

L'hétérophilie : un cauchemar pour la classification de nœuds via des GNNs ?

Is Heterophily A Real Nightmare For Graph Neural Networks To Do Node Classification?

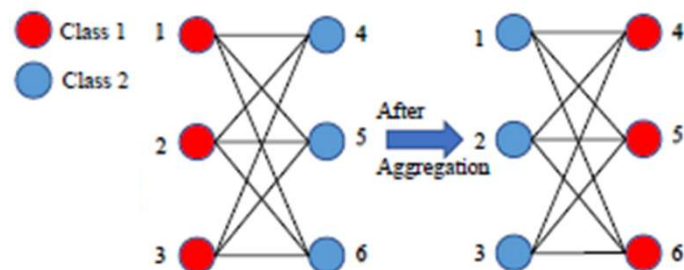
Sitao Luan^{1,2}, Chenqing Hua¹, Qincheng Lu¹, Jiaqi Zhu¹, Mingde Zhao^{1,2}, Shuyuan Zhang^{1,2},
 Xiao-Wen Chang¹, Doina Precup^{1,2,3}
 {sitao.luan@mail, chenqing.hua@mail, qincheng.lu@mail, jiaqi.zhu@mail, mingde.zhao@mail,
 shuyuan.zhang@mail, chang@cs, dprecup@cs}.mcgill.ca
¹McGill University; ²Mila; ³DeepMind

2021, <https://arxiv.org/abs/2109.05641>

- Hétérophilie et GNNs :
 - L'hétérophilie est-elle toujours un problème ?
 - Comment mesurer l'hétérophilie ?
 - Hétérophilie et GNN : comment la structure d'un graphe affecte-t-elle les performances prédictives du modèle ?
- Hétérophilie et filtrage du signal
 - Rappel d'analyse spectrale des graphes : transformée de Fourier et opérateur Laplacien
 - Quels liens entre hétérophilie et filtrage du signal de graphe ?
 - La diversification : une solution pour résoudre les problèmes liés à l'hétérophilie ? (introduction au framework ACM – Adaptive Channel Mixing)

Hétérophilie et GNNs

- Le principe d'agrégation de *features* dans les GNNs repose sur un principe d'homophilie (un nœud et son voisinage partagent, en moyenne, des caractéristiques communes, en particulier en ce qui concerne leur classe).
- L'hétérophilie serait l'une des raisons des contre-performances de GNNs basés sur le principe de *message passing* (intuitivement, des features de nœuds de classes différentes sont mixées 'de manière incorrecte' et cela a pour conséquence de rendre ces nœuds non-distinguables)
- Mais l'hétérophilie n'est pas toujours un problème :



Mesures de l'hétérophilie

Graphe $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$

Matrice d'adjacence symétrique $A \in \mathbb{R}^{N \times N}$

Vecteur de features $X \in \mathbb{R}^{N \times F}$

Matrice diagonale des degrés D , $D_{i,i} = d_i = \sum_j A_{i,j}$

Voisinage d'un nœud $\mathcal{N}_i = \{j : e_{ij} \in \mathcal{E}\}$

Label encoding matrix (1-hot encoding) $Z \in \mathbb{R}^{N \times C}$

Mesure de la proportion de liens qui connectent 2 nœuds de même classe :

$$H_{\text{edge}}(\mathcal{G}) = \frac{|\{e_{uv} \mid e_{uv} \in \mathcal{E}, Z_{u,:} = Z_{v,:}\}|}{|\mathcal{E}|}$$

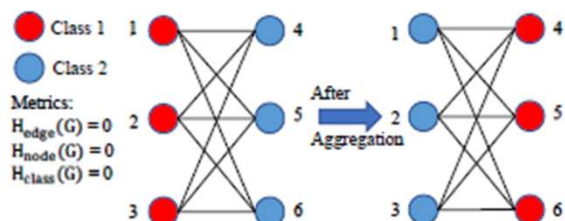
Mesure de la proportion moyenne de nœuds de même classe dans le voisinage d'un nœud :

$$H_{\text{node}}(\mathcal{G}) = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \frac{|\{u \mid u \in \mathcal{N}_v, Z_{u,:} = Z_{v,:}\}|}{d_v}$$

Evaluation de la surreprésentation des classes identiques à aux nœuds dans leur voisinage par rapport à la distribution des classes sur dans le graphe :

$$H_{\text{class}}(\mathcal{G}) = \frac{1}{C-1} \sum_{k=1}^C \left[h_k - \frac{|\{v \mid Z_{v,k} = 1\}|}{N} \right]_+, \quad h_k = \frac{\sum_{v \in \mathcal{V}} |\{u \mid Z_{u,k} = 1, u \in \mathcal{N}_v, Z_{u,:} = Z_{v,:}\}|}{\sum_{v \in \{v \mid Z_{v,k} = 1\}} d_v}$$

- Varient dans l'intervalle [0,1]
- Une valeur proche de 1 correspond à une forte homophilie
- Une valeur proche de 0 correspond à une forte hétérophilie



Les 3 mesures ci-dessus ne permettent pas de distinguer les cas d'hétérophilie « non-problématiques » (*non-harmful heterophily*) des cas « problématiques » (*harmful heterophily*)

Mesures de l'hétérophilie

Vers une nouvelle mesure de l'hétérophilie qui prend en compte la structure du graphe ?

Rappel sur le GCN :

$$Y = \text{softmax}(\hat{A}_{\text{sym}} \text{ReLU}(\hat{A}_{\text{sym}} X W_0) W_1)$$

Soit un opérateur général d'agrégation

Simplifions le GCN (1) en supprimant les $\hat{A} \in \mathbb{R}^{N \times N}$ on obtient:

$$Y = \text{softmax}(\hat{A} X W) = \text{softmax}(Y')$$

Essayons maintenant de regarder comment évolue Y à chaque étape de la descente de gradient :

$$\Delta Y' = \hat{A} X \Delta W = \gamma \hat{A} X \frac{d\mathcal{L}}{dW} \propto \hat{A} X \frac{d\mathcal{L}}{dW} = \boxed{\hat{A} X X^T \hat{A}^T} (Z - Y) \quad \mathcal{L} = -\text{trace}(Z^T \log Y) \text{ (Cross-entropie)}$$

Pour la suite, on utilisera $S(\hat{A}, X)$ pour désigner une matrice de similarité des nœuds post-agrégation

$$S(\hat{A}, X) \equiv \hat{A} X (\hat{A} X)^T$$

Mesures de l'hétérophilie

Vers une nouvelle mesure de l'hétérophilie qui prend en compte la structure du graphe ?

Définition : *aggregation similarity score*

$$S_{agg}(S(\hat{A}, X)) = \frac{|\{v \mid \text{Mean}_u(\{S(\hat{A}, X)_{v,u} \mid Z_{u,:} = Z_{v,:}\}) \geq \text{Mean}_u(\{S(\hat{A}, X)_{v,u} \mid Z_{u,:} \neq Z_{v,:}\})\}|}{|\mathcal{V}|}$$

Ce score mesure la proportion de nœuds qui assignent, de manière relative, des poids de similarité plus forts aux nœuds qui sont de même classe qu'à ceux d'autres classe après agrégation.

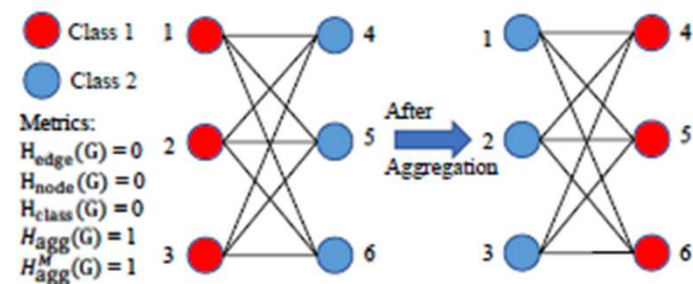
$$S_{agg}(S(\hat{A}, X)) \in [0, 1]$$

Mais en pratique, pour la plupart des jeux de données : $S_{agg}(S(\hat{A}, X)) \geq 0.5$

Et donc les auteurs proposent d'utiliser (en pratique) :

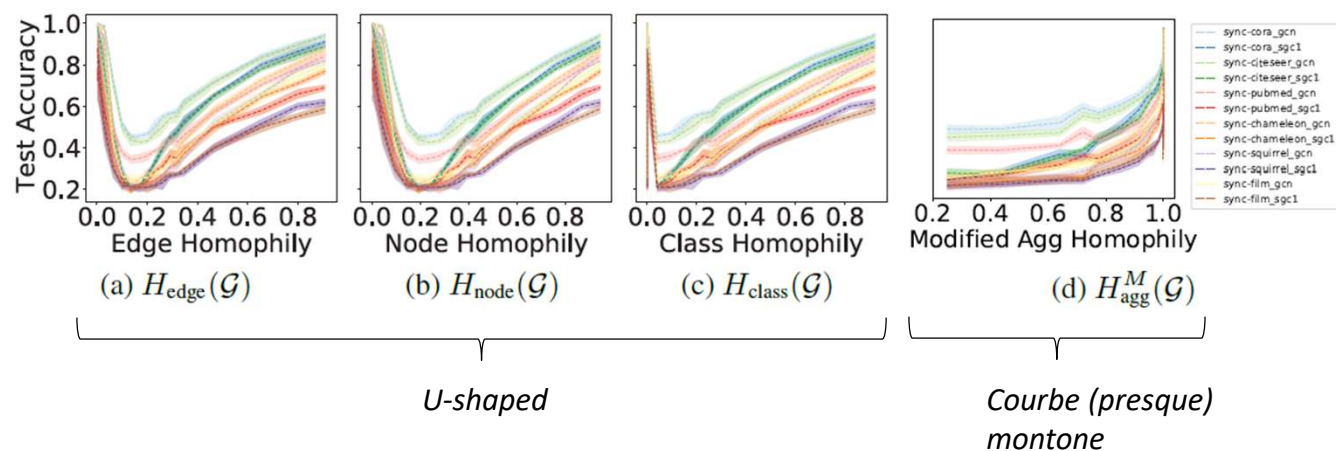
$$S_{agg}^M(S(\hat{A}, X)) = [2S_{agg}(S(\hat{A}, X)) - 1]_+$$

$$H_{agg}(\mathcal{G}) = S_{agg}(S(\hat{A}, Z)), \quad H_{agg}^M(\mathcal{G}) = S_{agg}^M(S(\hat{A}, Z))$$



Hétérophilie et performance des GCNs

Evaluation et comparaison sur des graphes synthétiques :



Rappel de théorie spectrale des graphes

Pour un signal temporel discret :

$$\Delta f(t) = f(t+1) - f(t)$$

On peut généraliser à une fonction *smooth arbitraire* $f : \mathbb{R}^d \rightarrow \mathbb{R}$:

$$\begin{aligned} \Delta f(\mathbf{x}) &= \nabla^2 f(\mathbf{x}) \\ &= \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} \end{aligned}$$

Propriété importante du Laplacien : ses fonctions propres sont les mêmes que celles qui permettent de définir les modes de fréquence d'une transformée de Fourier

$$-\Delta(e^{2\pi i s t}) = -\frac{\partial^2 (e^{2\pi i s t})}{\partial t^2} = (2\pi s)^2 e^{2\pi i s t}$$

Rappel de théorie spectrale des graphes

Dans le domaine des graphes, le Laplacien est défini par :

$$L = I - A$$

On peut généraliser la transformation de Fourier aux graphes en considérant les valeurs et vecteurs propres du Laplacien :

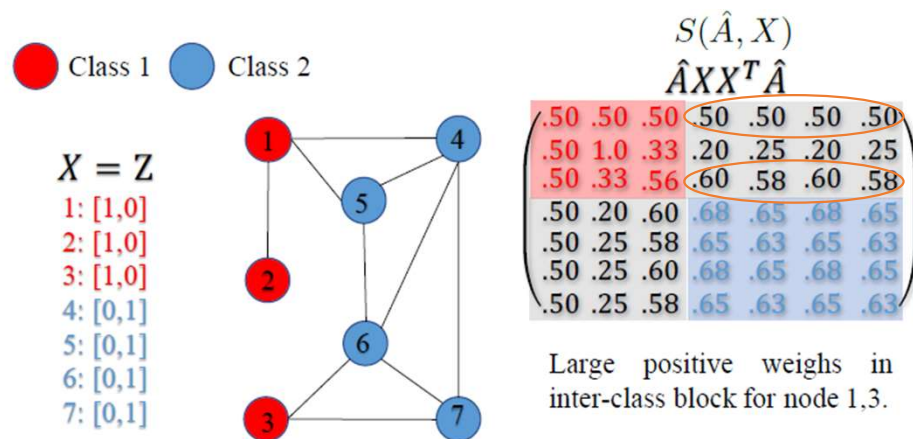
$$\begin{aligned} L &= U\Lambda U^T \\ A &= U(I - \Lambda)U^T \end{aligned}$$

Une convolution d'un signal f par un filtre h sur un graphe s'exprime donc ainsi :

$$\begin{aligned} f \star_{\mathcal{G}} h &= U (U^T f \circ U^T h) \\ &= (U \text{diag}(\theta_h) U^T) f \quad \text{où} \quad \theta_h = U^T h \in \mathbb{R}^{|\mathcal{V}|} \\ &= (U p_N(\Lambda) U^T) f \\ &= p_N(L) f \end{aligned}$$

Pour le GCN, la simple convolution est $I + A$

Hétérophilie et filtres



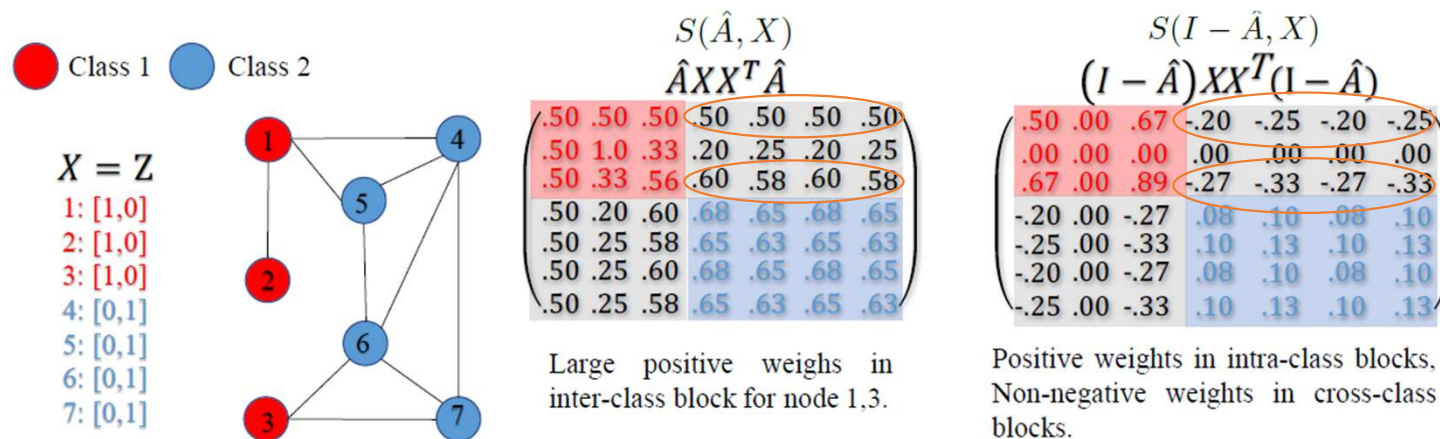
Les nœuds 1,3 assignent des poids positifs larges aux nœuds de classe 2 après agrégation, cela les rends donc difficile à distinguer de ces nœuds de classe 2.

Cependant, on peut les distinguer par la structure de leur voisinage :

- les nœuds 1 et 3 sont différents (en terme de classe) de la plupart de leurs voisins.
- les nœuds 4, 5, 6, 7 sont similaires à la plupart de leurs voisins.

Idée : réaliser une *opération de diversification* (i.e. utiliser un filtre passe haut : $I - A$)

Hétérophilie et filtres



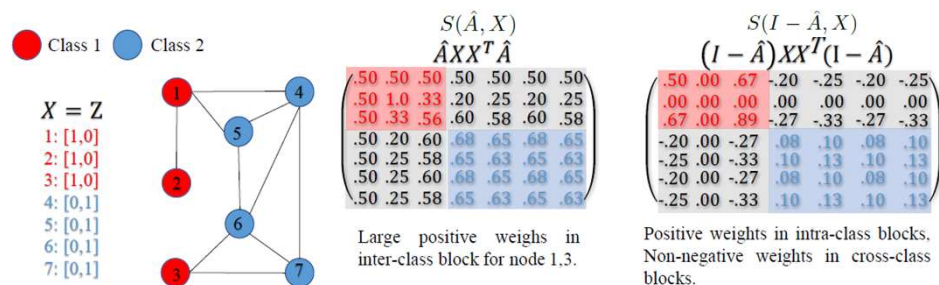
Les nœuds 1,3 assignent des poids positifs larges aux nœuds de classe 2 après agrégation, cela les rends donc difficile à distinguer de ces nœuds de classe 2.

Cependant, on peut les distinguer par la structure de leur voisinage :

- les nœuds 1 et 3 sont différents (en terme de classe) de la plupart de leurs voisins.
- les nœuds 4, 5, 6, 7 sont similaires à la plupart de leurs voisins.

Idée : réaliser une *opération de diversification* (i.e. utiliser un filtre passe haut : $I - A$)

Hétérophilie et filtres



Definition 2. *Diversification Distinguishability (DD) based on $S(I - \hat{A}, X)$.*

Given $S(I - \hat{A}, X)$, a node v is diversification distinguishable if the following two conditions are satisfied at the same time,

1. $\text{Mean}_u \left(\{S(I - \hat{A}, X)_{v,u} | u \in \mathcal{V} \wedge Z_{u,:} = Z_{v,:}\} \right) \geq 0$;
2. $\text{Mean}_u \left(\{S(I - \hat{A}, X)_{v,u} | u \in \mathcal{V} \wedge Z_{u,:} \neq Z_{v,:}\} \right) \leq 0$

Then, graph diversification distinguishability value is defined as

$$\text{DD}_{\hat{A}, X}(\mathcal{G}) = \frac{1}{|\mathcal{V}|} |\{v | v \text{ is diversification distinguishable}\}|$$

We can see that $\text{DD}_{\hat{A}, X}(\mathcal{G}) \in [0, 1]$.

ACM – Adaptive Channel Mixing

On se dote de filtres spatiaux :

- Passe-bas H_{LP} (les différentes versions de la matrice d'adjacence par exemple)
- Passe-haut H_{HP} (les différentes versions du Laplacien par exemple)

La plupart des modèles de GNN reposent sur une architecture de type filtrage uni-canal (le canal étant un filtre passe-haut ou un filtre passe-bas)

L'architecture ACM propose de prendre en compte et combiner l'information venant de plusieurs canaux : un filtre passe haut, un filtre passe-bas et l'identité.

Step 1. Feature Extraction for Each Channel:

Option 1: $H_L^l = \text{ReLU}(H_{LP}H^{l-1}W_L^{l-1})$, $H_H^l = \text{ReLU}(H_{HP}H^{l-1}W_H^{l-1})$, $H_I^l = \text{ReLU}(IH^{l-1}W_I^{l-1})$;

Option 2: $H_L^l = H_{LP}\text{ReLU}(H^{l-1}W_L^{l-1})$, $H_H^l = H_{HP}\text{ReLU}(H^{l-1}W_H^{l-1})$, $H_I^l = I \text{ReLU}(H^{l-1}W_I^{l-1})$;

$W_L^{l-1}, W_H^{l-1}, W_I^{l-1} \in \mathbb{R}^{F_{l-1} \times F_l}$;

Step 2. Feature-based Weight Learning

$\tilde{\alpha}_L^l = \sigma(H_L^l \tilde{W}_L^l)$, $\tilde{\alpha}_H^l = \sigma(H_H^l \tilde{W}_H^l)$, $\tilde{\alpha}_I^l = \sigma(H_I^l \tilde{W}_I^l)$, $\tilde{W}_L^{l-1}, \tilde{W}_H^{l-1}, \tilde{W}_I^{l-1} \in \mathbb{R}^{F_l \times 1}$

$[\alpha_L^l, \alpha_H^l, \alpha_I^l] = \text{Softmax}([\tilde{\alpha}_L^l, \tilde{\alpha}_H^l, \tilde{\alpha}_I^l] W_{\text{Mix}}^l / T)$, $W_{\text{Mix}}^l \in \mathbb{R}^{3 \times 3}$, $T \in \mathbb{R}$ is the temperature;

Step 3. Node-wise Channel Mixing:

$H^l = (\text{diag}(\alpha_L^l)H_L^l + \text{diag}(\alpha_H^l)H_H^l + \text{diag}(\alpha_I^l)H_I^l)$.

ACM – Adaptive Channel Mixing

	Cornell	Wisconsin	Texas	Film	Chameleon	Squirrel	Deezer-Europe	Cora	CiteSeer	PubMed	
#nodes	183	251	183	7,600	2,277	5,201	28,281	2,708	3,327	19,717	
#edges	295	499	309	33,544	36,101	217,073	92,752	5,429	4,732	44,338	
#features	1,703	1,703	1,703	931	2,325	2,089	31,241	1,433	3,703	500	
#classes	5	5	5	5	5	5	2	7	6	3	
H _{edge}	0.5669	0.4480	0.4106	0.3750	0.2795	0.2416	0.5251	0.8100	0.7362	0.8024	
H _{node}	0.3855	0.1498	0.0968	0.2210	0.2470	0.2156	0.5299	0.8252	0.7175	0.7924	
H _{class}	0.0468	0.0941	0.0013	0.0110	0.0620	0.0254	0.0304	0.7657	0.6270	0.6641	
Data Splits	60%/20%/20%	60%/20%/20%	60%/20%/20%	60%/20%/20%	60%/20%/20%	60%/20%/20%	50%/25%/25%	60%/20%/20%	60%/20%/20%	60%/20%/20%	
H _{aggM(G)}	0.8032	0.7768	0.694	0.6822	0.61	0.3566	0.5790	0.9904	0.9826	0.9432	
Test Accuracy (%) of State-of-the-art Models, Baseline GNN Models and ACM-GNN models											
											Rank
MLP-2*	91.30 ± 0.70	93.87 ± 3.33	92.26 ± 0.71	38.58 ± 0.25	46.72 ± 0.46	31.28 ± 0.27	66.55 ± 0.72	76.44 ± 0.30	76.25 ± 0.28	86.43 ± 0.13	18.60
GAT*	76.00 ± 1.01	71.01 ± 4.66	78.87 ± 0.86	35.98 ± 0.23	63.9 ± 0.46	42.72 ± 0.33	61.09 ± 0.77	76.70 ± 0.42	67.20 ± 0.46	83.28 ± 0.12	21.40
APPNP*	91.80 ± 0.63	92.00 ± 3.59	91.18 ± 0.70	38.86 ± 0.24	51.91 ± 0.56	34.77 ± 0.34	67.21 ± 0.56	79.41 ± 0.38	68.59 ± 0.30	85.02 ± 0.09	18.00
GPRGNN*	91.36 ± 0.70	93.75 ± 2.37	92.92 ± 0.61	39.30 ± 0.27	67.48 ± 0.40	49.93 ± 0.53	66.90 ± 0.50	79.51 ± 0.36	67.63 ± 0.38	85.07 ± 0.09	14.40
H2GCN	86.23 ± 4.71	87.5 ± 1.77	85.90 ± 3.53	38.85 ± 1.17	52.30 ± 0.48	30.39 ± 1.22	67.22 ± 0.90	87.52 ± 0.61	79.97 ± 0.69	87.78 ± 0.28	17.00
MixHop	60.33 ± 28.53	77.25 ± 7.80	76.39 ± 7.66	33.13 ± 2.40	36.28 ± 10.22	24.55 ± 2.60	66.80 ± 0.58	65.65 ± 11.31	49.52 ± 13.35	87.04 ± 4.10	23.50
GCN+JK	66.56 ± 13.82	62.50 ± 15.75	80.66 ± 1.91	32.72 ± 2.62	64.68 ± 2.85	53.40 ± 1.90	60.99 ± 0.14	86.90 ± 1.51	73.77 ± 1.85	90.09 ± 0.68	18.80
GAT+JK	74.43 ± 10.24	69.50 ± 3.12	75.41 ± 7.18	35.41 ± 0.97	68.14 ± 1.18	52.28 ± 3.61	59.66 ± 0.92	89.52 ± 0.43	74.49 ± 2.76	89.15 ± 0.87	16.70
FAGCN	88.03 ± 5.6	89.75 ± 6.37	88.85 ± 4.39	31.59 ± 1.37	49.47 ± 2.84	42.24 ± 1.2	66.86 p. 0.53	88.85 ± 1.36	82.37 ± 1.46	89.98 ± 0.54	14.10
GraphSAGE	71.41 ± 1.24	64.85 ± 5.14	79.03 ± 1.20	36.37 ± 0.21	62.15 ± 0.42	41.26 ± 0.26	OOM	86.58 ± 0.26	78.24 ± 0.30	86.85 ± 0.11	20.89
Geom-GCN†	60.81	64.12	67.57	31.63	60.9	38.14	NA	85.27	77.99	90.05	22.67
SGC-1	70.98 ± 8.39	70.38 ± 2.85	83.28 ± 5.43	25.26 ± 1.18	64.86 ± 1.81	47.62 ± 1.27	59.73 ± 0.12	85.12 ± 1.64	79.66 ± 0.75	85.5 ± 0.76	20.10
SGC-2	72.62 ± 9.92	74.75 ± 2.89	81.31 ± 3.3	28.81 ± 1.11	62.67 ± 2.41	41.25 ± 1.4	61.56 ± 0.51	85.48 ± 1.48	80.75 ± 1.15	85.36 ± 0.52	20.70
GCNII	89.18 ± 3.96	83.25 ± 2.69	82.46 ± 4.58	40.82 ± 1.79	60.35 ± 2.7	38.81 ± 1.97	66.38 ± 0.45	88.98 ± 1.33	81.58 ± 1.3	89.8 ± 0.3	14.80
GCNII*	90.49 ± 4.45	89.12 ± 3.06	88.52 ± 3.02	41.54 ± 0.99	62.8 ± 2.87	38.31 ± 1.3	66.42 ± 0.56	88.93 ± 1.37	81.83 ± 1.78	89.98 ± 0.52	12.30
GCN	82.46 ± 3.11	75.5 ± 2.92	83.11 ± 3.2	35.51 ± 0.99	64.18 ± 2.62	44.76 ± 1.39	62.23 ± 0.53	87.78 ± 0.96	81.39 ± 1.23	88.9 ± 0.32	16.30
Snowball-2	82.62 ± 2.34	74.88 ± 3.42	83.11 ± 3.2	35.97 ± 0.66	64.99 ± 2.39	47.88 ± 1.23	OOM	88.64 ± 1.15	81.53 ± 1.71	89.04 ± 0.49	15.22
Snowball-3	82.95 ± 2.1	69.5 ± 5.01	83.11 ± 3.2	36.00 ± 1.36	65.49 ± 1.64	48.25 ± 0.94	OOM	89.33 ± 1.3	80.93 ± 1.32	88.8 ± 0.82	14.78
ACM-SGC-1	93.77 ± 1.91	93.25 ± 2.92	93.61 ± 1.55	39.33 ± 1.25	63.68 ± 1.62	46.4 ± 1.13	66.67 ± 0.56	86.63 ± 1.13	80.96 ± 0.93	87.75 ± 0.88	12.60
ACM-SGC-2	93.77 ± 2.17	94.00 ± 2.61	93.44 ± 2.54	40.13 ± 1.21	60.48 ± 1.55	40.91 ± 1.39	66.53 ± 0.57	87.64 ± 0.99	80.93 ± 1.16	88.79 ± 0.5	13.40
ACM-GCNII	92.62 ± 3.13	94.63 ± 2.96	92.46 ± 1.97	41.37 ± 1.37	58.73 ± 2.52	40.9 ± 1.58	66.39 ± 0.56	89.1 ± 1.61	82.28 ± 1.12	90.12 ± 0.4	10.40
ACM-GCNII*	93.44 ± 2.74	94.37 ± 2.81	93.28 ± 2.79	41.27 ± 1.24	61.66 ± 2.29	38.32 ± 1.5	66.6 ± 0.57	89.00 ± 1.35	81.69 ± 1.25	90.18 ± 0.51	10.10
ACM-GCN	94.75 ± 3.8	95.75 ± 2.03	94.92 ± 2.88	41.62 ± 1.15	69.04 ± 1.74	58.02 ± 1.86	67.01 ± 0.38	88.62 ± 1.22	81.68 ± 0.97	90.66 ± 0.47	4.80
ACM-Snowball-2	95.08 ± 3.11	96.38 ± 2.59	95.74 ± 2.22	41.4 ± 1.23	68.51 ± 1.7	55.97 ± 2.03	OOM	88.83 ± 1.49	81.58 ± 1.23	90.81 ± 0.52	4.44
ACM-Snowball-3	94.26 ± 2.57	96.62 ± 1.86	94.75 ± 2.41	41.27 ± 0.8	68.4 ± 2.05	55.73 ± 2.39	OOM	89.59 ± 1.58	81.32 ± 0.97	91.44 ± 0.59	4.44
ACMII-GCN	95.9 ± 1.83	96.62 ± 2.44	95.08 ± 2.07	41.84 ± 1.15	68.38 ± 1.36	54.53 ± 2.09	67.15 ± 0.41	89.00 ± 0.72	81.79 ± 0.95	90.74 ± 0.5	3.40
ACMII-Snowball-2	95.25 ± 1.55	96.63 ± 2.24	95.25 ± 1.55	41.1 ± 0.75	67.83 ± 2.63	53.48 ± 0.6	OOM	88.95 ± 1.04	82.07 ± 1.04	90.56 ± 0.39	4.78
ACMII-Snowball-3	93.61 ± 2.79	97.00 ± 2.63	94.75 ± 3.09	40.31 ± 1.6	67.53 ± 2.83	52.31 ± 1.57	OOM	89.36 ± 1.26	81.56 ± 1.15	91.31 ± 0.6	5.89