12.29

A100操作

代码的修改

因为A100所使用的torch版本的问题，需要对原有代码进行修改

```
#修改 GF_dataset.py 因为networkx的版本使用新的版本了
def _process_single_graph(self, graph: GraphData) -> Tuple[np.ndarray, np.ndarray, np.ndarray]:
    try:
        # 首先添加节点
        for i in range(graph.node_num):
            graph.adj.add_node(i)

        # 准备特征矩阵
        feature_dim = self.args.graph_init_dim
        feature_matrix = np.zeros((self.max_nodes, feature_dim))
        if len(graph.features.shape) == 1:
            graph.features = graph.features.reshape(1, -1)
        feature_matrix[:graph.node_num, :] = graph.features

        # 准备邻接矩阵
        adj_matrix = nx.adjacency_matrix(graph.adj).toarray()
        adj_matrix = adj_matrix + np.eye(adj_matrix.shape[0])
        adj_padded = np.zeros((self.max_nodes, self.max_nodes))
        adj_padded[:adj_matrix.shape[0], :adj_matrix.shape[1]] = adj_matrix

        # 准备掩码
        mask = np.zeros(self.max_nodes)
        mask[:graph.node_num] = 1

        return feature_matrix, adj_padded, mask
    except Exception as e:
        print(f"Error processing graph {graph.name}: {str(e)}")
        print(f"Graph info: nodes={graph.node_num}, features shape={graph.features.shape}")
        raise
#GF_config.py
parser.add_argument('--gpu_index', type=str, default='1',
                    help="GPU index to use")
parser.add_argument('--log_path', type=str, default='../logs/',
```

```
                    help='path for log files')
parser.add_argument('--seed', type=int, default=42,
                    help='random seed for reproducibility')
```

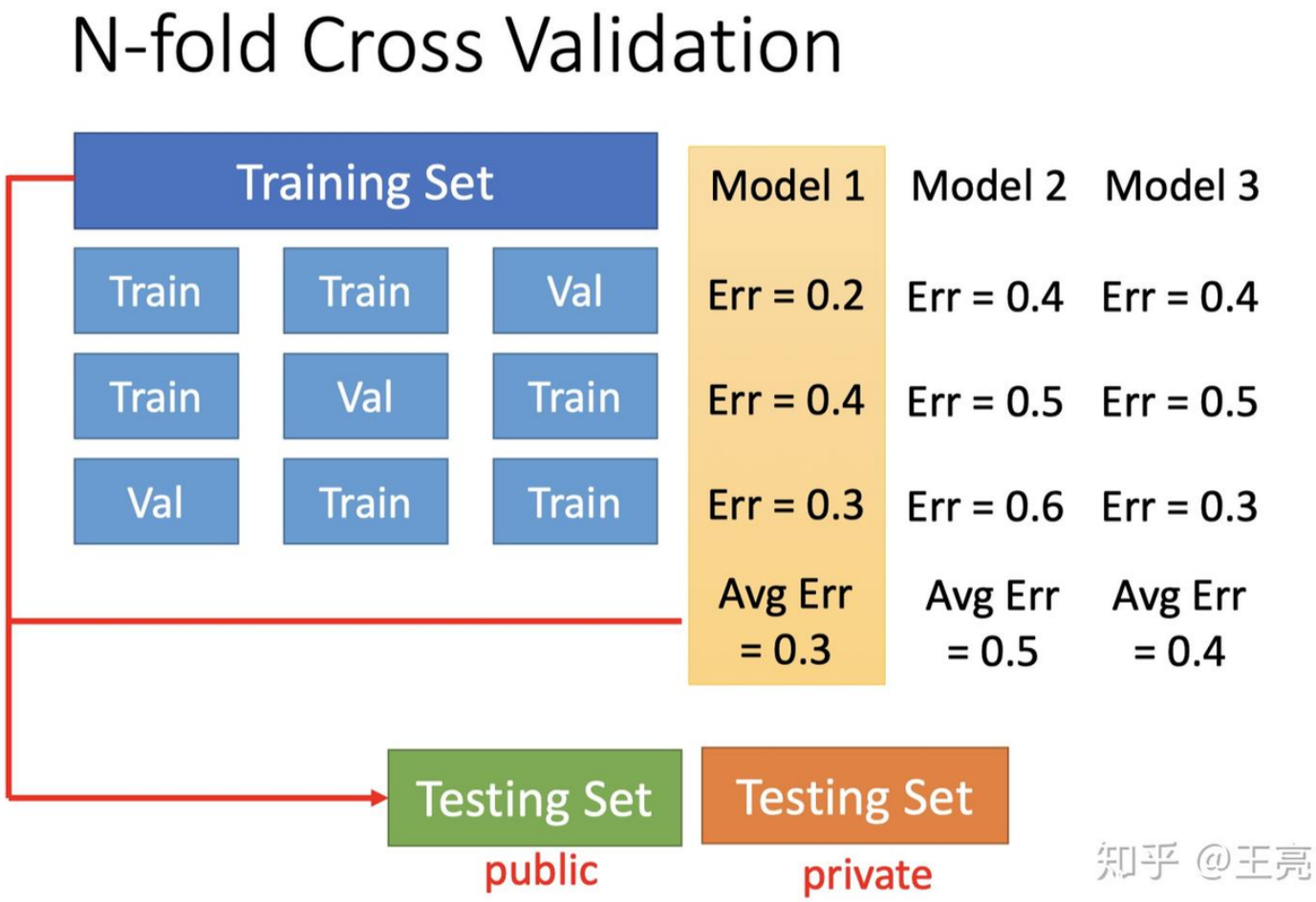
远程连接

```
pip install -i https://mirrors.aliyun.com/pypi/simple/ networkx
torch_geometric wandb scikit-learn
pip install --upgrade pip -i https://mirrors.aliyun.com/pypi/simple/
wandb login
wandb status#查看wandb相关配置信息
export WANDB_ENTITY="102201525-fuzhou-university"#需要针对性修改
export WANDB_PROJECT="GF"#需要针对性修改
cd /home/vllm/encode/
python GF_main.py
```

```
#后台训练
sudo apt-get install screen
'''
#创建或修改screen的配置文件 ~/.screenrc:
# 在 ~/.screenrc 中添加以下内容
defutf8 on
defencoding utf8
encoding utf8 utf8
'''
screen -U -S 241127  # "241127"是会话名称，你可以换成任何名字
python GF_main.py
Ctrl + A #然后+D,退出到命令行界面，把模型放后台运行
screen -ls #查看会话
screen -r 241127#进行会话
screen -S 54806.241127 -X quit#终止会话（慎用，我的在训练）
screen -U -S 241210
./run_main.sh
```

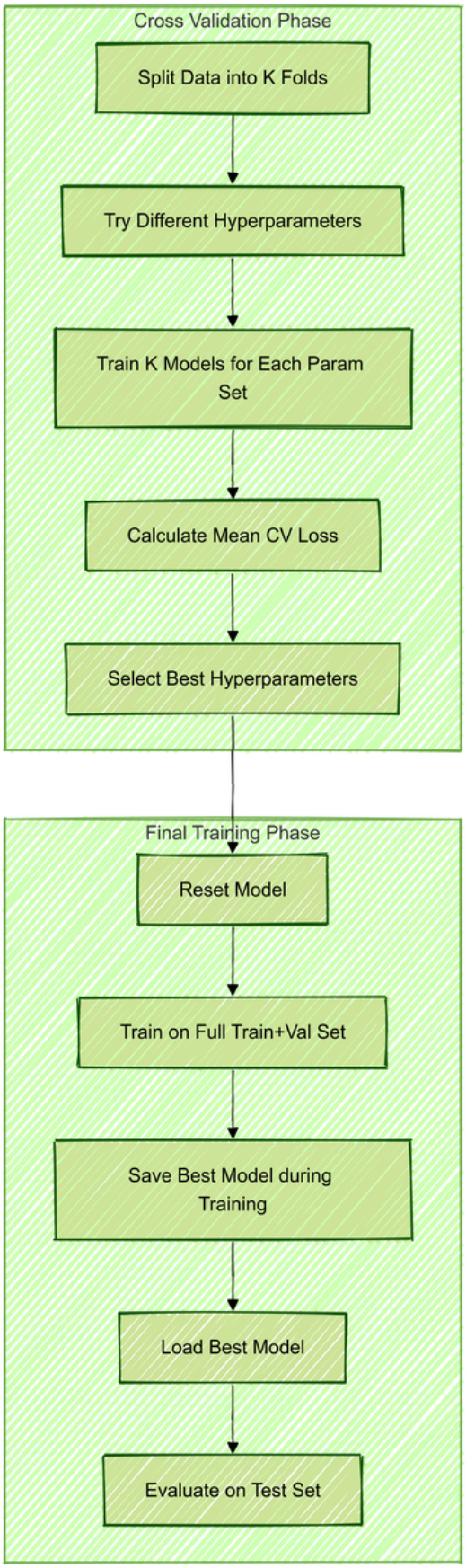
k折交叉验证的修正

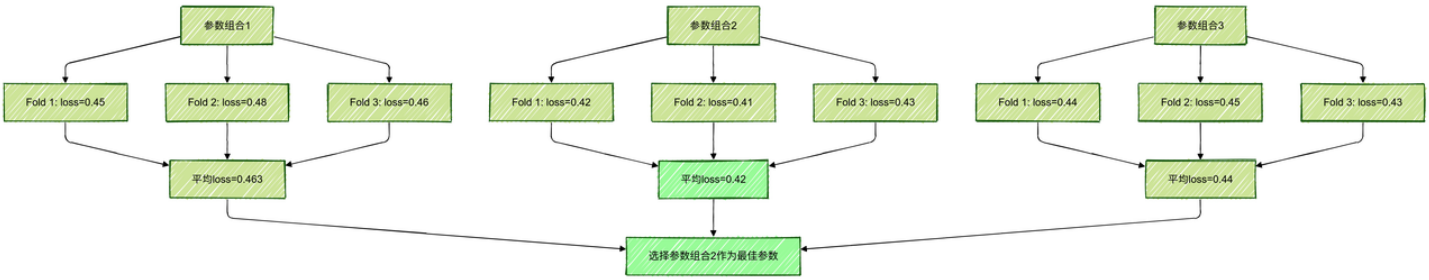
目的：用于超参数选择，从多种模型中选择出泛化能力最好的（即最不容易发生过拟合）的模型，消除单次划分时数据划分得不平衡而造成的不良影响（因为这种不良影响在小规模数据集上更容易出现）



训练流程

```
数据集划分:
- 总数据集先划分为80%训练验证集和20%测试集
- 80%的训练验证集再进行10折交叉验证划分
超参数组合:
- lr: [0.0001, 0.001]
- dropout: [0.1, 0.3]
- tau: [0.5, 0.7]
总共有 2 × 2 × 2 = 8 组超参数组合
对每组超参数:
- 进行完整的10折交叉验证
- 记录每一折的验证损失
- 计算平均验证损失和标准差
最终选择:
- 选择平均验证损失最低的超参数组合作为最佳参数
- 使用最佳参数在完整训练集重新训练最终模型
- 使用预留的测试集进行最终评估
```


Cross Validation Phase
Split Data into K Folds
Try Different Hyperparameters
Train K Models for Each Param Set
Calculate Mean CV Loss
Select Best Hyperparameters
Final Training Phase
Reset Model
Train on Full Train+Val Set
Save Best Model during Training
Load Best Model
Evaluate on Test Set



GF Training results

Cross-validation loss是8组超参数的交叉损失的平均值，每组超参数得到交叉损失是10折交叉验证的平均值

GF-Dismantle	Finished	Add notes	102201525	1w ago	7d 20h 57m 9s	-	gcn
GF-Dismantle	Finished	Add notes	102201525	2w ago	4d 19h 29m 25s	-	gcn

batch_size=256

Config parameters: {} 10 keys

- architecture: "gcn"
- batch_size: 256
- dataset: "PROTEINS"
- dropout: 0.1
- epochs: 50
- filters: "100_100_100"
- graph_init_dim: 6
- learning_rate: 0.0001
- model_parameters: 152,100
- patience: 30

Summary metrics: {} 86 keys

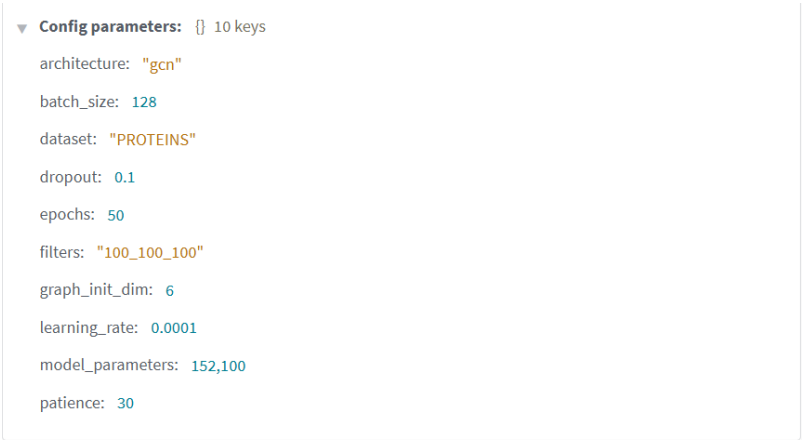
- final_cv_mean_loss: 5.9078251198175895
- final_cv_std_loss: 0.02515972198159657
- final_test_loss: 5.988961420978064
- final_training/epoch: 49
- final_training/train_loss: 5.988961513393636
- fold_1/best_val_loss: 5.9889614189239895
- fold_1/epoch: 49
- fold_1/learning_rate: 0.001
- fold_1/params.dropout: 0.3
- fold_1/params.lr: 0.001
- fold_1/params.tau: 0.7
- fold_1/train_loss: 5.9889614862859695
- fold_1/val_loss: 5.988961419504554
- fold_2/best_val_loss: 5.831152406411682

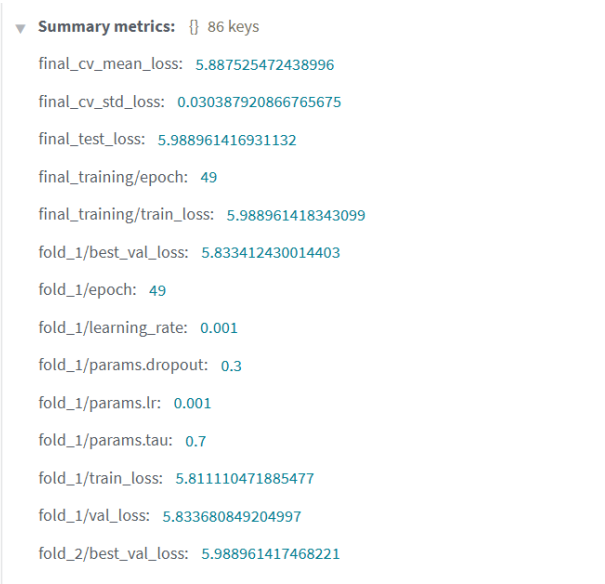
```
Hyperparameter search results:
2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.0001, 'dropout': 0.1, 'tau': 0.5}, Mean loss: 5.942475 ± 0.048375
2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.0001, 'dropout': 0.1, 'tau': 0.7}, Mean loss: 5.925134 ± 0.052241
2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.0001, 'dropout': 0.3, 'tau': 0.5}, Mean loss: 5.913993 ± 0.044732
2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.0001, 'dropout': 0.3, 'tau': 0.7}, Mean loss: 5.931278 ± 0.050139
2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.001, 'dropout': 0.1, 'tau': 0.5}, Mean loss: 5.879149 ± 0.056608
2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.001, 'dropout': 0.1, 'tau': 0.7}, Mean loss: 5.895398 ± 0.062881
```

```
2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.001, 'dropout': 0.3, 'tau': 0.5}, Mean loss: 5.863298 ± 0.065924
2024-11-26 23:13:33,890 - GF - INFO - Params: {'lr': 0.001, 'dropout': 0.3, 'tau': 0.7}, Mean loss: 5.911876 ± 0.077226
2024-11-26 23:13:34,602 - GF - INFO -
Training final model with best parameters: {'lr': 0.001, 'dropout': 0.3, 'tau': 0.5}
2024-11-27 00:35:42,137 - GF - INFO - Final test loss: 5.988961
2024-11-27 00:35:42,139 - GF - INFO - Training completed!
2024-11-27 00:35:42,139 - GF - INFO - Cross-validation loss: 5.907825 ± 0.025160
2024-11-27 00:35:42,139 - GF - INFO - Test loss: 5.988961

```

batch_size=128






```
4026 2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.0001, 'dropout': 0.1, 'tau': 0.5}, Mean loss: 5.942475 ± 0.048375
4027 2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.0001, 'dropout': 0.1, 'tau': 0.7}, Mean loss: 5.925134 ± 0.052241
4028 2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.0001, 'dropout': 0.3, 'tau': 0.5}, Mean loss: 5.913993 ± 0.044732
4029 2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.0001, 'dropout': 0.3, 'tau': 0.7}, Mean loss: 5.931278 ± 0.050139
4030 2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.001, 'dropout': 0.1, 'tau': 0.5}, Mean loss: 5.879149 ± 0.056608
4031 2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.001, 'dropout': 0.1, 'tau': 0.7}, Mean loss: 5.895398 ± 0.062881
4032 2024-11-26 23:13:33,889 - GF - INFO - Params: {'lr': 0.001, 'dropout': 0.3, 'tau': 0.5}, Mean loss: 5.863298 ± 0.065924
```

```
4033 2024-11-26 23:13:33,890 - GF - INFO - Params: {'lr': 0.001, 'dropout': 0.3, 'tau': 0.7}, Mean loss: 5.911876 ± 0.077226
4034 2024-11-26 23:13:34,602 - GF - INFO -
4035 Training final model with best parameters: {'lr': 0.001, 'dropout': 0.3, 'tau': 0.5}
4129 2024-12-05 11:35:16,727 - GF - INFO - Cross-validation loss: 5.887525 ± 0.030388
4130 2024-12-05 11:35:16,727 - GF - INFO - Test loss: 5.988961
```

两组结果参数选择都一致'lr': 0.001, 'dropout': 0.3, 'tau': 0.5，但是都过拟合了