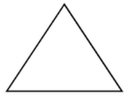
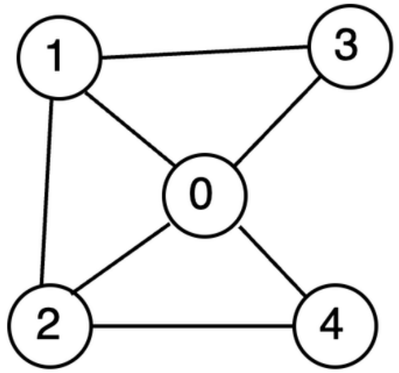
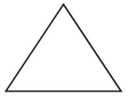
Cenia Meeting 26-07-24
Compositional Generalization: Triangles Dataset

Dataset Generation

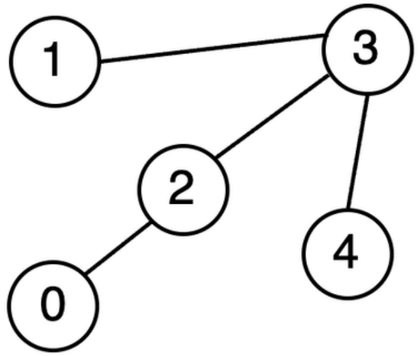
Con 

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$



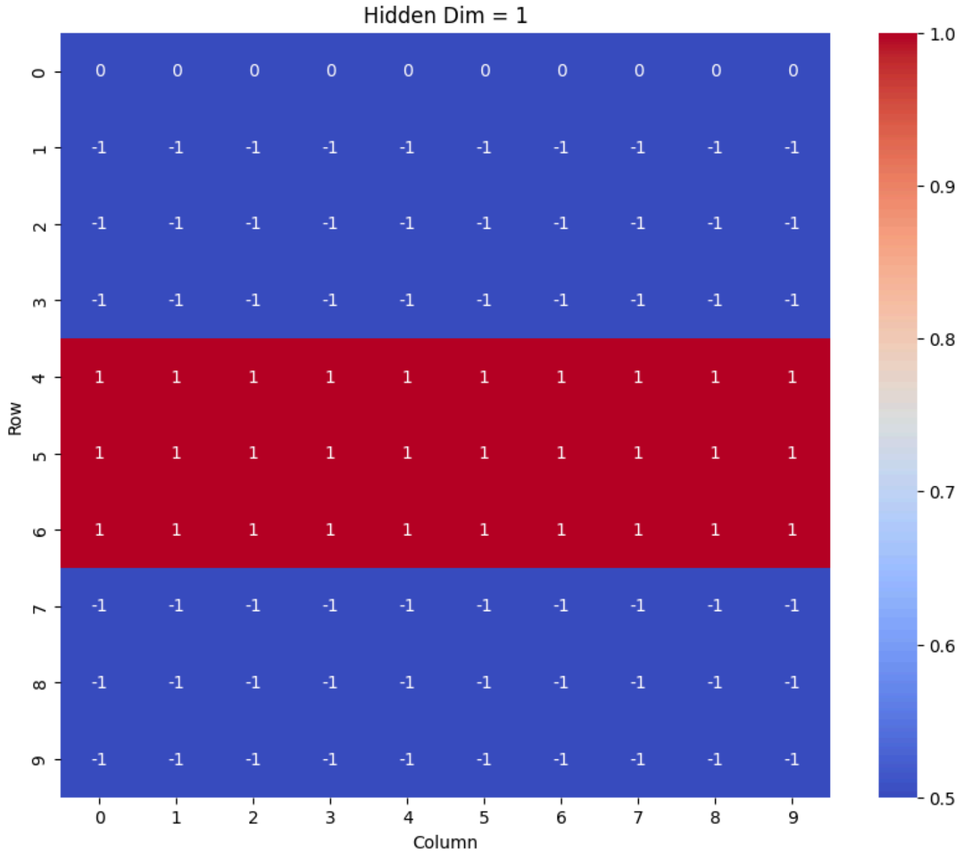
Sin 

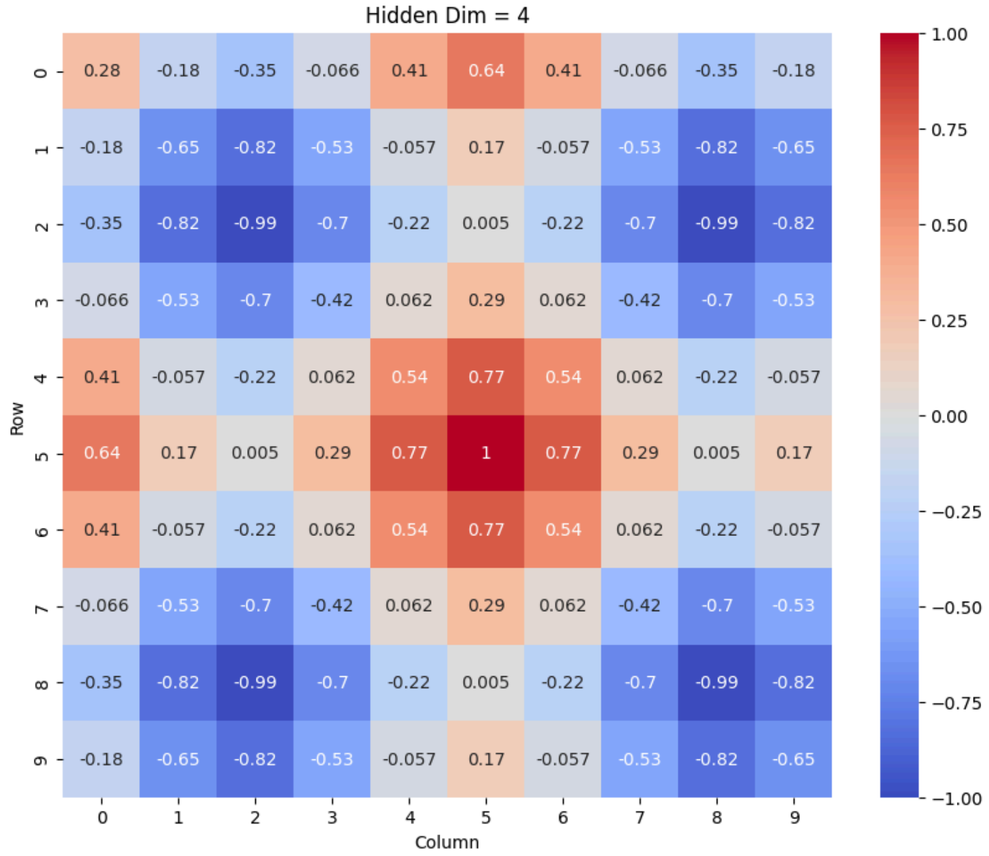
$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$



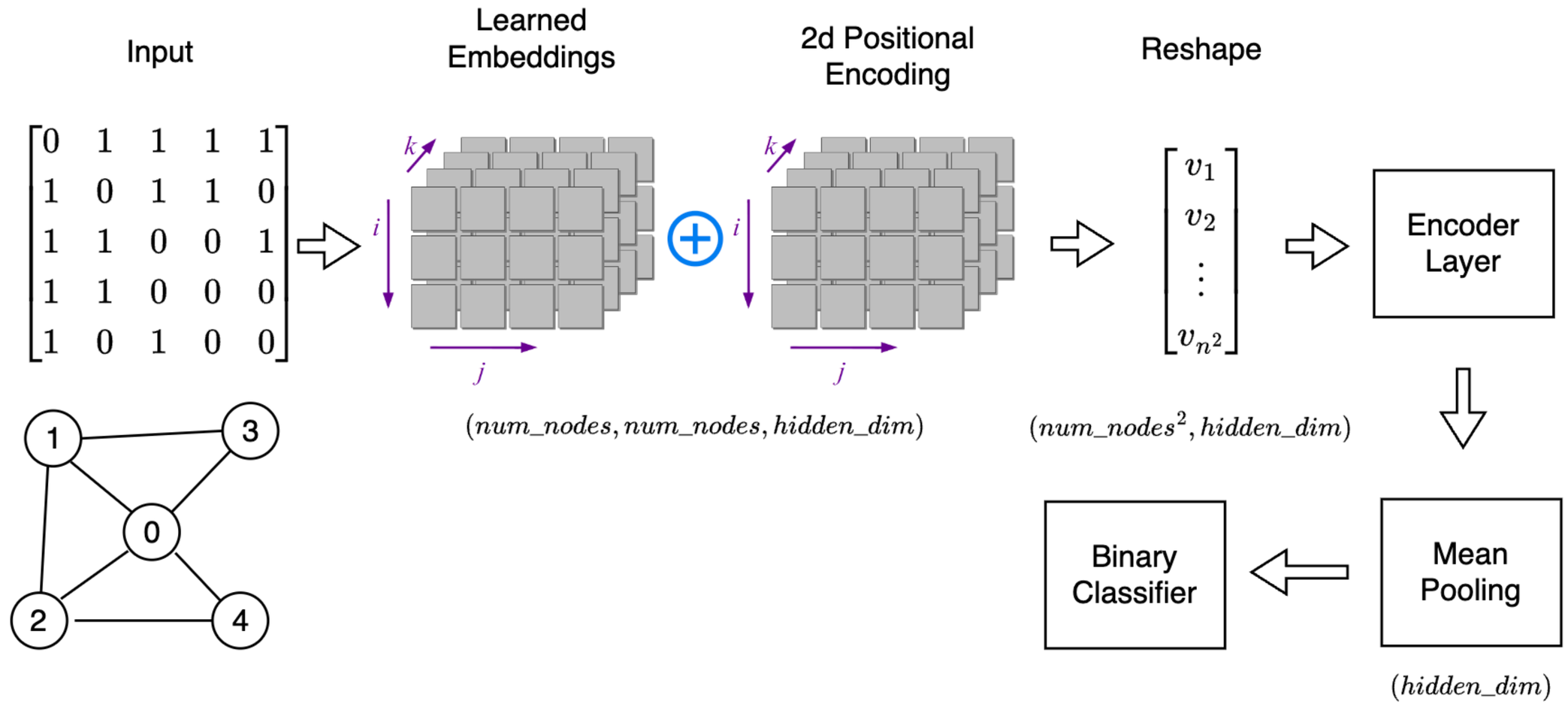
$$num\_nodes \in \{5, \ldots, 10\}$$

2d Positional Encoding





Encoder Experiment



Encoder Experiment

Architecture:

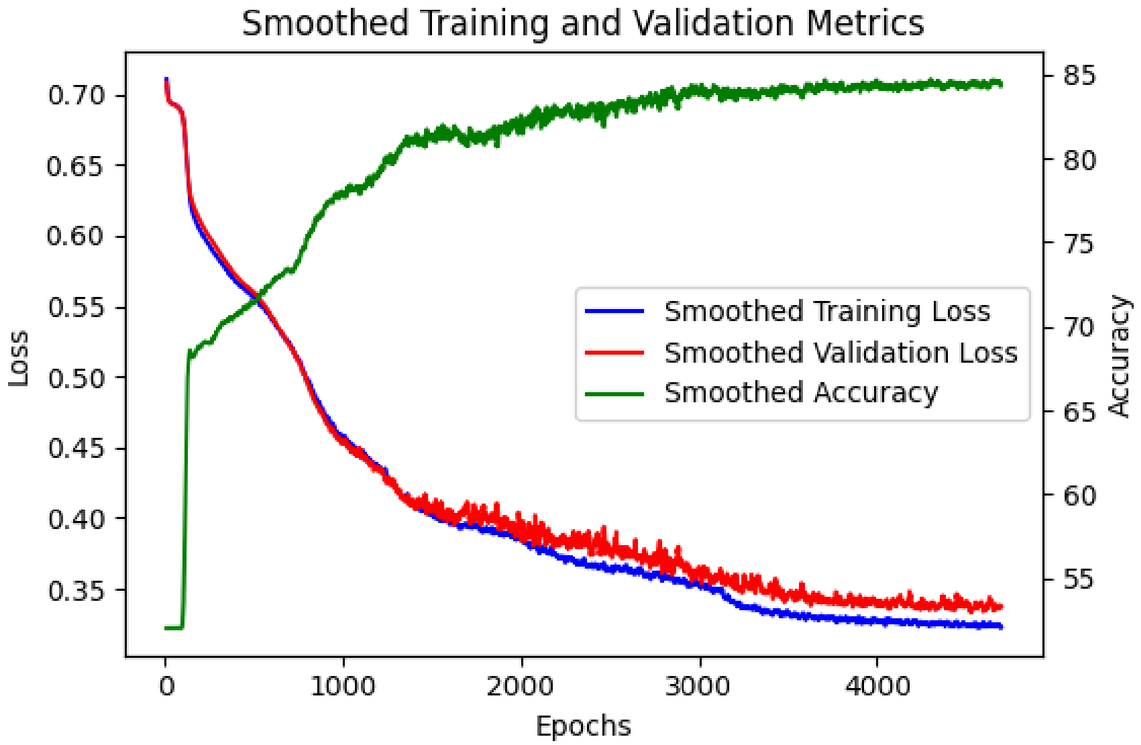
- Input_projection -> Adj. Matrix to nn.Linear
- Positional Encoding -> Positional encoding 2D
- Standard Pytorch Attention Layer
- Binary Classifier -> Sigmoid

Hyperparameters:

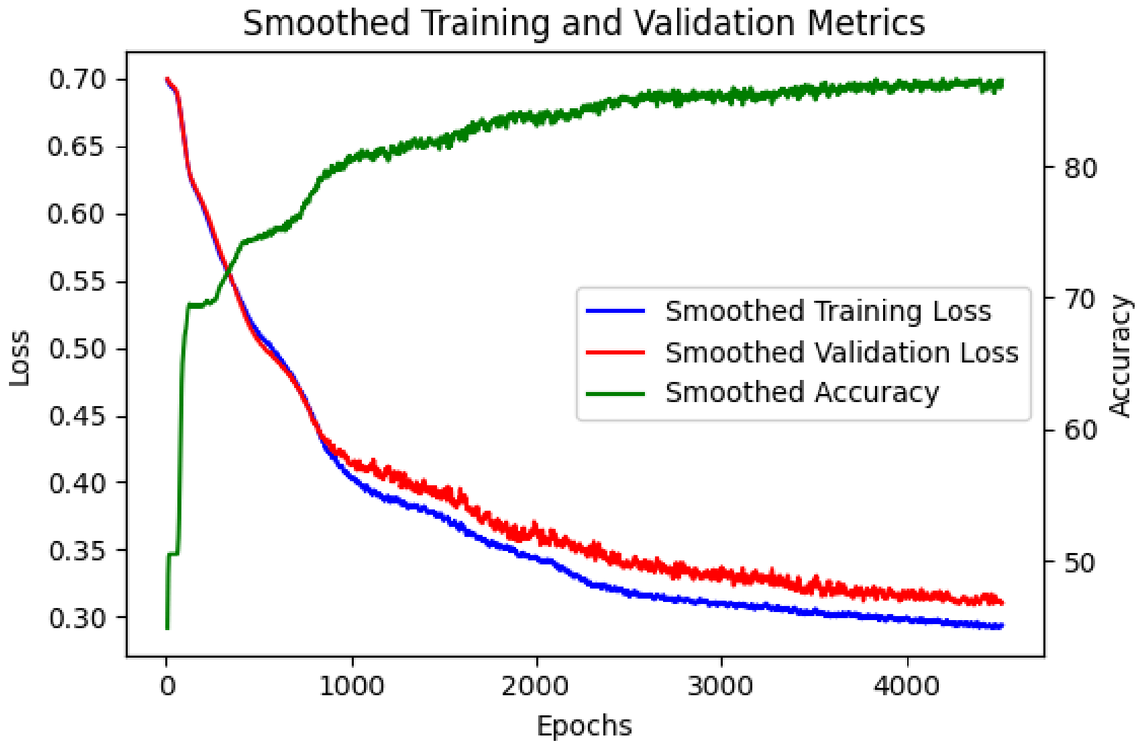
- Hidden_dimension -> 4, 6, 32
- Attention_heads -> 4, 6, 8
- Learning_rate -> 1e-3
- Epochs -> 4500
- Max_nodes -> 10

Results

hidden_dim = 4
num_heads = 4
num_layers = 1

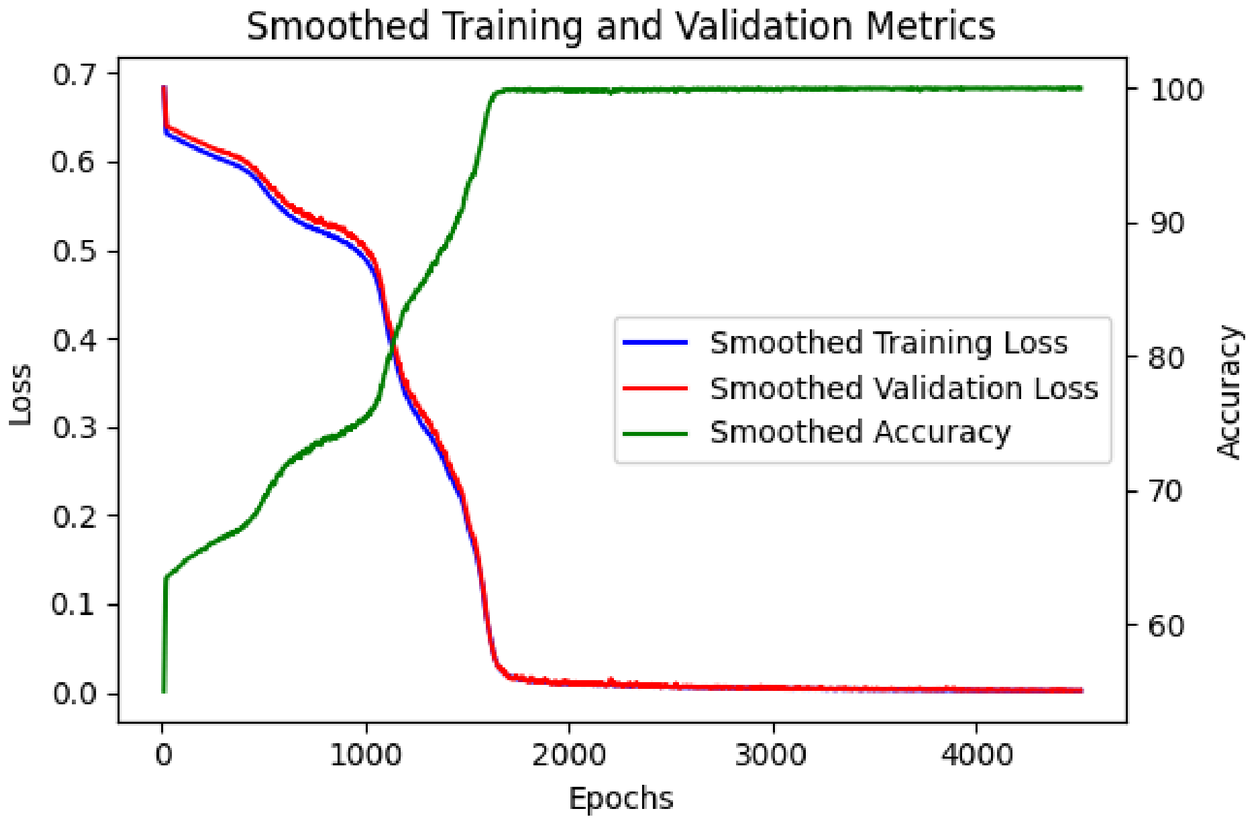


hidden_dim = 6
num_heads = 6
num_layers = 1



Experiment 2 Results

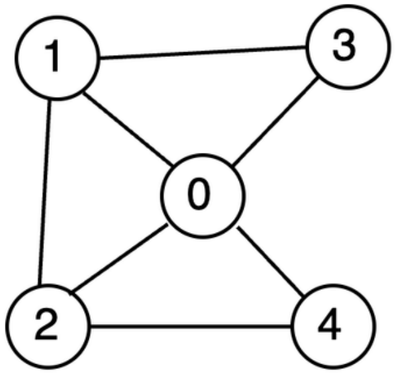
hidden_dim = 32
num_heads = 8
num_layers = 1



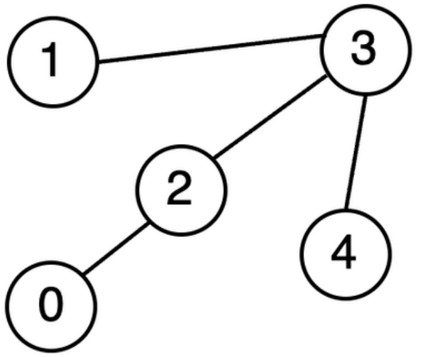
Discussion

When tokenizing an adjacency matrix each token represents a relation between nodes

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$



$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

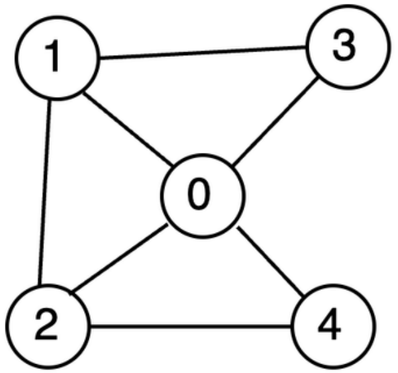


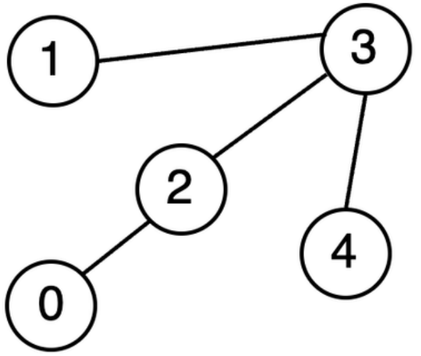
Discussion

Then attention is directly learning “relations between relations”

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$



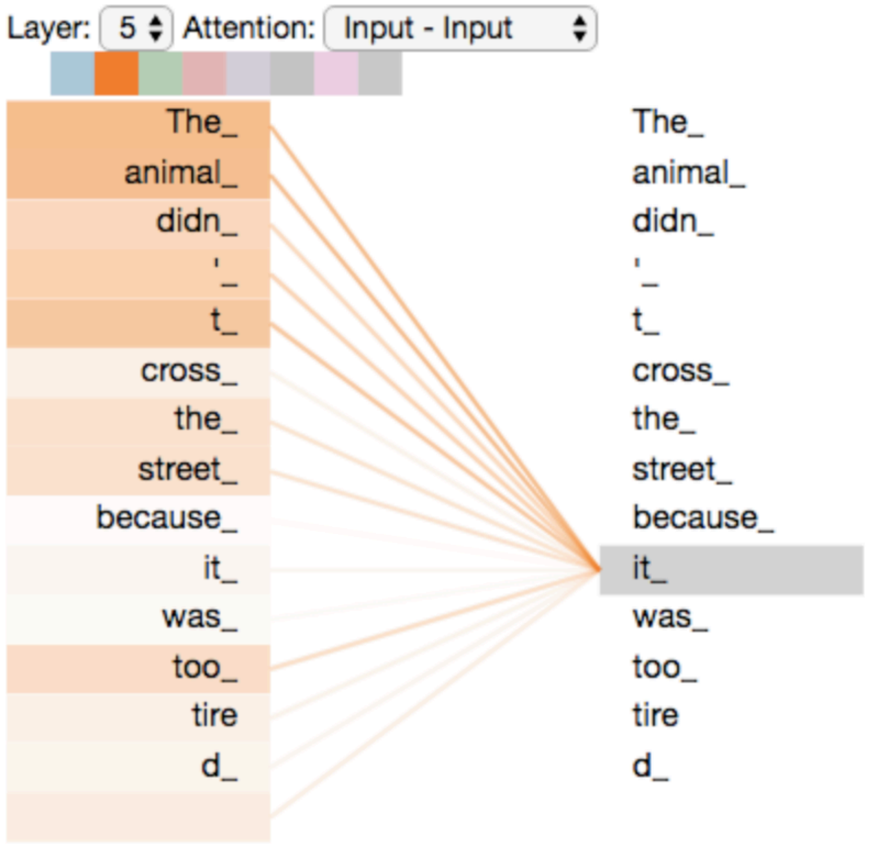


Discussion

On NLP transformers, each input token represents a set of characters, but not directly their relationships (the relationships are learned)

Discussion

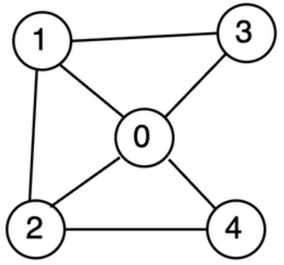
Token relationships are learned through attention



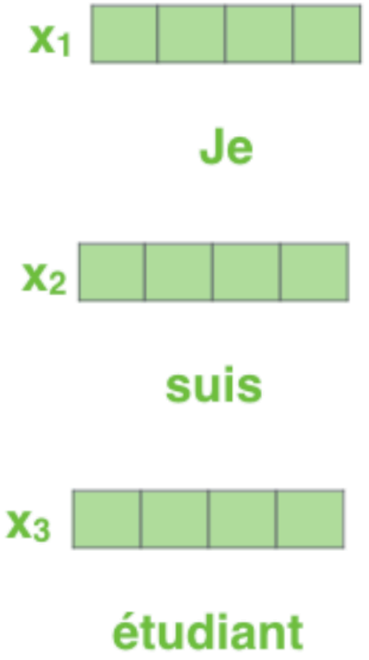
Discussion

The mismatch between the tokenized information might imply that our test is not assessing composition

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$



≠



Options?

Maybe our dataset should feature both nodes and relations as tokens expressed as a set of “fact” sequences

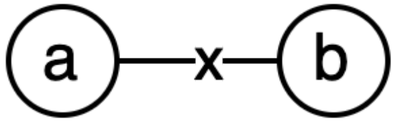
Node tokens: $\{a, b, c, d\}$

Relation tokens: $\{x, y\}$

Input sequence:

t1 t2 t3

$$”a \quad x \quad b”$$



Options?

And also include compositions as tokens, all of this expressed as a set of “fact” sequences

Node tokens: $\{a, b, c, d\}$

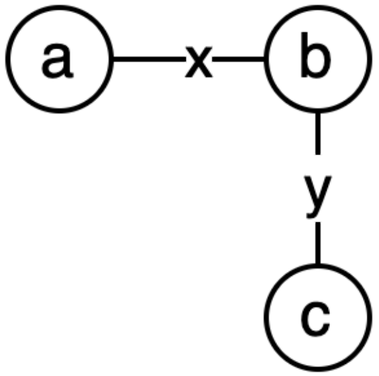
Relation tokens: $\{x, y\}$

Composition token: $\{z\}$

$$z = (x \circ y)$$

t1 t2 t3

sequence: $"a \;\; z \;\; c"$



$$(a, c) \in (x \circ y)$$