Docket #: S26-222

A Graph Foundation Model for Cross-Domain Node Representation Learning on Heterogeneous Graphs

Stanford researchers have developed a machine learning framework that uses graph diffusion to generate transferable node representations across heterogeneous networks, without retraining the encoder.

Many systems in science can be represented as graphs, which are networks in where individual entities (nodes) are linked by relationships (edges), such as interacting proteins within a cell. Graph neural networks (GNNs) are machine learning models designed to learn from such data, and they have achieved strong performance across a range of tasks in various fields, including commerce. Unfortunately, their performance is largely confined to the graph on which a model is trained. A GNN typically cannot be readily applied to another graph, because datasets differ in how entities are described, in what those descriptions signify, and in their scale and pattern of connectivity. Existing graph foundation models are generally confined to a single domain, depend on large, labeled datasets, or fail to generalize across differing network structures and feature schemas. A unified framework that is pretrained once on abundant, low-cost graphs and then applied to heterogeneous targets without retraining its encoder is therefore needed. The Islam and Xing labs at Stanford have a developed a novel framework that characterizes each node through graph diffusion processes, yielding structural descriptors that are comparable across datasets of different sizes, densities, and domains. By reducing the labeled data and computational effort required to analyze new networks, the Graph Foundation Model has the potential to accelerate discovery in computational biology and other network-driven fields.

Stage of Development: *proof of concept*

Applications

- The framework is a general-purpose method applicable to any task involving node representation learning on graphs
    - Computational biology, biomedical research, recommendation systems, social network analysis, citation and co-author networks, infrastructure and sensor networks

Advantages

- Learns transferable node representations from a single pretraining run on diverse source graphs and applies them to unrelated target domains without retraining the encoder
- First graph foundation model to achieve cross-domain transfer from non-biological pretraining graphs to biological prediction tasks at parity with supervised baselines trained directly on the target
- First method to render every node of any graph as a fixed-size diffusion-based image suitable for standard vision backbones, scalable to graphs with millions of nodes through Chebyshev polynomial approximation

Innovators

- Md Tauhidul Islam
- Sakib Mostafa
- Lei Xing

Licensing Contact

Matthew Ryan

Licensing Manager

Email