Supplementary materials for paper: scE2EGAE: Enhancing Single-cell RNA-Seq Data Analysis through an End-to-End Cell-Graph-Learnable Graph Autoencoder with Differentiable Edge Sampling

Shuo Wang

October 17, 2024

1 Hyperparameters used by scE2EGAE and AblationGAE

Table 1: Hyperparameters used in scE2EGAE and AblationGAE. Definitions: **Num_Epochs** - maximum number of epochs with early-stopping; **Patience** - patience parameter for early-stopping; **K** - number of neighboring cells for edge sampling (in Ablation, it decides the number of neighbors in the KNN algorithm); **Distance_Measure** - function used to calculate cell-by-cell distances matrix; **AE_Dim** - dimension of hidden representations in the DCA module (exclusive to scE2EGAE); **GAE_Dim** - dimension of hidden representations in the graph encoder; **Dropout_GAE** - dropout rate in the graph encoder's encoder layer; **LR** - learning rate; **Alpha** - weighting parameter between ZINB loss in DCA module and MSE/MSLE loss in graph encoder (exclusive to scE2EGAE); **MSLE** - indicates whether MSLE loss function is used (True) or MSE loss function is used (False).

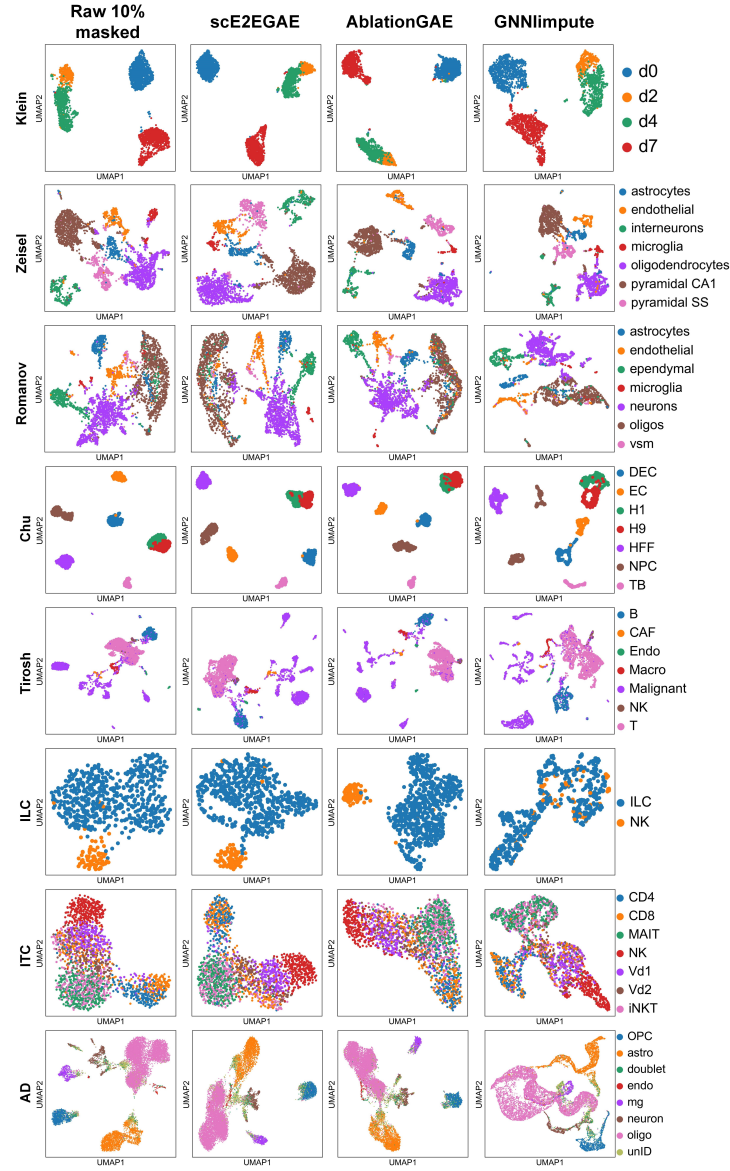| Datasets	Num_Epochs	Patience	K	Distance_Measure	AE_Dim	GAE_Dim	Dropout_GAE	LR	Alpha	MSLE
Klein	500	20	3	Hyperbolic	128	2000	0	0.003	0.0005	False
Zeisel	700	20	3	Hyperbolic	128	2000	0.1	0.003	0.001	True
Romanov	500	20	3	Hyperbolic	128	2000	0.1	0.003	0.0005	False
ITC	500	20	3	Hyperbolic	128	2000	0.1	0.003	0.001	False
Chu	2000	20	3	Euclidean	128	2000	0	0.003	0.001	False
ILC	4000	10	1	Hyperbolic	128	2000	0	0.003	0.001	False
Tirosh	800	10	1	Hyperbolic	128	128	0	0.003	0.001	False
AD	1500	20	1	Hyperbolic	128	64	0	0.003	0.001	False

2 UMAP visualization for the masked datasets

Figure 1: UMAP visualization of denoising results of scE2EGAE, AblationGAE, and GNNimpute on eight scRNA-Seq datasets with 10% masking rate. All of the results undergo dimensionality reduction to 40 dimensions, and the number of neighbors is set to 30.
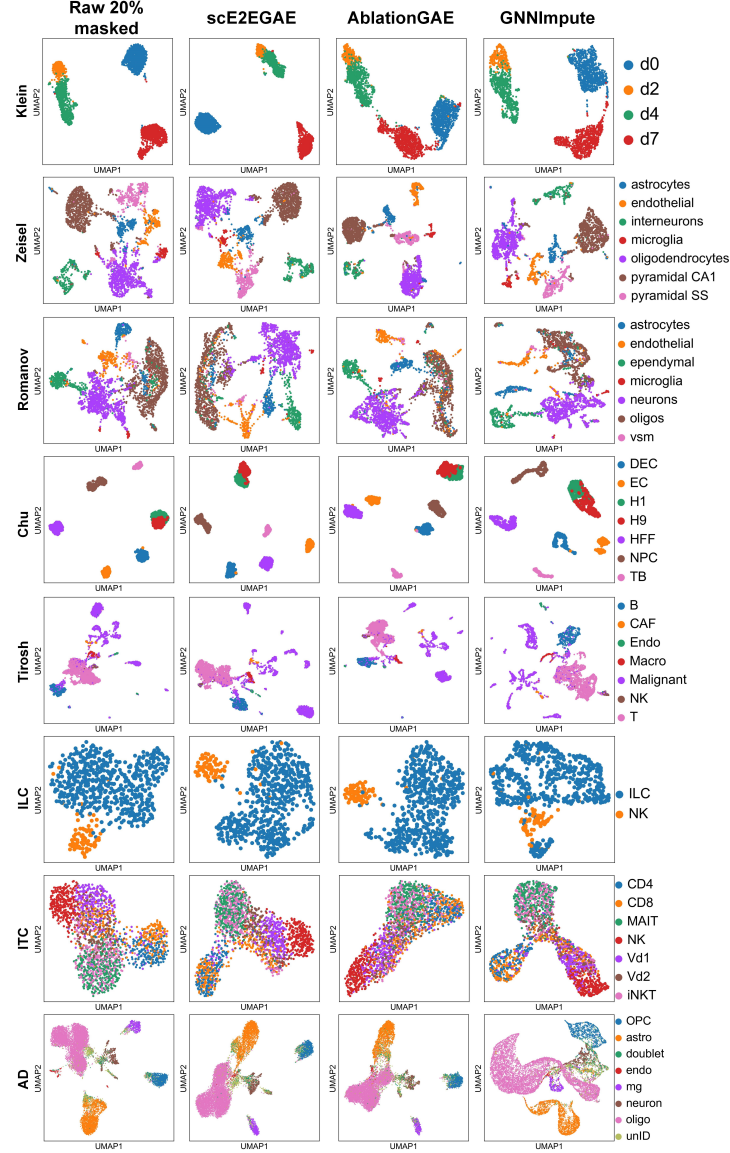
Figure 2: UMAP visualization of denoising results of scE2EGAE, AblationGAE, and GNNimpute on eight scRNA-Seq datasets with 20% masking rate. All of the results undergo dimensionality reduction to 40 dimensions, and the number of neighbors is set to 30.
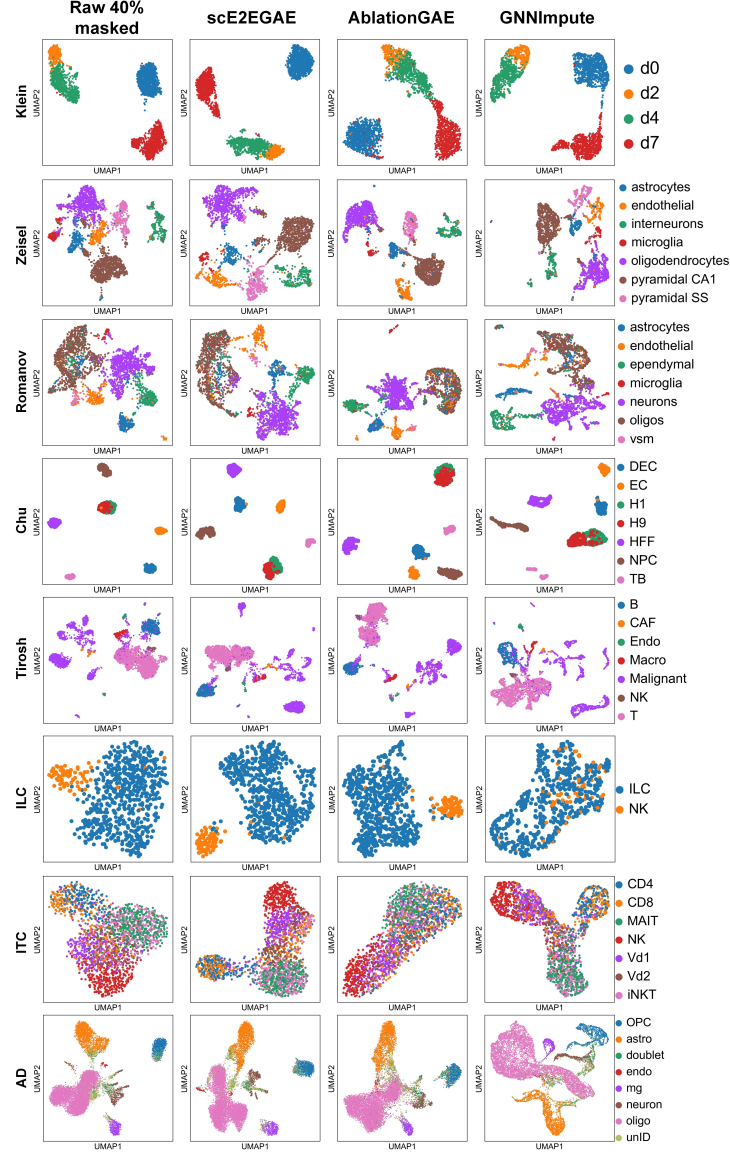
4

Figure 3: UMAP visualization of denoising results of scE2EGAE, AblationGAE, and GNNimpute on eight scRNA-Seq datasets with 40% masking rate. All of the results undergo dimensionality reduction to 40 dimensions, and the number of neighbors is set to 30.