A Method Supplement

A.1 Notations

The used notations in paper are summarized in Table 1.

Table 1: Notations and descriptions

Notation	Description	Notation	Description
$\boldsymbol{A}_v$	Adjacency matrix in the v-th view	$\boldsymbol{X}_v$	Attribute matrix in the v-th view
L_{com}	Common Loss	n	Number of nodes
V	Number of views	$\overline{\boldsymbol{X}}^v$	Imputed Attributes
$\boldsymbol{W_2}$	Weighted Maping Matrix	$\hat{\mathbf{X}}_v$	Reconstructed Attributes
$\boldsymbol{M}^v$	Missing position marker matrix in the v-th view	$\boldsymbol{R}^v$	Missing Position Reminder Matrix in the v-th view
$\boldsymbol{K}^v$	Random Variable Matrix	$\boldsymbol{Z}_{com}$	Common Represetation
h	Reminder Rate	$\boldsymbol{W_1}$	Dimension Interaction Matrix
k	Number of clusters	L_{CE}	Common Extractor Loss
L_{re}^v	Reconstruct loss in the v-th view	L_D^v	Loss for discriminator in the v-th view
L_G^v	Loss for generator in the v-th view	Z_v	Latent Representations in the v-th view
Q_v	Similarity Score in the v-th view	P_v	Common Category Assignment Distribution in the v-th view
λ	Attention Rate	m	Missing Rate

A.2 Proof

Theorem 1. *In intra-view adversarial games, incomplete attributes can hinder the generator from generating the desired distribution (the unique true attributed distribution) when achieving the equilibrium state, as multiple data distributions may satisfy the equilibrium conditions.*

Proof. Based on [Yoon *et al.*2018] and [Miao *et al.*2021], we give the proof as follows. In traditional intra-view adversarial games without the missing position reminder matrix, the objective function of dynamic game mechanism for the v-th view is

$$\begin{aligned}
&V(G^v, D^v) = E_{\overline{\boldsymbol{X}}^v,\mathbf{M}^v,\mathbf{R}^v}[\mathbf{M}^v \odot \log D^v(\overline{\boldsymbol{X}}^v) + (1-\mathbf{M}^v)\odot \\
&\log(1-D^v(\overline{\boldsymbol{X}}^v))] \\
=&\int_{\mathbb{X}^v} \sum_{\mathbf{M}^v\in\{0,1\}^{d\times n}} (\mathbf{M}^v \odot \log D^v(\mathbf{X}^v)) + (1-\mathbf{M}^v)\odot\log(1-D^v(\mathbf{X}^v)) \\
&p(\mathbf{X}^v,\mathbf{M}^v)d\mathbf{X}^v \\
=&\int_{\mathbb{X}^v} \sum_{i=1}^{n}\sum_{j=1}^{d}\Big(\sum_{\mathbf{M}^v\in\mathbf{M}_{1,v}^{ij}} \log D^v(\mathbf{X}^v)_i^j + \sum_{\mathbf{M}^v\in\mathbf{M}_{0,v}^{ij}} \log(1-D^v(\mathbf{X}^v)_i^j)\Big) \\
&p(\mathbf{X}^v,\mathbf{M}^v)d\mathbf{X}^v \\
=&\int_{\mathbb{X}^v} \sum_{i=1}^{n}\sum_{j=1}^{d}(\log D^v(\mathbf{X}^v)_i^j p(\mathbf{X}^v, m_{v,i}^j=1) + \log(1-D^v(\mathbf{X}^v)_i^j)) \\
&p(\mathbf{X}^v, m_{v,i}^j=0)d\mathbf{X}^v
\end{aligned} \tag{1}$$

where we divide $\mathbf{M}^v$ into two submatrices $\mathbf{M}_{1,v}^{ij}$ and $\mathbf{M}_{0,v}^{ij}$ in which all elements are zeros meaning missing. When fixing G^v, Eq. (1) obtains the maximum when the discriminator D^v is

$$\begin{aligned}
D^v(\mathbf{X}^v) &= \frac{p(\mathbf{X}^v, m_{v,i}^j=1)}{p(\mathbf{X}^v, m_{v,i}^j=1)+p(\mathbf{X}^v, m_{v,i}^j=0)} \\
&= p_m(m_{v,j}^i=1|\mathbf{X}^v)
\end{aligned} \tag{2}$$

where p_m is the marginal distribution of M^v.
We bring the optimal discriminator into Eq. (1) and we get

$$\begin{aligned}
&V(G^v, D^v) = E_{\overline{\boldsymbol{X}}^v,\mathbf{M}^v}[\mathbf{M}^v \odot \log D^v(\overline{\boldsymbol{X}}^v) + (1-\mathbf{M}^v)\odot \\
&\log(1-D^v(\overline{\boldsymbol{X}}^v))] \\
=&E_{\overline{\boldsymbol{X}}^v,\mathbf{M}^v}\Big[\sum_{(j,i):m_{v,i}^j=1} \log p_m(m_{v,j}^i=1|\overline{\boldsymbol{X}}^v)+ \\
&\sum_{(j,i):m_{v,i}^j=0} \log p_m(m_{v,j}^i=0|\overline{\boldsymbol{X}}^v)\Big].
\end{aligned} \tag{3}$$

Assuming that p_r is the marginal distribution of R, the above equation can be transformed to

$$\begin{aligned}
&E_{\overline{\boldsymbol{X}}^v,\mathbf{M}^v}\Big[\sum_{(j,i):m_{v,i}^j=1} \log p_m(m_{v,j}^i=1|\overline{\boldsymbol{X}}^v)+ \\
&\sum_{(j,i):m_{v,i}^j=0} \log p_m(m_{v,j}^i=0|\overline{\boldsymbol{X}}^v)\Big] \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}}\int_{\mathbb{X}^v} p(\mathbf{X}^v, m_i^j=e)\log p_m(m_{v,j}^i=e|\mathbf{X}^v)d\mathbf{X}^v \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}}\int_{\mathbb{X}^v} p(\mathbf{X}^v, m_i^j=e)\log\frac{p_m(\mathbf{X}^v, m_{v,j}^i=e)}{\hat{p}(\mathbf{X}^v)}d\mathbf{X}^v \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}}\int_{\mathbb{X}^v} p(\mathbf{X}^v, m_i^j=e)\log\frac{p(X, m_{v,j}^i=e)p(m_{v,j}^i=e)}{p(m_{v,j}^i=e)\hat{p}(\mathbf{X}^v)}d\mathbf{X}^v \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}}\int_{\mathbb{X}^v} p(\mathbf{X}^v, m_i^j=e)(\log p(m_{v,j}^i=e) \\
&+p(\mathbf{X}^v, m_i^j=e)\log\frac{p(\mathbf{X}^v|m_{v,j}^i=e)}{\hat{p}(\mathbf{X}^v)})d\mathbf{X}^v \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}}\int_{\mathbb{X}^v} p(\mathbf{X}^v, m_i^j=e)\log p_m(m_{v,j}^i=e) \\
&+p_m(m_{v,j}^i=e)\hat{p}(\mathbf{X}^v|m_{v,j}^i=e)\log\frac{p(\mathbf{X}^v|m_{v,j}^i=e)}{\hat{p}(\mathbf{X}^v)}d\mathbf{X}^v \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}}\int_{\mathbb{X}^v} (p(m_{v,j}^i=e)\log p_m(m_{v,j}^i=e)) \\
&+p_m(m_{v,j}^i=e)D_{KL}(\hat{p}(\mathbf{X}^v|m_{v,j}^i=e)||\hat{p}(\mathbf{X}^v))d\mathbf{X}^v
\end{aligned} \tag{4}$$

where D_{KL} is the Kullback-Leibler divergence, and Eq. (4) obtains the minimum when

$$\hat{p}(\mathbf{X}^v|m_{v,j}^i=e)=\hat{p}(\mathbf{X}^v). \tag{5}$$

Eq. (5) holds if and only if the following conditions are satisfied,

$$\hat{p}(\mathbf{X}^v|m_{v,j}^i=1)=\hat{p}(\mathbf{X}^v|m_{v,j}^i=0), \tag{6}$$

for $i\in\{1,...,n\}$ and $j\in\{1,...,d\}$. Since the number of parameters specified by the generator exceeds the total number of linear equations, there exist multiple data distributions that satisfy Eq. (5).

□

Theorem 2. *Given the missing position reminder matrix R^v, the equilibrium of each view after the dynamic game can be uniquely determined, where the distribution of generated attributes $\hat{p}(\boldsymbol{X}^v|\boldsymbol{M}^v)$ is the same as the real distribution of attributes $\hat{p}(\boldsymbol{X}^v|\mathbf{1})$ in each view, that is $\hat{p}(\boldsymbol{X}^v|\boldsymbol{M}^v) = \hat{p}(\boldsymbol{X}^v|\mathbf{1})$.*

Proof. We prove theorem 1 based on [Yoon *et al.*2018] and [Miao *et al.*2021]. The proof can be divided into three steps: first, we determine the optimal discriminator D^v for each view with fixed generator G^v; then we prove that the missing position reminder matrix ensures that the view-specific generator can generate a unique data distribution that is the same as the real data distribution for each view. The detailed proof is as follows.

In Stage 1, generators and discriminators in each view are trained independently to extract view-specific information without view interaction. Therefore, we can analyze each view independently in this proof. We take the v-th view as an example to give the correlation analysis.

Step 1: Determine the optimal discriminator with the fixed generator.

The objective function of dynamic game mechanism is

$$\begin{aligned}
&V(G^v, D^v) = E_{\overline{\boldsymbol{X}}^v, \mathbf{M}^v, \mathbf{R}^v}[\mathbf{M}^v \odot \log D^v(\overline{\boldsymbol{X}}^v, \mathbf{R}^v) + (1 - \mathbf{M}^v)\odot \\
&\log(1 - D^v(\overline{\boldsymbol{X}}^v, \mathbf{R}^v))] \\
=&\int_{\mathbb{X}^v} \sum_{\mathbf{M}^v \in \{0,1\}^{d\times n}} \int_{\mathbb{R}^v} (\mathbf{M}^v \odot \log D^v(\mathbf{X}^v, \mathbf{R}^v)) + (1 - \mathbf{M}^v)\odot \\
&\log(1 - D^v(\mathbf{X}^v, \mathbf{R}^v))p(\mathbf{X}^v, \mathbf{M}^v, \mathbf{R}^v) d\mathbf{R}^v d\mathbf{X}^v \\
=&\int_{\mathbb{X}^v}\int_{\mathbb{R}^v} \sum_{i=1}^{n}\sum_{j=1}^{d}(\sum_{\mathbf{M}^v \in \mathbf{M}^{ij}_{1,v}} \log D^v(\mathbf{X}^v, \mathbf{R}^v)^j_i \\
&+ \sum_{\mathbf{M}^v \in \mathbf{M}^{ij}_{0,v}} \log(1 - D^v(\mathbf{X}^v, \mathbf{R}^v)^j_i))p(\mathbf{X}^v, \mathbf{M}^v, \mathbf{R}^v) d\mathbf{R}^v d\mathbf{X}^v \\
=&\int_{\mathbb{X}^v}\int_{\mathbb{R}^v} \sum_{i=1}^{n}\sum_{j=1}^{d}(\log D^v(\mathbf{X}^v, \mathbf{R}^v)^j_i p(\mathbf{X}^v, \mathbf{R}^v, m^j_{v,i} = 1) \\
&+ \log(1 - D^v(\mathbf{X}^v, \mathbf{R}^v)^j_i))p(\mathbf{X}^v, \mathbf{R}^v, m^j_{v,i} = 0) d\mathbf{R}^v d\mathbf{X}^v
\end{aligned} \tag{7}$$

where we divide $\mathbf{M}^v$ into two submatrices $\mathbf{M}^{ij}_{1,v}$ and $\mathbf{M}^{ij}_{0,v}$ in which all elements are zeros meaning missing. When fixing G^v, Eq. (7) obtains the maximum when the discriminator D^v is

$$\begin{aligned}
D^v(\mathbf{X}, \mathbf{R}) &= \frac{p(\mathbf{X}^v, \mathbf{R}^v, m^j_{v,i} = 1)}{p(\mathbf{X}^v, \mathbf{R}^v, m^j_{v,i} = 1) + p(\mathbf{X}^v, \mathbf{R}^v, m^j_{v,i} = 0)} \\
&= p_m(m^i_{v,j} = 1|\mathbf{X}^v, \mathbf{R}^v)
\end{aligned} \tag{8}$$

where p_m is the marginal distribution of M^v.

Step 2: Prove that the missing position reminder matrix ensures that the view-specific generator can generate a unique data distribution that is the same as the real data distribution for each view.

Based on the optimal discriminator obtained in Step 1, we can change Eq. (7) as follows:

$$\begin{aligned}
&V(G^v, D^v) = E_{\overline{\boldsymbol{X}}^v, \mathbf{M}^v, \mathbf{R}^v}[\mathbf{M}^v \odot \log D^v(\overline{\boldsymbol{X}}^v, \mathbf{R}^v) + (1 - \mathbf{M}^v)\odot \\
&\log(1 - D^v(\overline{\boldsymbol{X}}^v, \mathbf{R}^v))] \\
=&E_{\overline{\boldsymbol{X}}^v, \mathbf{M}^v, \mathbf{R}^v}\Big[\sum_{(j,i):m^j_{v,i}=1} \log p_m(m^i_{v,j} = 1|\overline{\boldsymbol{X}}^v, \mathbf{R}^v) + \\
&\sum_{(j,i):m^j_{v,i}=0} \log p_m(m^i_{v,j} = 0|\overline{\boldsymbol{X}}^v, \mathbf{R}^v)\Big].
\end{aligned} \tag{9}$$

Assuming that p_r is the marginal distribution of R, Eq. (9) can transform to

$$\begin{aligned}
&E_{\overline{\boldsymbol{X}}^v, \mathbf{M}^v, \mathbf{R}^v}\Big[\sum_{(j,i):m^j_{v,i}=1} \log p_m(m^i_{v,j} = 1|\overline{\boldsymbol{X}}^v, \mathbf{R}^v) + \\
&\sum_{(j,i):m^j_{v,i}=0} \log p_m(m^i_{v,j} = 0|\overline{\boldsymbol{X}}^v, \mathbf{R}^v)\Big] \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}} \int_{\mathbb{R}^{ji}_{e,v}} \int_{\mathbb{X}^v} p(\mathbf{X}^v, \mathbf{R}^v, m^j_i = e) \\
&\log p_m(m^i_{v,j} = e|\mathbf{X}^v, \mathbf{R}^v) d\mathbf{R}^v d\mathbf{X}^v \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}} \int_{\mathbb{R}^{ji}_{e,v}} \int_{\mathbb{X}^v} p(\mathbf{X}^v, \mathbf{R}^v, m^j_i = e) \\
&\log \frac{p_m(m^i_{v,j} = e|\mathbf{R}^v) p(\mathbf{X}^v, m^i_{v,j} = e|\mathbf{R}^v)}{p_m(m^i_{v,j} = e|\mathbf{R}^v)\hat{p}(\mathbf{X}^v|\mathbf{R}^v)} d\mathbf{R}^v d\mathbf{X}^v \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}} \int_{\mathbb{R}^{ji}_{e,v}} \int_{\mathbb{X}^v} p(\mathbf{X}^v, \mathbf{R}^v, m^j_i = e)(\log p_m(m^i_{v,j} = e \\
&|\mathbf{R}^v) + \log \frac{p(\mathbf{X}^v|m^i_{v,j} = e, \mathbf{R}^v)}{\hat{p}(\mathbf{X}^v|\mathbf{R}^v)}) d\mathbf{R}^v d\mathbf{X}^v \\
=&\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{e\in\{0,1\}} \int_{\mathbb{R}^{ji}_{e,v}} (p(m^i_{v,j} = e, \mathbf{R}^v)\log p_m(m^i_{v,j} = e|\mathbf{R}^v)) \\
&+ p_m(m^i_{v,j} = e, \mathbf{R}^v) D_{KL}(\hat{p}(.|\mathbf{R}^v, m^i_{v,j} = e)||\hat{p}(.|\mathbf{R}^v)) d\mathbf{R}^v
\end{aligned} \tag{10}$$

where D_{KL} is the Kullback-Leibler divergence, and Eq. (10) obtains the minimum when

$$\hat{p}(.|\mathbf{R}^v, m^i_{v,j} = e) = \hat{p}(.|\mathbf{R}^v). \tag{11}$$

With the missing position reminder matrix, we can obtain that

$$\begin{aligned}
&\hat{p}(\mathbf{X}^v|\mathbf{R}^v, m^i_{v,j} = e) = \hat{p}(\mathbf{X}^v|\mathbf{M}^{v,e}, \mathbf{K}^v) \\
=&\frac{\hat{p}(\mathbf{X}^v|\mathbf{M}^{v,e})}{\hat{p}(\mathbf{K}^v|\mathbf{M}^{v,e})} = \frac{\hat{p}(\mathbf{X}^v|\mathbf{M}^{v,e})}{p_k(\mathbf{K}^v)}.
\end{aligned} \tag{12}$$

Taking Eq. (12) into Eq. (11), we can obtain that

$$\hat{p}(\mathbf{X}^v|\mathbf{M}^{v,0}) = \hat{p}(\mathbf{X}^v|\mathbf{M}^{v,1}), \tag{13}$$

where $\mathbf{M}^{v,0}$ and $\mathbf{M}^{v,1}$ are two missing position reminder matrices with only one element difference and they share the same K^v. p_k is the marginal distribution of K^v

Suppose M^v_a and M^v_b are any two missing position reminder matrices in the v-th view, so there is a series of matrices $M_1 ... M_k$, where M_{k-1} and M_k differ in only one element

Table 2: Clustering results on the **Aminer** dataset in different missing rates.

Missing Rate	m=0.1				m=0.3				m=0.5				m=0.7			
	ACC	NMI	F1	ARI	ACC	NMI	F1	ARI	ACC	NMI	F1	ARI	ACC	NMI	F1	ARI
AGCN0	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000
AGCNM	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000
SDCN0	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000
SDCNM	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000	0.308	0.001	0.118	0.000
ASD-VAE	0.231	0.017	0.159	0.012	0.197	0.011	0.135	0.005	0.196	0.009	0.132	0.004	0.201	0.004	0.134	0.001
MvAGC0	0.329	0.016	0.180	0.003	0.309	0.006	0.214	0.004	0.298	0.002	0.166	-0.002	0.295	0.001	0.174	0.001
MvAGCM	0.313	0.011	0.195	0.007	0.303	0.004	0.162	-0.001	0.309	0.003	0.176	0.001	0.293	0.000	0.198	0.000
MAGC0	0.293	0.020	0.202	-0.001	0.291	0.018	0.210	-0.001	0.298	0.020	0.209	0.000	0.289	0.014	0.202	-0.001
MAGCM	0.294	0.019	0.218	-0.001	0.296	0.020	0.186	0.001	0.299	0.017	0.197	0.002	0.290	0.014	0.177	-0.001
Our	0.332	0.026	0.295	0.013	0.329	0.024	0.307	0.014	0.307	0.019	0.286	0.012	0.312	0.012	0.274	0.010

Table 3: Clustering results on the **Citeseer** dataset in different missing rates.

Missing Rate	m=0.1				m=0.3				m=0.5				m=0.7			
	ACC	NMI	F1	ARI	ACC	NMI	F1	ARI	ACC	NMI	F1	ARI	ACC	NMI	F1	ARI
AGCN0	0.643	0.383	0.554	0.356	0.624	0.356	0.573	0.368	0.547	0.298	0.470	0.294	0.473	0.225	0.397	0.204
AGCNM	0.639	0.386	0.548	0.360	0.626	0.356	0.576	0.371	0.585	0.330	0.526	0.333	0.457	0.212	0.367	0.195
SDCN0	0.504	0.269	0.453	0.235	0.397	0.186	0.334	0.147	0.406	0.167	0.382	0.124	0.272	0.043	0.236	0.034
SDCNM	0.433	0.231	0.376	0.186	0.377	0.119	0.301	0.098	0.349	0.111	0.322	0.079	0.258	0.043	0.231	0.021
ASD-VAE	0.350	0.186	0.353	0.147	0.271	0.117	0.274	0.081	0.205	0.003	0.134	0.001	0.185	0.002	0.136	0
MvAGC0	0.469	0.174	0.355	0.164	0.442	0.137	0.386	0.117	0.350	0.098	0.281	0.050	0.211	0.004	0.060	0.000
MvAGCM	0.432	0.172	0.363	0.144	0.397	0.124	0.343	0.107	0.333	0.064	0.297	0.050	0.213	0.003	0.082	-0.001
MAGC0	0.589	0.361	0.499	0.322	0.584	0.360	0.495	0.317	0.558	0.340	0.472	0.283	0.494	0.333	0.378	0.240
MAGCM	0.635	0.384	0.550	0.362	0.624	0.367	0.543	0.338	0.609	0.356	0.531	0.316	0.528	0.323	0.406	0.273
Our	0.659	0.391	0.615	0.403	0.648	0.378	0.603	0.383	0.631	0.369	0.596	0.363	0.591	0.317	0.559	0.323

and $M_a^v = M_1$, $M_b^v = M_k$. Then, we can obtain that

$$\hat{p}(\mathbf{X}^v|\mathbf{M}_a^v) = \hat{p}(\mathbf{X}^v|\mathbf{M}_1) = ... = \hat{p}(\mathbf{X}^v|\mathbf{M}_k) = \hat{p}(\mathbf{X}^v|\mathbf{M}_b^v) \tag{14}$$

and so in particular, for any M_v we have that

$$\hat{p}(\mathbf{X}^v|\mathbf{M}^v) = \hat{p}(\mathbf{X}^v|1). \tag{15}$$

Since $\hat{p}(\mathbf{X}^v|1)$ is the unique true distribution of attributes in the v-th view, thus the view-specific generator can generate a unique data distribution that is the same as the real data distribution with the missing position reminder matrix in each view. □

A.3 Time complexity analysis

Time complexity analysis. The time complexity of the TOTF algorithm is near $\mathcal{O}(n^2)$. The time complexity of CSAGNN is $\mathcal{O}(nd_v\hat{d}_l^v + nd_v^2 + nd + n^2\hat{d}_l^v)$. As for Stage 1, the time complexity of calculating R^v is $\mathcal{O}(nd)$, and each view can perform independently. As discriminators consist of several linear layers, the time complexity of discriminators is near to $\mathcal{O}(n)$. Thus, the whole time complexity of Stage 1 is near to $\mathcal{O}(n^2)$. In Stage 2, the time complexity of the common extractor is $\mathcal{O}(nVd^2)$ where d is the common dimension. Considering the encoders and decoders, the time complexity of Stage 2 is near to $\mathcal{O}(n^2)$. In summary, the time complexity of the TOTF algorithm is near to $\mathcal{O}(n^2)$. Particularly, for sparse graphs, the overall complexity of CSAGNN is near linear complexity, the same as the traditional GNNs, with no additional cost. Since the adjacency matrix is sparse, assuming that the number of non-zero elements in the matrix is n_z, where n_z is much smaller than n^2, the overall complexity of CSAGNN is $\mathcal{O}(nd_v\hat{d}_l^v + nd_v^2 + nd + n_z\hat{d}_l^v)$, which is near linear complexity.

B Experiment Supplement

B.1 Experiment Setting

We use four evaluation metrics to assess clustering results: ACC, F1-score, NMI, and ARI [Strehl and Ghosh2002, Fan *et al.*2020]. Additionally, we conduct experiments on five datasets covering different types of multi-view attributed graphs. Detailed information of datasets is shown in Table 4. The ACM[3] and AMiner [Hong *et al.*2020] datasets each contain one attributed matrix and multiple adjacency matrices. Cora and Citeseer datasets have one adjacency matrix and several attributed matrices, with attributed matrices in the second view constructed via a log-scale. The Wiki dataset [Fettal *et al.*2023] includes multiple attributed and adjacency matrices, with additional views derived from original attributed and adjacency matrices.

Model Setting: In the experiments, the three-layer parameters of view encoder are 128, 256, and 64. The three-layer parameters of the view decoder are 64, 256, and 128. The common extractor consists of two fully connected layers with 256 nodes and two exponential linear units. The discriminator comprises two linear layers with 256 nodes along with two active layers. The total number of epochs is set to 200.

B.2 Clustering Performance

We compare our method with representative approaches in recent years across five datasets. Tables 2 and 3 list the clustering performance on the Aminer and Citeseer datasets.

[3]http://dl.acm.org

Table 4: Multi-view attributed graph datasets.

Dataset	**Views**	**Nodes**	**Attributes**	**Edges**	**Clusters**
ACM	2	3,025	1,830	29,281 2,210,761	3
AMiner	2	8,052	5	29,646 31,224	4
Wiki	4	2,405	4,973	24,357 12,025	17
Cora	2	2,708	1,433	5,429	7
Citeseer	2	3,327	3,703	4,732	6

References

[Fan *et al.*, 2020] Shaohua Fan, Xiao Wang, Chuan Shi, Emiao Lu, Ken Lin, and Bai Wang. One2multi graph autoencoder for multi-view graph clustering. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pages 3070–3076, Taipei, April 2020. ACM / IW3C2.

[Fettal *et al.*, 2023] Chakib Fettal, Lazhar Labiod, and Mohamed Nadif. Simultaneous linear multi-view attributed graph representation learning and clustering. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, WSDM 2023, Singapore, 27 February 2023 - 3 March 2023*, pages 303–311, Singapore, February 2023. ACM.

[Hong *et al.*, 2020] Huiting Hong, Hantao Guo, Yucheng Lin, Xiaoqing Yang, Zang Li, and Jieping Ye. An attention-based graph neural network for heterogeneous structural learning. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 4132–4139, New York, NY, USA, February 2020. AAAI Press.

[Miao *et al.*, 2021] Xiaoye Miao, Yangyang Wu, Jun Wang, Yunjun Gao, Xudong Mao, and Jianwei Yin. Generative semi-supervised learning for multivariate time series imputation. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 8983–8991, Virtual Event, February 2021. AAAI Press.

[Strehl and Ghosh, 2002] Alexander Strehl and Joydeep Ghosh. Cluster ensembles-A knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, 3:583–617, 2002.

[Yoon *et al.*, 2018] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. GAIN: missing data imputation using generative adversarial nets. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 5675–5684, Stockholm, Sweden, July 2018. PMLR.