A Datasets

- **UK Biobank (UKB)** [24]: Consisting of 40,913 samples, the UK Biobank dataset stands as one of the most comprehensive fMRI datasets available. The dataset includes extensive demographic information, such as gender and age (between 40 and 70 years), which are valuable for various pre-training tasks.
- **Adolescent Brain Cognitive Development (ABCD)** [5]: This dataset comprises 9,111 samples from children and adolescents aged 9 to 11 years, focusing on their development. It includes demographic information on gender and age, useful for developmental studies and can be utilized alongside the UKB dataset for pre-training.
- **Human Connectome Project (HCP) Young Adults** [27]: The HCP Young Adults dataset includes 1,093 samples from participants aged 22 to 37 years, providing a valuable resource for studying brain connectivity in young adults.
- **Human Connectome Project (HCP) Aging** [3]: The HCP-A dataset, with 724 samples, focuses on older adults aged 36 to 90 years, offering insights into brain changes and development in this age group.
- **Human Connectome Project (HCP) Development** [23]: The HCP-D dataset, consisting of 632 samples, targets the developmental stages of children and adolescents, encompassing ages from 8 to 21 years. It provides gender and age data for detailed developmental analyses.
- **Autism Brain Imaging Data Exchange (ABIDE)** [7]: The ABIDE dataset includes 884 clinical samples and provides Autism Spectrum Disorder (ASD) labels, making it useful for benchmarking psychiatric diagnosis classification tasks.
- **ADHD200** [4]: This dataset includes 669 clinical samples and contains labels for Normal and ADHD conditions, serving as a useful resource for benchmarking psychiatric diagnosis classification.

B Baseline Graph Self-supervised Methods

- **Deep Graph Infomax (DGI)** [26]: DGI aims to maximize the mutual information between node representations and global graph representations. A discriminator is trained to differentiate between the original graph and a permuted version, thereby learning meaningful node and graph representations.
- **Graph Auto-Encoder (GAE)** [16]: GAE employs an autoencoder architecture to reconstruct the original graph from node representation. It utilizes Graph Convolutional Network (GCN) encoder to infer node embedding $\boldsymbol{Z}$ using node feature $\boldsymbol{X}$ and adjacency matrix $\boldsymbol{A}$, and uses them to reconstruct the original links of graph.
- **Variational Graph Auto-Encoder (VGAE)** [16]: VGAE extends GAE by introducing stochasticity in the encoder layer. The encoder outputs the mean and standard deviation of embedding $\boldsymbol{z}$, from which node representations are sampled. These sampled representations are then used to reconstruct the original graph. The reconstruction is given by $\hat{\boldsymbol{A}} = \sigma(\boldsymbol{Z}\boldsymbol{Z}^T)$, where $\boldsymbol{Z} = \mathrm{GCN}(\boldsymbol{X}, \boldsymbol{A})$).
- **SimGRACE** [28]: Unlike traditional Graph Contrastive Learning (GCL) methods that use graph augmentations to create multiple views, SimGRACE perturbs the model weights to generate different views. This approach eliminates the need for dataset-specific augmentations, making it a more universally applicable method.
- **Spatio-Temporal Deep Graph Infomax (ST-DGI)** [21]: ST-DGI extends DGI to spatio-temporal graphs by maximizing mutual information between node embeddings $\boldsymbol{h}_v(t)$ and subsequent time step features $\boldsymbol{x}_v(t+k)$, while minimizing it with corrupted features, thus capturing both spatial and temporal dynamics of the graph, where k serves as a hyperparameter that determines the time step difference.
- **Graph Masked AutoEncoder (GraphMAE)** [11]: GraphMAE advances self-supervised learning in graph domains by focusing on masked feature reconstruction rather than traditional edge reconstruction. It introduces a unique masking strategy and utilizes scaled cosine error for feature reconstruction. Its successor, GraphMAE2, further enhances the model by introducing additional regularization techniques for better performance [12].