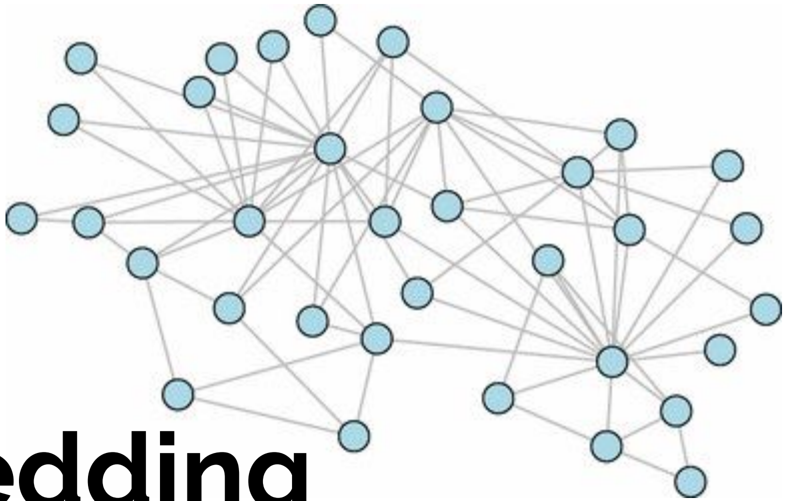


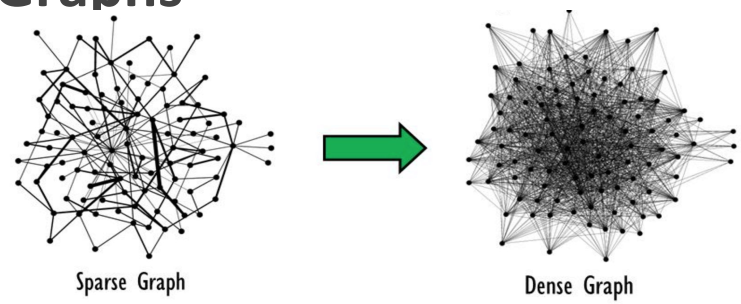
Neural Graph Embedding Methods for NLP

Shikhar Vashishth
Indian Institute of Science

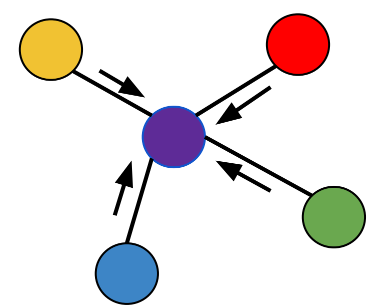
Advised by:
Dr. Partha Talukdar (IISc)
Prof. Chiranjib Bhattacharyya (IISc)
Dr. Manaal Faruqui (Google Research)

Outline

- **Addressing Sparsity in Knowledge Graphs**
  - KG Canonicalization
  - Relation Extraction
  - Link Prediction



- **Exploiting Graph Convolutional Networks in NLP**
  - Document Timestamping
  - Word Representation

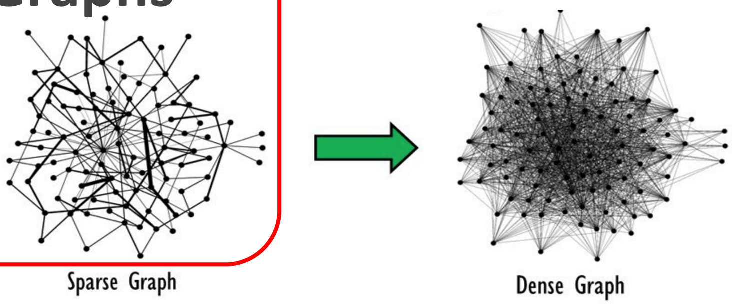


- **Addressing Limitations of Existing GCN Architectures**
  - Unrestricted Influence Neighborhood
  - Applicability to restricted class of graphs

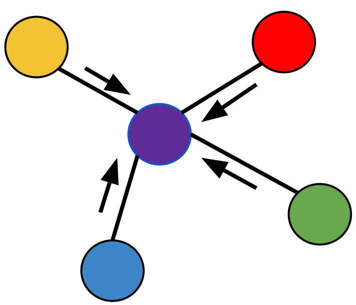
- Conclusion and Future work

Outline

- **Addressing Sparsity in Knowledge Graphs**
  - KG Canonicalization
  - Relation Extraction
  - Link Prediction



- **Exploiting Graph Convolutional Networks in NLP**
  - Document Timestamping
  - Word Representation

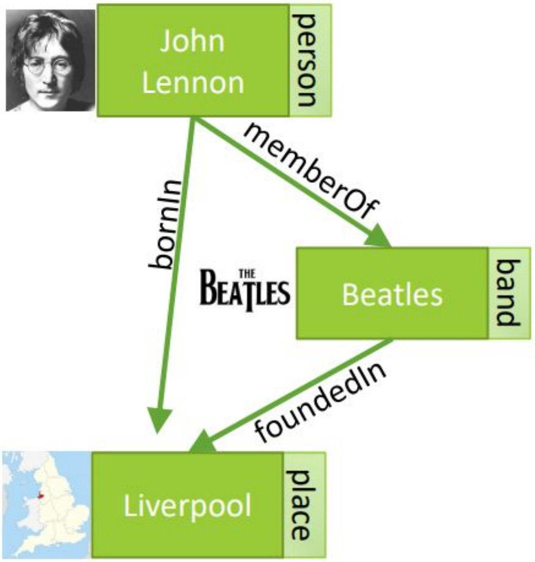


- **Addressing Limitations of Existing GCN Architectures**
  - Unrestricted Influence Neighborhood
  - Applicability to restricted class of graphs

- Conclusion and Future work

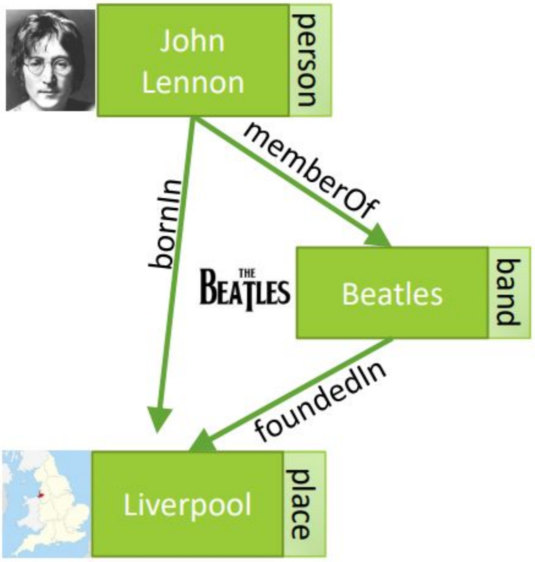
Knowledge Graphs

- Knowledge in graph form
- Nodes represent entities
- Edges represent relationships
- Examples: Freebase, Wikidata …



Knowledge Graphs

- Knowledge in graph form
- Nodes represent entities
- Edges represent relationships
- Examples: Freebase, Wikidata …
- **Use cases:**
  - Question Answering
  - Dialog systems
  - Web Search



Sparsity in Knowledge Graphs

- Most KGs are highly sparse
- For instance, NELL has 1.34 facts/entity
- Restricts applicability to real-world problems

Sparsity in Knowledge Graphs

- Most KGs are highly sparse
- For instance, NELL has 1.34 facts/entity
- Restricts applicability to real-world problems
- **Solutions:**
  - Identify and merge same entities (Canonicalization)
  - Extract more facts (Relation Extraction)
  - Infer new facts (Link Prediction)

Knowledge Graph Canonicalization

Noun Phrases	**Relation phrases:**
Barack Obama	born_in
Obama	took_birth_in
George Bush	is_employed_in
New York City	works_for
NYC	capital_of

Knowledge Graph Canonicalization

Noun Phrases

- Barack Obama
- Obama

- George Bush

- New York City
- NYC

Relation phrases:

- born_in
- took_birth_in

- is_employed_in
- works_for

- capital_of

Open Knowledge Graphs

- KGs with entities and relations **not restricted to a defined set.**
- **Construction**: Automatically extracting (noun-phrase, relation-phrase, noun-phrase) from unstructured text.
  - Obama was the President of US. $\longrightarrow$ (Obama, was president of, US)
  - **Examples**: TextRunner, ReVerb, Ollie etc.

Issues with existing methods

- **Surface form not sufficient** for disambiguation
  - E.g. (US, America)
- **Manual feature engineering** is **expensive** and often **sub-optimal**
- **Sequentially canonicalizing** of noun and relation phrases can lead to **error propagation**

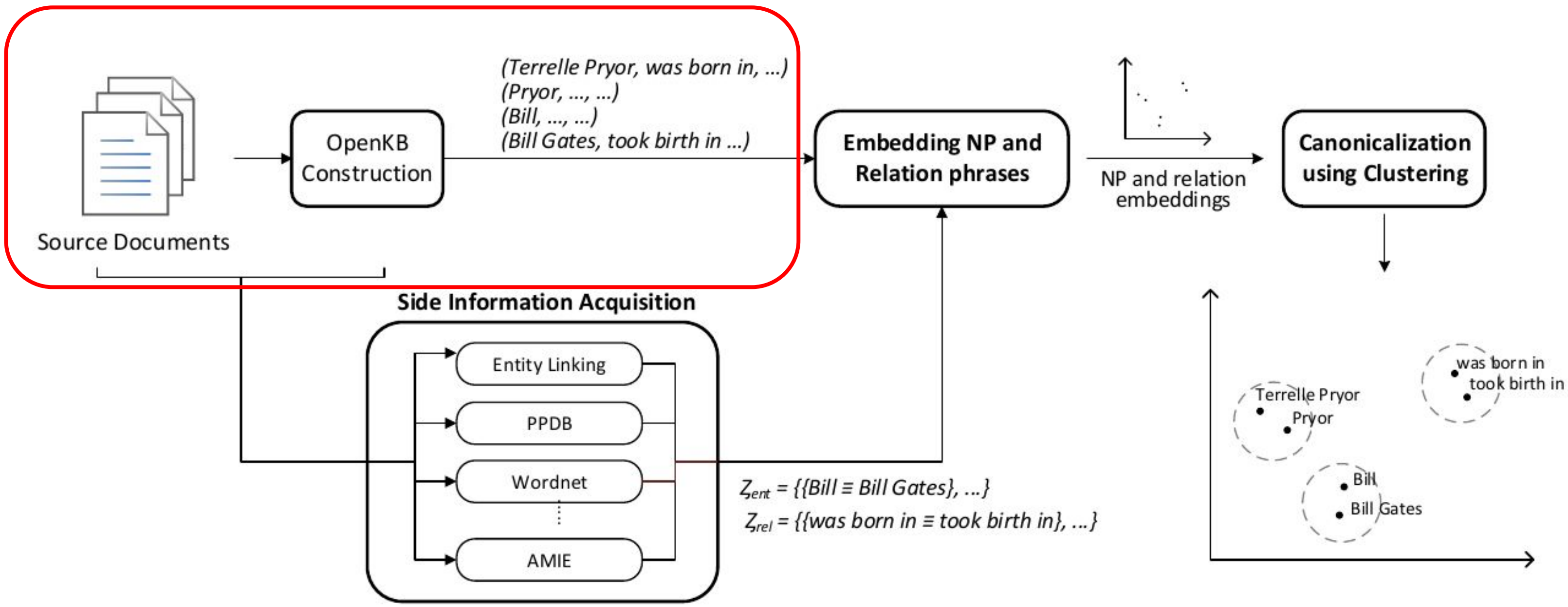
Contributions

- We propose **CESI**, a novel method for canonicalizing Open KBs using **learned embeddings.**

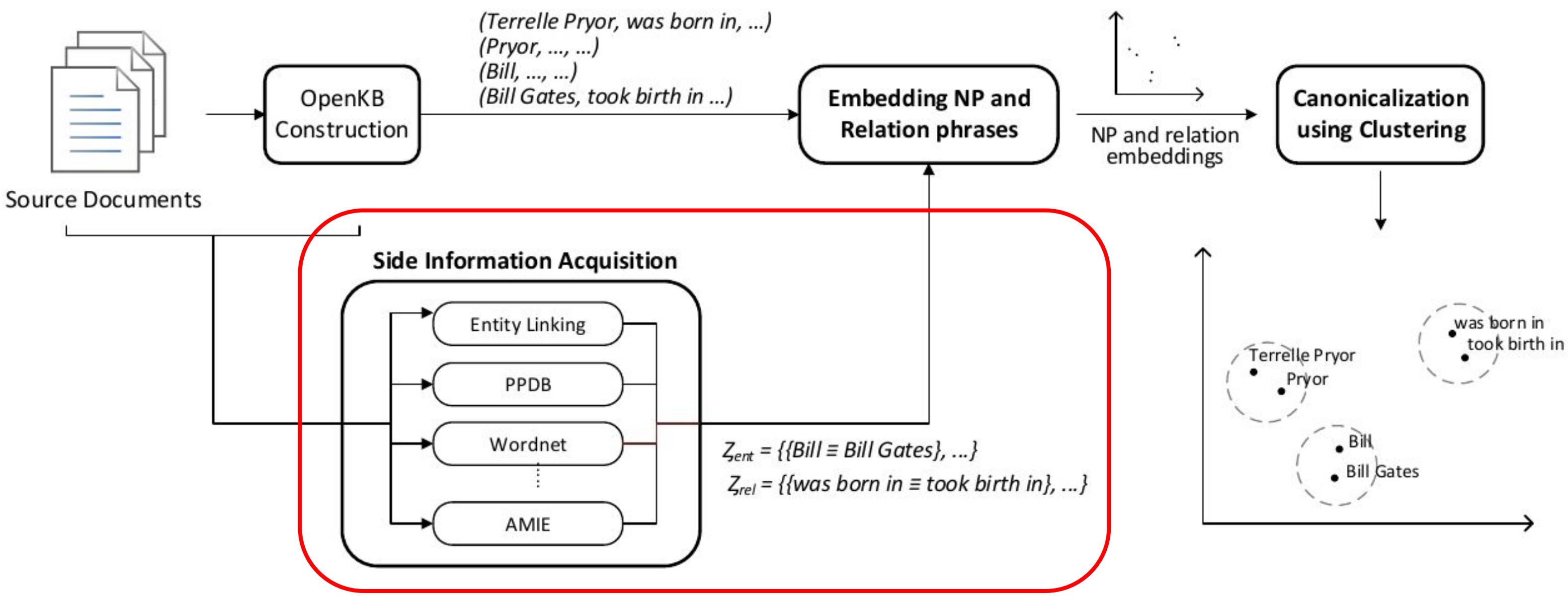
- CESI **jointly canonicalize** both noun phrase (NP) and relation phrase using **relevant side information.**

- Propose a new dataset, **ReVerb45K** for the task. It consists of **20x more NPs** than the previous biggest dataset.

CESI: Overview

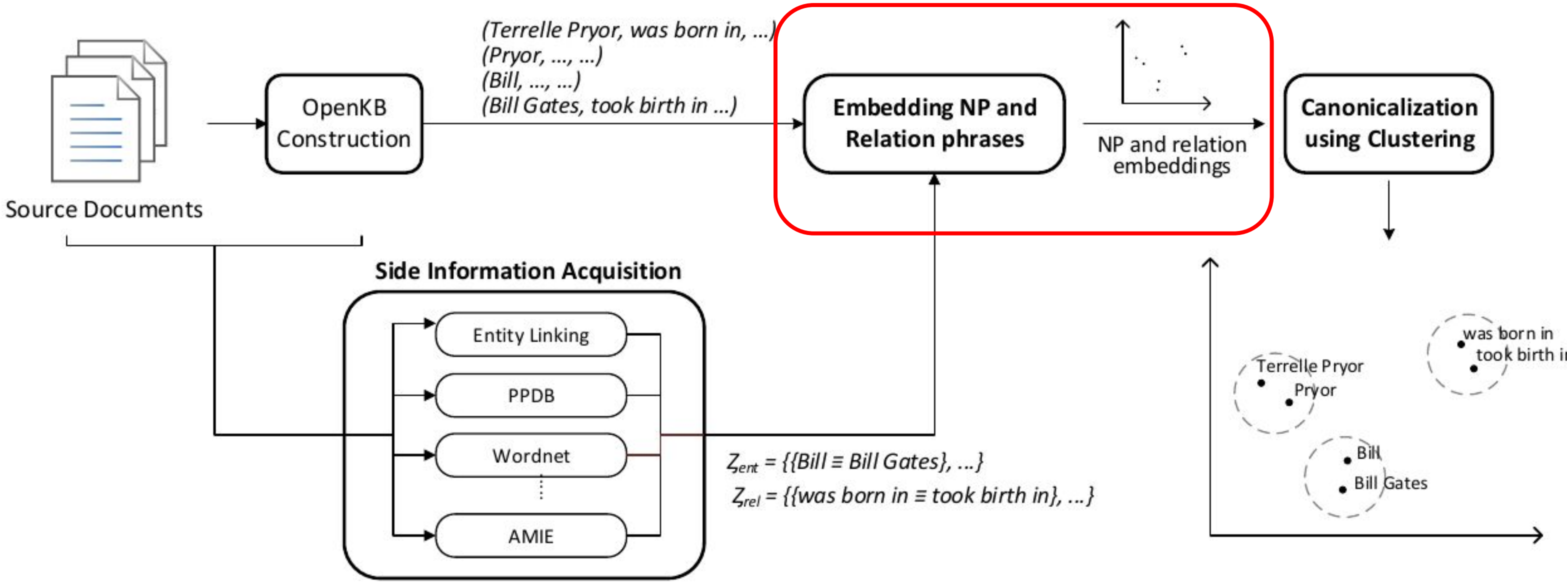


CESI: Overview



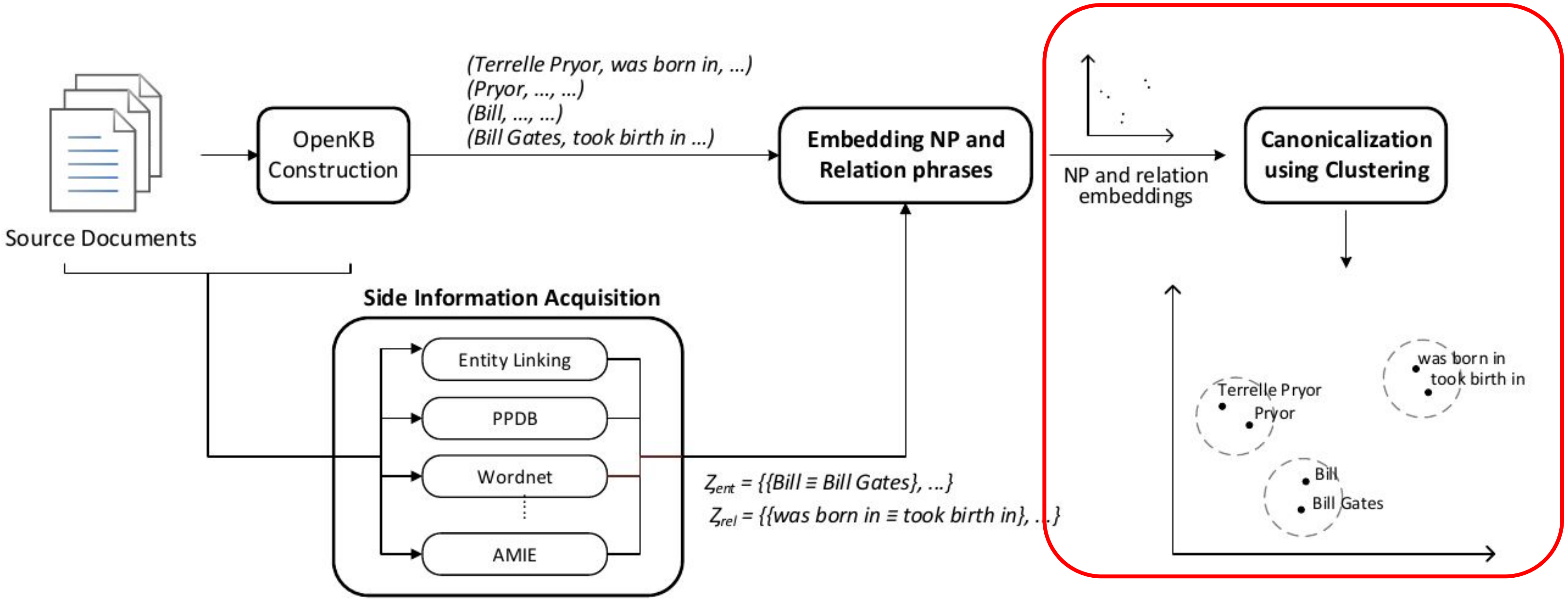
- **Side Information Acquisition:**
  - Gathers various noun and relation phrase side Information

CESI: Overview



- **Side Information Acquisition:**
  - Gathers various noun and relation phrase side Information
- **Embeddings Noun and relation phrases:**
  - Learns a specialized vector embeddings

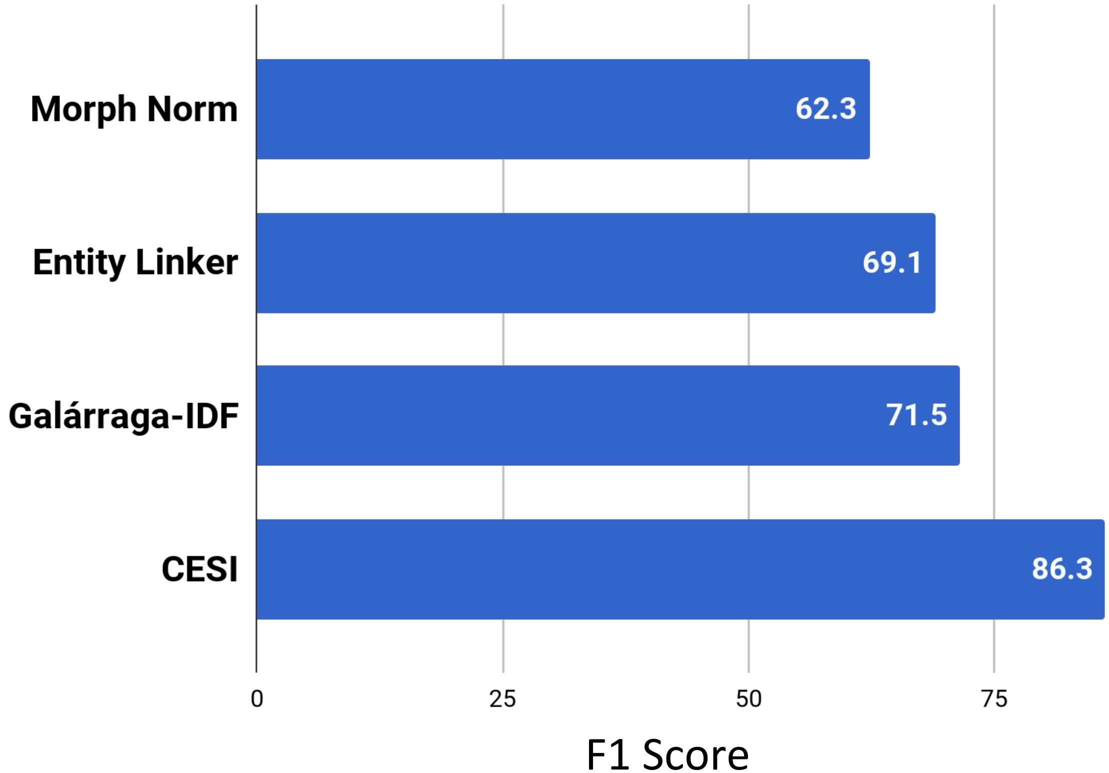
CESI: Overview



- **Side Information Acquisition:**
  - Gathers various noun and relation phrase side Information
- **Embeddings Noun and relation phrases:**
  - Learns a specialized vector embeddings
- **Clustering Embeddings and Canonicalization:**
  - Clusters embeddings and assigns a representative to cluster

Results: Noun Phrase Canonicalization

- CESI **outperforms** others in **noun phrase canonicalization**



Results: Relation Canonicalization

- CESI produces **more** and **better** relation **canonicalized clusters**

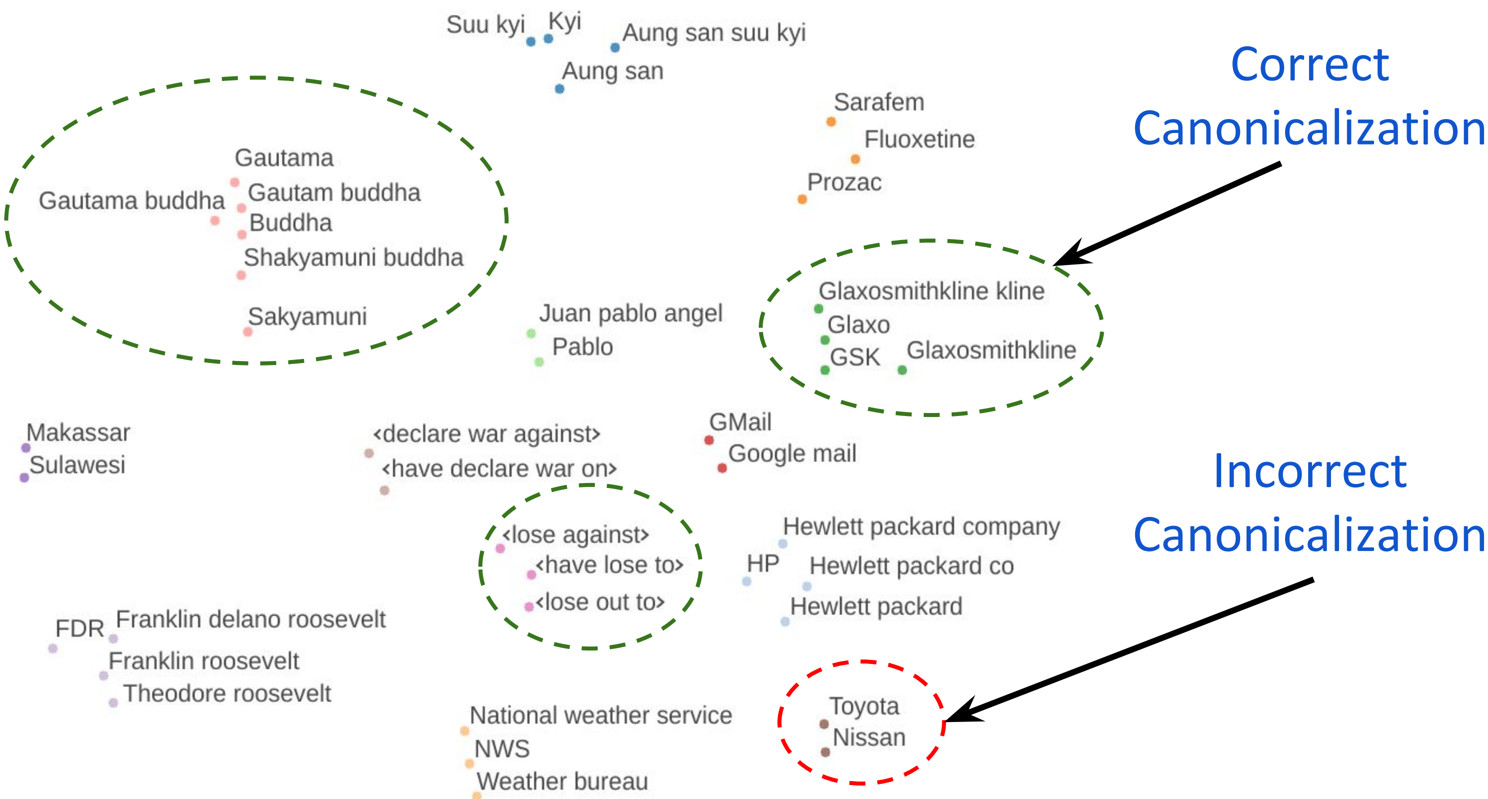
	Macro Precision	Micro Precision	Pairwise Precision	Induced Relation Clusters
		Base Dataset		
AMIE	42.8	63.6	43.0	7
CESI	**88.0**	**93.1**	**88.1**	**210**
		Ambiguous Dataset		
AMIE	55.8	64.6	23.4	46
CESI	**76.0**	**91.9**	**80.9**	**952**
		ReVerb45K		
AMIE	69.3	84.2	66.2	51
CESI	**77.3**	**87.8**	**72.6**	**2116**

Results: Qualitative Evaluation (t-sne)

Correct Canonicalization

Incorrect Canonicalization

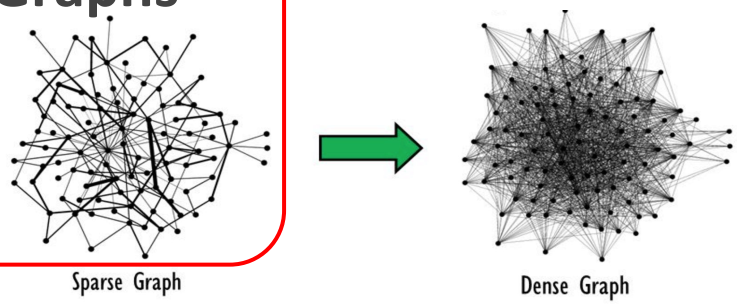
Results: Qualitative Evaluation (t-sne)



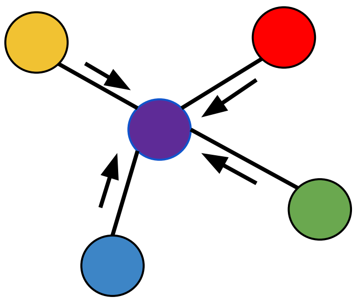
Shikhar Vashishth, Prince Jain, and Partha Talukdar.
"CESI: Canonicalizing Open Knowledge Bases using Embeddings and Side Information".
*In Proceedings of the **World Wide Web Conference (WWW), 2018**.*

Outline

- **Addressing Sparsity in Knowledge Graphs**
  - ✓ KG Canonicalization
  - ○ Relation Extraction
  - ○ Link Prediction



- **Exploiting Graph Convolutional Networks in NLP**
  - ○ Document Timestamping
  - ○ Word Representation



- **Addressing Limitations of Existing GCN Architectures**
  - ○ Unrestricted Influence Neighborhood
  - ○ Applicability to restricted class of graphs

- Conclusion and Future work

Relation Extraction

- Identify **relation** between entities.

- Google was **founded** in California in 1998.
  - **Founding-year** (Google, 1998)
  - **Founding-location** (Google, California)

Relation Extraction

- Identify **relation** between entities.

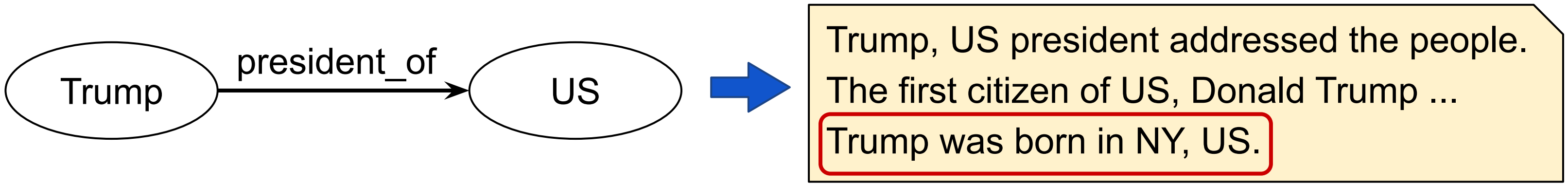
- Google was **founded** in California in 1998.
  - **Founding-year** (Google, 1998)
  - **Founding-location** (Google, California)

- **Used for**
  - Knowledge base population
  - Biomedical knowledge discovery
  - Question answering

Distant Supervision

- Alleviates the problem of lack of annotated data.

- **Distant Supervision (DS) assumption:** [Mintz et al., 2009]
  "If two entities have a relationship in a KB, then all sentences mentioning the entities express the same relation"



Motivation

- **KGs** contain information which **can improve RE**
  - Limiting supervision from KG to dataset creation

- **Dependency tree based features** have been found **relevant for RE** [Mintz et al. 2009]
  - Instead of defining hand-crafted features can employ Graph Convolutional Networks (GCNs).

Contributions

- Propose **RESIDE**, a novel method which utilizes additional supervision from KB in a principled manner for improving distant supervised RE.

- RESIDE uses **GCNs for modeling syntactic information** and performs competitively even with limited side information.

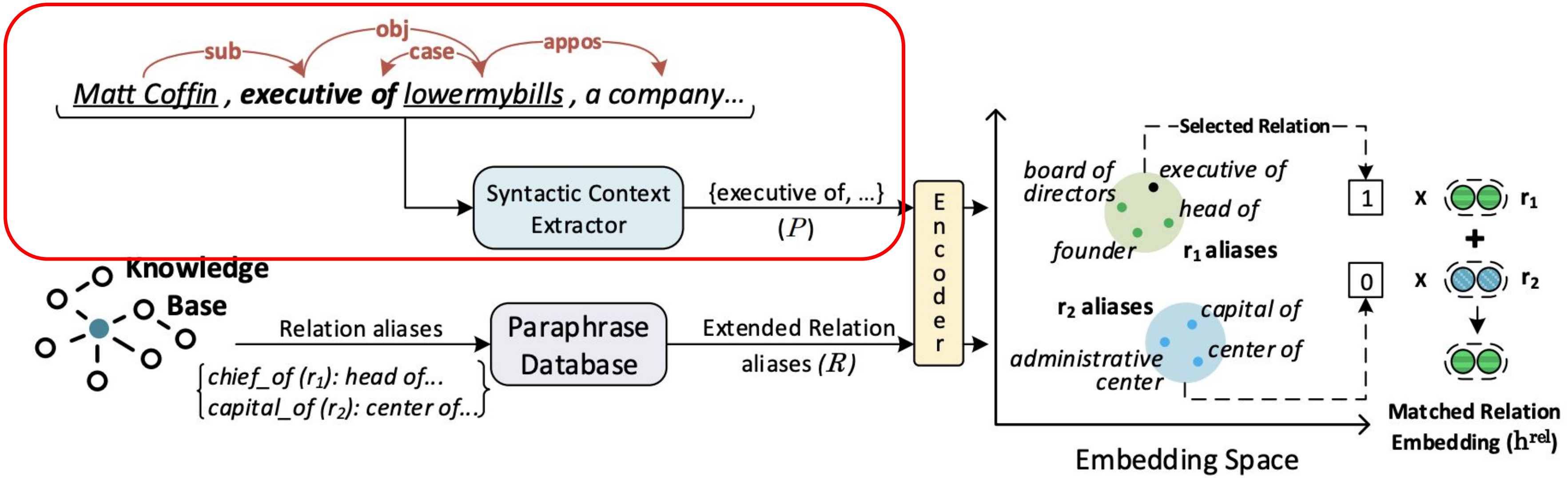
RESIDE: Side Information

- **Entity Type Information:**
  - All relations are constrained by the entity types
  - president_of(X, Y) => X = Person Y = Country

RESIDE: Side Information

- **Entity Type Information:**
  - All relations are constrained by the entity types
  - president_of(X, Y) => X = Person Y = Country

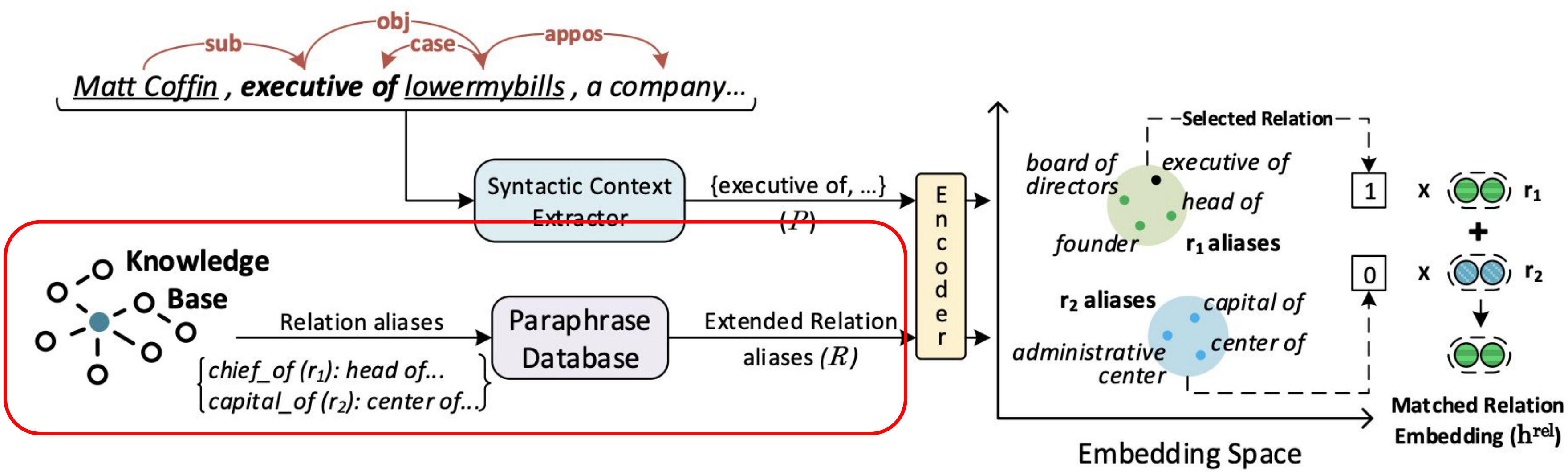
- **Relation Alias Information:**
  - Utilize relation aliases provided by KGs.



RESIDE: Side Information

- **Entity Type Information:**
  - All relations are constrained by the entity types
  - president_of(X, Y) => X = Person Y = Country

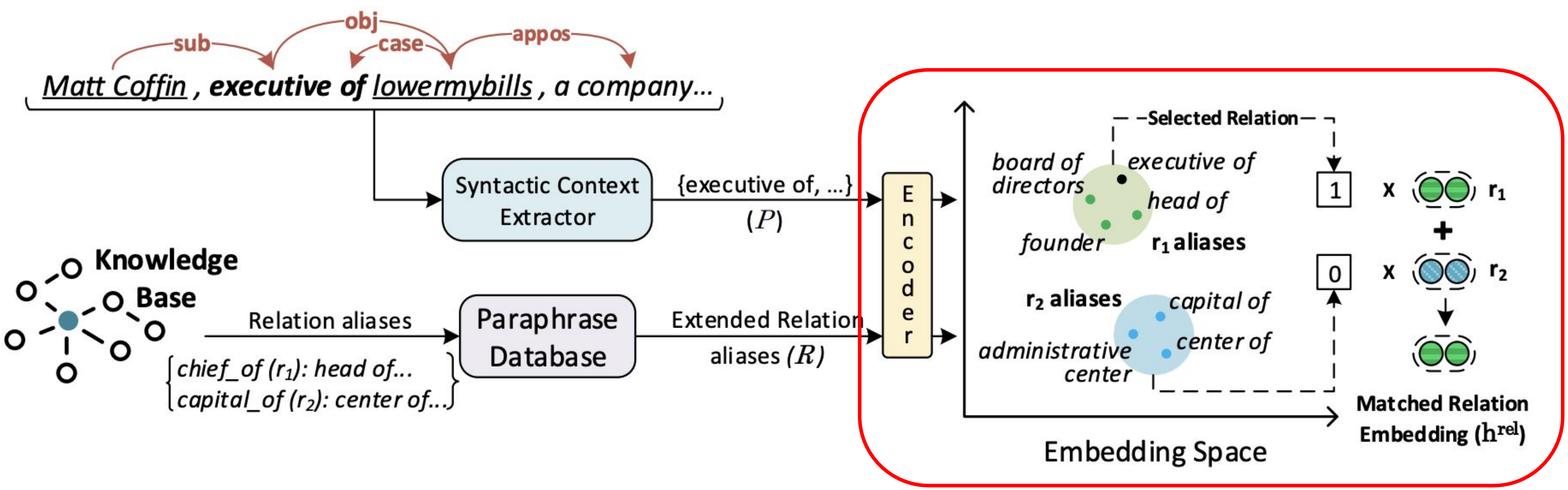
- **Relation Alias Information:**
  - Utilize relation aliases provided by KGs.



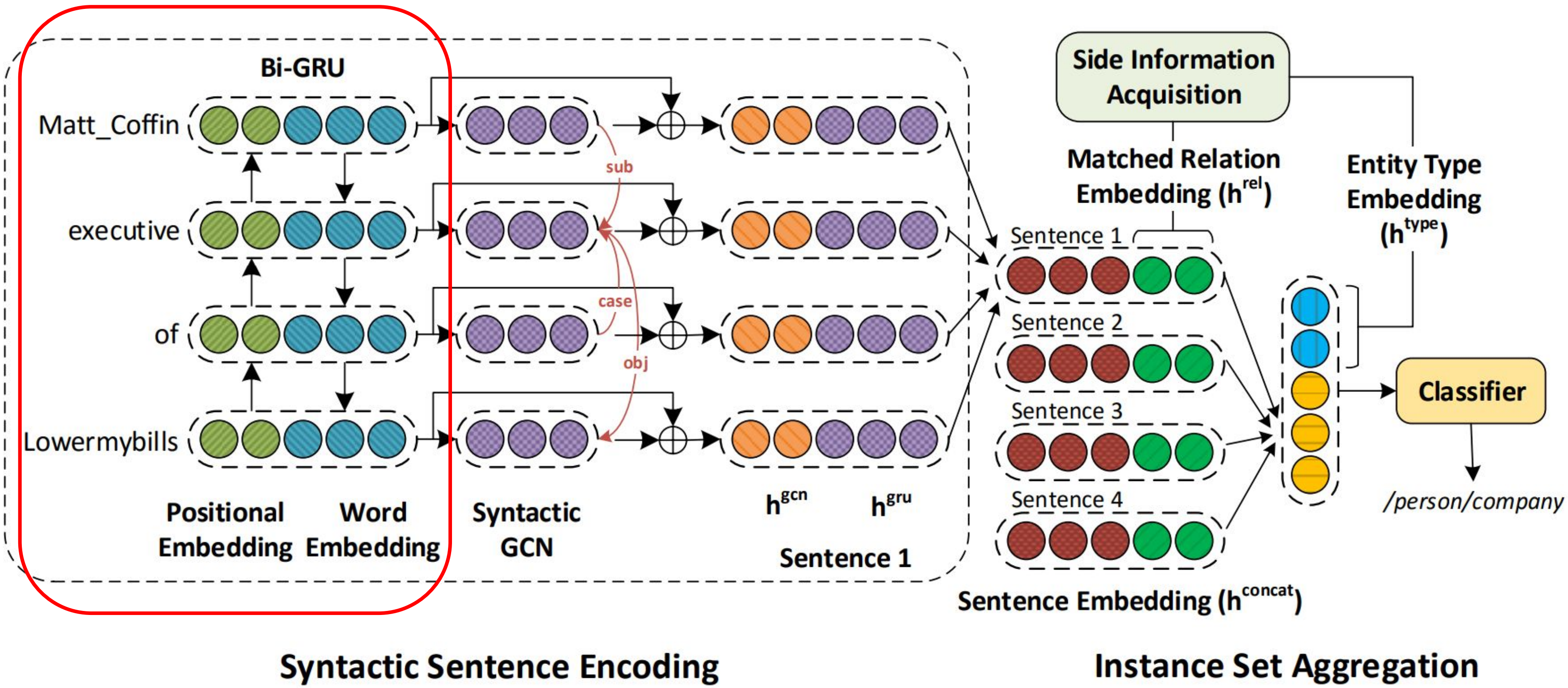
RESIDE: Side Information

- **Entity Type Information:**
  - All relations are constrained by the entity types
  - president_of(X, Y) => X = Person Y = Country

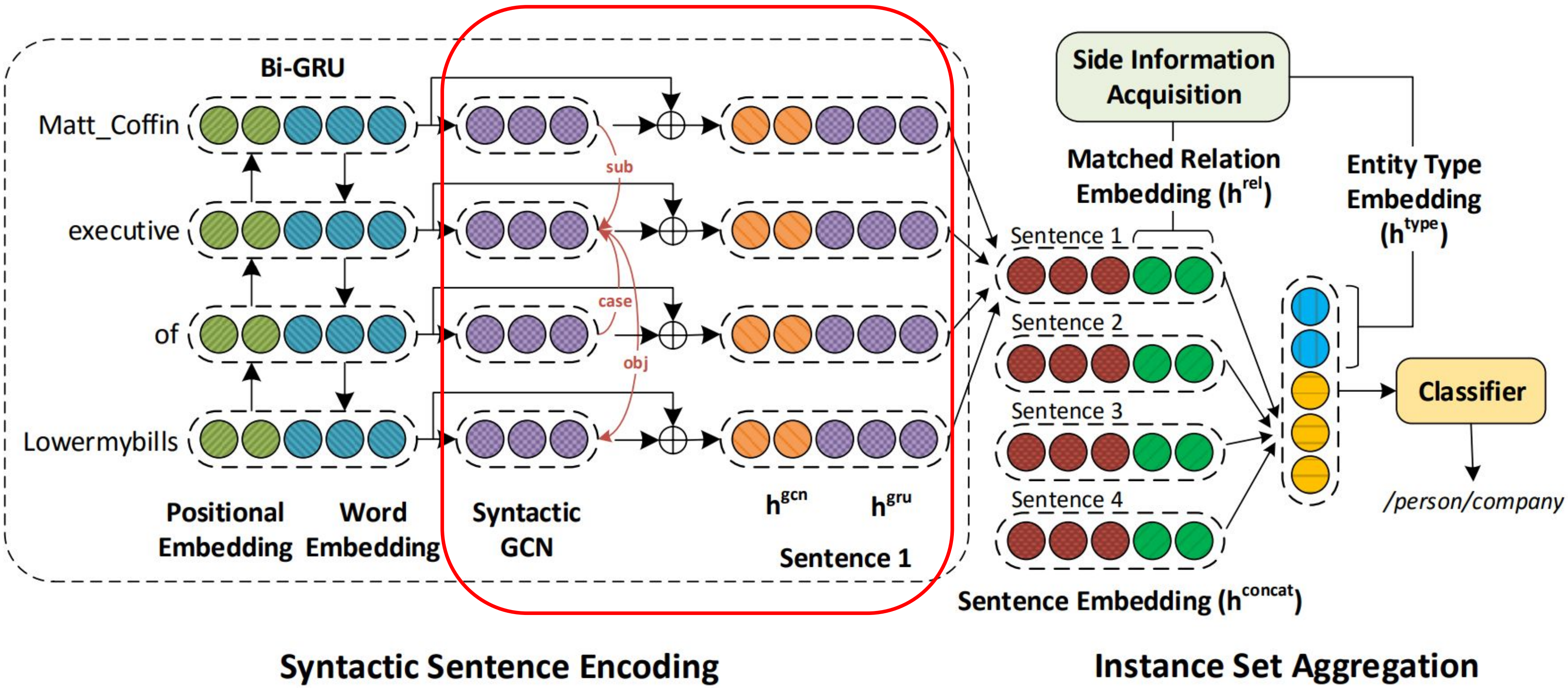
- **Relation Alias Information:**
  - Utilize relation aliases provided by KGs.



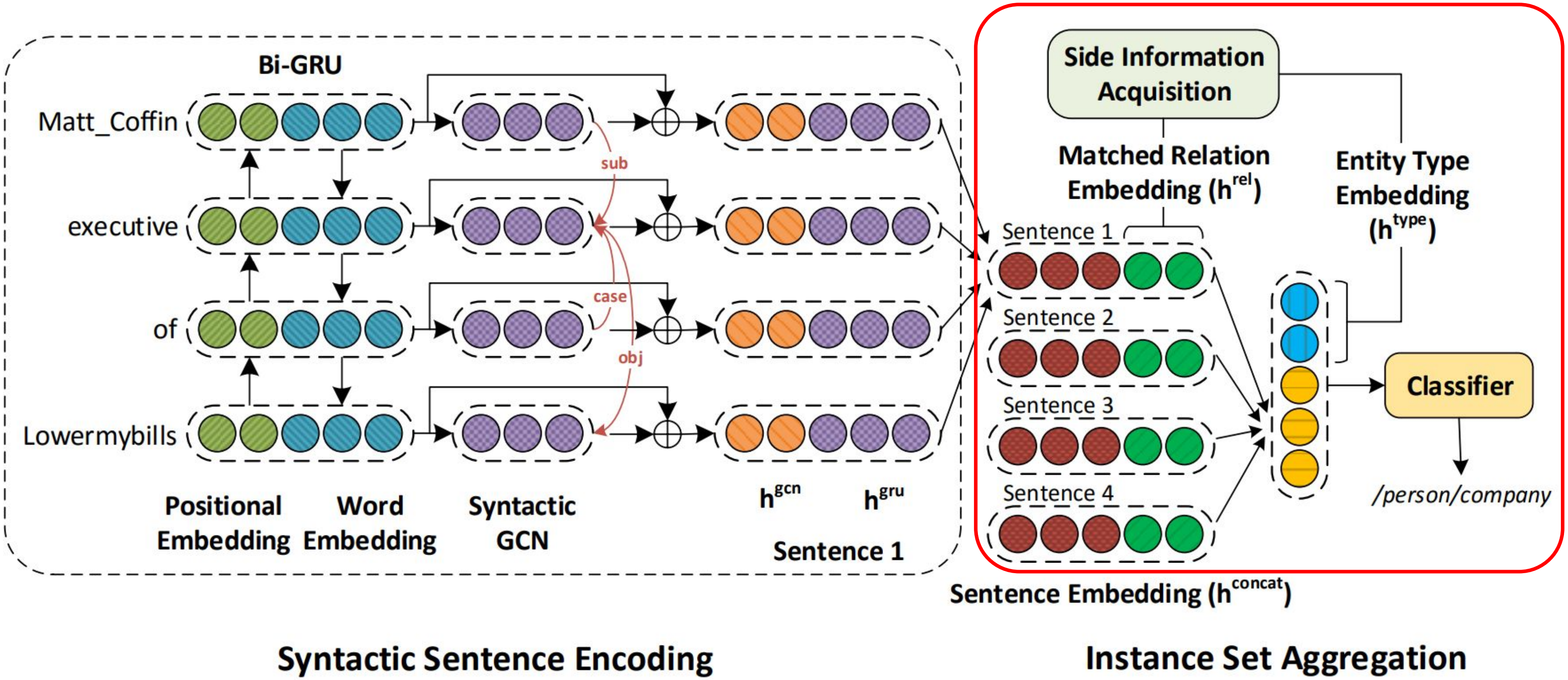
RESIDE: Architecture



RESIDE: Architecture

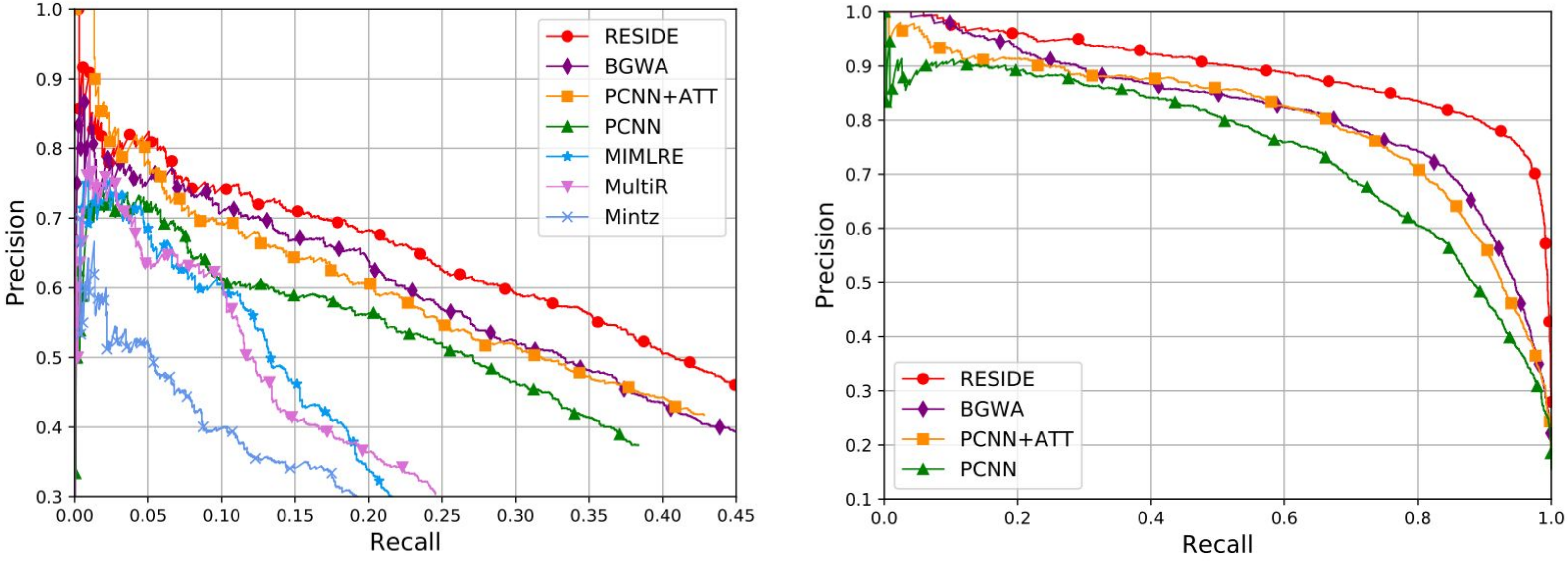


RESIDE: Architecture



Results: Performance Comparison

Comparison of Precision Recall curves

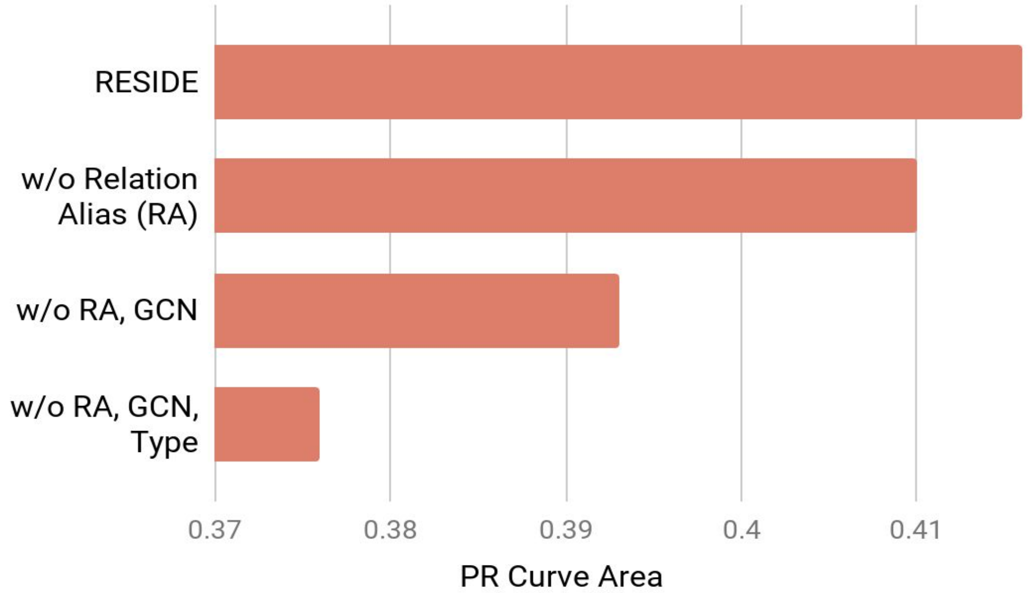


(a) Riedel dataset

(b) GIDS dataset

RESIDE achieves **higher precision** over the **entire recall range**.

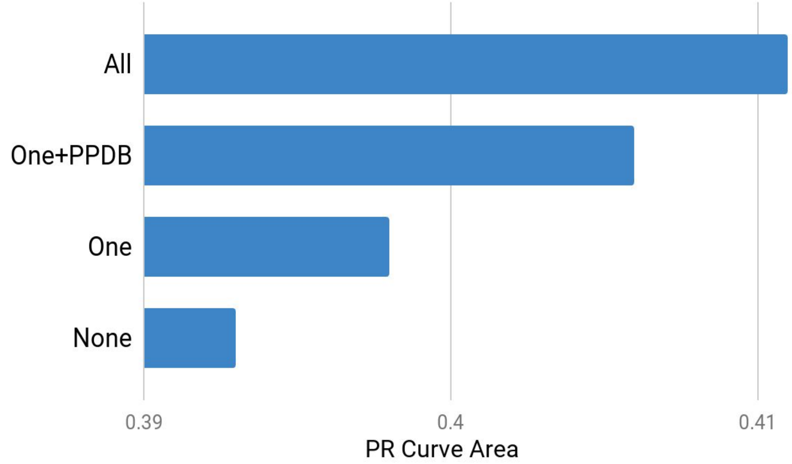
Results: Ablation Study



- Comparison of different **ablated version of RESIDE**
  - **Cumulatively removing** different **side information**
  - **Side information** helps **improve performance.**

Results: Effect of Relation Alias Information

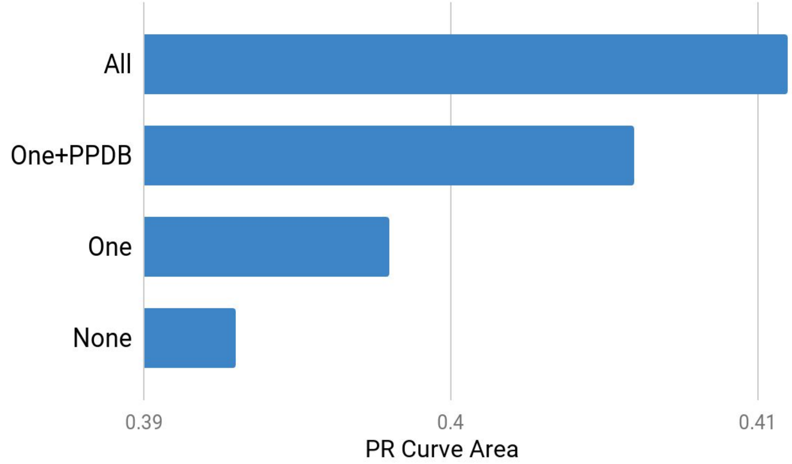
- **Performance on different settings:**
  - **None:** Relation aliases not available
  - **One:** Name of relations used as aliases
  - **One+PPDB:** Relation names extended using Paraphrase DB
  - **All:** Relation aliases from KG



RESIDE performs **comparable** with **limited side information.**

Results: Effect of Relation Alias Information

- **Performance on different settings:**
  - **None**: Relation aliases not available
  - **One**: Name of relations used as aliases
  - **One+PPDB**: Relation names extended using Paraphrase DB
  - **All**: Relation aliases from KG

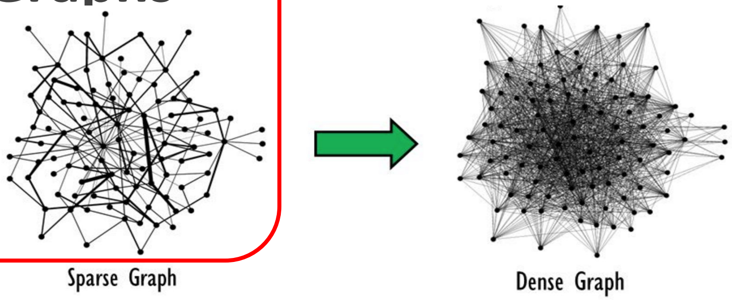


RESIDE performs **comparable** with **limited side information.**

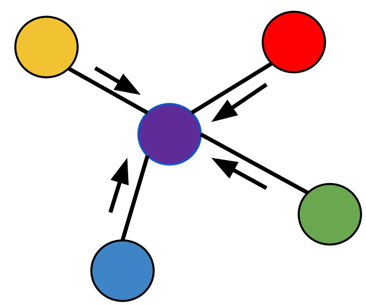
*S. Vashishth, R. Joshi, S. S. Prayaga, C. Bhattacharyya, and P. Talukdar. "**RESIDE: Improving Distantly-Supervised Neural Relation Extraction using Side Information**". In Proceedings of the **Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018**.*

Outline

- Addressing Sparsity in Knowledge Graphs
  - ✓ KG Canonicalization
  - ✓ Relation Extraction
  - ○ Link Prediction



- Exploiting Graph Convolutional Networks in NLP
  - Document Timestamping
  - Word Representation



- Addressing Limitations of Existing GCN Architectures
  - Unrestricted Influence Neighborhood
  - Applicability to restricted class of graphs

- Conclusion and Future work

Link Prediction

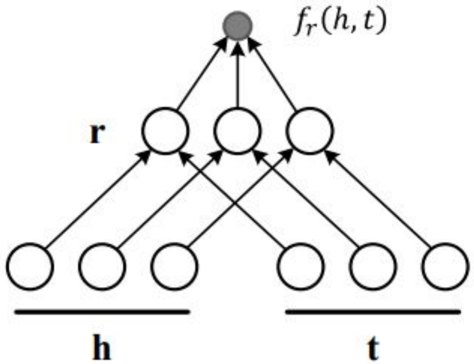
- **Definition:**
  Task of inferring missing facts based on known ones.

- **Example:**
  - (Barack Obama, spouse_of, Michelle Obama)
  - (Sasha Obama, child_of, Mitchell Obama)
  - (Sasha Obama, child_of, Barack Obama)

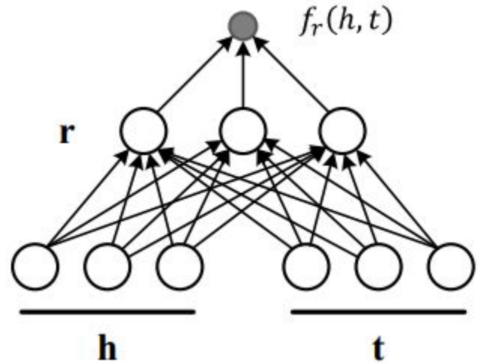
- **General technique** involves learning a representation for all entities and relations in KG.

Motivation

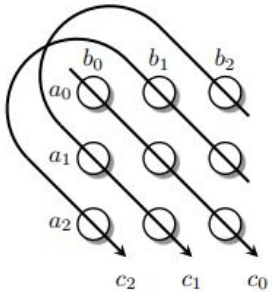
- **Increasing** interactions helps



DistMult.



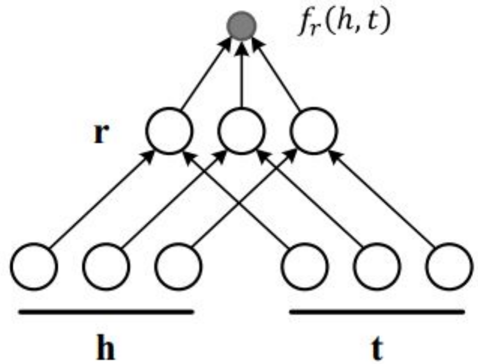
HolE.



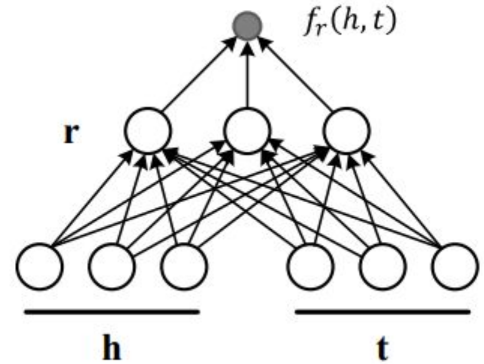
Circular Convolution

Motivation

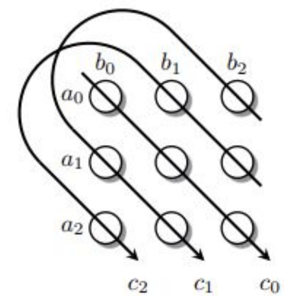
- **Increasing interactions helps**



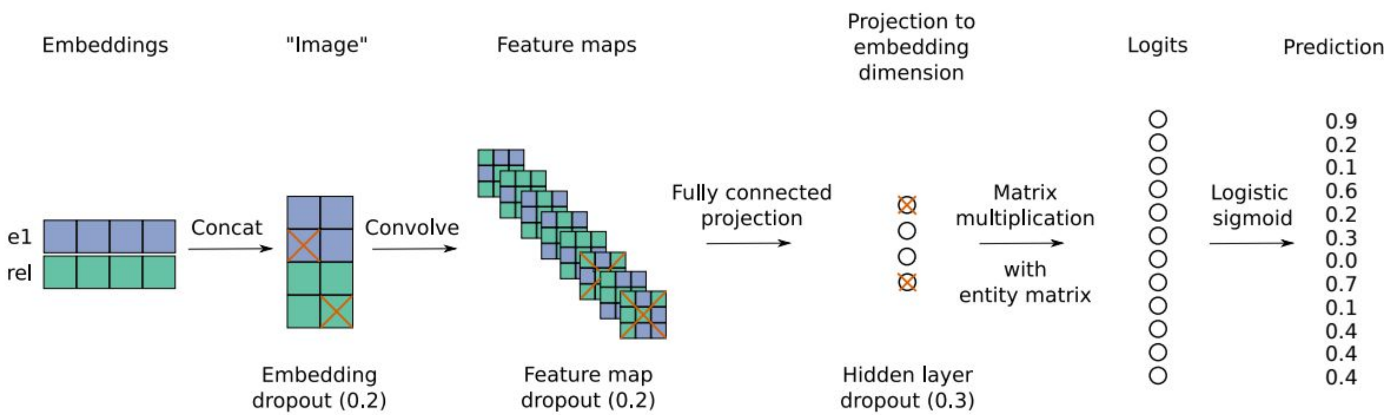
DistMult.



HolE.



Circular Convolution



ConvE

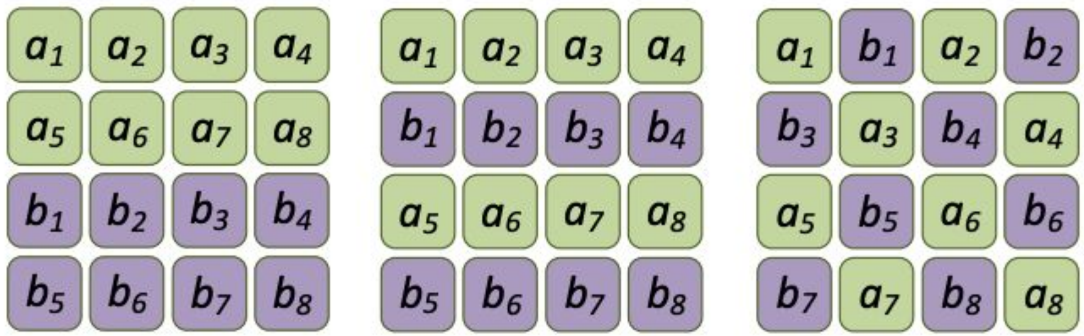
Contributions

- We propose **InteractE**, a method that augments the expressive power of ConvE through three key ideas – feature permutation, "checkered" feature reshaping, and circular convolution.

- Establish correlation between number of interactions and link prediction performance. Theoretically show that InteractE increases interactions compared to ConvE.

InteractE: Reshaping Function

- InteractE uses Chequer reshaping.



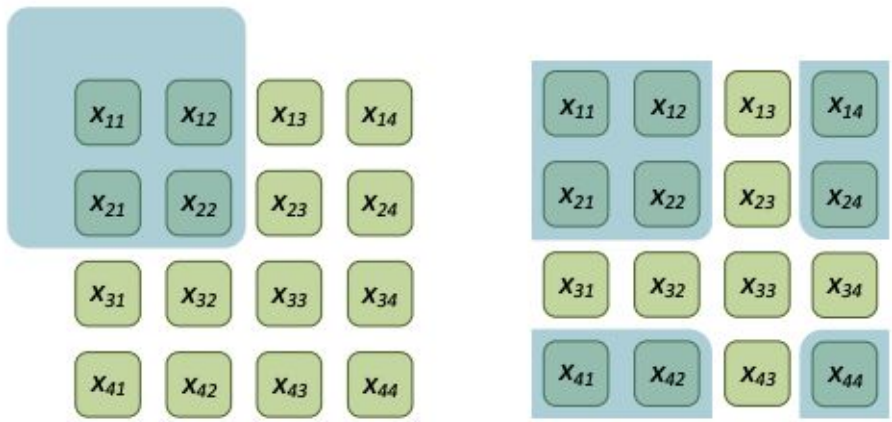
(a) Stack (b) Alternate (c) Chequer

Proposition 7.3. *For any kernel w of size k and for all reshaping functions $\phi : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{n \times n}$, the following statement holds:*

$$\mathcal{N}_{het}(\phi_{chk}, k) \geq \mathcal{N}_{het}(\phi, k)$$

InteractE: Reshaping Function

- InteractE uses Circular Convolution.

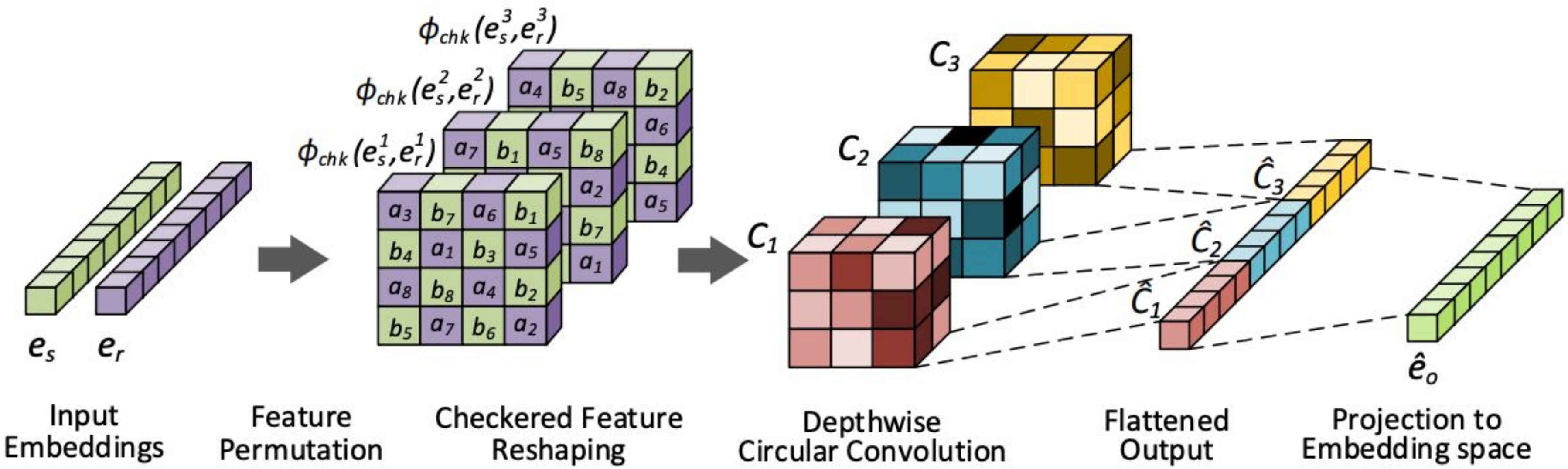


(a) Standard Convolution (b) Circular Convolution

Proposition 7.4. *Let Ω_0, $\Omega_c : \mathbb{R}^{n\times n} \rightarrow \mathbb{R}^{(n+p)\times(n+p)}$ denote zero padding and circular padding functions respectively, for some $p > 0$. Then for any reshaping function ϕ,*

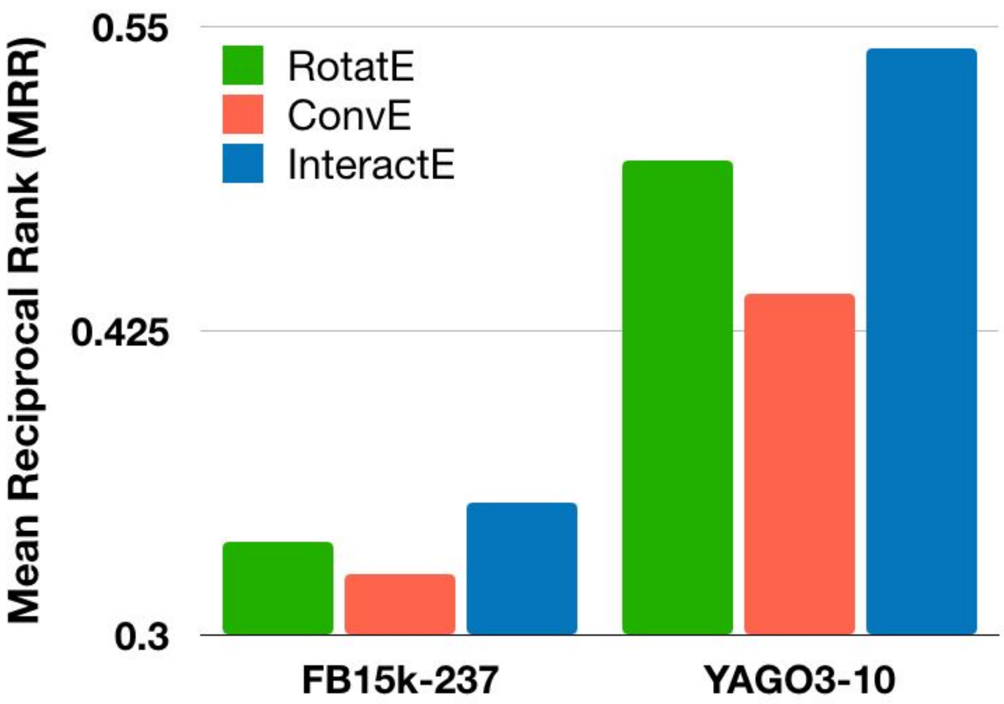
$$\mathcal{N}_{het}(\Omega_c(\phi), k) \geq \mathcal{N}_{het}(\Omega_0(\phi), k)$$

InteractE: Overview



InteractE: Results

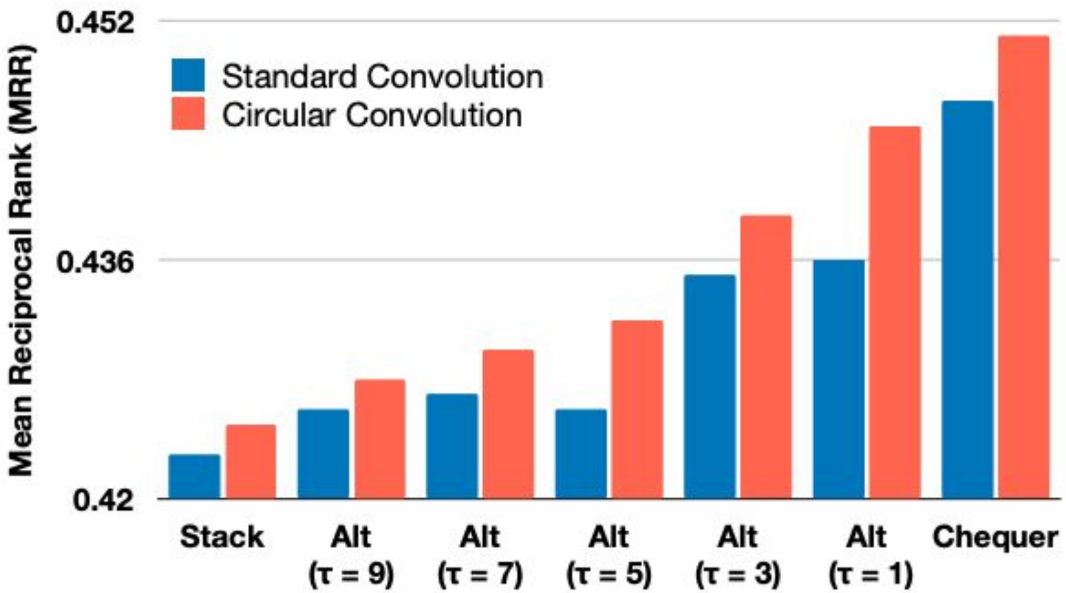
- **Performance Comparison** (MRR)



InteractE gives substantial improvement over **ConvE** and **RotatE** (SOTA)

InteractE: Results

- **Effect of Feature Reshaping function**

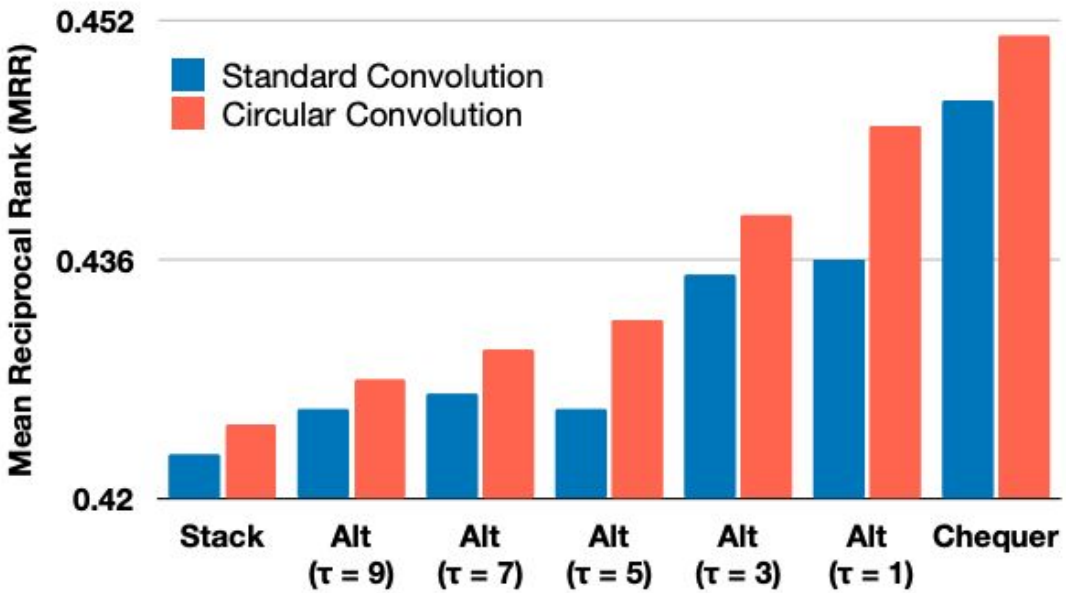


Empirical verification of our claim:

Increasing interactions improves link prediction

InteractE: Results

- **Effect of Feature Reshaping function**



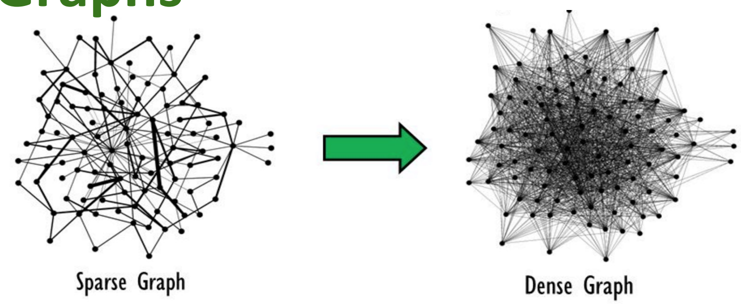
Empirical verification of our claim:

Increasing interactions improves link prediction

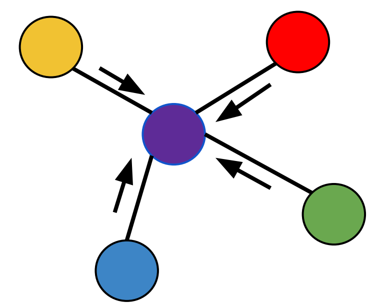
S. Vashishth, S. Sanyal*, V. Nitin, N. Agarwal, and P. Talukdar. "**InteractE: Improving Convolution-based Knowledge Graph Embeddings by Increasing Feature Interactions**". **[Under Submission]***

Outline

- **Addressing Sparsity in Knowledge Graphs**
  - ✓ KG Canonicalization
  - ✓ Relation Extraction
  - ✓ Link Prediction



- **Exploiting Graph Convolutional Networks in NLP**
  - Document Timestamping
  - Word Representation

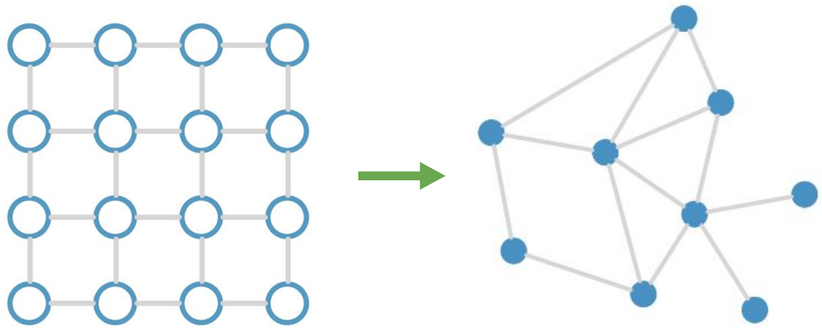


- **Addressing Limitations of Existing GCN Architectures**
  - Unrestricted Influence Neighborhood
  - Applicability to restricted class of graphs

- Conclusion and Future work

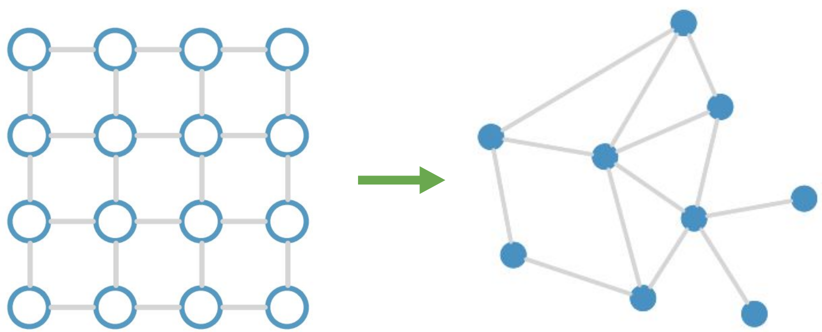
Graph Convolutional Networks (GCNs)

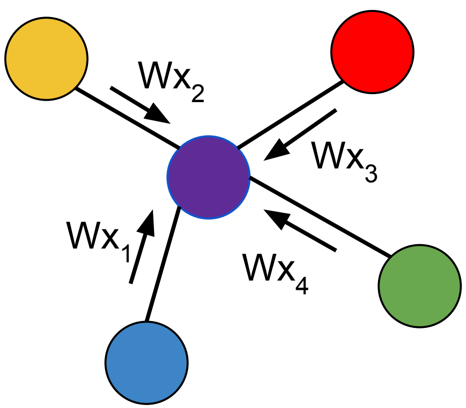
- **Generalization of CNNs over Graphs.**



Graph Convolutional Networks (GCNs)

- **Generalization of CNNs over Graphs.**





GCN First-order approximation

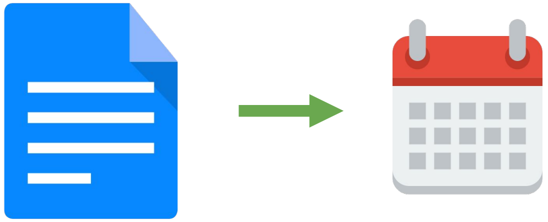
(Kipf et. al. 2016)

$$h_v = f\left(\frac{1}{|\mathcal{N}(v)|}\sum_{u\in\mathcal{N}(v)} W x_u + b\right), \quad \forall v \in \mathcal{V}.$$

Document Time-stamping

- **Problem:**

  Predicting the creation time of the document

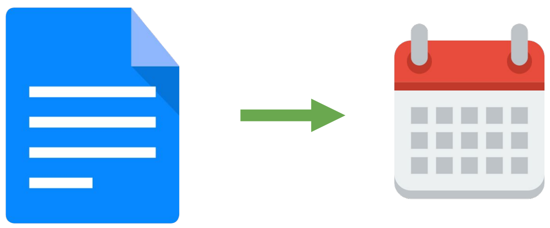


- **Applications:**
  - Information Extraction
  - Temporal reasoning
  - Text Summarization
  - Event detection ...

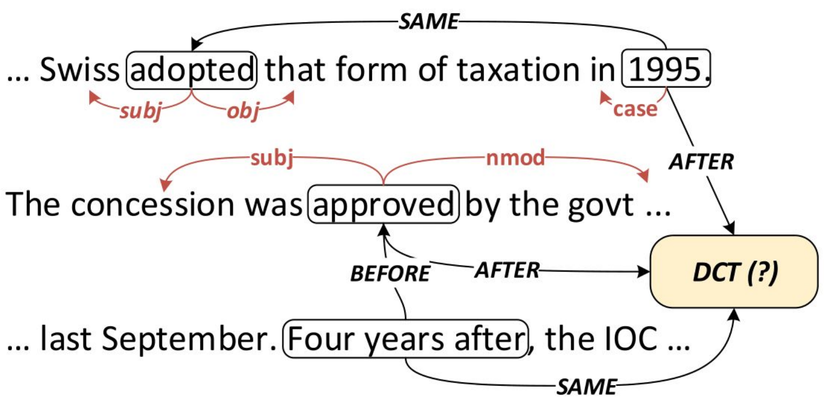
Document Time-stamping

- **Problem:**

  Predicting the creation time of the document



- **Applications:**
  - Information Extraction
  - Temporal reasoning
  - Text Summarization
  - Event detection ...

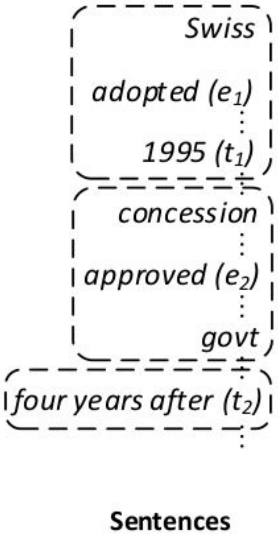


Contributions

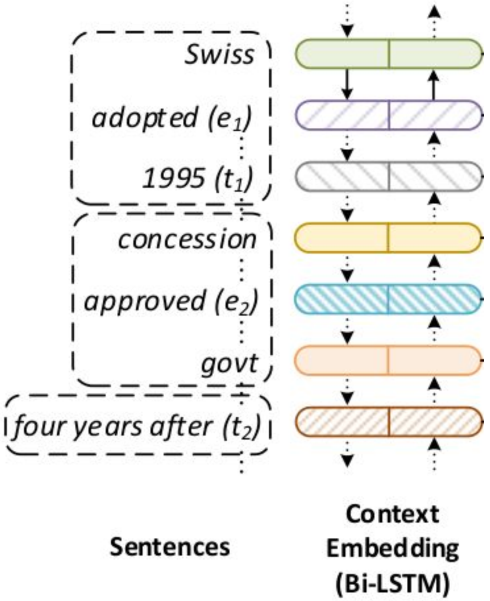
- We propose **NeuralDater**, a Graph Convolutional based approach for document dating. It is the first application of GCNs and neural network-based method for the problem.

- NeuralDater exploits **syntactic** as well as **temporal structure** of the document, all within a principled joint model.

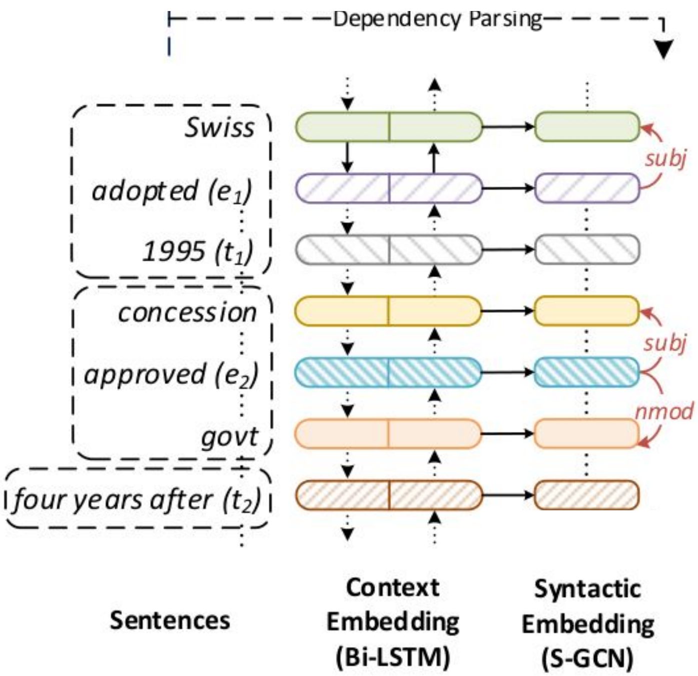
NeuralDater: Overview



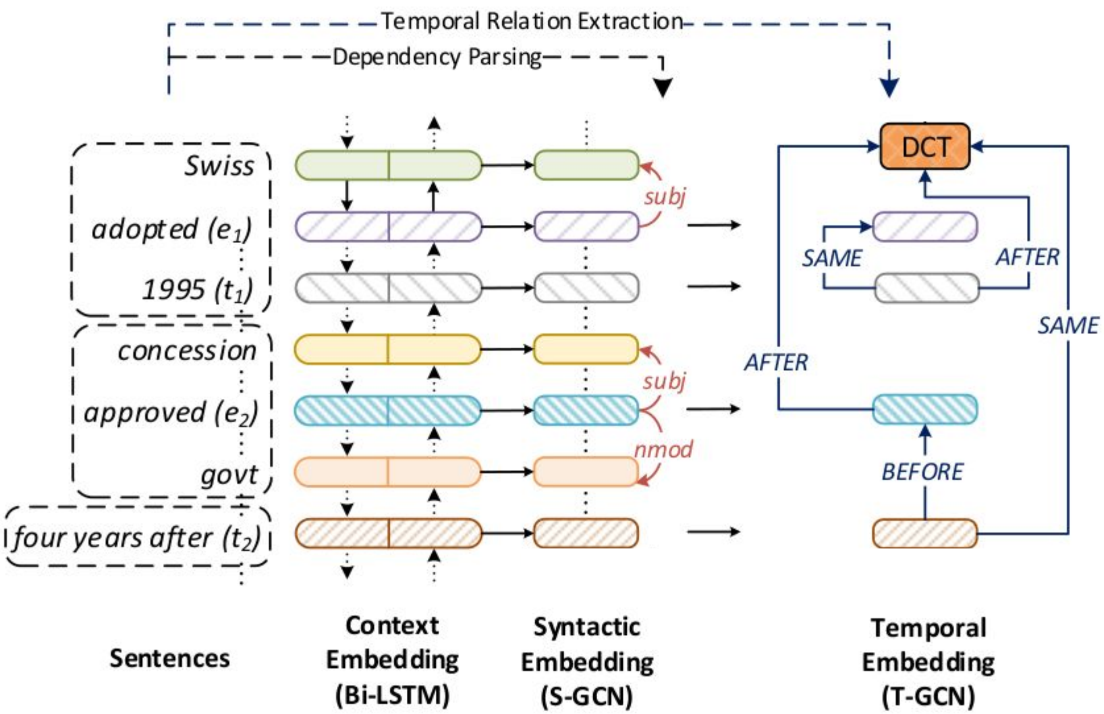
NeuralDater: Overview



NeuralDater: Overview

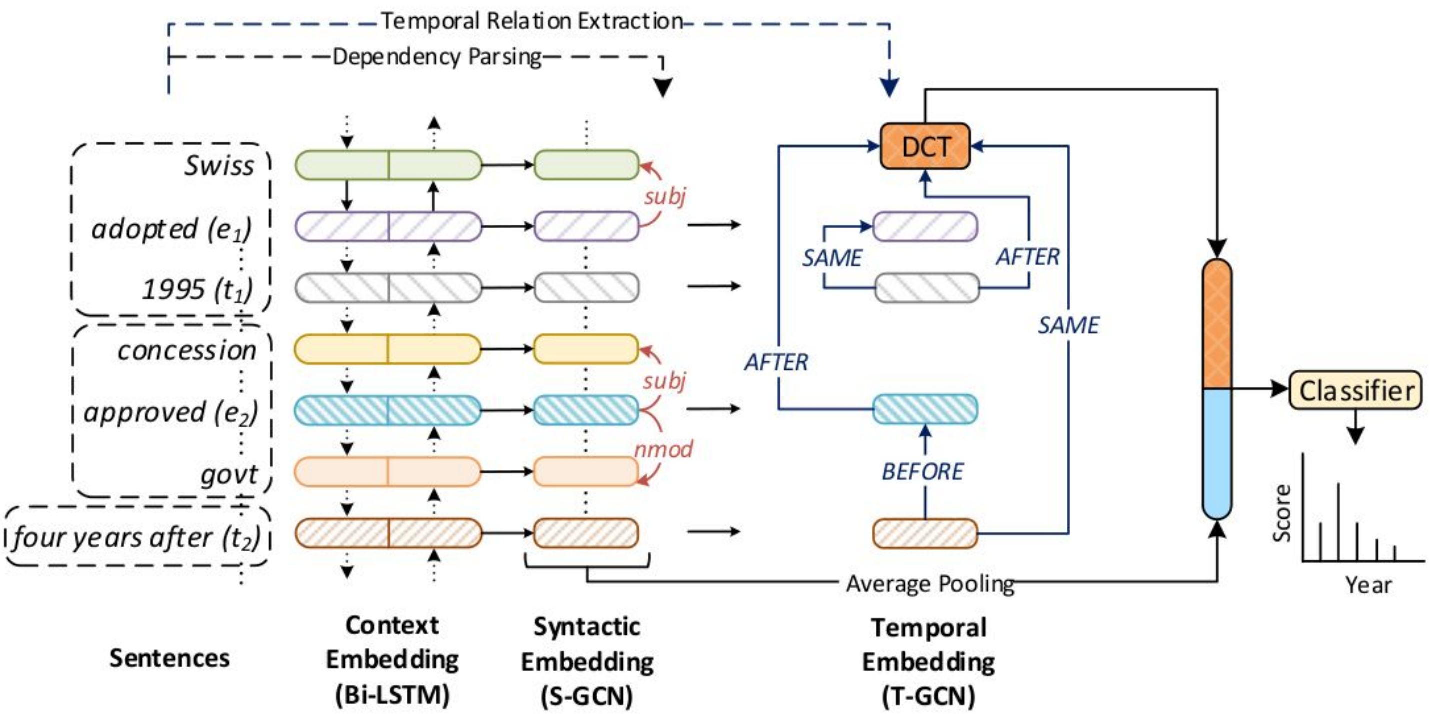


NeuralDater: Overview



CATENA [Mirza et al., COLING'16]

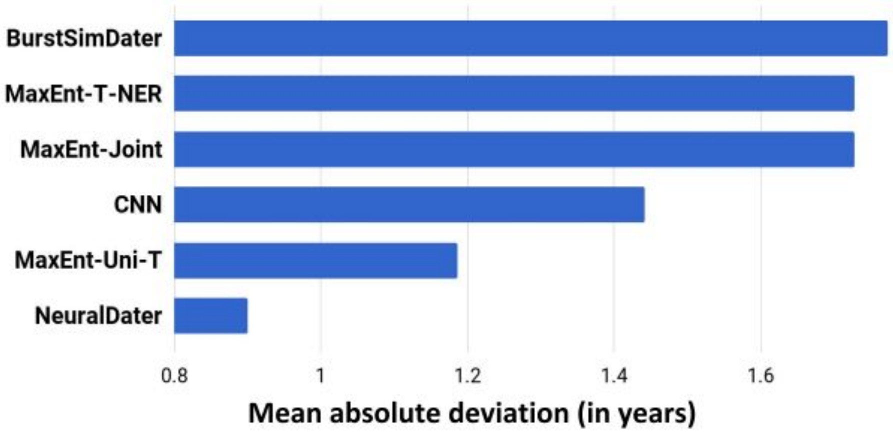
NeuralDater: Overview



NeuralDater: Results

- **Accuracy** and **Mean absolute deviation** on APW & NYT datasets

Method	APW	NYT
BurstySimDater	45.9	38.5
MaxEnt-Time+NER	52.5	42.3
MaxEnt-Joint	52.5	42.5
MaxEnt-Uni-Time	57.5	50.5
CNN	56.3	50.4
NeuralDater	**64.1**	**58.9**



NeuralDater outperforms all the existing methods on the **task.**

NeuralDater: Ablation Study

- Effect of different components of NeuralDater

Method	Accuracy
T-GCN	57.3
S-GCN + T-GCN ($K=1$)	57.8
S-GCN + T-GCN ($K=2$)	58.8
S-GCN + T-GCN ($K=3$)	**59.1**
Bi-LSTM	58.6
Bi-LSTM + CNN	59.0
Bi-LSTM + T-GCN	60.5
Bi-LSTM + S-GCN + T-GCN (no gate)	62.7
Bi-LSTM + S-GCN + T-GCN ($K=1$)	**64.1**
Bi-LSTM + S-GCN + T-GCN ($K=2$)	63.8
Bi-LSTM + S-GCN + T-GCN ($K=3$)	63.3

Incorporation of **Context**, **Syntactic,** and **Temporal** structure achieves **best performance**.

NeuralDater: Ablation Study

- Effect of different components of NeuralDater

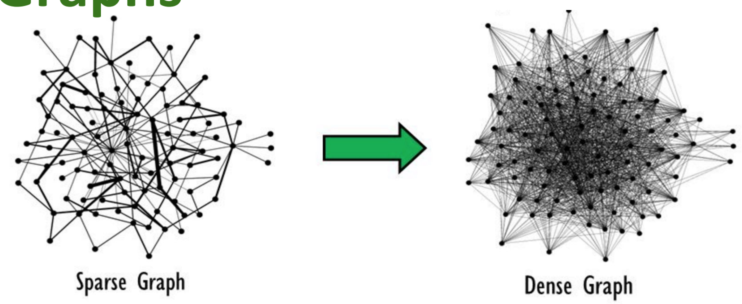
Method	Accuracy
T-GCN	57.3
S-GCN + T-GCN ($K = 1$)	57.8
S-GCN + T-GCN ($K = 2$)	58.8
S-GCN + T-GCN ($K = 3$)	**59.1**
Bi-LSTM	58.6
Bi-LSTM + CNN	59.0
Bi-LSTM + T-GCN	60.5
Bi-LSTM + S-GCN + T-GCN (no gate)	62.7
Bi-LSTM + S-GCN + T-GCN ($K = 1$)	**64.1**
Bi-LSTM + S-GCN + T-GCN ($K = 2$)	63.8
Bi-LSTM + S-GCN + T-GCN ($K = 3$)	63.3

Incorporation of **Context**, **Syntactic,** and **Temporal** structure achieves **best performance**.

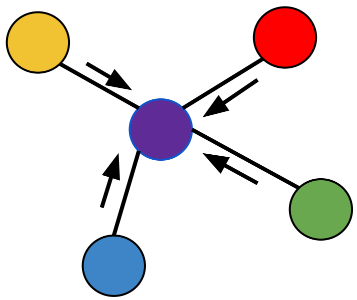
*Shikhar Vashishth, Shib Shankar Dasgupta, Swayambhu Nath Ray, and Partha Talukdar. "**Dating Documents using Graph Convolution Networks**". In Proceedings of the* ***56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018****.*

Outline

- **Addressing Sparsity in Knowledge Graphs**
  - ✓ KG Canonicalization
  - ✓ Relation Extraction
  - ✓ Link Prediction



- **Exploiting Graph Convolutional Networks in NLP**
  - ✓ Document Timestamping
  - ○ Word Representation

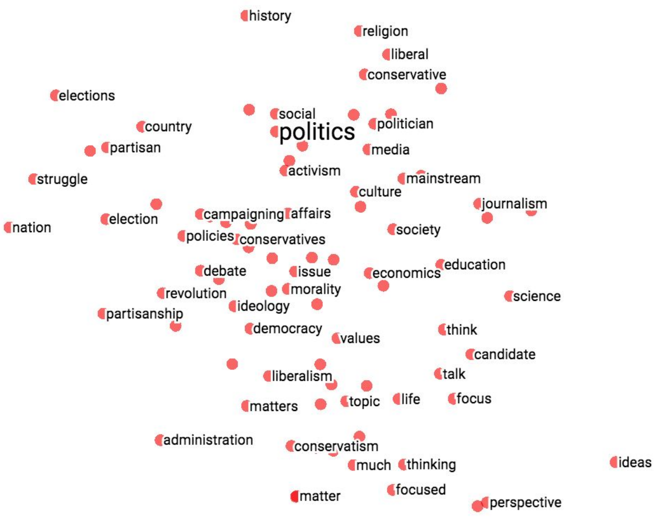


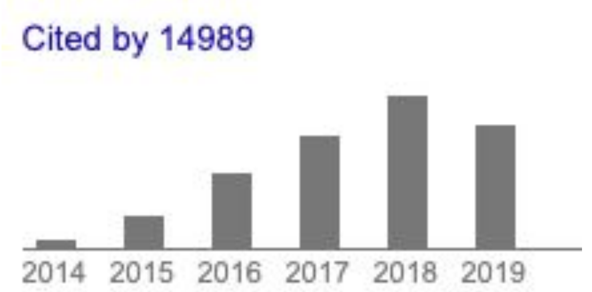
- **Addressing Limitations of Existing GCN Architectures**
  - ○ Unrestricted Influence Neighborhood
  - ○ Applicability to restricted class of graphs

- Conclusion and Future work

Word Representation Learning

- **Problem:**
  Learning a vector representation of words in text.
- Widely used across **all NLP applications**





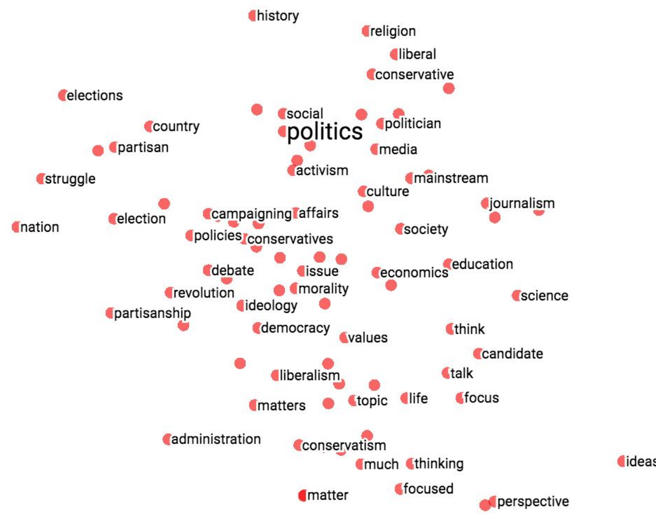
References word2vec

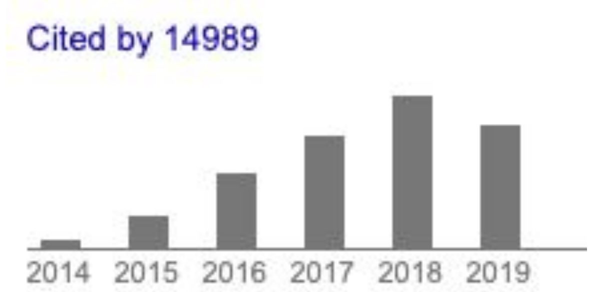
Word Representation Learning

- **Problem:**
  Learning a vector representation of words in text.

- Widely used across **all NLP applications**

- However, most techniques restricted to sequential context
  - Methods using syntactic context suffers from vocabulary explosion
  - Explodes to 1.3 million for 220k words.





References word2vec

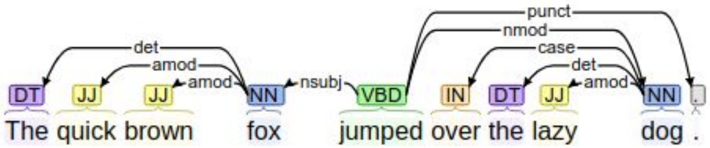
Contributions

- **SynGCN**, a GCN based method for learning word embeddings. Unlike previous methods, SynGCN utilizes syntactic context for learning word representations without increasing vocabulary.

- We also present **SemGCN**, a framework for incorporating diverse semantic knowledge e.g. synonyms, antonyms, hypernyms etc.

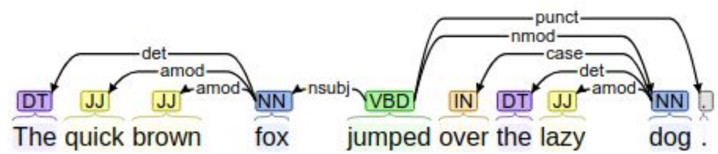
Method: SynGCN

- Given a sentence, $s = (w_1, w_2 \ldots w_n)$. We obtain its dependency parse.

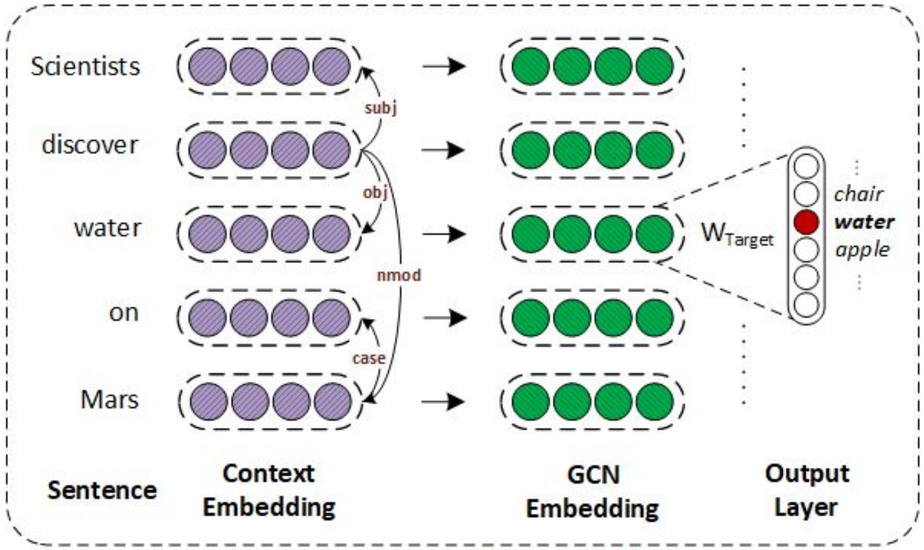


Method: SynGCN

- Given a sentence, $s = (w_1, w_2 \ldots w_n)$. We obtain its dependency parse.



- Utilize syntactic context for predicting a given word w_i.

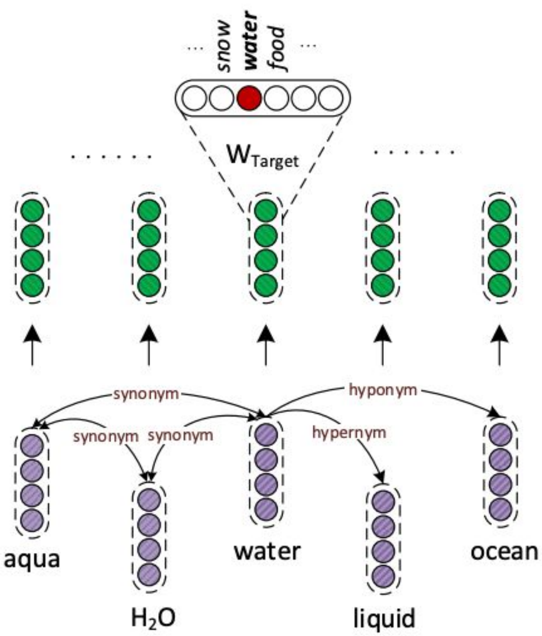


SynGCN (Sentence-level)

$$h_i^{k+1} = f\left(\sum_{j \in \mathcal{N}(i)} g^k_{l_{ij}} \times \left(W^k_{l_{ij}} h^k_j + b^k_{l_{ij}}\right)\right)$$

Method: SemGCN

- Incorporates semantic knowledge in pre-trained word embeddings
- Unlike prior approaches, SemGCN can utilize any kind of semantic knowledge like synonym, antonym, hypernym etc. jointly



SemGCN (Corpus-level)

SynGCN: Results

- Evaluation results on intrinsic and extrinsic tasks.

Intrinsic Tasks

	Word Similarity			
Method	WS353S	WS353R	SimLex999	RW
Word2vec	71.4	52.6	38.0	30.0
GloVe	69.2	**53.4**	36.7	29.6
Deps	65.7	36.2	39.6	33.0
EXT	69.6	44.9	43.2	18.6
SynGCN	**73.2**	45.7	**45.5**	**33.7**

Extrinsic Tasks

POS	SQuAD	NER	Coref
95.0 ± 0.1	78.5 ± 0.3	89.0 ± 0.2	65.1 ± 0.3
94.6 ± 0.1	78.2 ± 0.2	89.1 ± 0.1	64.9 ± 0.2
95.0 ± 0.1	77.8 ± 0.3	88.6 ± 0.3	64.8 ± 0.1
94.9 ± 0.2	$\mathbf{79.6\pm0.1}$	88.0 ± 0.1	64.8 ± 0.1
$\mathbf{95.4\pm0.1}$	$\mathbf{79.6\pm0.2}$	$\mathbf{89.5\pm0.1}$	$\mathbf{65.8\pm0.1}$

SynGCN performs **comparably or outperforms** all word embedding approaches **across several tasks**.

SemGCN: Results

- Evaluation results on intrinsic and extrinsic tasks.

Intrinsic Tasks

Init Embeddings (=X)	SynGCN		
Datasets	WS353	AP	MSR
Performance of X	61.7	69.3	52.8
Retro-fit (X,1)	61.2	67.1	51.4
Counter-fit (X,2)	55.2	66.4	31.7
JointReps (X,4)	60.9	68.2	24.9
SemGCN (X,4)	$\boxed{65.3}$	$\boxed{69.3}$	$\boxed{54.4}$

Extrinsic Tasks

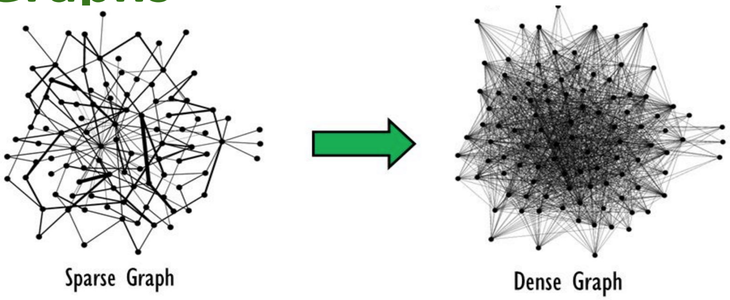
POS	SQuAD	NER	Coref
95.4±0.1	79.6±0.2	**89.5±0.1**	65.8±0.1
94.8±0.1	79.6±0.1	88.8±0.1	66.0±0.2
94.7±0.1	79.8±0.1	88.3±0.3	65.7±0.3
95.4±0.1	79.4±0.3	89.1±0.3	65.6±0.1
95.5±0.1	**80.4±0.1**	**89.5±0.1**	**66.1±0.1**

SemGCN+SynGCN gives best performance across **multiple tasks**.

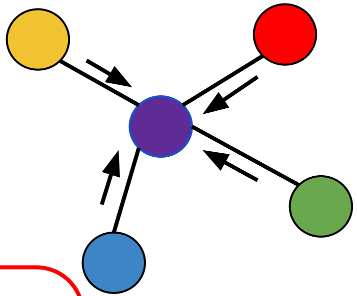
*S. Vashishth, M. Bhandari, P. Yadav, P. Rai, C. Bhattacharyya, and P. Talukdar. "**Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Networks**". In Proceedings of the **57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019.***

Outline

- **Addressing Sparsity in Knowledge Graphs**
  - ✓ KG Canonicalization
  - ✓ Relation Extraction
  - ✓ Link Prediction



- **Exploiting Graph Convolutional Networks in NLP**
  - ✓ Document Timestamping
  - ✓ Word Representation



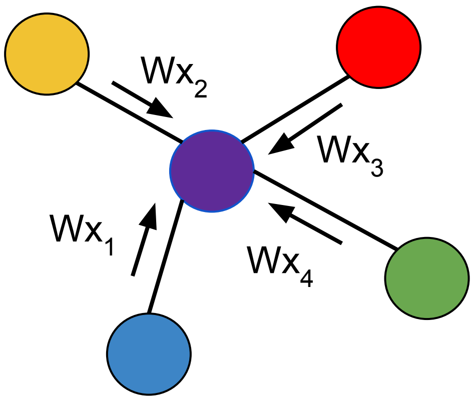
- **Addressing Limitations of Existing GCN Architectures**
  - ○ Unrestricted Influence Neighborhood
  - ○ Applicability to restricted class of graphs

- Conclusion and Future work

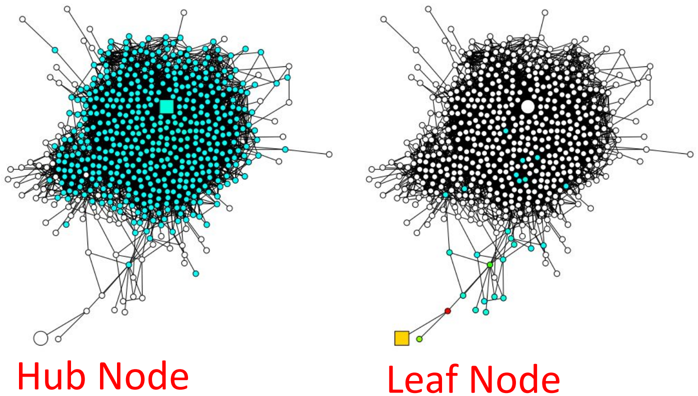
Neighborhood Aggregations in GCNs

- Standard GCN neighborhood aggregation

$$h_v = f\left(\frac{1}{|\mathcal{N}(v)|}\sum_{u\in\mathcal{N}(v)} Wx_u + b\right), \quad \forall v \in \mathcal{V}.$$



- No restriction on influence neighborhood



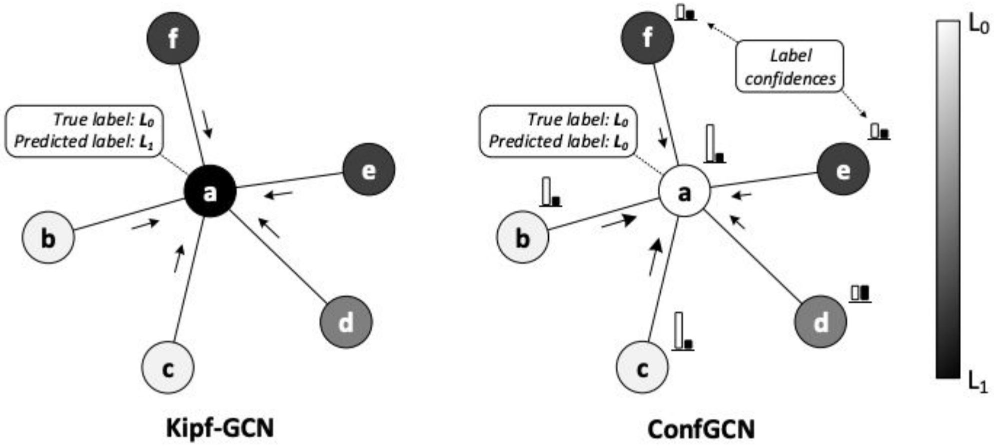
Contributions

- Propose **ConfGCN**, a Graph Convolutional Network (GCN) framework for semi-supervised learning which models label distribution and their confidences for each node in the graph.

- ConfGCN utilize label confidences to estimate influence of one node on another in a label-specific manner during neighborhood aggregation of GCN learning.

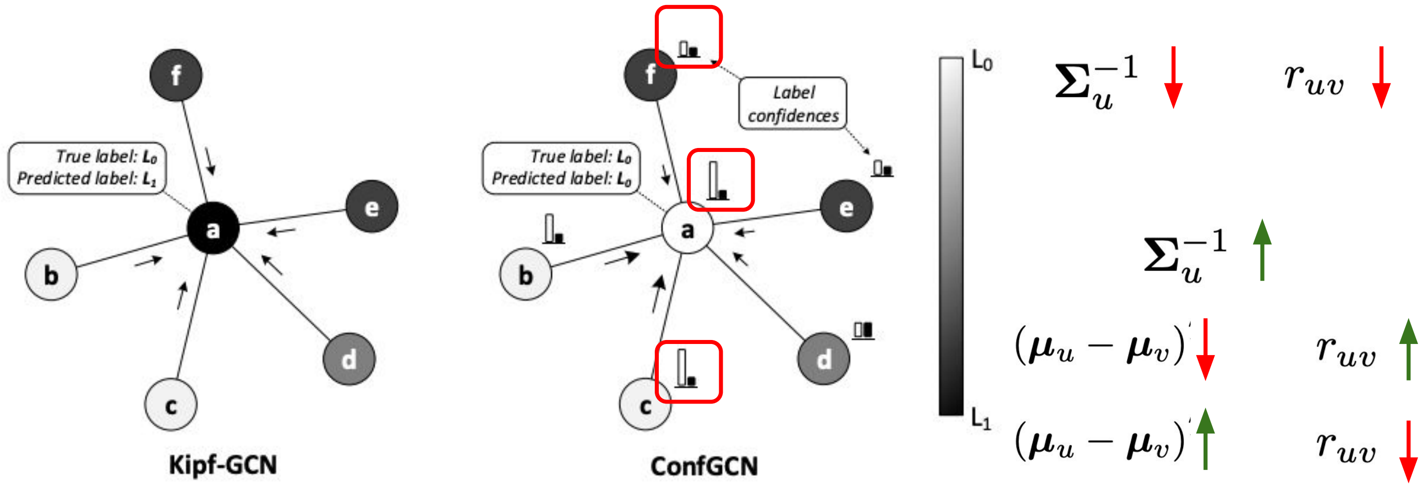
Confidence-based GCN

- Comparison with standard GCN model



Confidence-based GCN

- Comparison with standard GCN model



- Importance for a node is calculated as:

$$r_{uv} = \frac{1}{d_M(u,v)}. \quad d_M(u,v) = (\boldsymbol{\mu}_u - \boldsymbol{\mu}_v)^T(\boldsymbol{\Sigma}_u^{-1} + \boldsymbol{\Sigma}_v^{-1})(\boldsymbol{\mu}_u - \boldsymbol{\mu}_v).$$

  - μ_u, μ_v are label distribution and Σ_u, Σ_v denote co-variance matrices.

ConfGCN: Results

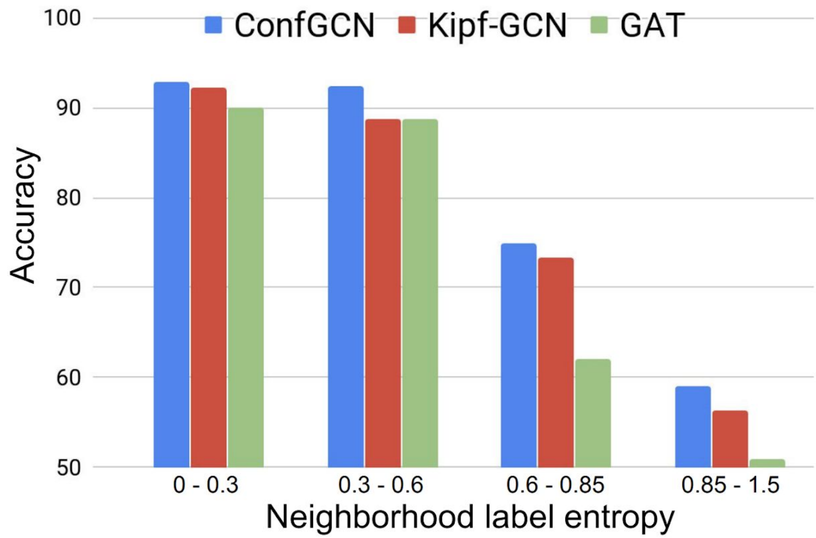
- Performance on **Semi-supervised Learning**

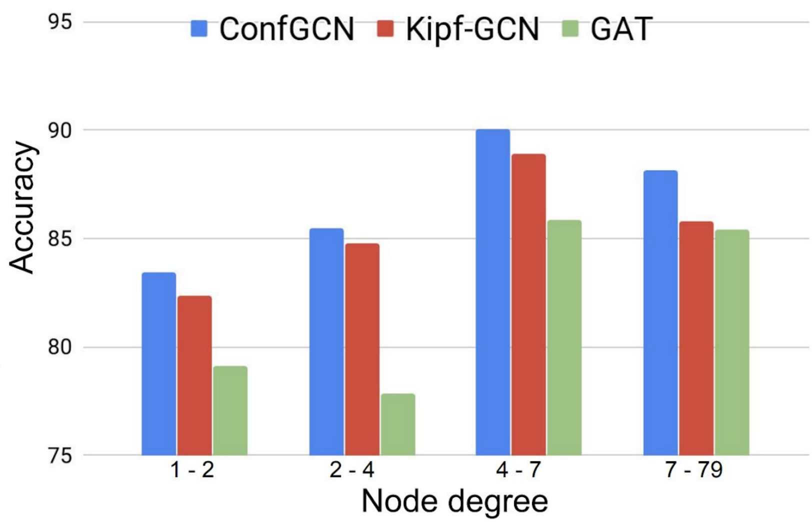
Method	Citeseer	Cora	Pubmed	Cora ML
Kipf-GCN (Kipf and Welling, 2016)	69.4 ± 0.4	80.9 ± 0.4	76.8 ± 0.2	85.7 ± 0.3
G-GCN (Marcheggiani and Titov, 2017)	69.6 ± 0.5	81.2 ± 0.4	77.0 ± 0.3	86.0 ± 0.2
GPNN (Liao et al., 2018)	68.1 ± 1.8	79.0 ± 1.7	73.6 ± 0.5	69.4 ± 2.3
GAT (Veličković et al., 2018)	72.5 ± 0.7	$\mathbf{83.0 \pm 0.7}$	79.0 ± 0.3	83.0 ± 0.8
ConfGCN (this paper)	$\mathbf{72.7 \pm 0.8}$	82.0 ± 0.3	$\mathbf{79.5 \pm 0.5}$	$\mathbf{86.5 \pm 0.3}$

ConfGCN performs **consistently better** across all the **datasets**

ConfGCN: Results

- **Effect of Neighborhood Entropy and Node Degree**

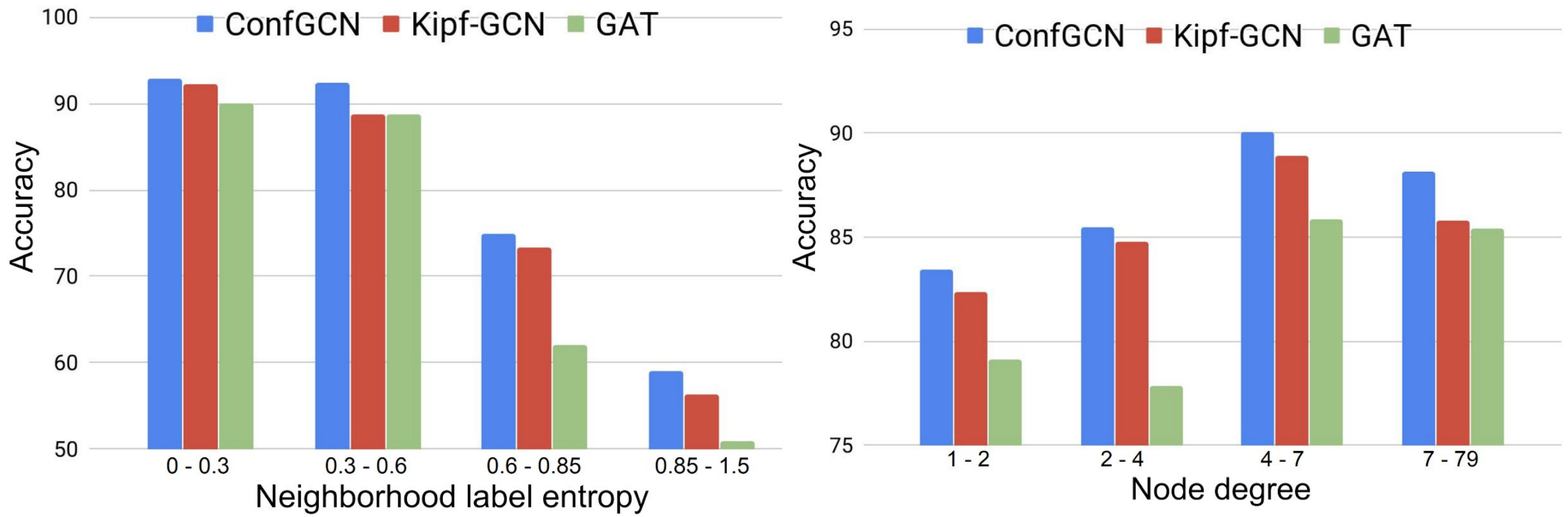




ConfGCN performs **better** than **Kipf-GCN** and **GAT** at all levels of node **entropy** and **degree**.

ConfGCN: Results

- **Effect of Neighborhood Entropy and Node Degree**

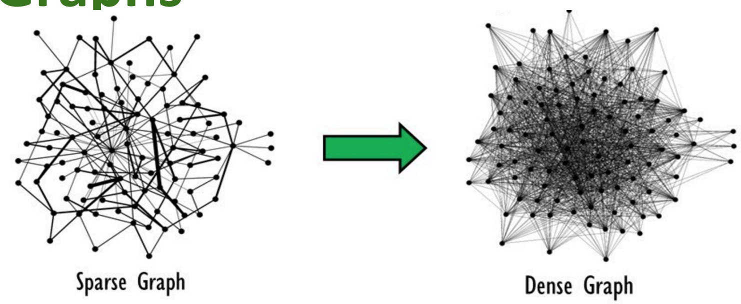


ConfGCN performs **better** than **Kipf-GCN** and **GAT** at all levels of node **entropy** and **degree**.

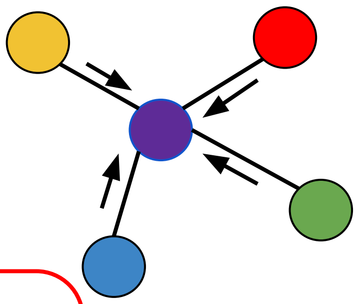
Shikhar Vashishth , Prateek Yadav* , Manik Bhandari* , and Partha Talukdar. "**Confidence-based Graph Convolutional Networks for Semi-Supervised Learning**". In Proceedings of the* ***International Conference on Artificial Intelligence and Statistics (AISTATS), 2019.***

Outline

- **Addressing Sparsity in Knowledge Graphs**
  - ✓ KG Canonicalization
  - ✓ Relation Extraction
  - ✓ Link Prediction



- **Exploiting Graph Convolutional Networks in NLP**
  - ✓ Document Timestamping
  - ✓ Word Representation

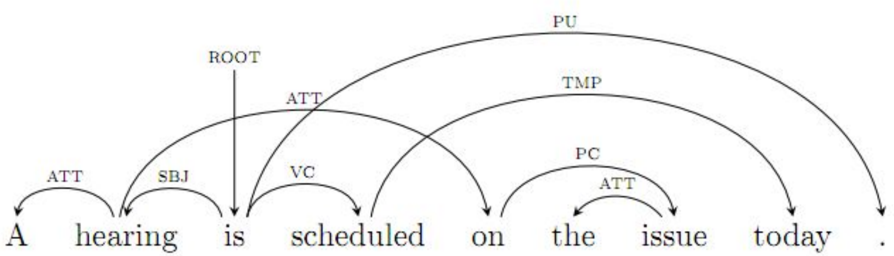


- **Addressing Limitations of Existing GCN Architectures**
  - ✓ Unrestricted Influence Neighborhood
  - ○ Applicability to restricted class of graphs

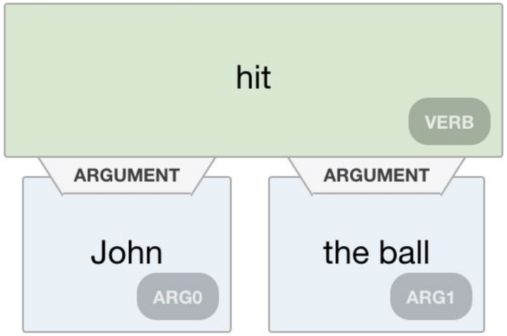
- Conclusion and Future work

Limitations of GCN models

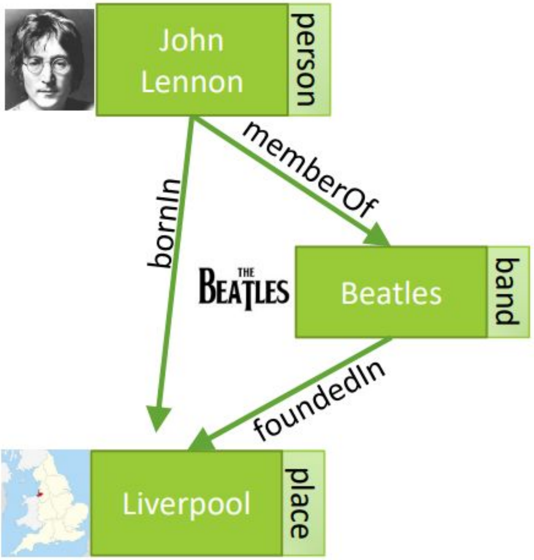
- Most GCNs formulation are for undirected graphs but multi-relational graphs are pervasive.



Dependency Parse



Semantic Role Labeling



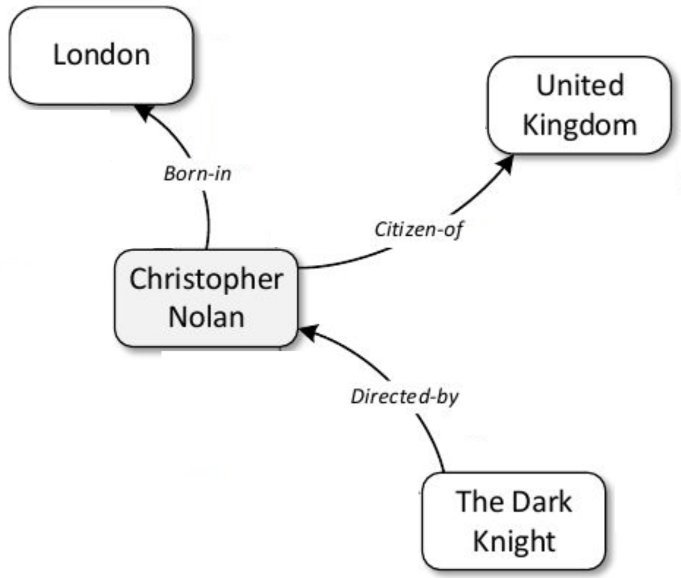
Knowledge Graphs

Contributions

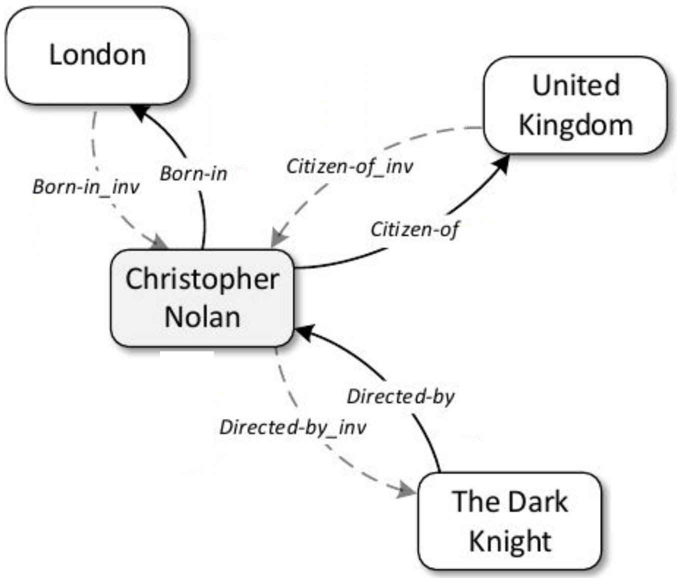
- We propose **CompGCN**, a novel framework for incorporating multi-relational information in GCNs which leverages a variety of composition operations from knowledge graph embedding techniques.

- Unlike previous GCN based multi-relational graph embedding methods, CompGCN jointly learns embeddings of both nodes and relations in the graph

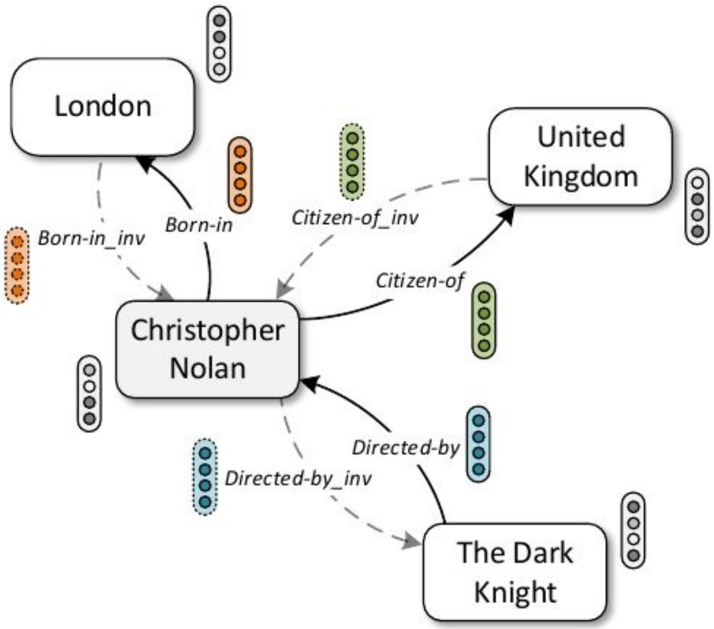
CompGCN: Overview



CompGCN: Overview

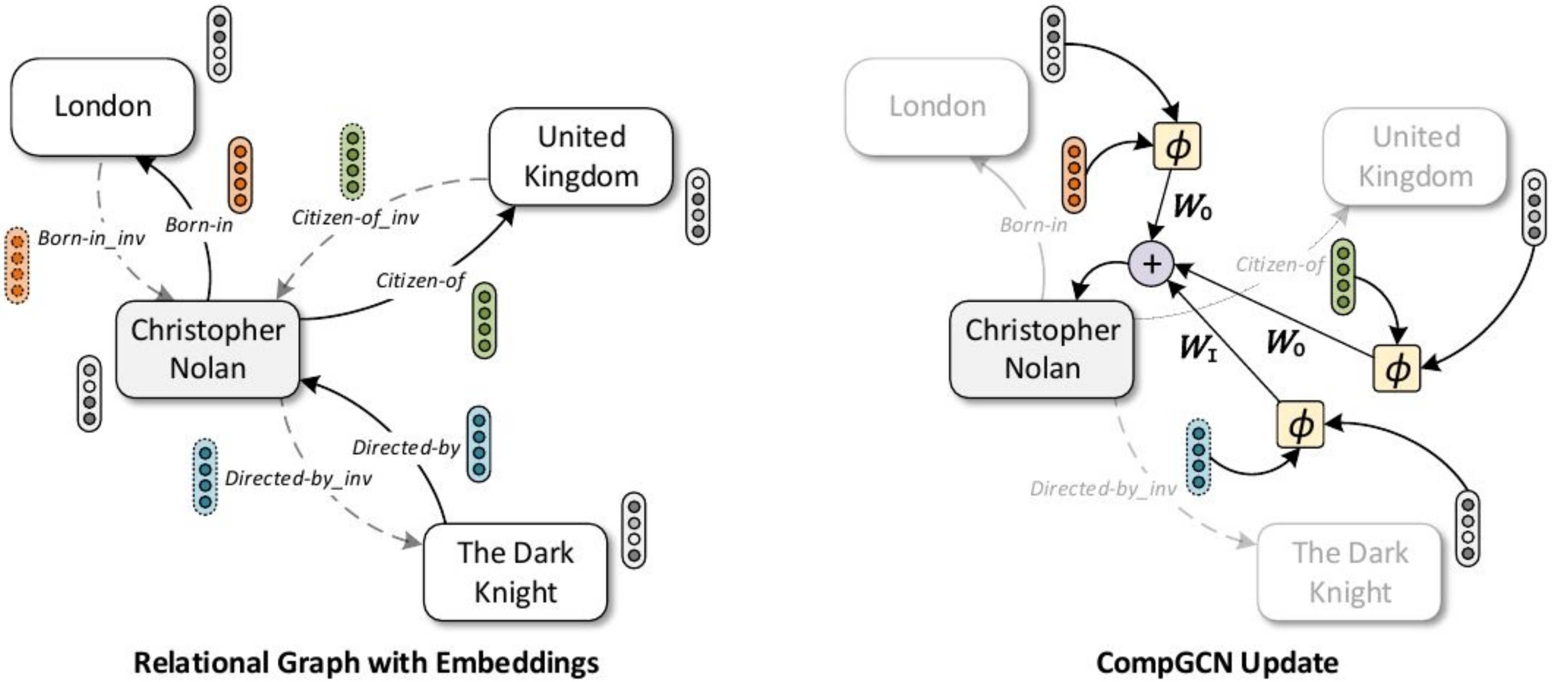


CompGCN: Overview



Relational Graph with Embeddings

CompGCN: Overview



$$h_v^{k+1} = f\left(\sum_{(u,r)\in\mathcal{N}(v)} W_{g(r)}^k \phi(h_u^k, h_r^k)\right) \qquad W_{g(r)} = \begin{cases} W_O, & r \in \mathcal{R} \\ W_I, & r \in \mathcal{R}_{inv} \\ W_S, & r = \top \end{cases}$$

$$h_r^{k+1} = W_{rel}^k \, h_r^k$$

CompGCN: Results

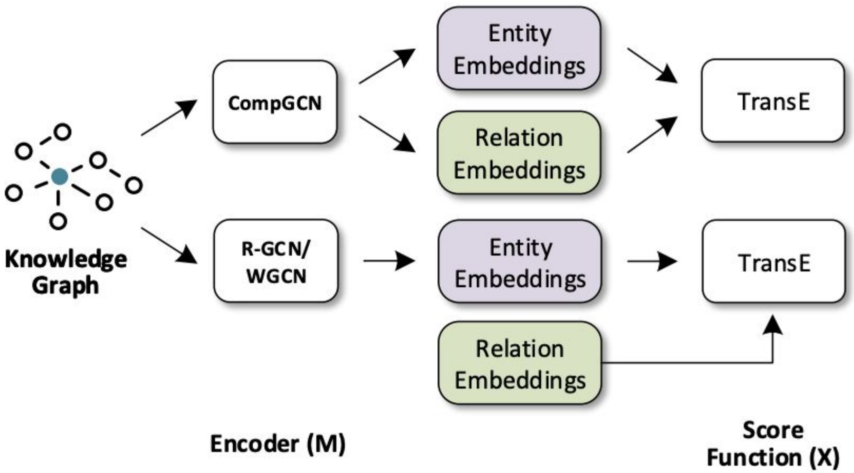
- Performance on Link Prediction

	FB15k-237					WN18RR				
	MRR	MR	H@10	H@3	H@1	MRR	MR	H@10	H@3	H@1
SACN (Shang et al., 2019)	.35	-	.54	.39	.26	.47	-	.54	.48	.43
HypER (Balažević et al., 2019)	.341	250	.520	.376	.252	.465	5798	.522	.477	.436
RotatE (Sun et al., 2019)	.338	**177**	.533	.375	.241	.476	**3340**	**.571**	.492	.428
ConvR (Jiang et al., 2019)	.350	-	.528	.385	.261	.475	-	.537	.489	**.443**
VR-GCN (Ye et al., 2019)	.248	-	.432	.272	.159	-	-	-	-	-
CompGCN (Proposed Method)	**.355**	197	**.535**	**.390**	**.264**	**.479**	3533	.546	**.494**	**.443**

CompGCN performs **consistent improvement** across all the **datasets**

CompGCN: Results

- Effect of different GCN models and composition operators

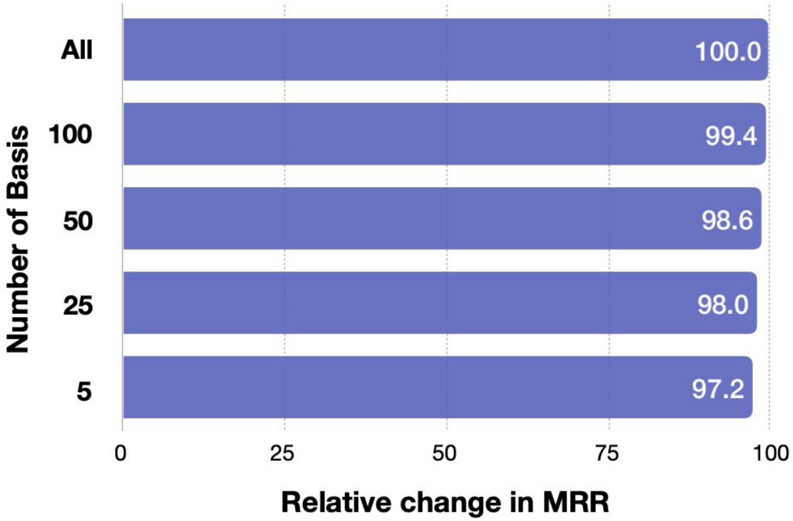


Scoring Function (=X) →	ConvE		
Methods ↓	MRR	MR	H@10
X	0.325	244	0.501
X + D-GCN	0.344	200	0.524
X + R-GCN	0.342	197	0.524
X + W-GCN	0.344	201	0.525
X + CompGCN (Sub)	0.352	199	0.530
X + CompGCN (Mult)	0.353	216	0.532
X + CompGCN (Corr)	**0.355**	197	**0.535**
X + CompGCN ($\mathcal{B} = 50$)	0.350	**193**	0.530

ConvE + CompGCN(Corr) gives **best performance** across all settings.

CompGCN: Results

- Performance with **different number of relation basis vectors** and on **node classification**



Node classification Performance

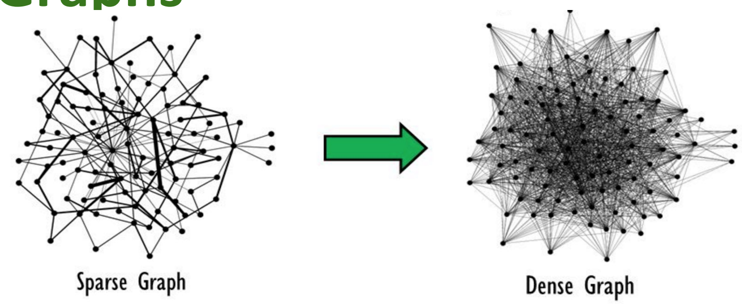
	MUTAG (Node)	AM
Feat*	77.9	66.7
WL*	80.9	87.4
RDF2Vec*	67.2	88.3
R-GCN*	73.2	89.3
SynGCN	74.8 ± 5.5	86.2 ± 1.9
WGCN	77.9 ± 3.2	90.2 ± 0.9
COMPGCN	$\mathbf{85.3 \pm 1.2}$	$\mathbf{90.6 \pm 0.2}$

COMPGCN gives **comparable performance** even with **limited parameters**

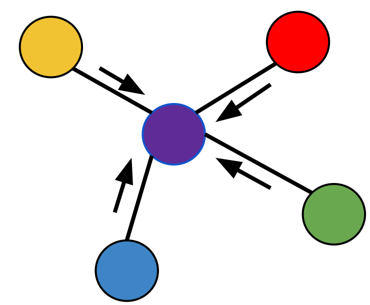
Shikhar Vashishth, Soumya Sanyal*, Vikram Nitin, and Partha Talukdar. "**Composition-based Multi-Relational Graph Convolutional Networks**". CoRR, abs/1909.11218, 2019. [Under Review]*

Outline

- **Addressing Sparsity in Knowledge Graphs**
  - ✓ KG Canonicalization
  - ✓ Relation Extraction
  - ✓ Link Prediction



- **Exploiting Graph Convolutional Networks in NLP**
  - ✓ Document Timestamping
  - ✓ Word Representation



- **Addressing Limitations of Existing GCN Architectures**
  - ✓ Unrestricted Influence Neighborhood
  - ✓ Applicability to restricted class of graphs

- Conclusion and Future work

Scope for Future Research

- **Addressing Sparsity in Knowledge Graphs**
  - Utilizing contextualized embeddings for canonicalization
    - Instead of GloVe, using models like ELMo, BERT.
  - Exploiting other signals from Knowledge graphs
    - Relationship between different entities
  - Extending idea of increase interactions to several existing models
    - Current work demonstrates improvement for one method

Scope for Future Research

- **Exploiting Graph Convolutional Networks in NLP**
  - Instead of restricting to input text, utilizing real world knowledge
    - More close to how humans timestamp a document
  - Utilizing GCNs for learning contextualized embeddings
    - Contextualized embeddings are superior to word2vec embeddings

Scope for Future Research

- **Exploiting Graph Convolutional Networks in NLP**
  - Instead of restricting to input text, utilizing real world knowledge
    - More close to how humans timestamp a document
  - Utilizing GCNs for learning contextualized embeddings
    - Contextualized embeddings are superior to word2vec embeddings

- **Addressing Limitations of Existing GCN Architectures**
  - Scaling GCNs to large graphs
  - Using spectral GCNs for different NLP tasks

Conclusion

- **Addressing Sparsity in Knowledge Graphs**
  - Canonicalization: CESI learns embeddings followed by clustering.
  - Relation Extraction: RESIDE, utilized signals from KG for improving RE
  - Link Prediction: Demonstrate effectiveness of increasing interactions

Conclusion

- **Addressing Sparsity in Knowledge Graphs**
  - Canonicalization: CESI learns embeddings followed by clustering.
  - Relation Extraction: RESIDE, utilized signals from KG for improving RE
  - Link Prediction: Demonstrate effectiveness of increasing interactions

- **Exploiting Graph Convolutional Networks in NLP**
  - NeuralDater for document timestamping which exploits syntactic and temporal graph structure
  - Use GCNs for utilizing syntactic context for learning word embeddings

Conclusion

- **Addressing Sparsity in Knowledge Graphs**
  - Canonicalization: CESI learns embeddings followed by clustering.
  - Relation Extraction: RESIDE, utilized signals from KG for improving RE
  - Link Prediction: Demonstrate effectiveness of increasing interactions
- **Exploiting Graph Convolutional Networks in NLP**
  - NeuralDater for document timestamping which exploits syntactic and temporal graph structure
  - Use GCNs for utilizing syntactic context for learning word embeddings
- **Addressing Limitations of Existing GCN Architectures**
  - Restricted influence neighborhood through confidence based GCN
  - Propose CompGCN for extending GCNs to relational graphs

Thank you

- **References:**

  - Vashishth, Shikhar, Prince Jain, and Partha Talukdar. "CESI: Canonicalizing Open Knowledge Bases using Embeddings and Side Information." Proceedings of the 2018 World Wide Web Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2018. https://arxiv.org/abs/1902.00172
  - Vashishth, Shikhar, et al. "RESIDE: Improving distantly-supervised neural relation extraction using side information." arXiv preprint arXiv:1812.04361 (2018). https://arxiv.org/abs/1812.04361
  - Vashishth, Shikhar, Shib Sankar Dasgupta, Swayambhu Nath Ray and Partha Pratim Talukdar. "Dating Documents using Graph Convolution Networks." ACL (2018). https://arxiv.org/abs/1902.00175
  - Vashishth, Shikhar, Manik Bhandari, Prateek Yadav, Piyush Rai, Chiranjib Bhattacharyya and Partha Pratim Talukdar. "Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Networks." ACL (2018). https://arxiv.org/abs/1809.04283
  - Vashishth, Shikhar, Soumya Sanyal, Vikram Nitin, Nilesh Agrawal and Partha Talukdar. "InteractE: Improving Convolution-based Knowledge Graph Embeddings by Increasing Feature Interactions". Under review in AAAI 2020. https://arxiv.org/abs/1911.00219
  - Vashishth, Shikhar, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. "Composition-based Multi-Relational Graph Convolutional Networks". Under review in International Conference on Learning Representations, 2020. https://openreview.net/forum?id=BylA_C4tPr
  - Vashishth, Shikhar, Prateek Yadav, Manik Bhandari and Partha Pratim Talukdar. "Confidence-based Graph Convolutional Networks for Semi-Supervised Learning." AISTATS (2019). https://arxiv.org/abs/1901.08255