How to Define the Homophily Ratio on Hypergraph

Homophily on Hypergraphs

Definition 1 (Homophily on Hypergraphs).

$$\frac{1}{m}\sum_{j=1}^{m}\frac{\sum_{(u,v)\in e_j}\mathbb{1}(y_u=y_v)}{\binom{n_j}{2}},$$

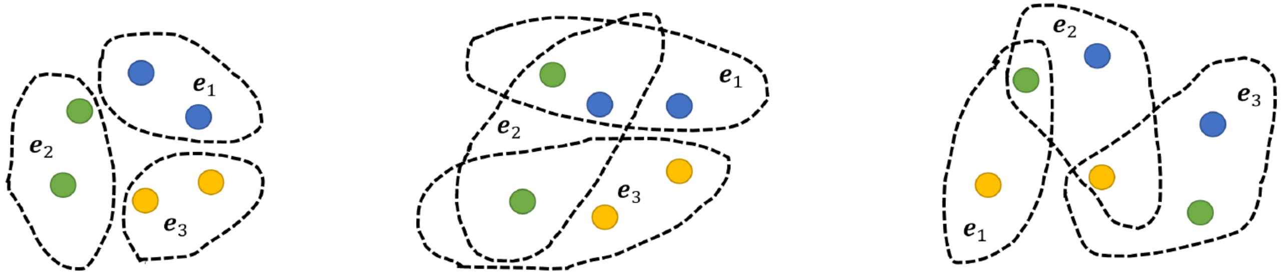
New Definition

where $\mathbb{1}(\cdot)$ is the indicator function (i.e., $\mathbb{1}(\cdot)=1$ if the condition holds, otherwise $\mathbb{1}(\cdot)=0$), and e_j $(j=1,2,\ldots,m)$ is the hyper edge, n_j represents the number of nodes within e_j.

As far as I know, there is **no existing work** studying the topic "**Beyond Homophily in Hypergraph Neural Networks (HGNN)**" or "**Hypergraph Neural Networks (HGNN) with Heterophily**"

: Class I
: Class II
: Class III

e_i $(i=1,2,3)$ denotes the hyper edge
ρ is the homophily ratio



Case 1
ρ=1

Case 2
ρ=2/3

Case 3
ρ=0

Fig. 1. Example showing three different cases for hypergraphs with different homophily ratios.

Initial Empirical Study on HyperSBM Model

1.2. **Hypergraph stochastic block models.** The *hypergraph stochastic block model*, first introduced in [26], is a generalization of the SBM for hypergraphs. We define the hypergraph stochastic block model (HSBM) as follows for d-uniform hypergraphs.

Definition 1.1 (Hypergraph)**.** A d-uniform hypergraph H is a pair $H = (V, E)$ where V is a set of vertices and $E \subset \binom{V}{d}$ is a set of subsets with size d of V, called hyperedges. when $d = 2$, it is the same as an ordinary graph.

Definition 1.2 (Hypergraph stochastic block model (HSBM))**.** Let $\mathcal{C} = \{C_1, \ldots C_k\}$ be a partition of the set $[n]$ into k sets of size $s = n/k$ (assume n is divisible by k), each $C_i, 1 \le i \le k$ is called a cluster. For constants $0 \le q < p < 1$, we define the d-uniform hypergraph SBM as follows:

For any set of d distinct vertices $i_1, \ldots i_d$, generate a hyperedge $\{i_1, \ldots i_d\}$ with probability p if the vertices $i_1, \ldots i_d$ are in the same cluster in $\mathcal{C}$. Otherwise, generate the hyperedge $\{i_1, \ldots i_d\}$ with probability q. We denote this distribution of random hypergraphs as $H(n, d, \mathcal{C}, p, q)$. When $d = 2$, it is the same as the stochastic block models for random graphs.

Ghoshdastidar, D., & Dukkipati, A. (2014). Consistency of spectral partitioning of uniform hypergraphs under planted partition model. *NIPS*.
Cole, S., & Zhu, Y. (2020). Exact recovery in the hypergraph stochastic block model: A spectral algorithm. *Linear Algebra and its Applications*, *593*, 45-73.

Initial Empirical Study on HyperSBM Model

HSBM: n, d, k, p, q = 100, 3, 2, [, 0.01, 0.6, 0.9, 0.99], [, 0.99, 0.5, 0.2, 0.01]				
hom. ratio	**0.3354**	**0.5182**	**0.7262**	**0.9795**
HGNN [1]	53.33%	56.00%	100%	100%
UniGCN [2]	44.44%	65.28%	100%	100%
UniSAGE [2]	51.39%	79.17%	100%	100%
UniGCNII [2]	48.61%	62.50%	100%	100%

'bad' results due to low Homophily ratio

'perfect' results due to high Homophily ratio

[1] Feng, Y., You, H., Zhang, Z., Ji, R., & Gao, Y. (2019). Hypergraph neural networks. AAAI, pp. 3558-3565.

[2] Huang, J., & Yang, J. (2021). UniGNN: a unified framework for graph and hypergraph neural networks. IJCAI, pp. 2563-2569.

Further Empirical Study on Benchmark Hypergraphs

Motivation: Researchers have used frequently these five benchmark datasets to evaluate newly-designed HGNNs, but paid less attention to their **Homophily Ratios.**

Statitics	DBLP (co-authorship)	Pubmed (co-citation)	Cora (co-authorship)	Cora (co-citation)	Citeseer (co-citation)
# hypernodes, $\lvert V \rvert$	43, 413	19, 717	2, 708	2, 708	3, 312
# hyperedges, $\lvert E \rvert$	22, 535	7, 963	1, 072	1, 579	1, 079
avg. hyperedge size	4.7 ± 6.1	4.3 ± 5.7	4.2 ± 4.1	3.0 ± 1.1	3.2 ± 2.0
# features, d	1, 425	500	1, 433	1, 433	3, 703
# classes, q	6	3	7	7	6
label rate, $\lvert V_L \rvert / \lvert V \rvert$	0.040	0.008	0.052	0.052	0.042
hom. ratio	**0.8656**	**0.7765**	**0.7797**	**0.7462**	**0.6814**

Our finding: these five datasets are all Homophily Hypergraphs, rather than Heterophily Hypergraphs

Further Empirical Study on Benchmark Hypergraphs

	Co-authorship Data		Co-citation Data		
Method	DBLP	Cora	Pubmed	Citeseer	Cora
MLP+HLR	63.6 ± 4.7	59.8 ± 4.7	64.7 ± 3.1	56.1 ± 2.6	61.0 ± 4.1
HGNN	69.2 ± 5.1	63.2 ± 3.1	66.8 ± 3.7	56.7 ± 3.8	**70.0 ± 2.9**
FastHyperGCN	68.1 ± 9.6	61.1 ± 8.2	65.7 ± 11.1	56.2 ± 8.1	61.3 ± 10.3
HyperGCN	70.9 ± 8.3	63.9 ± 7.3	68.3 ± 9.5	57.3 ± 7.3	62.5 ± 9.7
HyperSAGE	77.4 ± 3.8	72.4 ± 1.6	72.9 ± 1.3	61.8 ± 2.3	69.3 ± 2.7
UniGAT	**88.7 ± 0.2**	75.0 ± 1.1	**74.7 ± 1.2**	**63.8 ± 1.6**	69.2 ± 2.9
UniGCN	**88.8 ± 0.2**	**75.3 ± 1.2**	74.4 ± 1.0	63.6 ± 1.3	**70.1 ± 1.4**
UniGIN	88.6 ± 0.3	74.8 ± 1.3	74.4 ± 1.1	63.3 ± 1.2	69.2 ± 1.5
UniSAGE	88.5 ± 0.2	75.1 ± 1.2	74.3 ± 1.0	**63.8 ± 1.3**	**70.2 ± 1.5**

Table 1: Testing accuracy (%) of UniGNNs and other hypergraph models on co-authorship and co-citation datasets for *Semi-supervised Hypernode Classification*. The best or competitive results are highlighted for each dataset.

Screenshot from [2] Huang, J., & Yang, J. (2021). UniGNN: a unified framework for graph and hypergraph neural networks. IJCAI, pp. 2563-2569.

We built four new benchmark datasets: Hypergraphs with Heterophily

Datasets	homo. ratio, $\mathcal{H}$	MLP	HGNN	HyperGCN	UniSAGE	UniGAT	UniGCNII
Actor (co-occurence)	0.4675	**79.82 ± 4.14**	64.02 ± 0.03	52.64 ± 8.68	47.57 ± 3.60	48.58 ± 6.46	54.23 ± 8.23
Amazon-ratings(co-purchasing)	0.3549	**24.16 ± 0.01**	17.66 ± 0.01	18.16 ± 3.03	21.86 ± 1.83	21.91 ± 0.46	24.11 ± 1.59
Pokec(co-friendship)	0.4529	**58.00 ± 0.55**	49.16 ± 0.08	52.27 ± 1.77	53.20 ± 0.74	49.13 ± 0.25	52.87 ± 0.44
Twitch-gamers(co-create)	0.4857	**53.67 ± 1.27**	50.00 ± 0.00	51.07 ± 0.83	51.96 ± 0.27	51.19 ± 0.30	51.61 ± 0.25

Subsequent Works….

1. Generate and build **Benchmark Heterophily Hypergraphs;(Five datasets are ready!)**
2. Develop advanced **Hypergraph neural networks (HGNN) with Heterophily**. Then, **evaluate and compare its performance on** the generated benchmark **Heterophily Hypergraphs (I think, all the existing HGNNs may fail in these Benchmark Heterophily Hypergraphs);**
3. (If possible) Theoretical studies on why and how the **Homophily Ratio** affect the ability of HGNN?