

Research Paper Recommendations



Team Members: Sharath Giri, Judy Fang, Jerry Jiang, Jacky Jiang, James Murphy, & Andres Villada
Mentors: Joao Pedro Mattos, Arko Barman

Motivation

Number of papers over years



Fig 1: Published papers ahead of annual NeurIPS conference

- Exponential increase in scientific literature
- Increasing difficulty in finding relevant literature
- More efficient recommendation system needed

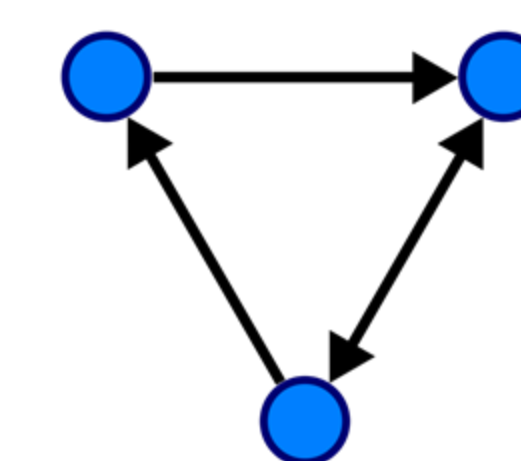
Model Exploration

Potential Models



Keyword Search

- Strength: Finding relevant data with few keywords
- Limitation: Terms are too broad and limiting for researcher exploration



Research Paper Similarity

- Strength: Compares entire papers
- Limitation: Papers contain many subjects of interest. Difficult to seek information on singular subject

Objective

Implement a **graph-based paper recommendation system** based on **contextual citation**

Data Overview

Source Paper Metadata



Source Citation Text



Cited Paper Title



Target Paper Metadata



Fulltext PeerRead (AI Peer Reviews)

Papers Cited: 3,693
Publish Years: 2007-2017
Citations: 16,669

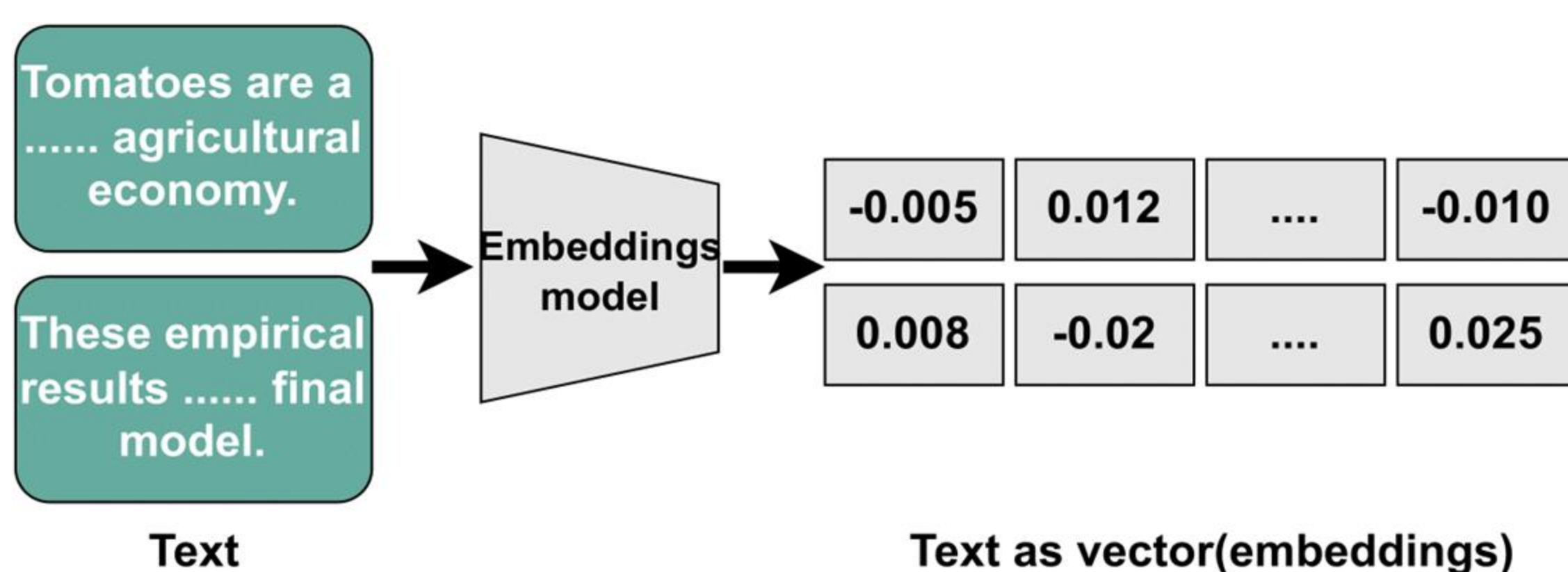
ArXivCS (CS Research Papers)

Papers Cited: 6,236
Publish Years: 1904-2017
Citations: 27,823

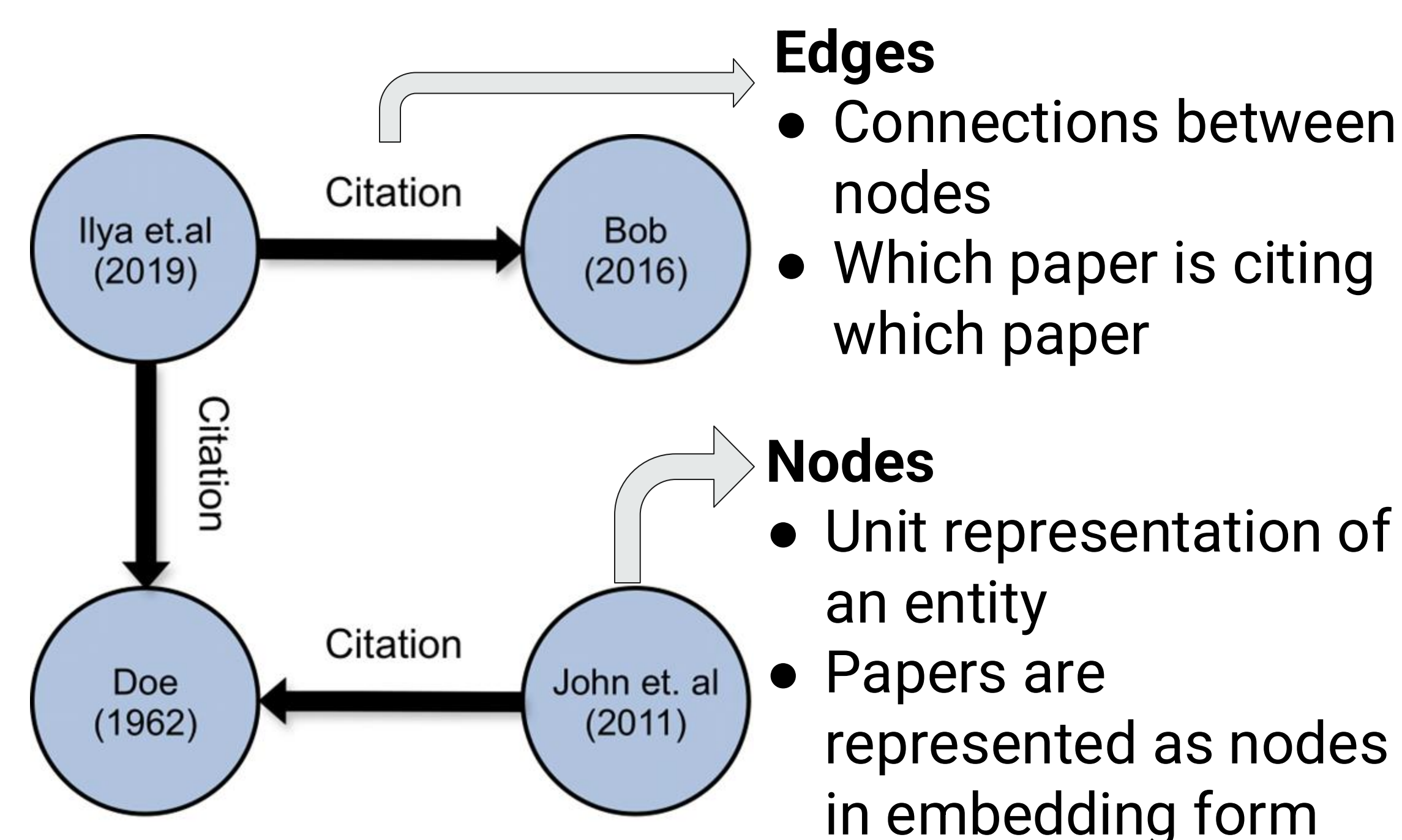
Key Concepts

Embeddings

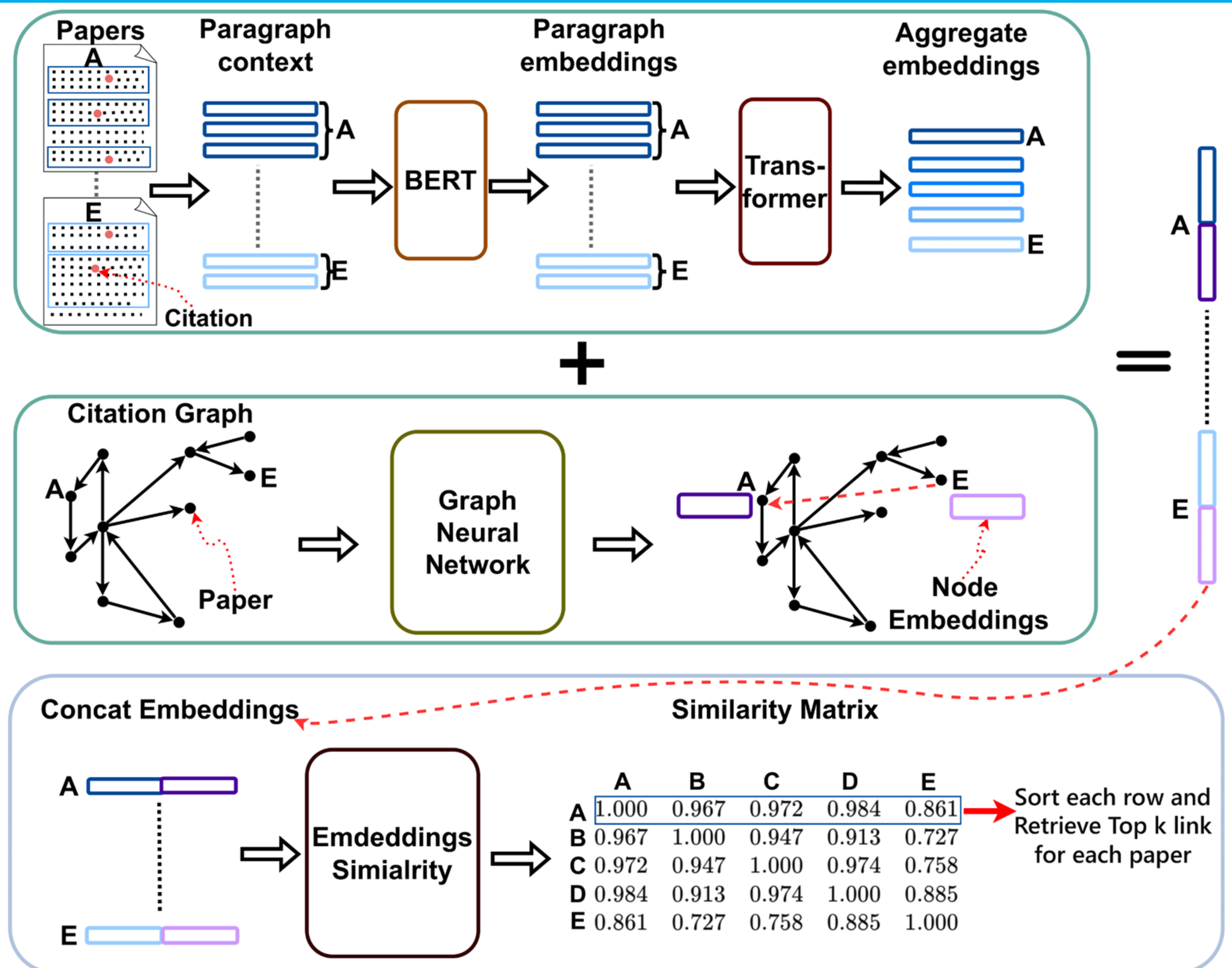
- Lower-dimensional representation of input data
- Captures meaningful relationships and patterns, enabling more effective processing.



Graphs



Model Architecture



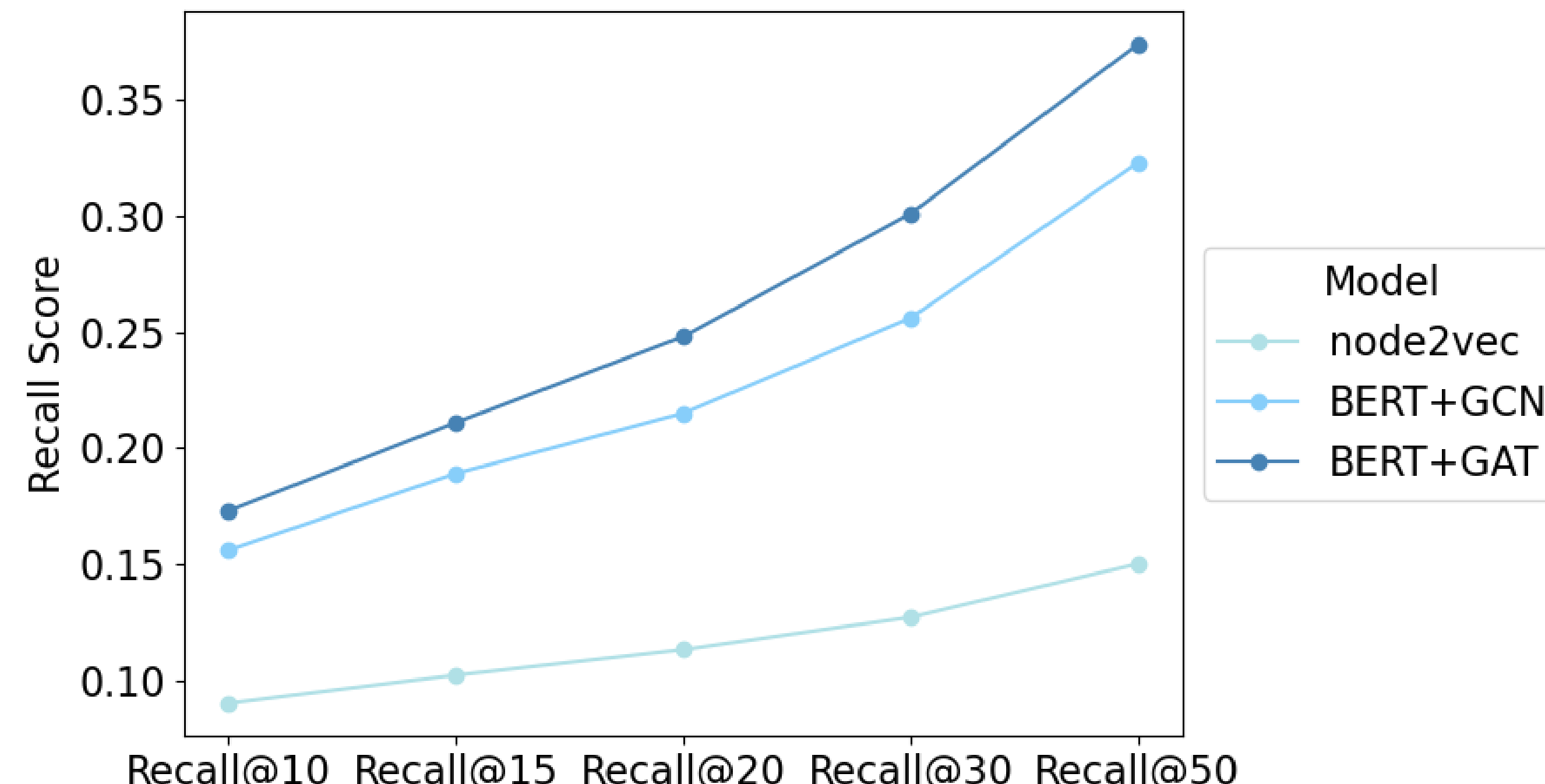
Evaluation Metrics & Results

Recall Scores for Different Models

$$Recall@k = \frac{true\ positives@k}{(true\ positives@k) + (false\ negatives@k)}$$

Recall@1 = $\frac{1}{3}$

Recall@3 = $\frac{2}{2+1} = \frac{2}{3}$



Conclusion

- BERT + GAT architecture produces the best Recall scores followed by BERT + GCN

Key Notes

- BERT (Bidirectional Encoder Representations from Transformers)
- GCN (Graph Convolutional Network)
- GAT (Graph Attention Network)
- Recall@K: K is the number of recommendations