Table 1: Classification performance (*i.e.,* Macro-F1 and Micro-F1) of the affinity matrix and self-attention mechanisms on heterogeneous graph datasets.

Method	**ACM**		**Yelp**		**DBLP**		**Aminer**	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
InfoNCE	91.3±1.1	91.2±0.8	92.4±0.7	92.0±0.8	**94.1±0.6**	94.6±0.5	76.8±0.4	85.4±0.3
Proposed	**92.7±0.6**	**92.6±0.5**	**93.0±0.7**	**92.8±0.4**	94.0±0.3	**94.7±0.4**	**77.5±0.9**	**86.8±0.7**

Table 2: Classification performance (*i.e.,* Macro-F1 and Micro-F1) of variant methods with the affinity matrix, cosine similarity and, self-attention mechanisms on heterogeneous graph datasets.

Method	**ACM**		**Yelp**		**DBLP**		**Aminer**	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
cosine similarity	85.3±0.9	85.1±1.1	88.2±0.4	87.7±0.7	89.3±0.7	90.5±0.8	67.6±0.5	75.4±0.6
self-attention	88.7±0.8	88.4±0.7	92.0±0.5	91.7±0.6	91.2±0.4	92.1±0.6	73.2±0.7	82.1±0.6
affinity matrix	**92.7±0.6**	**92.6±0.5**	**93.0±0.7**	**92.8±0.4**	**94.0±0.3**	**94.7±0.4**	**77.5±0.9**	**86.8±0.7**

Table 3: Settings for the dimensions of encoders (i.e., $g_\phi \in \mathbb{R}^{f \times d_1}$ and $f_\theta \in \mathbb{R}^{f \times d_1}$) and projection heads (i.e., $p_\varphi \in \mathbb{R}^{d_1 \times c}$ and $q_\gamma \in \mathbb{R}^{d_1 \times d_2}$) on all datasets.

Settings	ACM	Yelp	DBLP	Aminer	Photo	Computers
f	1,902	82	334	128	745	767
d_1	512	256	128	256	1024	1024
d_2	64	256	256	256	256	256
c	3	3	4	4	8	10

Table 4: Classification performance (*i.e.,* Macro-F1 and Micro-F1) of variant methods with and without (w/o) the simplex constraint on the columns of $\mathbf{Y}$.

Method	**ACM**		**Yelp**		**DBLP**		**Aminer**	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
with constraint	92.3±0.5	92.1±0.6	92.8±0.8	92.7±0.9	93.1±0.3	94.0±0.4	75.1±0.7	84.2±0.5
w/o constraint	**92.7±0.6**	**92.6±0.5**	**93.0±0.7**	**92.8±0.4**	**94.0±0.3**	**94.7±0.4**	**77.5±0.9**	**86.8±0.7**

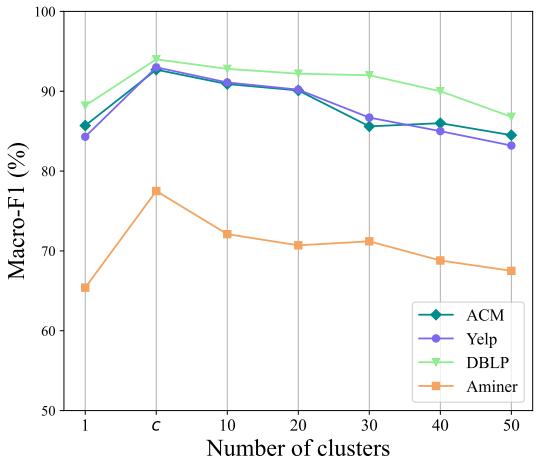

Figure 1: Classification performance (i.e., Macro-F1) of the proposed method under different clusters.