**User Guide**

We used a Google Colab environment in Python to run the scripts of the MaskGAE model. You can find the Colab notebook link here:

<https://colab.research.google.com/drive/1zCf-xiN3_OkUpfkQVHhyaKJotRJpEa0R?usp=sharing>

**How to Run the MaskGAE scripts**

Follow these steps to run the MaskGAE model (exactly like in the Colab notebook):

1. (first cell) Install numpy, torch and other python libraries. Also, install all the torch extended libraries with the recent TORCH and CUDA versions.
2. (second cell) Cloning from git the repository and to enter the folder (you can check with `!ls` that the folder was cloned) or Pull the repository if you already cloned.
3. Customizing parameters to the script. Here is a list of parameters and their descriptions and default values:

"--dataset" - set the dataset (default: Cora)

"--mask" - set the Masking strategy, `Path`, `Edge` or `None` (default: Path)

"--seed" - set seed for model and dataset (default: 2022)

"--bn" - set whether to use batch normalization for GNN encoder (default: False)

"--layer" - set type of GNN layer (default: gcn)

"--encoder\_activation" - set activation function for GNN encoder (default: elu)

"--encoder\_channels" - set number of channels of GNN encoder (default: 128)

"--hidden\_channels" - set number of channels of hidden representation (default: 128)

"--decoder\_channels" - set number of channels of decoder (default: 64)

"--encoder\_layers" - set number of layers of encoder (default: 1)

"--decoder\_layers" - set number of layers for decoders (default: 2)

"--encoder\_dropout" - set dropout probability of encoder (default: 0.7)

"--decoder\_dropout" - set dropout probability of decoder (default: 0.3)

"--alpha" - set loss weight for degree prediction (default: 0.003)

"--lr" - set learning rate for training (default: 0.001)

"--weight\_decay" - set weight\_decay for training (default: 5e-5)

"--grad\_norm" - set grad\_norm for training (default: 1.0)

"--batch\_size" - set number of batch size (default: 2\*\*16)

"--start" - set which type to sample starting nodes for random walks (default: edge)

"--p" - set mask ratio or sample ratio for MaskEdge MaskPath (default: 0.7)

"--epochs" - set number of training epochs (default: 300)

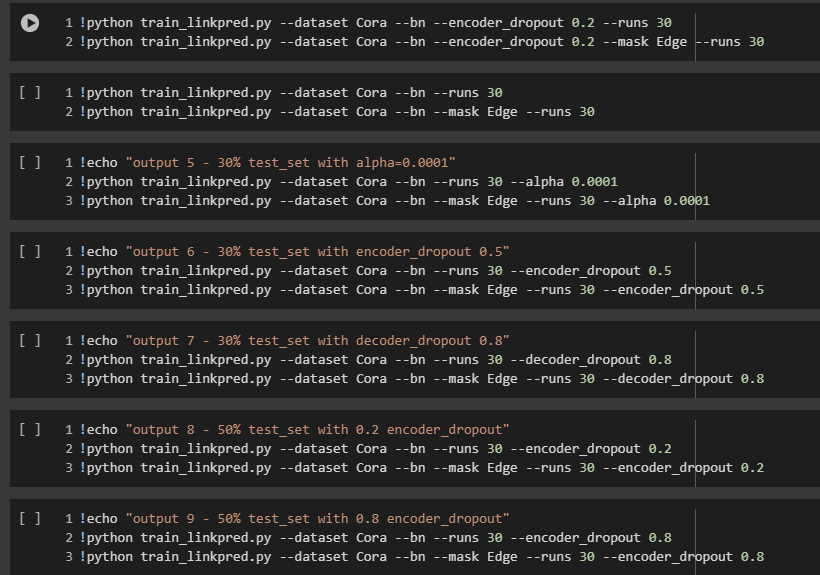
"--runs" - set number of runs (default: 10)

"--eval\_period" - set evaluation period (default: 10)

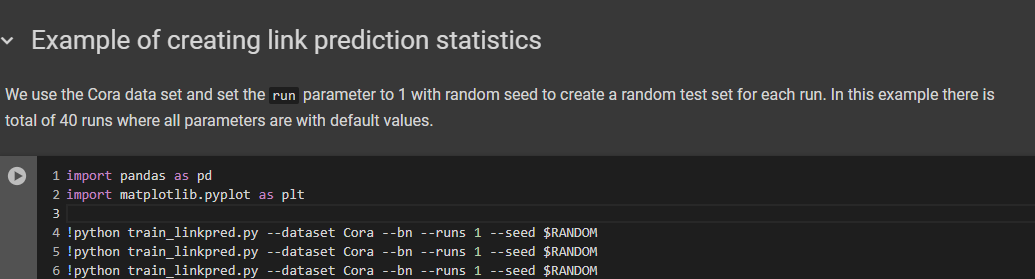
"--save\_path" - set path for model (default="MaskGAE-LinkPred.pt")

"--device" - set device (default=0)

You can view several examples of iterations in the next cells in the colab notebook:



1. If you wish to create a histogram that shows the distribution of correct predictions for each link, please move to following cell:



You will see in this cell 40 lines of calling the 'train\_linkpred.py' script with `runs` parameter set to '1' and the `seed` set to $RANDOM for making sure the test set is created randomly.

1. After you ran this cell, you will have the 'link\_statistics.csv' file which you can parse in the following cell to create the histogram.

