Graph attribution methods applied to understanding immunogenicity in glycans

Somesh Mohapatra [1] **Joyce An** [1] **Rafael Gómez-Bombarelli** [1]

Abstract

Macromolecules, such as naturally occurring and synthetic proteins and glycans, have diverse chemical structures, varying in monomer composition, connecting bonds and topology. In addition to the chemical diversity, macromolecules usually have opaque structure-activity relationships, making activity prediction and model attribution hard tasks. Recently, we proposed macromolecule graph representation learning, achieving state-of-the-art results in immunogenicity classification of glycans. Here, we extend this framework to include attribution methods for graph neural networks. We evaluated the performance of 2 attribution methods over 3 model architectures, and an attention attribution for the attention-based model, and demonstrated it for an immunogenic glycan. Our work has two-fold implications - (1) provides attribution-backed chemical insights at the monomer and chemical substructure level, and (2) informs further *in silico* and wet-lab experiments.

1. Introduction

Graphs are a natural representation for social networks, molecules, and biological interactomes, amongst others (Battaglia et al., 2018). Unsupervised and supervised learning over graph representations have enabled significant advancements, achieving state-of-the-art results across several fields (Hamilton et al., 2017). Attribution methods have been evaluated for a number of graph neural networks (GNNs) (Sanchez-lengeling et al., 2020). In chemistry and life sciences, GNNs have been used for property prediction and design of small molecules (Yang et al., 2019; Jin et al., 2020) and periodic crystals (Xie & Grossman, 2018).

Macromolecule is a result of monomer composition, and (complex) bonds connecting different monomers, similar to atoms and bonds in small molecules. We have reported that fingerprint-based representations worked well for macromolecule property prediction (Schissel et al., 2020). Recently, we introduced graph representations for macromolecules featurized using fingerprints, which we leveraged to obtain state-of-the-art results for supervised classification of glycans (Mohapatra et al., 2021).

In this study, we extend our macromolecule graph representation learning framework to include attribution methods for feature importance analysis in macromolecule property prediction. We apply our tools to the study of immunogenicity in glycans.

2. Methodology

2.1. Macromolecule graph representation

We represented the macromolecule as undirected, attributed graph, $\mathcal{G}(V, E)$, where V represents vertices/nodes, and E represents edges (Figure 1A). Each node corresponds to a monomer, and edge to a bond. Both nodes and edges are featurized using stereochemical extended connectivity fingerprints, capturing the inherent chemistry of the monomer/bond molecule (Rogers & Hahn, 2010).

2.2. Graph neural networks

We used pre-trained GNNs, specifically, molecular graph convolution networks (GraphConv) (Kearnes et al., 2016), message passing neural networks (MPNN) (Gilmer et al., 2017), and graph attention networks (GraphAtt) (Xiong et al., 2020), for immunogenicity classification in glycans, as reported in Mohapatra et al. (2021) (Figure 1B).

Briefly, these models were trained with 60:20:20 train:valid:test splits, and optimized for 1000 hyperparameter iterations. Ultimately, an ensemble of 25 models per model architecture consisting of the top 5 hyperparameter sets retrained with 5 random weight initialization seeds, was used to make the predictions. The model performance metrics on the test data set have been reported in Table 1.

[1]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Correspondence to: Rafael Gómez-Bombarelli <rafagb@mit.edu>.

The 2021 ICML Workshop on Computational Biology.

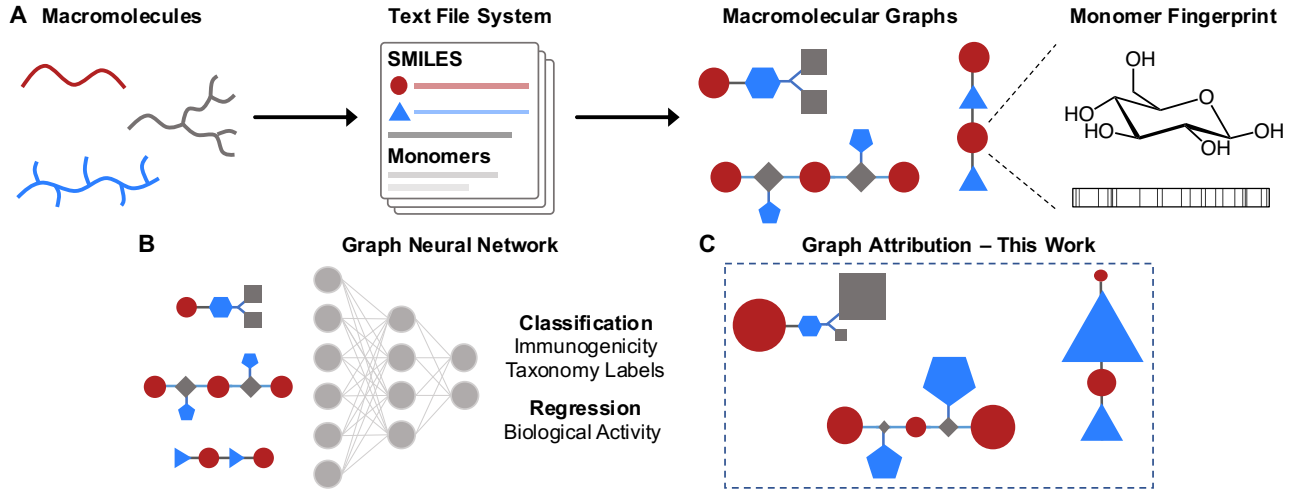

Figure 1. **Macromolecule graph representation enables supervised learning and attribution. A.** Chemical structures of macromolecules are represented as text files, which are parsed into NetworkX graphs. In the graphs, nodes and edges are featurized using extended connectivity fingerprints of monomers and bonds, respectively. The text file lists string representations (SMILES) for monomers and bonds. **B.** Supervised learning over macromolecule graph representations is done for a variety of classification and regression tasks using different model architectures. **C.** Graph attribution provides an insight into the decision-making process of the model, by highlighting the relative importance of different nodes/monomers (denoted as node size) in making a specific prediction.

2.3. Attribution methods

We used integrated gradients (IGs) (Sundararajan et al., 2017) and Input x Grad (InpGrad) (Shrikumar et al., 2017) attribution methods for the analysis of GNNs (Figure 1C). The notation follows Sanchez-lengeling et al. (2020).

IGs interpolate between the input graph and a baseline graph, where all features are zero, and accumulate the gradient values for each node.

$$\mathcal{G}_A = (\mathcal{G} - \mathcal{G}') \int_{\alpha=0}^{1} \frac{dy\left(\mathcal{G}' + \alpha\left(\mathcal{G} - \mathcal{G}'\right)\right)}{d\mathcal{G}} d\alpha$$

InpGrad is the element-wise product of the input graph and the gradient.

$$\mathcal{G}_A = \left(\frac{d\hat{y}}{d\mathcal{G}}\right)^T \mathcal{G}$$

For the attention-based GNN, GraphAtt, in addition to IGs and InpGrad, we evaluated attribution using attention weights, where the node attention weights are obtained by averaging over the attention scores of the adjacent nodes.

For each attribution method, we obtained the node weights by multiplying the positive weights with the input fingerprint vectors -

$$\mathbf{n} = \sum_{nodes} \mathcal{G}_A^+ \, \mathcal{G}$$

The node weights were normalized to the maximum node weight to obtain the normalized weights -

$$\mathbf{n_{norm}} = \frac{\mathbf{n}}{max(\mathbf{n})}$$

Consistency of node weights, as defined in Sanchez-lengeling et al. (2020), was used for evaluation of different attribution methods and model architectures.

We have demonstrated the results using an immunogenic glycan from the test data set. This glycan was correctly classified by all 3 model architectures.

Table 1. Classification metrics for different model architectures on held-out test data set. Abbreviations: ROC-AUC, Receiver Operating Characteristic - Area Under Curve; BCE Loss, Binary Cross-entropy Loss.

MODEL	ROC-AUC	ACCURACY	BCE LOSS
GRAPHATT	0.99± 0.01	0.95± 0.01	0.13± 0.10
MPNN	0.99± 0.01	0.96± 0.01	0.14± 0.10
GRAPHCONV	0.99± 0.01	0.96± 0.01	0.13± 0.11

3. Results

3.1. Evaluation of IG and InpGrad attribution methods

The node weights calculated using IG and InpGrad have similar trends, but varying attribution consistency, both across attribution methods and model architectures (Figure 2).

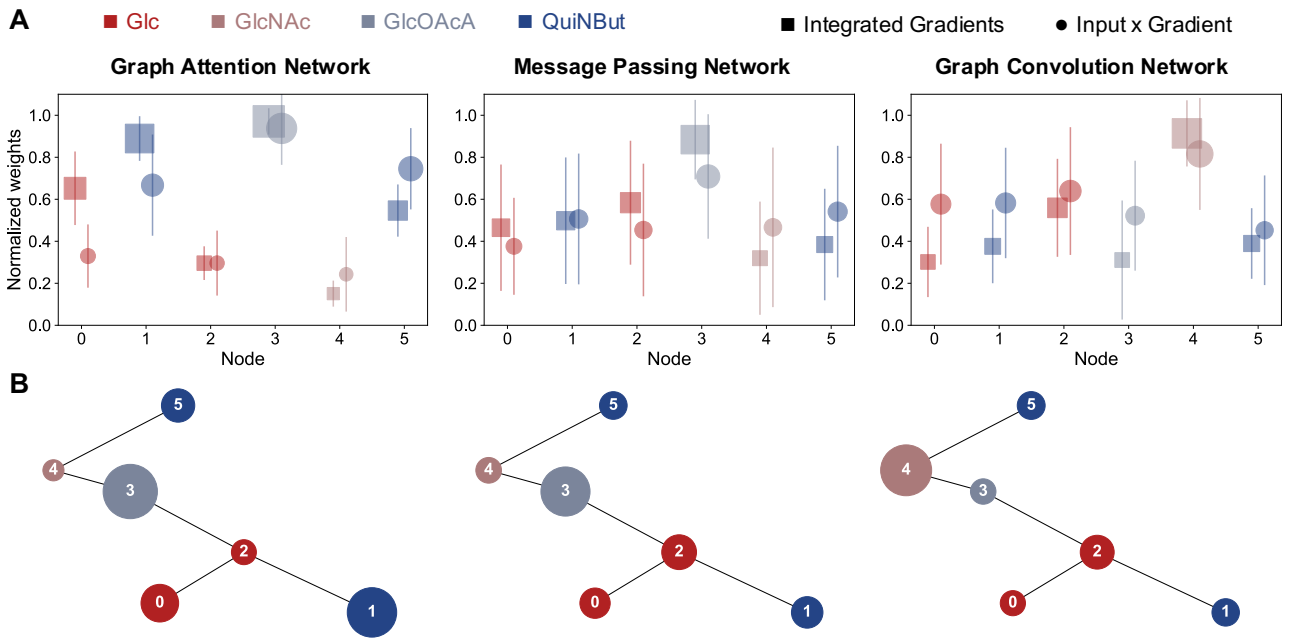

Figure 2. **Model architectures with similar classification performance have varying attribution consistency. A.** Mean node weights, obtained using IG and InpGrad, are denoted in the scatter plot, with error bars representing the standard deviation. The mean is calculated from the node weights of respective model architecture, across top 5 hyperparameter sets and 5 random weight initialization. The data points are colored by the monomer, and shaped according to the attribution method. **B.** Visualization of IG node weights obtained using different model architectures for an immunogenic glycan. The size of the node corresponds to the importance, and the color corresponds to the monomer, consistent with the pattern in A.

IG node weights are relatively more consistent than InpGrad weights, as noted from the smaller standard deviation error bars. This observation is similar to the reports in Sanchez-lengeling et al. (2020), and in line with IG satisfying both sensitivity and implementation invariance axioms, unlike other attribution methods. For a single model architecture, the relative magnitudes and trends of node weights, obtained from IG and InpGrad, are similar.

GraphAtt has the most consistent node weights amongst all model architectures. The rank order of the nodes is similar, for both IG and InpGrad weights, in GraphAtt and MPNN. In GraphConv, the highest and lowest nodes, by weights, are swapped, with the rest of the nodes being of relatively similar weight, in comparison to the other 2 model architectures.

Among the different combinations of the 2 attribution methods and 3 model architectures evaluated for the single example, IG is a better attribution method than InputGrad, and GraphAtt is the most invariant across different model implementations. Similar evaluations of different glycans in the data set need to be done to firmly ascertain this attribution-architecture choice.

3.2. Attention weights for GraphAtt models

Node attention weights obtained from GraphAtt models follow a similar trend as IG and InpGrad weights (Figure 3). However, the variation in the magnitudes of the mean values is not as significant, and the standard deviations are relatively larger, in comparison to the other attribution methods.

4. Discussion and Future Work

Graph attribution methods help in cracking open the black-box of GNN model predictions, and provide explanations in the absence of ground-truth understanding. In this case, we evaluated 2 attribution methods for 3 model architectures for an immunogenic glycan. Across multiple attribution-architecture combinations, we observed that the 3^{rd} node, a GlcOAcA monomer, contributed most to the immunogenicity of the glycan.

In the near future, we aim to extend our analysis of the attribution methods to different glycans in the data set. Through a combination of n-grams-like composition analysis and attribution weights, we hope to uncover the underlying principles of glycan immunogenicity.

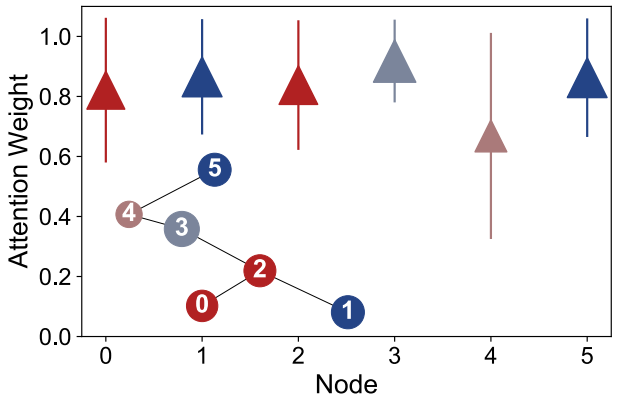

Figure 3. **Visualization of attention weights.** Scatter plot shows the mean node attention weights, and error bars denote the standard deviation. The glycan graph, with node size corresponding to attention weight, is shown as an inset image. The color palette is consistent with Figure 2.

5. Conclusion

As demonstrated in the study, robust application of model training and attribution methods can help in elucidation of fundamental design principles. For biochemical properties, such as immunogenicity or efficacy of a drug, the ability to determine the importance of relevant chemical substructures, monomers and bonds, will significantly improve the understanding of the system. Moreover, such approaches can inform further studies to probe the mechanism of action in wet-lab analyses, and ultimately drive design of better (biological) macromolecules.

References

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G. E., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., and Pascanu, R. Relational inductive biases, deep learning, and graph networks. *arXiv: 1806.01261*, 2018.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural Message Passing for Quantum Chemistry. *arXiv:1704.01212v2*, 2017.

Hamilton, W. L., Ying, R., and Leskovec, J. Representation Learning on Graphs: Methods and Applications. *arxiv:1709.05584*, 2017.

Jin, W., Barzilay, R., and Jaakkola, T. S. Hierarchical generation of molecular graphs using structural motifs. *arXiv:2002.03230*, 2020.

Kearnes, S., McCloskey, K., Berndl, M., Pande, V., and Riley, P. F. Molecular graph convolutions: moving beyond fingerprints. *Journal of Computer-Aided Molecular Design*, 30(8):595–608, 2016.

Mohapatra, S., An, J., and Gómez-Bombarelli, R. Chemistry-informed Macromolecule Graph Representation for Similarity Computation and Supervised Learning. *arXiv: 2103.02565*, 2021.

Rogers, D. and Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5), 2010.

Sanchez-lengeling, B., Wei, J., Lee, B., Reif, E., Wang, P. Y., Qian, W., Mccloskey, K., Colwell, L., and Wiltschko, A. Evaluating Attribution for Graph Neural Networks. *Proceeedings of the Neural Information Processing Systems Conference*, (NeurIPS), 2020.

Schissel, C. K., Mohapatra, S., Wolfe, J. M., Fadzen, C. M., Bellovoda, K., Wu, C.-L., Wood, J. A., Malmberg, A. B., Loas, A., Gómez-Bombarelli, R., and Pentelute, B. L. Interpretable Deep Learning for De Novo Design of Cell-Penetrating Abiotic Polymers. *bioRxiv:2020.04.10.036566*, 2020.

Shrikumar, A., Greenside, P., and Kundaje, A. Learning important features through propagating activation differences. *arXiv:1704.02685*, 2017.

Sundararajan, M., Taly, A., and Yan, Q. Axiomatic attribution for deep networks. *arXiv:1703.01365*, 2017.

Xie, T. and Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters*, 120(14):145301, 2018. ISSN 10797114.

Xiong, Z., Wang, D., Liu, X., Zhong, F., Wan, X., Li, X., Li, Z., Luo, X., Chen, K., Jiang, H., and Zheng, M. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of Medicinal Chemistry*, 63(16):8749–8760, 2020.

Yang, K., Swanson, K., Jin, W., Coley, C. W., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T. S., Jensen, K. F., and Barzilay, R. Analyzing Learned Molecular Representations for Property Prediction. *Journal of Chemical Information and Modeling*, 59(8), 2019.