Cell Segmentation Explainer

Ziqiao Ma

August 2020

1 Introduction

Many biological data can be represented as a typical graph data. Combining the node features and the graph structures, Graph Neural Networks were proved to be a powerful model for machine learning on biological data...

[To be written... –Martin]

2 Theoretic Background

To formulate the problem, we introduce a uniform set of notations for the following discussions.

2.1 Graph Neural Networks

Given a large graph $G = (V, E)$, each node $v \in V$ is associated with a d dimensional feature vector $x_v \in X \subseteq \mathbb{R}^d$, and a label $y_v \in \{1, 2, \cdots, C\}$.

Denote the input features as a matrix $X \in \mathbb{R}_{n\times d}$, and the labels as a vector $Y \in \mathbb{R}_n$. We also define a loss function $l(\Phi|G, X, Y)$ for a model Φ, which maps V, together with the feature and graph information (G, X) to a prediction vector $\tilde{Y} \in \mathbb{R}_n$. Thus, GNN's prediction $\tilde{y}_v$ for a node v is given by $\tilde{y} = \Phi(G_v, X_v)$.

Graph Neural Networks (GNNs) are a class of such model M that defines a multilayer feature propagation process. We represent the k^{th} layer as $X(k)$, and $X(0) \in \mathbb{R}_{n\times d}$ is the input node features. GNNs defines a recursive function f for the next-layer representation,

$$X_{(k+1)} = f(X_{(k)}; G, W_{(k)}), \tag{1}$$

where $W_{(k)}$ is the parameter for the k^{th} layer.

In this work, we will consider the Graph Convolutional Network (GCN) model [1]. Given the original adjacency matrix A and the diagonal degree matrix D, the normalized adjacency matrix S is given as

$$S = (I + D)^{-\frac{1}{2}}(A + I)(I + D)^{-\frac{1}{2}} \tag{2}$$

The recursive function for GCNs is defined as,

$$X_{(k+1)} = \mathrm{ReLU}(SX_{(k)}W_{(k)}) \tag{3}$$

where ReLU is the element-wise rectified-linear unit activation function [2].

2.2 GNN-Explainer

Interpretability has been address by several recent work recently. Especially, GNN-Explainer is a model-agnostic approach for providing interpretation explanations for predictions of any GNN-based model on any graph-based machine learning task [3]. Given a trained GNN model and the prediction, the GNN-explainer will generate an explanation by identifying a subgraph and a subset of features that are most influential for the model's prediction. GNN-Explainer is typically suitable for single node explanation.

Denote the subgraph as $G_s \subseteq G$ of size K_M, and a subset of features $X_s^F \subseteq X_s$ of size K_F, where X_s is the associated features of G_s and F is a binary selection function that mask out only highly relevant features.

The metric measuring the importance is mutual information MI, and formulate the GNN-Explainer as a jointly optimization problem:

$$\max_{G_s,F} MI(Y,(G_s,F)) = H(Y) - H(Y|G=G_s, X=X_s^F) \tag{4}$$

where $H(\cdot)$ is a fixed entropy term. Thus, the problem becomes

$$\min_{G_s,F} \mathbb{E}_{Y|G_s,X_s}[\log P_\phi(Y|G=G_s, X=X_s^F)] \tag{5}$$

We assume the convexity of GNNs and bound $H(Y|G=G_s, X=X_s^F)$ by Jensen's inequality. Let $G_s \sim \mathcal{G}$, the optimization problem is reformed as

$$\min_{\mathcal{G},F} \mathbb{E}_{Y|G_s,X_s}[\log P_\phi(Y|G=\mathbb{E}_{\mathcal{G}}[G_s], X=X_s^F)] \tag{6}$$

By multivariate Bernoulli decomposition, the above optimization problem can be replacing $\mathbb{E}_{\mathcal{G}}[G_s]$ by masking the adjacency matrix A,

$$\min_{\mathcal{G},F} \mathbb{E}_{Y|G_s,X_s}[\log P_\phi(Y|G=A\odot\sigma(M), X=X_s^F)] \tag{7}$$

where $M \in \mathbb{R}_{n\times n}$ is a mask to learn and σ is the sigmoid function that maps the mask to binaries.

To learn the selector F, we marginalize over all X_s with a Monte Carlo estimate to sample from empirical marginal distribution for nodes and backpropagate to the feature mask F by through a d-dimensional random variable X. We reparametrize X as:

$$X = Z + (X_s Z) \ s.t. \ \sum_j F_j \leq K_F \tag{8}$$

where Z is a random variable sampled from the empirical distribution.

2.3 Label Explanation via reference Node Selection

The original work address the global explanation of class c by choosing a reference node v_c by computing the mean embedding of all nodes assigned to c and take explanation $G_s(v_c)$ for reference v_c and align it to explanations of other nodes assigned to class c [3]. This choice is general and works poorly on biological dataset. Here we propose 2 alternative methods to choose a subset of reference node V_s of size K_N from the set of all nodes V, and then further span a subgraph with the chosen nodes.

Uncertainty sampling queries the labels for the nodes which current model is least certain with represent to classification prediction [4]. To stay uniform with single node selection procedure, we use the information entropy H as well as our informativeness metric. The information entropy of node v_i is calculated as:

$$H(v_i) = P_\Phi(Y_{ic}=1|G,X)\log P(Y_{ic}=1|G,X) \tag{9}$$

The larger $H(v_i)$ is, the more uncertain current model is regarding to v_i.

To avoid the noisy cells and utilizing the fact that cells cluster in terms of islets, we can use density to find the nodes that locate dense regions of the training nodes in the embedded space [5]. The density score for node v_i is calculated by first applying Kmeans on the embeddings of the model and then compute the Euclidean distance between each node and its cluster center. Let L be the number of layers of the GNN and $C(v_i)$ be the cluster center of the v_i on embedding. The density is the calculated by:

$$\rho(v_i) = \frac{1}{1 + ||X_i^{L-1} - C(v_i)||_2} \tag{10}$$

To combine the metrics, we define the influence score of the node v_i by subtraction, which gives us the node that the model is most confident about and is the most representative.

$$I(v_i) = \rho(v_i) - H(v_i) \tag{11}$$

The problem has now become to choose the K_N nodes with largest influence score.

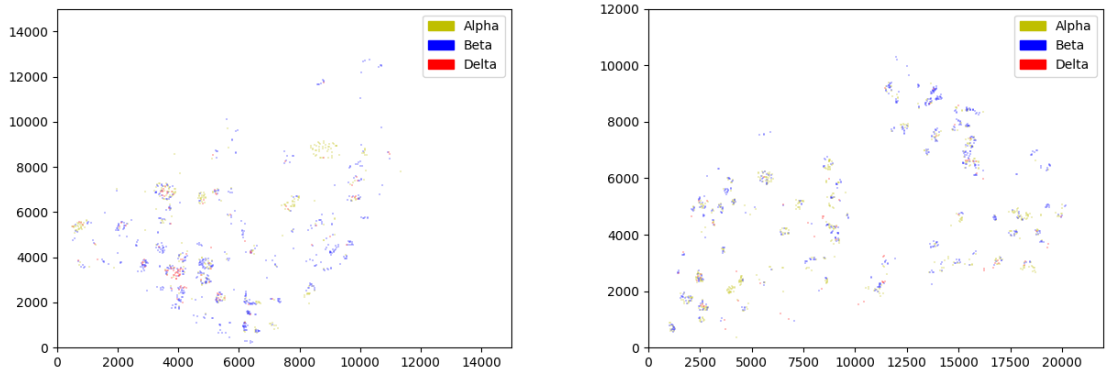
3 Experiment Setup

3.1 Image Segmentation

We use CODEX to segment an classify markers for 8 donors from 4 normal (ND) and 4 type 2 diabetes (T2D).

[To be written... –Martin]

The following are the visualizations for 2 of donor islets, colored according to the cell types.



(a) The visualization for ABHQ115-T2D-Islet

(b) The visualization for AFG1440-ND-Islet

Figure 1: The visualizations for 2 of donor islets

For each donor, we obtained the a list of segmented cells with their markers and coordinates. For each single cell, we obtain the $X_{\min}$, $X_{\max}$, $Y_{\min}$, $Y_{\max}$ coordinates of the box that encloses the indicated cell.

For islets, we are mainly concerned with the following 3 markers:

- Cpep: C-peptide, beta cells;
- GCG: glucagon, alpha cells;
- SST: somatostatin, delta cells.

For each marker, there are 2 values associated:

- Cytoplasm Intensity: Average intensity in the "cytoplasmic region", which we defined as 1um ring around the nucleus;
- Cytoplasm Completeness: Percent of cytoplasm with intensity exceeding the user-defined intensity threshold to call a cell positive. Usually set to around 60-70%.

3.2 Data Preprocessing

For experiment, we choose 3 donor samples: ABHQ115-T2D-Islet, ABIC495-T2D-Islet and AFG1440-ND-Islet. We filtered out all the unclassified cells (Either one of alpha, beta, delta) and obtained 3563, 2893, 5483 valid cells for each donor sample.

The feature x_i has dimension $d = 6$, composed by the Cytoplasm Intensity and Cytoplasm Completeness of Cpep, GCG and SST.

The classification is binary, *i.e.* $C = 2$. Thus, we use 0 to label cells from the normal donors and 1 to label the cells from the T2D donors: $y_i \in \{0, 1\}$.

The graph information is represented by the adjacency matrix of the data, constructed by connecting nearby cells (set the entry to 1). The number of edge is set to 10 times the total number of cells.

3.3 Model Training

In the past semi-supervised setting of citation networks, a two-layer GCN is the optimal structure for the node classification task [1]. Therefore, we constructed a GCN with 32 hidden neurons and no drop-out layer. An Adam optimizer is used with learning rate of 10^{-3} and default weight decay [6].

The data is partitioned evenly to training, validation and test sets. We train our model with 1000 epochs with patience of 50 epoch of growing cross entropy loss to avoid overfit.

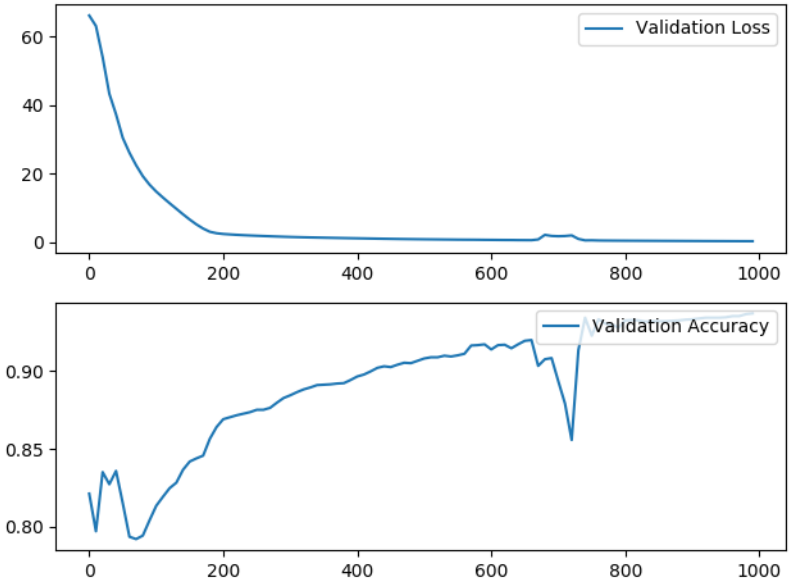
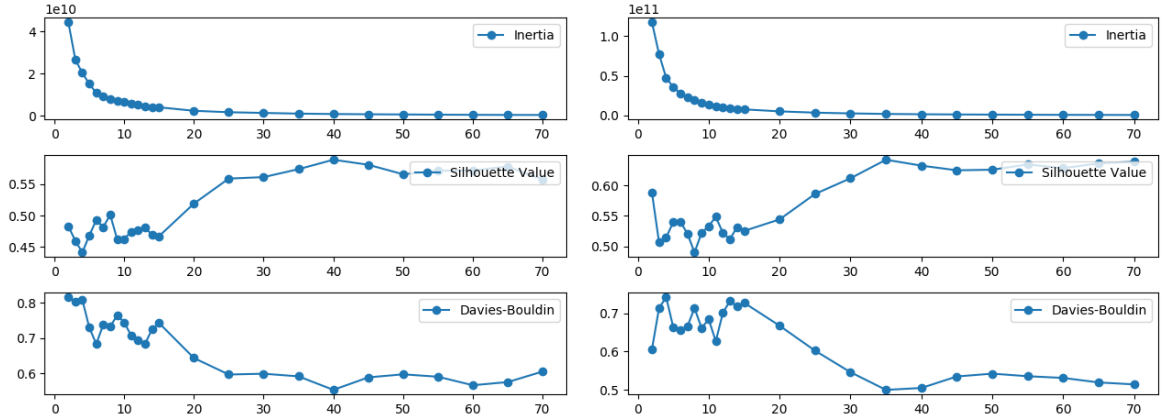


Figure 2: The GCN validation loss and accuracy

The final loss on test dataset is 1.544 and the accuracy on test dataset is 93.79%, which valify the ability of GCN model in our dataset.

3.4 Model Explanation

The number of representative cells K_N is approximated by the number of islets in each donor samples. This can be done by applying a KMeans clustering on the cell locations, and minimizing inertia and Davies-Bouldin score [7], and maximize the Silhouette score [8].



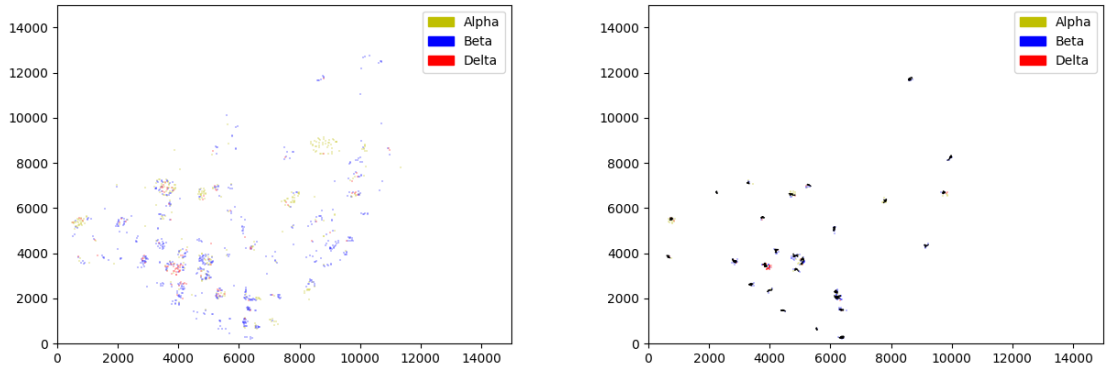
(a) The Kmeans results on ABHQ115-T2D-Islet (b) The Kmeans results on ABHQ115-T2D-Islet

Figure 3: The Kmeans results on samples

By observation, $K_N = 40$ for T2D sample, and $K_N = 35$ for ND sample.

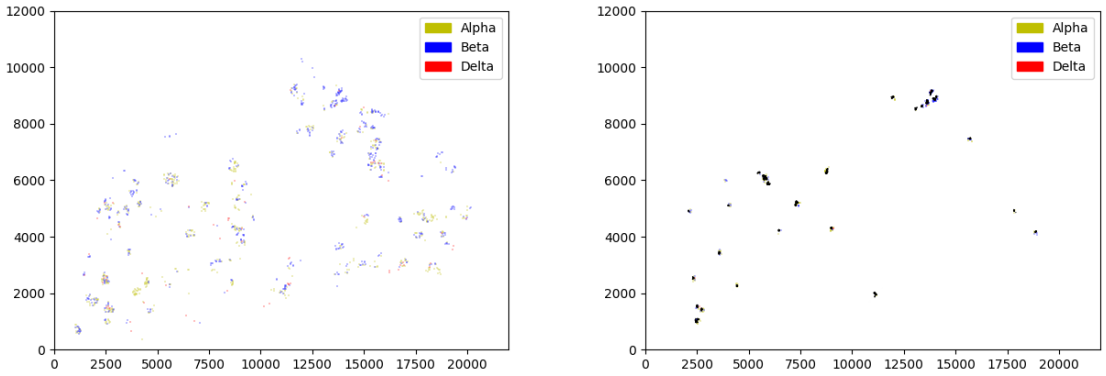
3.5 Preliminary Results

From the overview following, we can see that highly influential islets are selected by the model.



(a) The original overview for ABHQ115-T2D-Islet (b) The explained overview for ABHQ115-T2D-Islet

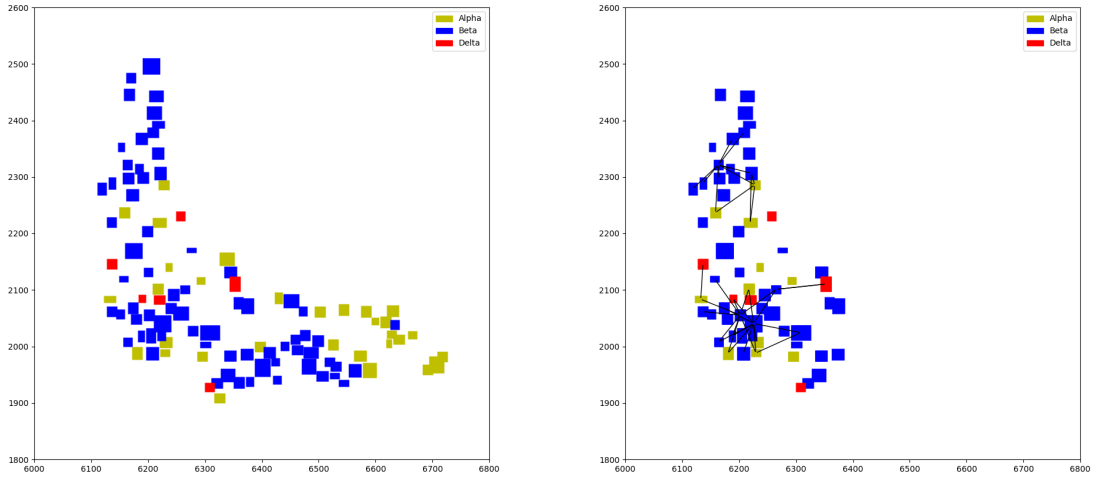
Figure 4: The visualizations for ABHQ115-T2D-Islet



(a) The original overview for AFG1440-ND-Islet (b) The explained overview for AFG1440-ND-Islet

Figure 5: The visualizations for AFG1440-ND-Islet

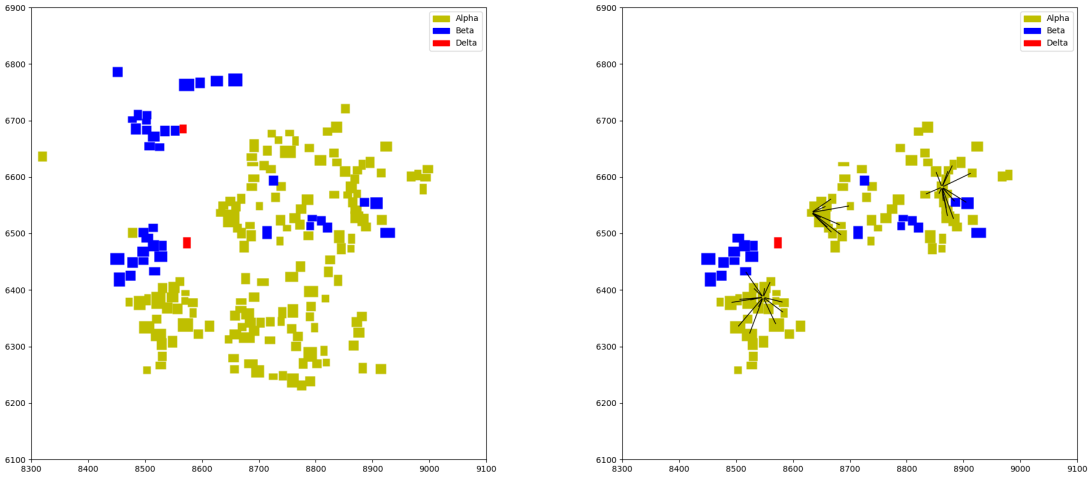
By zooming in and viewing specific islets in the T2D samples, one may conclude that clusters of beta cells are considered as the key local features determining if a sample has type-II diabete.



(a) The original view for a specific islet from ABHQ115-T2D-Islet (b) The explained view for a specific islet from ABHQ115-T2D-Islet

Figure 6: The visualizations for a specific islet from ABHQ115-T2D-Islet

Similarly, in the ND samples, one may conclude that clusters of alpha cells are considered as the key local features determining if a sample is normal.



(a) The original view for a specific islet from AFG1440-ND-Islet

(b) The explained view for a specific islet from AFG1440-ND-Islet

Figure 7: The visualizations for a specific islet from AFG1440-ND-Islet

By normalizing the cumulative marker strength, the following visualizing leads to the conclusion that Cpep intensity, the marker for beta cells, is responsible for GCN's decision of T2D classification, while GCG and SST intensity, the marker for alpha and delta cells, is responsible for GCN's decision of normal classification.

(a) The marker importance of normal label $c = 0$

(b) The marker importance of T2D label $c = 1$

Figure 8: The marker importance of each label, from left to right are Cpep intensity, Cpep Completeness, GCG intensity, GCG Completeness, SST intensity, SST Completeness

4 Conclusion

- Clusters of beta cells are considered as the key local features determining if a sample has type-II diabete, while clusters of alpha cells are considered as the key local features determining if a sample is normal.
- Cpep intensity, the marker for beta cells, is responsible for GCN's decision of T2D classification, while GCG and SST intensity, the marker for alpha and delta cells, is responsible for GCN's decision of normal classification.

References

[1] Thomas N. Kipf and Max Welling. *Semi-Supervised Classification with Graph Convolutional Networks*. 2016. arXiv: 1609.02907 [cs.LG].

[2] V. Nair and Geoffrey E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *ICML*. 2010.

[3] Rex Ying et al. *GNNExplainer: Generating Explanations for Graph Neural Networks*. 2019. arXiv: 1903.03894 [cs.LG].

[4] Burr Settles and Mark Craven. "An Analysis of Active Learning Strategies for Sequence Labeling Tasks". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. EMNLP '08. Honolulu, Hawaii: Association for Computational Linguistics, 2008, pp. 1070–1079.

[5] Hongyun Cai, Vincent W. Zheng, and Kevin Chen-Chuan Chang. *Active Learning for Graph Embedding*. 2017. arXiv: 1705.05085 [cs.LG].

[6] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: 1412.6980 [cs.LG].

[7] D. L. Davies and D. W. Bouldin. "A Cluster Separation Measure". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1.2 (1979), pp. 224–227.

[8] Peter J. Rousseeuw. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65. ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(87)90125-7. URL: http://www.sciencedirect.com/science/article/pii/0377042787901257.