IEEE PACIFIC VIS 2021

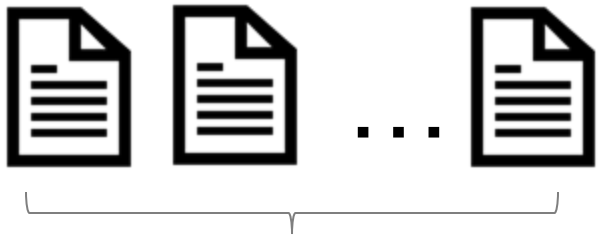
KeywordMap: Attention-based Visual Exploration for Keyword Analysis

Yamei Tu, *Jiayi Xu, Han-Wei Shen*

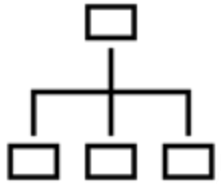
GRAphics & **VI**sualization s**T**ud**Y** Research Group, The Ohio State University

IEEE 天津大学 Tianjin University

Introduction



Document Collections



Document Categorization

Summarization

Indexing

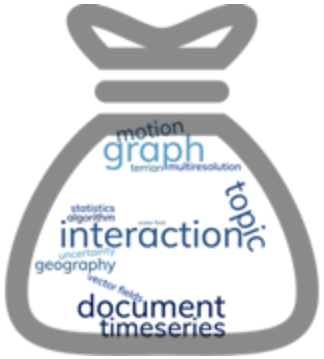
Topic Modeling

Document Collections

Number of Topics

?

Topic Modeling



Topic 1

cat, dog, bird

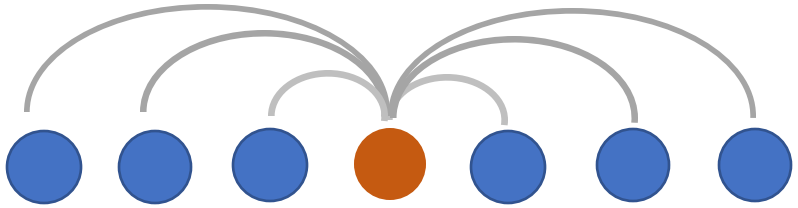
?

Topic 2

apple, peach, orange

Challenges

- Utilizing Word Context Information



*"He cashed a check at the **bank.**"*

*"The hotel is located on the **bank** of the river."*

- Identifying the correlated keywords

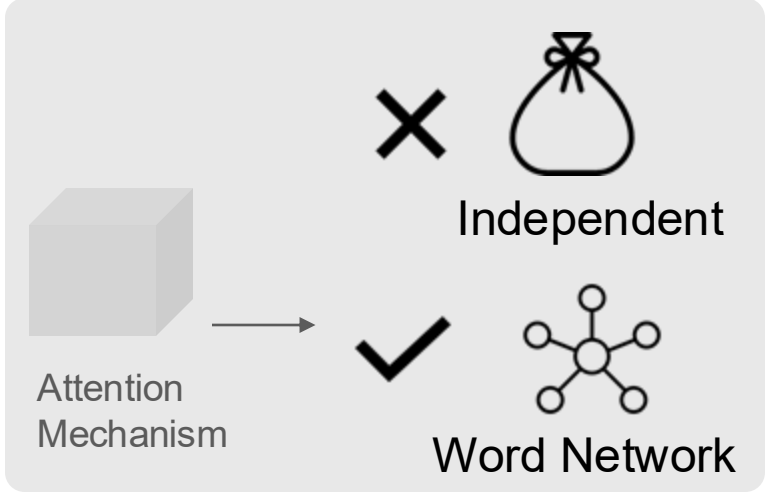
- Determining the number of topics

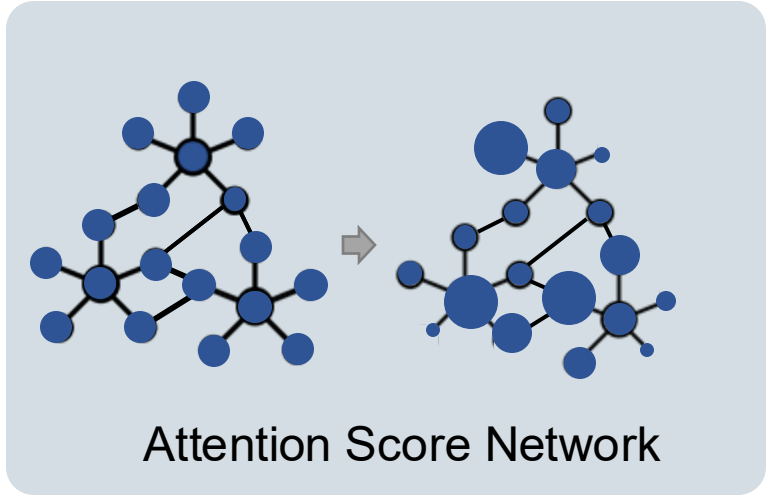
Method

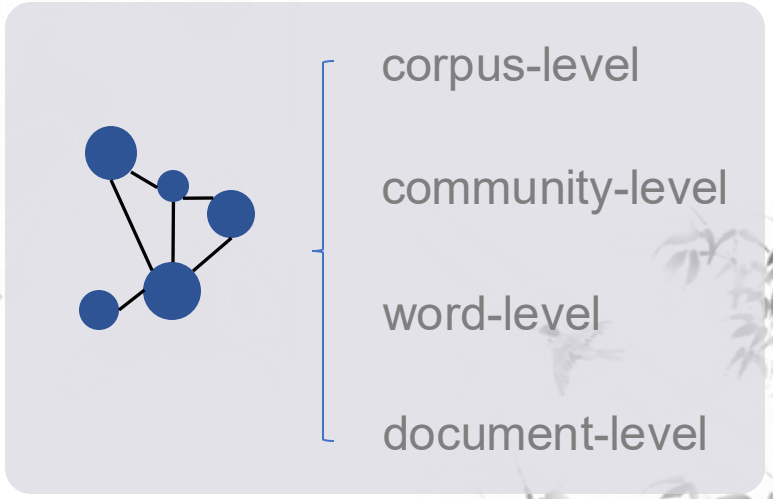
Challenge1: Utilize word context

Challenge2: Identify Keywords

Challenge3: Identify Topics







Building Attention-based Word Network

Attention-based Word Influence Algorithm

KeywordMap: Interactive Visual System

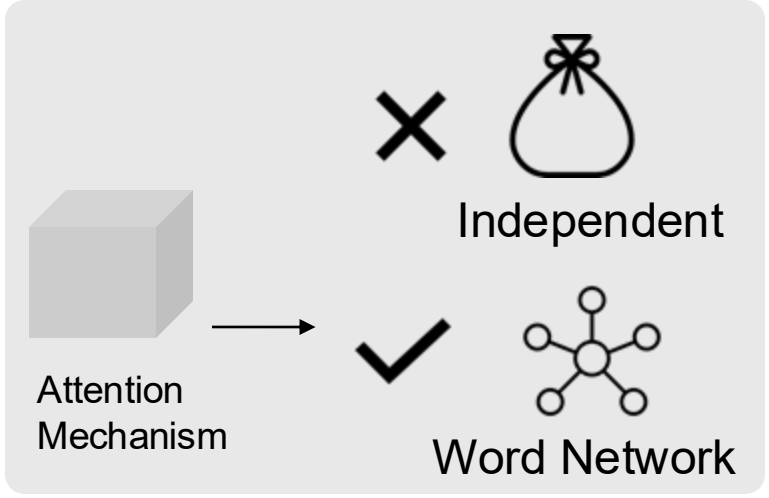
IEEE PACIFIC VIS 2021

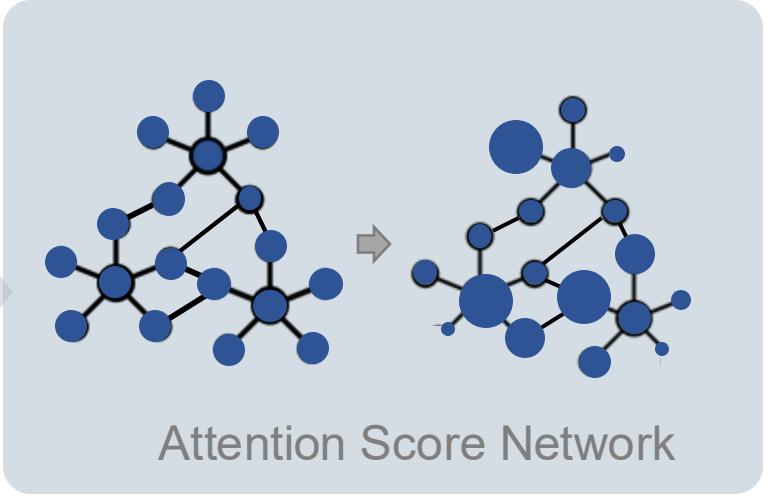
Method

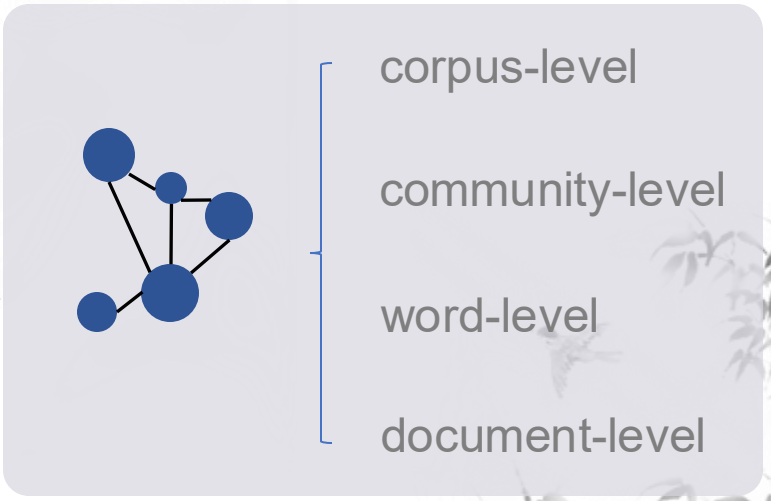
Challenge1: Utilize word context

Challenge2: Identify Keywords

Challenge3: Identify Topics







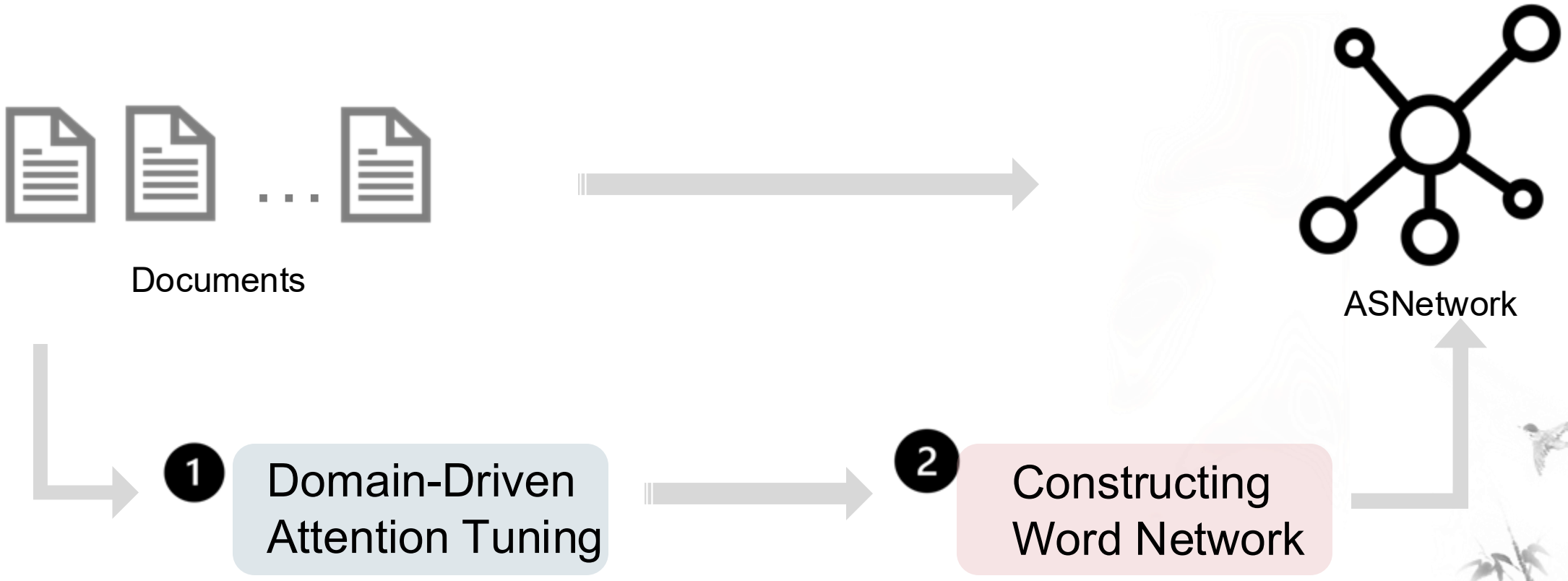
Building Attention-based Word Network

Attention-based Word Influence Algorithm

KeywordMap: Interactive Visual System

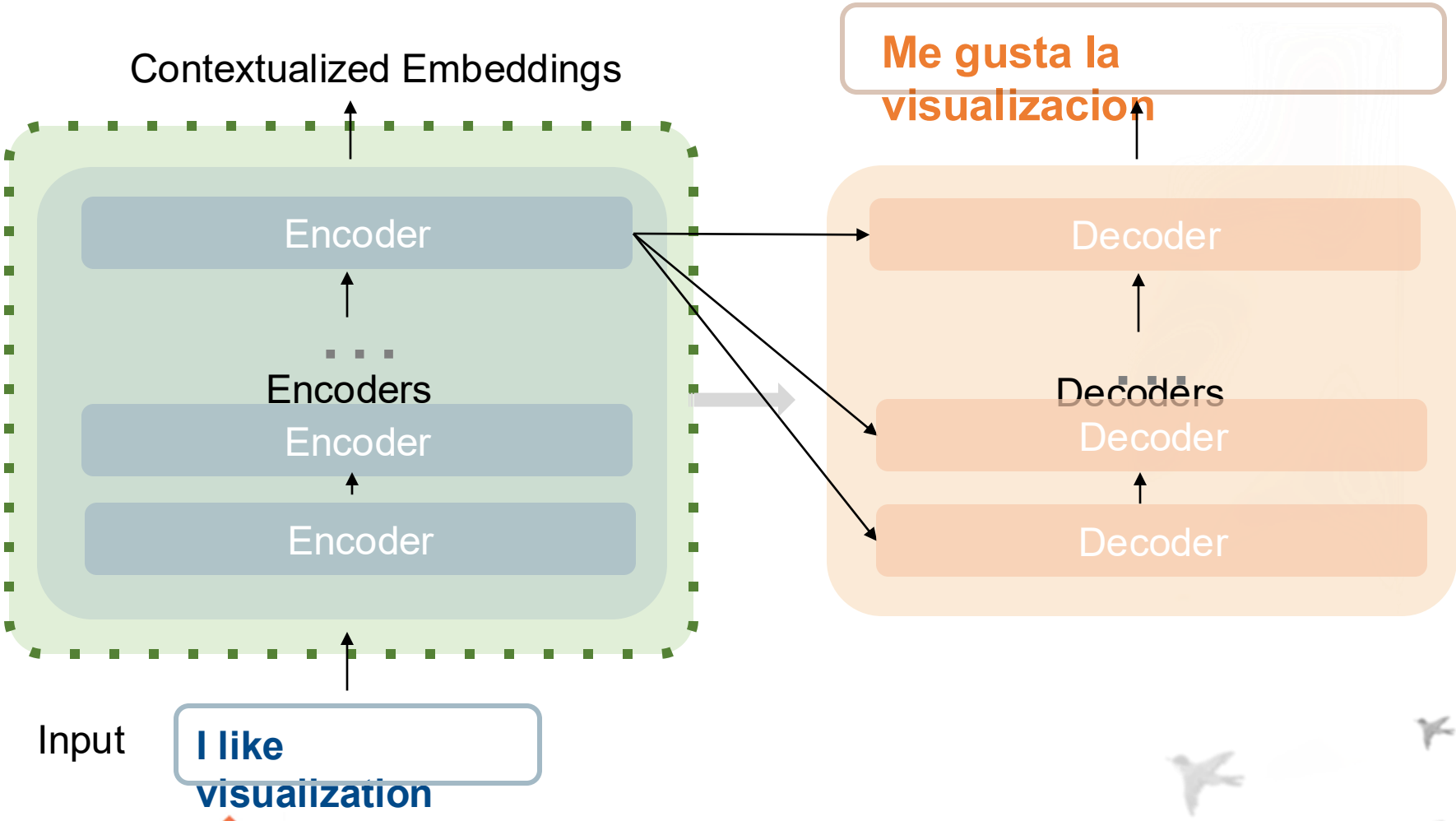
IEEE PACIFIC VIS 2021

Building Attention-based Word Network



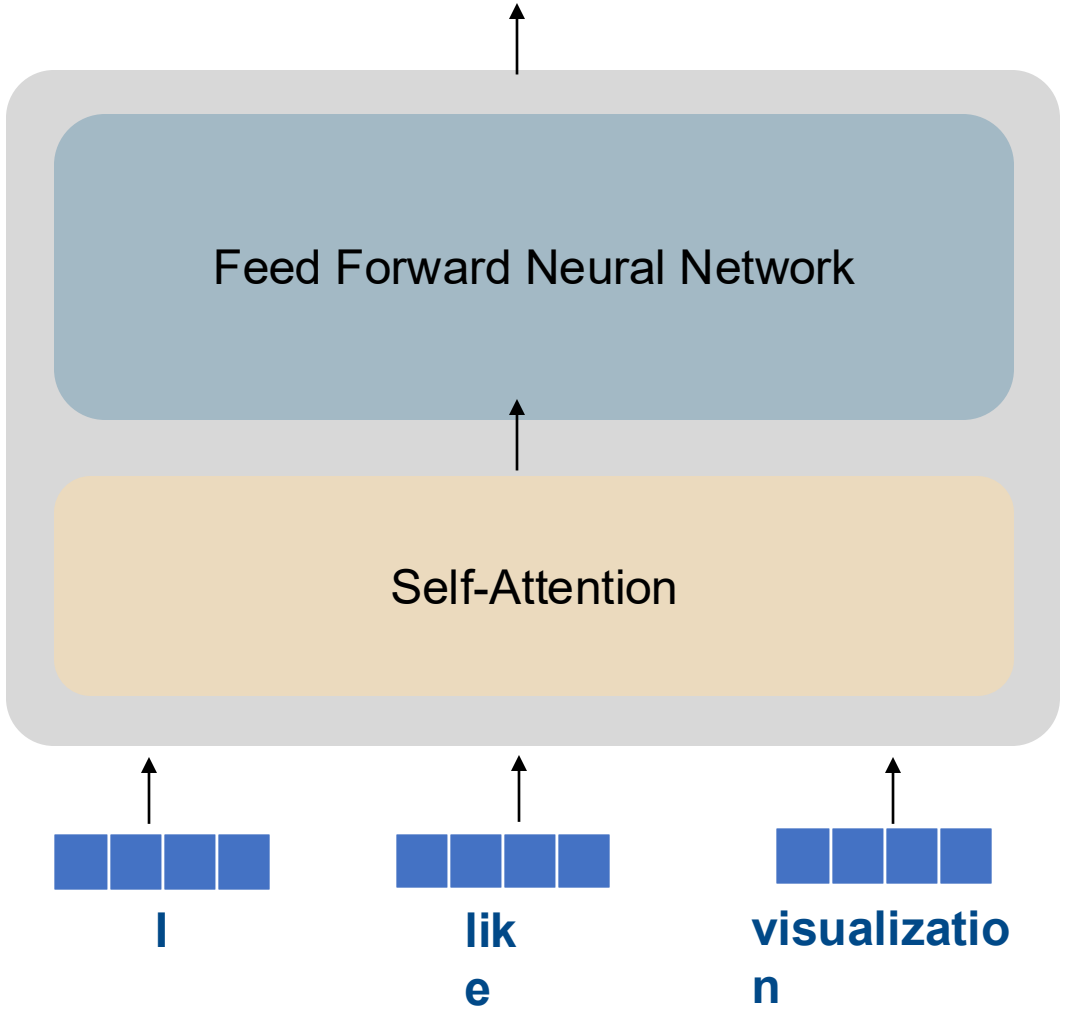
Transformer

1 Domain-Driven Attention Tuning → 2 Constructing Word Network



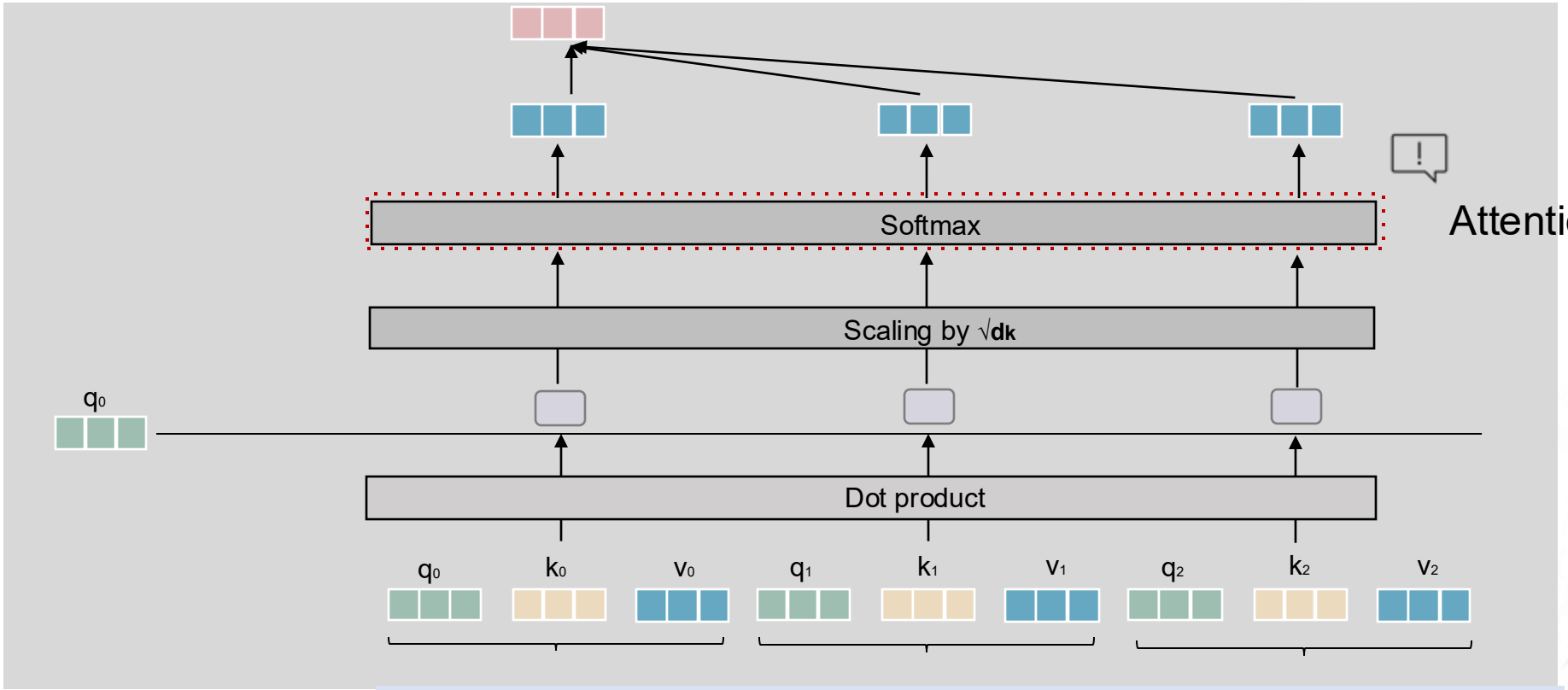
Encoder

1 Domain-Driven Attention Tuning → 2 Constructing Word Network



Self-Attention

1 Domain-Driven Attention Tuning → 2 Constructing Word Network

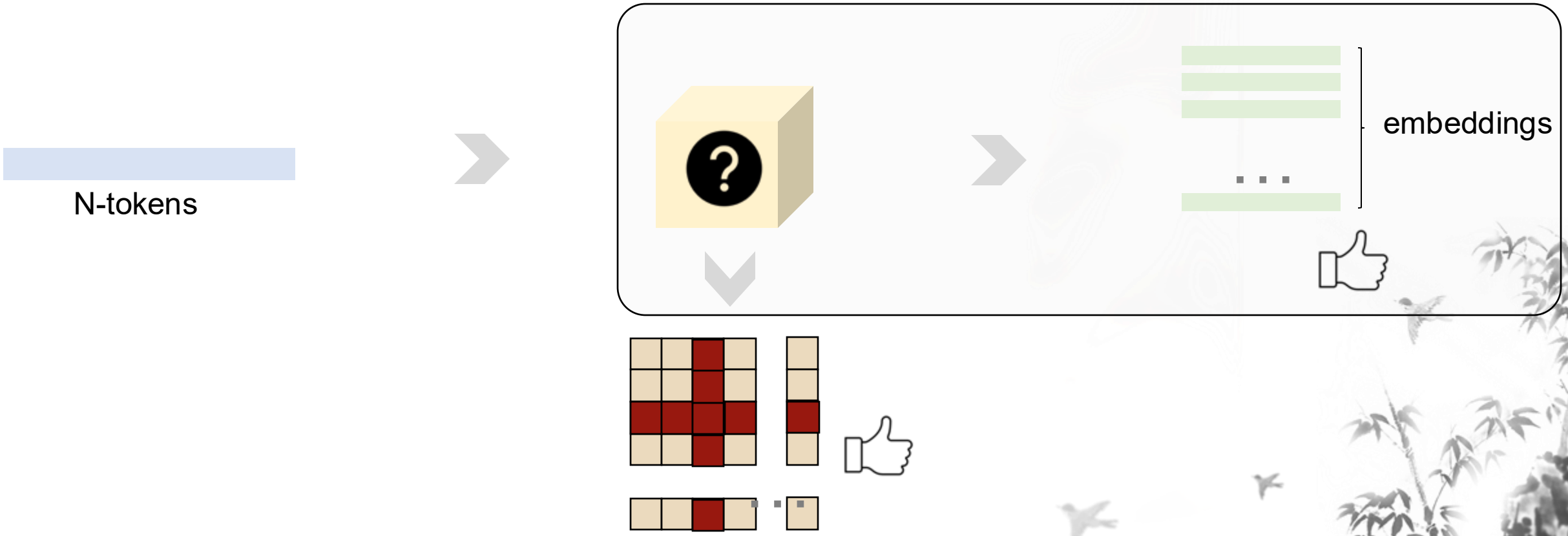


Attention Map

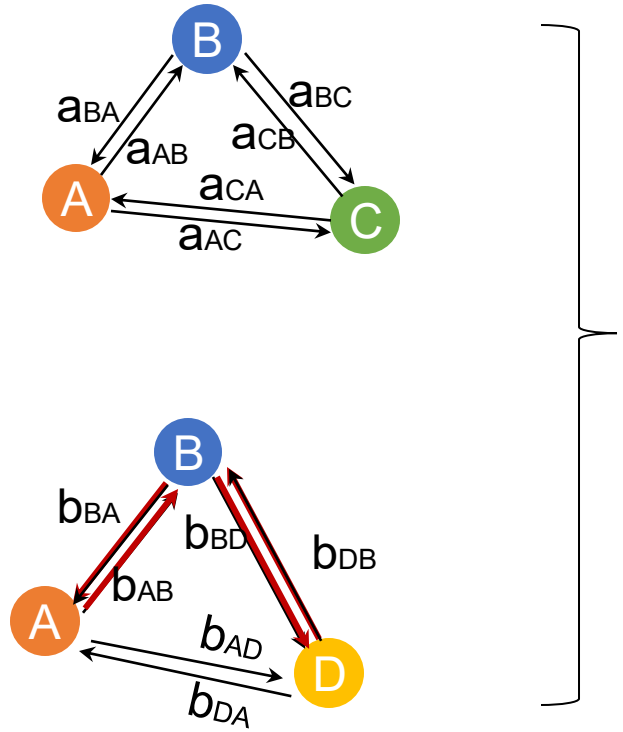
https://towardsdatascience.com/illustrated-self-attention-2d627e33b20a

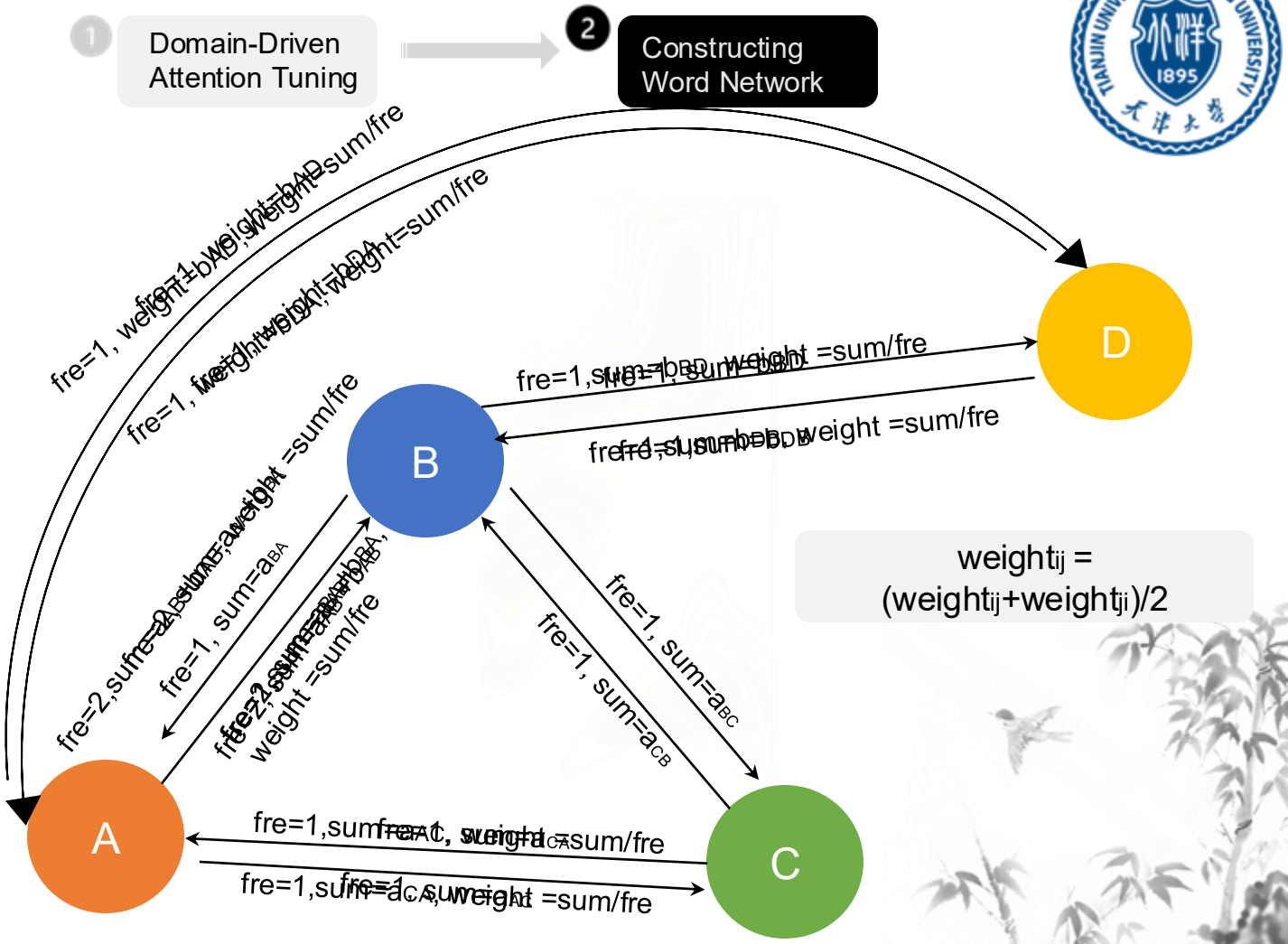
Domain-Driven Attention Tuning

1 Domain-Driven Attention Tuning → 2 Constructing Word Network



Constructing ASNetwork



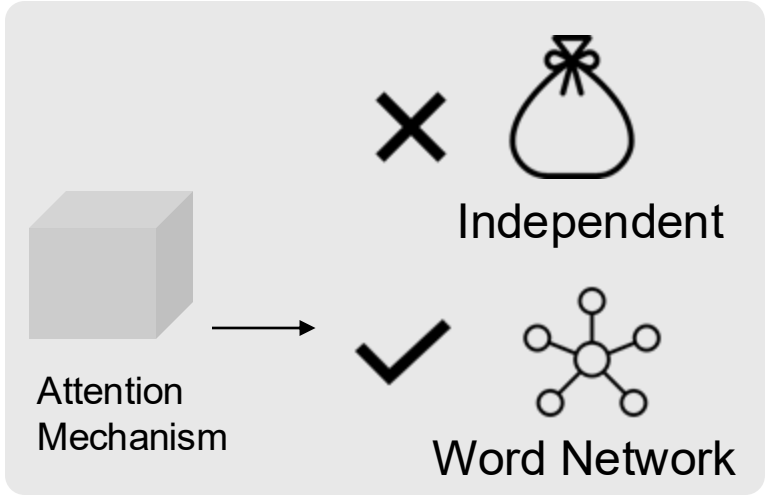


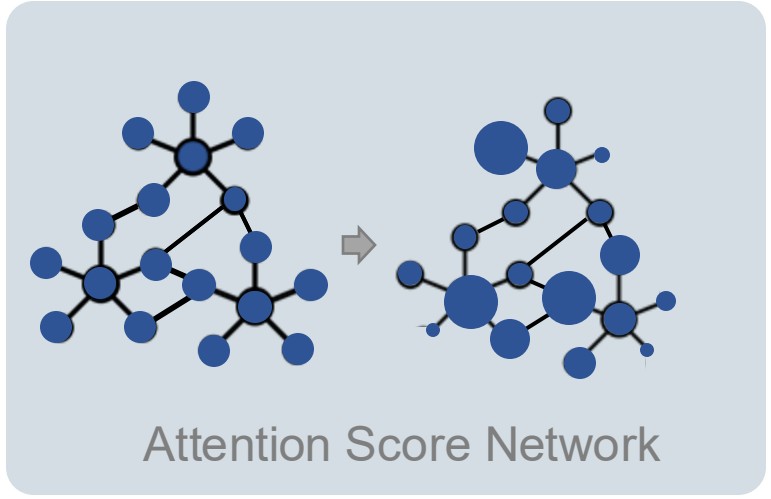
Method

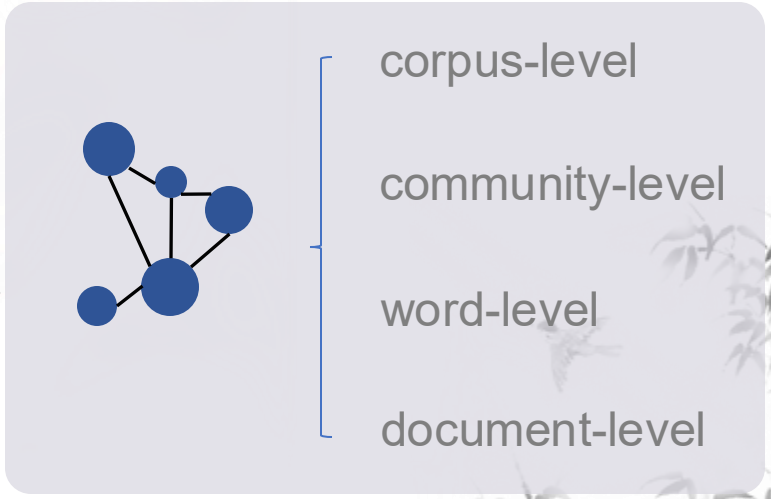
Challenge1: Utilize word context

Challenge2: Identify Keywords

Challenge3: Identify Topics







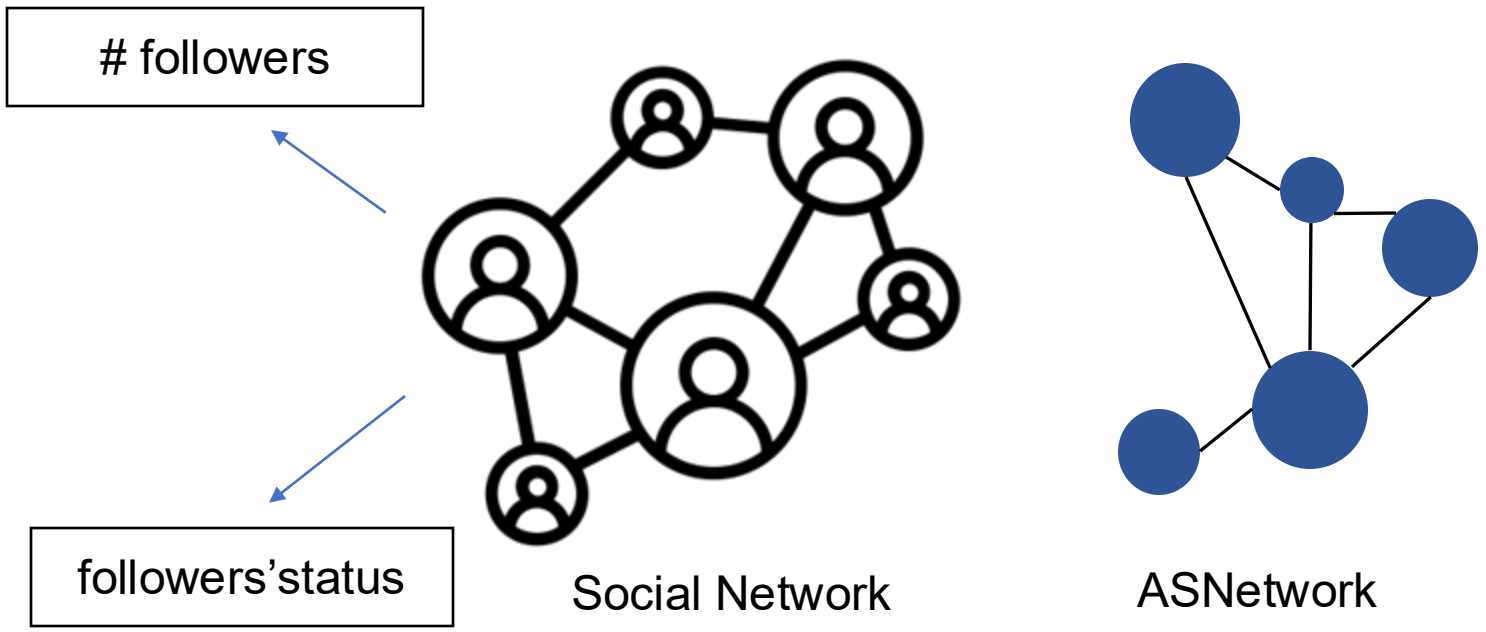
Building Attention-based Word Network

Attention-based Word Influence Algorithm

KeywordMap: Interactive Visual System

IEEE PACIFIC VIS 2021

Attention-based Word Influence Algorithm



Attention-based Word Influence Algorithm

Algorithm 2: Attention-based Word Influence

Input: ASNetwork

Output: Influence I_i for each word $v_i \in V$.

1 $e_{k \to i}$: the edge weight, denoting how much attention is transitioned from $word_k$ to $word_i$.

2 **repeat**

  for $v_i \in V$ **do**

    $I_i = \sum_{(k \to i \in G)} e_{k \to i} \times I_k$

    $\text{softmax}(I_i) = \frac{e^{I_i}}{\sum_{j=1}^{N} e^{I_j}}$

  end

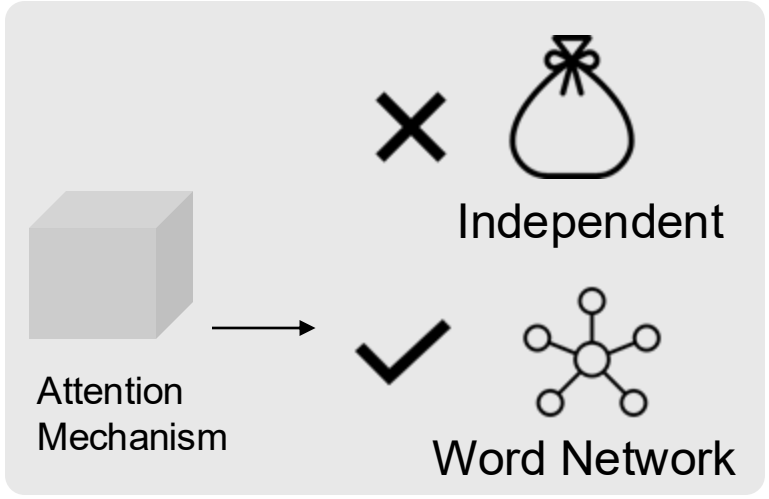
until I_i *converges*;

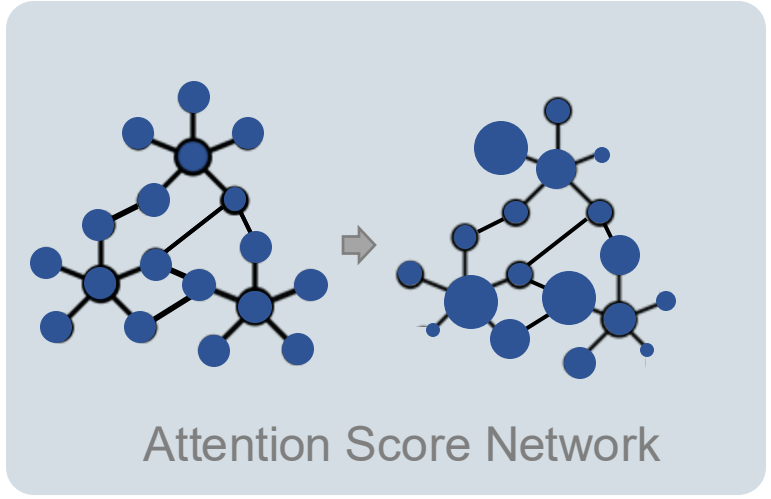
Method

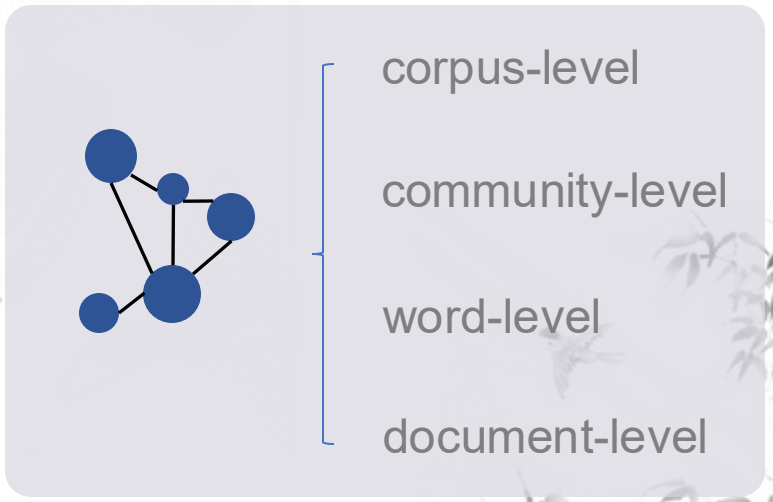
Challenge1: Utilize word context

Challenge2: Identify Keywords

Challenge3: Identify Topics





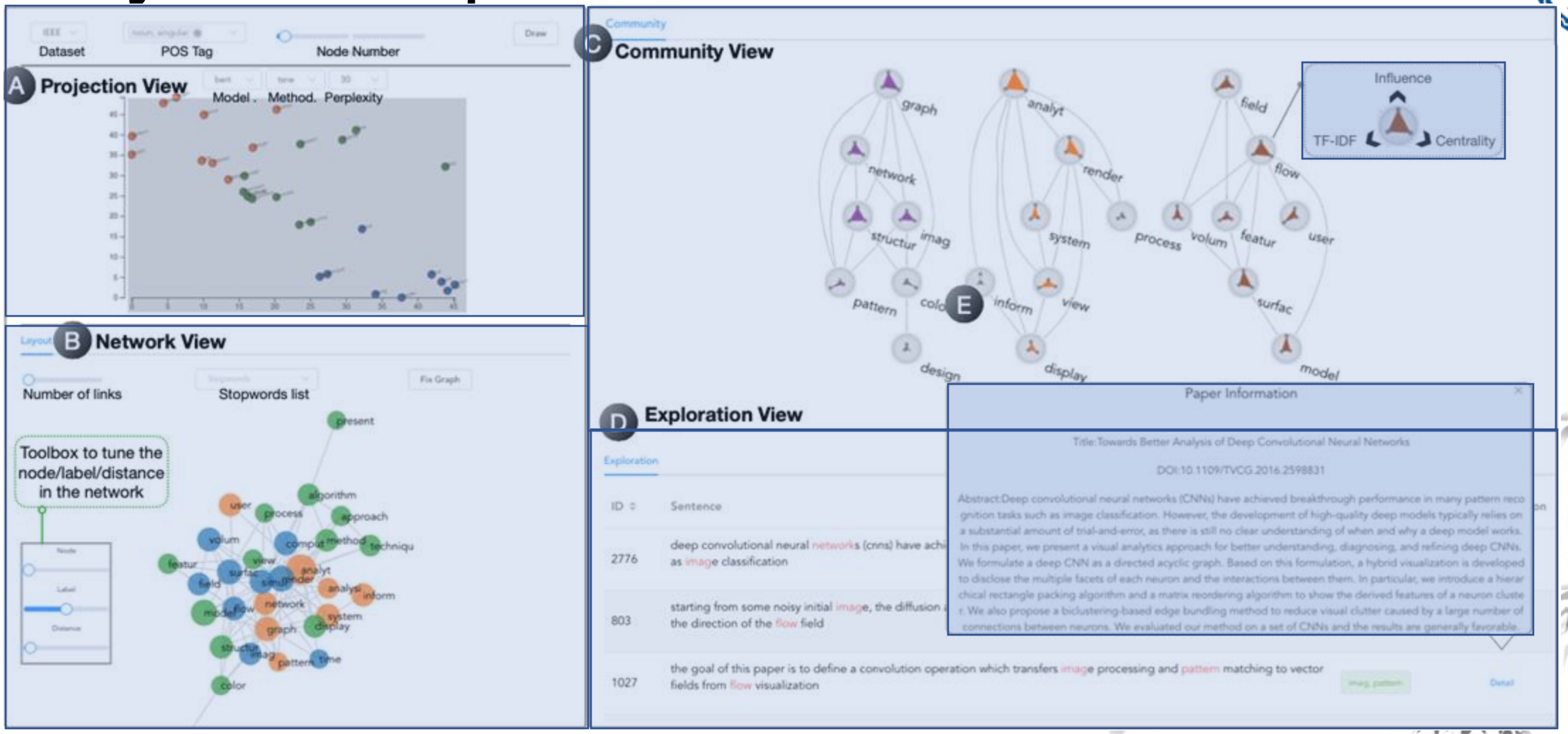


Building Attention-based Word Network

Attention-based Word Influence Algorithm

KeywordMap: Interactive Visual System

KeywordMap


Dataset
POS Tag
Node Number
A Projection View
Model . Method. Perplexity
B Network View
Number of links
Stopwords list
Toolbox to tune the node/label/distance in the network
C Community View
Influence
TF-IDF
Centrality
graph
network
structur
imag
pattern
design
analyt
render
system
display
field
flow
process
volum
featur
surfac
user
model
E
D Exploration View
Paper Information
Title:Towards Better Analysis of Deep Convolutional Neural Networks
DOI:10.1109/TVCG.2016.2598831
Abstract:Deep convolutional neural networks (CNNs) have achieved breakthrough performance in many pattern recognition tasks such as image classification. However, the development of high-quality deep models typically relies on a substantial amount of trial-and-error, as there is still no clear understanding of when and why a deep model works. In this paper, we present a visual analytics approach for better understanding, diagnosing, and refining deep CNNs. We formulate a deep CNN as a directed acyclic graph. Based on this formulation, a hybrid visualization is developed to disclose the multiple facets of each neuron and the interactions between them. In particular, we introduce a hierarchical rectangle packing algorithm and a matrix reordering algorithm to show the derived features of a neuron cluster. We also propose a biclustering-based edge bundling method to reduce visual clutter caused by a large number of connections between neurons. We evaluated our method on a set of CNNs and the results are generally favorable.
ID
Sentence
2776
deep convolutional neural networks (cnns) have achi
as image classification
803
starting from some noisy initial image, the diffusion
the direction of the flow field
1027
the goal of this paper is to define a convolution operation which transfers image processing and pattern matching to vector fields from flow visualization

Evaluation

Documents

ASNetwork

Attention-based Word Influence Algorithm

KeywordMap System

1 Domain-Driven Attention Tuning

2 Constructing Word Network

Embedding + Attention Map

Co-occurrence Matrix

TextRank

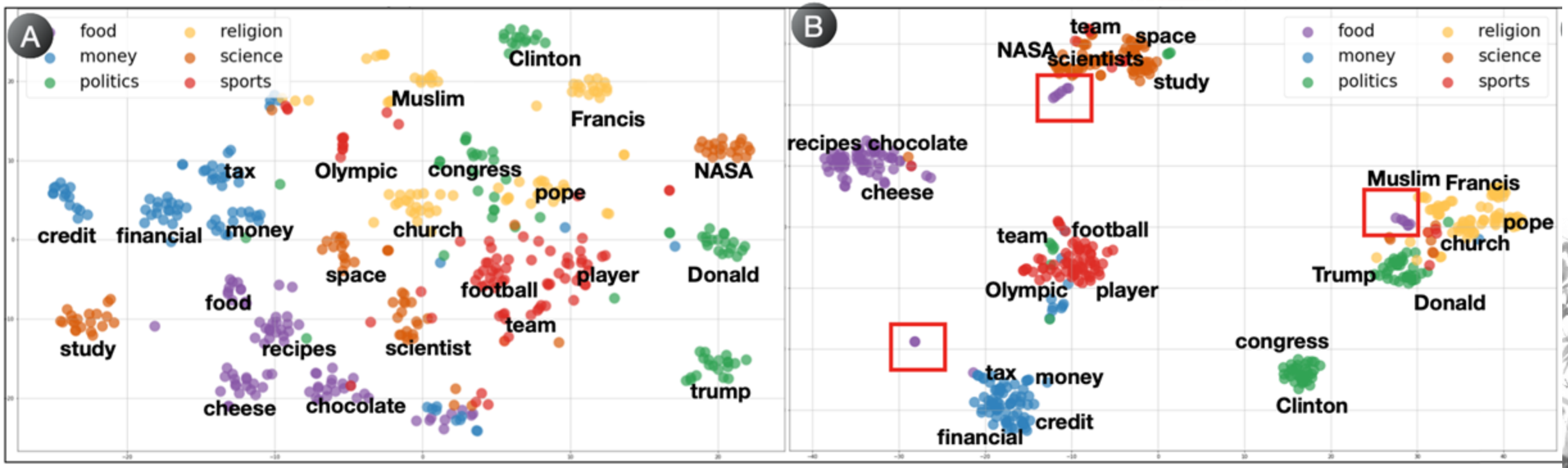
Tuning Details

- 3 Datasets
  - VisPubdata
  - The News (pre-select 6 fields)
  - Arxiv (pre-select 3 sub-categories under CS)
- 2 Models (BERT + XLNET)
- Hyper-parameters:
  - Learning rate = $2e-5$,
  - Batch-size = 32

Table 1: Epochs and performance of models trained in this paper.

Dataset	Model	Epochs	Valid. Acc.
VIS	BERT	4	0.78
VIS	XLNet	4	0.74
NEWS	BERT	2	0.88
NEWS	XLNet	1	0.84
ARXIV	BERT	2	0.91
ARXIV	XLNet	2	0.90

Clusters of Word Embeddings- Qualitative



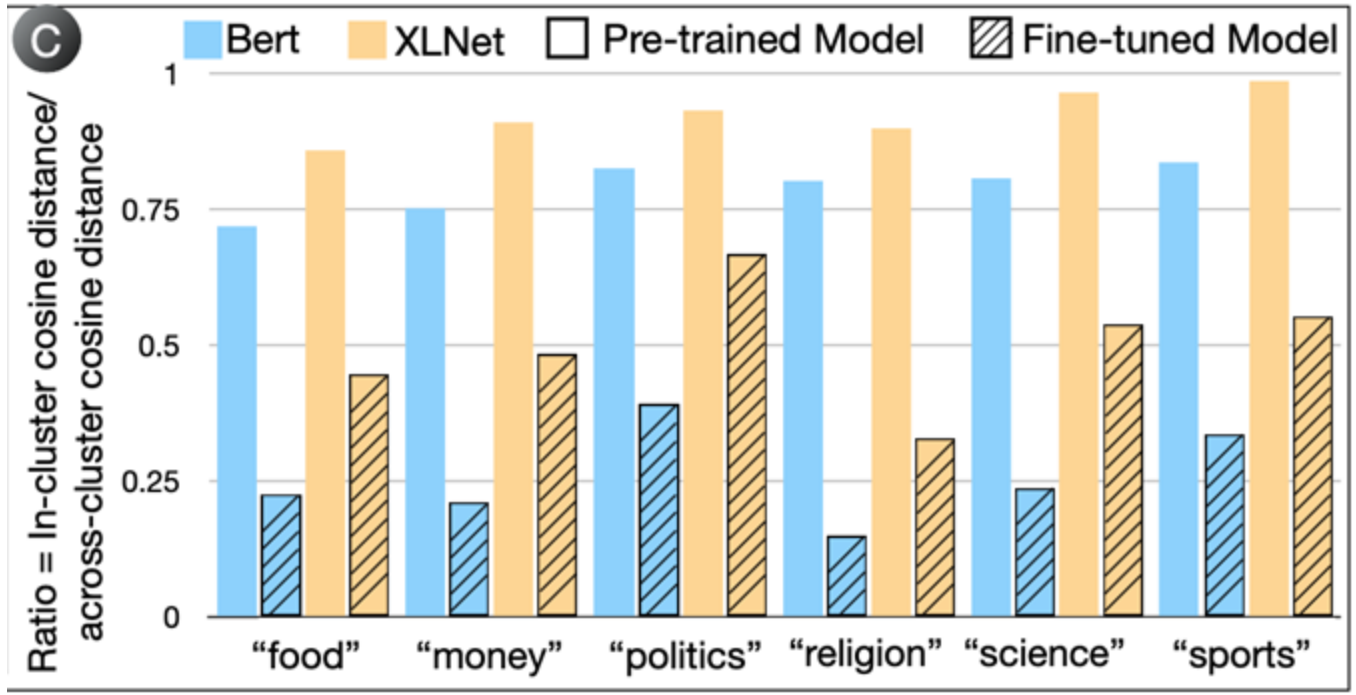
Before

After

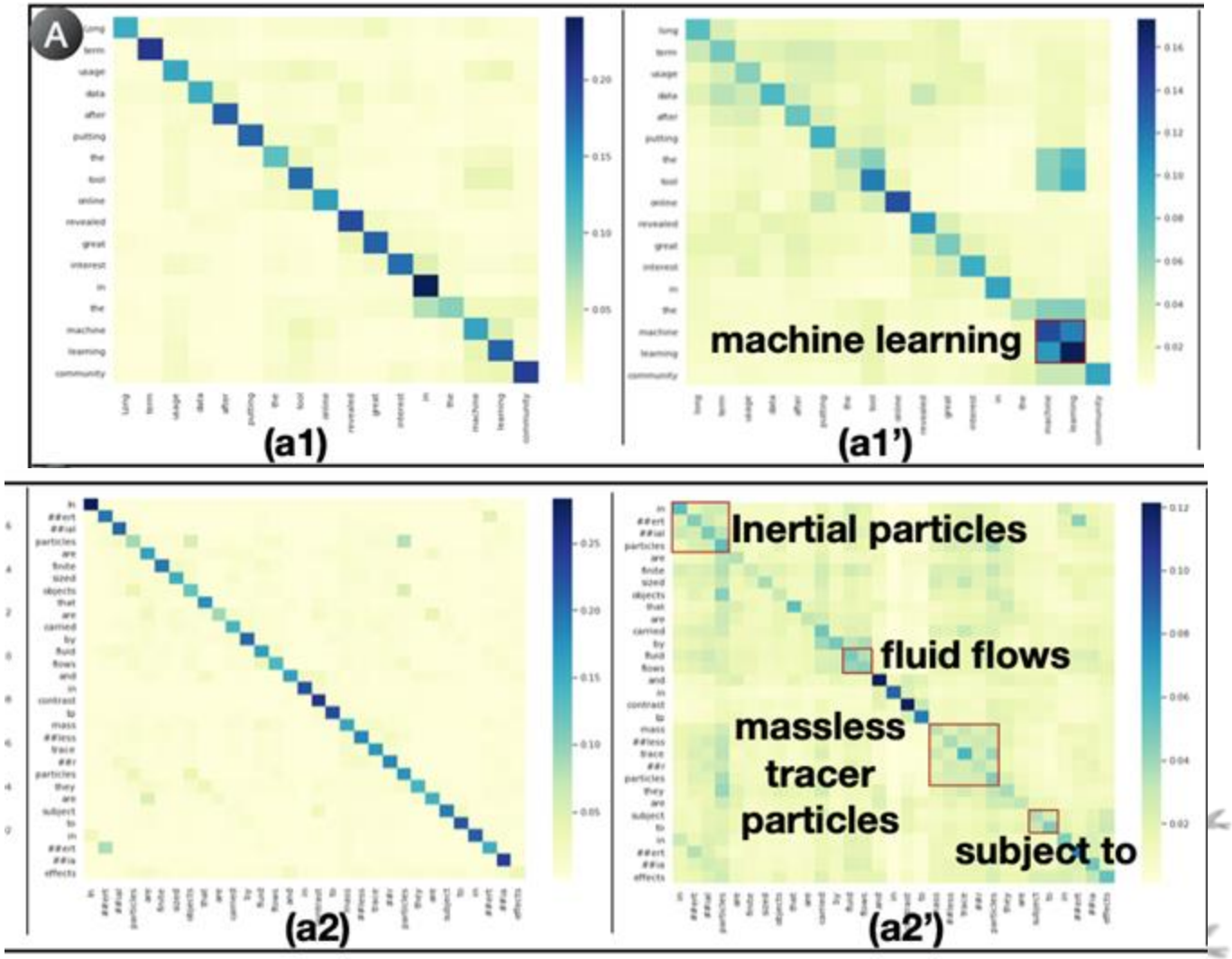
Clusters of Word Embeddings-Quantitative

- 6 topics, where each topic has 80(4×20) word vectors.

$$R_j = \frac{\sum_{v_i \in C_j} cos(v_i, In_j)}{\sum_{v_i \in C_j} cos(v_i, Out_j)} (j = 1, 2..6)$$



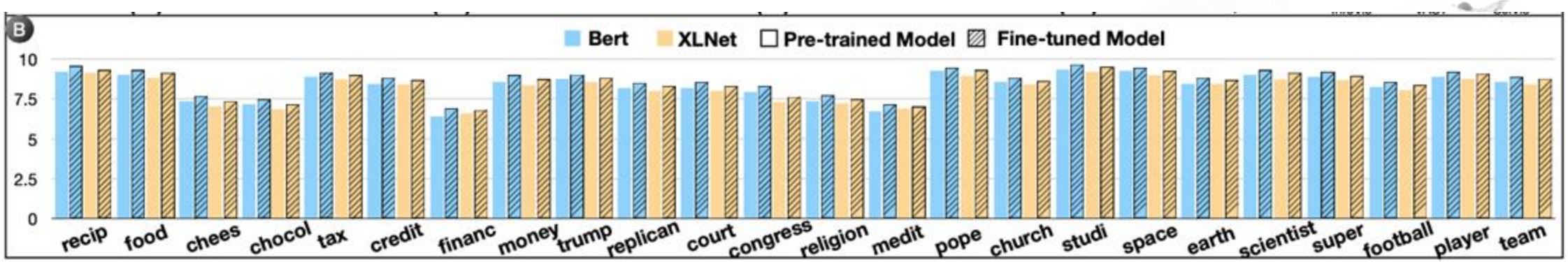
Pattern Change of Attention Maps



Pattern Change of Attention Maps

- Reason: Validate the pattern of the consolidated ASNetwork
- Step:
  - Convert to the matrix;
  - normalize the sum of outgoing attention to 1 in each row.
    [cell c_{ij} is considered as the probability of wordi attends to word j , defined as P(j|i)]

$$H(i) = -\sum_{j=1}^{|V|} P(j|i) log P(j|i)$$



Evaluating the ASNetwork

- co-occurrence matrix is constructed by computing how many times the two words are co-occurring in a sliding window.

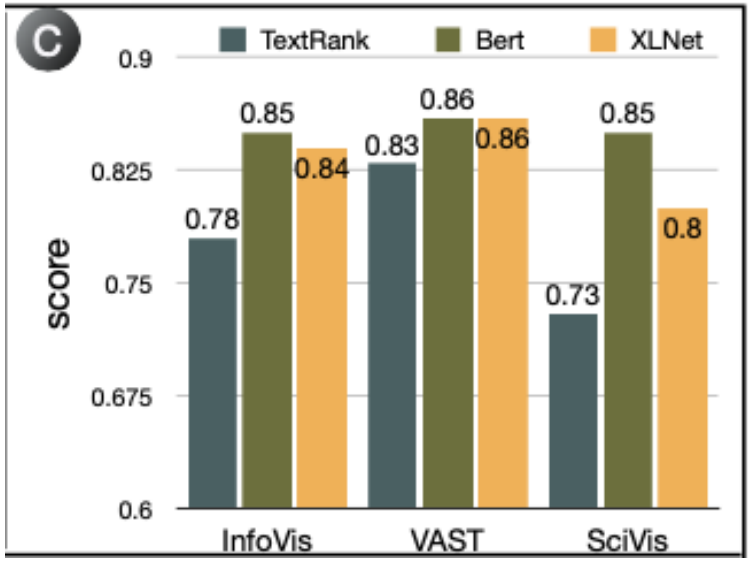
Table 2: Top10 Keywords with Different Methods

Method	Rank#1	Rank#2	Rank#3	Rank#4	Rank#5	Rank#6	Rank#7	Rank#8	Rank#9	Rank#10
Co-occurrence(window=2)	visual	data	techniqu	interact	user	present	method	system	analysi	approach
Co-occurrence(window=5)	visual	data	techniqu	interact	user	method	present	system	analysi	model
ASNetwork(XLNet+last layer)	visual	data	system	inform	network	graph	cluster	analysi	user	tree
ASNetwork(BERT+last layer)	data	visual	analyt	model	render	surfac	graph	system	comput	volum

Evaluating **the AWI Algorithm**

- TextRank is a popular method based on Google's PageRank algorithm.
- Normalized Discounted Cumulative Gain:

$$DCG_K = \sum_{i=1}^{K} \frac{2^{rel_i} - 1}{log_2(i+1)} \qquad NCDG_K = \frac{DCG_K}{IDCG_K}$$



Evaluating Hyper Parameters

- Hyper-parameters
  - Model type: BERT + XLNet
  - Constructing ASNetwork:
    - last encoder
    - averaging over all encoders

Table 2: Top10 Keywords with Different Methods

Method	Rank#1	Rank#2	Rank#3	Rank#4	Rank#5	Rank#6	Rank#7	Rank#8	Rank#9	Rank#10
ASNetwork(XLNet+last layer)	visual	data	system	inform	network	graph	cluster	analysi	user	tree
ASNetwork(BERT+last layer)	data	visual	analyt	model	render	surfac	graph	system	comput	volum
ASNetwork(XLNet+all layers)	data	method	model	techniqu	system	algorithm	inform	analysi	approach	structur
ASNetwork(BERT+all layers)	data	model	surfac	system	method	user	algorithm	volum	techniqu	visual

Case Study

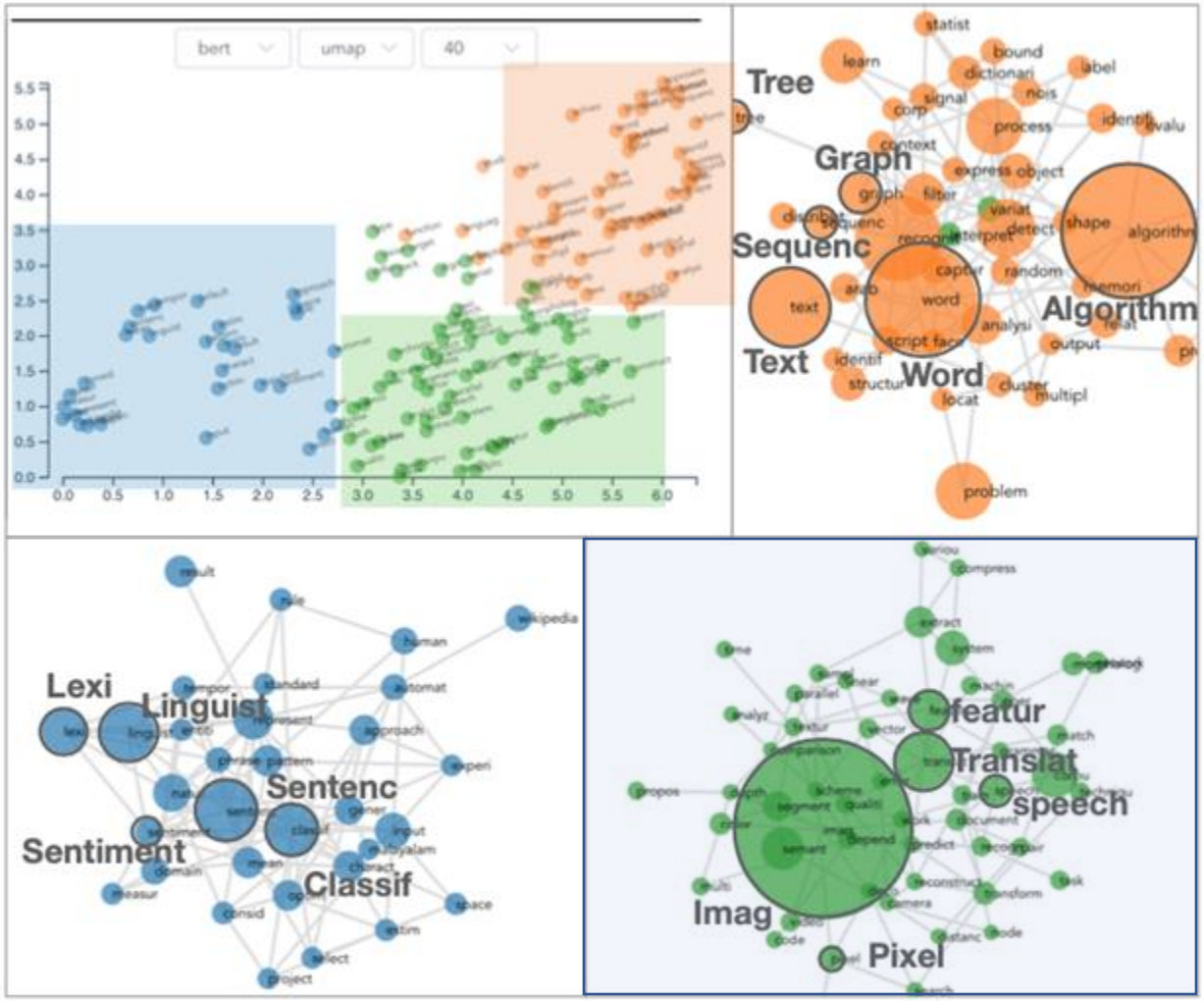


Figure 5: Brushing each clustering to see the local keyword structure in the *network* view.

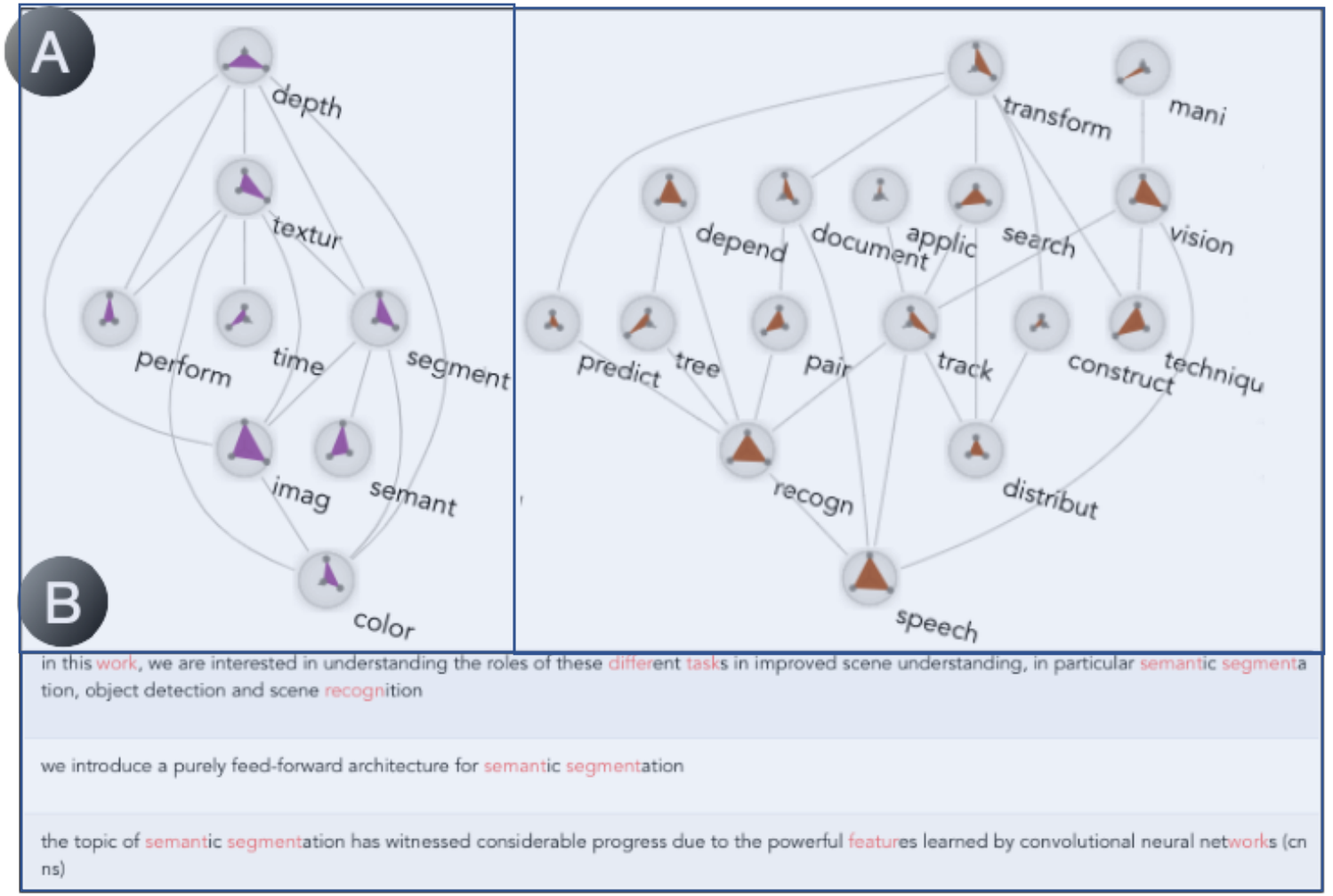
Data structure and Algorithm

Computation and language

Computer vision

IEEE PACIFIC VIS 2021

Case Study


A
depth
textur
perform
time
segment
imag
semant
color
B
transform
mani
depend
document
applic
search
vision
predict
tree
pair
track
construct
techniqu
recogn
distribut
speech
in this work, we are interested in understanding the roles of these different tasks in improved scene understanding, in particular semantic segmenta tion, object detection and scene recognition
we introduce a purely feed-forward architecture for semantic segmentation
the topic of semantic segmentation has witnessed considerable progress due to the powerful features learned by convolutional neural networks (cn ns)

Conclusion

- Domain-driven attention tuning
- Attention-based word influence algorithm
- Quantitative and qualitative evaluations
- Interactive system: KeywordMap

Q&A

Thanks 

? *Any Question*