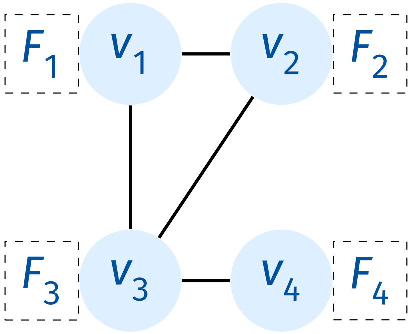
CISPA
HELMHOLTZ CENTER FOR
INFORMATION SECURITY

GNNs Getting ComFy: Community and Feature Similarity Guided Rewiring

Celia Rubio-Madrigal*, Adarsh Jamadandi*, Rebekka Burkholz
International Conference on Learning Representations — ICLR 2025

Graph learning

We consider a graph $G = (V, E)$ with a set of nodes V and a set of edges E, also equipped with some data (*features*) F_v on each node $v \in V$.

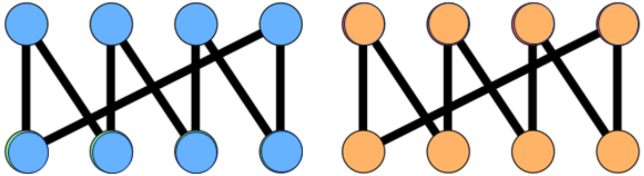


- **Node prediction:** predict a label y_v per node $v \in V$.
- **Transductive** — Train: partially labelled graph. Test: partially unlabelled same graph.

$\Rightarrow G = (V_{train} \cup V_{test}, E)$ where only nodes $v \in V_{train}$ are labelled with a known y_v.

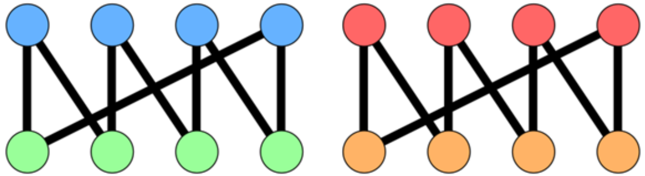
Homophily vs. Heterophily

Homophilic graphs:
most neighbors have **same label**

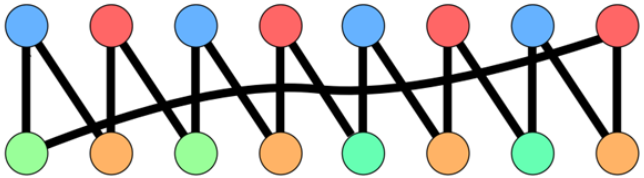


Platonov et al., "Characterizing Graph Datasets for Node Classification: Homophily–Heterophily Dichotomy and Beyond" (2023)

Heterophilic graphs:
many neighbors have **different labels**

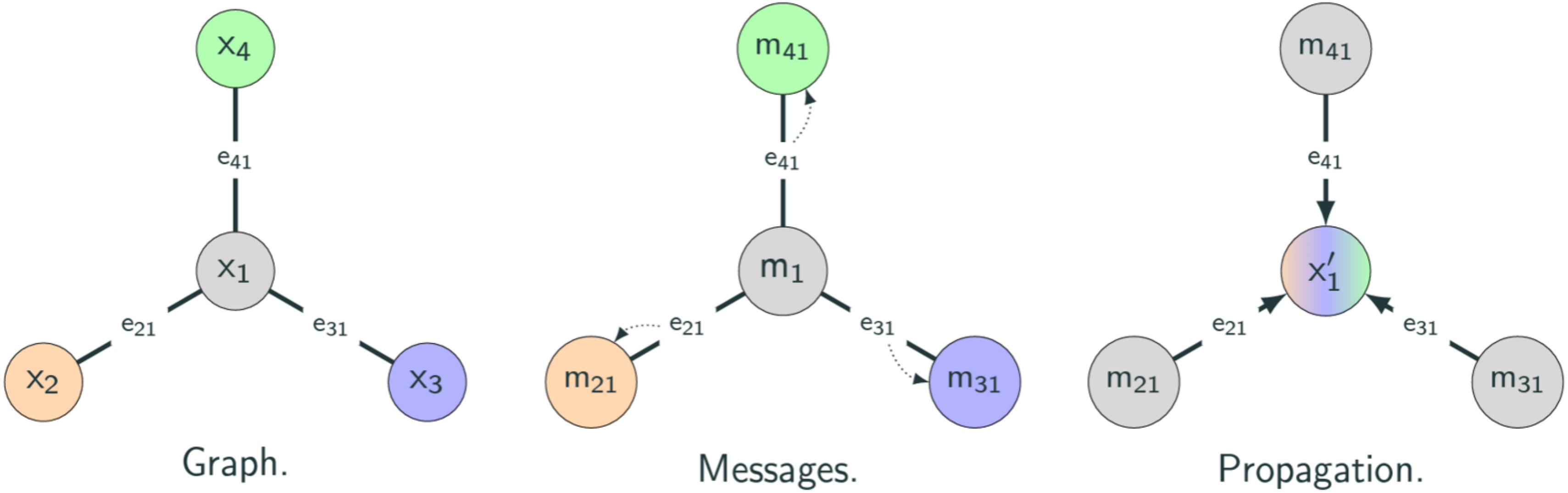


(a) Informative neighbors



(b) Less informative neighbors

Message passing



D. Grattarola, “Graph Deep Learning” (2021)

Challenges of message passing

The input graph has 2 roles:

- provides **relational information**
- serves as the **computational structure**

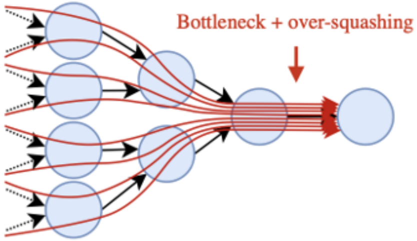
HOWEVER, not all graphs propagate information easily!

$\Rightarrow$ Limitations related to information flow, e.g. **over-squashing**.

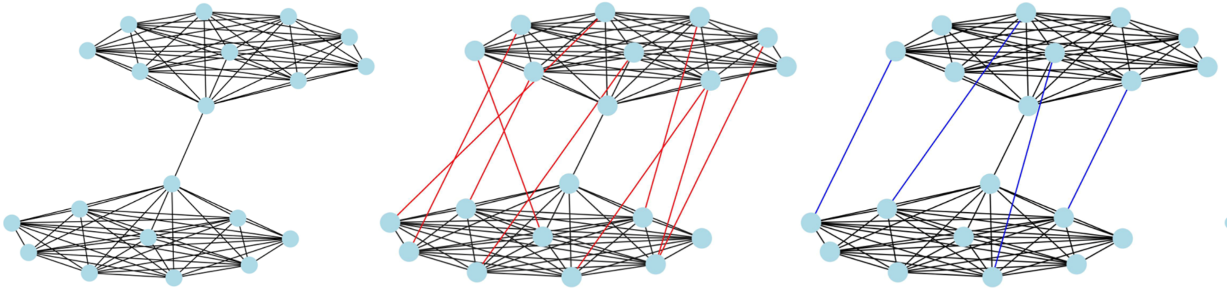
Over-squashing

Bottlenecks obstruct the flow of information.

- Measure: low **spectral gap** $\lambda_1 - \lambda_0$ $(= \lambda_1)$ of the normalized Laplacian: $L_G = I - D^{-1/2}AD^{-1/2}$.



Alon et al., "On the Bottleneck of Graph Neural Networks and its Practical Implications" (2021)



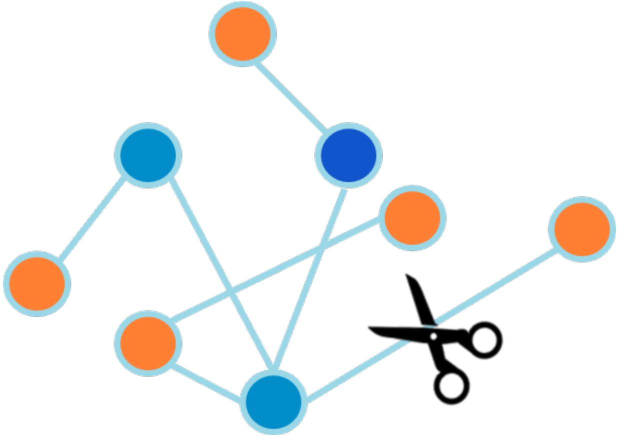
Mitigation of over-squashing

Common mitigation: Maximizing λ_1 by *rewiring* the input graph.

Spectral Graph Pruning Against Over-Squashing and Over-Smoothing

NeurIPS 2024 — Adarsh Jamadandi*, Celia Rubio-Madrigal*, and Rebekka Burkholz

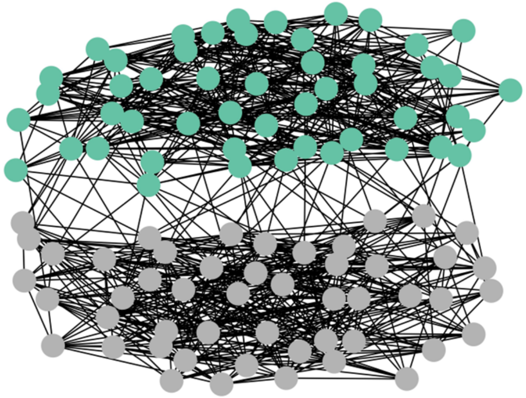
$\Rightarrow$ Deleting edges to maximize λ_1 also improves over-smoothing.



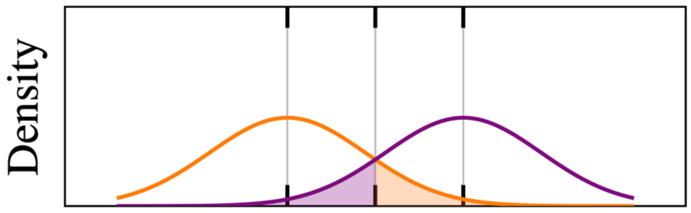
Is spectral gap maximization always beneficial?

Not without considering node features and labels.

- Let's see this with a simple test bed: *Planted partition model*



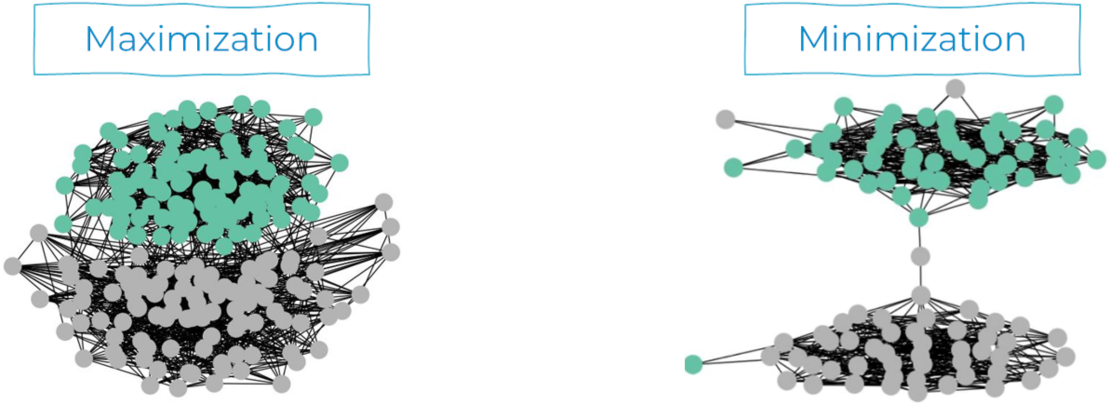
Stochastic block model (SBM)



Feature distributions

The effect of the spectral gap on community structure

- Spectral gap *maximization* **reduces** the strength of the community structure.
- Spectral gap *minimization* **increases** the structure.



- If labels = communities (**aligned** task-graph), maximizing is BAD.

Theory for SBMs

- Theorem 1 (previous slide): A **less pronounced** community structure corresponds to a **higher spectral gap**. In fact, λ_1 grows approximately like $-(p - q)/(q + p)$.

Theory for SBMs

- Theorem 1 (previous slide): A **less pronounced** community structure corresponds to a **higher spectral gap**. In fact, λ_1 grows approximately like $-(p-q)/(q+p)$.
- Theorem 2: A **less pronounced** community structure **harms performance** —if high task-graph alignment.

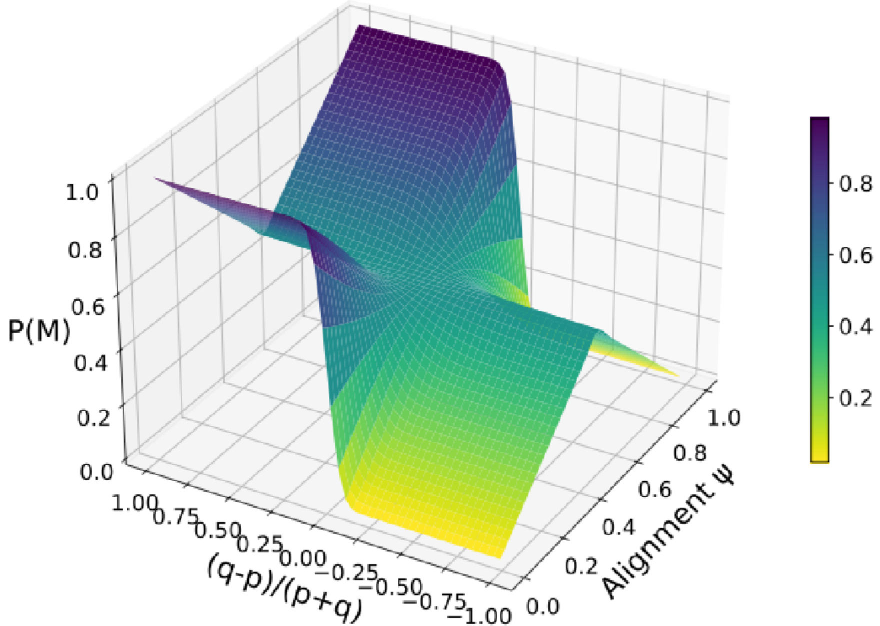
Theory for SBMs

- Theorem 1 (previous slide): A **less pronounced** community structure corresponds to a **higher spectral gap**. In fact, λ_1 grows approximately like $-(p-q)/(q+p)$.
- Theorem 2: A **less pronounced** community structure **harms performance** —if high task-graph alignment.
- Theorem 3: Different alignment levels affect error more.

$$P(M) \approx 1 - \psi + (2\psi - 1)\Phi\left(\frac{\frac{N}{2}(2\psi - 1)(p - q)}{\sqrt{\frac{N}{2}(p + q + p(1 - p) + q(1 - q) + 2(p - q)^2\psi(1 - \psi))}}\right)$$

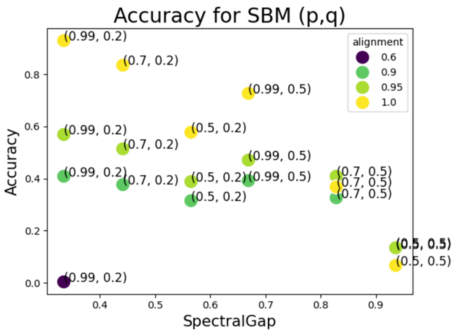
Theoretical insights

Intuition: High alignment + clear communities = low error

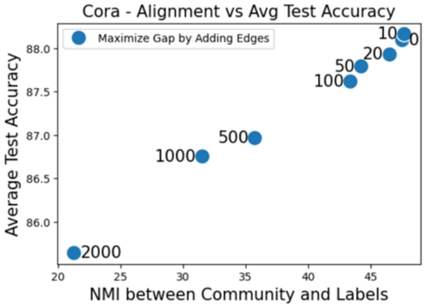


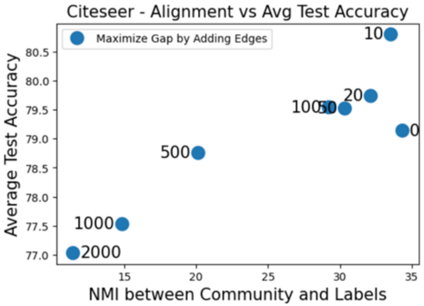
Experiments in support of the theory

- Alignment is crucial for GCN performance
  $\Rightarrow$ The number of same-labelled edges changes, indirectly reducing homophily.



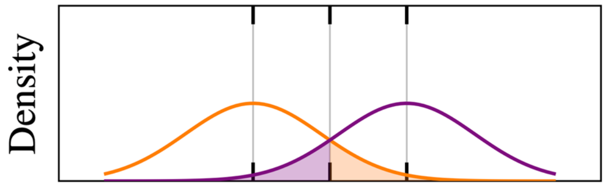
- Maximizing the spectral gap on homophilic datasets harms performance.



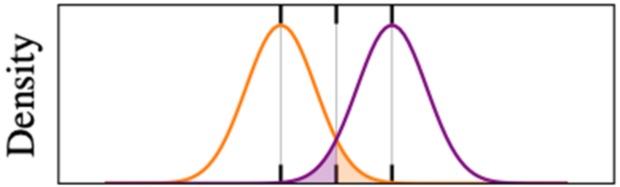


Alignment's mechanism: Feature denoising

If nodes with similar features are connected
⇒ there is a **variance reduction** of the feature distribution.

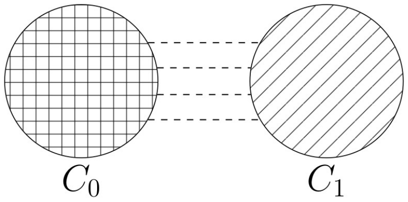


⇒ Aggregation ⇒



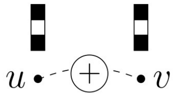
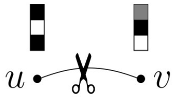
Rewire based on feature similarity

2-cluster graph:



Sizes A, B, and C of the three edge areas:

$A := |C_0| \cdot |C_0|$ $\quad$ $B := |C_0| \cdot |C_1|$ $\quad$ $C := |C_1| \cdot |C_1|$

	$\texttt{ComMa}(k)$		$\texttt{FeaSt}(k)$		$\texttt{ComFy}(k)$	
	`LowerComMa`	`HigherComMa`	`Add`	`Del`	`Add`	`Del`
$C_0 \times C_0$	If `Del`, draw and delete $\left\lfloor \frac{A}{A+C} \cdot k \right\rfloor$ edges	If `Add`, draw and add $\left\lfloor \frac{A}{A+C} \cdot k \right\rfloor$ edges	Add top k of $\mathrm{sim}(u,v)$ for $(u,v) \in \bar{\mathcal{E}}$ 	Delete bottom k of $\mathrm{sim}(u,v)$ for $(u,v) \in \mathcal{E}$ 	$\texttt{FeaSt Add}\left(\left\lfloor \frac{A}{A+B+C} \cdot k \right\rfloor\right)$	$\texttt{FeaSt Del}\left(\left\lfloor \frac{A}{A+B+C} \cdot k \right\rfloor\right)$
$C_0 \times C_1$	If `Add`, draw and add k edges	If `Del`, draw and delete k edges			$\texttt{FeaSt Add}\left(\left\lfloor \frac{B}{A+B+C} \cdot k \right\rfloor\right)$	$\texttt{FeaSt Del}\left(\left\lfloor \frac{B}{A+B+C} \cdot k \right\rfloor\right)$
$C_1 \times C_1$	If `Del`, draw and delete $\left\lfloor \frac{C}{A+C} \cdot k \right\rfloor$ edges	If `Add`, draw and add $\left\lfloor \frac{C}{A+C} \cdot k \right\rfloor$ edges			$\texttt{FeaSt Add}\left(\left\lfloor \frac{C}{A+B+C} \cdot k \right\rfloor\right)$	$\texttt{FeaSt Del}\left(\left\lfloor \frac{C}{A+B+C} \cdot k \right\rfloor\right)$

Comparison of methods

Table 1: Accuracy on node classification comparing different rewiring schemes.

Method	Cora	Citeseer	Pubmed	Cornell	Texas	Wisconsin	Chameleon	Squirrel	Actor
GCN	86.12±0.36	77.83±0.35	85.57±0.11	35.14±1.63	35.14±1.50	38.00±1.47	39.33±0.59	31.69±0.42	27.24±0.21
GCN+BORF	87.50±0.20	73.80±0.20	NA	50.80±1.10	NA	50.30±0.90	**61.50±0.40**	NA	NA
GCN+FoSR	83.50±0.39	75.47±0.31	86.08±0.10	40.54±1.47	51.35±1.75	54.00±1.46	41.01±0.63	32.36±0.37	27.57±0.21
GCN+ProxyAddMin	84.10±0.39	78.77±0.40	86.15±0.10	45.95±1.50	48.65±1.45	42.00±1.23	39.33±0.55	33.71±0.40	28.03±0.22
GCN+ProxyAddMax	85.92±0.43	79.25±0.35	86.41±0.11	48.65±1.41	40.54±1.64	50.00±1.25	38.20±0.70	35.06±0.44	25.99±0.20
GCN+ProxyDelMin	85.92±0.37	79.01±0.34	86.28±0.11	45.95±1.50	48.65±1.63	44.00±1.13	39.89±0.59	34.83±0.45	26.58±0.25
GCN+ProxyDelMax	86.32±0.38	**81.84±0.38**	85.95±0.11	54.05±1.67	48.65±1.35	52.00±1.33	39.33±0.70	34.61±0.39	27.30±0.22
GCN+HigherComMaAdd	83.64±0.38	77.13±0.38	85.86±0.10	49.93±1.34	52.66±1.47	50.55±1.24	41.23±0.72	34.51±0.40	30.92±0.21
GCN+HigherComMaDel	83.82±0.31	77.31±0.41	85.90±0.11	49.03±1.26	48.57±1.53	50.32±1.38	40.44±0.69	34.66±0.39	30.71±0.24
GCN+LowerComMaAdd	83.41±0.37	77.15±0.36	85.85±0.09	51.08±1.67	50.29±1.71	50.95±1.29	40.61±0.64	34.48±0.39	30.79±0.23
GCN+LowerComMaDel	83.61±0.35	77.39±0.37	85.90±0.10	49.69±1.43	50.59±1.52	50.61±1.35	40.43±0.71	34.76±0.40	30.79±0.22
GCN+FeaStAdd	87.73±0.39	78.54±0.34	86.43±0.09	59.46±1.49	54.05±1.51	60.00±1.09	43.26±0.62	**39.33±0.73**	31.25±0.22
GCN+FeaStDel	**90.74±0.39**	81.60±0.39	**86.76±0.10**	51.35±1.63	**64.86±1.43**	60.00±1.27	42.70±0.69	36.40±0.36	31.97±0.21
GCN+ComFyAdd	87.73±0.26	77.36±0.38	86.74±0.10	67.57±1.68	62.16±1.52	62.00±1.12	41.57±0.83	36.85±0.38	32.30±0.25
GCN+ComFyDel	88.13±0.27	78.07±0.35	86.23±0.11	**70.27±1.50**	**64.86±1.51**	**66.00±1.34**	45.51±0.76	39.10±0.43	31.12±0.19

Spectral

Comparison of methods

Table 1: Accuracy on node classification comparing different rewiring schemes.

Method	Cora	Citeseer	Pubmed	Cornell	Texas	Wisconsin	Chameleon	Squirrel	Actor
GCN	86.12±0.36	77.83±0.35	85.57±0.11	35.14±1.63	35.14±1.50	38.00±1.47	39.33±0.59	31.69±0.42	27.24±0.21
GCN+BORF	87.50±0.20	73.80±0.20	NA	50.80±1.10	NA	50.30±0.90	**61.50±0.40**	NA	NA
GCN+FoSR	83.50±0.39	75.47±0.31	86.08±0.10	40.54±1.47	51.35±1.75	54.00±1.46	41.01±0.63	32.36±0.37	27.57±0.21
GCN+ProxyAddMin	84.10±0.39	78.77±0.40	86.15±0.10	45.95±1.50	48.65±1.45	42.00±1.23	39.33±0.55	33.71±0.40	28.03±0.22
GCN+ProxyAddMax	85.92±0.43	79.25±0.35	86.41±0.11	48.65±1.41	40.54±1.64	50.00±1.25	38.20±0.70	35.06±0.44	25.99±0.20
GCN+ProxyDelMin	85.92±0.37	79.01±0.34	86.28±0.11	45.95±1.50	48.65±1.63	44.00±1.13	39.89±0.59	34.83±0.45	26.58±0.25
GCN+ProxyDelMax	86.32±0.38	**81.84±0.38**	85.95±0.11	54.05±1.67	48.65±1.35	52.00±1.33	39.33±0.70	34.61±0.39	27.30±0.22
GCN+HigherComMaAdd	83.64±0.38	77.13±0.38	85.86±0.10	49.93±1.34	52.66±1.47	50.55±1.24	41.23±0.72	34.51±0.40	30.92±0.21
GCN+HigherComMaDel	83.82±0.31	77.31±0.41	85.90±0.11	49.03±1.26	48.57±1.53	50.32±1.38	40.44+0.69	34.66±0.39	30.71±0.24
GCN+LowerComMaAdd	83.41±0.37	77.15±0.36	85.85±0.09	51.08±1.67	50.29±1.71	50.95±1.29	40.61±0.64	34.48±0.39	30.79±0.23
GCN+LowerComMaDel	83.61±0.35	77.39±0.37	85.90±0.10	49.69±1.43	50.59±1.52	50.61±1.35	40.43±0.71	34.76±0.40	30.79±0.22
GCN+FeaStAdd	87.73±0.39	78.54±0.34	86.43±0.09	59.46±1.49	54.05±1.51	60.00±1.09	43.26±0.62	**39.33±0.73**	31.25±0.22
GCN+FeaStDel	**90.74±0.39**	81.60±0.39	**86.76±0.10**	51.35±1.63	**64.86±1.43**	60.00±1.27	42.70±0.69	36.40±0.36	31.97±0.21
GCN+ComFyAdd	87.73±0.26	77.36±0.38	86.74±0.10	67.57±1.68	62.16±1.52	62.00±1.12	41.57±0.83	36.85±0.38	32.30±0.25
GCN+ComFyDel	88.13±0.27	78.07±0.35	86.23±0.11	**70.27±1.50**	64.86±1.51	**66.00±1.34**	45.51±0.76	39.10±0.43	31.12±0.19

feature similarity

- Feature similarity max. even boosts performance for homophilic graphs.

Comparison of methods

Table 1: Accuracy on node classification comparing different rewiring schemes.

Method	Cora	Citeseer	Pubmed	Cornell	Texas	Wisconsin	Chameleon	Squirrel	Actor
GCN	86.12±0.36	77.83±0.35	85.57±0.11	35.14±1.63	35.14±1.50	38.00±1.47	39.33±0.59	31.69±0.42	27.24±0.21
GCN+BORF	87.50±0.20	73.80±0.20	NA	50.80±1.10	NA	50.30±0.90	**61.50±0.40**	NA	NA
GCN+FoSR	83.50±0.39	75.47±0.31	86.08±0.10	40.54±1.47	51.35±1.75	54.00±1.46	41.01±0.63	32.36±0.37	27.57±0.21
GCN+ProxyAddMin	84.10±0.39	78.77±0.40	86.15±0.10	45.95±1.50	48.65±1.45	42.00±1.23	39.33±0.55	33.71±0.40	28.03±0.22
GCN+ProxyAddMax	85.92±0.43	79.25±0.35	86.41±0.11	48.65±1.41	40.54±1.64	50.00±1.25	38.20±0.70	35.06±0.44	25.99±0.20
GCN+ProxyDelMin	85.92±0.37	79.01±0.34	86.28±0.11	45.95±1.50	48.65±1.63	44.00±1.13	39.89±0.59	34.83±0.45	26.58±0.25
GCN+ProxyDelMax	86.32±0.38	**81.84±0.38**	85.95±0.11	54.05±1.67	48.65±1.35	52.00±1.33	39.33±0.70	34.61±0.39	27.30±0.22
GCN+HigherComMaAdd	83.64±0.38	77.13±0.38	85.86±0.10	49.93±1.34	52.66±1.47	50.55±1.24	41.23±0.72	34.51±0.40	30.92±0.21
GCN+HigherComMaDel	83.82±0.31	77.31±0.41	85.90±0.11	49.03±1.26	48.57±1.53	50.32±1.38	40.44+0.69	34.66±0.39	30.71±0.24
GCN+LowerComMaAdd	83.41±0.37	77.15±0.36	85.85±0.09	51.08±1.67	50.29±1.71	50.95±1.29	40.61±0.64	34.48±0.39	30.79±0.23
GCN+LowerComMaDel	83.61±0.35	77.39±0.37	85.90±0.10	49.69±1.43	50.59±1.52	50.61±1.35	40.43±0.71	34.76±0.40	30.79±0.22
GCN+FeaStAdd	87.73±0.39	78.54±0.34	86.43±0.09	59.46±1.49	54.05±1.51	60.00±1.09	43.26±0.62	**39.33±0.73**	31.25±0.22
GCN+FeaStDel	**90.74±0.39**	81.60±0.39	**86.76±0.10**	51.35±1.63	**64.86±1.43**	60.00±1.27	42.70±0.69	36.40±0.36	31.97±0.21
GCN+ComFyAdd	87.73±0.26	77.36±0.38	86.74±0.10	67.57±1.68	62.16±1.52	62.00±1.12	41.57±0.83	36.85±0.38	32.30±0.25
GCN+ComFyDel	88.13±0.27	78.07±0.35	86.23±0.11	**70.27±1.50**	**64.86±1.51**	**66.00±1.34**	45.51±0.76	39.10±0.43	31.12±0.19

feature similarity + community

- Feature similarity max. even boosts performance for homophilic graphs.
- Community struct. + feature similarity tends to fare better for heterophily.

Comparison of methods

Table 1: Accuracy on node classification comparing different rewiring schemes.

Method	Cora	Citeseer	Pubmed	Cornell	Texas	Wisconsin	Chameleon	Squirrel	Actor
GCN	86.12±0.36	77.83±0.35	85.57±0.11	35.14±1.63	35.14±1.50	38.00±1.47	39.33±0.59	31.69±0.42	27.24±0.21
GCN+BORF	87.50±0.20	73.80±0.20	NA	50.80±1.10	NA	50.30±0.90	**61.50±0.40**	NA	NA
GCN+FoSR	83.50±0.39	75.47±0.31	86.08±0.10	40.54±1.47	51.35±1.75	54.00±1.46	41.01±0.63	32.36±0.37	27.57±0.21
GCN+ProxyAddMin	84.10±0.39	78.77±0.40	86.15±0.10	45.95±1.50	48.65±1.45	42.00±1.23	39.33±0.55	33.71±0.40	28.03±0.22
GCN+ProxyAddMax	85.92±0.43	79.25±0.35	86.41±0.11	48.65±1.41	40.54±1.64	50.00±1.25	38.20±0.70	35.06±0.44	25.99±0.20
GCN+ProxyDelMin	85.92±0.37	79.01±0.34	86.28±0.11	45.95±1.50	48.65±1.63	44.00±1.13	39.89±0.59	34.83±0.45	26.58±0.25
GCN+ProxyDelMax	86.32±0.38	**81.84±0.38**	85.95±0.11	54.05±1.67	48.65±1.35	52.00±1.33	39.33±0.70	34.61±0.39	27.30±0.22
GCN+HigherComMaAdd	83.64±0.38	77.13±0.38	85.86±0.10	49.93±1.34	52.66±1.47	50.55±1.24	41.23±0.72	34.51±0.40	30.92±0.21
GCN+HigherComMaDel	83.82±0.31	77.31±0.41	85.90±0.11	49.03±1.26	48.57±1.53	50.32±1.38	40.44±0.69	34.66±0.39	30.71±0.24
GCN+LowerComMaAdd	83.41±0.37	77.15±0.36	85.85±0.09	51.08±1.67	50.29±1.71	50.95±1.29	40.61±0.64	34.48±0.39	30.79±0.23
GCN+LowerComMaDel	83.61±0.35	77.39±0.37	85.90±0.10	49.69±1.43	50.59±1.52	50.61±1.35	40.43±0.71	34.76±0.40	30.79±0.22
GCN+FeaStAdd	87.73±0.39	78.54±0.34	86.43±0.09	59.46±1.49	54.05±1.51	60.00±1.09	43.26±0.62	**39.33±0.73**	31.25±0.22
GCN+FeaStDel	**90.74±0.39**	81.60±0.39	**86.76±0.10**	51.35±1.63	**64.86±1.43**	60.00±1.27	42.70±0.69	36.40±0.36	31.97±0.21
GCN+ComFyAdd	87.73±0.26	77.36±0.38	86.74±0.10	67.57±1.68	62.16±1.52	62.00±1.12	41.57±0.83	36.85±0.38	32.30±0.25
GCN+ComFyDel	88.13±0.27	78.07±0.35	86.23±0.11	**70.27±1.50**	**64.86±1.51**	**66.00±1.34**	45.51±0.76	39.10±0.43	31.12±0.19

- Feature similarity max. even boosts performance for homophilic graphs.
- Community struct. + feature similarity tends to fare better for heterophily.
- Edge deletions most effective.

Thank you!

Celia Rubio-Madrigal

Adarsh Jamadandi

Rebekka Burkholz