1 Appendix A

1.1 Derivation of $\mathcal{L}^{(pre)}_{ELBO}$: We prove that $\mathcal{L}^{(pre)}_{ELBO}$ (as stated below) is a lower bound of the input graph $\mathcal{G}$ log-likelihood:

$$\mathcal{L}^{(pre)}_{\text{ELBO}}(X,\,A,\,W) = \sum_{i,j=1}^{N} \mathbb{E}_{z_i,\,z_j\sim q(.\,|\,X,\,A)}\Big[\log\big(p(a^{gen}_{ij}\,|\,z_i,\,z_j)\big)\Big] - 2\,N\sum_{i=1}^{N}\text{KL}\big(q(z_i\,|\,X,\,A)\,||\,p(z_i)\big).$$

$$\begin{aligned}
\log\big(p(A^{gen})\big) &= \log\big(\prod_{i,j=1}^{N} p(a^{gen}_{ij})\big),\\
&= \sum_{i,j=1}^{N} \log\big(p(a^{gen}_{ij})\big),\\
&= \sum_{i,j=1}^{N} \log\Big(\int_{z_i}\int_{z_j} \frac{p(a^{gen}_{ij}\,|\,z_i,\,z_j)\,p(z_i,\,z_j)}{q(z_i,\,z_j\,|\,X,\,A)}\,q(z_i,\,z_j\,|\,X,\,A)\,dz_i\,dz_j\Big),\\
&= \sum_{i,j=1}^{N} \log\Big(\mathbb{E}_{z_i,z_j\sim q(.|X,A)}\Big[\frac{p(a^{gen}_{ij}\,|\,z_i,\,z_j)\,p(z_i)\,p(z_j)}{q(z_i\,|\,X,\,A)\,q(z_j\,|\,X,\,A)}\Big]\Big),\\
&\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j\sim q(.|X,A)}\Big[\log\Big(\frac{p(a^{gen}_{ij}|z_i,z_j)\,p(z_i)\,p(z_j)}{q(z_i|X,A)\,q(z_j|X,A)}\Big)\Big], \quad \text{(Jensen's inequality)}\\
&\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j\sim q(.|X,A)}\Big[\log\Big(p(a^{gen}_{ij}|z_i,z_j)\Big) + \log\Big(\frac{p(z_i)}{q(z_i|X,A)}\Big) + \log\Big(\frac{p(z_j)}{q(z_j|X,A)}\Big)\Big],\\
&\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j\sim q(.|X,A)}\Big[\log\Big(p(a^{gen}_{ij}|z_i,z_j)\Big)\Big] + N\sum_{i=1}^{N}\mathbb{E}_{z_i\sim q(.|X,A)}\Big[\log\Big(\frac{p(z_i)}{q(z_i|X,A)}\Big)\Big]\\
&+ N\sum_{j=1}^{N}\mathbb{E}_{z_j\sim q(.|X,A)}\Big[\log\Big(\frac{p(z_j)}{q(z_j|X,A)}\Big)\Big],\\
&\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j\sim q(.|X,A)}\Big[\log\Big(p(a^{gen}_{ij}|z_i,z_j)\Big)\Big] +\ 2\,N\sum_{i=1}^{N}\mathbb{E}_{z_i\sim q(.|X,A)}\Big[\log\Big(\frac{p(z_i)}{q(z_i|X,A)}\Big)\Big],\\
&\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j\sim q(.|X,A)}\Big[\log\Big(p(a^{gen}_{ij}|z_i,z_j)\Big)\Big] -\ 2\,N\sum_{i=1}^{N} KL\Big(q(z_i|X,A)||p(z_i)\Big),\\
&\geqslant \mathcal{L}^{(pre)}_{\text{ELBO}}(X,\,A,\,W).
\end{aligned}$$

1.2 Derivation of $\mathcal{L}^{(clus)}_{ELBO}$: We prove that $\mathcal{L}^{(clus)}_{ELBO}$ (as stated below) is a lower bound of the input graph $\mathcal{G}$ log-likelihood:

$$\mathcal{L}^{(clus)}_{\text{ELBO}}(X,\,A,\,W) = \mathcal{L}^{(pre)}_{\text{ELBO}}(X,\,A,\,W) - 2N\sum_{i=1}^{N}\mathbb{E}_{z_i\sim q(.|X,A)}\Big[\text{KL}\Big(q(c_i|z_i)||p(c_i|z_i)\Big)\Big].$$

$$
\begin{aligned}
\log\big(p(A^{gen})\big) &= \log\big(\prod_{i,j=1}^{N} p(a_{ij}^{gen})\big),\\
&= \sum_{i,j=1}^{N} \log\big(p(a_{ij}^{gen})\big),\\
&= \sum_{i,j=1}^{N} \log\Big(\sum_{c_i}\sum_{c_j}\int_{z_i}\int_{z_j} \frac{p(a_{ij}^{gen}|z_i,z_j)\,p(c_i|z_i)\,p(c_j|z_j)\,p(z_i)p(z_j)}{q(z_i|X,A)\,q(c_i|z_i)\,q(z_j|X,A)\,q(c_j|z_j)}\,q(z_i|X,A)\,q(c_i|z_i)\,q(z_j|X,A)\,q(c_j|z_j)\,dz_i\,dz_j\Big),\\
&= \sum_{i,j=1}^{N} \log\Big(\mathbb{E}_{\substack{c_i\sim q(.|z_i),\\ c_j\sim q(.|z_j),\\ z_i,z_j\sim q(.|X,A)}}\Big[\frac{p(a_{ij}^{gen}|z_i,z_j)\,p(c_i|z_i)\,p(c_j|z_j)\,p(z_i)\,p(z_j)}{q(z_i|X,A)\,q(c_i|z_i)\,q(z_j|X,A)\,q(c_j|z_j)}\Big]\Big),\\
&\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{\substack{c_i\sim q(.|z_i),\\ c_j\sim q(.|z_j),\\ z_i,z_j\sim q(.|X,A)}}\Big[\log\Big(\frac{p(a_{ij}^{gen}|z_i,z_j)\,p(c_i|z_i)\,p(z_i)\,p(c_j|z_j)\,p(z_j)}{q(c_i|z_i)\,q(z_i|X,A)\,q(c_j|z_j)\,q(z_j|X,A)}\Big)\Big], \quad \text{(Jensen's inequality)}\\
&\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{\substack{c_i\sim q(.|z_i),\\ c_j\sim q(.|z_j),\\ z_i,z_j\sim q(.|X,A)}}\Big[\log\Big(p(a_{ij}^{gen}|z_i,z_j)\Big) + \log\Big(\frac{p(z_i)}{q(z_i|X,A)}\Big) + \log\Big(\frac{p(z_j)}{q(z_j|X,A)}\Big) + \log\Big(\frac{p(c_i|z_i)}{q(c_i|z_i)}\Big)\\
&+ \log\Big(\frac{p(c_j|z_j)}{q(c_j|z_j)}\Big)\Big],
\end{aligned}
$$

$$
\begin{aligned}
\log\big(p(A^{gen})\big) &\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j\sim q(.|X,A)}\Big[\log\Big(p(a_{ij}^{gen}|z_i,z_j)\Big)\Big] + N\sum_{i=1}^{N}\mathbb{E}_{z_i\sim q(.|X,A)}\Big[\log\Big(\frac{p(z_i)}{q(z_i|X,A)}\Big)\Big]\\
&+ N\sum_{j=1}^{N}\mathbb{E}_{z_j\sim q(.|X,A)}\Big[\log\Big(\frac{p(z_j)}{q(z_j|X,A)}\Big)\Big] + N\sum_{i=1}^{N}\mathbb{E}_{\substack{c_i\sim q(.|z_i),\\ z_i\sim q(.|X,A)}}\Big[\log\Big(\frac{p(c_i|z_i)}{q(c_i|z_i)}\Big)\Big]\\
&+ N\sum_{i=1}^{N}\mathbb{E}_{\substack{c_j\sim q(.|z_j),\\ z_j\sim q(.|X,A)}}\Big[\log\Big(\frac{p(c_j|z_j)}{q(c_j|z_j)}\Big)\Big],\\
&\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j\sim q(.|X,A)}\Big[\log\Big(p(a_{ij}^{gen}|z_i,z_j)\Big)\Big] + 2N\sum_{i=1}^{N}\mathbb{E}_{z_i\sim q(.|X,A)}\Big[\log\Big(\frac{p(z_i)}{q(z_i|X,A)}\Big)\Big]\\
&+ 2N\sum_{i=1}^{N}\mathbb{E}_{\substack{c_i\sim q(.|z_i),\\ z_i\sim q(.|X,A)}}\Big[\log\Big(\frac{p(c_i|z_i)}{q(c_i|z_i)}\Big)\Big],\\
&\geqslant \sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j\sim q(.|X,A)}\Big[\log\Big(p(a_{ij}^{gen}|z_i,z_j)\Big)\Big] - 2N\sum_{i=1}^{N} KL\Big(q(z_i|X,A)||p(z_i)\Big)\\
&- 2N\sum_{i=1}^{N}\mathbb{E}_{z_i\sim q(.|X,A)}\Big[KL\Big(q(c_i|z_i)||p(c_i|z_i)\Big)\Big],\\
&\geqslant \mathcal{L}_{ELBO}^{(pre)}(X,\,A,\,W) - 2N\sum_{i=1}^{N}\mathbb{E}_{z_i\sim q(.|X,A)}\Big[KL\Big(q(c_i|z_i)||p(c_i|z_i)\Big)\Big] = \mathcal{L}_{ELBO}^{(clus)}(X,\,A,\,W).
\end{aligned}
$$

1.3 Derivation of $\mathcal{L}_{BELBO}^{(clus)}$

THEOREM 1.1. *Given the design choices, we derive a lower bound $\mathcal{L}_{BELBO}^{(clus)}$ that verifies:*

$$
\mathcal{L}_{ELBO}^{(clus)}(X,\,A^{clus},\,U) \leqslant \mathcal{L}_{BELBO}^{(clus)}(X,\,A^{clus},\,A,\,U) \leqslant log\big(p(A^{gen})\big),
$$

$$\mathcal{L}^{(clus)}_{BELBO}(X,\ A^{clus},\ A,\ U,\ W) = \mathcal{L}^{(clus)}_{ELBO}(X,\ A^{clus},\ U) + 2\ N\ \sum_{i=1}^{N} KL\Big(q_u(z_i|X, A^{clus}) \parallel q_w(z_i|X, A)\Big).$$

Proof. We start by proving two lemmas: Lemma 1.1 and Lemma 1.2.

LEMMA 1.1. *Given the design choices for the generative and inference models of the second auto-encoder, the likelihood of the generated graph structure A^{clus} can be expressed as:*

$$log(p(A^{gen})) = \sum_{i,j=1}^{N} KL\Big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \parallel p(z_i, z_j, c_i, c_j|a^{gen}_{ij})\Big) + \mathcal{L}^{(clus)}_{ELBO}(X,\ A^{clus},\ U).$$

Proof.

$$\begin{aligned}
KL\big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \| p(z_i, z_j, c_i, c_j|a^{gen}_{ij})\big) &= -\sum_{c_i,c_j} \int_{z_i}\int_{z_j} q(z_i, z_j, c_i, c_j|X, A^{clus}) \log\big(\frac{p(z_i, z_j, c_i, c_j|a^{gen}_{ij})}{q(z_i, z_j, c_i, c_j|X, A^{clus})}\big)\ dz_i\ dz_j, \\
&+ \sum_{c_i,c_j} \int_{z_i}\int_{z_j} q(z_i, z_j, c_i, c_j|X, A^{clus}) \log\big(p(a^{gen}_{ij})\big)\ dz_i\ dz_j, \\
&= -\sum_{c_i,c_j} \mathbb{E}_{z_i,z_j\sim q_u(.|X,A^{clus})}\Big[q(c_i|z_i)\ q(c_j|z_j) \log\big(\frac{p(z_i, z_j, c_i, c_j, a^{gen}_{ij})}{q(z_i, z_j, c_i, c_j|X, A^{clus})}\big)\Big] \\
&+ \log\big(p(a^{gen}_{ij})\big).
\end{aligned}$$

Hence, we can write:

$$\begin{aligned}
\log\big(p(A^{gen})\big) &= \log\big(\prod_{i,j=1}^{N} p(a^{gen}_{ij})\big), \\
&= \sum_{i,j=1}^{N} \log\big(p(a^{gen}_{ij})\big), \\
&= \sum_{i,j=1}^{N} KL\Big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \parallel p(z_i, z_j, c_i, c_j|a^{gen}_{ij})\Big) \\
&+ \sum_{i,j=1}^{N}\sum_{c_i,c_j} \mathbb{E}_{z_i,z_j\sim q_u(.|X,A^{clus})}\Big[q(c_i|z_i)\ q(c_j|z_j) \log\big(\frac{p(z_i, z_j, c_i, c_j, a^{gen}_{ij})}{q(z_i, z_j, c_i, c_j|X, A^{clus})}\big)\Big], \\
&= \sum_{i,j=1}^{N} KL\Big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \parallel p(z_i, z_j, c_i, c_j|a^{gen}_{ij})\Big) \\
&+ \mathbb{E}_{z_i,z_j\sim q_u(.|X,A^{clus})}\Big[\log\big(p(a^{gen}_{ij}|z_i, z_j)\big)\Big] \\
&- 2\ KL\Big(q_u(z_i|X, A^{clus}) \parallel p(z_i)\Big) - 2\ \mathbb{E}_{z_i\sim q_u(.|X,A^{clus})}\Big[KL\Big(q(c_i|z_i) \parallel p(c_i|z_i)\Big)\Big],
\end{aligned}$$

$$\implies \log\big(p(A^{gen})\big) = \sum_{i,j=1}^{N} KL\Big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \parallel p(z_i, z_j, c_i, c_j|a^{gen}_{ij})\Big) + \mathcal{L}^{(clus)}_{ELBO}(X,\ A^{clus},\ U).$$

□

LEMMA 1.2.

$$\textit{If } q_w(z_i, z_j|X, A) = p(z_i, z_j|a_{ij}^{gen}),$$

$$\textit{then } KL\Big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \,\|\, p(z_i, z_j, c_i, c_j|a_{ij}^{gen})\Big) \geqslant 2\, KL\Big(q_u(z_i|X, A^{clus}) \,\|\, q_w(z_i|X, A)\Big).$$

Proof.

$$\begin{aligned}
KL\Big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \,\|\, p(z_i, z_j, c_i, c_j|a_{ij}^{gen})\Big) &= \sum_{c_i,c_j} \int_{z_i}\int_{z_j} q(z_i, z_j, c_i, c_j|X, A^{clus}) \log\big(\frac{q(z_i, z_j, c_i, c_j|X, A^{clus})}{p(z_i, z_j, c_i, c_j|a_{ij}^{gen})}\big)\, dz_i\, dz_j, \\
&= \sum_{c_i,c_j} \int_{z_i}\int_{z_j} q(z_i, z_j, c_i, c_j|X, A^{clus}) \log\big(\frac{q(z_i, z_j, c_i, c_j|X, A^{clus})\, q_w(z_i, z_j|X, A)}{p(z_i, z_j, c_i, c_j|a_{ij}^{gen})\, q_w(z_i, z_j|X, A)}\big)\, dz_i\, dz_j, \\
&= \int_{z_i}\int_{z_j} q_u(z_i|X, A^{clus})\, q_u(z_j|X, A^{clus}) \log\big(\frac{q_u(z_i|X, A^{clus})\, q_u(z_j|X, A^{clus})}{q_w(z_i|X, A)\, q_w(z_j|X, A)}\big)\, dz_i\, dz_j \\
&+ \sum_{c_i,c_j} \int_{z_i}\int_{z_j} q(z_i, z_j, c_i, c_j|X, A^{clus}) \log\big(\frac{q_w(z_i, z_j|X, A)\, q(c_i|z_i)\, q(c_j|z_j)}{p(z_i, z_j|a_{ij}^{gen})\, p(c_i|z_i)\, p(c_j|z_j)}\big)\, dz_i\, dz_j, \\
&= \int_{z_i} q_u(z_i|X, A^{clus}) \log\big(\frac{q_u(z_i|X, A^{clus})}{q_w(z_i|X, A)}\big)\, dz_i + \int_{z_j} q_u(z_j|X, A^{clus}) \log\big(\frac{q_u(z_j|X, A^{clus})}{q_w(z_j|X, A)}\big)\, dz_j \\
&+ \underbrace{\sum_{c_i,c_j} \int_{z_i}\int_{z_j} q(z_i, z_j, c_i, c_j|X, A^{clus}) \log\big(\frac{q_w(z_i, z_j|X, A)}{p(z_i, z_j|a_{ij}^{gen})}\big)\, dz_i\, dz_j}_{=0} \\
&+ \sum_{c_i} \int_{z_i} q(z_i, c_i, |X, A^{clus}) \log\big(\frac{q(c_i|z_i)}{p(c_i|z_i)}\big)\, dz_i + \sum_{c_j} \int_{z_j} q(z_j, c_j, |X, A^{clus}) \log\big(\frac{q(c_j|z_j)}{p(c_j|z_j)}\big)\, dz_j, \\
&= 2\, KL\Big(q_u(z_i|X, A^{clus}) \,\|\, q_w(z_i|X, A)\Big) + \underbrace{2\, N \sum_{i=1}^{N} \mathbb{E}_{z_i \sim q_u(.|X, A^{clus})}\Big[KL\Big(q(c_i|z_i) \,\|\, p(c_i|z_i)\Big)\Big]}_{\geqslant 0}, \\
&\geqslant 2\, KL\Big(q_u(z_i|X,\ A^{clus}) \,\|\, q_w(z_i|X,\ A)\Big).
\end{aligned}$$

□

Since maximizing $\mathcal{L}_{ELBO}^{(pre)}(X,\ A,\ W)$ during the pretraining phase makes the variational distribution $q_w(z_i, z_j|X, A)$ approximate the distribution $p(z_i, z_j|a_{ij}^{gen})$, then according to Lemma 1.2, we have

$$\begin{aligned}
&KL\Big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \,\|\, p(z_i, z_j, c_i, c_j|a_{ij}^{gen})\Big) \geqslant 2\, KL\Big(q_u(z_i|X, A^{clus}) \,\|\, q_w(z_i|X, A)\Big), \\
&\implies \mathcal{L}_{ELBO}^{(clus)}(X,\ A^{clus},\ U) + \sum_{i,j=1}^{N} KL\Big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \,\|\, p(z_i, z_j, c_i, c_j|a_{ij}^{gen})\Big) \geqslant \mathcal{L}_{ELBO}^{(clus)}(X,\ A^{clus},\ U) + 2\sum_{i=1}^{N} KL\Big(q_u(z_i|X, A^{clus}) \,\|\, q_w(z_i|X, A)\Big).
\end{aligned}$$

Based on Lemma 1.1, we have $\log(A^{gen}) = \sum_{i,j=1}^{N} KL\Big(q(z_i, z_j, c_i, c_j|X, A^{clus}) \,\|\, p(z_i, z_j, c_i, c_j|a_{ij}^{gen})\Big) + \mathcal{L}_{ELBO}^{(clus)}(X,\ A^{clus},\ U)$.

$$\implies \log(A^{gen}) \geqslant \mathcal{L}_{ELBO}^{(clus)}(X,\ A^{clus},\ U) + 2\,\sum_{i=1}^{N} KL\Big(q_u(z_i|X, A^{clus}) \,\|\, q_w(z_i|X, A)\Big).$$

Or

$$2\,\sum_{i=1}^{N} KL\Big(q_u(z_i|X, A^{clus}) \,\|\, q_w(z_i|X, A)\Big) \geqslant 0.$$

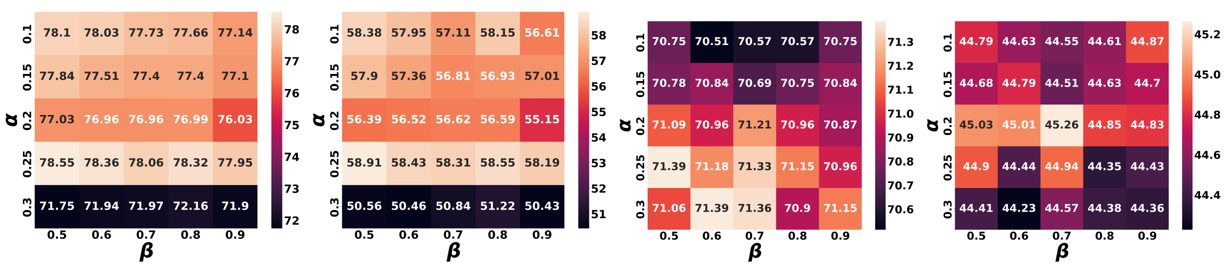
Then, we conclude

$$\log(A^{gen}) \geqslant \mathcal{L}_{\text{BELBO}}^{(clus)}(X,\ A^{clus},\ A,\ U,\ W) \geqslant \mathcal{L}_{ELBO}^{(clus)}(X,\ A^{clus},\ U).$$

□

Table 1: Comparing the node clustering results for different graph self-supervision methods. Best method in bold.

Method	Cora			Citeseer			Pubmed		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
GMI [5]	63.4	50.3	38.8	63.8	38.1	37.5	67.1	26.2	26.8
AGC [8]	68.9	53.7	48.6	67.0	41.1	41.9	69.8	31.6	31.9
GALA [4]	74.6	57.7	53.2	69.3	44.1	44.6	69.4	32.7	32.1
DGI [7]	71.3	56.4	51.1	68.8	44.4	45.0	58.9	27.7	31.5
MVGRL [2]	73.2	56.2	51.9	68.7	43.7	44.3	67.0	31.6	29.4
GRACE [9]	65.8	51.7	44.0	67.5	41.9	42.1	70.2	**36.7**	33.6
BGRL [6]	73.8	54.7	51.1	65.6	38.4	38.7	58.7	24.9	23.1
AFGRL [3]	74.6	58.4	57.6	67.4	42.2	42.7	63.9	27.6	25.4
AGE [1]	76.1	**59.7**	54.5	70.1	44.3	45.4	70.9	30.8	32.9
BELBO-VGAE	**78.5**	58.4	**58.6**	**71.4**	**44.2**	**46.5**	**73.9**	34.3	**37.4**



ACC on Cora | NMI on Cora | ACC on Citeseer | NMI on Citeseer

Figure 1: Sensitivity of BELBO-VGAE to the hyperparameters α and β in terms of ACC and NMI.

2 Appendix B

2.1 Comparison with non-variational methods: In Table 1, we compare the clustering performance of BELBO-VGAE with state-of-the-art non-variational methods, including the most recent self-supervised approaches such as GRACE [9], BGRL [6], AFGRL [3]. As we can see, our model outperforms all the considered methods in terms of ACC and ARI. For example, the difference in clustering performance between BELBO-VGAE and the most competitive approach on Cora (i.e., AGE) amounts to 2.4% in terms of ACC and 4.1% in terms of ARI. These results confirm the suitability of our model.

Sensitivity analysis: We explore the sensitivity of BELBO-VGAE to the data-dependent hyperparameters (α and β). The remaining hyperparameters and design choices are fixed for all datasets. As we can see in Figure 1, our model shows consistent results in terms of ACC and NMI in a wide range of values of α and β. Especially, fixing α, BELBO-VGAE yields stable clustering results as β varies, while the ACC spikes when α is equal to 0.25 and 0.2 on Cora and Citeseer, respectively.

References

[1] G. Cui, J. Zhou, C. Yang, and Z. Liu, *Adaptive graph encoder for attributed graph embedding*, in SIGKDD, 2020, pp. 976–985.
[2] K. Hassani and A. H. Khasahmadi, *Contrastive multi-view representation learning on graphs*, in ICML, 2020, pp. 3451–3461.
[3] N. Lee, J. Lee, and C. Park, *Augmentation-free self-supervised learning on graphs*, in AAAI, 2022.
[4] J. Park, M. Lee, H. J. Chang, K. Lee, and J. Y. Choi, *Symmetric graph convolutional autoencoder for unsupervised graph representation learning*, in ICCV, 2019, pp. 6519–6528.
[5] Z. Peng, W. Huang, M. Luo, Q. Zheng, Y. Rong, T. Xu, and J. Huang, *Graph representation learning via graphical mutual information maximization*, in WWW, 2020, p. 259–270.
[6] S. Thakoor, C. Tallec, M. G. Azar, M. Azabou, E. L. Dyer, R. Munos, P. Veličković, and M. Valko, *Large-scale representation learning on graphs via bootstrapping*, in ICLR, 2022.
[7] P. Velickovic, W. Fedus, W. L. Hamilton, P. Liò, Y. Bengio, and R. D. Hjelm, *Deep graph infomax*, ICLR, (2019).
[8] X. Zhang, H. Liu, Q. Li, and X.-M. Wu, *Attributed graph clustering via adaptive graph convolution*, in IJCAI, 2019, pp. 4327–4333.
[9] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, *Deep graph contrastive representation learning*, in ICML Workshop, 2020.