Interactive Visual Pattern Search on Graph Data via Graph Representation Learning

Objective:
The goal of this study is to support human-in-the-loop, example-based graph pattern search in a graph database containing hundreds to thousands of distinct graphs. Based on graph representation learning, the authors offer a novel framework for interactive visual graph pattern search. On the basis of the presented framework, they create the visual analytics system called GraphQ.

Contributions:
There are 3 major contributions by the authors of this paper:

1. Presented a visual analytics platform for human-in-the-loop, example-based graph pattern search using graph representation learning.
2. Presented a unique technique (NeuroAlign) for pairwise node-alignment based on graph representation learning that gives 10x-100x faster results than the baseline combinatorial algorithm and 19%-29% more accurate results than the previous deep learning-based approach.
3. Presented, GraphQ, a prototype implementation of the framework that includes interactive query specification, query result presentation with different levels of detail, and user feedback methods for query refining. Also provided two use examples demonstrating the wide applicability and efficacy of the proposed solution.

VIS designs/techniques used:
Histograms or scatterplots are used to depict the distribution of important graph statistics. Semantic scene graphs are used to reference the relevant literature in computer vision. Directed acyclic graphs are also used to represent query results.

ML methods used:
Graph neural networks (GNNs) are used in this research to encode topological and node attribute information in a graph as fixed-length vectors. They used two distinct GNNs to solve 1) the decision problem of determining whether a query pattern exists in a graph and 2) the node-alignment problem of determining the one-to-one node correspondence between the query pattern and the query targets. The framework's essential components are the two GNNs, NeuroMatch and NeuroAlign.

Strengths:
NeuroAlign's intuitive and explainable visual signals are combined with unique visual and interaction designs to assist users in navigating the retrieval results and extracting insights. When compared to the baseline combinatorial Hungarian algorithm, NeuroAlign provides the result 10x-100x faster. They also provide results with 19-29% more accuracy as compared to the existing deep learning-based approach.

Shortcomings:
NeuroAlign's training and inference currently focus only on a single instance of subgraph isomorphism. In reality, however, the query nodes may be mapped to many groups of nodes in the same matched target network. The authors discovered that hard negative samples are essential for achieving a high accuracy rate during NeuroMatch training. To be sure that the subgraph link does not exist, sampled or perturbed queries must be checked using exact matching methods which are time-consuming to compute. At the moment, this technique works with undirected or connected graphs as the query pattern.

Future work:
More studies can be done to add node alignment for multiple subgraph isomorphism. The current approach may be improved to be more scalable to very large query graphs. Further work may be done to make it capable of handling directed or unconnected query patterns. In addition, given the great diversity of graph-structured data, the present work can be expanded to include other usage scenarios such as social network analysis.

Question for discussion: In unidirectional graphs, the vector embeddings will have one additional dimension. So how will this be handled by this GNN algorithm?