

# Text Generation from Knowledge Graphs with Graph Transformers

NAACL19

**Rik Koncel-Kedziorski** , Dhanush Bekal , Yi Luan , Mirella Lapata , and Hannaneh Hajishirzi

**University of Washington**

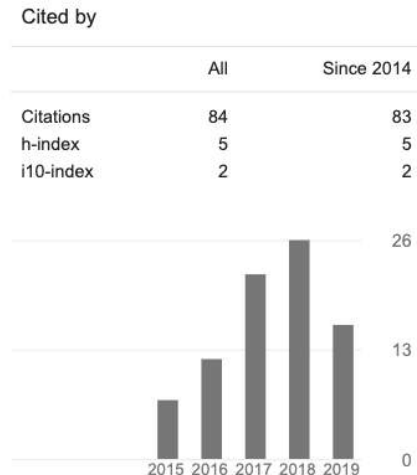
**University of Edinburgh**

**Allen Institute for Artificial Intelligence**

<https://www.youtube.com/watch?v=BiRyvB2NmCM>

# Outline

- Author
- Motivation
- Task
- Dataset
- Model
- Experiments
- Conclusion



# Author

- Rik Koncel-Kedziorski
- Lives on a sailboat
- University of Washington Ph.D. Winter 2019

## Selected Publications

Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata, and Hannaneh Hajishirzi.  
**Text Generation from Knowledge Graphs.** Under Review

Sachin Metha, Rik Koncel-Kedziorski, Mohammad Rastegari, and Hannaneh Hajishirzi.  
**Pyramidal Recurrent Units for Language Modeling.** EMNLP 2018

Rik Koncel-Kedziorski, Ioannis Konstas, Luke Zettlemoyer, and Hannaneh Hajishirzi.  
**A Theme-Rewriting Approach for Generating Math Word Problems.** EMNLP 2016

Aaron Jaech, Rik Koncel-Kedziorski, and Mari Ostendorf.  
**Phonological Pun-derstanding.** NAACL 2016

Rik Koncel-Kedziorski, Subhro Roy, Aida Amini, Nate Kushman, and Hannaneh Hajishirzi.  
**MAWPS: A Math Word Problem Repository.** NAACL 2016

Rik Koncel-Kedziorski, Hannaneh Hajishirzi, Ashish Sabharwal, Oren Etzioni, and Siena Dumas Ang.  
**Parsing Algebraic Word Problems into Equations.** TACL 2015.

R. Koncel-Kedziorski, Hannaneh Hajishirzi, and Ali Farhadi. 2014.  
**Multi-Resolution Language Grounding with weak supervision.** EMNLP 2014.

# Knowledge



World Events

**USED-FOR**

Scientific Information Extraction with Semi-supervised Neural Tagging

**Task**                      **Method**

Luanyl, Ostendorf, Hannaneh Hajjaj, et al.  
 Department of Electrical Engineering, University of Washington  
 {luanyl, ostendor, hannaneh}@uw.edu

**Abstract**

This paper addresses the problem of extracting keyphrases from scientific articles and categorizing them as corresponding to a task, process, or material. We cast the problem as sequence tagging and introduce semi-supervised methods to a neural tagging model, which builds on recent advances in named entity recognition. Since annotated training data is scarce in this domain, we introduce a graph-based semi-supervised algorithm together with a data selection scheme to leverage unannotated articles. Both inductive and transductive semi-supervised learning strategies outperform state-of-the-art information extraction performance on the 2017 SemEval Task 10 ScienceIE task.

**Computer Science:**  
 This paper addresses the task of [named entity recognition]<sub>TASK</sub>, using [conditional random fields]<sub>PROCESS</sub>. Our method is evaluated on the (CoNLL NER Corpus)<sub>MATERIAL</sub>.

**Physics:**  
 [Local field effects]<sub>PROCESS</sub> on spontaneous emission rates within [nanostucture photonic material]<sub>MATERIAL</sub> for example are familiar, and have been well used.

**Material Science:**  
 The [Kelvin probe force microscopy technique]<sub>PROCESS</sub> allows [detection of local EWF]<sub>TASK</sub> between an [atomic force microscopy]<sub>MATERIAL</sub> and [metal surface]<sub>MATERIAL</sub>.

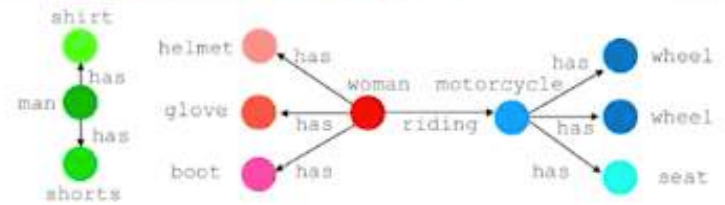
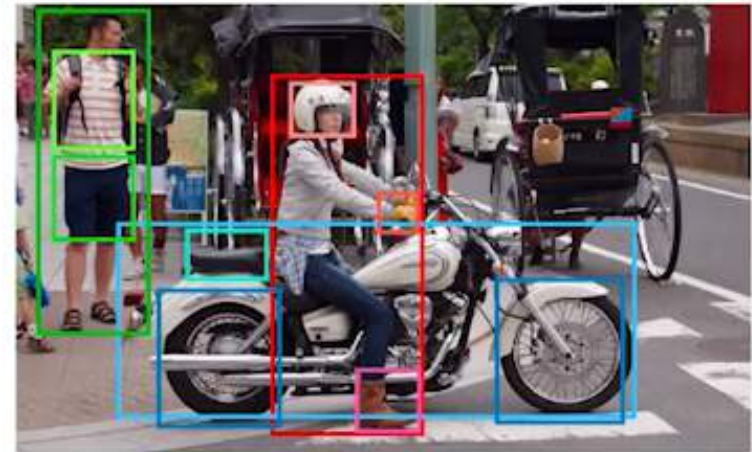
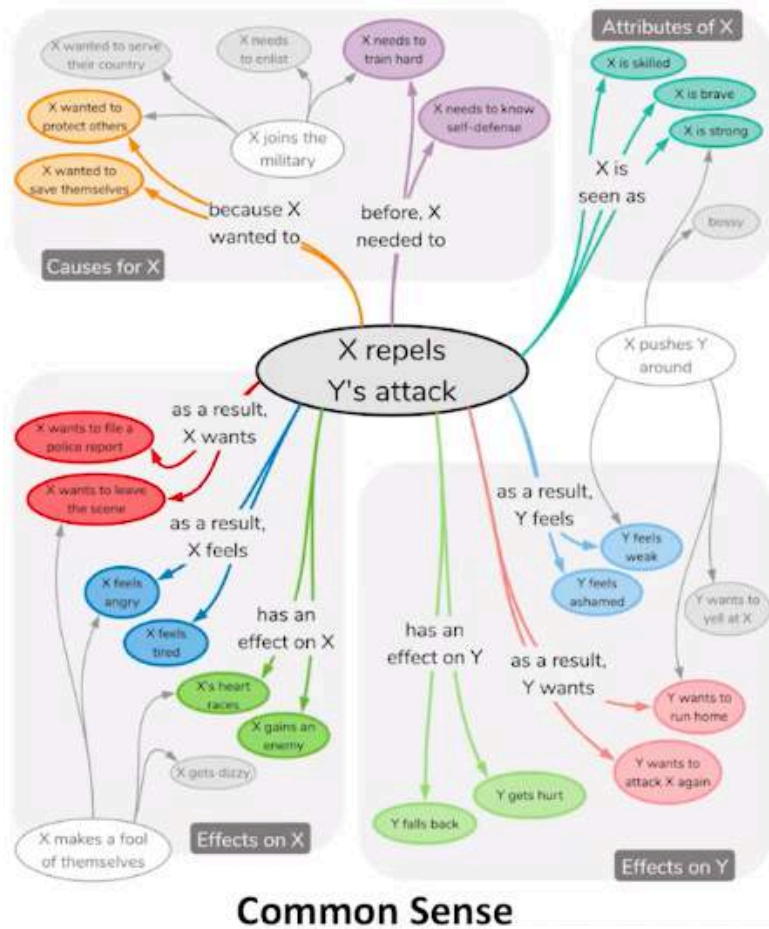
Figure 1: Annotated ScienceIE examples.

scientific paragraphs with keyphrase annotations for three categories: TASK, PROCESS, MATERIAL across three scientific domains, Computer Science (CS), Material Science (MS), and Physics (Phy).

**Dataset**

Science

# Knowledge



Multi-media

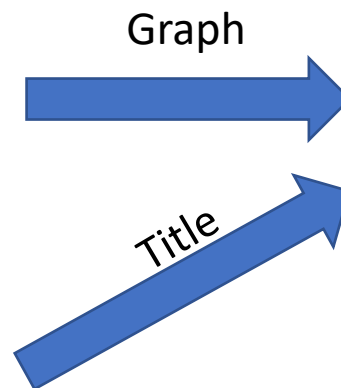
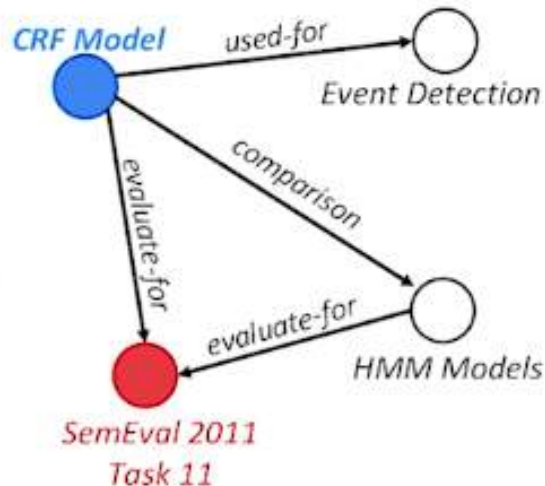
# Task

- **Input**

- **Title** of a scientific article;
- **Knowledge graph** constructed by an automatic information extraction system;

- **Output**

- **Abstract** (text);



## Abstract

We present a CRF model for Event Detection tasks. Our model utilizes such and such features and can outperform standard HMM models by 110% on SemEval Task 11 Dataset. ...

Title: Event Detection with Conditional Random Fields

# Dataset

- **Abstract GENeration DAataset (AGENDA) Dataset**
- 12 top AI conferences
- **SciE** system : a state-of-the-art science domain information extraction system.
  - NER、 Co-Reference、 Relations

	Title	Abstract	KG
Vocab	29K	77K	54K
Tokens	413K	5.8M	1.2M
Entities	-	-	518K
Avg Length	9.9	141.2	-
Avg #Vertices	-	-	12.42
Avg #Edges	-	-	4.43

# Dataset

## Title: Event Detection with Conditional Random Fields

### Abstract

We present a **CRF Model** for Event Detection.

*used-for*

We evaluate **this model** on **SemEval 2010 Task 11**

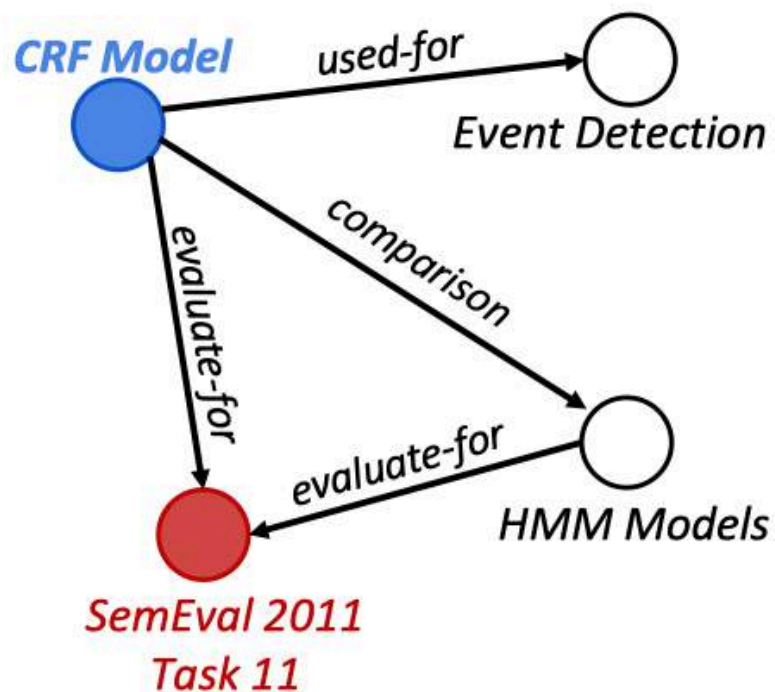
*evaluate-for*

**Our Model** outperforms HMM models by 15% on **this data**.

*comparison*

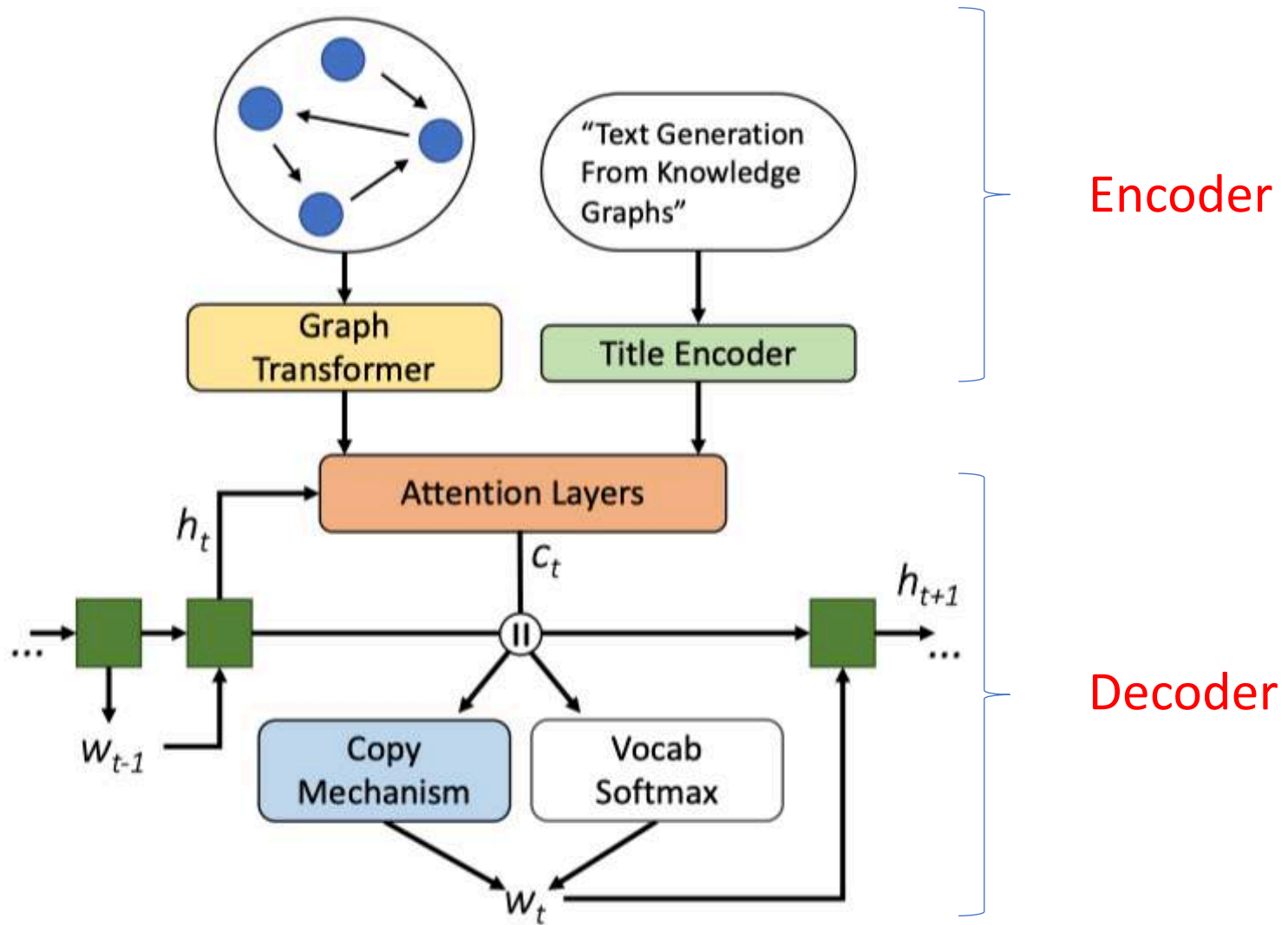
*evaluate-for*

### Graph

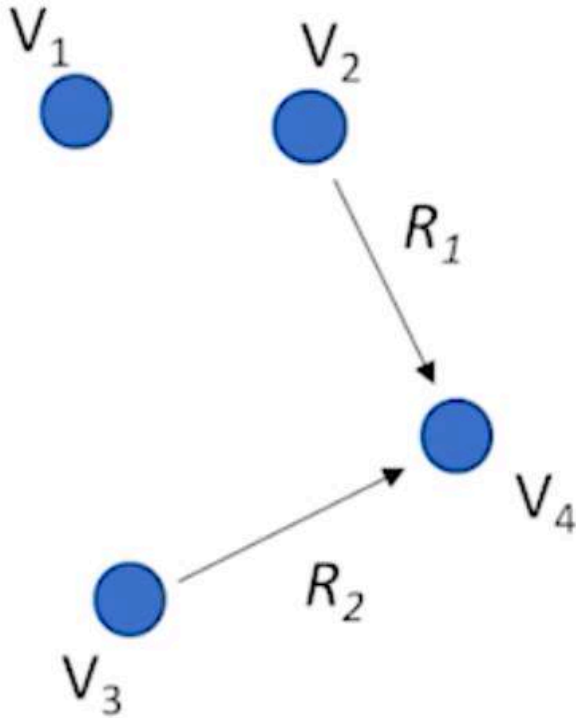




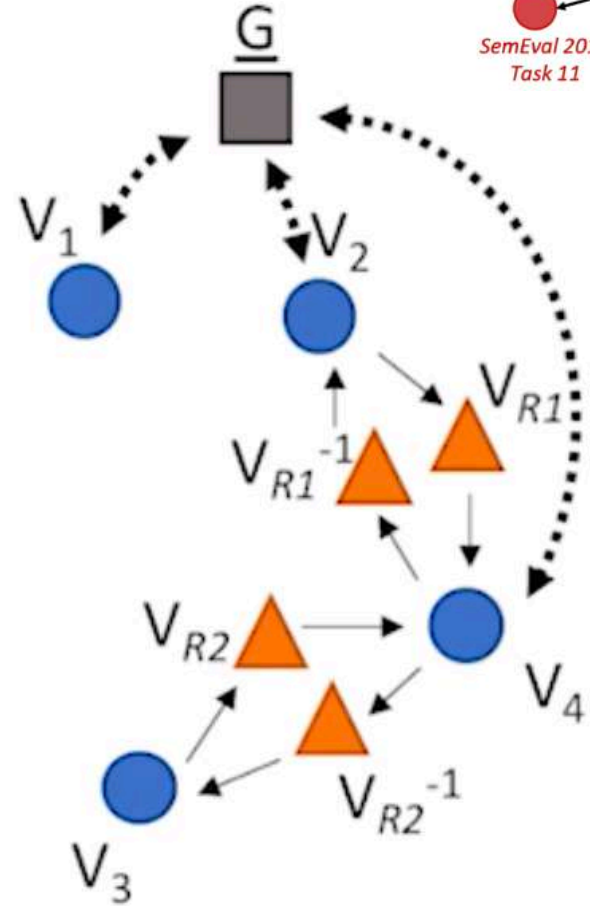
# Model-GraphWriter



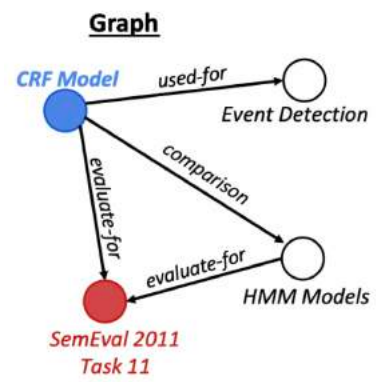
# Graph Preparation



disconnected labeled graph



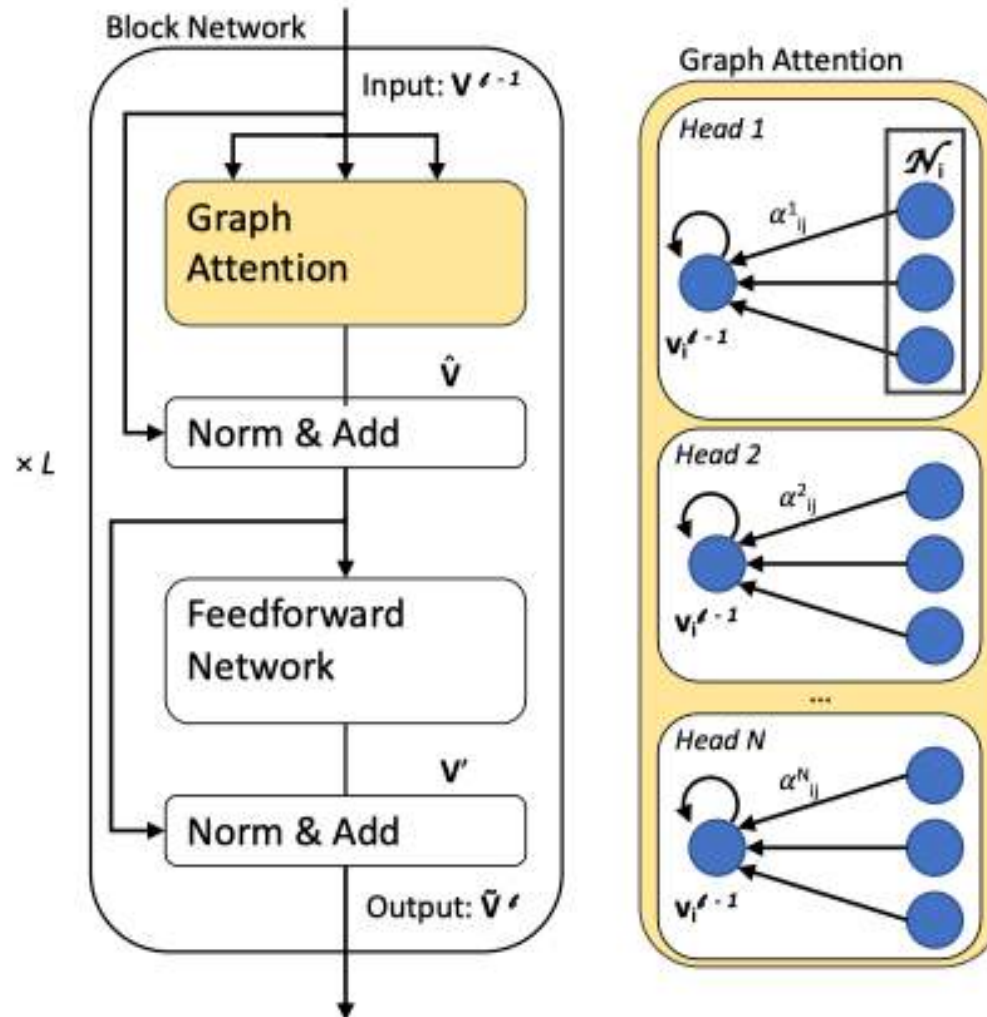
connected unlabeled graph



# Embedding Vertices, Encoding Title

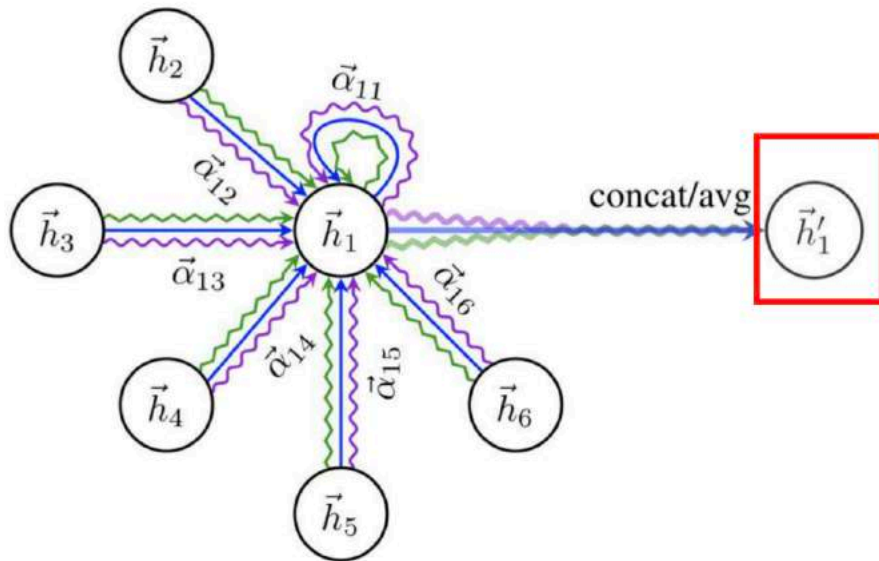
- **Relation** : forward- and backward-looking, two embeddings per relation
- **Entities** correspond to scientific terms which are often multi-word expressions.
- **Bidirectional RNN** run over embeddings of each word
- The **title** input is also a short string, and so we encode it with another **BiRNN**

# Graph Transformer



# GAT

- Graph attention networks ICLR 2018 GAT

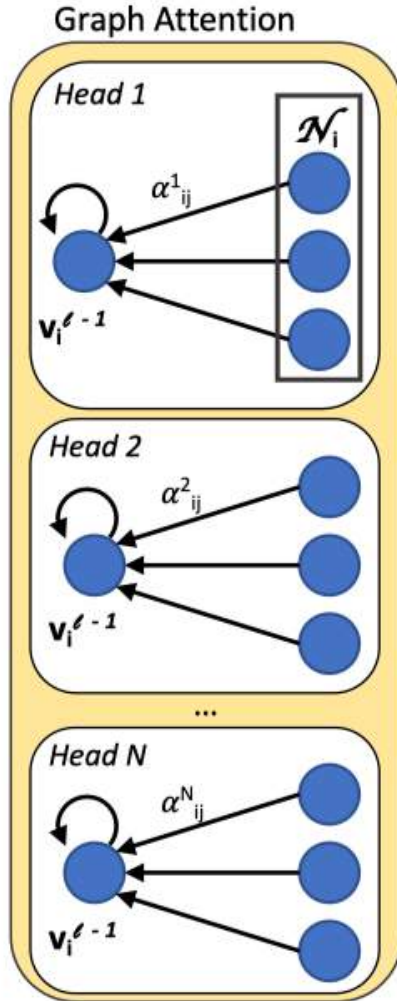


[Figure from Veličković et al. (ICLR 2018)]

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i\|\mathbf{W}\vec{h}_j]\right)\right)}{\sum_{k \in \mathcal{N}_i} \exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i\|\mathbf{W}\vec{h}_k]\right)\right)}$$

$$\vec{h}'_i = \sigma\left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{N}_i} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j\right)$$

# Graph Attention



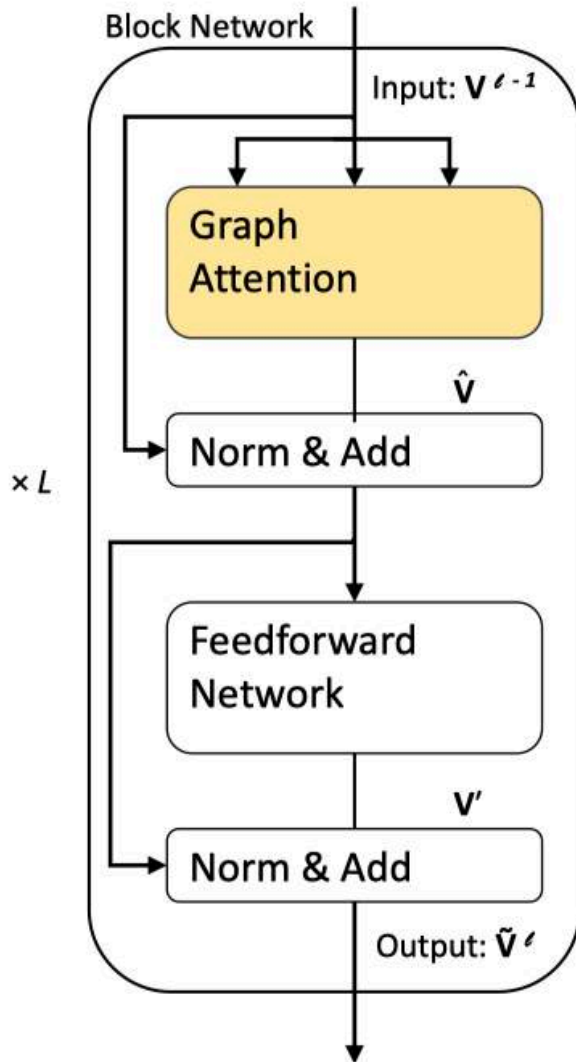
$$\hat{\mathbf{v}}_i = \mathbf{v}_i + \parallel \sum_{n=1}^N \alpha_{ij}^n \mathbf{W}_V^n \mathbf{v}_j$$

concat

$$\alpha_{ij}^n = a^n(\mathbf{v}_i, \mathbf{v}_j)$$

$$a(\mathbf{q}_i, \mathbf{k}_j) = \frac{\exp((\mathbf{W}_K \mathbf{k}_j)^\top \mathbf{W}_Q \mathbf{q}_i)}{\sum_{z \in \mathcal{N}_i} \exp((\mathbf{W}_K \mathbf{k}_z)^\top \mathbf{W}_Q \mathbf{q}_i)}$$

# Block networks



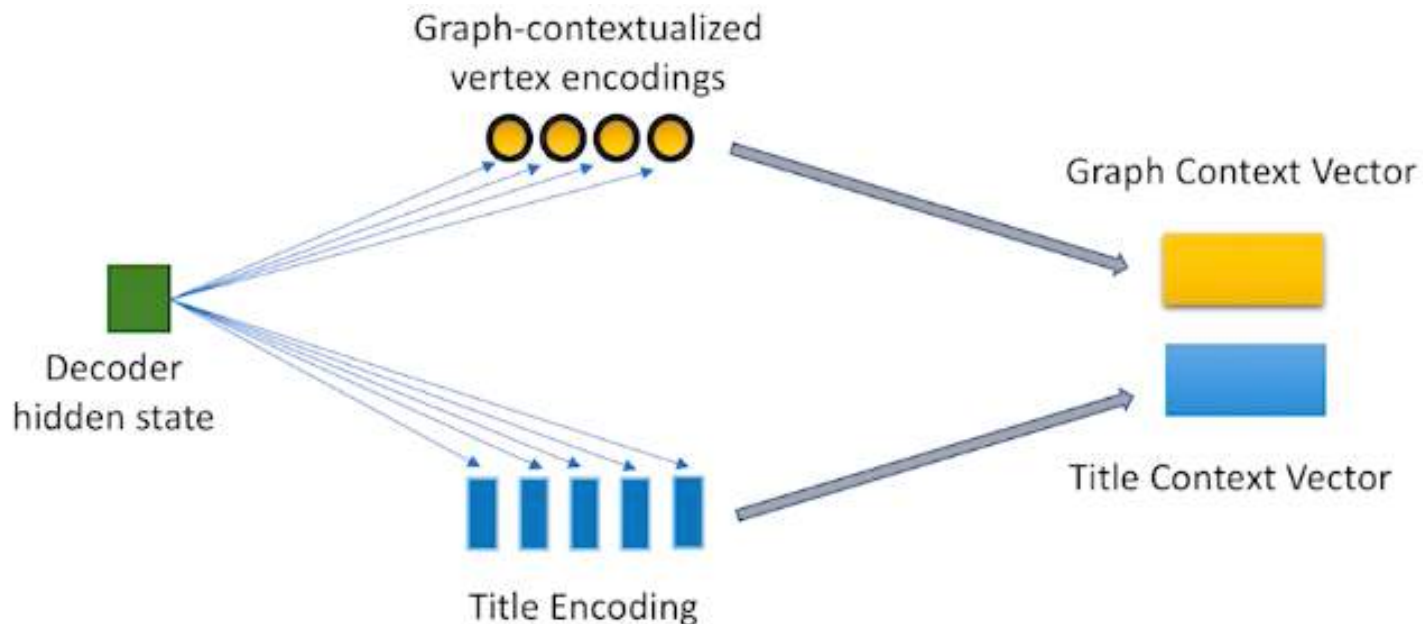
global contextualization

$$\tilde{\mathbf{v}}_i = \text{LayerNorm}(\mathbf{v}'_i + \text{LayerNorm}(\hat{\mathbf{v}}_i))$$

$$\mathbf{v}'_i = \text{FFN}(\text{LayerNorm}(\hat{\mathbf{v}}_i))$$

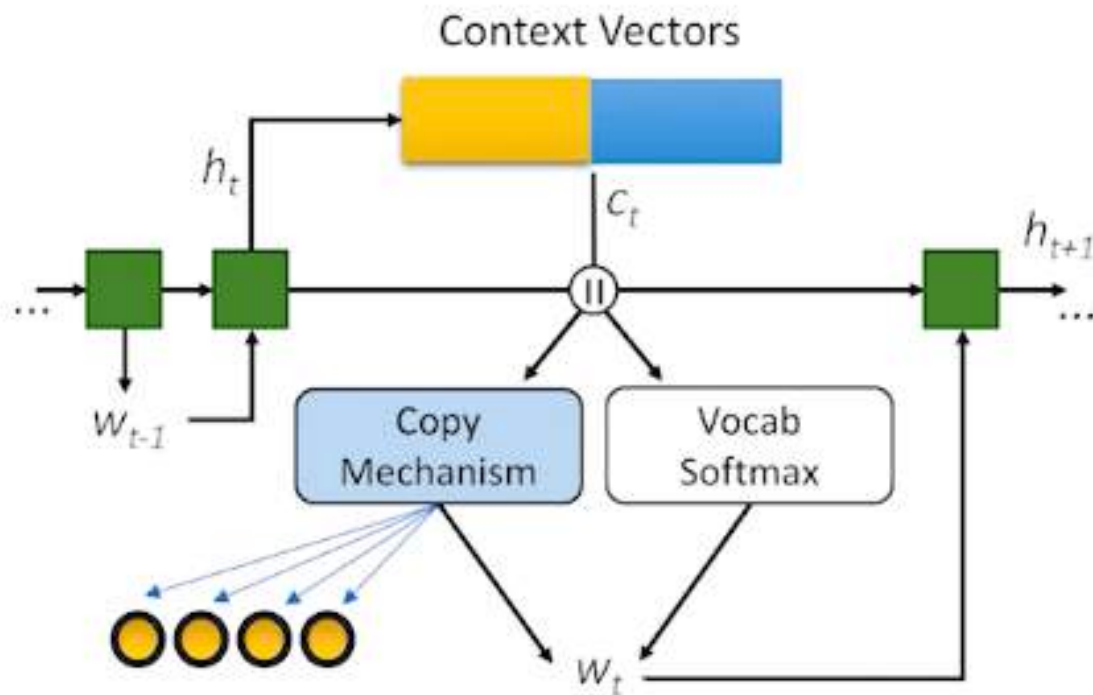
# Decoder

- At each decoding timestep  $t$  we use decoder hidden state  $h_t$  to compute context vectors  $c_g$  and  $c_s$  for the graph and title sequence





# Copy



$$\mathbf{c}_t = [\mathbf{c}_g || \mathbf{c}_s]$$

$$p = \sigma(\mathbf{W}_{copy}[\mathbf{h}_t || \mathbf{c}_t] + b_{copy})$$

$$p * \alpha^{copy} + (1 - p) * \alpha^{vocab}$$

entities

# Experiments

- Evaluation Metrics
- Human evaluation
  - Grammar
  - Fluency
  - Coherence
  - Informativeness
- Automatic metrics
  - BLEU
  - METEOR

# Baselines

- **GAT** : PReLU activations stacked between 6 self-attention layers.
- **EntityWriter** : uses only entities and title (no graph)
- **Rewriter** : uses only the document title

	BLEU	METEOR
GraphWriter	<b>14.3</b> $\pm$ 1.01	<b>18.8</b> $\pm$ 0.28
GAT	12.2 $\pm$ 0.44	17.2 $\pm$ 0.63
EntityWriter	10.38	16.53
Rewriter	1.05	8.38

# Does Knowledge Help?

	Best	Worst
Rewriter (No knowledge)	12%	64%
GraphWriter (Knowledge)	24%	36%
Human Authored	64%	0%

Table 3: Does knowledge improve generation? Human evaluations of best and worst abstract.

	Win	Lose	Tie
Structure	63%	17%	20%
Informativeness	43%	23%	33%
Grammar	63%	23%	13%
Overall	63%	17%	20%

Table 4: Human Judgments of GraphWriter and EntityWriter models.

# Conclusion

- Propose a new graph transformer encoder that applies the successful sequence transformer to graph structured inputs.
- Provide a large dataset of knowledge graphs paired with scientific texts for further study.

**Thanks!**