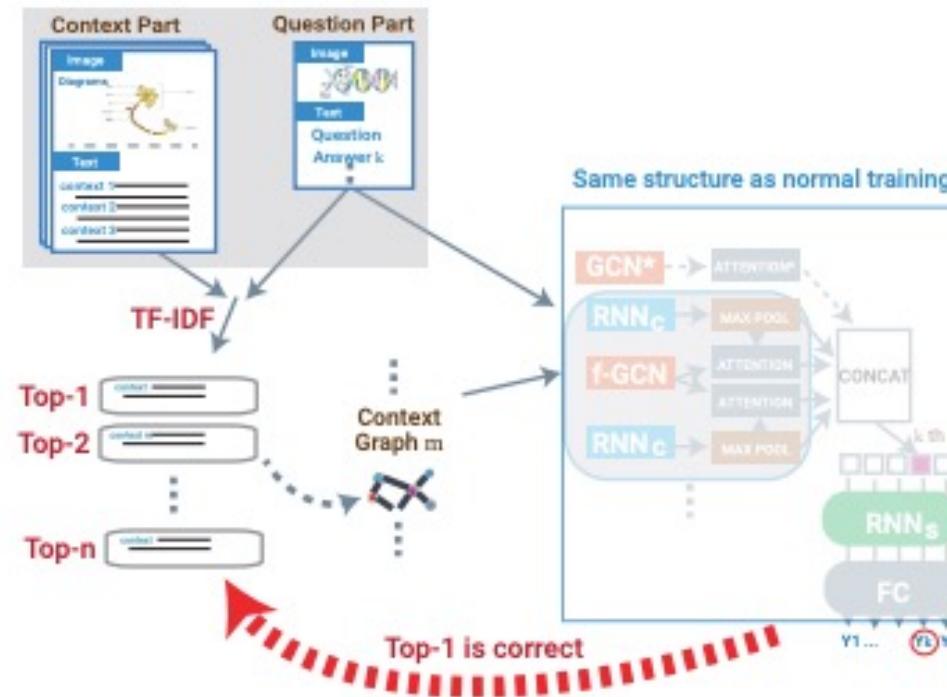


Kim et al. **Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension** ACL'19

TQA: Textbook Question Answering

[Kim et al., ACL'19]

- Self-training: reads and understands a textbook and problems in advance.



Kim et al. **Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension** ACL'19

TQA: Textbook Question Answering

[Kim et al., ACL'19]

Our full model

w/o SSOC(VAL)

w/o SSOC(TR+VAL)

w/o f-GCN & SSOC(TR+VAL)

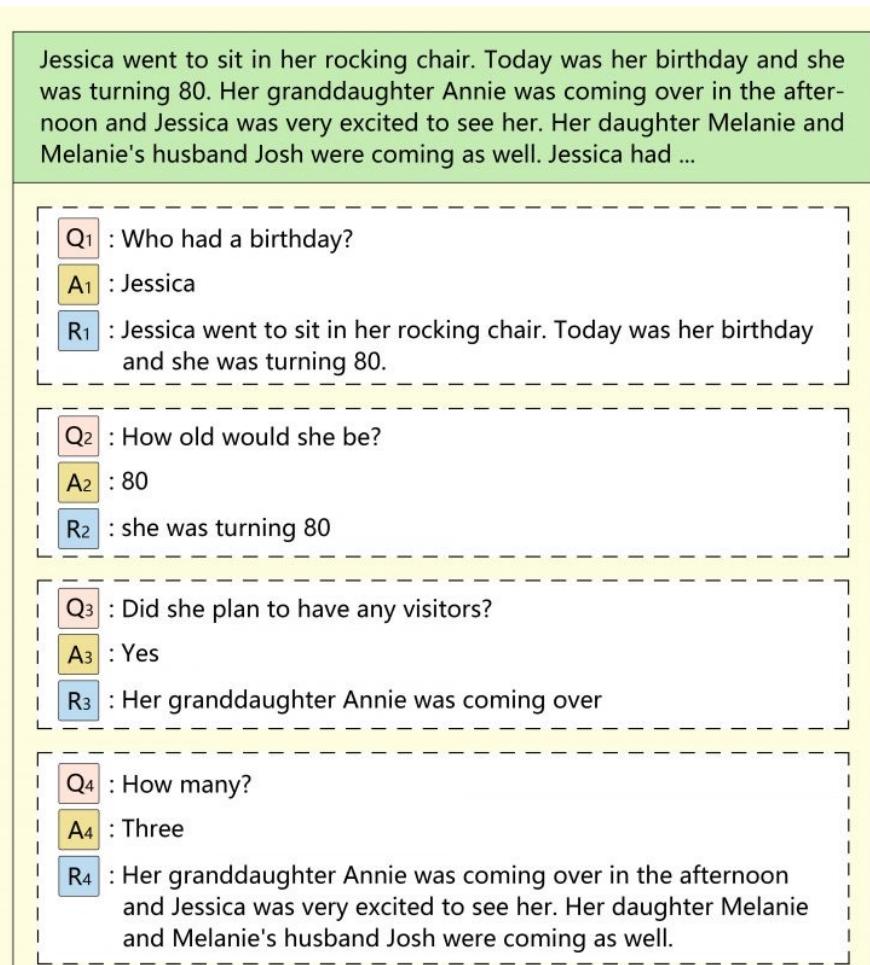
45.77
45.39
43.97
42.74

Self-training, even only
on training, is useful

Context Graph is
effective

Conversational Machine Comprehension

[Chen et al, IJCAI'20]

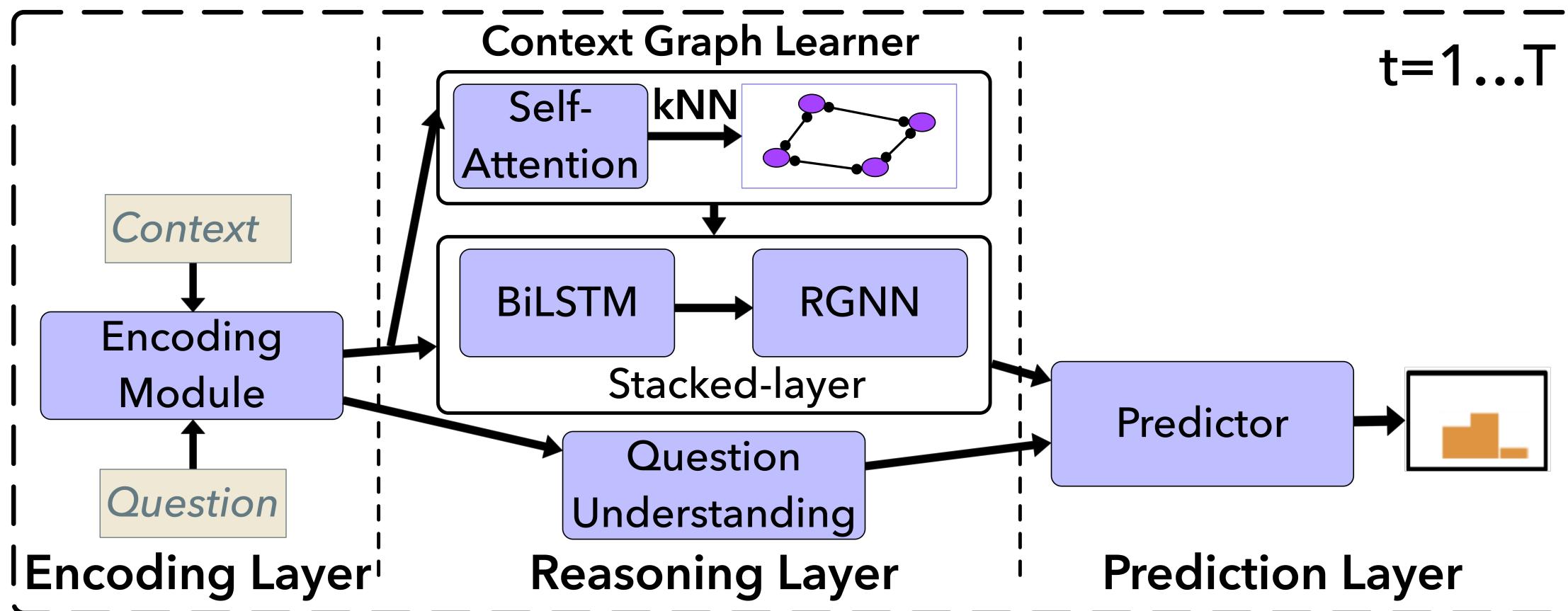


Challenges:

- Focus shift
- Coreference or ellipsis

Conversational Machine Comprehension

[Chen et al, IJCAI'20]

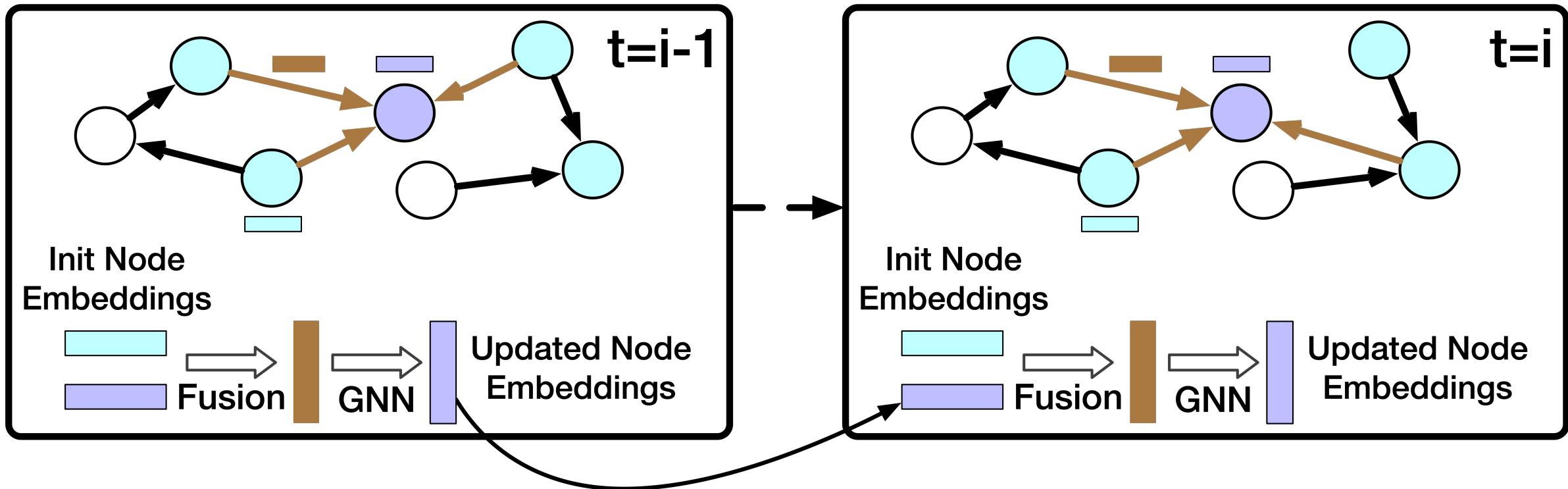


Chen et al. GraphFlow: Exploiting Conversation Flow with Graph Neural Networks for Conversational Machine Comprehension IJCAI'20

Conversational Machine Comprehension

[Chen et al, IJCAI'20]

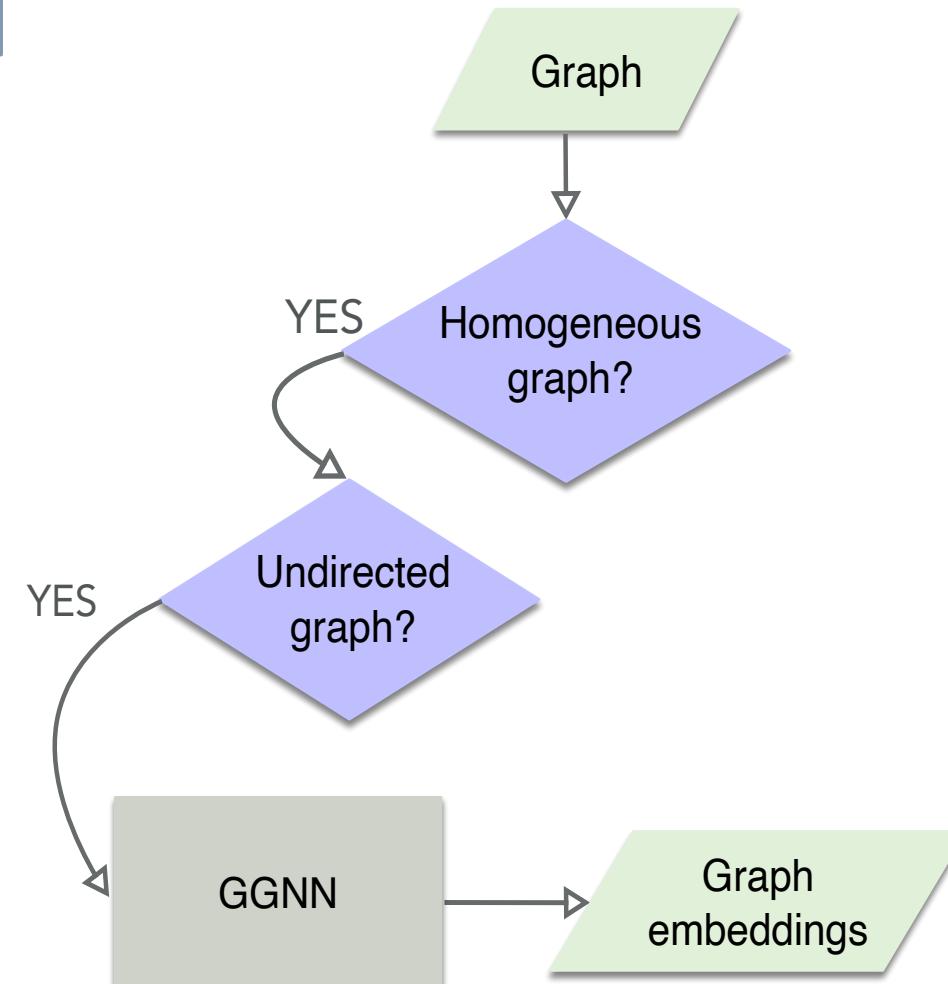
Dynamically construct a question and conversation history aware context graph at each turn.



Chen et al. **GraphFlow: Exploiting Conversation Flow with Graph Neural Networks for Conversational Machine Comprehension** IJCAI'20

Conversational Machine Comprehension

[Chen et al, IJCAI'20]



Chen et al. **GraphFlow: Exploiting Conversation Flow with Graph Neural Networks for Conversational Machine Comprehension** IJCAI'20

Conversational Machine Comprehension

[Chen et al, IJCAI'20]

GraphFlow's **focus shifts** between consecutive question turns.

Q1: Who went to the farm? -> Q2: Why?

Billy went to the farm to buy some beef for his brother 's birthday .
When he arrived there he saw that all six of the cows were sad and had brown spots . The cows were all eating their breakfast in a big grassy meadow . He thought that the spots looked very strange so he went closer to the cows to get a better look ...

Q2: Why? -> Q3: For what?

Billy went to the farm to buy some beef for his brother 's birthday .
When he arrived there ... After Billy got a good look at the cows he went to the farmer to buy some beef . The farmer gave him four pounds of beef for ten dollars . Billy thought that ...

Q3: For what? -> Q4: How many cows did he see there?

Billy went to the farm to buy some beef for his brother 's birthday .
When he arrived there he saw that all six of the cows were sad and had brown spots . The cows were ...

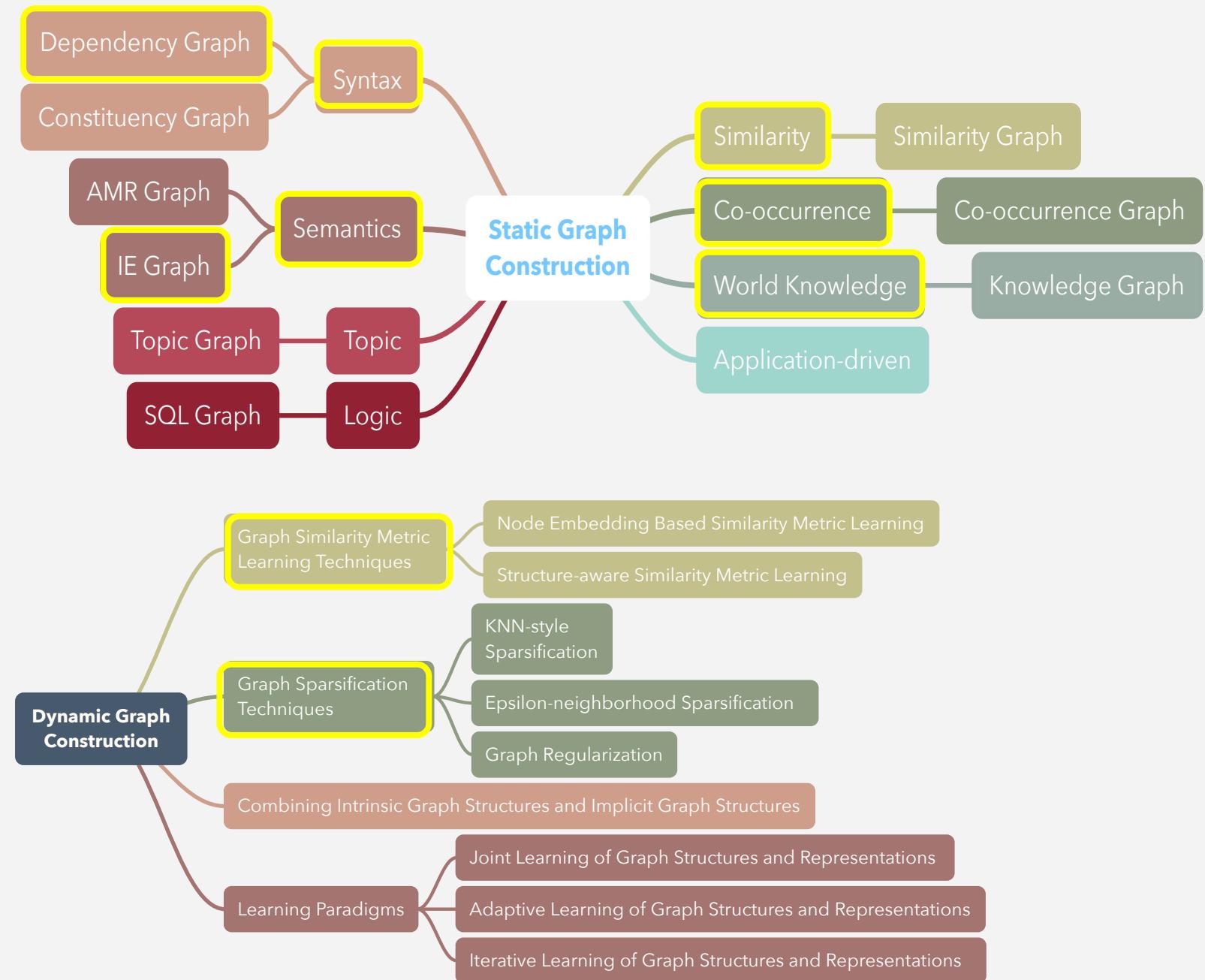
Q4: How many cows did he see there? -> Q5: Did they have spots?

Billy went to ... When he arrived there he saw that all six of the cows were sad and had brown spots . The cows were all eating ...

Summary

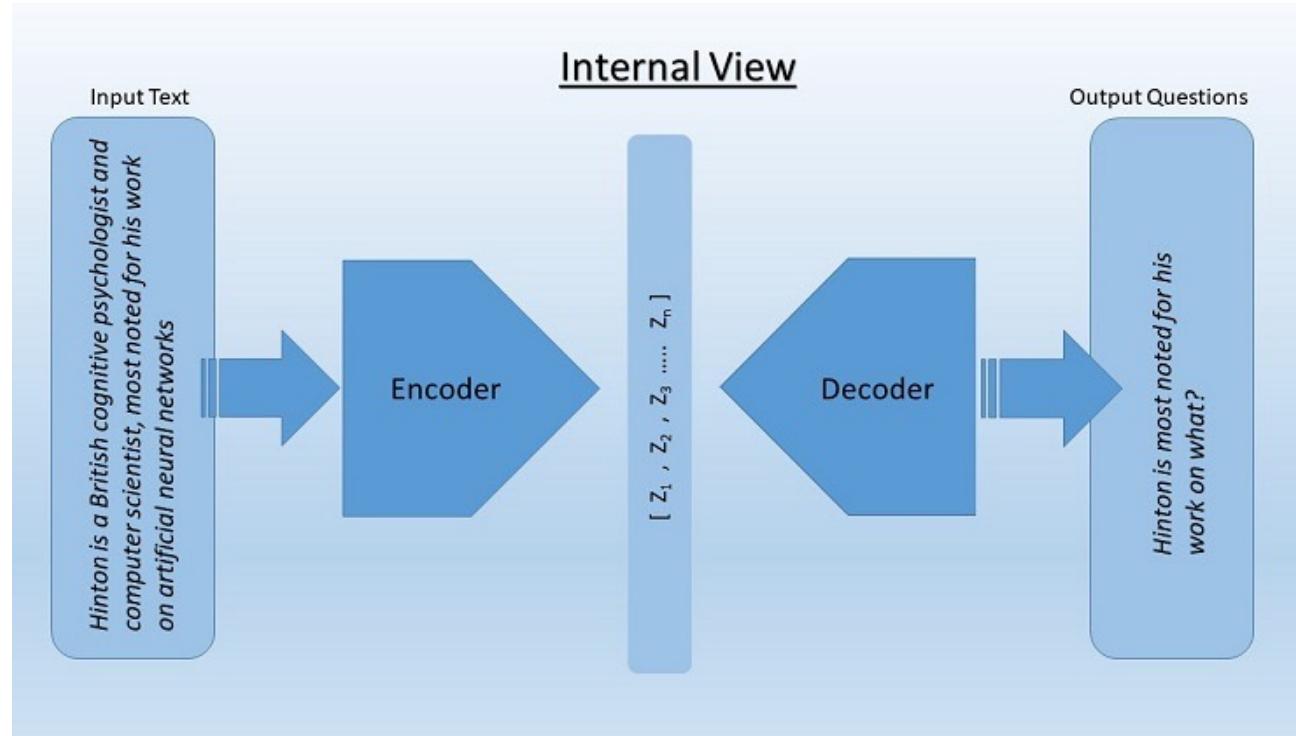
GNN enables:

- Multi-hop reasoning
- Encode heterogenous information



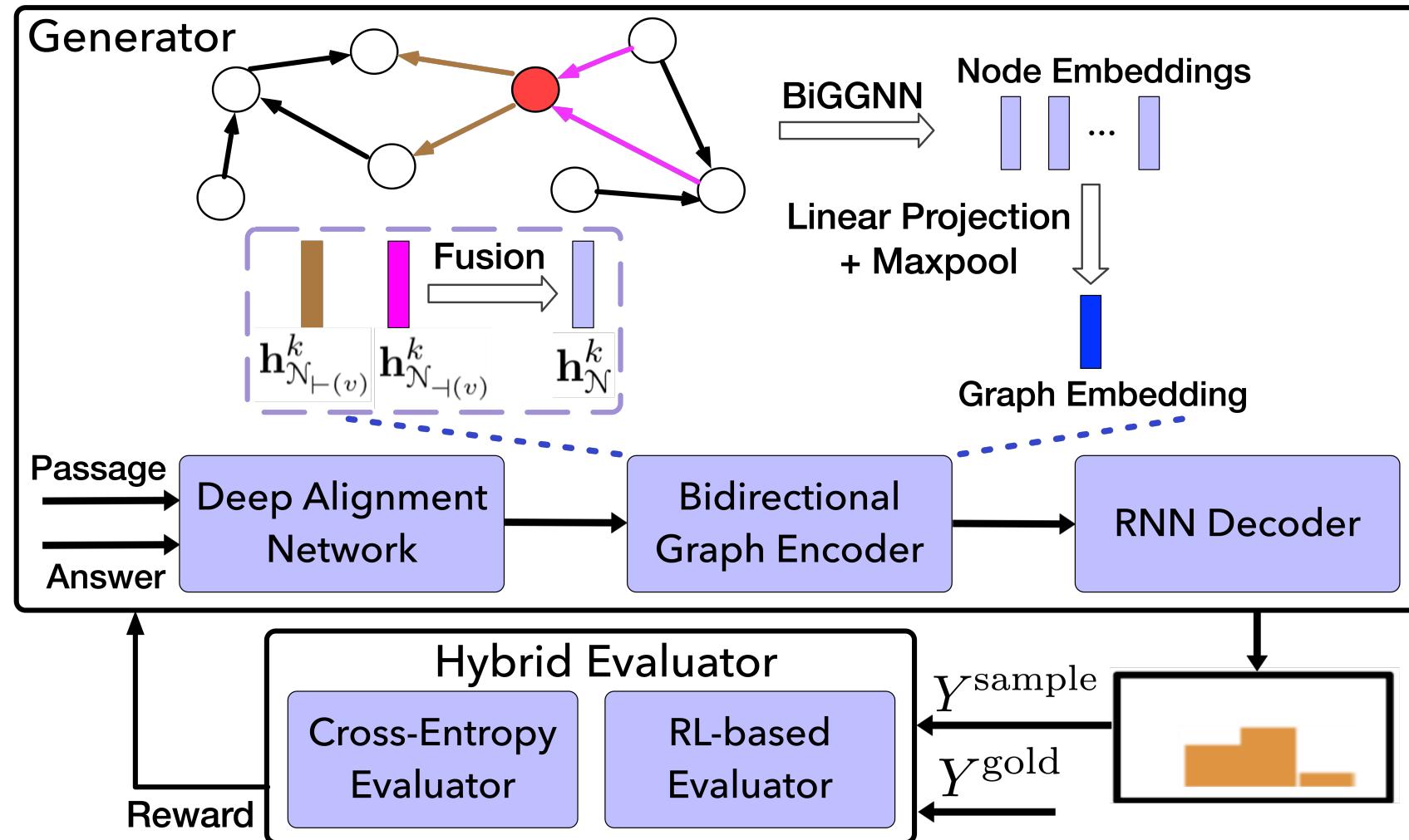
Natural Language Generation

Natural Question Generation



- **Input**
 - A text passage $X^p = \{x_1^p, x_2^p, \dots, x_N^p\}$
 - A target answer $X^a = \{x_1^a, x_2^a, \dots, x_L^a\}$
- **Output**
 - A natural language question
$$\hat{Y} = \{y_1, y_2, \dots, y_T\}$$
which maximizes the conditional likelihood
$$\hat{Y} = \arg \max_Y P(Y|X^p, X^a)$$

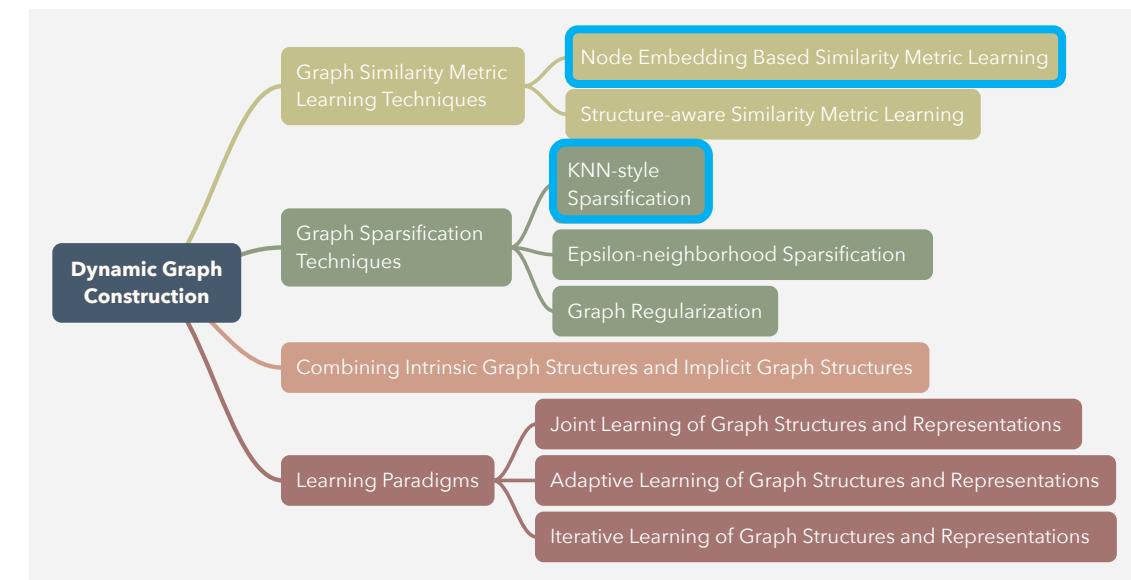
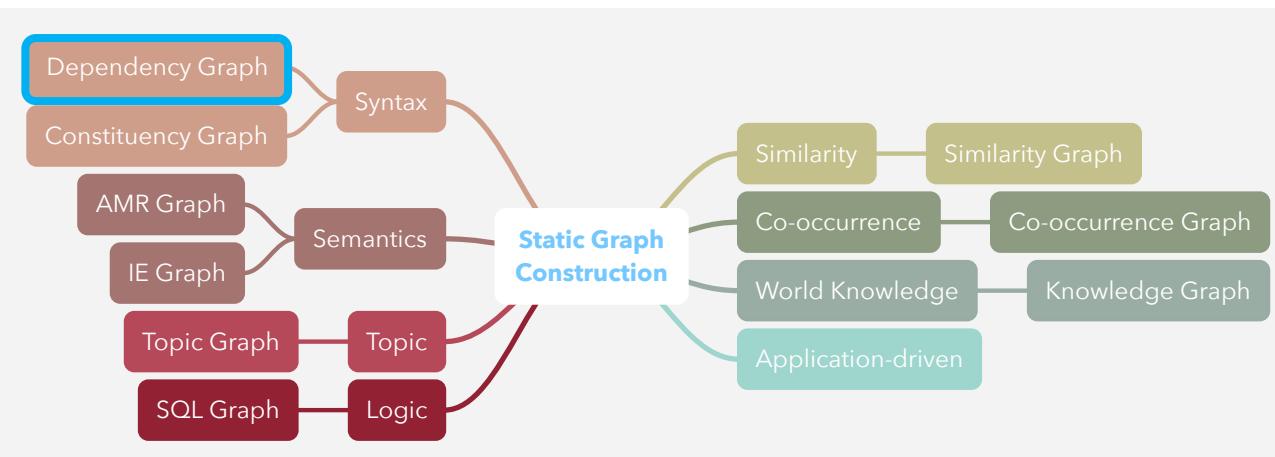
RL-based Graph2Seq for QG [Chen et al. ICLR'20]



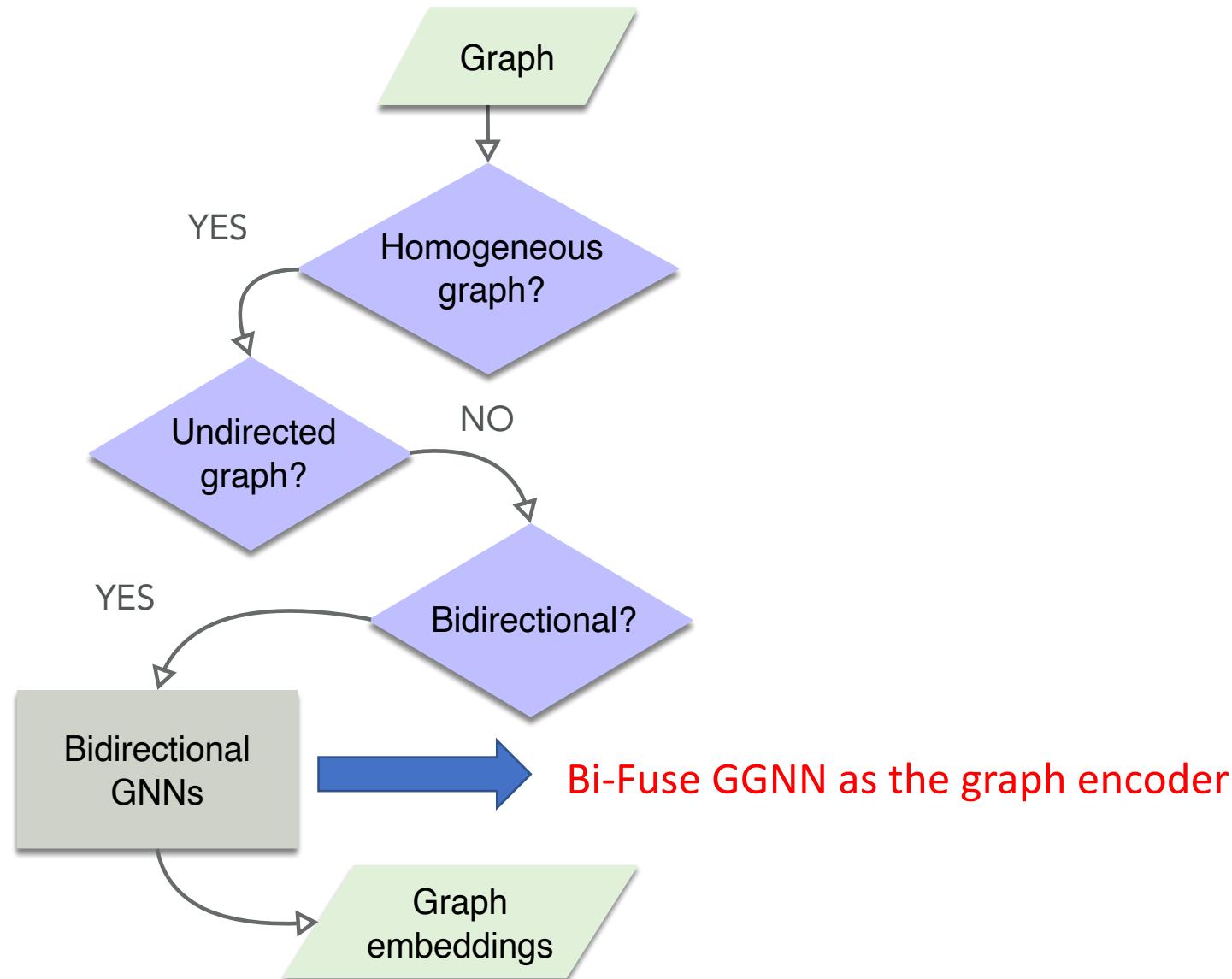
RL-based Graph2Seq for QG [Chen et al. ICLR'20]

Two graph construction strategies:

- 1) Syntax-based **static** passage graph construction
- 2) Semantics-aware **dynamic** passage graph construction



RL-based Graph2Seq for QG [Chen et al. ICLR'20]



RL-based Graph2Seq for QG [Chen et al. ICLR'20]

Methods	BLEU-4	Methods	BLEU-4
G2S _{dyn} +BERT+RL	18.06	G2S _{dyn} w/o feat	16.51
G2S _{sta} +BERT+RL	18.30	G2S _{sta} w/o feat	16.65
G2S _{sta} +BERT-fixed+RL	18.20	G2S _{dyn} w/o DAN	12.58
G2S _{dyn} +BERT	17.56	G2S _{sta} w/o DAN	12.62
G2S _{sta} +BERT	18.02	G2S _{sta} w/ DAN-word only	15.92
G2S _{sta} +BERT-fixed	17.86	G2S _{sta} w/ DAN-contextual only	16.07
G2S _{dyn} +RL	17.18	G2S _{sta} w/ GGNN-forward	16.53
G2S _{sta} +RL	17.49	G2S _{sta} w/ GGNN-backward	16.75
G2S _{dyn}	16.81	G2S _{sta} w/o BiGGNN, w/ Seq2Seq	16.14
G2S _{sta}	16.96	G2S _{sta} w/o BiGGNN, w/ GCN	14.47

Ablation study on the SQuAD split-2 test set.

Static graph construction
performs slightly better

Bidirectional GNN
performs better

Graph2Seq performs
better than Seq2Seq

Code Summarization



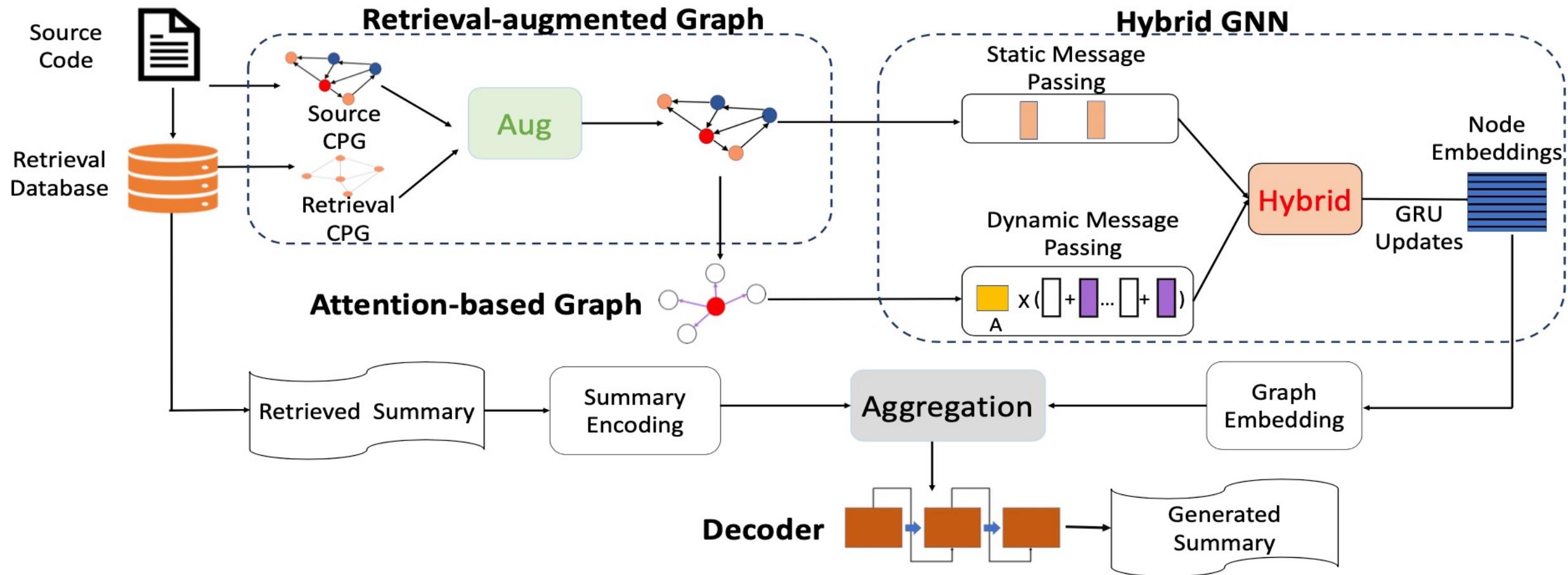
Source Code:

```
int pdc_tod_set(unsigned long sec, unsigned long usec) {
    int retval; unsigned long flags;
    spin_lock_irqsave(&pdc_lock, flags);
    retval = mem_pdc_call(PDC_TOD, PDC_TOD_WRITE, sec, usec);
    spin_unlock_irqrestore(&pdc_lock, flags);
    return retval;
}
```

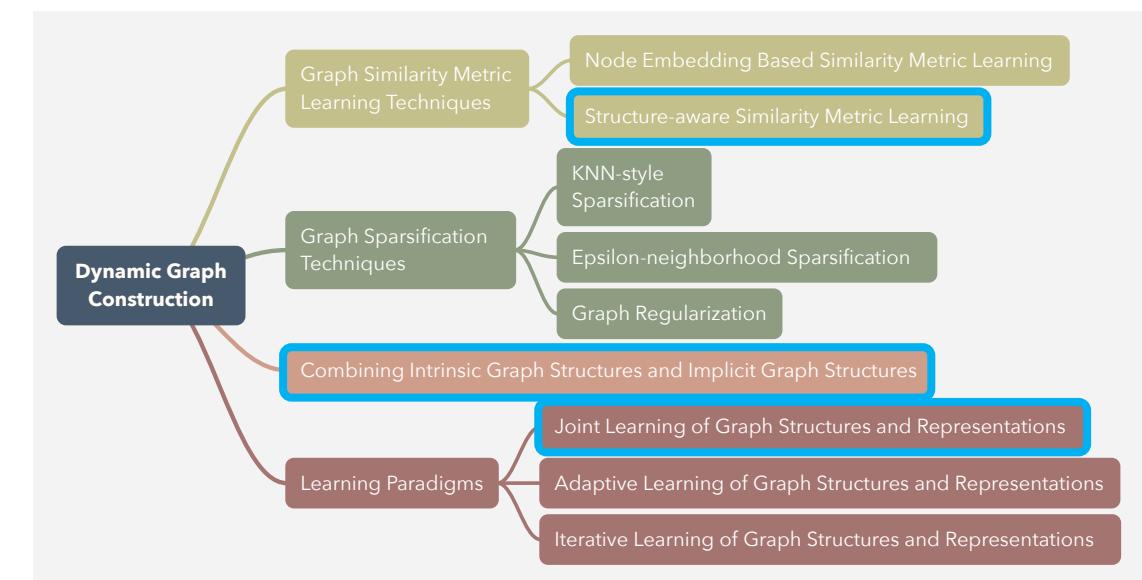
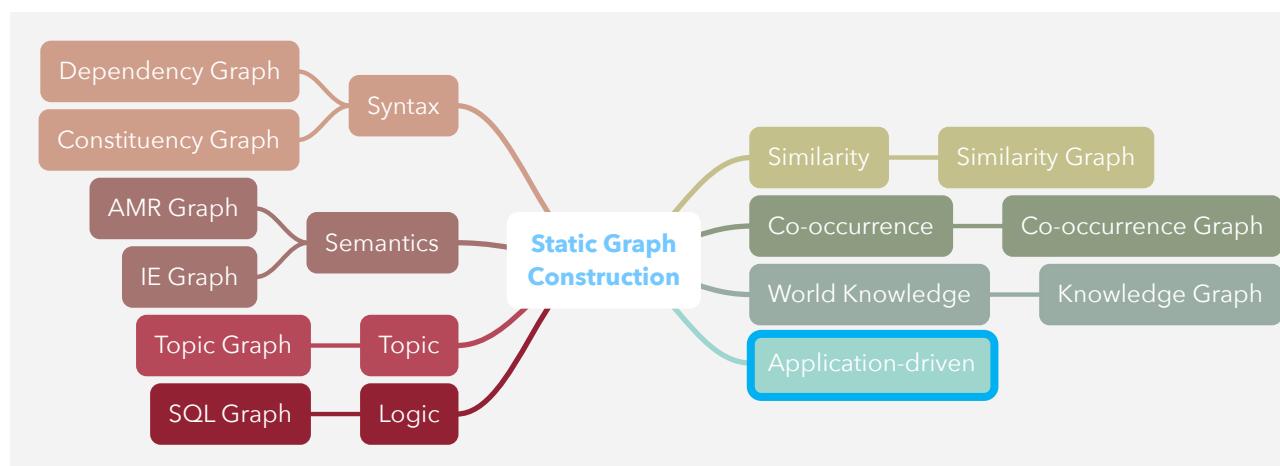
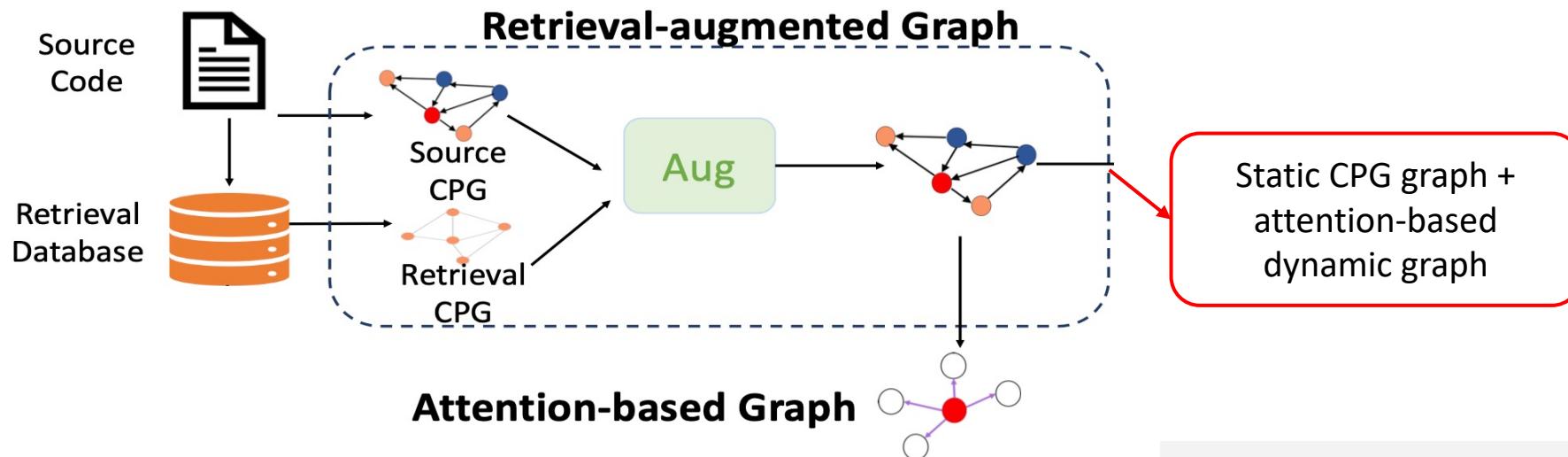
Ground-Truth: set the time of day clock

- Input
 - A piece of code
- Output
 - A natural language sentence

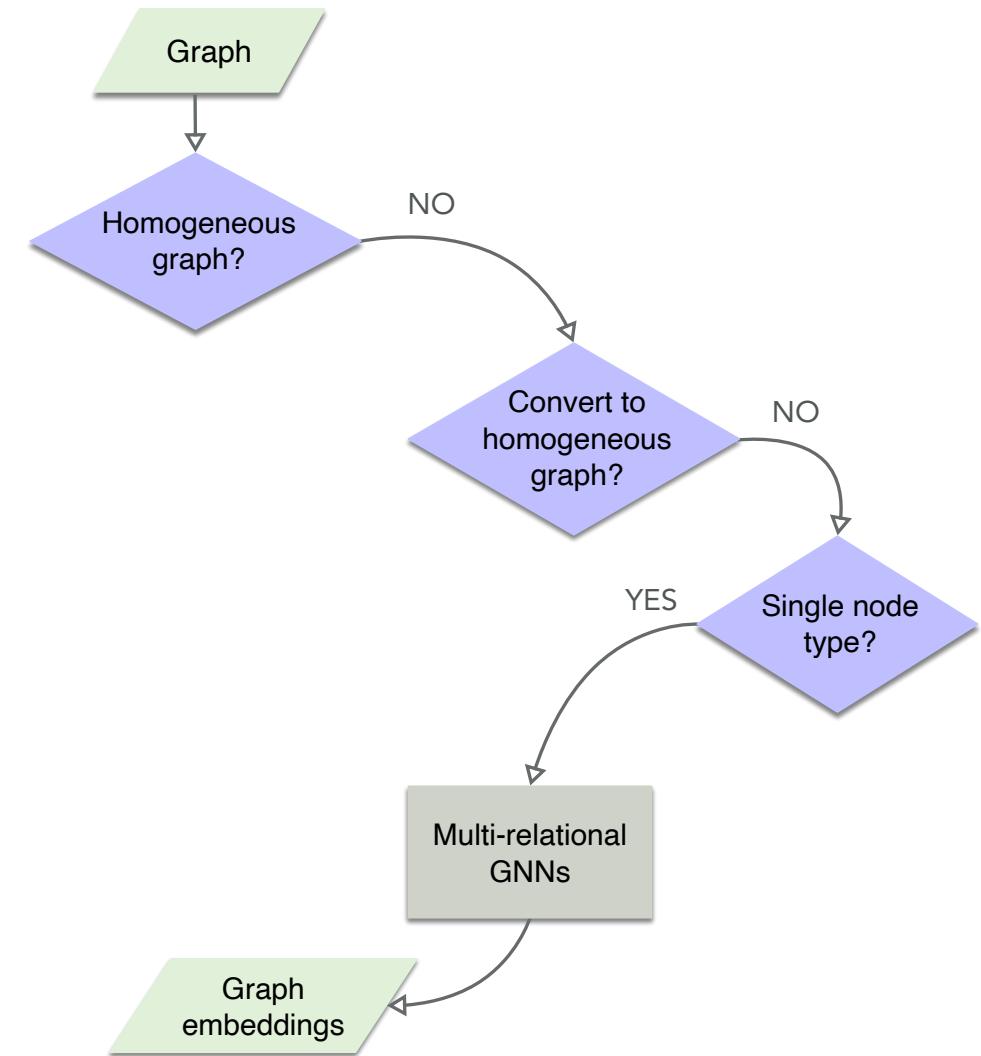
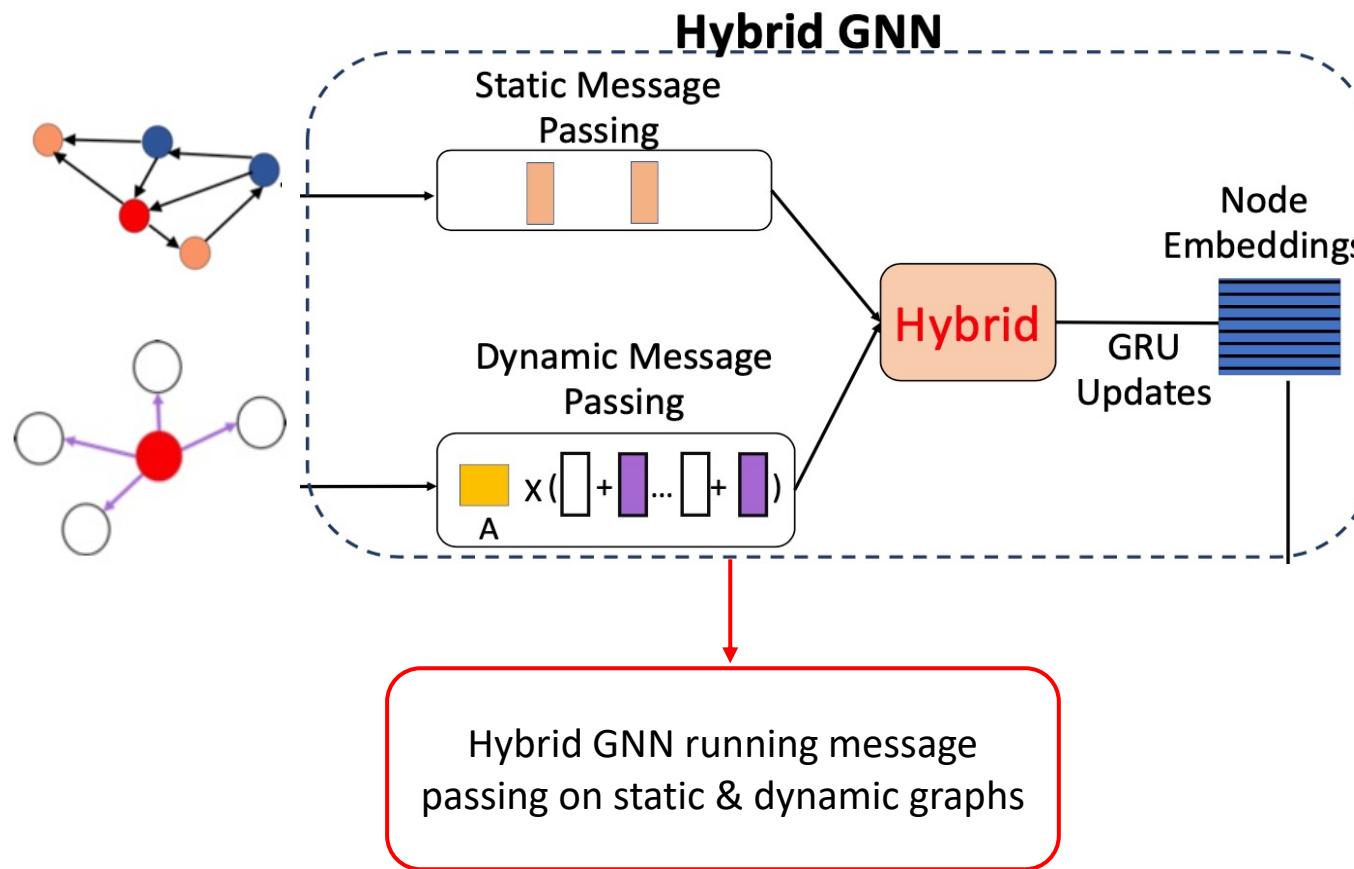
GNN for Code Summarization [Liu et al. ICLR'21]



GNN for Code Summarization [Liu et al. ICLR'21]



GNN for Code Summarization [Liu et al. ICLR'21]



GNN for Code Summarization [Liu et al. ICLR'21]

Methods	In-domain			Out-of-domain			Overall		
	BLEU-4	ROUGE-L	METEOR	BLEU-4	ROUGE-L	METEOR	BLEU-4	ROUGE-L	METEOR
TF-IDF	15.20	27.98	13.74	5.50	15.37	6.84	12.19	23.49	11.43
NNGen	15.97	28.14	13.82	5.74	16.33	7.18	12.76	23.93	11.58
CODE-NN	10.08	26.17	11.33	3.86	15.25	6.19	8.24	22.28	9.61
Hybrid-DRL	9.29	30.00	12.47	6.30	24.19	10.30	8.42	28.64	11.73
Transformer	12.91	28.04	13.83	5.75	18.62	9.89	10.69	24.65	12.02
Dual Model	11.49	29.20	13.24	5.25	21.31	9.14	9.61	26.40	11.87
Rencos	14.80	31.41	14.64	7.54	23.12	10.35	12.59	28.45	13.21
GCN2Seq	9.79	26.59	11.65	4.06	18.96	7.76	7.91	23.67	10.23
GAT2Seq	10.52	26.17	11.88	3.80	16.94	6.73	8.29	22.63	10.00
SeqGNN	10.51	29.84	13.14	4.94	20.80	9.50	8.87	26.34	11.93
<i>HGNN w/o augment & static</i>	11.75	29.59	13.86	5.57	22.14	9.41	9.98	26.94	12.05
<i>HGNN w/o augment & dynamic</i>	11.85	29.51	13.54	5.45	21.89	9.59	9.93	26.80	12.21
<i>HGNN w/o augment</i>	12.33	29.99	13.78	5.45	22.07	9.46	10.26	27.17	12.32
<i>HGNN w/o static</i>	15.93	33.67	15.67	7.72	24.69	10.63	13.44	30.47	13.98
<i>HGNN w/o dynamic</i>	15.77	33.84	15.67	7.64	24.72	10.73	13.31	30.59	14.01
HGNN	16.72	34.29	16.25	7.85	24.74	11.05	14.01	30.89	14.50

Automatic evaluation results (in %) on the CCSD test set.

Combining static + dynamic graphs performs better

Text Clustering and Matching

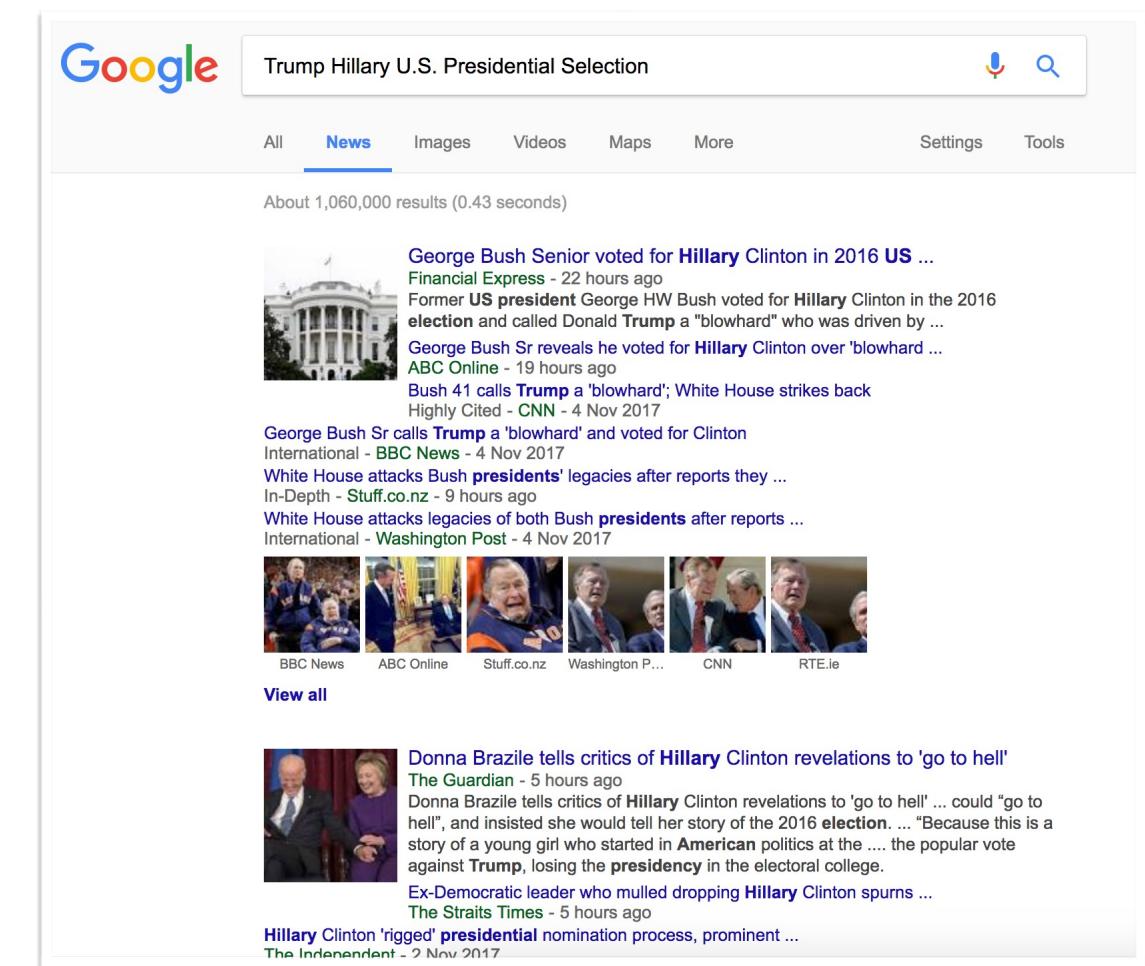
News Reading: Search Engines & Feeds Streams

Disadvantages of existing systems

- Messed document lists
- Extremely fine-grained (articles)
- Redundant useless information
- Unstructured information



Information Explosion



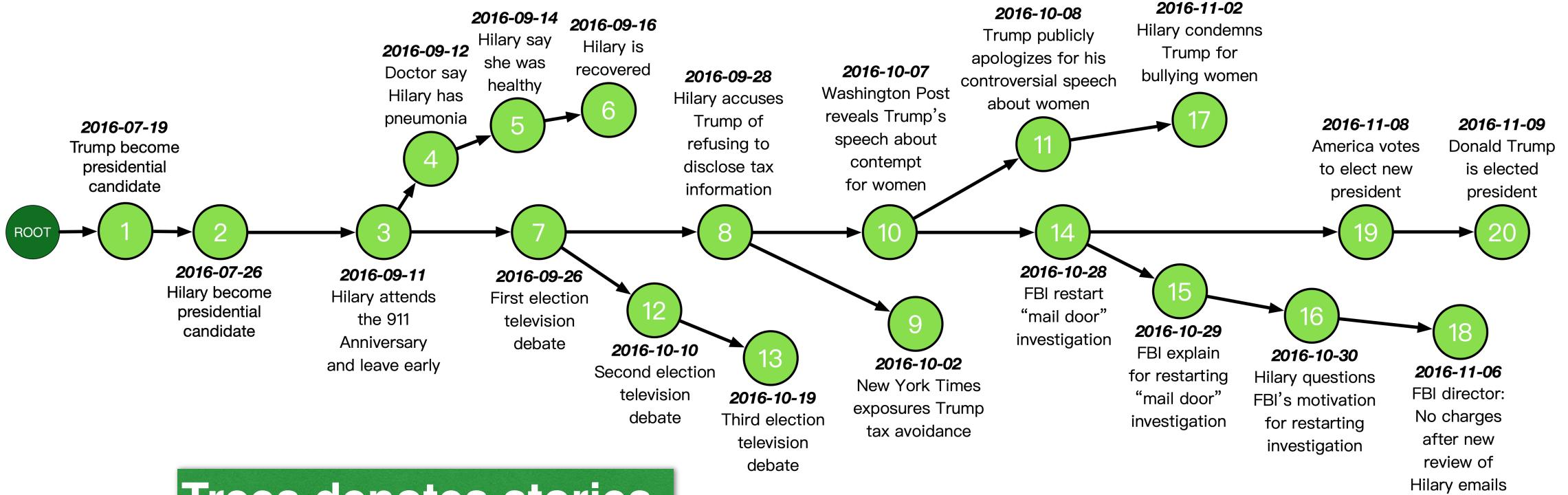
Google News search results for "Trump Hillary U.S. Presidential Selection". The results show approximately 1,060,000 results from 0.43 seconds ago. The top stories include:

- George Bush Senior voted for Hillary Clinton in 2016 US ... - Financial Express - 22 hours ago
- Former US president George HW Bush voted for Hillary Clinton in the 2016 election and called Donald Trump a "blowhard" who was driven by ... - George Bush Sr reveals he voted for Hillary Clinton over 'blowhard' ... - ABC Online - 19 hours ago
- Bush 41 calls Trump a 'blowhard'; White House strikes back - Bush 41 calls Trump a 'blowhard' and voted for Clinton - Highly Cited - CNN - 4 Nov 2017
- George Bush Sr calls Trump a 'blowhard' and voted for Clinton - International - BBC News - 4 Nov 2017
- White House attacks Bush presidents' legacies after reports they ... - In-Depth - Stuff.co.nz - 9 hours ago
- White House attacks legacies of both Bush presidents after reports ... - International - Washington Post - 4 Nov 2017

Below the stories are thumbnail images of political figures and logos for BBC News, ABC Online, Stuff.co.nz, Washington Post, CNN, and RTE.ie.

Story Forest

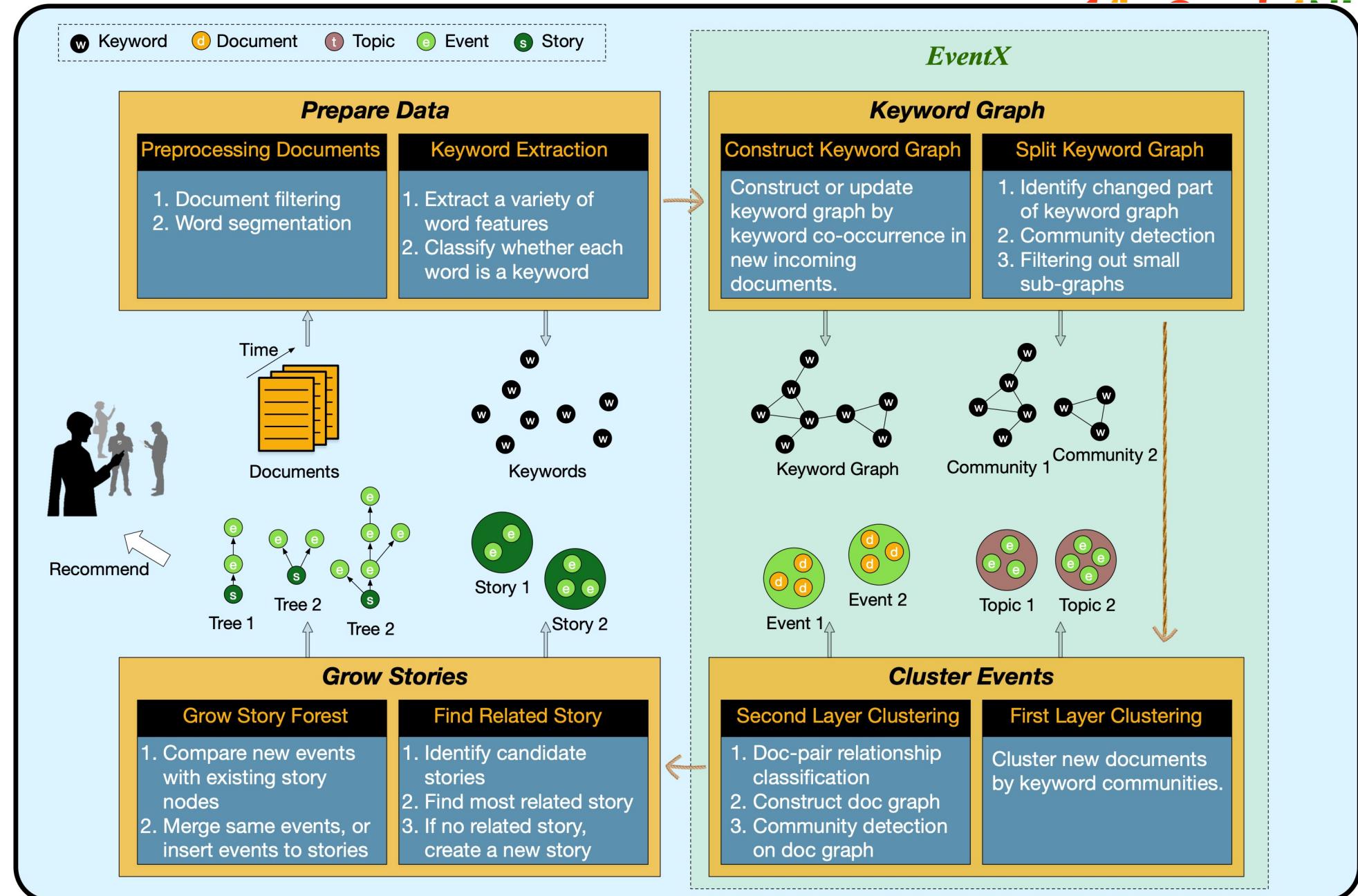
Detect events automatically from massive news articles



Trees denotes stories,
nodes denotes events

Edges in the tree denotes
events evolving relationship

Story Forest



Applied to QQ Browser Hot Topic List

热搜榜 猜你想搜 **Beijing Winter Olympics close**

1 北京冬奥闭幕 2026米兰见 热	571.8w
2 谷爱凌苏翊鸣闭幕式同框 心	519.8w
3 普京直言俄罗斯被耍了	350.6w
4 高亭宇扛着徐梦桃进场 热	298.1w
5 康辉说了3遍请运动员落座... 热	274.0w
6 中式浪漫 折柳送别运动员	261.2w
7 中国队冬奥会9金4银2铜收官	220.8w
8 鸟巢上空现"天下一家"焰火	198.9w
9 2008年五环和2022年火炬同框	187.1w
10 郑州楼市危机?代市长回应	181.2w
11 普京亲自坐镇指挥核演习 热	179.6w
12 冰墩墩下班啦 晕	171.8w
13 特鲁索娃cos神奇女侠再现蟹步	136.4w
14 乌克兰内政部长车队在乌东遇袭	131.5w
15 广州下雪?专家回应来了	130.0w
16 新浪微博裁员波及数百人	75.9w



**The story line that contains the event
“Beijing Winter Olympics close”**

2022 Beijing Winter Olympics

2022年北京冬奥会 热度359万

更新至2022-02-20 08:51

事件脉络 只看梗概 倒序

今天 08:51 NEW 圆满闭幕 北京冬奥会主火炬熄灭 2026米 兰见
今天 07:37 参加北京冬奥闭幕式运动员数破纪录
今天 06:52 北京冬奥会闭幕式节目单曝光
今天 00:43 张艺谋称冬奥闭幕式是个大party
今天 00:30 祝贺！中国队9金4银2铜15枚奖牌收官
昨天 21:30 高亭宇徐梦桃任冬奥闭幕式旗手
昨天 07:16 巴赫将奥林匹克杯授予中国人民 
昨天 04:04 张艺谋剧透闭幕式空灵且浪漫
昨天 01:26 冬奥会闭幕式将于2月20日晚亮相鸟巢

Social Event Detection

Raw Messages

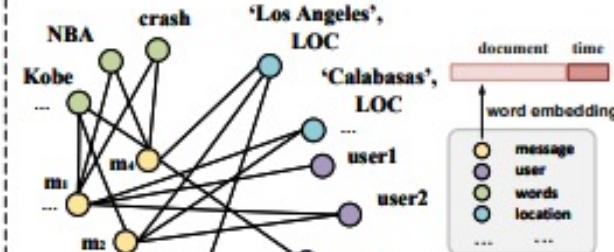
m₁
@user1 Jan 27th, 2020
NBA legend Kobe Bryant was one of five people killed in a helicopter crash in Calabasas, California, ...@user2

m₂
@user2 Jan 27th, 2020
Kobe Bryant avait 41 ans. Une femme, 4 filles, dont une de seulement 7 mois. Il avait pris sa retraite en 2018. ...

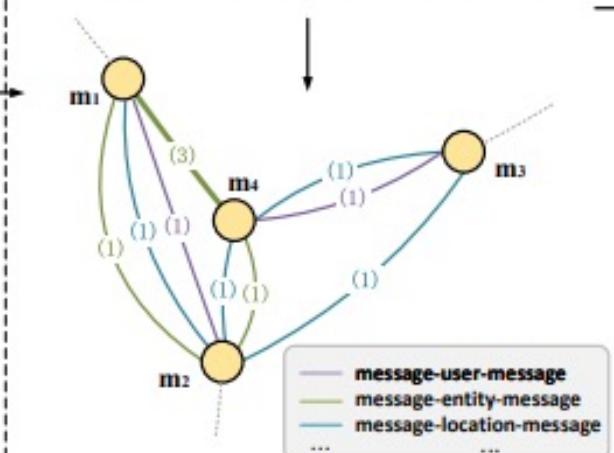
m₃
@user3 Jan 27th, 2020
The Los Angeles County Sheriff Department's Recruitment Unit is hosting its Inaugural Women's Symposium... ...

m₄
@user3 Jan 28th, 2020
BREAKING: Kobe Bryant, the Los Angeles Lakers superstar and future NBA Hall of Famer, died on Sunday in a helicopter crash. He was 41.

I. Preprocessing

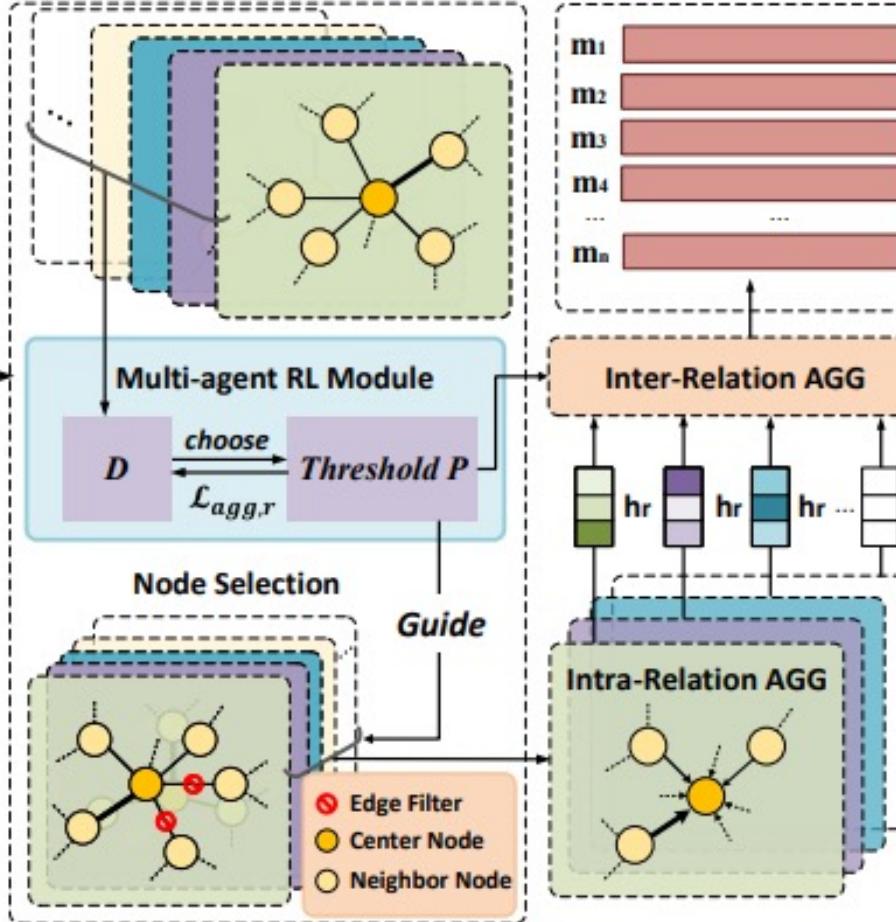


a) Heterogeneous Social Message Graph

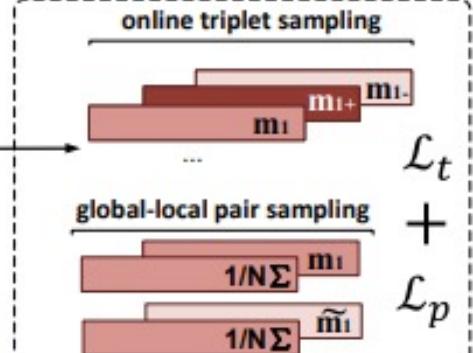


b) Weighted Multi-relational Message Graphs

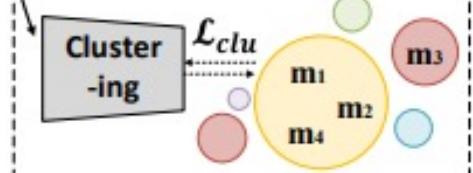
II. MarGNN



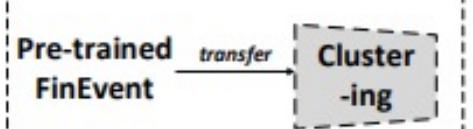
III. BasCL



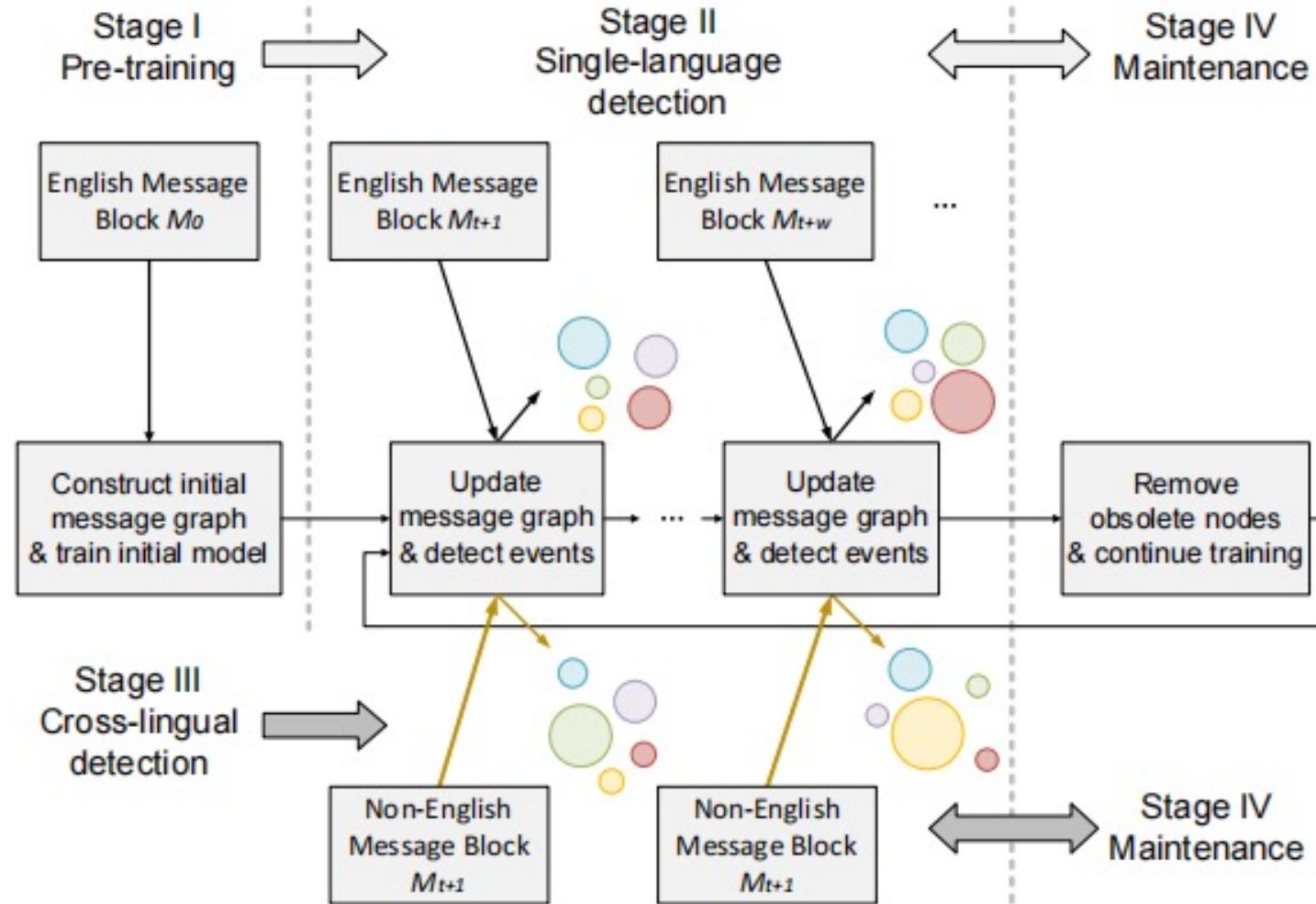
IV. DRL-DBSCAN



V. CrIme



Social Event Detection



Experiments

TABLE 2: Offline evaluation results on the Twitter dataset. The best results are marked in **bold**.

Metrics	Word2vec	LDA	WMD	BERT	BiLSTM	PP-GCN	EventX	KPGNN _t	KPGNN	FinEvent _k	FinEvent _d
NMI	.44±.00	.29±.00	.65±.00	.64±.00	.63±.00	.68±.02	.72±.00	.69±.01	.70±.01	.79±.01	.80±.01 ↑ .08
AMI	.13±.00	.04±.00	.50±.00	.44±.00	.41±.00	.50±.02	.19±.00	.51±.00	.52±.01	.62±.01	.69±.01 ↑ .19
ARI	.02±.00	.01±.00	.06±.00	.07±.00	.17±.00	.20±.01	.05±.00	.21±.01	.22±.01	.24±.01	.48±.01 ↑ .28

TABLE 3: Incremental evaluation NMIs. The best results are marked in **bold** and second-best in *italic*.

Blocks	Word2vec	LDA	WMD	BERT	BiLSTM	PP-GCN	EventX	KPGNN _t	KPGNN	FinEvent _k	FinEvent _d
<i>M</i> ₁	.19±.00	.11±.00	.32±.00	.36±.00	.24±.00	.23±.00	.36±.00	.38±.01	.39±.00	.38±.01	.84±.01 ↑ .49
<i>M</i> ₂	.50±.00	.27±.01	.71±.00	.78±.00	.50±.00	.57±.02	.68±.00	.78±.01	.79±.01	.81±.00	.84±.01 ↑ .06
<i>M</i> ₃	.39±.00	.28±.00	.67±.00	.75±.00	.39±.00	.55±.01	.63±.00	.77±.00	.76±.00	.83±.00	.89±.00 ↑ .14
<i>M</i> ₄	.34±.00	.25±.00	.50±.00	.60±.00	.40±.00	.46±.01	.63±.00	.68±.01	.67±.00	.71±.01	.71±.01 ↑ .08
<i>M</i> ₅	.41±.00	.26±.00	.61±.00	.72±.00	.41±.00	.48±.01	.59±.00	.73±.01	.73±.01	.76±.00	.83±.00 ↑ .11
<i>M</i> ₆	.53±.00	.32±.00	.61±.00	.78±.00	.50±.00	.57±.01	.70±.00	.81±.00	.82±.01	.84±.00	.83±.00 ↑ .06
<i>M</i> ₇	.25±.00	.18±.01	.46±.00	.54±.00	.33±.00	.37±.00	.51±.00	.54±.01	.55±.01	.56±.00	.73±.01 ↑ .02
<i>M</i> ₈	.46±.00	.37±.01	.67±.00	.79±.00	.49±.00	.55±.02	.71±.00	.79±.01	.80±.00	.87±.01	.87±.02 ↑ .08
<i>M</i> ₉	.35±.00	.34±.00	.55±.00	.70±.00	.43±.00	.51±.02	.67±.00	.74±.01	.74±.02	.78±.02	.79±.01 ↑ .09
<i>M</i> ₁₀	.51±.00	.44±.01	.61±.00	.74±.00	.50±.00	.55±.02	.68±.00	.79±.01	.80±.01	.81±.01	.82±.01 ↑ .08
<i>M</i> ₁₁	.37±.00	.33±.01	.50±.00	.68±.00	.49±.00	.50±.01	.65±.00	.73±.00	.74±.01	.76±.00	.75±.00 ↑ .08
<i>M</i> ₁₂	.30±.00	.22±.01	.60±.00	.59±.00	.39±.00	.45±.01	.61±.00	.69±.01	.68±.01	.76±.01	.67±.01 ↑ .15
<i>M</i> ₁₃	.37±.00	.27±.00	.54±.00	.63±.00	.46±.00	.47±.01	.58±.00	.68±.01	.69±.01	.67±.00	.79±.00 ↑ .18
<i>M</i> ₁₄	.36±.00	.21±.00	.66±.00	.64±.00	.44±.00	.44±.01	.57±.00	.68±.01	.69±.00	.74±.00	.82±.00 ↑ .16
<i>M</i> ₁₅	.27±.00	.21±.00	.51±.00	.54±.00	.40±.00	.39±.01	.49±.00	.57±.01	.58±.00	.64±.00	.69±.01 ↑ .15
<i>M</i> ₁₆	.49±.00	.35±.01	.60±.00	.75±.00	.53±.00	.55±.01	.62±.00	.78±.01	.79±.01	.80±.00	.90±.01 ↑ .15
<i>M</i> ₁₇	.33±.00	.19±.00	.55±.00	.63±.00	.45±.00	.48±.00	.58±.00	.69±.01	.70±.01	.73±.00	.83±.00 ↑ .20
<i>M</i> ₁₈	.29±.00	.18±.00	.63±.00	.57±.00	.44±.00	.47±.01	.59±.00	.68±.01	.68±.02	.72±.01	.74±.01 ↑ .11
<i>M</i> ₁₉	.37±.00	.29±.01	.54±.00	.66±.00	.44±.00	.51±.02	.60±.00	.73±.00	.73±.01	.76±.02	.66±.01 ↑ .10
<i>M</i> ₂₀	.38±.00	.35±.00	.58±.00	.68±.00	.48±.00	.51±.01	.67±.00	.73±.00	.72±.02	.73±.00	.80±.00 ↑ .12
<i>M</i> ₂₁	.31±.00	.19±.00	.58±.00	.59±.00	.41±.00	.41±.02	.53±.00	.59±.01	.60±.00	.65±.01	.74±.01 ↑ .15

Long Document Matching

Identify the relationship
between documents



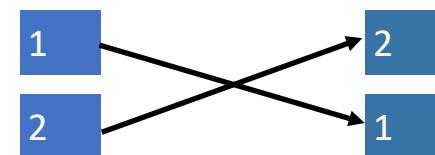
Divide-and-Conquer

Our strategies

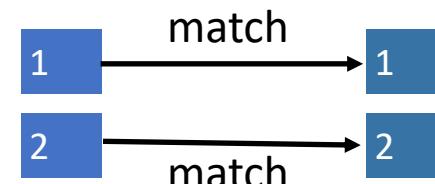
◆ Divide



◆ Align



◆ Distributed match



Limitations

◆ Hard to encode

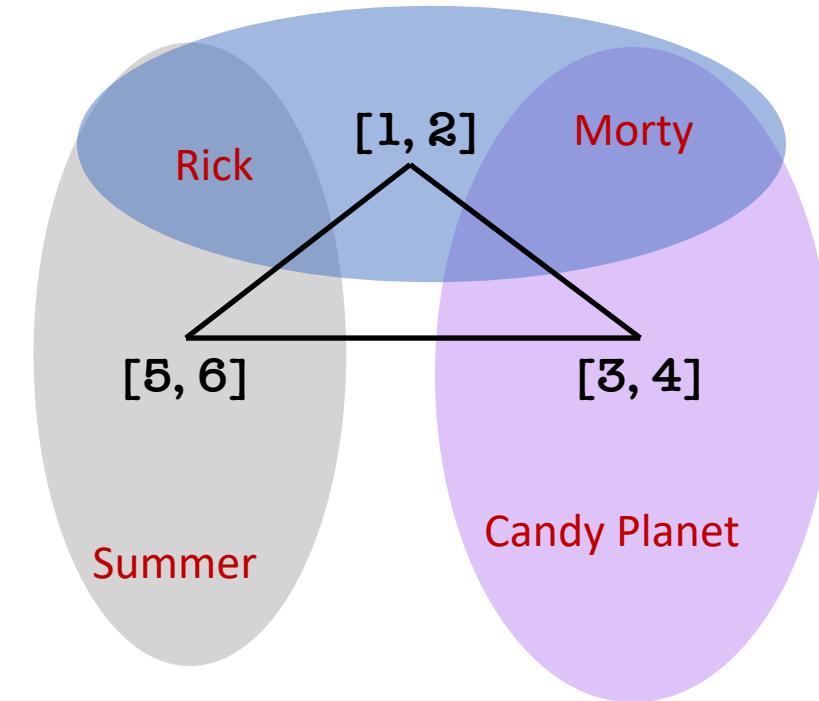
◆ Flexible order

◆ Time complexity



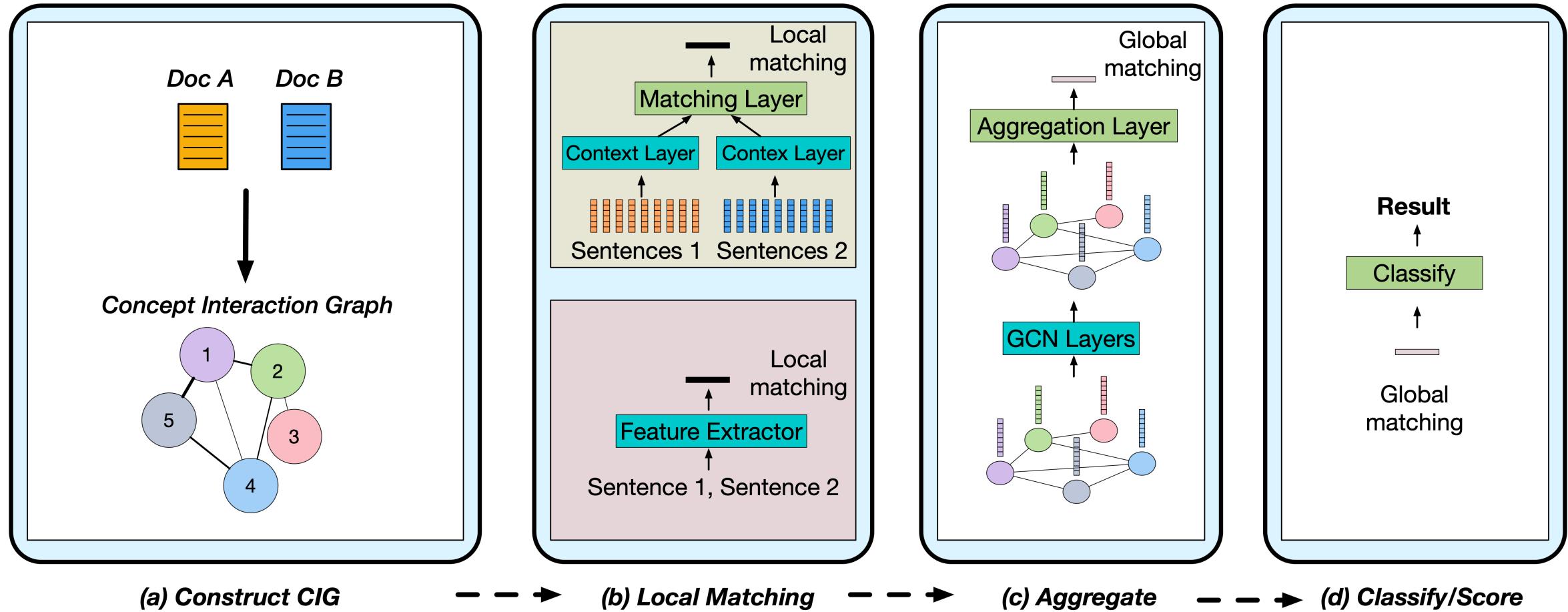
Concept Interaction Graph

- [1] **Rick** asks **Morty** to travel with him in the universe.
- [2] **Morty** doesn't want to go as **Rick** always brings him dangerous experiences.
- [3] However, the destination of this journey is the **Candy Planet**, which is a fascinating place that attracts **Morty**.
- [4] The planet is full of delicious candies.
- [5] **Summer** wishes to travel with **Rick**.
- [6] However, **Rick** doesn't like to travel with **Summer**.



- STEP 1:** Extract keywords
- STEP 2:** Group keywords
- STEP 3:** Assign sentences
- STEP 4:** Construct edges

Graph Decomposition for Document Matching



Experiments

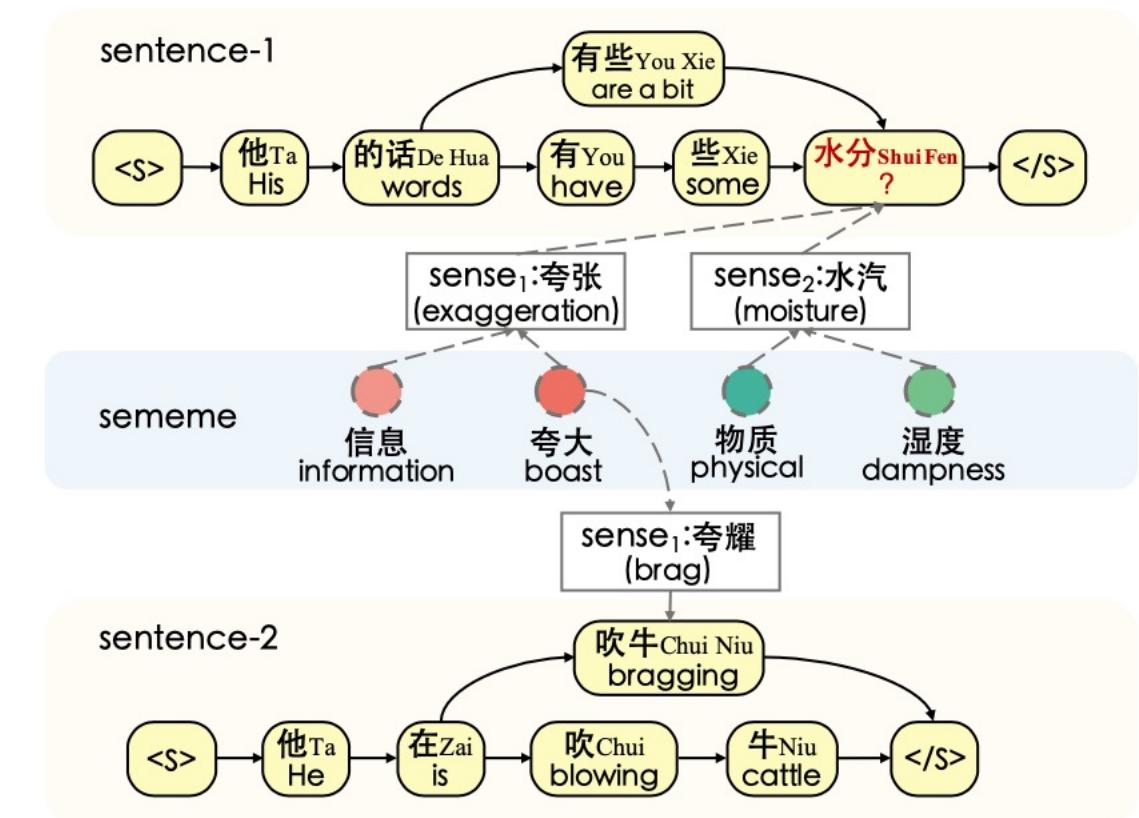
Baselines	CNSE		CNSS		Our models	CNSE		CNSS	
	Acc	F1	Acc	F1		Acc	F1	Acc	F1
I. ARC-I	53.84	48.68	50.10	66.58	XI. CIG-Siam	74.47	73.03	75.32	78.58
II. ARC-II	54.37	36.77	52.00	53.83	XII. CIG-Siam-GCN	74.58	73.69	78.91	80.72
III. DUET	55.63	51.94	52.33	60.67	XIII. CIG _{cd} -Siam-GCN	73.25	73.10	76.23	76.94
IV. DSSM	58.08	64.68	61.09	70.58	XIV. CIG-Sim	72.58	71.91	75.16	77.27
V. C-DSSM	60.17	48.57	52.96	56.75	XV. CIG-Sim-GCN	83.35	80.96	87.12	87.57
VI. MatchPyramid	66.36	54.01	62.52	64.56	XVI. CIG _{cd} -Sim-GCN	81.33	78.88	86.67	87.00
VII. BM25	69.63	66.60	67.77	70.40	XVII. CIG-Sim&Siam-GCN	84.64	82.75	89.77	90.07
VIII. LDA	63.81	62.44	62.98	69.11	XVIII. CIG-Sim&Siam-GCN-Sim ^g	84.21	82.46	90.03	90.29
IX. SimNet	71.05	69.26	70.78	74.50	XIX. CIG-Sim&Siam-GCN-BERT ^g	84.68	82.60	89.56	89.97
X. BERT fine-tuning	81.30	79.20	86.64	87.08	XX. CIG-Sim&Siam-GCN-Sim ^g &BERT ^g	84.61	82.59	89.47	89.71

♦ **Graph Representation:** greatly improves performance. (IX vs. XI)
(+4% Acc, F1)

♦ **Graph Convolution:** greatly improves performance. (XIV vs. XV)
(+10% Acc, F1)

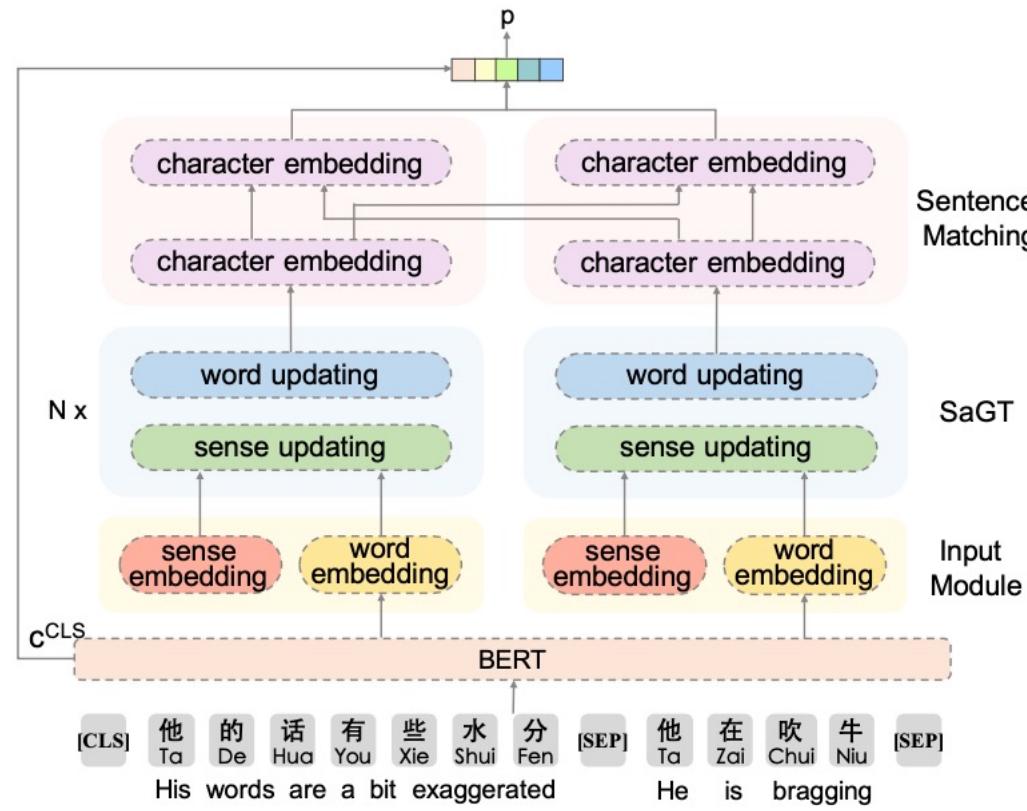
Short Text Matching

- Some Chinese words are **polysemous**, and semantic information is not fully utilized.
- Some models suffer potential issues caused by **word segmentation**.

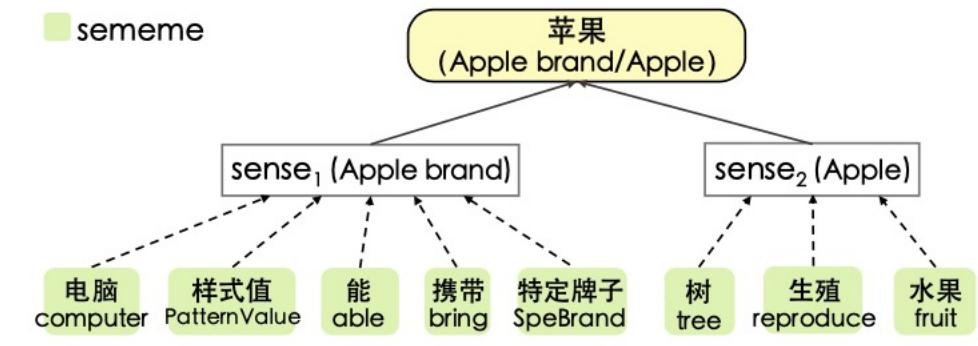


An example of word segmentation and the potential word ambiguity.

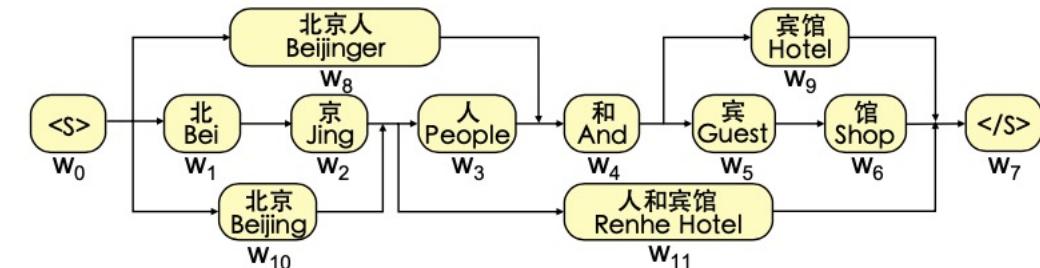
Short Text Matching



The framework of Linguistic knowledge Enhanced graph Transformer (LET) model



An example of the HowNet structure



An example of lattice graph.

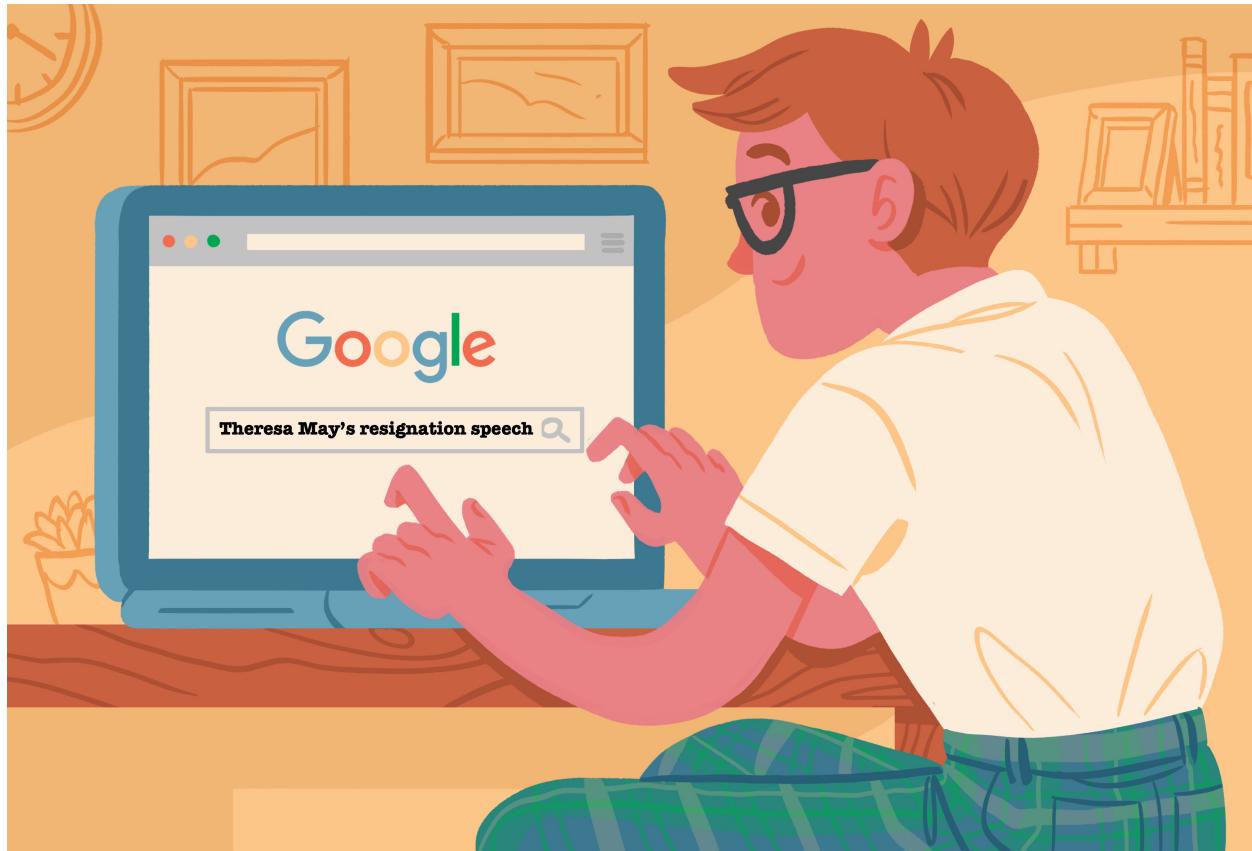
Experiments

Models	Pre-Training	Interaction	BQ		LCQMC	
			ACC.	F1	ACC.	F1
Text-CNN(He et al. 2016)	×	×	68.52	69.17	72.80	75.70
BiLSTM(Mueller and Thyagarajan 2016)	×	×	73.51	72.68	76.10	78.90
Lattice-CNN (Lai et al. 2019)	×	×	78.20	78.30	82.14	82.41
BiMPM (Wang, Hamza, and Florian 2017)	×	✓	81.85	81.73	83.30	84.90
ESIM (Chen et al. 2017)	×	✓	81.93	81.87	82.58	84.49
LET (Ours)	×	✓	83.22	83.03	84.81	86.08
BERT-wwm (Cui et al. 2019)	✓	✓	84.89	84.29	86.80	87.78
BERT-wwm-ext (Cui et al. 2019)	✓	✓	84.71	83.94	86.68	87.71
ERNIE (Sun et al. 2019)	✓	✓	84.67	84.20	87.04	88.06
BERT(Devlin et al. 2019)	✓	✓	84.50	84.00	85.73	86.86
LET-BERT (Ours)	✓	✓	85.30	84.98	88.38	88.85

Table 1: Performance of various models on LCQMC and BQ test datasets. The results are average scores using 5 different seeds. All the improvements over baselines are statistically significant ($p < 0.05$).

Text Mining and Classification

What are users interested in?

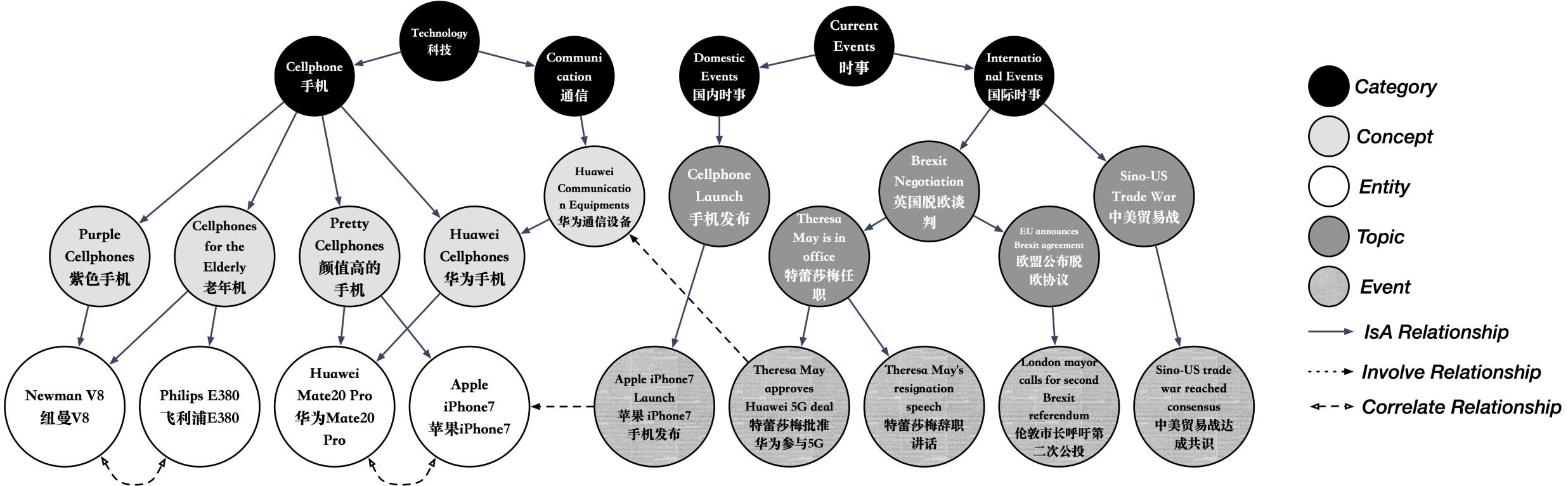


Query: “Theresa May’s resignation speech”

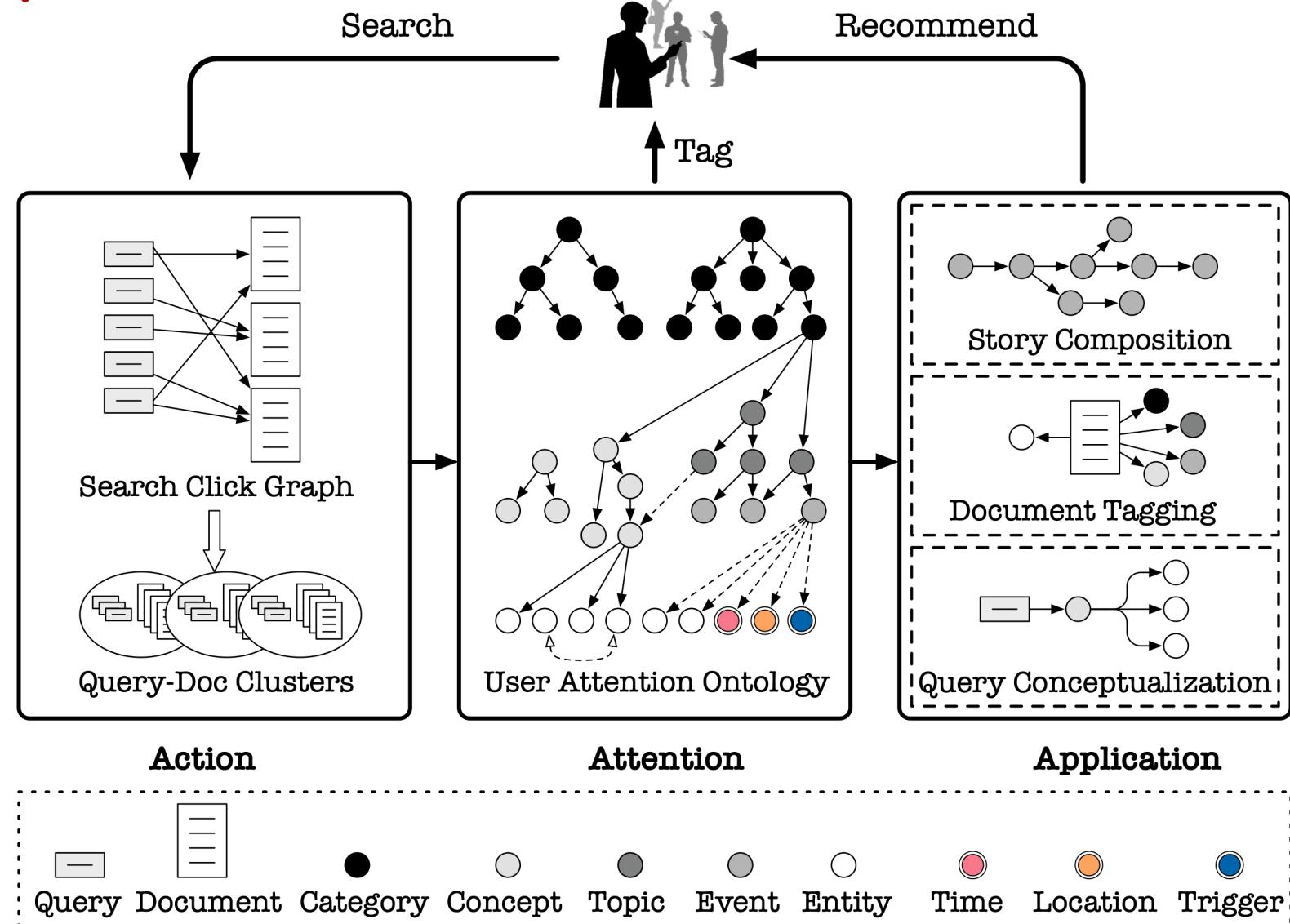
Infer users’ interests

Attention Ontology

Create a web-scale ontology to represent user interests and document topics.



GIANT System



Heterogeneous Phrase Mining

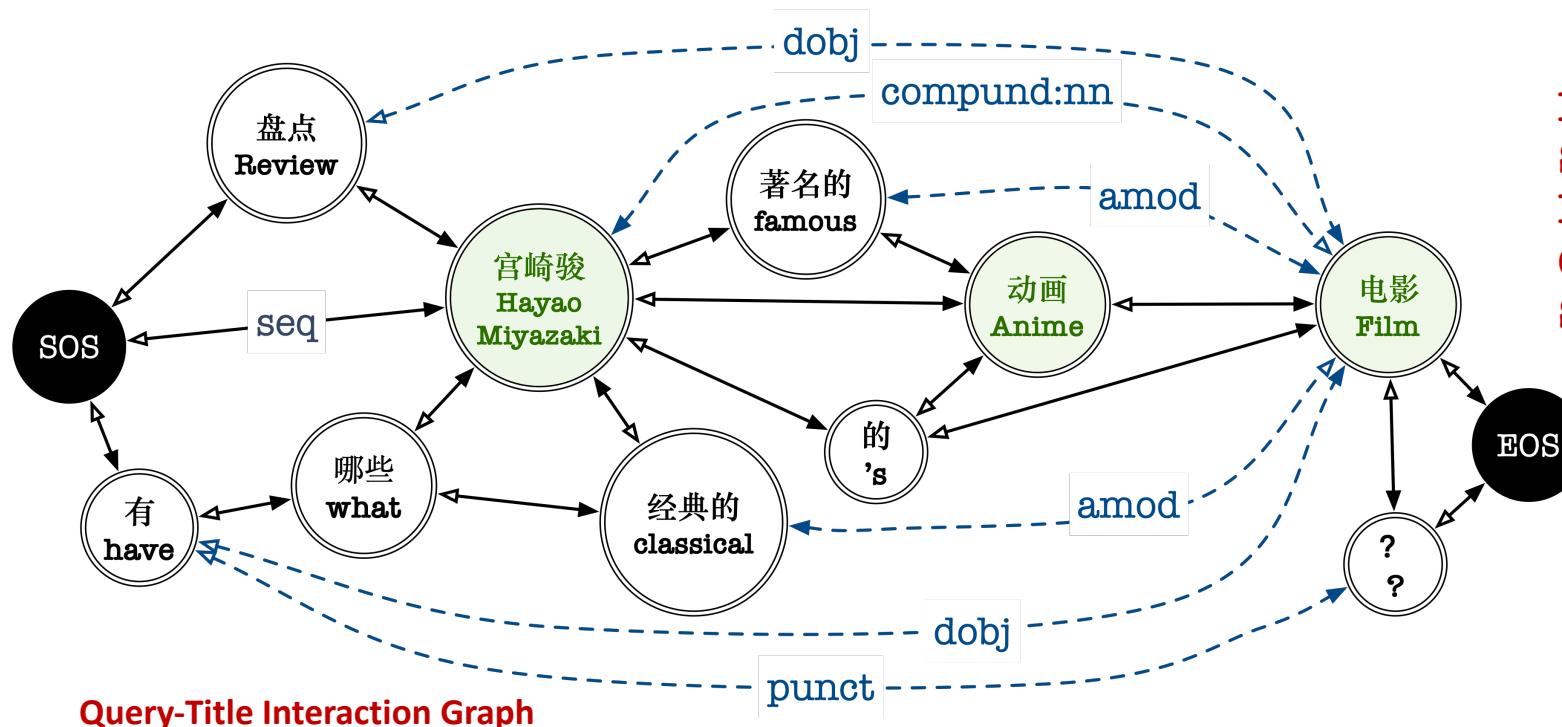
Query: What are the Hayao Miyazaki's animated film (有哪些 宫崎骏 的 动画 电影)

Titles: Review Hayao Miyazaki's animated film (盘点 宫崎骏 动画 电影)

The famous animated films of Hayao Miyazaki (宫崎骏 著名的 动画 电影)

What are the classic Miyazaki's movies? (有哪些 经典的 宫崎骏 的 电影?)

Concept: Hayao Miyazaki animated film (宫崎骏 动画 电影)



Patterns: graph structure

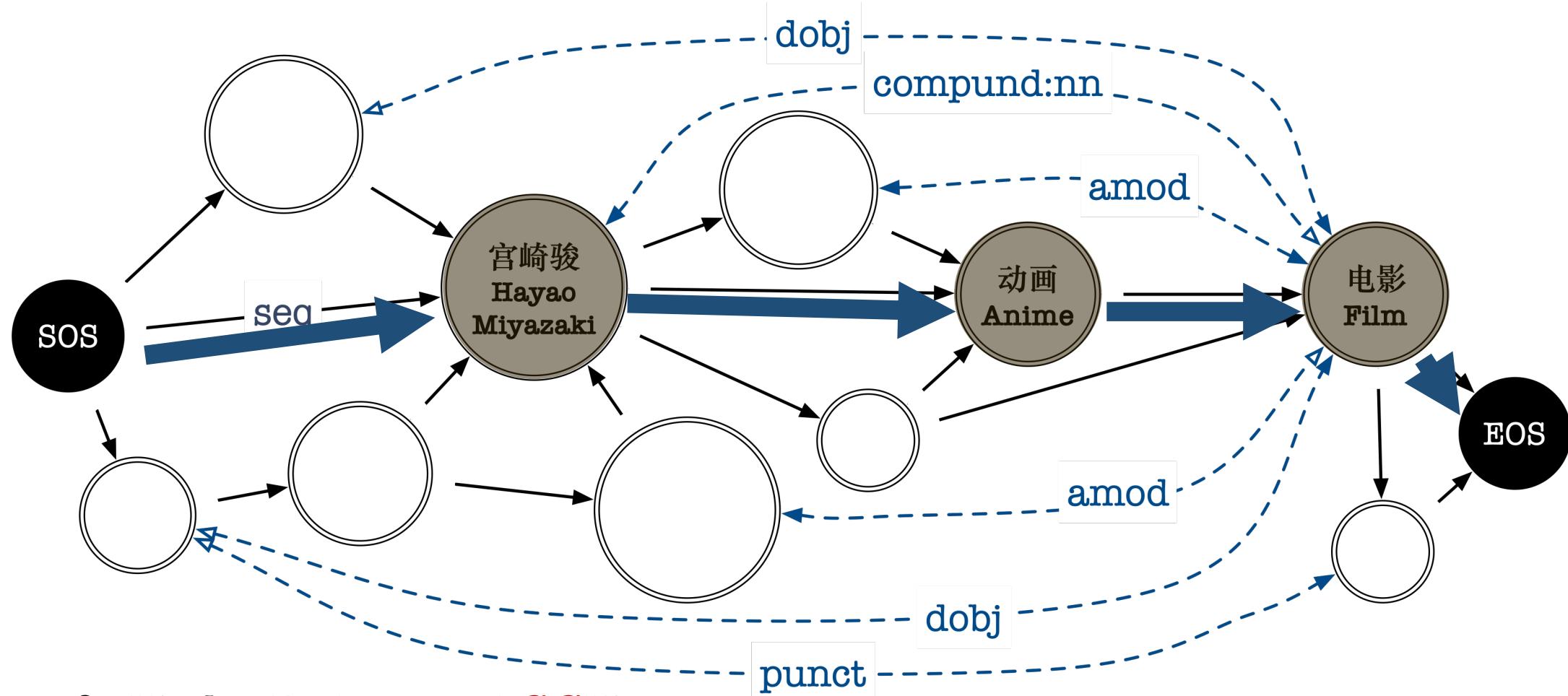
Show up multiple times: node feature

NER/Part-of-Speech tags: node feature

Continuous chunk: seq edge

Syntactic dependency: syntactic edge

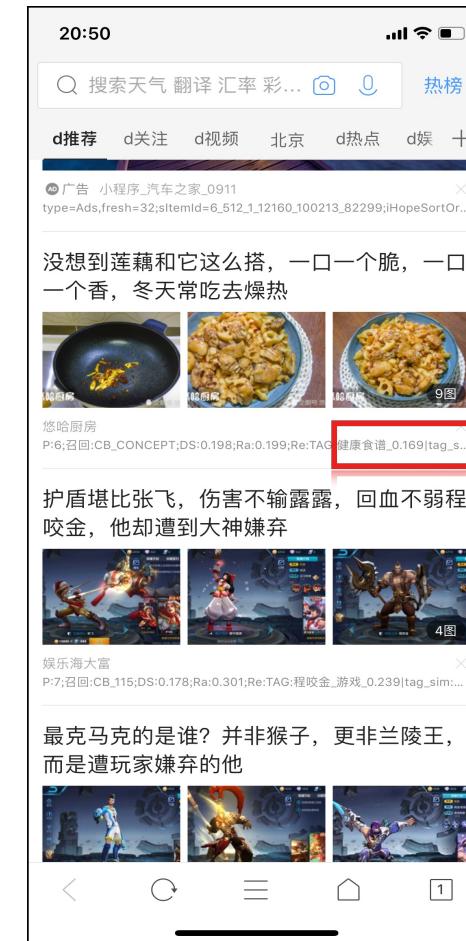
Heterogeneous Phrase Mining



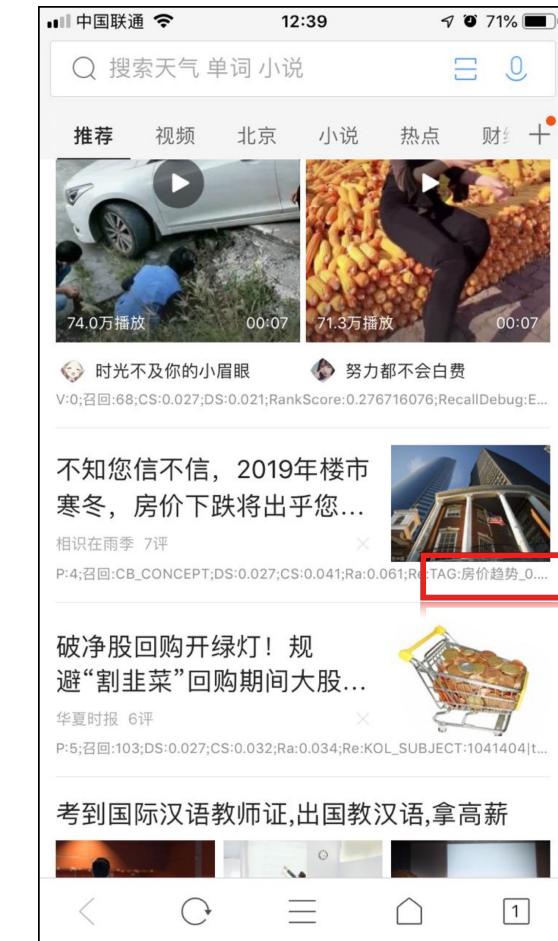
Classify Node: **Relational GCN**

Sort Node: **Asymmetric Traveling Salesman Problem**

Applied to feeds news recommendation



Healthy
recipes

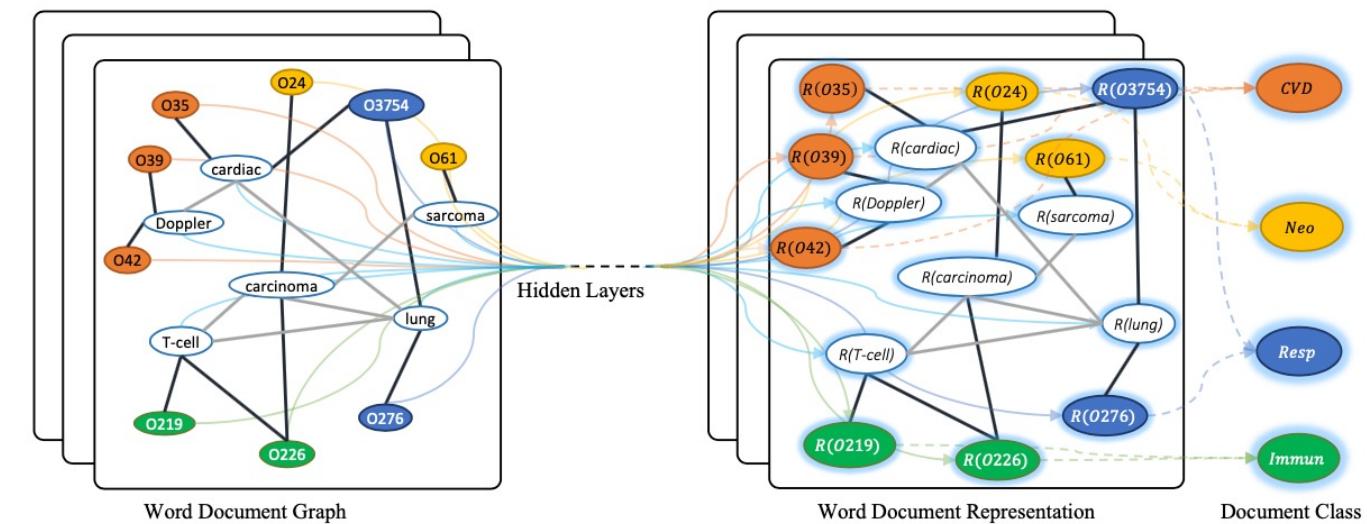


House
Price
trends



Transductive Text Classification with TextGCN

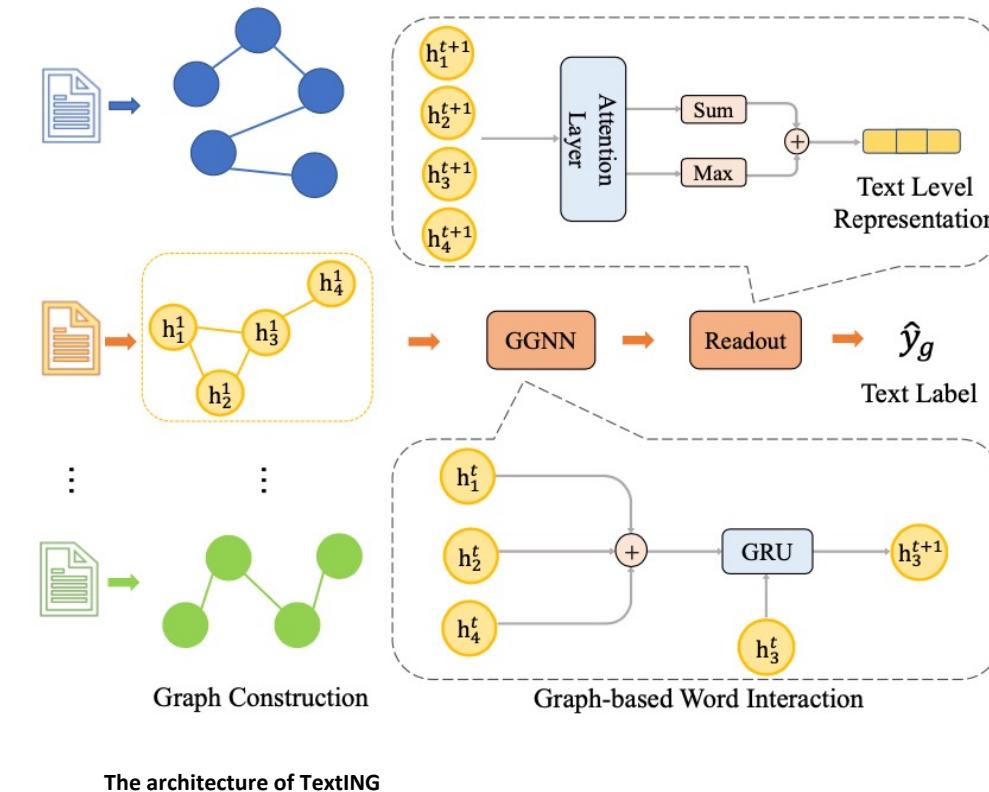
- **Nodes**: words and documents
- **Edges**: co-occurrence (word-word) and TFIDF (word-doc)
- Model the graph with a **Graph Convolutional Network (GCN)** (Kipf and Welling 2017)



$$A_{ij} = \begin{cases} \text{PMI}(i, j) & i, j \text{ are words, } \text{PMI}(i, j) > 0 \\ \text{TF-IDF}_{ij} & i \text{ is document, } j \text{ is word} \\ 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

Inductive Text Classification with TextING

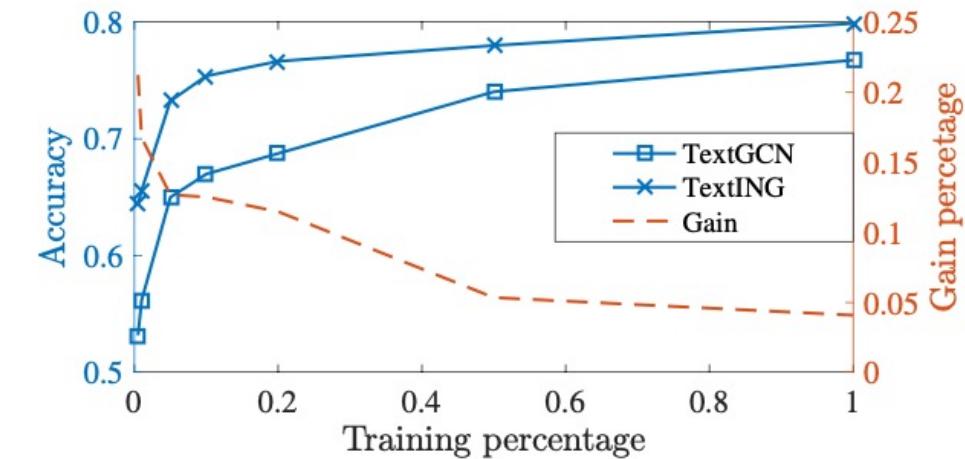
- **Nodes:** words in a document
- **Edges:** co-occurrence (word-word)
- Model the graph with a **Gated Graph Neural Networks** (Li et al., 2015)
- Each document is an individual graph and text level word interactions can be learned in it.
- It can generalize to new words that absent in training, therefore applicable for inductive circumstances.



Inductive Text Classification with TextING

Model	MR*	Ohsumed*
TextGCN	53.15	47.24
TextING	64.43	57.11
# Words in Training	465	7,009
# New Words in Test	18,299	7,148

Accuracy (%) of TextGCN and TextING on MR and Ohsumed



Test performance and gain with different percent of training data on MR.

Short Texts Classification

- Short texts and short text classification tasks

Tweets

The hard truth about the United States is that the money other countries spend on health and infrastructure, we spend on war.

Reviews

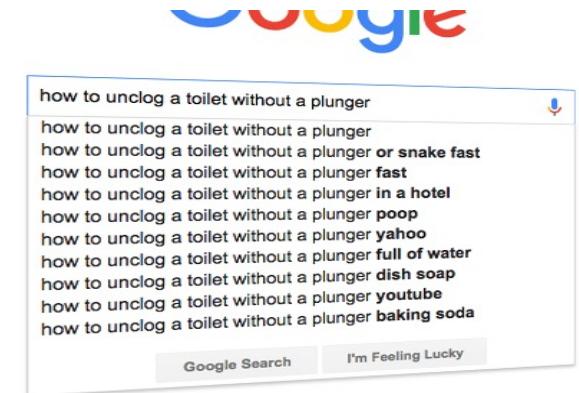
★★★★★ Home secured.
By M. Wolff on January 25, 2014
I purchased this along with 4 turtles and a rat.
18 years worth of karate lessons later, I finally feel safe to leave my house at night.

News

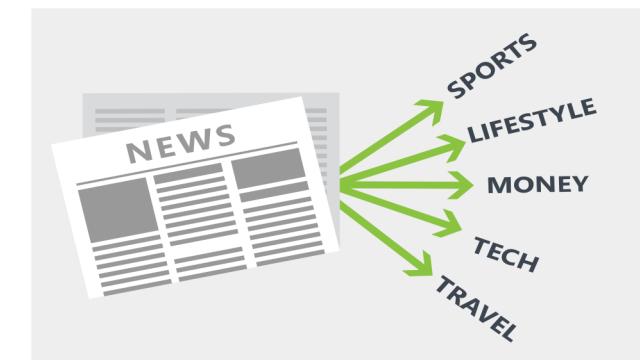
Trump claims he's suing 'various people' for violating confidentiality agreements
By Veronica Stracqualursi and Pamela Brown, CNN
Updated 4:37 PM ET, Sat August 31, 2019

Queries

how to use google effectively?



Query Intent Classification



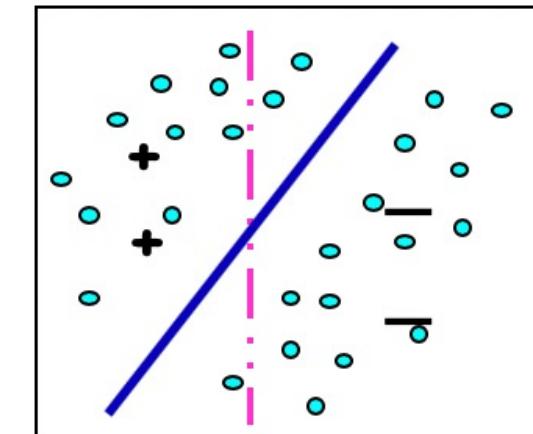
New Categorization

Challenges of Short Texts Classification

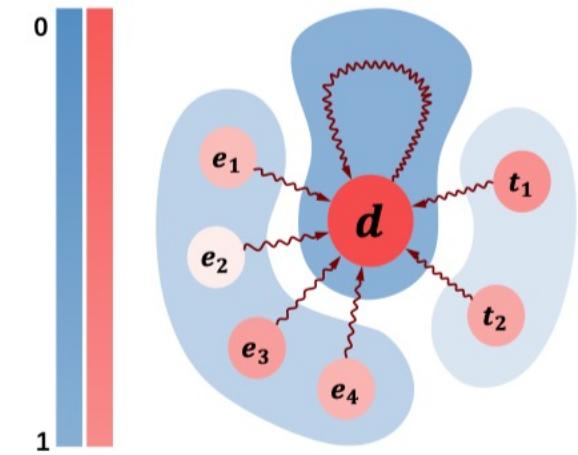
- How to deal with **semantically sparse** and ambiguous.
- How to utilize **limited labeled** training data.
- How to discover **different importance** of different information?



Ambiguous



Limited Label

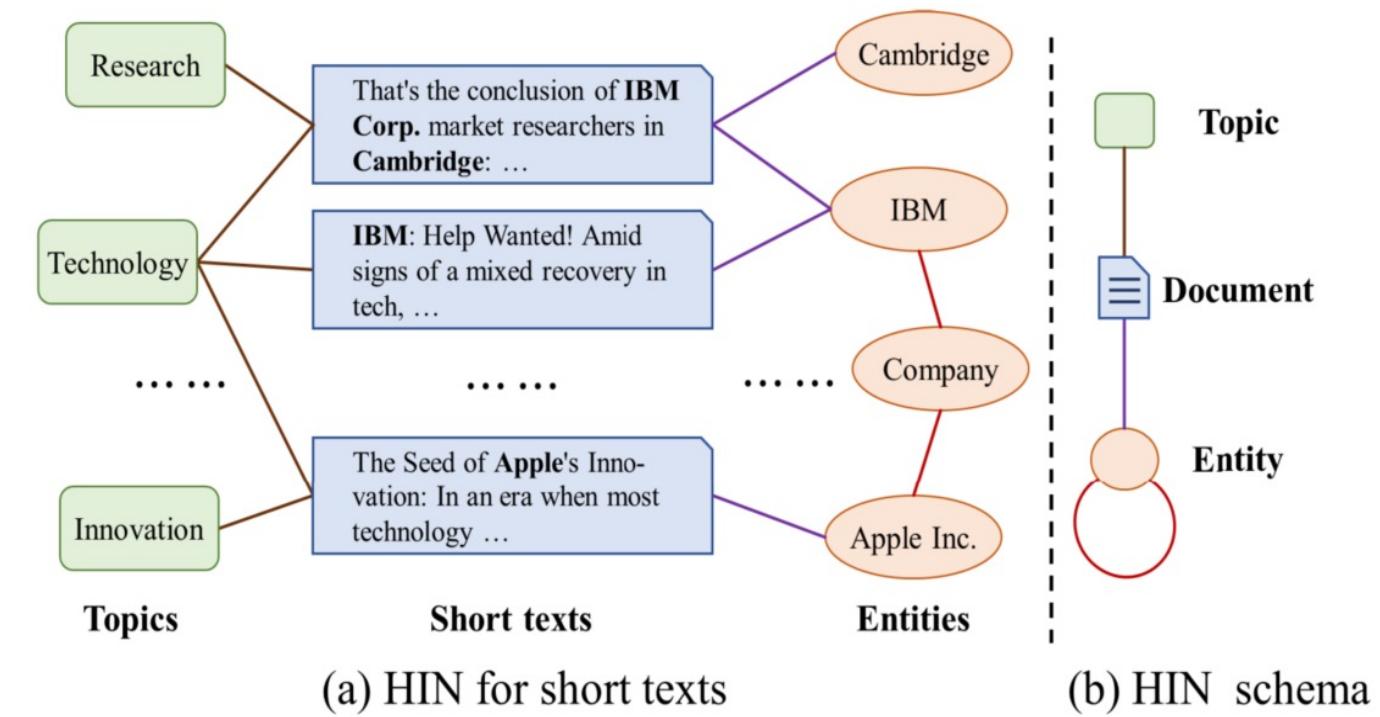


Different Importance

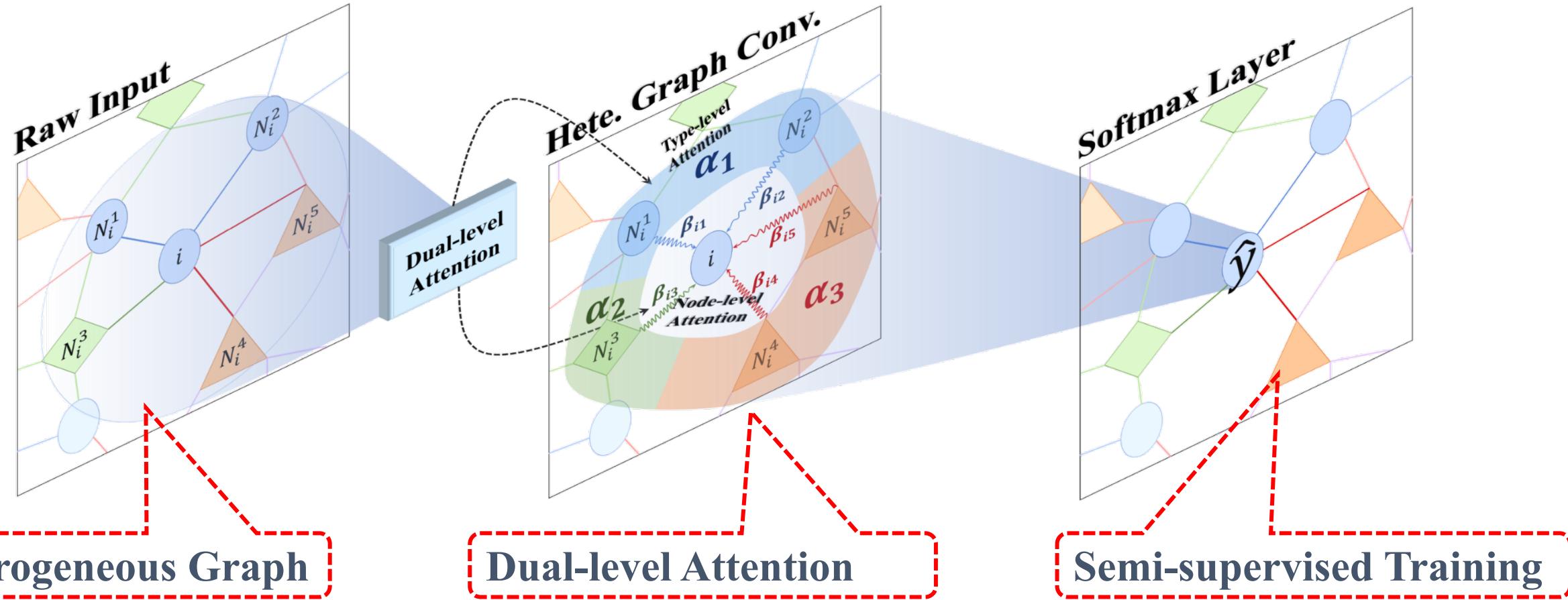
HGAT for Semi-supervised Short Text

- **Nodes**
 - **Text:** Document in the corpus.
 - **Topic:** Mined by LDA.
 - **Entity:** Recognized by an entity linking tool, named TagMe

- **Edges**
 - **Text-Topic:** The text is assigned to the Top P topics.
 - **Text-Entity:** The text contains the entity.
 - **Entity-Entity:** The similarity score is above threshold.



Heterogeneous Graph ATtention Network (HGAT)



Experiments

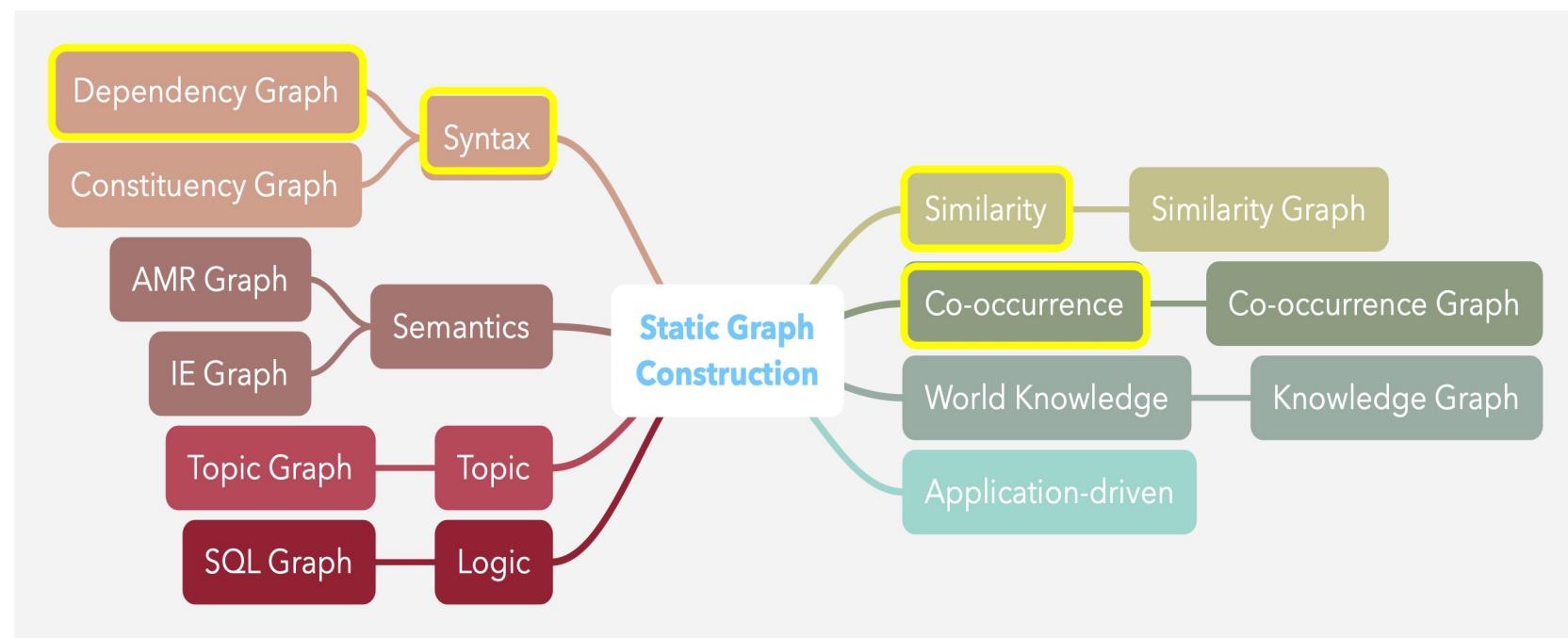
Dataset	AGNews	Snippets	Ohsumed	TagMyNews	MR	Twitter
GCN-HIN	70.87	76.69	40.25	56.33	60.81	61.59
GAT-HIN	70.92	71.56	39.88	56.49	60.24	60.19
HGAT w/o ATT	70.97	80.42	41.31	59.41	62.13	62.35
HGAT-Type	71.54	81.68	41.95	60.78	62.27	62.95
HGAT-Node	71.76	81.93	42.17	61.29	62.31	62.45
HGAT w/o Orphan	71.79	82.01	42.23	61.27	62.40	62.03
HGAT	72.10	82.36	42.68	61.72	62.75	63.21

Test accuracy of HGAT and its variants

Summary

GNN enables:

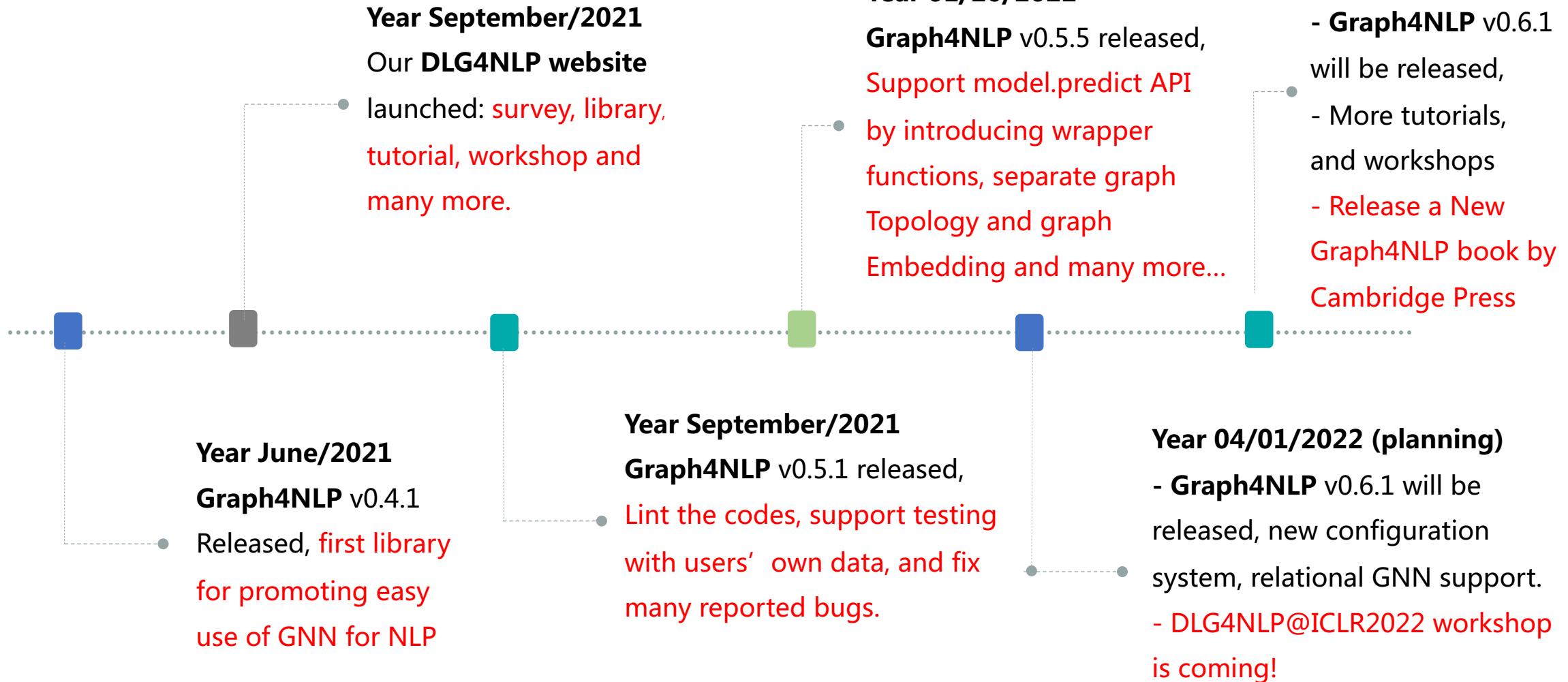
- Encode multi-scale information
- Encode heterogenous information



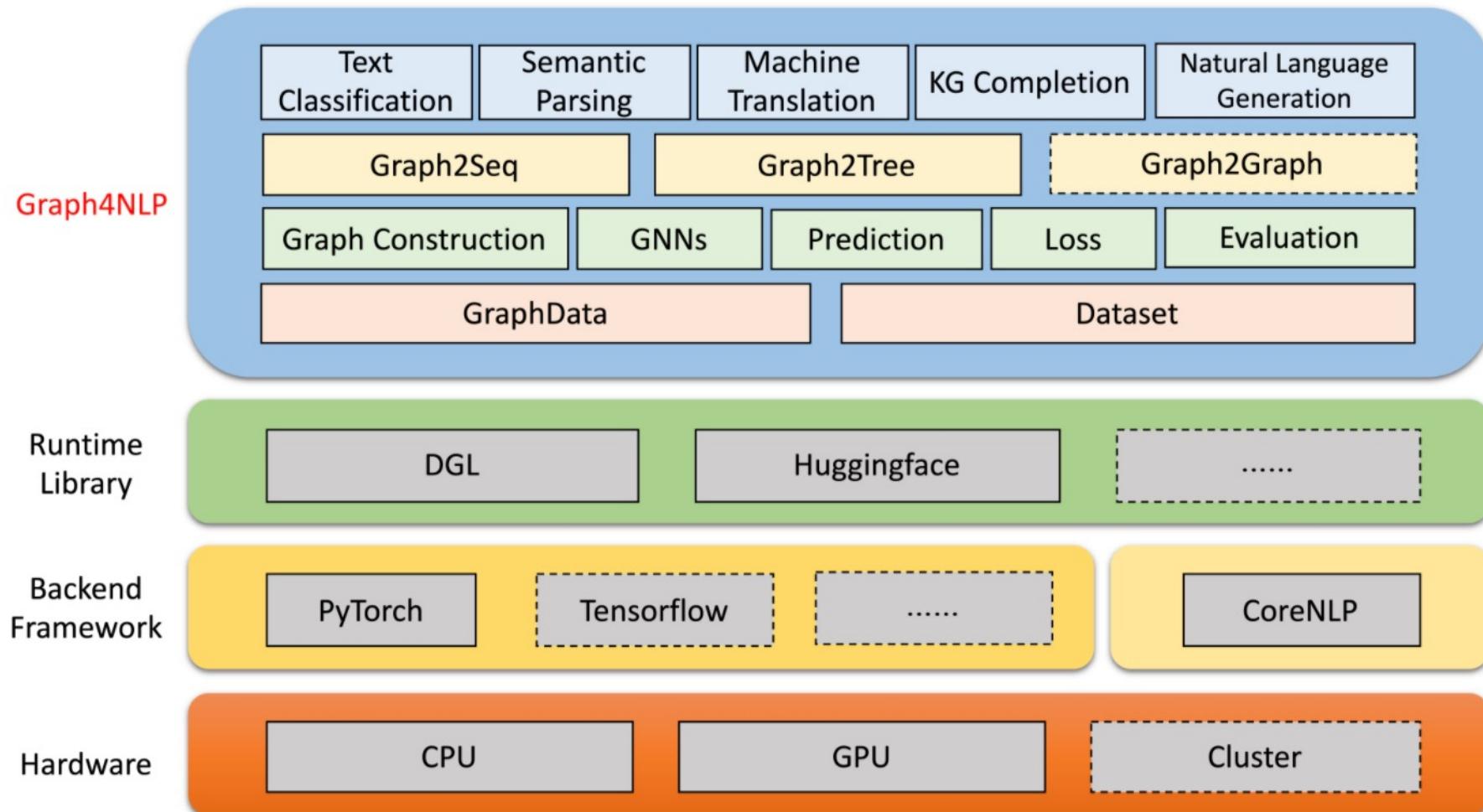
Hands-on Demonstration

Graph4NLP: A Library for Deep Learning on Graphs for NLP

Graph4NLP: A Brief History and Future



Overall Architecture of Graph4NLP Library



Key Features and Future Releases

Easy-to-use and Flexible

Provides both full implementations of state-of-the-art models and also flexible interfaces to build customized models with whole-pipeline support

Rich Set of Learning Resources

Provide a variety of learning materials including code demos, code documentations, research tutorials and videos, and paper survey

High Running Efficiency and Extensibility

Build upon highly-optimized runtime libraries including DGL and provide highly modularization blocks

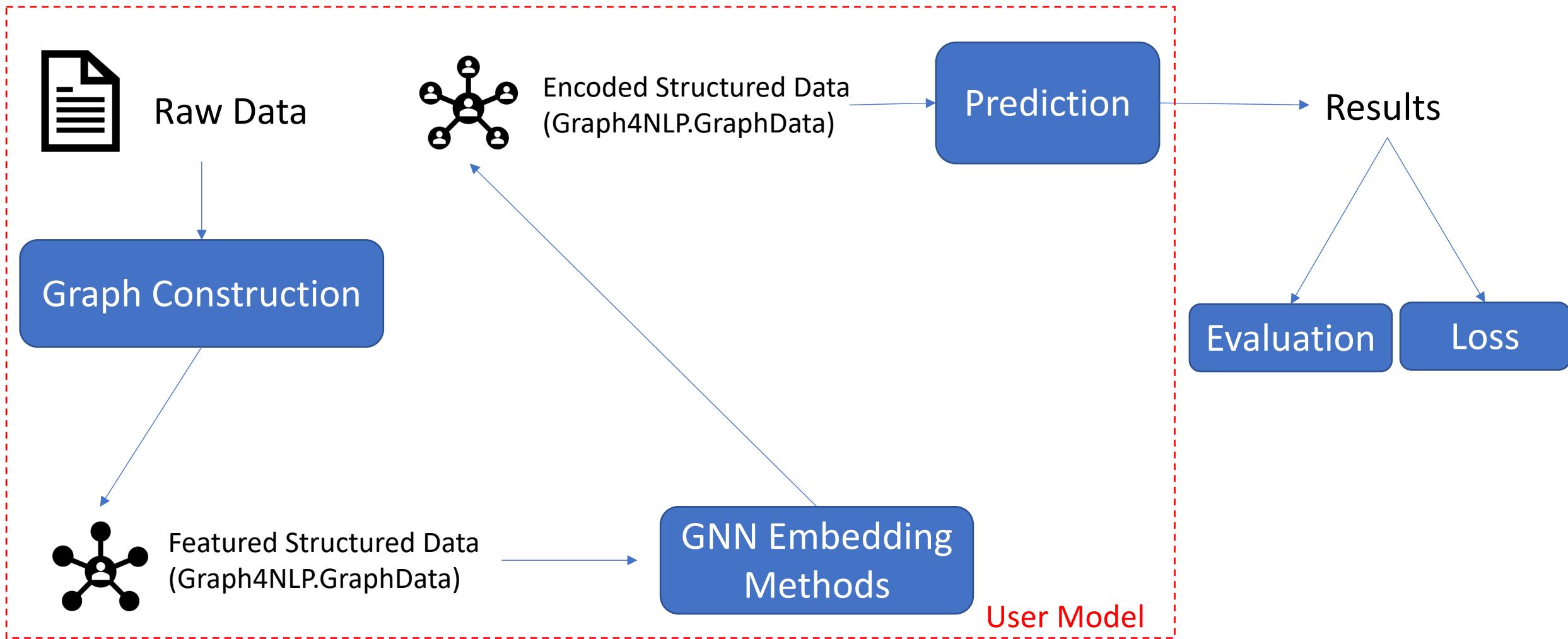
Comprehensive Code Examples

Provide a comprehensive collection of NLP applications and the corresponding code examples for quick-start

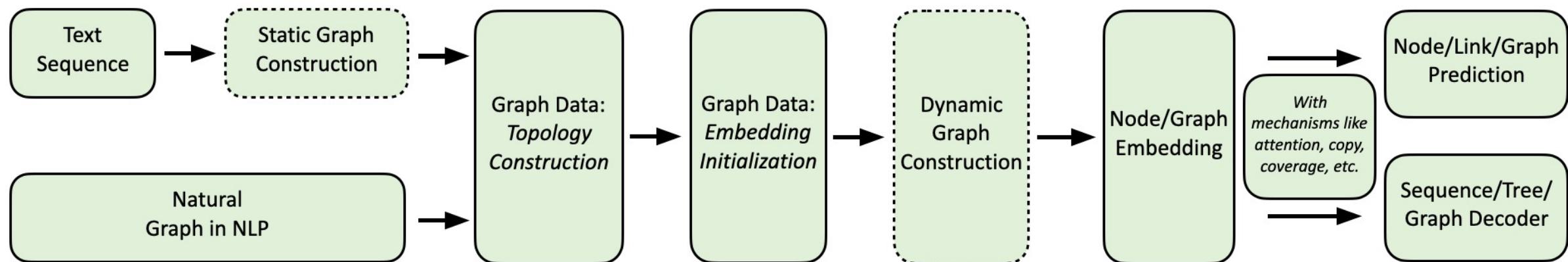
Future Releases

- **V0.6 new features:** new configuration system, relational GNN support, etc.
- **Future todo:** Native multi-GPU support, better support for customized graph construction, etc.

Data Flow of Graph4NLP



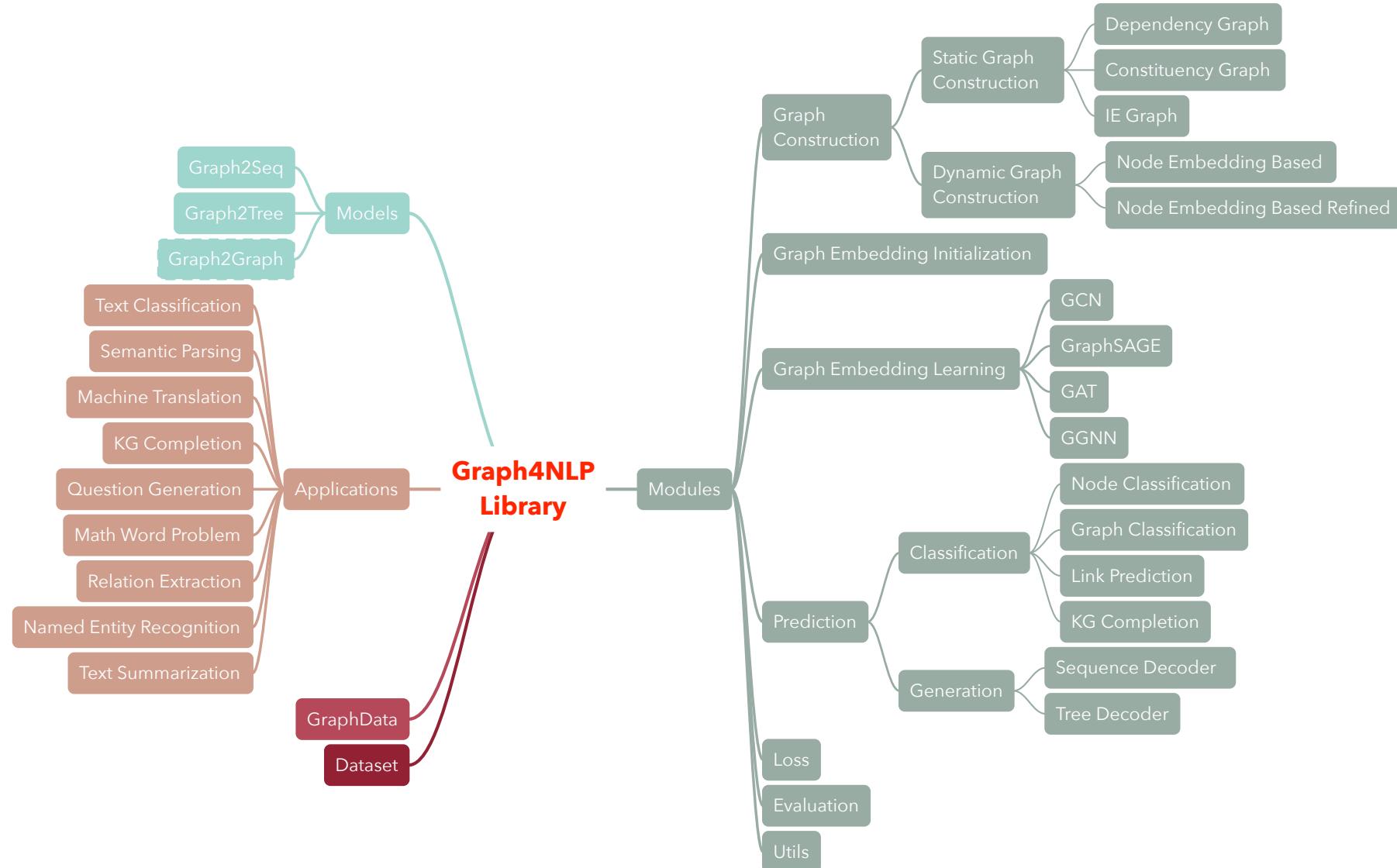
Computing Flow of Graph4NLP



Performance of Built-in NLP Tasks

Task	Dataset	GNN Model	Graph construction	Evaluation	Performance
Text classification	TRECT	GAT	Dependency	Accuracy	0.948
	CAirline				0.769
	CNSST				0.538
Semantic Parsing	JOBS	SAGE	Constituency	Execution accuracy	0.936
Question generation	SQuAD	GGNN	Dependency	BLEU-4	0.15175
Machine translation	IWSLT14	GCN	Dynamic	BLEU-4	0.3212
Summarization	CNN(30k)	GCN	Dependency	ROUGE-1	26.4
Knowledge graph completion	Kinship	GCN	Dependency	MRR	82.4
Math word problem	MAWPS	SAGE	Dynamic	Solution accuracy	76.4
	MATHQA			Exact match	61.07

Dive Into Graph4NLP Library



Graph Construction Module

- Static graph construction
 - Dependency graph construction
 - Constituency graph construction
 - IE graph construction
- Dynamic graph construction
 - Node embedding based
 - Node embedding based refined (i.e., static & dynamic hybrid)

```
self.graph_topology = NodeEmbeddingBasedGraphConstruction(
    sim_metric_type=config["gl_metric_type"],
    num_heads=config["gl_num_heads"],
    top_k_neigh=config["gl_top_k"],
    epsilon_neigh=config["gl_epsilon"],
    smoothness_ratio=config["gl_smoothness_ratio"],
    connectivity_ratio=config["gl_connectivity_ratio"],
    sparsity_ratio=config["gl_sparsity_ratio"],
    input_size=config["num_hidden"],
    hidden_size=config["gl_num_hidden"],
)
```

```
self.graph_topology = NodeEmbeddingBasedRefinedGraphConstruction(
    config["init_adj_alpha"],
    sim_metric_type=config["gl_metric_type"],
    num_heads=config["gl_num_heads"],
    top_k_neigh=config["gl_top_k"],
    epsilon_neigh=config["gl_epsilon"],
    smoothness_ratio=config["gl_smoothness_ratio"],
    connectivity_ratio=config["gl_connectivity_ratio"],
    sparsity_ratio=config["gl_sparsity_ratio"],
    input_size=config["num_hidden"],
    hidden_size=config["gl_num_hidden"],
)
```

Graph Embedding Initialization Module

- Single-token & multi-token node/edge
- Various built-in strategies for node/edge embedding initialization (non-exhaustive list)
 - 'w2v'
 - 'w2v_bilstm'
 - 'bert'
 - 'bert_bilstm'
 - 'w2v_bert'
 - 'w2v_bert_bilstm'

```
self.graph_initializer = GraphEmbeddingInitialization(
    word_vocab=self.vocab_model.in_word_vocab,
    embedding_style=embedding_style,
    hidden_size=config["num_hidden"],
    word_dropout=config["word_dropout"],
    rnn_dropout=config["rnn_dropout"],
    fix_word_emb=not config["no_fix_word_emb"],
    fix_bert_emb=not config.get("no_fix_bert_emb", False),
)
```

```
embedding_style = {
    "single_token_item": True if self.graph_name != "ie" else False,
    "emb_strategy": config.get("emb_strategy", "w2v_bilstm"),
    "num_rnn_layers": 1,
    "bert_model_name": config.get("bert_model_name", "bert-base-uncased"),
    "bert_lower_case": True,
}
```

Graph Embedding Learning Module

- Common GNN variants
 - GCN
 - GAT
 - GraphSAGE
 - GGNN
- direction_option
 - ‘undirected’
 - ‘bi_fuse’
 - ‘bi_sep’
- use_edge_weight

```
self.gnn = GGNN(  
    config["gnn_num_layers"],  
    config["num_hidden"],  
    config["num_hidden"],  
    config["num_hidden"],  
    feat_drop=config["gnn_dropout"],  
    direction_option=config["gnn_direction_option"],  
    bias=True,  
    use_edge_weight=use_edge_weight,  
)
```

Prediction Module

- Classification
 - Node classification
 - Graph classification
 - Link prediction
 - KG completion
 - Pooling: avg_pool, max_pool
- Generation
 - Sequence decoder
 - Tree decoder
 - Attention, copy, coverage mechanisms

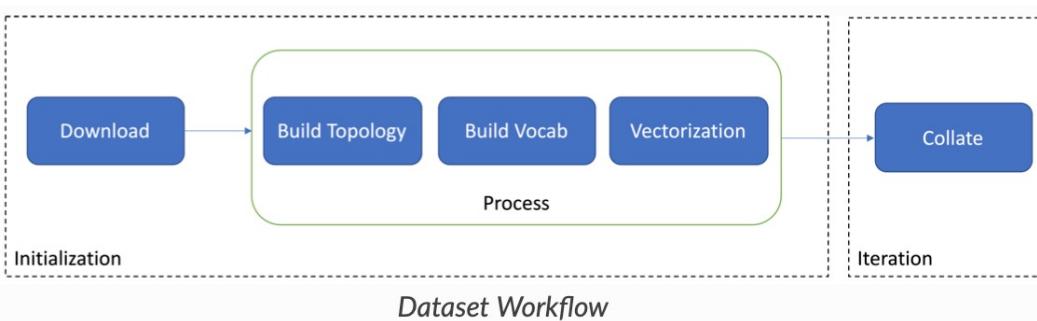
```
self.seq_decoder = StdRNNDecoder(
    rnn_type=rnn_type,
    max_decoder_step=decoder_length,
    input_size=input_size,
    hidden_size=hidden_size,
    graph_pooling_strategy=graph_pooling_strategy,
    word_emb=self.dec_word_emb,
    vocab=vocab_model.out_word_vocab,
    attention_type=attention_type,
    fuse_strategy=fuse_strategy,
    node_type_num=node_type_num,
    rnn_emb_input_size=rnn_input_size,
    use_coverage=use_coverage,
    use_copy=use_copy,
    tgt_emb_as_output_layer=tgt_emb_as_output_layer,
    dropout=rnn_dropout,
)
```

```
self.decoder = StdTreeDecoder(
    attn_type=dec_attention_type,
    embeddings=self.enc_word_emb.word_emb_layer
    if self.use_share_vocab
    else self.tgt_word_embedding,
    enc_hidden_size=gnn_hidden_size,
    dec_emb_size=self.tgt_vocab.embedding_dims,
    dec_hidden_size=dec_hidden_size,
    output_size=self.output_size,
    criterion=self.criterion,
    teacher_force_ratio=dec_teacher_forcing_rate,
    use_sibling=dec_use_sibling,
    use_copy=self.use_copy,
    dropout_for_decoder=dec_dropout,
    max_dec_seq_length=dec_max_decoder_step,
    max_dec_tree_depth=dec_max_tree_depth,
    tgt_vocab=self.tgt_vocab,
)
```

Built-in high-level
[Graph2Seq](#),
[Graph2Tree](#) APIs.
 Config in, model
 out.

Dataset

- Built-in dataset types
 - Text2TextDataset
 - TextToTreeDataset
 - Text2LabelDataset
 - SequenceLabelingDataset
 - DoubleText2TextDataset



```
class TrecDataset(Text2LabelDataset):
    @property
    def raw_file_names(self):
        """3 reserved keys: 'train', 'val' (optional), 'test'. Represent the split of dataset."""
        return {"train": "train.txt", "test": "test.txt"}

    @property
    def processed_file_names(self):
        """At least 3 reserved keys should be fiiled: 'vocab', 'data' and 'label'."""
        return {"vocab": "vocab.pt", "data": "data.pt", "label": "label.pt"}

    def __init__(
```

```
dataset = TrecDataset(
    root_dir=self.config["graph_construction_args"]["graph_construction_share"]["root_dir"],
    topology_subdir=topology_subdir,
    graph_name=self.graph_name,
    dynamic_init_graph_name=self.config["graph_construction_args"][
        "graph_construction_private"
    ]["dynamic_init_graph_name"],
    dynamic_init_topology_aux_args={"dummy_param": 0},
    pretrained_word_emb_name=self.config["pretrained_word_emb_name"],
    merge_strategy=self.config["graph_construction_args"]["graph_construction_private"][
        "merge_strategy"
    ],
    edge_strategy=self.config["graph_construction_args"]["graph_construction_private"][
        "edge_strategy"
    ],
    min_word_vocab_freq=self.config.get("min_word_freq", 1),
    word_emb_size=self.config.get("word_emb_size", 300),
    seed=self.config["seed"],
    thread_number=self.config["graph_construction_args"]["graph_construction_share"][
        "thread_number"
    ],
    port=self.config["graph_construction_args"]["graph_construction_share"]["port"],
    timeout=self.config["graph_construction_args"]["graph_construction_share"]["timeout"],
    reused_label_model=None,
)
```

Inference

- Inference wrapper
 - classifier_inference_wrapper
 - generator_inference_wrapper
 - generator_inference_wrapper_for_tree

```
self.inference_tool = GeneratorInferenceWrapper(
    cfg=self.config, model=self.model,
    dataset=DoubleText2TextDataset,
    data_item=DoubleText2TextDataItem,
    beam_size=self.config["beam_size"],
    topk=1, lower_case=True,
    tokenizer=word_tokenize,
    share_vocab=True,
)
```

```
self.inference_tool = ClassifierInferenceWrapper(
    cfg=self.config,
    model=self.model,
    label_names=self.model.label_model.le.classes_.tolist(),
    dataset=Text2LabelDataset,
    data_item=Text2LabelDataItem,
    lower_case=True,
    tokenizer=word_tokenize,
)
```



Demo 1: Text Classification Application

- 1) git clone https://github.com/graph4ai/graph4nlp_demo
- 2) follow Get Started instructions in README



Demo 1: Building a Text Classification Application

```
def forward(self, graph_list, tgt=None, require_loss=True):
    # graph embedding initialization
    batch_gd = self.graph_initializer(graph_list)

    # run dynamic graph construction if turned on
    if hasattr(self, "graph_topology") and hasattr(self.graph_topology, "dynamic_topology"):
        batch_gd = self.graph_topology.dynamic_topology(batch_gd)

    # run GNN
    self.gnn(batch_gd)

    # run graph classifier
    self.clf(batch_gd)
    logits = batch_gd.graph_attributes["logits"]

    if require_loss:
        loss = self.loss(logits, tgt)
        return logits, loss
    else:
        return logits
```

Demo 1: Building a Text Classification Application

Graph embedding initialization API, various built-in options, can be customized

```
embedding_style = {
    "single_token_item": True if self.graph_name != "ie" else False,
    "emb_strategy": config.get("emb_strategy", "w2v_bilstm"),
    "num_rnn_layers": 1,
    "bert_model_name": config.get("bert_model_name", "bert-base-uncased"),
    "bert_lower_case": True,
}

self.graph_initializer = GraphEmbeddingInitialization(
    word_vocab=self.vocab_model.in_word_vocab,
    embedding_style=embedding_style,
    hidden_size=config["num_hidden"],
    word_dropout=config["word_dropout"],
    rnn_dropout=config["rnn_dropout"],
    fix_word_emb=not config["no_fix_word_emb"],
    fix_bert_emb=not config.get("no_fix_bert_emb", False),
)
```

Demo 1: Building a Text Classification Application

```
self.graph_topology = NodeEmbeddingBasedGraphConstruction(  
    sim_metric_type=config["gl_metric_type"],  
    num_heads=config["gl_num_heads"],  
    top_k_neigh=config["gl_top_k"],  
    epsilon_neigh=config["gl_epsilon"],  
    smoothness_ratio=config["gl_smoothness_ratio"],  
    connectivity_ratio=config["gl_connectivity_ratio"],  
    sparsity_ratio=config["gl_sparsity_ratio"],  
    input_size=config["num_hidden"],  
    hidden_size=config["gl_num_hidden"],  
)
```

Graph construction API,
various built-in options,
can be customized

Demo 1: Building a Text Classification Application

Graph embedding learning API, various built-in options, can be customized

```
self.gnn = GraphSAGE(config['gnn_num_layers'],
                      config['num_hidden'],
                      config['num_hidden'],
                      config['num_hidden'],
                      config['graphsage_aggregate_type'],
                      direction_option=config['gnn_direction_option'],
                      feat_drop=config['gnn_dropout'],
                      bias=True,
                      norm=None,
                      activation=F.relu,
                      use_edge_weight=use_edge_weight)
```

Demo 1: Building a Text Classification Application

Prediction API, various built-in options, can be customized

```
self.clf = FeedForwardNN(2 * config['num_hidden'] \
    if config['gnn_direction_option'] == 'bi_sep' \
    else config['num_hidden'],
    config['num_classes'],
    [config['num_hidden']],
    graph_pool_type=config['graph_pooling'],
    dim=config['num_hidden'],
    use_linear_proj=config['max_pool_linear_proj'])
```

Demo 1: Building a Text Classification Application

```
dataset = TrecDataset(  
    root_dir=self.config["graph_construction_args"]["graph_construction_share"]["root_dir"],  
    topology_subdir=topology_subdir,  
    graph_name=self.graph_name,  
    dynamic_init_graph_name=self.config["graph_construction_args"] [  
        "graph_construction_private"  
    ] ["dynamic_init_graph_name"],  
    dynamic_init_topology_aux_args={"dummy_param": 0},  
    pretrained_word_emb_name=self.config["pretrained_word_emb_name"],  
    merge_strategy=self.config["graph_construction_args"]["graph_construction_private"] [  
        "merge_strategy"  
    ],  
    edge_strategy=self.config["graph_construction_args"]["graph_construction_private"] [  
        "edge_strategy"  
    ],  
    min_word_vocab_freq=self.config.get("min_word_freq", 1),  
    word_emb_size=self.config.get("word_emb_size", 300),  
    seed=self.config["seed"],  
    thread_number=self.config["graph_construction_args"]["graph_construction_share"] [  
        "thread_number"  
    ],  
    port=self.config["graph_construction_args"]["graph_construction_share"]["port"],  
    timeout=self.config["graph_construction_args"]["graph_construction_share"]["timeout"],  
    reused_label_model=None,  
)
```

Dataset API, various built-in options, can be customized

Demo 2: Building a Semantic Parsing Application

- 1) git clone https://github.com/graph4ai/graph4nlp_demo
- 2) follow Get Started instructions in README



Demo 2: Building a Semantic Parsing Application

Sentence: *What is the capital of Germany?*

(Semantic parsing)



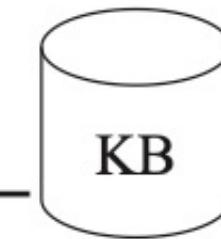
Logical form: $\lambda x . \text{capital}(\text{Germany}, x)$

(Execution)



Result:

{*Berlin*} ←



Demo 2: Building a Semantic Parsing Application

Graph2Seq API

```
def _build_model(self):
    self.model = Graph2Seq.from_args(self.opt, self.vocab).to(self.device)
```

DLG4NLP: Future Directions and Conclusions

Future Directions

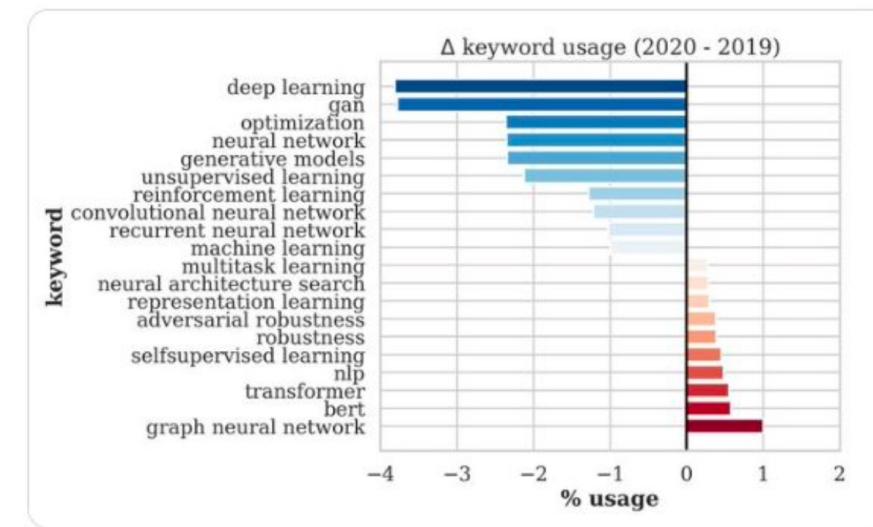
[Vashisht et al. EMNLP'19 Tutorial]

- The Rise of **GNN + NLP**

#ICLR2020 submissions on graph neural networks, NLP and robustness have the greatest growth. [@iclr_cont](#)
[@openreviewnet](#)

- Graph Construction for NLP

- Dynamic graph construction are largely underexplored!
- How to effectively combine advantages of static graph and dynamic graph?
- How to construct heterogeneous dynamic graph?
- How to make dynamic graph construction itself scalable?



Future Directions

- **Scaling GNNs to Large Graphs**
 - Most existing multi-relational or heterogeneous GNNs will have scalability issues when applied to large graphs in NLP such as KGs (> 1m)
- **GNNs + Transformer in NLP**
 - How to effectively combine the advantages of GNNs and Transformer?
 - Is graph transformer the best way to utilize?
- **Pretraining GNNs for NLP**
 - Information Retrieval/ Search

Future Directions

- **Graph-to-graph Learning in NLP**
 - How to effectively develop Graph-to-Graph models for solving graph transformation problem in NLP (i.e. information extraction)?
- **Joint Text and KG Reasoning in NLP**
 - Joint text and KG reasoning is less explored although GNNs for multi-hop reasoning gains popularity
- **Incorporate Source and Context into Knowledge Graph Construction and Verification**

Conclusions

- Deep Learning on Graphs for NLP is a fast-growing area today!
- Since graph can naturally encode complex information, it could bridge a gap by combining both **empirical domain knowledges and the power of deep learning**.
- For a NLP task,
 - how to convert text sequence into the best graph (directed, multi-relation, heterogeneous)
 - how to determine proper graph representation learning technique?
- **Our Graph4NLP library aims to make easy use of GNNs for NLP:**
 - Website: <https://dlg4nlp.github.io/index.html>
 - Survey: <https://arxiv.org/abs/2106.06090>
 - Library: <https://github.com/graph4ai/graph4nlp>
 - Demo: https://github.com/graph4ai/graph4nlp_demo
 - Github literature list: https://github.com/graph4ai/graph4nlp_literature

FAQ

- GNNs vs Transformers
- GNNs vs BERT
- Why is GNN helpful?
- Pretraining GNNs for NLP
- How to train GNNs on large-scale graphs?
- Which graph construction method is the best? How to choose?