SPAGAttE:

Graph Embeddings from Ratings Criteria Documents

William Watson, Lawrence Yong
Ratings Data Science
S&P Global Ratings

S&P Global
Ratings

Problem Statement

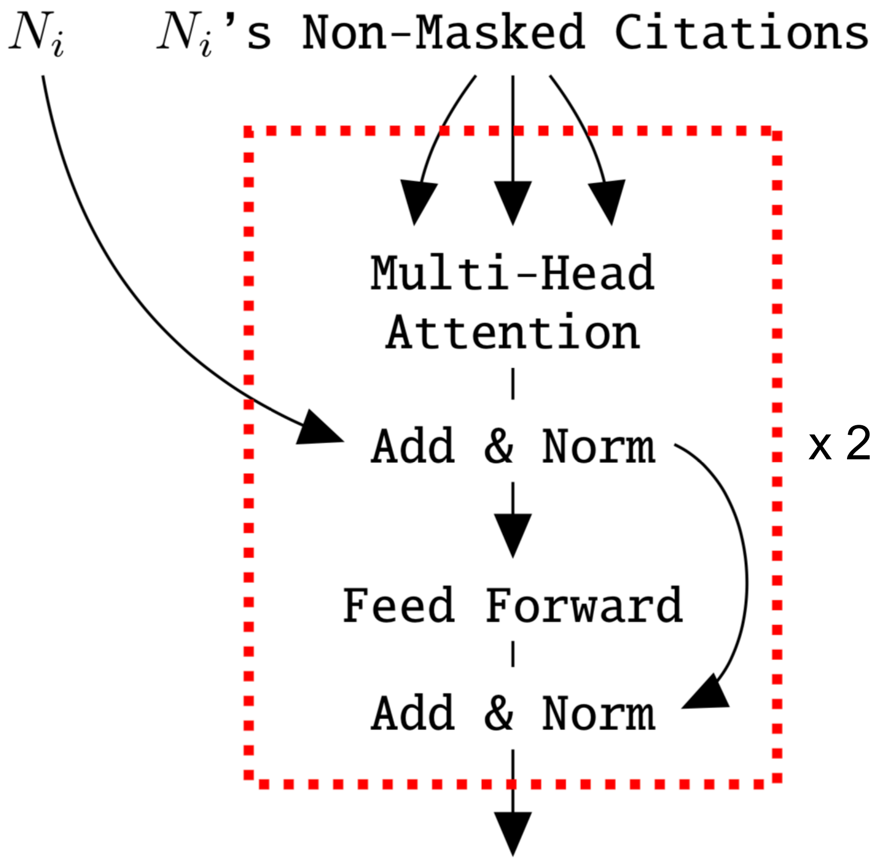
- S&P Global, as the leading issuer of credit ratings, maintains an analytical framework for prescribing said ratings
- This is akin to an academic citation graph, similar to *Cora, Citeseer,* or *Pubmed*
- It is **critical** to keep the citation graph up-to-date
- Hence, we present **SPAGAttE**, a system to build document embeddings from the citation graph
  - And use these embeddings to *reconstruct* the citations for a queried criteria document

Data: The S&P Criteria Methodology Corpus

- **Nodes:** 2,247 Criteria Documents
- **Links:** 13,959 Directed Citations
  - No self-references allowed, but cycles may be present
  - 22.9% of nodes do not have direct citations, but may be cited
- **Vocabulary:** 10,428 Lemmatized Nouns
  - Reduced to the **300 most frequent** lemmatized nouns
  - Converted to normalized TF-IDF vectors

Methodology: Model Components
Graph Attention Networks

- Based on the Transformer Encoder
- The mask is the adjacency matrix
- Each layer attends to an increasing neighborhood
  - First Hop: direct citations
  - Second Hop: direct citations and their direct citations
- Employ dropout on the adjacency matrix to **simulate an incomplete** citation graph

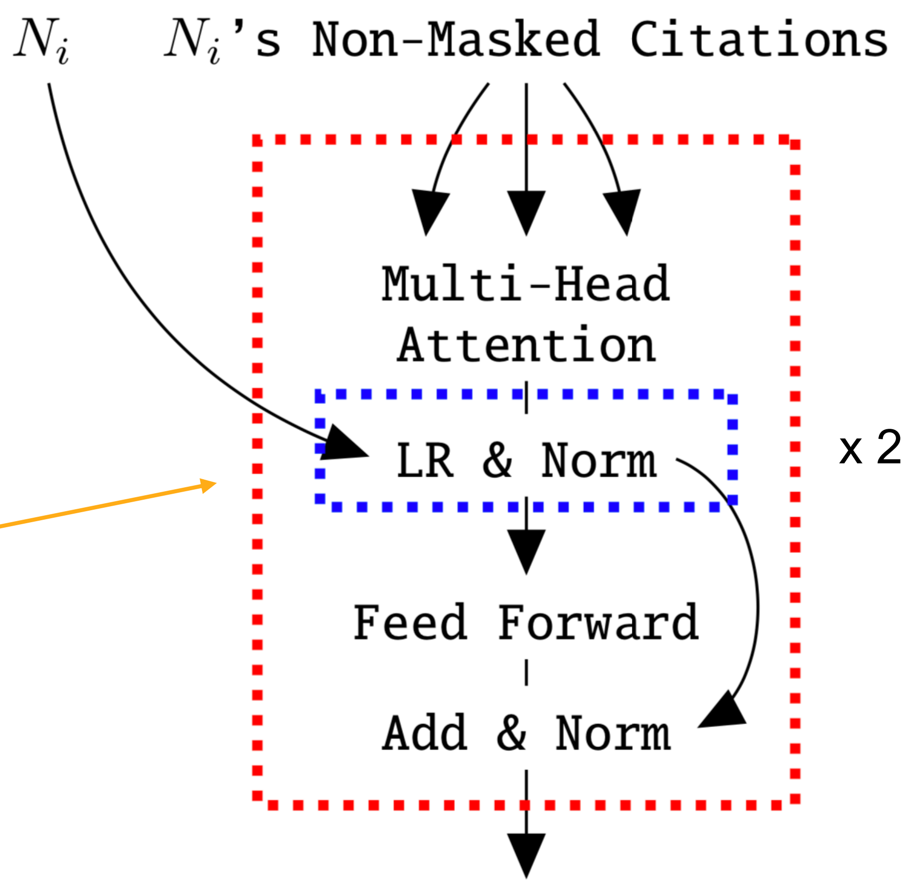


[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
[Veličković *et al.*, 2017] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2017.

Methodology: Model Components
Learned Residual

- The original Transformer equally weights the current node embedding with the attended neighborhood
- Add a learned residual component to **control the influence** of the graph structure on the node embedding

$$z_a = \sigma\left(v_a^T \tanh\left(W_a o_t + U_a n_t + b_a\right)\right)$$
$$r_t = z_a \odot o_t + (1 - z_a) \odot n_t$$



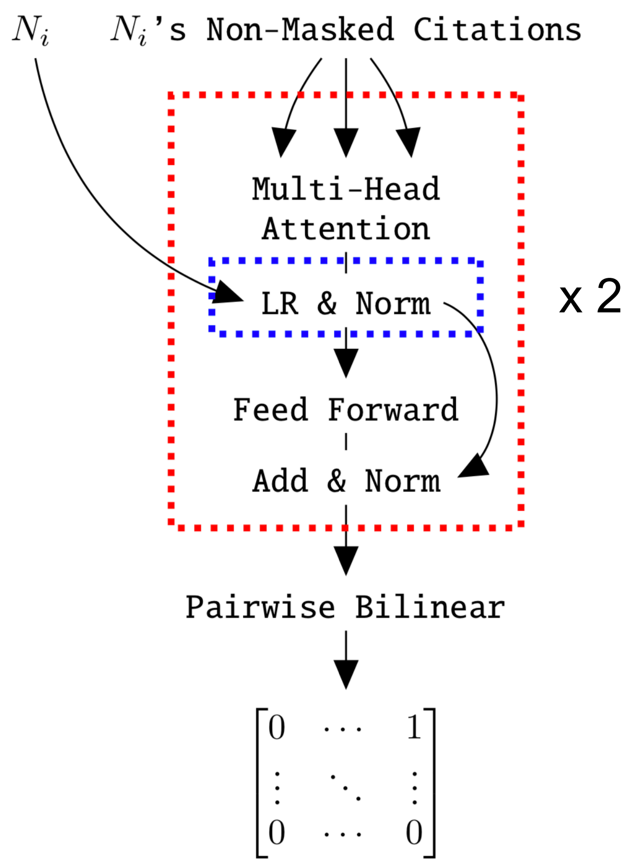
[Bahdanau *et al.*, 2014] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate, 2014.

Methodology: Model Components
Bilinear Scoring and Logistic Loss

- Allows for non-symmetric predictions
- Creates a **pairwise** citation prediction matrix

$$f(e_i, e_j) \quad = \quad e_i^T W_b e_j$$

- Trained using logistic loss
  - Negative Sampling



[Yang *et al.*, 2014] Bishan Yang, Wentau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases, 2014.
[Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality, 2013.

Training: Node Splitting

- **Train:** 1,472 Nodes (65.5%)
- **Validation:** 260 Nodes (11.6%)
- **No Direct Citations:** 515 Nodes (22.9%)
- **Note:** faithful to the transductive experimental setup for graph networks, we allow access to all node features

[Yang *et al.*, 2016] Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings, 2016.

Training: Experimental Setup

- Validation node rows are **voided** during training
- We **do not update** our parameters on the validation predictions
  - Only on the training nodes
- Adam Optimizer
  - Learning Rate is 0.001
  - Number of Updates: **1,920**
  - Saved the model with the best MAR at k=20
  - Number Parameters: **90,434**

[Kingma and Ba, 2014] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

Evaluation Metrics

- Similarity:
  - Mean Squared Error: *How close are my citations?*
  - Cosine Similarity: *How well oriented are my citations?*
- Recovery:
  - Mean Average Recall at k
    - Proportion of relevant documents retrieved within the top k documents
  - Mean Average Precision at k
    - Proportion of retrieved documents that are relevant in the top k
- These metrics quantify how well our embeddings can be used for citation ranking, and how dense the embedding space is

Ablations Studies: Model Flows

Input	1st Embedding	2nd Embedding	3rd Embedding	4th Embedding	Scoring
TF-IDF					
TF-IDF	Bilinear				
TF-IDF	Linear	Graph Transformer	Graph Transformer		
TF-IDF	Linear	GT w/ Learned Residual	GT w/ Learned Residual		
TF-IDF	Linear	GT w/ Learned Residual	GT w/ Learned Residual	Bilinear	
TF-IDF	Linear	GT w/ Learned Residual	GT w/ Learned Residual	Bilinear	Logit Scoring

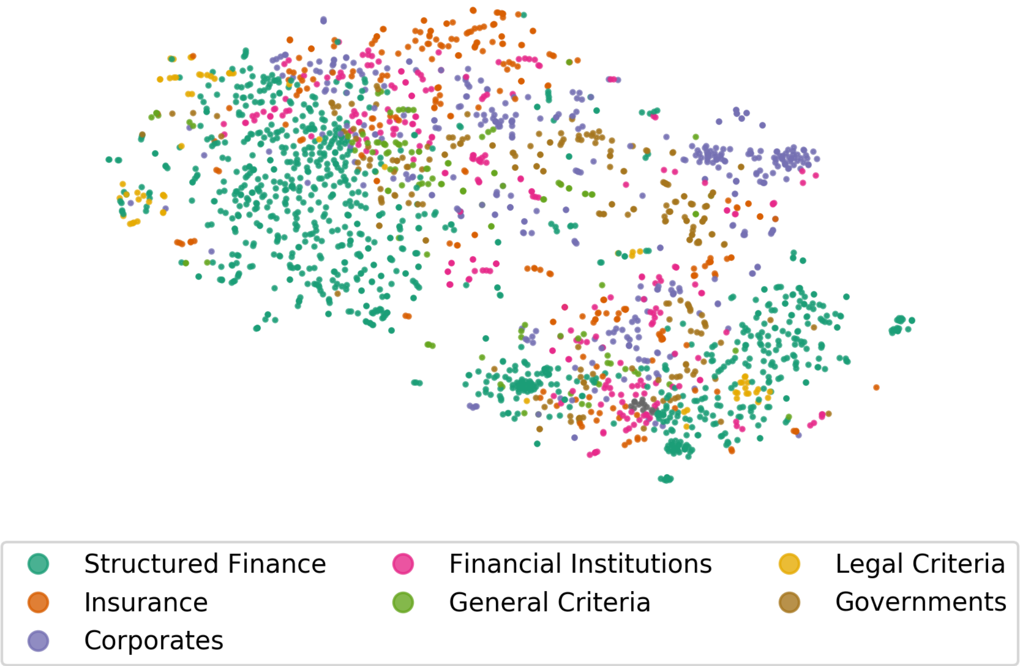
Ablation Studies: Results

- Our validation results show that our modeling choices in conjunction with the training objective represent a good proxy for citation recommendation

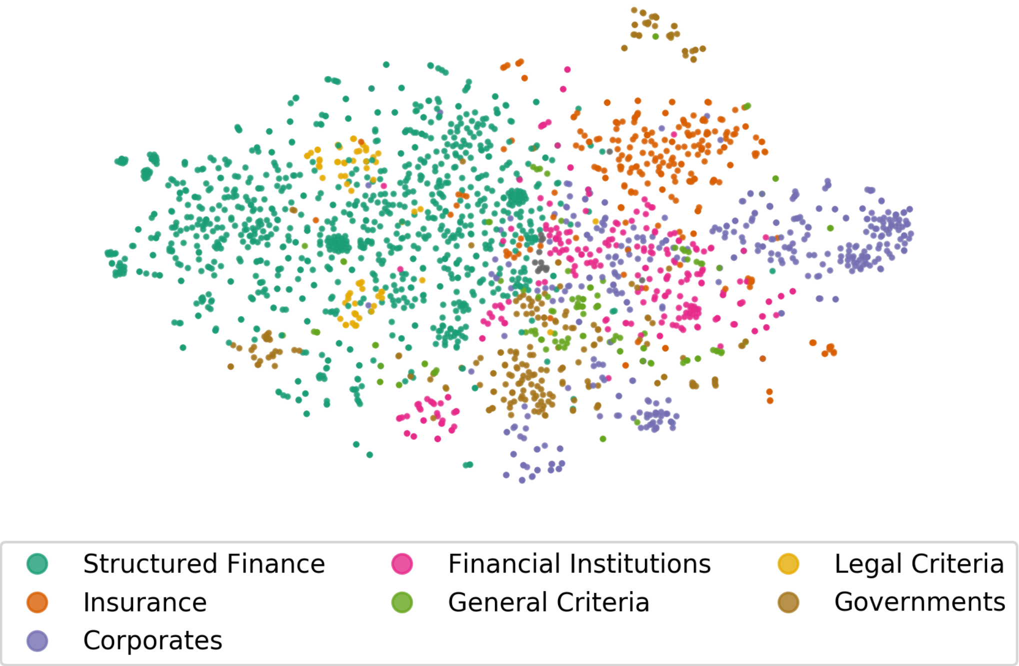
		Similarity		Recovery	
	Model (Embedding Size)	**MSE**	**Cosine**	**MAP@k**	**MAR@k**
	TF-IDF (300)	22.4	0.18	0.100	0.335
ABL	Pairwise Bilinear (300)	4.4	0.17	0.078	0.300
ABL	Graph Transformer (64)	5.0	0.26	0.131	0.378
ABL	+ Learned Residual (64)	**3.6**	0.38	0.152	0.463
INT	Linear Embedding (64)	6.2	0.44	0.125	0.392
INT	Graph Embedding (64)	7.0	0.59	0.156	0.487
INT	Bilinear Embedding (64)	3.7	**0.72**	0.164	0.508
	Logit Scoring (1)	-	-	**0.259**	**0.690**

Visual: t-SNE Embeddings

TF-IDF Embeddings



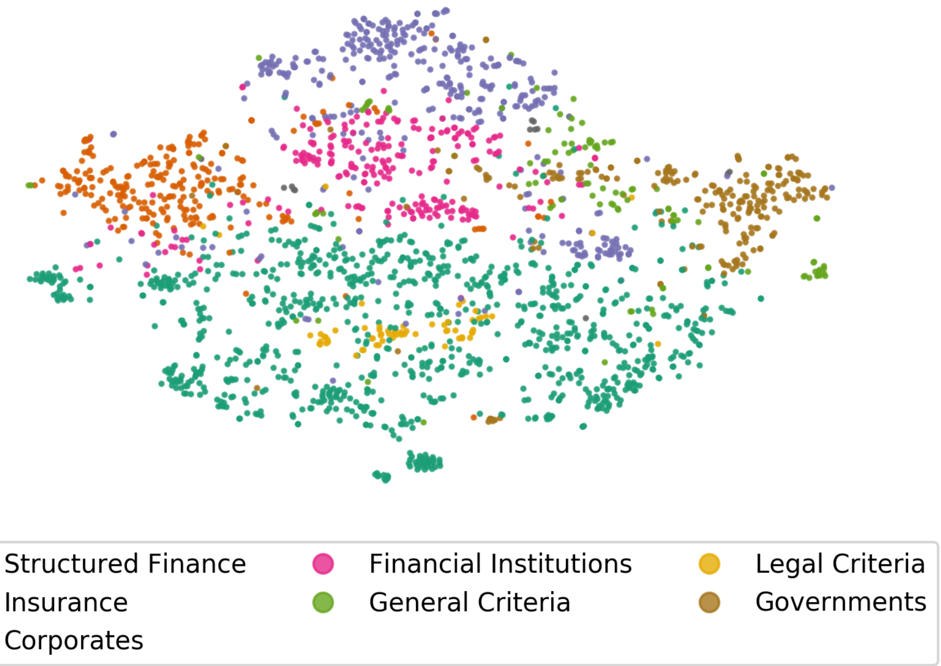
Linear Embeddings



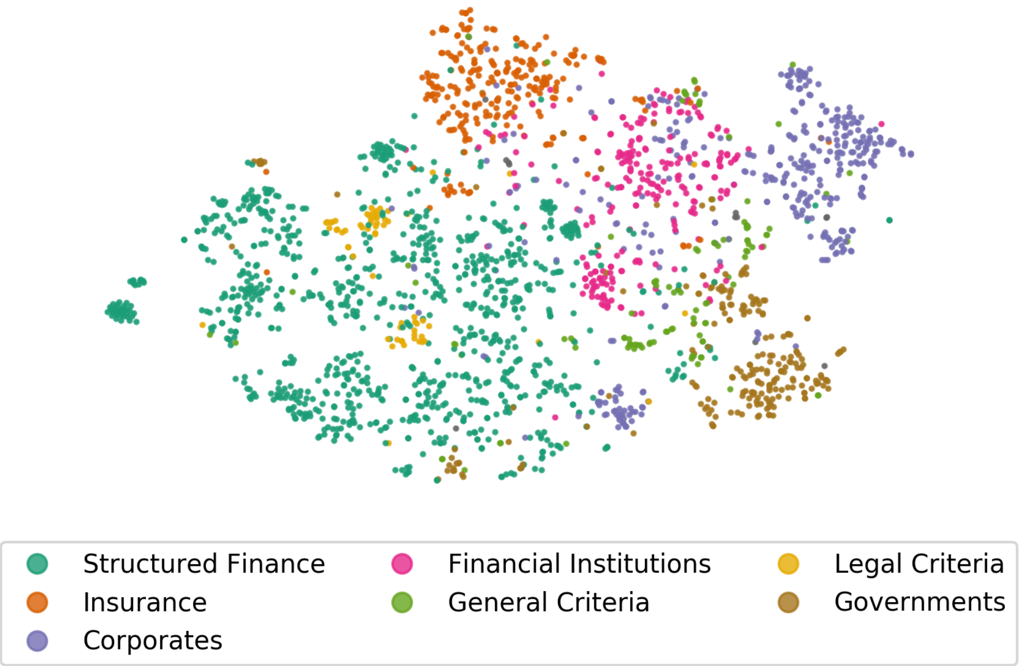
[Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

Visual: t-SNE Embeddings

Graph Layer Embeddings



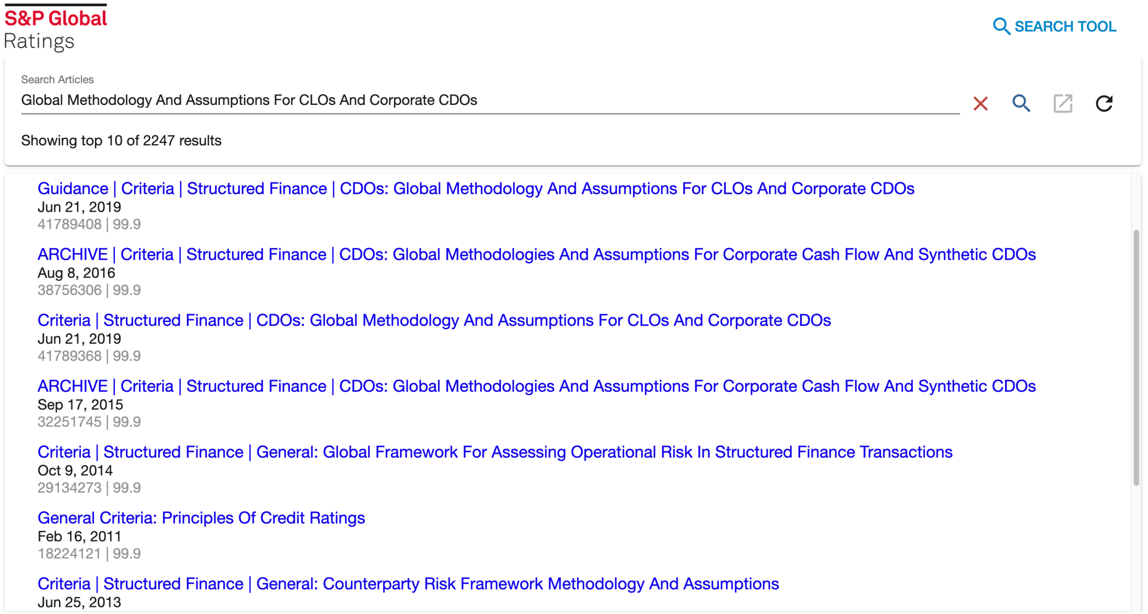
Bilinear Embeddings



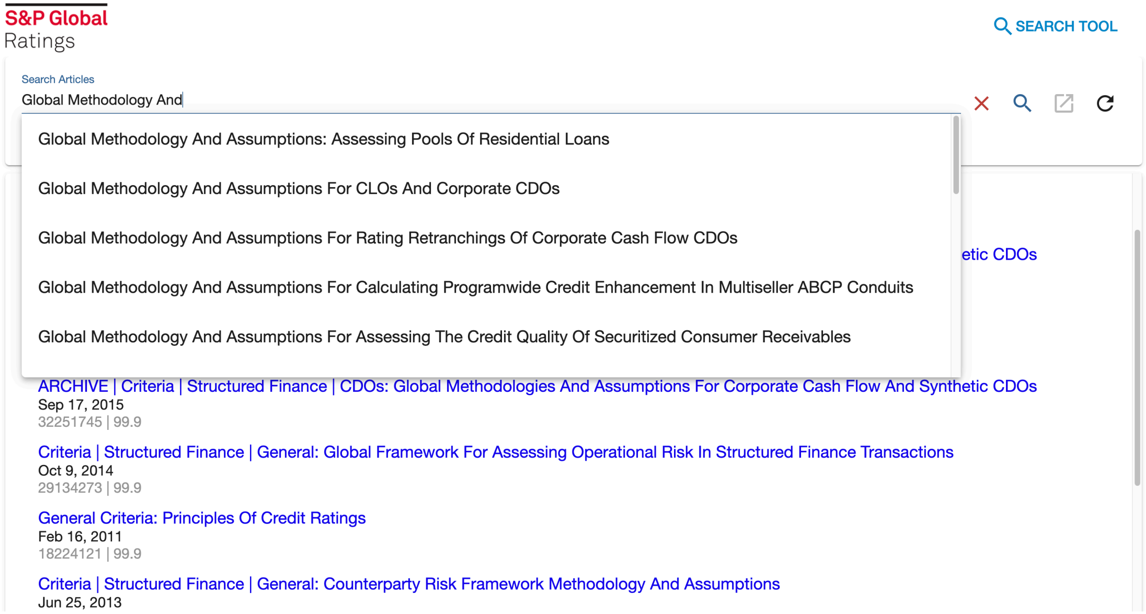
[Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

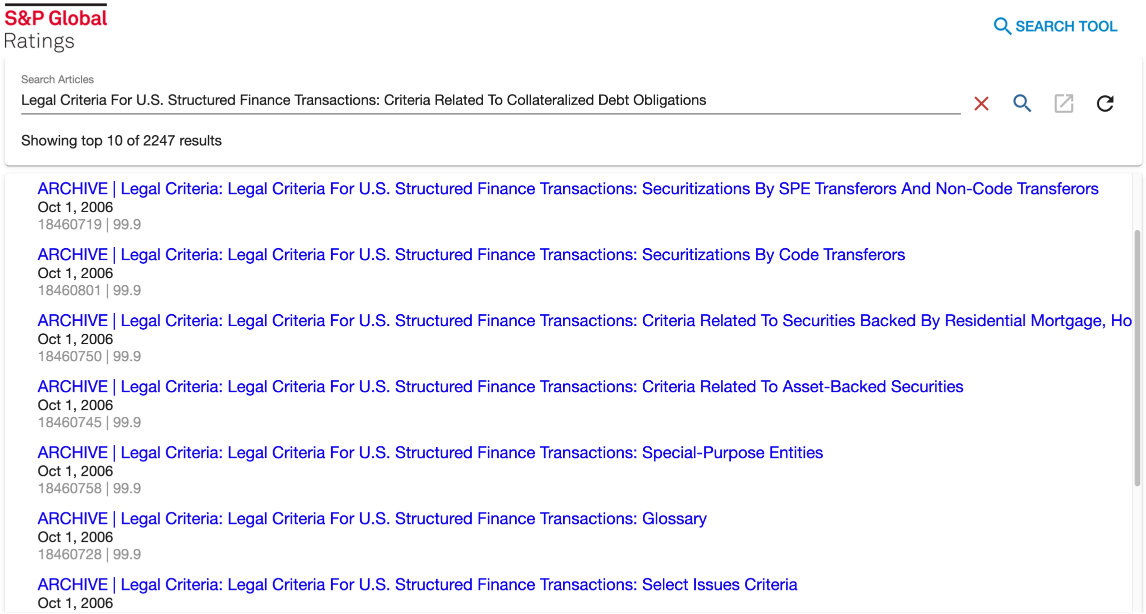
System Components

- **Frontend:**
  - Angular interface
  - Autocomplete Search Bar
  - Ranked-ordered list of relevant citations
- **Backend:**
  - RESTful Service
  - Accepts a Criteria ID request
  - Performs lookup with our learned model
  - Returns relevant metadata to frontend



System Components: User Interface Demo





Conclusion & Future Work

- **SPAGAttE** is a tool to model our criteria citation graph and perform link prediction with the explicit task of *recommending* citations
- Good case study for applying graph neural networks
- We hope to expand **SPAGAttE's** utility to:
  - *Business Relationships*
  - *Supply Chain Networks*

Project Links

- Github Source Code:
  – https://github.com/nextBillyonair/SPAGAttE
- Demo Video:
  – https://vimeo.com/394545958
  – Password: 55water

References

- [Bahdanau *et al.*, 2014] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate, 2014.
- [Clevert *et al.*, 2015] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus), 2015.
- [Kingma and Ba, 2014] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.
- [Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- [Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality, 2013.
- [Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
- [Veličković *et al.*, 2017] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2017.
- [Yang *et al.*, 2014] Bishan Yang, Wentau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases, 2014.
- [Yang *et al.*, 2016] Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings, 2016.

Thank You!