A IMPLEMENTATION DETAILS

A.1 DATASETS

We evaluate BRGCL on five public benchmarks that are widely used for node representation learning, namely Cora, Citeseer, PubMed (Sen et al., 2008), Coauthor CS, and ogbn-arxiv (Hu et al., 2020). Cora, Citeseer and PubMed are three most widely used citation networks. Coauthor CS is co-authorship graph. The ogan-arxiv is a directed citation graph. We summarize the statistics of the datasets in Table 3. Among the five benchmarks, ogbn-arxiv is known for its larger scale, and is more challenging to deal with. For all our experiments, we follow the default separation of training, validation, and test sets on each benchmark.

Table 3: The statistics of the datasets.

Dataset	**Nodes**	**Edges**	**Features**	**Classes**
Cora	2,708	5,429	1,433	7
CiteSeer	3,327	4,732	3,703	6
PubMed	19,717	44,338	500	3
Coauthor CS	18,333	81,894	6,805	15
ogbn-arxiv	169,343	1,166,243	128	40

A.2 MORE DETAILS ABOUT NODE CLASSIFICATION

The robust node representations are used to perform node classification and node clustering mentioned in Section 3 of the main paper. More details about node classification are introduced in this subsection. As the connected neighbors in a graph usually show similar semantic information, we generate soft labels of nodes via label propagation on the graph to take the advantage of the information from the neighborhood. The classifier for node classification is trained with soft labels instead of hard labels.

First, we define the one-hot hard label matrix $\mathbf{Y} \in \mathbb{R}^{N \times K}$, where $\mathbf{Y}_{ij} = 1$ if and only if node v_i is in class j for $i \in [N]$ and $j \in [K]$. If a node $v_i \in \mathcal{V}_{\mathcal{L}}$, then $\mathbf{Y}_{ij} = 1$ if the ground truth label of v_i is j. If $v_i \notin \mathcal{V}_{\mathcal{L}}$, we initialize $\mathbf{Y}_{ij} = 0$ for all $j \in [K]$. Then the soft labels of all the nodes are generated by graph label propagation. Similar to (3), after T aggregation steps of label propagation, we have $\mathbf{Y}^{(t+1)} = (1-\beta)\tilde{\mathbf{A}}\mathbf{Y}^{(t)} + \beta\mathbf{Y}^{(0)}, t = 1, ..., T-1$, where $\mathbf{Y}^{(0)} = \mathbf{Y}$, β is the teleport probability. The soft labels are then obtained by $\tilde{\mathbf{Y}} = \text{softmax}(\mathbf{Y}^{(T)})$. We denote the i-th row of $\tilde{\mathbf{Y}}$ by $\tilde{\boldsymbol{y}}_i$, which is the soft label of node v_i. $f(\cdot)$ is a classifier built by a two-layer MLP followed by a softmax function, which is trained by minimizing the standard loss function for classification, $\mathcal{L}_{\text{cls}} = \frac{1}{|\mathcal{V}_{\mathcal{L}}|} \sum_{v_i \in \mathcal{V}_{\mathcal{L}}} H(\tilde{\boldsymbol{y}}_i, f(\mathbf{h}_i))$, where H is the cross-entropy function.

A.3 TUNING HYPER-PARAMETERS BY CROSS-VALIDATION

In this section, we show the tuning procedures on the hyper-parameters ξ from Equation (2) and γ_0 from Equation (5). We perform cross-validations on 20% of training data to decide the value of ξ and γ_0. The value of ξ is selected from $\{0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0, 5\}$. The value of γ_0 is selected from $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. The selected values for ξ and γ_0 on each dataset are shown in Table 4.

Table 4: Selected hyper-parameters for each dataset.

Dataset	Cora	Citeseer	PubMed	Coauthor CS	ogbn-arxiv
ξ	0.20	0.15	0.35	0.40	0.25
γ_0	0.3	0.5	0.7	0.4	0.4

B Additional Experiment Results

B.1 Additional Node Classification Results with Label Noise and Attribute Noise.

The results for node classification with symmetric label noise and asymmetric label noise on Cora, Citeseer, and Coauthor CS are shown in Table 5. In addition, the results for node classification with attribute noise on Cora, Citeseer, and Coauthor CS are shown in Table 6. It is observed that BRGCL outperforms all the baselines for node classification with both label noise and attribute noise on all three benchmark datasets.

Table 5: Performance comparison on node classification with symmetric label noise and asymmetric label noise.

Dataset	Methods	Noise Level			
		0	40	60	80
		-	Asymmetric / Symmetric	Asymmetric / Symmetric	Asymmetric / Symmetric
Cora	GCN *	0.817±0.005	0.547±0.015 / 0.636±0.007	0.405±0.014 / 0.517±0.010	0.265±0.012 / 0.354±0.014
	S^{2}GC *	0.831±0.002	0.569±0.007 / 0.664±0.007	0.422±0.010 / 0.535±0.010	0.279±0.014 / 0.366±0.014
	GCE	0.819±0.004	0.573±0.011 / 0.652±0.008	0.449±0.011 / 0.509±0.011	0.280±0.013 / 0.353±0.013
	UnionNET	0.820±0.006	0.569±0.014 / 0.664±0.007	0.452±0.010 / 0.541±0.010	0.283±0.014 / 0.370±0.011
	NRGNN	0.822±0.006	0.571±0.019 / 0.676±0.007	0.470±0.014 / 0.548±0.014	0.282±0.022 / 0.373±0.012
	GraphCL	0.815±0.005	0.560±0.011 / 0.661±0.009	0.450±0.017 / 0.541±0.012	0.270±0.018 / 0.368±0.017
	MVGRL	0.829±0.007	0.566±0.009 / 0.672±0.009	0.455±0.014 / 0.545±0.014	0.275±0.014 / 0.379±0.014
	MERIT	0.831±0.006	0.560±0.008 / 0.670±0.008	0.467±0.013 / 0.547±0.013	0.277±0.013 / 0.385±0.013
	SUGRL	**0.834±0.005**	0.564±0.011 / 0.674±0.012	0.468±0.011 / 0.552±0.011	0.280±0.012 / 0.381±0.012
	BRGCL	0.822±0.006	**0.584±0.009 / 0.694±0.007**	**0.484±0.013 / 0.567±0.013**	**0.295±0.012 / 0.394±0.012**
Citeseer	GCN *	0.703±0.005	0.475±0.023 / 0.501±0.013	0.351±0.014 / 0.341±0.014	0.291±0.022 / 0.281±0.019
	S^{2}GC *	0.727±0.005	0.488±0.013 / 0.528±0.013	0.363±0.012 / 0.367±0.014	0.304±0.024 / 0.284±0.019
	GCE	0.705±0.004	0.490±0.016 / 0.512±0.014	0.362±0.015 / 0.352±0.010	0.309±0.012 / 0.285±0.014
	UnionNET	0.706±0.006	0.499±0.015 / 0.547±0.014	0.379±0.013 / 0.399±0.013	0.322±0.021 / 0.302±0.013
	NRGNN	0.710±0.006	0.498±0.015 / 0.546±0.015	0.382±0.016 / 0.412±0.016	0.336±0.021 / 0.309±0.018
	GraphCL	0.715±0.008	0.479±0.017 / 0.534±0.016	0.373±0.015 / 0.411±0.014	0.331±0.017 / 0.297±0.016
	MVGRL	0.726±0.007	0.491±0.013 / 0.541±0.013	0.379±0.013 / 0.420±0.013	0.341±0.016 / 0.301±0.016
	MERIT	**0.740±0.007**	0.496±0.012 / 0.536±0.012	0.383±0.011 / 0.425±0.011	0.344±0.014 / 0.301±0.014
	SUGRL	0.730±0.005	0.493±0.011 / 0.541±0.011	0.376±0.009 / 0.421±0.009	0.339±0.010 / 0.305±0.010
	BRGCL	0.722±0.004	**0.510±0.013 / 0.569±0.013**	**0.403±0.012 / 0.433±0.014**	**0.359±0.013 / 0.321±0.014**
Coauthor CS	GCN *	0.918±0.001	0.645±0.009 / 0.656±0.006	0.511±0.013 / 0.501±0.009	0.429±0.022 0.389±0.011
	S^{2}GC*	**0.928±0.001**	0.657±0.012 / 0.663±0.006	0.516±0.013 / 0.514±0.009	0.437±0.020 0.396±0.010
	GCE	0.922±0.003	0.662±0.017 / 0.659±0.007	0.515±0.016 / 0.502±0.007	0.443±0.017 0.389±0.012
	UnionNET	0.918±0.002	0.669±0.023 / 0.671±0.013	0.525±0.011 / 0.529±0.011	0.458±0.015 0.401±0.011
	NRGNN	0.919±0.002	0.678±0.014 / 0.689±0.009	0.545±0.021 / 0.556±0.011	0.461±0.012 0.410±0.012
	GraphCL	0.905±0.005	0.664±0.018 / 0.679±0.014	0.541±0.017 / 0.550±0.015	0.441±0.015 0.396±0.014
	MVGRL	0.913±0.001	0.675±0.008 / 0.685±0.008	0.550±0.014 / 0.560±0.014	0.453±0.013 0.405±0.013
	MERIT	0.924±0.004	0.679±0.011 / 0.689±0.008	0.552±0.014 / 0.562±0.014	0.452±0.013 0.403±0.013
	SUGRL	0.922±0.005	0.675±0.010 / 0.695±0.010	0.550±0.011 / 0.560±0.011	0.449±0.011 0.411±0.011
	BRGCL	0.920±0.003	**0.690±0.012 / 0.710±0.008**	**0.566±0.014 / 0.572±0.011**	**0.461±0.011 / 0.428±0.015**

Table 6: Performance comparison on node classification with attribute noise.

Dataset	Methods	Noise Level					
		0	40	50	60	70	80
Cora	GCN *	0.817±0.005	0.639±0.008	0.510±0.006	0.439±0.012	0.371±0.014	0.317±0.013
	S^{2}GC *	0.831±0.002	0.661±0.007	0.521±0.008	0.454±0.011	0.371±0.010	0.320±0.013
	NRGNN	0.822±0.006	0.654±0.009	0.517±0.009	0.449±0.014	0.385±0.012	0.322±0.013
	SUGRL	0.829±0.007	0.655±0.011	0.522±0.007	0.445±0.012	0.381±0.011	0.330±0.014
	MVGRL	0.831±0.006	0.671±0.009	0.531±0.008	0.450±0.014	0.385±0.010	0.335±0.009
	MERIT	**0.834±0.005**	0.675±0.009	0.528±0.011	0.452±0.012	0.388±0.012	0.338±0.014
	BRGCL	0.822±0.006	**0.690±0.010**	**0.540±0.010**	**0.469±0.013**	**0.399±0.010**	**0.356±0.011**
Citeseer	GCN *	0.703±0.005	0.529±0.009	0.468±0.012	0.372±0.011	0.313±0.011	0.290±0.014
	S^{2}GC *	0.727±0.005	0.553±0.008	0.491±0.011	0.390±0.013	0.310±0.012	0.2880±0.011
	NRGNN	0.710±0.006	0.540±0.007	0.501±0.013	0.384±0.014	0.317±0.009	0.287±0.012
	SUGRL	0.726±0.007	0.540±0.008	0.501±0.008	0.386±0.011	0.315±0.005	0.282±0.011
	MVGRL	**0.740±0.007**	0.542±0.010	0.505±0.007	0.387±0.008	0.311±0.007	0.295±0.009
	MERIT	0.730±0.005	0.544±0.010	0.503±0.008	0.388±0.009	0.314±0.011	0.300±0.009
	BRGCL	0.722±0.004	**0.562±0.007**	**0.514±0.012**	**0.399±0.012**	**0.331±0.012**	**0.312±0.010**
Coauthor CS	GCN *	0.918±0.001	0.702±0.010	0.628±0.012	0.531±0.010	0.455±0.011	0.415±0.013
	S^{2}GC *	**0.928±0.001**	0.713±0.010	0.638±0.010	0.556±0.009	0.476±0.012	0.422±0.012
	NRGNN	0.919±0.002	0.710±0.012	0.632±0.013	0.560±0.008	0.469±0.011	0.423±0.012
	SUGRL	0.913±0.001	0.706±0.008	0.633±0.008	0.561±0.008	0.465±0.009	0.412±0.008
	MVGRL	0.924±0.004	0.709±0.005	0.634±0.007	0.562±0.011	0.466±0.005	0.426±0.005
	MERIT	0.922±0.005	0.714±0.006	0.639±0.009	0.561±0.007	0.471±0.007	0.429±0.008
	BRGCL	0.920±0.003	**0.722±0.009**	**0.653±0.011**	**0.575±0.013**	**0.488±0.010**	**0.442±0.012**

B.2 STATISTICAL SIGNIFICANCE OF IMPROVEMENTS FOR NODE CLASSIFICATION WITH LABEL NOISE.

To validate the statistical significance of the improvements of BRGCL over competing methods, we further calculate the p-values of t-test between BRGCL and the second best baseline for each noise level and dataset for symmetric label noise, asymmetric label noise, and attribute label noise. The p-values for node classification with label noise are shown in Table 7. The p-values for node classification with attribute noise are shown in Table 8. We run all the experiments for 10 times with random initialization and injected symmetric and asymmetric label noise. The p-values for all datasets with all noise levels are less than 0.05, suggesting the statistically significant improvement of BRGCL over baseline methods.

Table 7: p-values of the t-test between BRGCL and the second best baseline on semi-supervised node classification with symmetric label noise and asymmetric label noise.

Dataset	Noise Level		
	40	60	80
	Asymmetric / Symmetric	Asymmetric / Symmetric	Asymmetric / Symmetric
Cora	0.0285 / 0.0021	0.0091 / 0.0038	0.0079 / 0.0217
Citeseer	0.0133 / 0.0024	0.0003 / 0.0013	0.0009 / 0.0401
PubMed	0.0051 / 0.0354	0.0219 / 0.0129	0.0279 / 0.0106
Coauthor CS	0.0341 / 0.0102	0.0121 / 0.0267	0.0317 / 0.0097
ogbn-arxiv	0.0393 / 0.0076	0.0039 / 0.0331	0.0095 / 0.0292

Table 8: p-values of the t-test between BRGCL and the second best baseline on semi-supervised node classification with attribute noise.

Dataset	Noise Level				
	40	50	60	70	80
Cora	0.0082	0.0193	0.0110	0.0267	0.0372
Citeseer	0.0019	0.0410	0.0394	0.0106	0.0289
PubMed	0.0286	0.0301	0.0371	0.0097	0.0165
Coauthor CS	0.0402	0.0122	0.0398	0.0051	0.0176
ogbn-arxiv	0.0219	0.0284	0.0314	0.0177	0.0120

B.3 NODE CLUSTERING WITH INPUT ATTRIBUTE NOISE

To further validate the performance of the node representation learned by BRGCL, we perform node clustering on clean benchmark datasets. We follow the same evaluation protocol as that in Section 5.2. K-means is then applied to the learned node representations to obtain the clustering results. It can be observed from Table 9 that BRGCL still outperforms all baseline methods for node clustering.

Table 9: Node clustering performance comparison on clean benchmark datasets.

Methods	Cora		Citeseer		PubMed		Coauthor CS		ogbn-arxiv	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
Supervised										
GCN	68.3±0.71	52.3±0.54	68.8±0.65	41.9±0.24	69.1±0.99	31.2±0.46	69.8±0.34	68.6±0.59	52.0±1.02	68.0±0.74
S^2GC	69.6±0.42	54.7±0.65	69.1±0.82	42.8±0.55	70.1±0.89	33.2±0.31	70.2±0.45	67.0±0.72	53.6±0.79	68.5±0.70
NRGNN	72.1±0.53	55.6±0.49	69.3±0.77	43.6±0.51	69.9±1.03	34.2±0.59	68.8±0.59	66.2±0.84	51.9±0.84	68.1±0.65
Unsupervised										
K-means	49.2±0.56	32.1±0.53	54.0±0.43	30.5±1.03	59.5±0.67	31.5±0.77	35.2±0.76	20.1±0.92	31.6±0.75	35.9±0.96
GAE	59.6±0.67	42.9±0.62	57.9±0.38	17.6±1.01	65.2±0.37	27.7±0.35	46.7±0.88	41.6±0.81	38.9±0.91	48.4±1.02
ARVGA	64.0±0.41	45.0±0.59	57.3±0.51	26.1±0.54	69.0±0.60	29.0±0.44	60.3±0.61	55.9±0.65	49.2±0.80	68.3±0.65
GALA	74.5±0.57	57.6±0.68	69.3±0.60	44.1±0.39	69.3±0.58	32.7±0.42	66.5±0.79	68.8±0.48	53.5±0.65	66.5±0.72
GraphCL	71.9±0.66	54.6±0.59	68.3±0.42	42.7±0.63	67.6±0.42	31.5±0.32	65.2±0.53	66.4±0.69	52.1±0.67	67.6±0.63
MVGRL	74.2±0.54	57.3±0.44	69.6±0.31	44.7±0.56	69.6±0.44	33.9±0.46	69.3±0.61	69.2±0.57	53.2±0.61	68.2±0.50
MERIT	74.1±0.67	57.6±0.55	69.7±0.55	45.1±0.41	69.8±0.30	33.7±0.40	69.8±0.60	69.5±0.59	54.3±0.54	68.5±0.75
BRGCL	**74.8±0.39**	**58.1±0.51**	**70.3±0.39**	**45.7±0.63**	**70.4±0.39**	**34.9±0.45**	**71.4±0.43**	**70.2±0.56**	**55.5±0.52**	**70.2±0.71**
p-value	0.0009	0.0254	0.0439	0.0395	0.0429	0.0189	0.0099	0.0302	0.0189	0.0151

B.4 COMPARISONS TO EXISTING SAMPLE SELECTION METHODS

In this subsection, we compare BRGCL against previous sample selection methods, including Co-teaching (Han et al., 2018) and Self-Training (Li et al., 2018) for node classification with symmetric label noise. Co-teaching maintains two networks to select clean samples for each other. Self-Training finds nodes with the most confident pseudo labels, and it augmented the labeled training data by incorporating confident nodes with their pseudo labels into the existing training data.The results are shown in Table 10. We can clearly see that BRGCL greatly outperforms competing sample selection methods.

Table 10: Performance comparison against Co-teaching (Han et al., 2018) and Self-training (Li et al., 2018) on node classification with different levels of symmetric label noise.

Dataset	Methods	Noise Level				
		40	50	60	70	80
Cora	Self-training	0.664±0.012	0.584±0.007	0.532±0.013	0.459±0.011	0.368±0.012
	Co-teaching	0.668±0.011	0.593±0.011	0.527±0.010	0.465±0.010	0.367±0.017
	BRGCL	**0.694±0.007**	**0.622±0.009**	**0.567±0.013**	**0.500±0.014**	**0.394±0.012**
Citeseer	Self-training	0.541±0.014	0.465±0.013	0.397±0.013	0.347±0.016	0.301±0.022
	Co-teaching	0.522±0.018	0.461±0.011	0.383±0.011	0.338±0.014	0.299±0.020
	BRGCL	**0.569±0.013**	**0.496±0.011**	**0.433±0.014**	**0.368±0.013**	**0.321±0.014**
PubMed	Self-training	0.597±0.019	0.507±0.011	0.419±0.021	0.380±0.020	0.345±0.023
	Co-teaching	0.584±0.013	0.499±0.015	0.403±0.014	0.371±0.011	0.342±0.022
	BRGCL	**0.632±0.010**	**0.530±0.010**	**0.468±0.010**	**0.399±0.012**	**0.349±0.013**
Coauthor CS	Self-training	0.672±0.010	0.614±0.012	0.542±0.013	0.462±0.015	0.397±0.015
	Co-teaching	0.666±0.012	0.610±0.011	0.529±0.015	0.451±0.013	0.404±0.019
	BRGCL	**0.710±0.008**	**0.638±0.009**	**0.572±0.011**	**0.480±0.011**	**0.428±0.015**
ogbn-arxiv	Self-training	0.462±0.012	0.413±0.014	0.368±0.018	0.328±0.014	0.276±0.020
	Co-teaching	0.437±0.024	0.406±0.011	0.359±0.016	0.322±0.012	0.282±0.025
	BRGCL	**0.482±0.006**	**0.432±0.009**	**0.399±0.009**	**0.344±0.012**	**0.296±0.013**

B.5 JOINT TRAINING VS. DECOUPLED TRAINING.

We study the effectiveness of our decoupled training framework compared with jointly training the encoder and the classifier. We compare the performance on node classification with 20% label noise level. The results are shown in Table 11. It can be observed that decoupling the training of classifier and encoder can mitigate the effects of label noise.

Table 11: Ablation study on contrastive components for node classification with label noise.

Method	Cora		Citeseer		PubMed	
	Confident	K-means	Confident	K-means	Confident	K-means
Joint	77.7±0.08	78.2±0.11	65.7±0.09	65.3±0.10	72.5±0.10	72.5±0.12
Decoupled	79.3±0.09	78.5±0.09	66.8±0.07	66.3±0.08	73.4±0.07	73.0±0.09

B.6 NUMBER OF CONFIDENT PROTOTYPES.

To further study the behavior of BRGCL, we show the number of robust prototypes estimated by BPL in Table 12. It can be observed from the results that the estimated number of robust prototypes is usually very close to the ground truth number of classes for different datasets, justifying the effectiveness of BPL. Because BEC is based on the pseudo labels estimated by BPL, the success of BPL leads to trustworthy estimation of confident nodes and robust prototypes by BEC.

Table 12: Number of robust prototypes inferred by BPL compared with the ground truth number of classes on different datasets

Datasets	Citeseer	Cora	PubMed	Coauthor CS	ogbn-arxiv
ξ in eq. (2)	0.15	0.20	0.35	0.30	0.4
Estimated K	6	8	3	19	48
Classes	6	7	3	15	40