

GNN Interpretation Thesis Project

Shalin Patel

GNNs and Computational Biology

- GNNs useful for creating multi-modal models of complex biological systems (e.g. Gene Expression)
 - Based off HiC contact data
 - Based off Gene Regulatory Networks (GRNs)
- These models often demonstrate state of the art performance
- Many biological datasets demonstrate graphical structure and are often coerced into unnatural forms
 - GNNs remove a lot of these issues

Interpretation of GNNs

- While GNNs give better performance than previous models, interpreting them is harder than before
- Many traditional models have well established interpretations
 - SVMs, Random Forests, Logistic Regression all have natural interpretations
 - CNNs also have many interpretation methods including saliency maps, LIME, SHAP, ...
- To derive understanding, and new biological value from the improved performance, interpretations for GNNs need to be generated
- Specifically, we want to understand what parts of the underlying graph structure are driving newfound performance

Problem Formulation

Let \mathcal{G} denote a graph on edges E and nodes V such that each node is associated with $\mathcal{X} = \{x_1, \dots, x_n\}$.

Let $f : V \mapsto \{1, \dots, C\}$ represent a classification task on nodes in V .

Specifically, for a node v , a GNN ϕ that approximates f learns the conditional distribution $P_\phi(Y \mid G_c(v), X_c(v))$ where G_c and X_c represent the computational graph of v (typically the n -hop neighborhood).

Predictions are given as $\hat{y} = \phi(G_c(v), X_c(v))$.

The goal of GNN interpretation is to find $\psi : V \mapsto \mathcal{G}$ where $\psi(v) = G_s(v)$ such that $G_s(v) \subseteq G_c(v)$ and

$$MI(\hat{y}, \phi(G_s(v), X_c(v)))$$

is maximized.

