#NAACL2022

KCD: Knowledge Walks and Textual Cues Enhanced Political Perspective Detection in News Media

Aug 18, 2022

Shangbin Feng

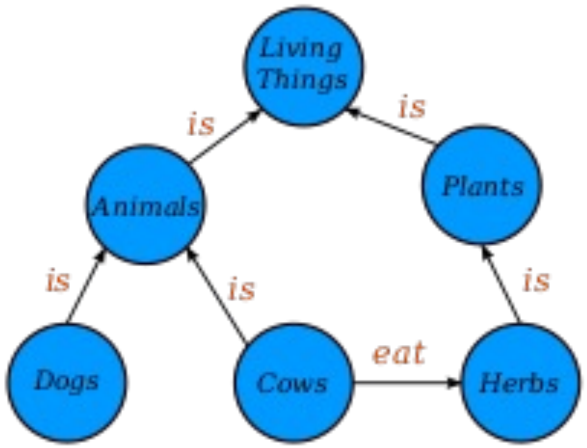
University of Washington

shangbin@cs.washington.edu

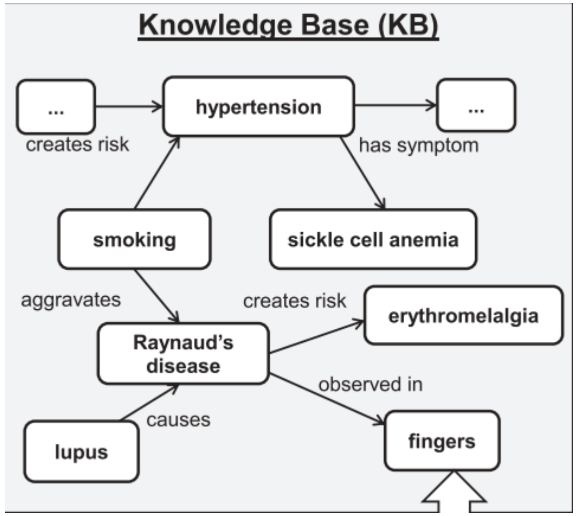
bunsenfeng.github.io

Knowledge Graphs (KGs)

- Structured representation of commonsense and domain knowledge



commonsense[1]



domain-specific[2]

[1]https://en.wikipedia.org/wiki/Knowledge_graph
[2]Ernst, Patrick, Amy Siu, and Gerhard Weikum. "Knowlife: a versatile approach for constructing a large knowledge graph for biomedical sciences." *BMC bioinformatics* 16.1 (2015): 1-13.

KGs in NLP

- Question Answering
  - QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering
  - GreaseLM: Graph REASoning Enhanced Language Models for Question Answering
- Text generation
  - Text generation from knowledge graphs with graph transformers
  - Language Generation with Multi-Hop Reasoning on Commonsense Knowledge Graph
- Knowledge and LMs
  - KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation
  - BertNet: Harvesting Knowledge Graphs from Pretrained Language Models
- Social text analysis
  - Compare to The Knowledge: Graph Neural Fake News Detection with External Knowledge
  - KCD: Knowledge Walks and Textual Cues Enhanced Political Perspective Detection in News Media
- …

Three lanes of using KG in NLP

- Feature extraction
    - Extract features with KG embedding models
    - Inject such features
- Enhance LMs
    - Incorporating KGs into pre-trained LMs
    - Mostly with adapters
- Graph and GNNs
    - KG subgraphs with GNNs
    - Document graph with GNNs

1/3 Feature Extraction

- CompareNet[3]

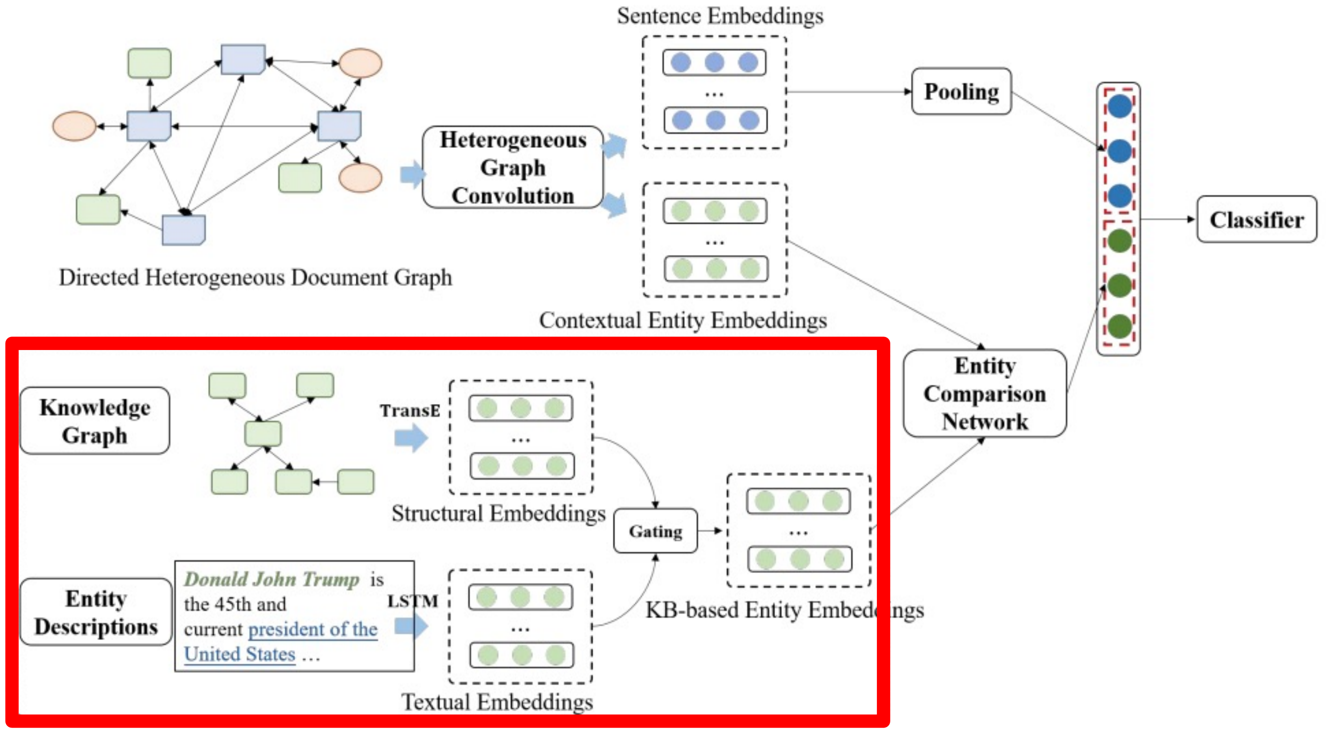


Figure 2: The overview of our proposed model CompareNet.

[3]Hu, Linmei, et al. "Compare to the knowledge: Graph neural fake news detection with external knowledge." *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2021.

2/3 Enhance LMs

- Mixture-of-Partitions[4]

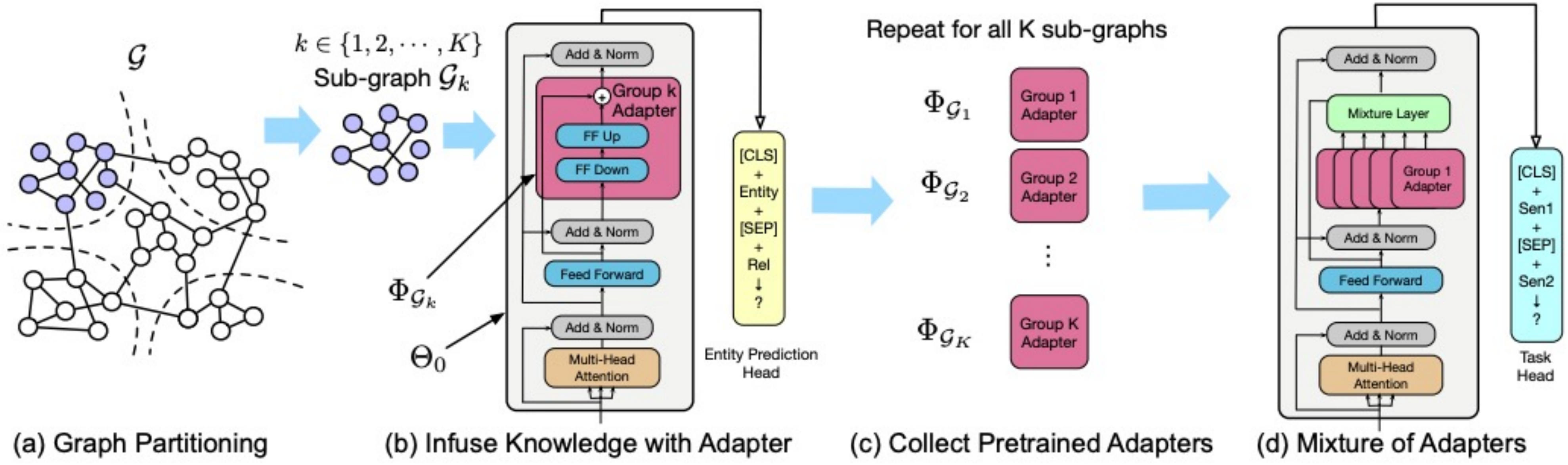
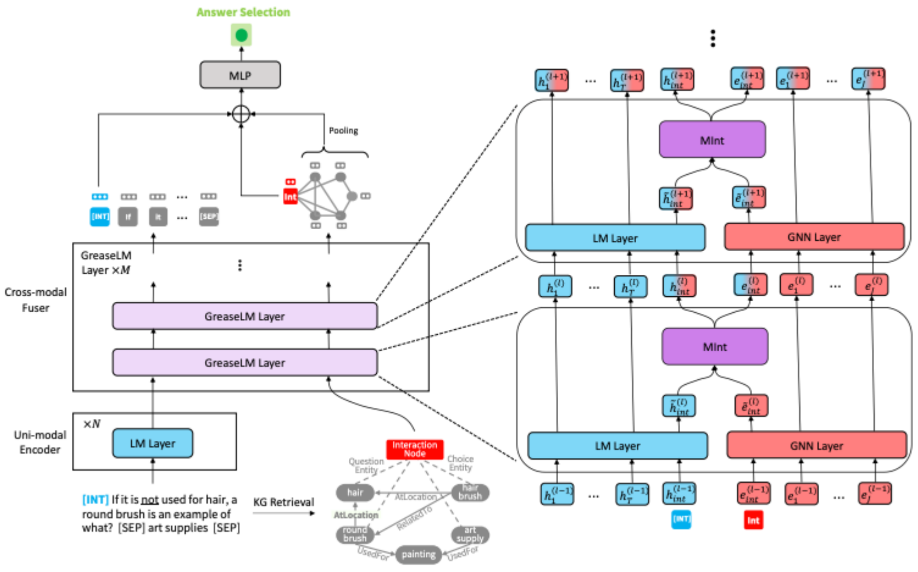


Figure 1: Overview of the proposed MoP.

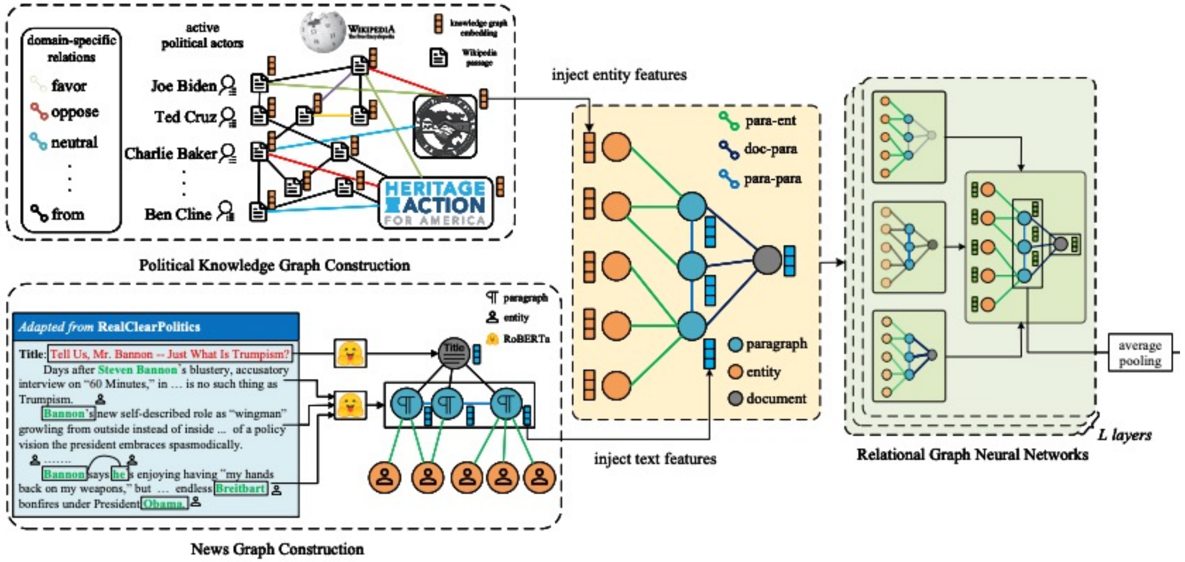
[4]Meng, Zaiqiao, et al. "Mixture-of-Partitions: Infusing Large Biomedical Knowledge Graphs into BERT." *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 2021.

3/3 Graph and GNNs

- GreaseLM[5] and KGAP[6]



GNNs on KG subgraphs



GNNs on "document graphs"

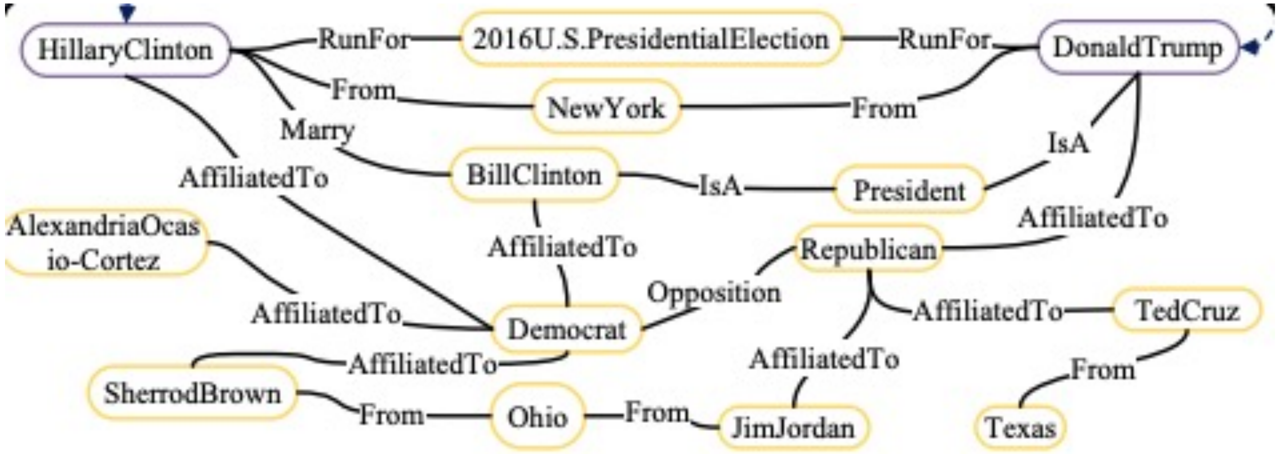
[5]Zhang, X., et al. "GreaseLM: Graph REASoning Enhanced Language Models for Question Answering." *International Conference on Representation Learning (ICLR)*. 2022.
[6]Feng, Shangbin, et al. "KGAP: Knowledge Graph Augmented Political Perspective Detection in News Media." *arXiv preprint arXiv:2108.03861* (2021).

However…

- Multi-hop knowledge reasoning is missing

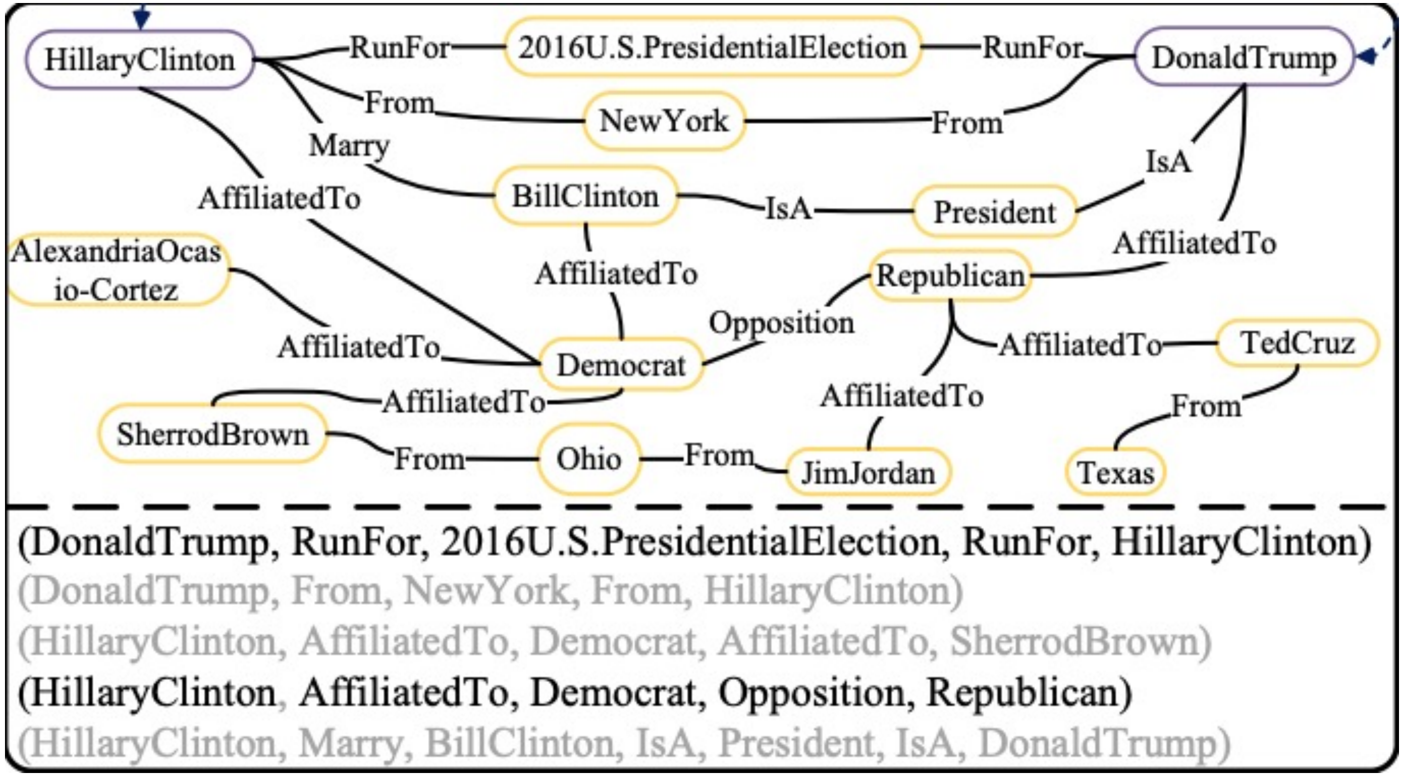
Source: Daily Kos   Stance: Left

CNN *is reporting* that the ***Trump*** campaign were offered access to Wikileaks documents, including special access to a Wikileaks website, a month before Wikileaks began publishing ……
The email, which apparently *slipped* the memory of history's most forgetful campaign team, included a decryption key and address for documents stolen by Russian hackers and later ……
He *claimed* that …… "…… That means that the vast amount of stuff that we are publishing about ***Clinton*** will have a much higher impact, because it won't be perceived as coming from ……"
……



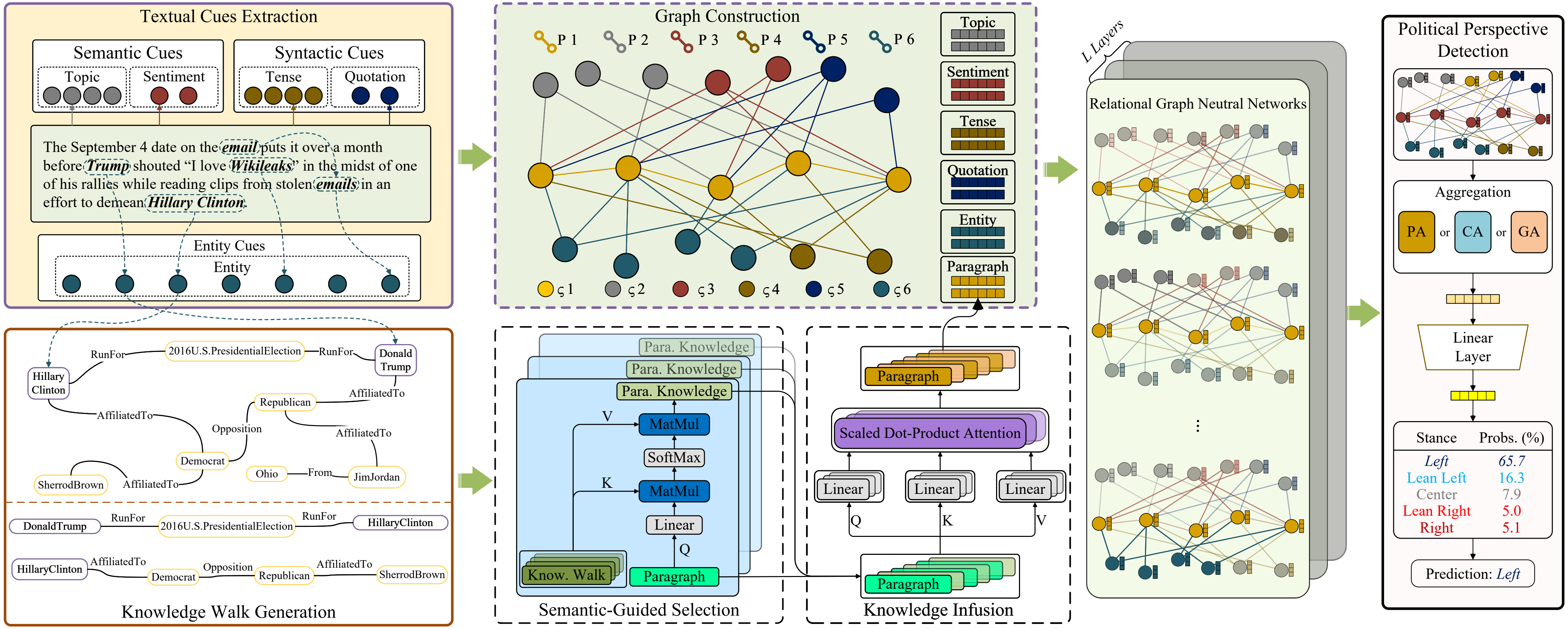
Knowledge Walk

- **Knowledge walk**: multi-hop path on knowledge graphs
- However, …



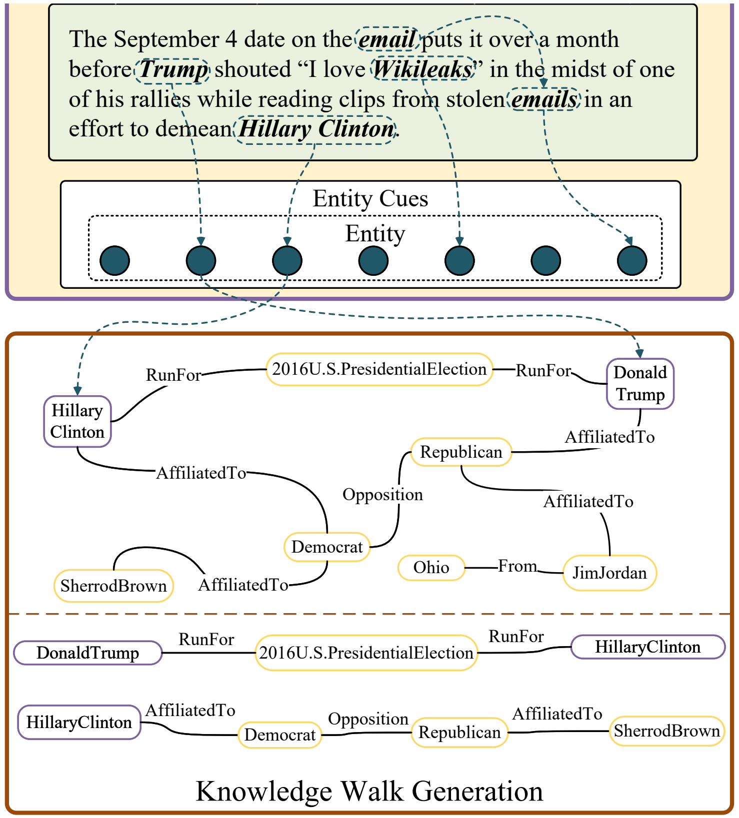
KCD

- Enable multi-hop knowledge reasoning with knowledge walks



1/5 knowledge walk (kw) generation

- Biased random walk on KGs



K-hop knowledge walk:

$$kw_i = \{e_{(0)}, r_{0,1}, e_{(1)}, \ldots, r_{K-1,K}, e_{(K)}\} \quad (2)$$

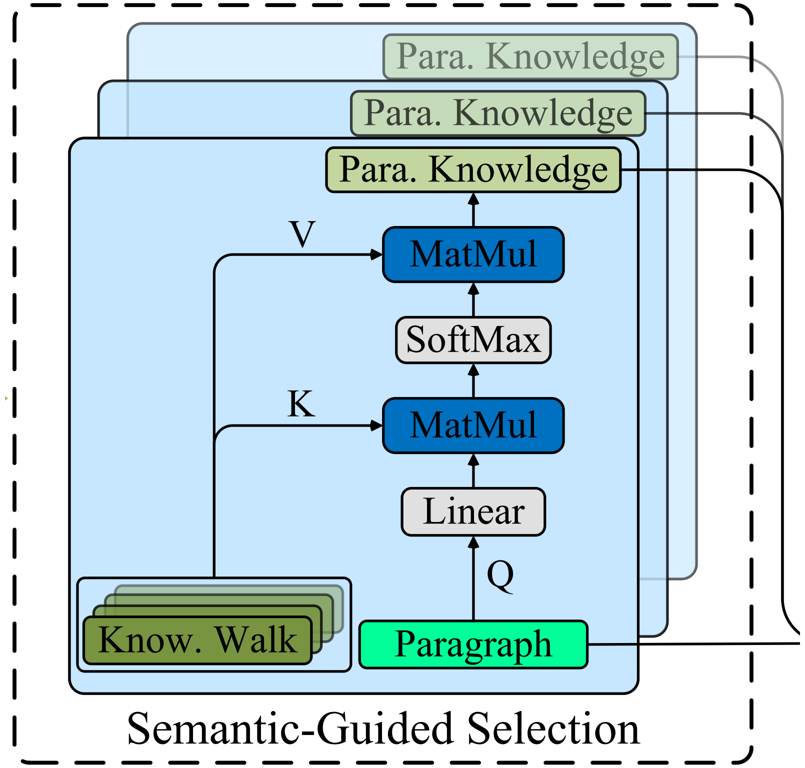
Arriving at $e_{(i)}$ from $e_{(i-1)}$:

$$P(e_{(i)}|e_{(i-1)}, r_{i-1,i}) = \frac{exp(p(r_{i-1,i}))}{\sum_{j=1}^{|N_r(i-1)|} exp(p(r_j))} \quad (3)$$

P(r): importance of relation *r*

2/5 kw importance estimation

- kw aggregation based on semantic relevance



For kw_i:

generate sentence t_i by concatenating the names of entities and relations

$$v_i^k = RoBERTa(t_i) \tag{4}$$

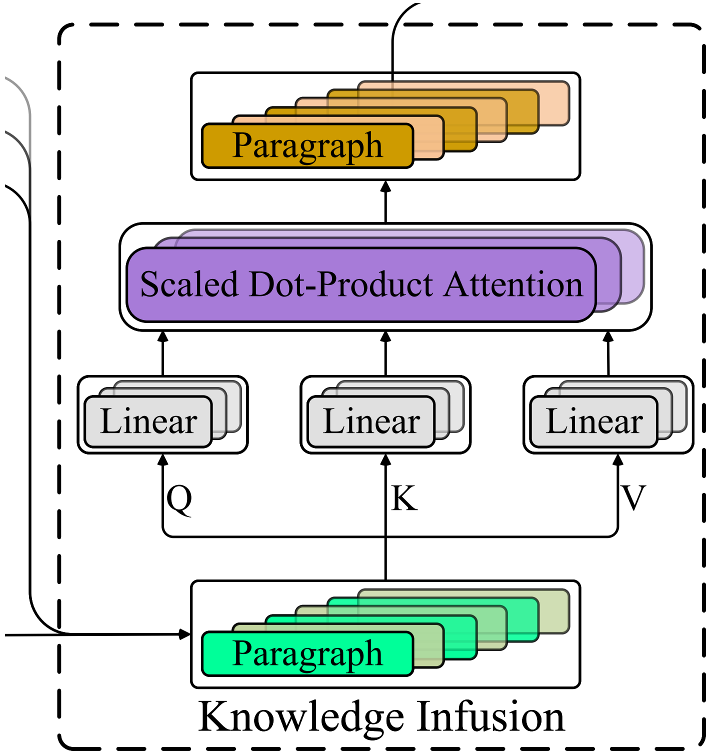
Use paragraph rep. v_i^s as query, (4) as k&v

$$v_i^p = \sum_{j=1}^{m} \frac{exp(\alpha \cdot v_{i,j}^k)}{\sum_{q=1}^{m} exp(\alpha \cdot v_{i,q}^k)} v_{i,j}^k \tag{5}$$

$$\alpha = \phi(W_a v_i^s + b_a) \tag{6}$$

3/5 knowledge infusion

- Infuse kw representations with document representations



v_i^s: paragraph *i* LM representation
v_i^p: paragraph *i* knowledge walk representation

$$T = concat([v_1^s, v_1^p, ..., v_n^s, v_n^p]) \quad (7)$$

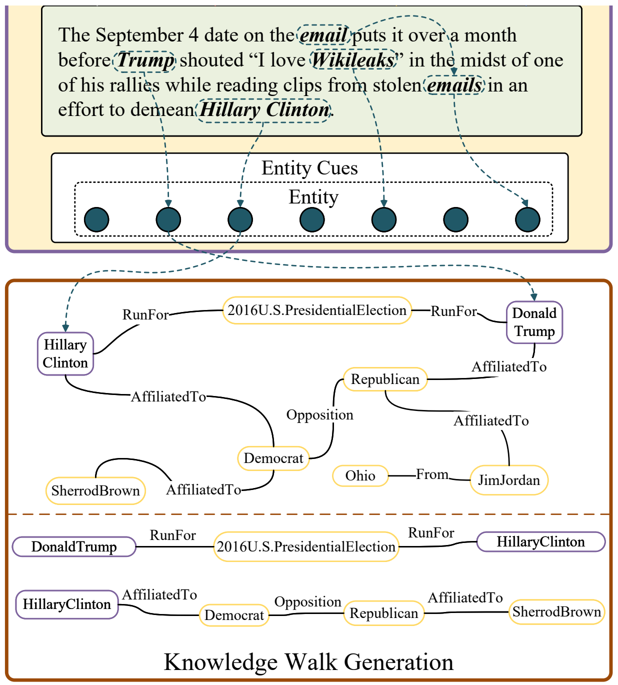
$$\tilde{T} = MultiHead(Q, K, V) \quad (8)$$

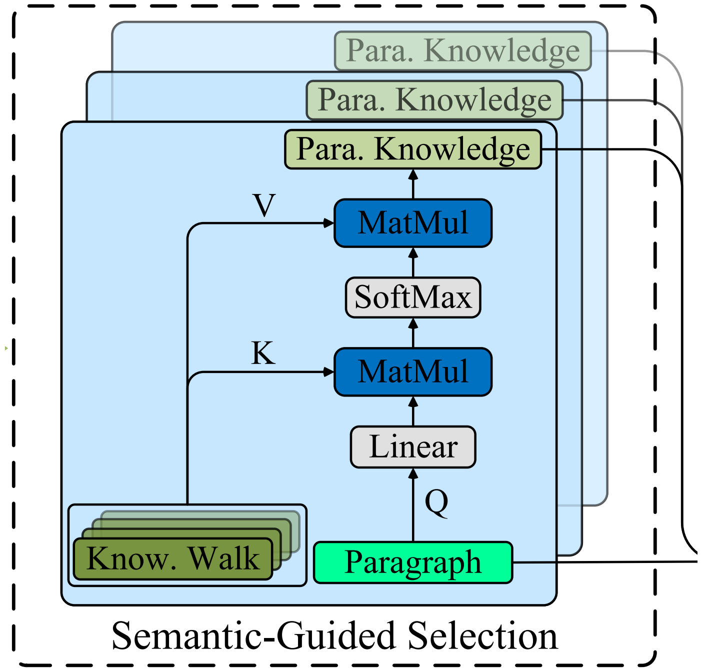
Where *Q=K=V=T*

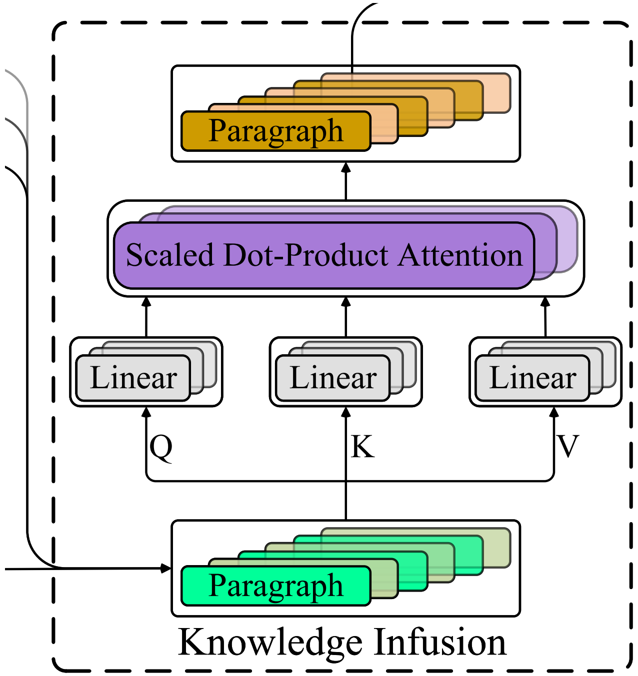
$$\tilde{T} = concat([\tilde{v_1^s}, \tilde{v_1^p}, ..., \tilde{v_n^s}, \tilde{v_n^p}])$$

$\{\tilde{v_i^s}\}_{i=1}^n$ Language representation enhanced with multi-hop knowledge reasoning

Quick recap







Generate knowledge walks
with
Biased random walks

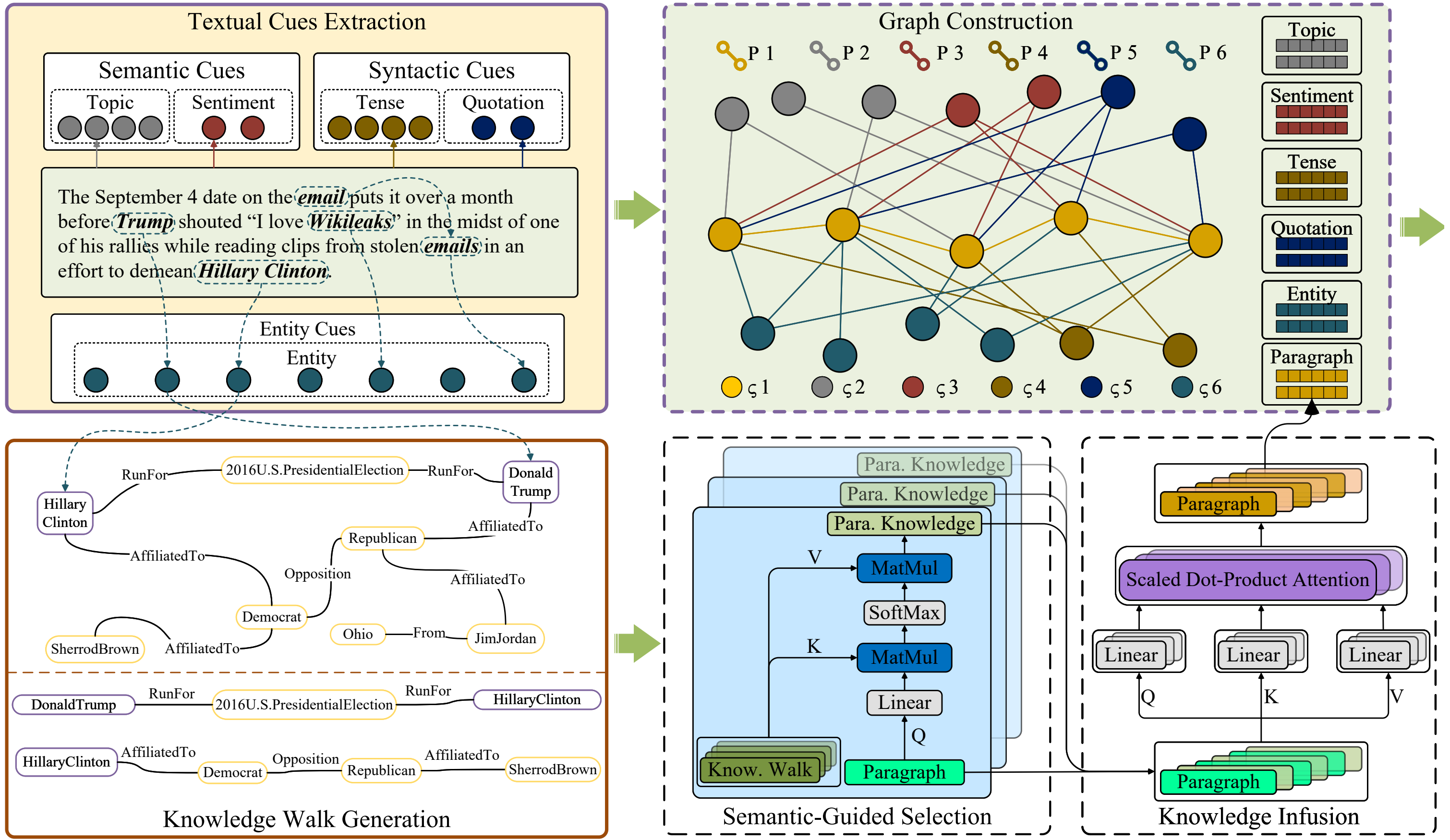
Select knowledge walks
with
Semantic relevance

Infuse knowledge walks
with
Multi-head attention

4/5 document graph construction

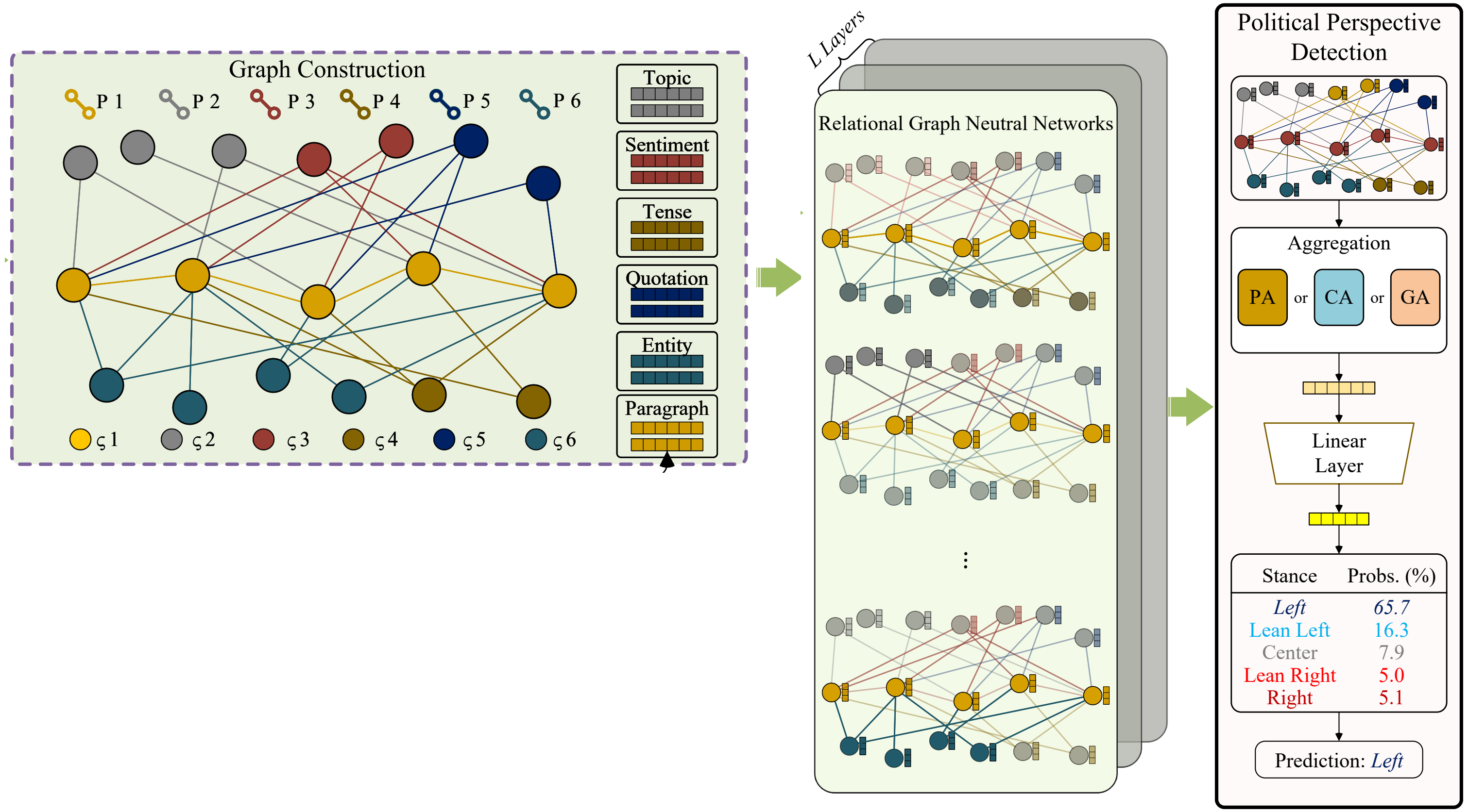
- Leverage ”textual cues”
  - External labels generated by tools such as NLTK and LDA
  - Topic
  - Sentiment
  - Tense
  - Quotation

- That’s beside the point



5/5 representation learning

- Relational GNNs



1/3 experiments

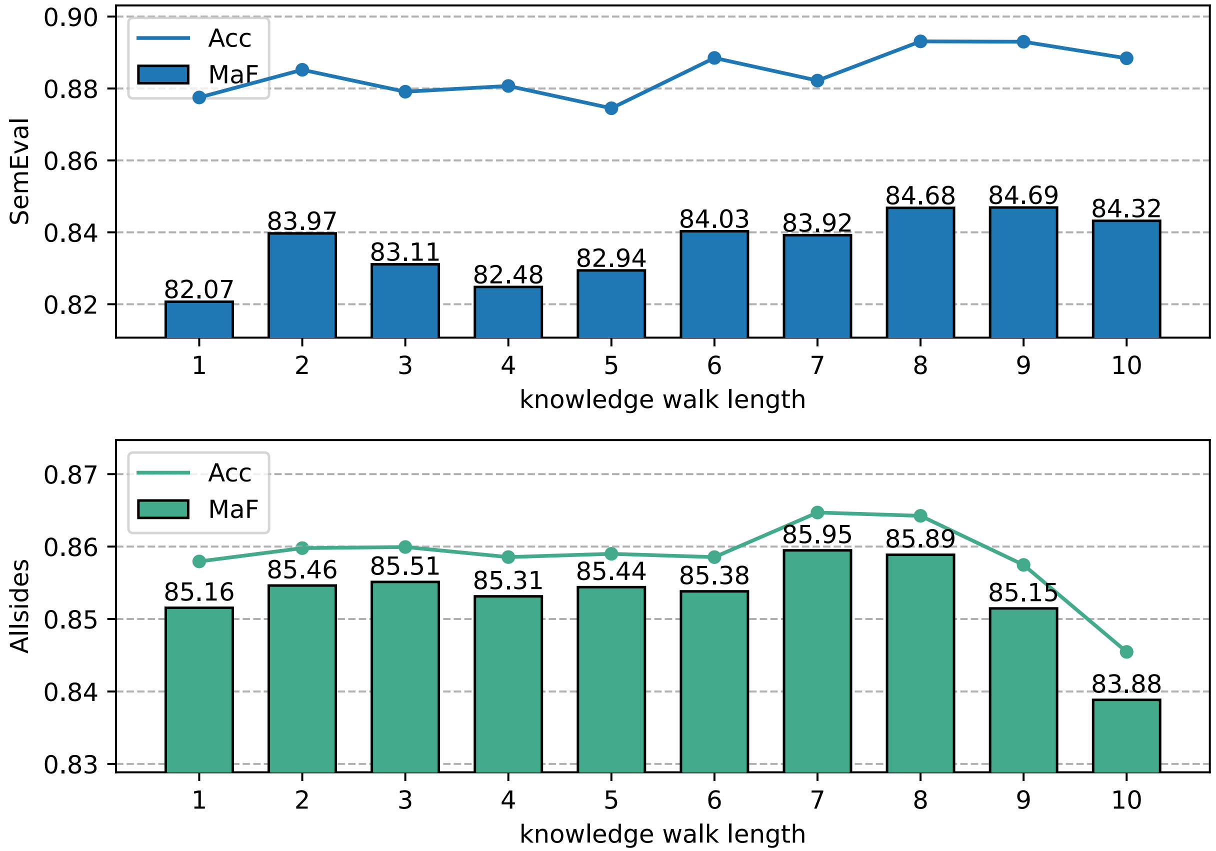
- KCD is good
- Knowledge walks are essential

Method	Setting	SemEval Acc	SemEval MaF	AllSides Acc	AllSides MaF
CNN	GloVe	79.63	N/A	N/A	N/A
	ELMo	84.04	N/A	N/A	N/A
HLSTM	GloVe	81.58	N/A	N/A	N/A
	ELMo	83.28	N/A	N/A	N/A
	Embed	81.71	N/A	76.45	74.95
	Output	81.25	N/A	76.66	75.39
Text Model	Word2Vec	70.27	39.37	48.58	34.33
	GloVe	80.71	63.64	71.01	69.81
	ELMo	86.78	80.46	81.97	81.15
	BERT	86.92	80.71	82.46	81.77
	RoBERTa	87.08	81.34	85.35	84.85
MAN	GloVe	81.58	79.29	78.29	76.96
	ELMo	84.66	83.09	81.41	80.44
	Ensemble	86.21	84.33	85.00	84.25
KGAP	GRGCN	89.56	84.94	86.02	85.52
KCD	GA	88.52	84.13	86.02	85.53
	CA	89.77	85.26	81.28	80.39
	PA	**90.87**	**87.87**	**87.38**	**87.14**
KCD (PA)	- w/o TC	88.22	83.53	86.08	85.58
	- w/o KW	87.29	81.77	85.51	85.00

Table 2: Political perspective detection performance on two benchmark datasets. Acc and MaF denote accuracy and macro-averaged F1-score. N/A indicates that the result is not reported in previous works. TC and KW indicate textual cues and knowledge walks respectively.

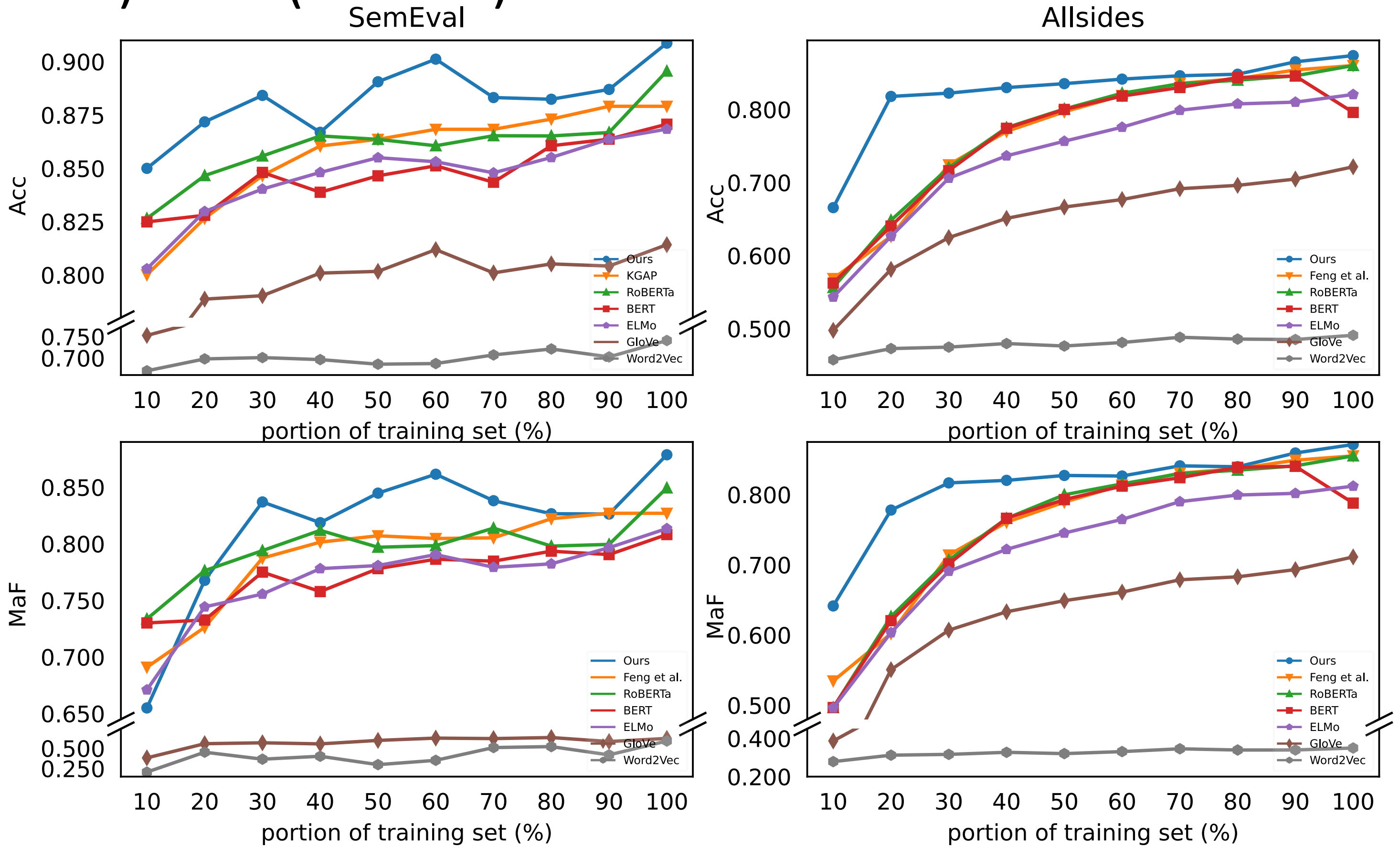
2/3 experiments

- Robust to knowledge walk length



3/3 experiments

- Improve data efficiency
  - size(SemEval) < size(Allsides)



Future Work

- Better understand the contribution of KGs in NLP tasks
  - Is KG important in Task A?
  - How do we assess the importance of KG in NLP task A?
  - How do we enable information exchange between LMs and KGs?
  - How does information flow between text and KGs?
  - …

- Hopefully …

Thank you

Aug 18, 2022

Shangbin Feng
University of Washington
shangbin@cs.washington.edu
bunsenfeng.github.io