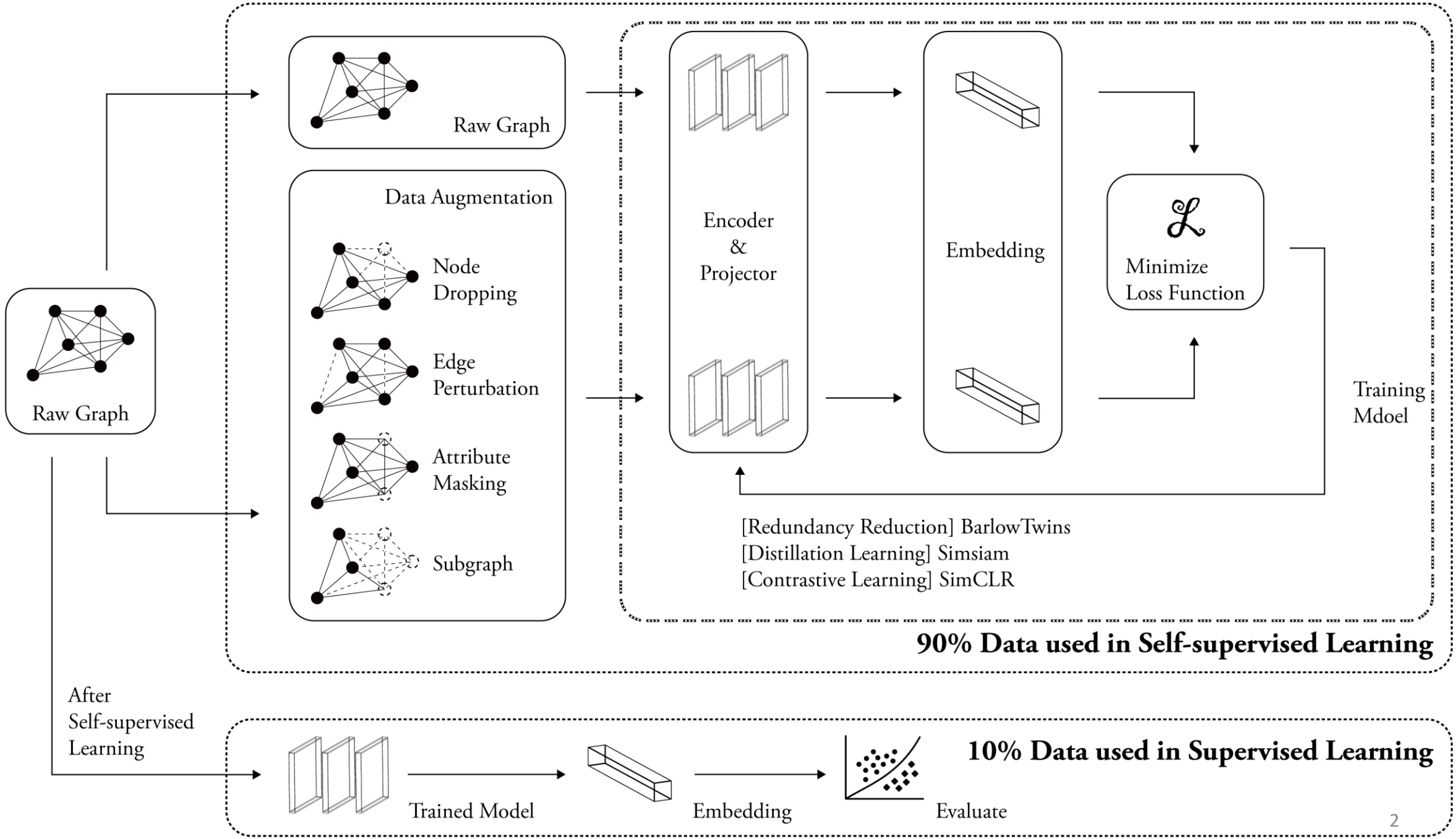
Project Report

HO Ching-Ru

October 25th, 2021


Raw Graph
Raw Graph
Data Augmentation
Node Dropping
Edge Perturbation
Attribute Masking
Subgraph
Encoder & Projector
Embedding
Minimize Loss Function
Training Mdoel
[Redundancy Reduction] BarlowTwins
[Distillation Learning] Simsiam
[Contrastive Learning] SimCLR
90% Data used in Self-supervised Learning
After Self-supervised Learning
Trained Model
Embedding
Evaluate
10% Data used in Supervised Learning

Detail of TUDataset

	MUTAG	PROTEINS	NCI1	DD	COLLAB	RDT-B
# Graphs	188	1113	4110	1178	5000	2000
Classes	2	2	2	2	3	2
Avg. Nodes	17.93	39.06	29.87	284.32	74.49	429.63
Avg. Edges	19.79	72.82	32.30	715.66	2457.78	497.75
Node Labels	+	+	+	+	-	-
Edge Labels	+	-	-	-	-	-
Node Attr.	-	+	+	-	-	-

Reference: https://chrsmrrs.github.io/datasets/docs/datasets/

Detail of Models (Total: 144)

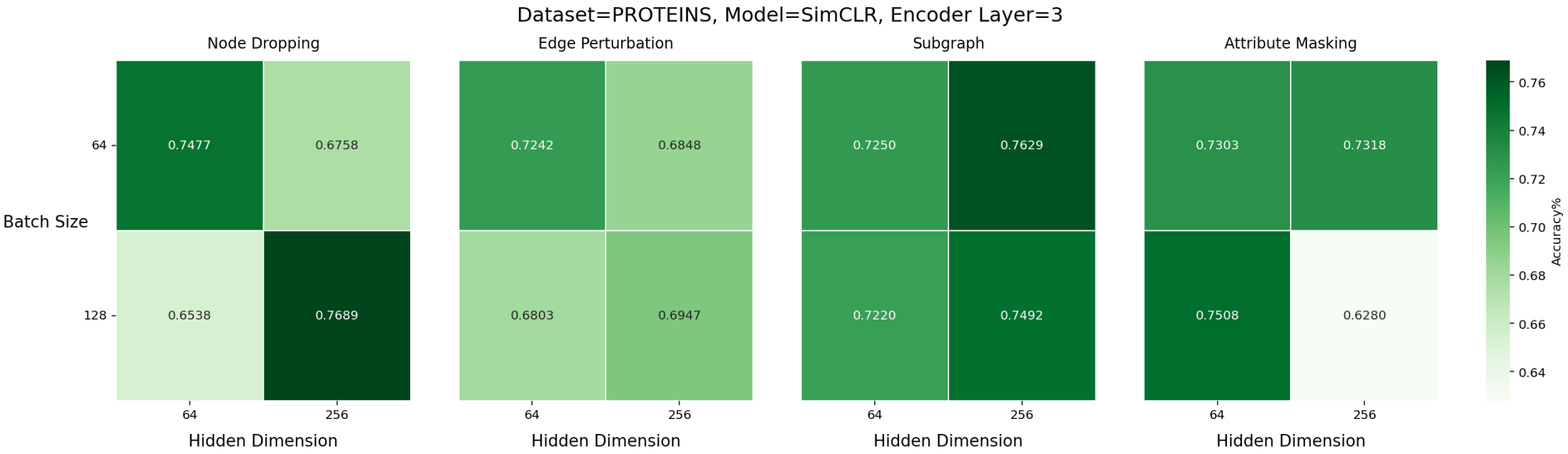
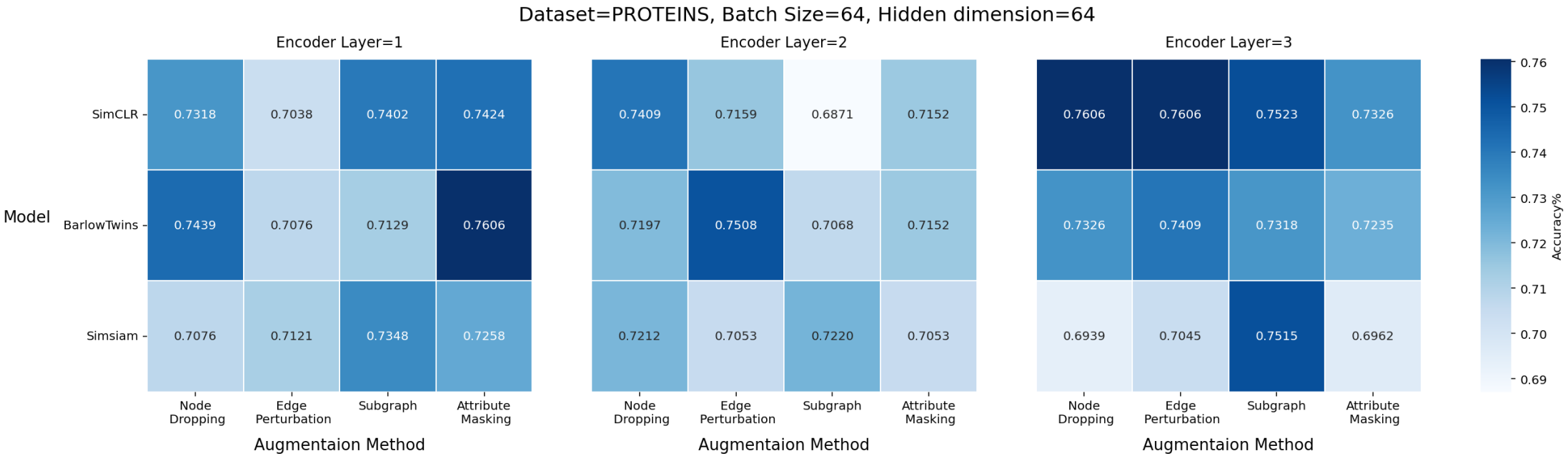
Model	Augmentation Method	Batch Size	Hidden Dimension	# Encoder Layers	# Projector Layers	Learning Rate	Epoch	Data Prop.
SimCLR	Node Dropping	64	64	1	3	0.001	200	90% Self-supervised
BarlowTwins	Edge Perturbation	256	512	2				10% Supervised
Simsiam	Attribute Masking			3				
	Subgraph			[Linear] [ReLU] [Linear]	[Linear] [BatchNorm] [ReLU]			

Detail of Experiment (144, repeat 5 times)

	MUTAG	PROTEINS	NCI1	DD	COLLAB	RDT-B
Top Ranking	100.00	84.91	87.20	95.67	95.62	93.15
10% Baseline		70.40±1.54	73.72±0.24	73.56±0.41	73.71±0.27	86.63±0.27
10% Aug.		70.29±0.64	73.59±0.32	74.30±0.81	74.19±0.13	87.74±0.39
10% GAE		70.51±0.17	74.36±0.24	74.54±0.68	75.09±0.19	87.69±0.40
10% Infomax		72.27±0.40	74.86±0.26	75.78±0.34	73.76±0.29	88.66±0.95
10% GraphCL		74.17±0.34	74.63±0.25	76.17±1.37	74.23±0.21	89.11±0.19
10% SimCLR	97.00	80.56	74.23*			
10% BarlowTwins	93.99	83.19	73.24*			
10% Simsiam	95.00	75.89	73.46*			
Average Cost Time (Training+Embed)	41.47 sec. per Exp.	438 sec. per Exp.	1.73 hr. per Exp.			

Reference: You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., & Shen, Y. (2020). Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, *33*, 5812-5823.

Result in Heatmap (PROTEINS)



Result in Heatmap (PROTEINS)