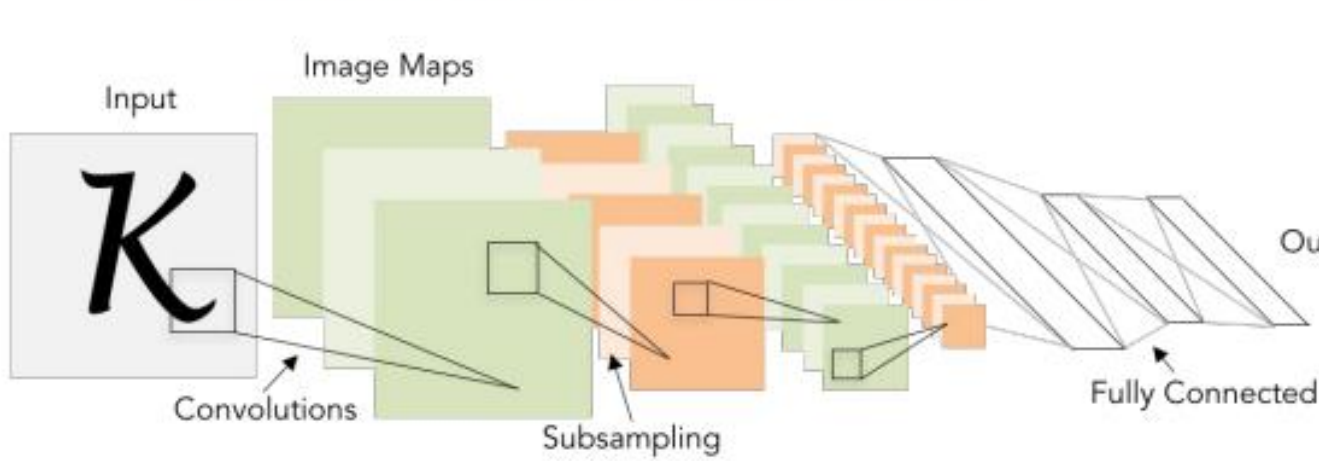


Graph Convolutional Networks for AST-Based Representations of Code

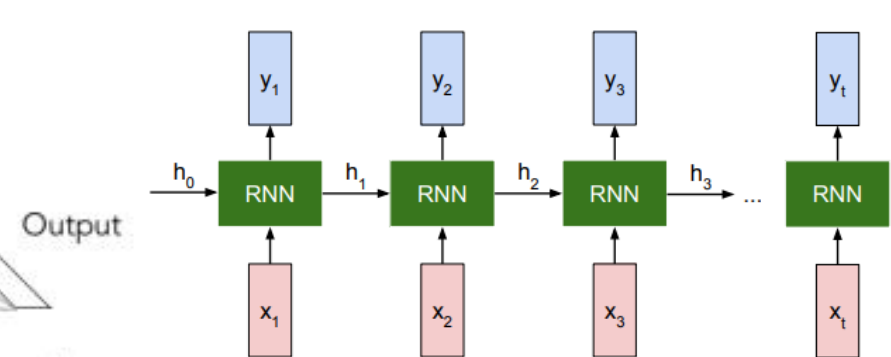
Rich Chen and Andy Stanciu

Neural network architectures can be specialized for various kinds of data (spatial, sequence) exist:

Convolutional Neural Networks



Recurrent Neural Network



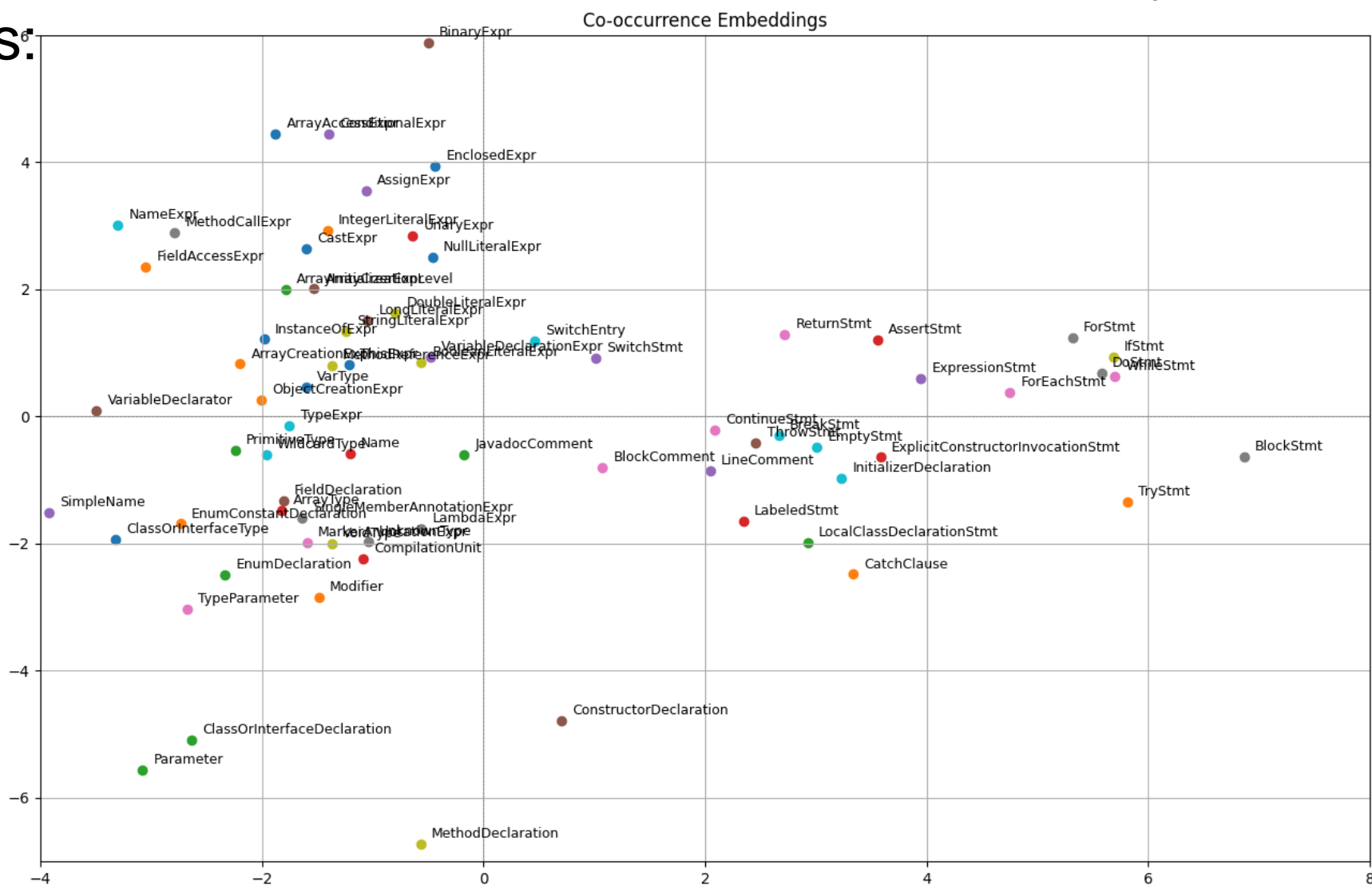
What about code for programs?

- We could model it as a sequence like regular text and use methods for that such as RNNs, LSTMs, Transformers, etc.
- But code has logical connections between statements that can't be easily represented with sequences

```
class Solution {
public int[] twoSum(int[] nums, int t) {
    int n = nums.length;
    // Loop through pairs of numbers
    for (int i = 0; i < n - 1; i++) {
        for (int j = i + 1; j < n; j++) {
            // If pair is found, return it
            if (nums[i] + nums[j] == t) {
                return new int[] { i, j };
            }
        }
    }
    // No solution found
    return new int[] {};
}
```

AST Representations for Neural Networks

- Construct **co-occurrence matrix** from abstract syntax tree nodes:



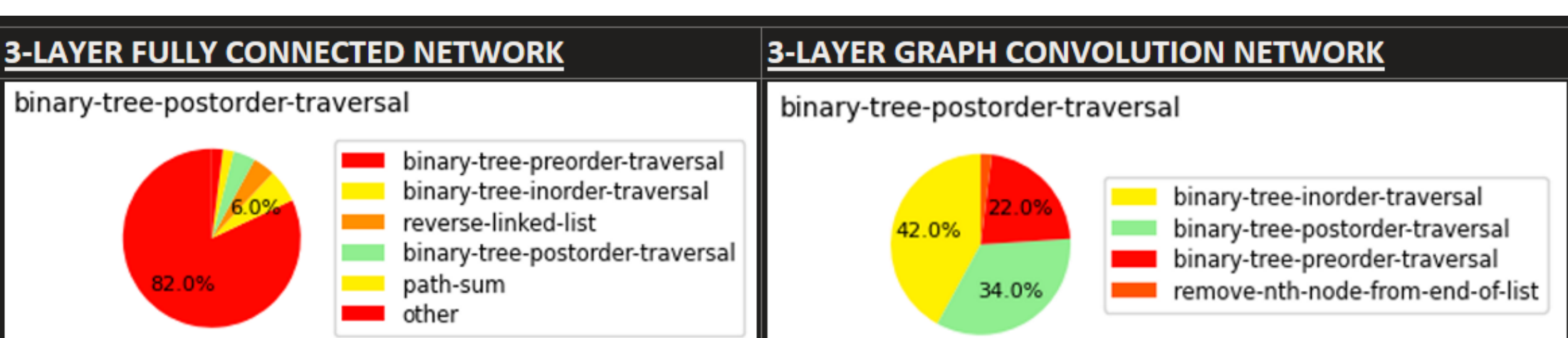
Experimentation

- Data: 100 LeetCode problems and solution examples for each
- Vectorize code examples into co-occurrence matrix
- Construct train set from 90% of the data and test set from 10%
- Train various architectures on train set
- Evaluate on test set with top-1, top-5, and top-10 accuracy:

Model	Top-1 Accuracy	Top-5 Accuracy	Top-10 Accuracy
Linear	85.90%	97.68%	99.06%
3-Layer MLP	90.64%	98.68%	99.30%
3-Layer GCN	94.04%	99.44%	99.76%

Adversarial Examples

- Certain problems were chosen to be very similar:
- Post-order, in-order, and pre-order traversal of binary tree
- GCN performance is much better in these cases

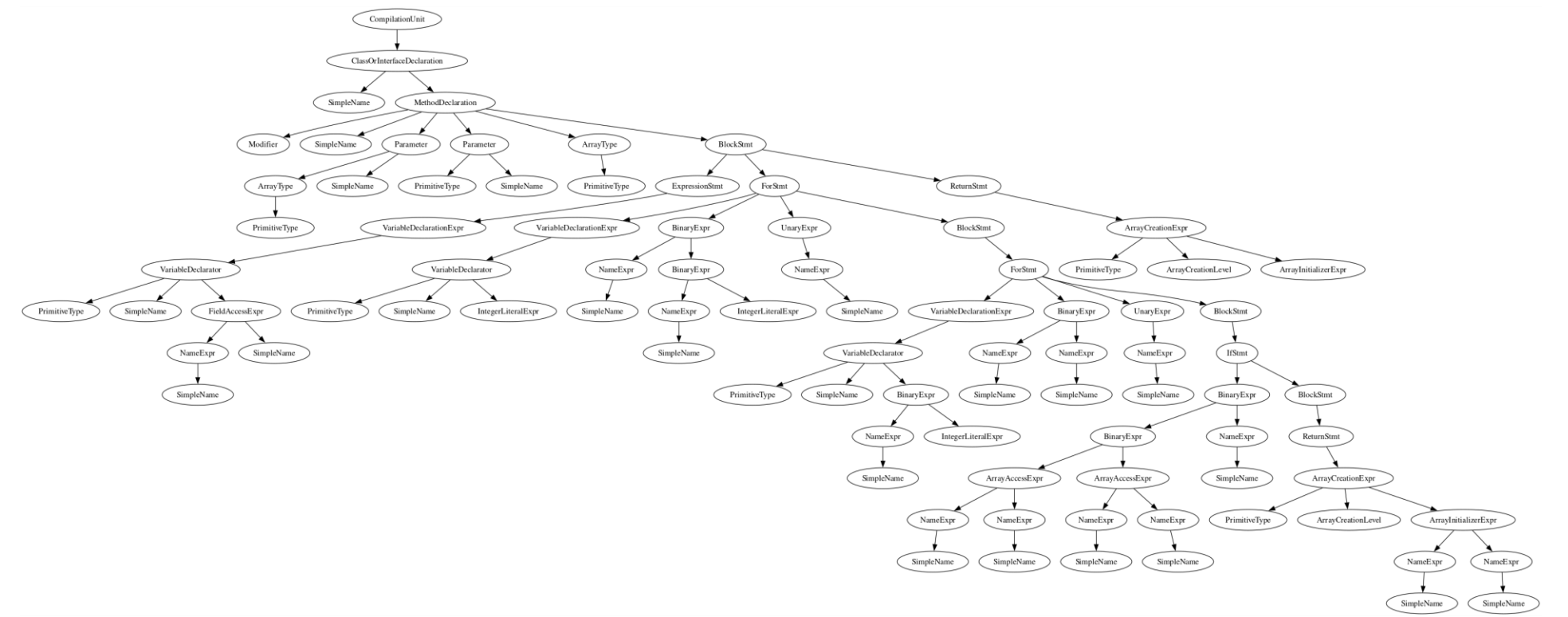


Graph Neural Networks

- Allow us to take in input data in the format of a graph and perform various operations such as graph-convolutions to learn properties of graph-based data.

Abstract Syntax Trees

- In compilers, **abstract syntax trees** are ubiquitous as an intermediate representation of code. Programs are compiled into this tree structure, preserving the fundamental structure and meaning of the code.



Graph Convolutional Networks (GCNs)

