Adaptive Node Feature Selection for Graph Neural Networks

Madeline Navarro, Ali Azizpour, and Santiago Segarra

Department of Electrical and Computer Engineering

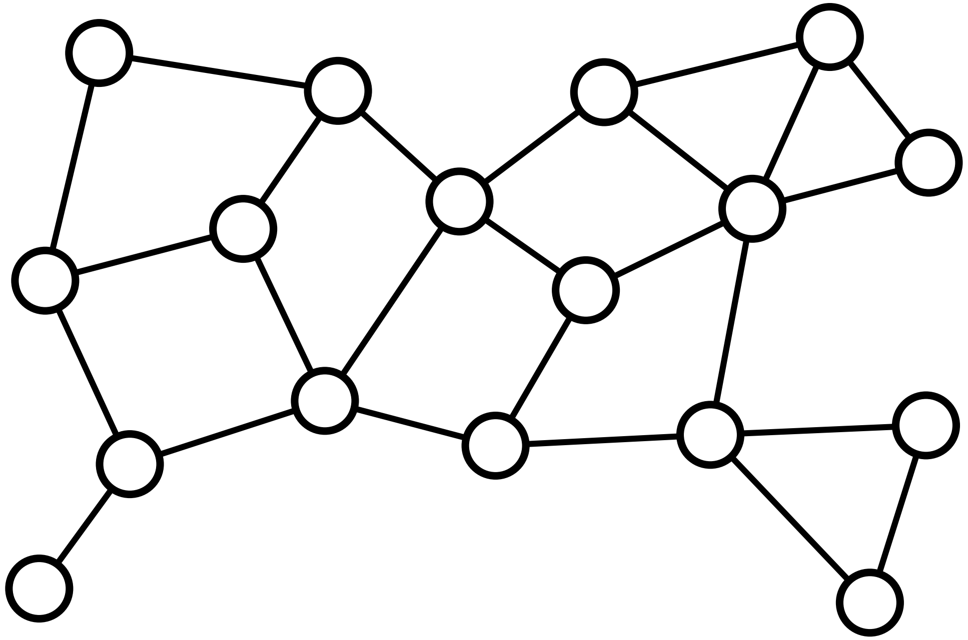
Rice University

9 June 2026

Graph Signal Processing Workshop 2026

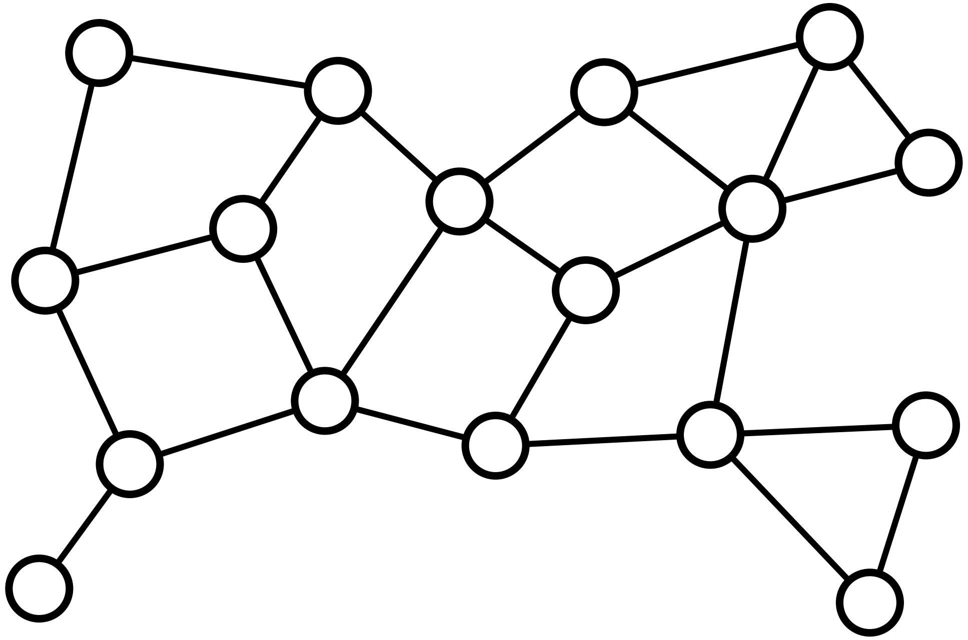
Contact:
Email: nav@rice.edu

Graph learning exploits connectivity for tasks involving complex data

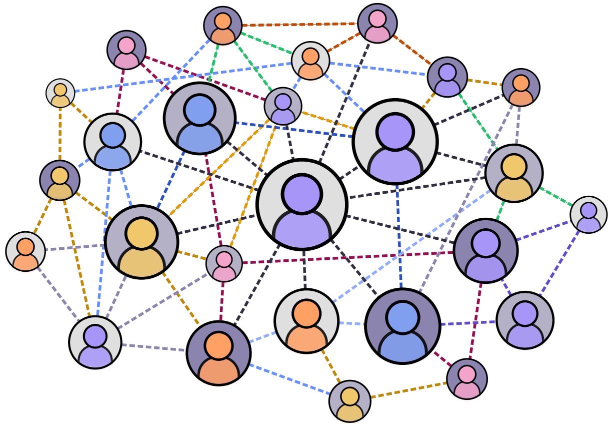


Graph $\mathcal{G}$

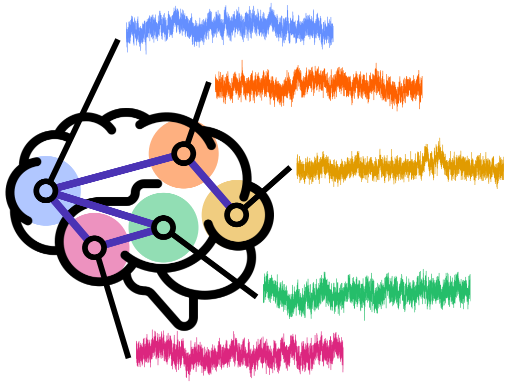
Graph learning exploits connectivity for tasks involving complex data



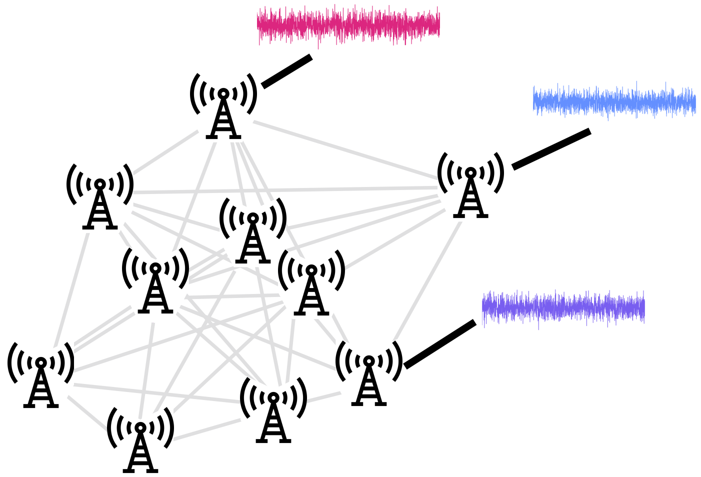
Graph $\mathcal{G}$



Social network analysis



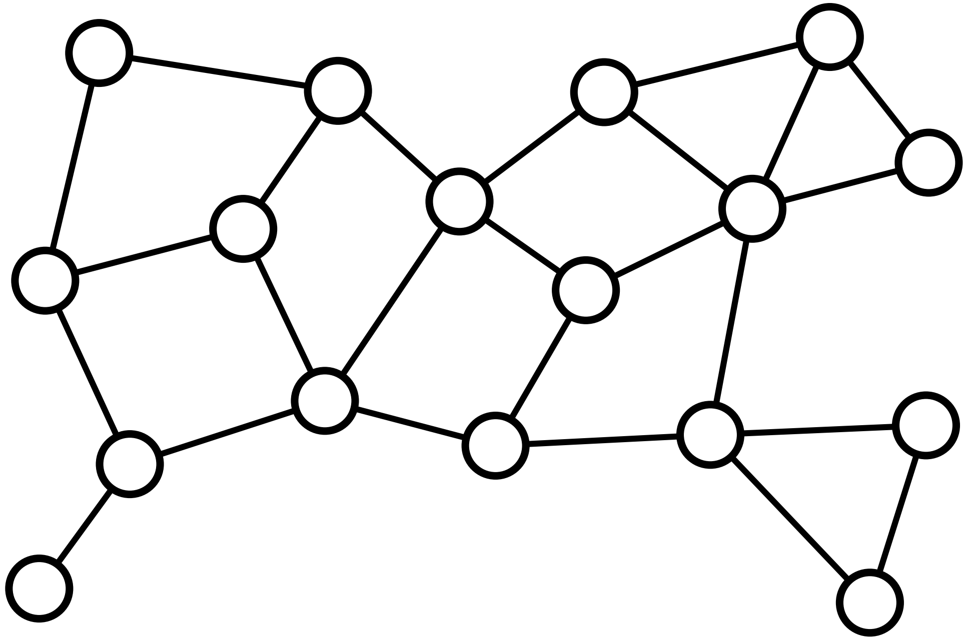
Brain connectivity



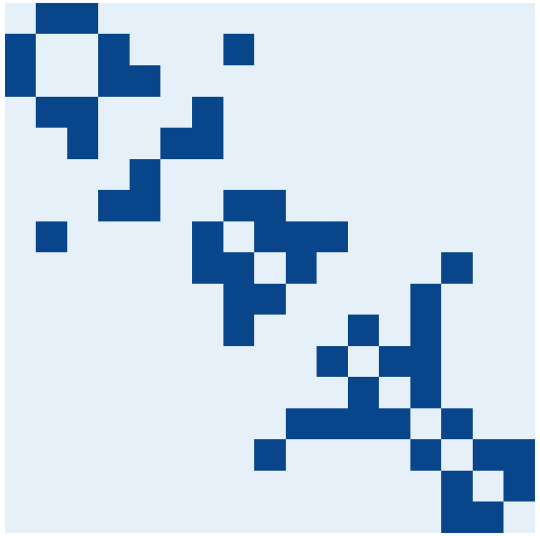
Wireless communications

Halberstam & Knight, “Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter”, J Public Econ 2016
Chu et al., “Function-specific and enhanced brain structural connectivity mapping via joint modeling of diffusion and functional MRI”, Sci Rep 2018
Nilforoshan et al., “Human mobility networks reveal increased segregation in large cities”, Nature 2023

Graph learning exploits connectivity for tasks involving complex data

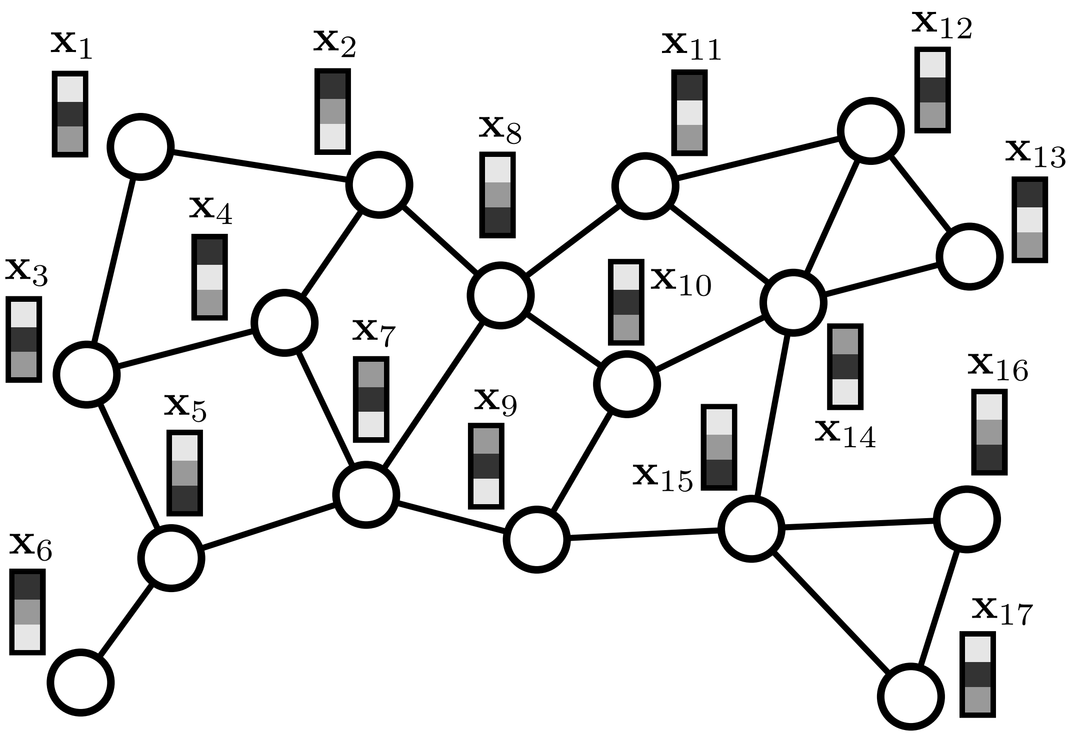


Graph $\mathcal{G}$



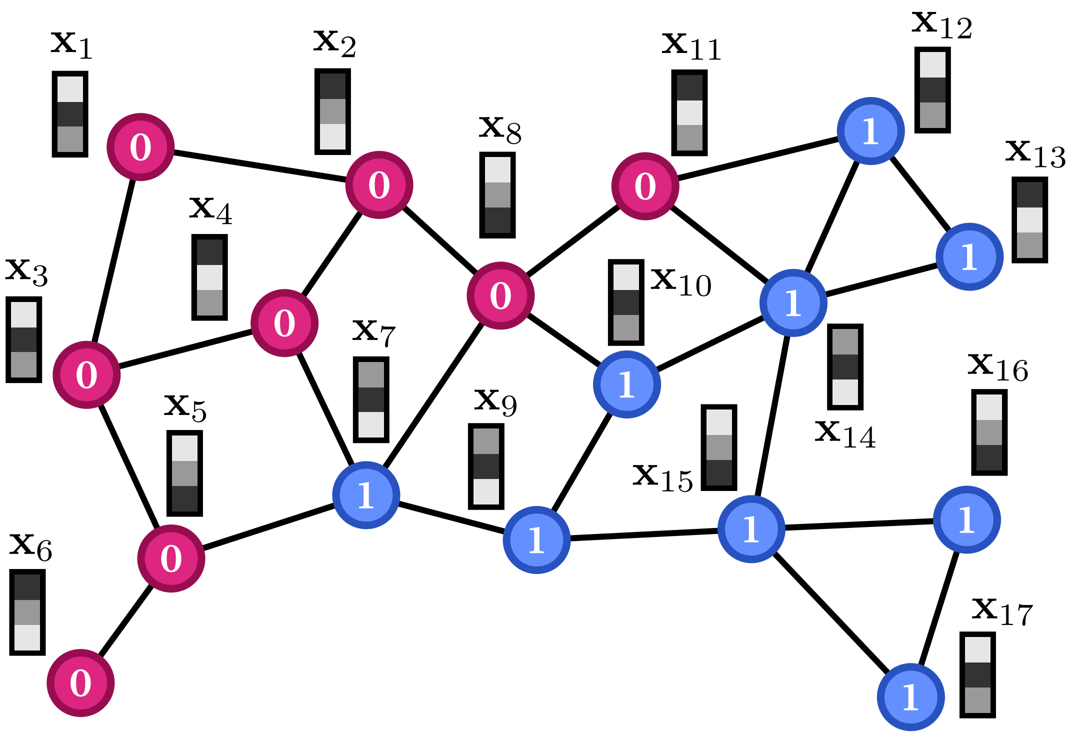
Adjacency matrix $\mathbf{A}$

Graph learning exploits connectivity for tasks involving complex data



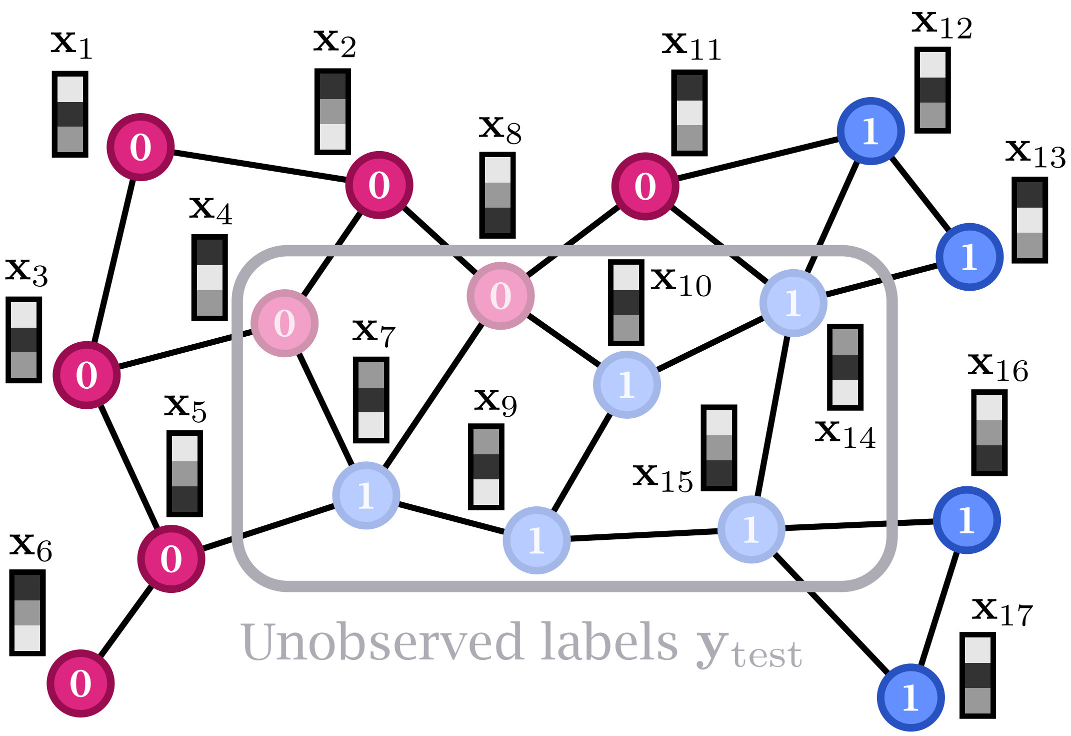
Graph $\mathcal{G}$

Graph learning exploits connectivity for tasks involving complex data



Graph $\mathcal{G}$ with
node features $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$
and node labels $\mathbf{y} \in \{0, 1\}^N$

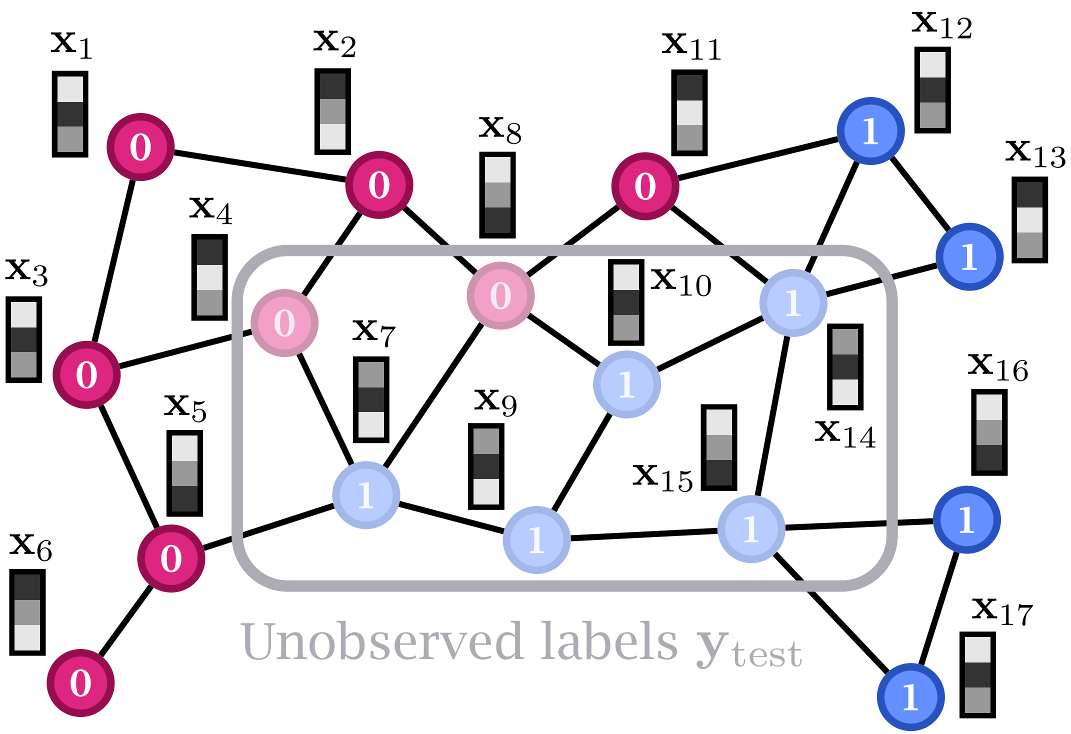
Graph learning exploits connectivity for tasks involving complex data



Graph $\mathcal{G}$ with
node features $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$
and observed labels $\mathbf{y}_{\text{train}}$

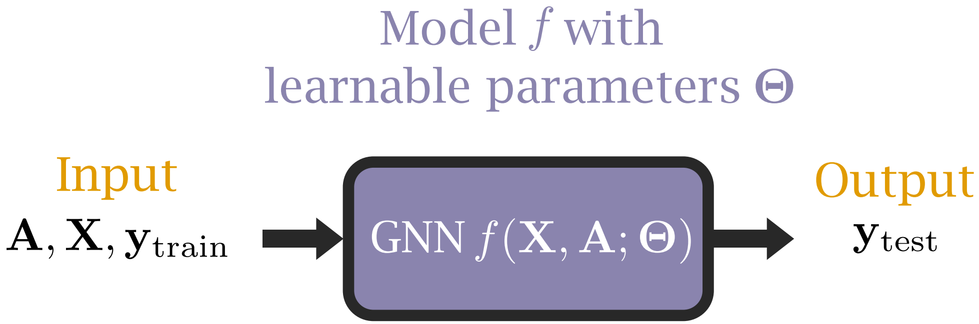
Goal: Predict unseen labels $\mathbf{y}_{\text{test}}$ from graph $\mathcal{G}$ and node features $\mathbf{X}$

Graph learning exploits connectivity for tasks involving complex data



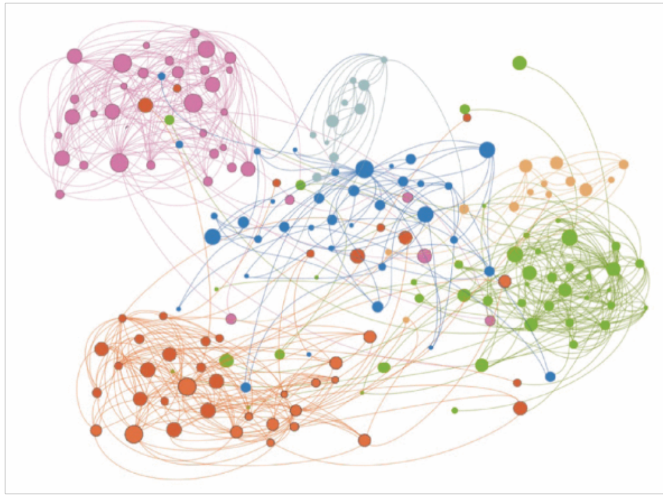
Graph $\mathcal{G}$ with
node features $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$
and observed labels $\mathbf{y}_{\text{train}}$

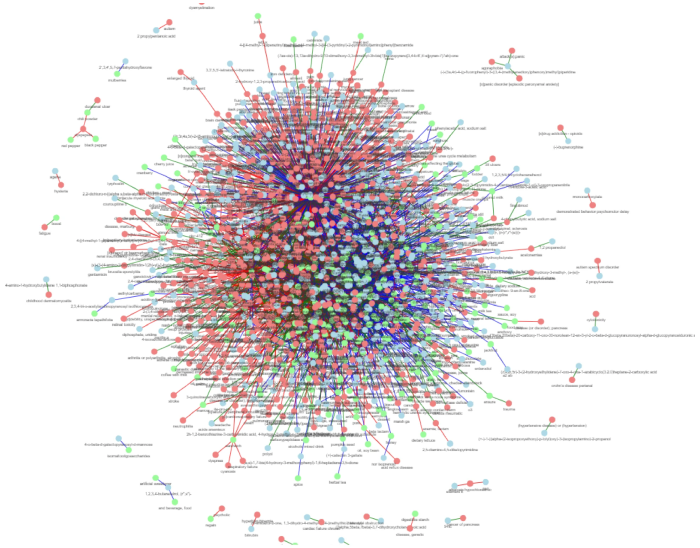
Goal: Predict unseen labels $\mathbf{y}_{\text{test}}$ from graph $\mathcal{G}$ and node features $\mathbf{X}$

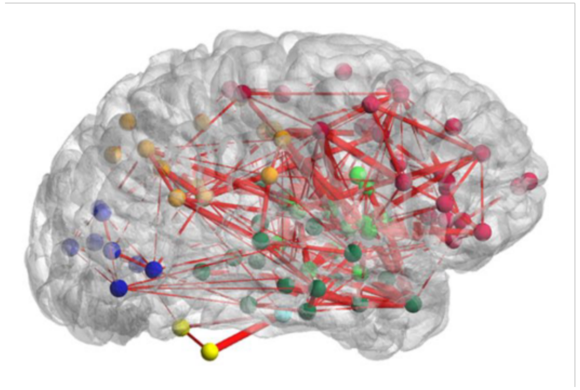


Approach: Optimize parameters $\boldsymbol{\Theta}$ of GNN f to predict labels from graph data

Real-world networks are growing increasingly large







Classify individuals from social network

Li et al., “Birds of a feather rumor together? Exploring homogeneity and conversation structure in social media for rumor detection”, IEEE Access 2020

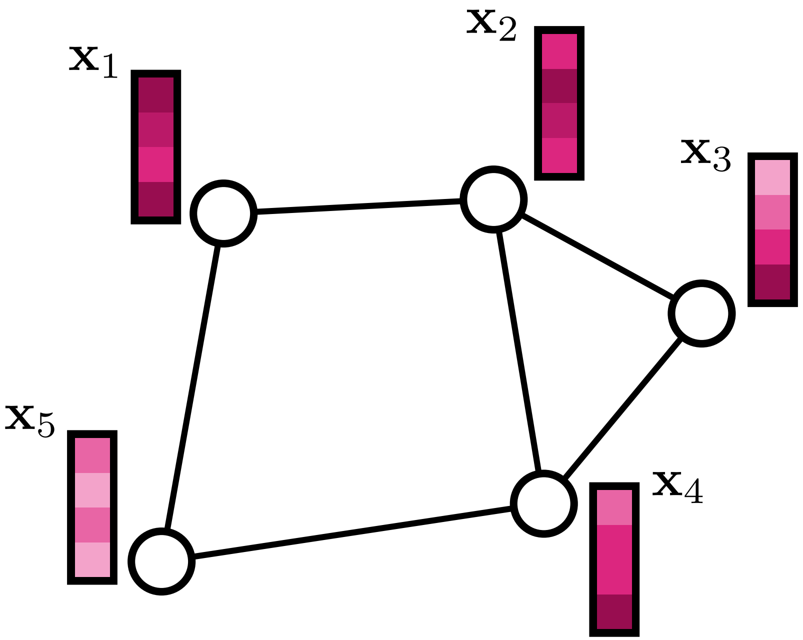
Assess nodes in heterogeneous or knowledge graphs

Cenikj et al., “From language models to large-scale food and biomedical knowledge graphs”, Sci Rep 2023

Identify regions of neural activity

Chu et al., “Function-specific and enhanced brain structural connectivity mapping via joint modeling of diffusion and functional MRI”, Sci Rep 2018

Feature selection reduces complexity by eliminating unnecessary variables for training



Publisher under fire after 'fake' citations found in AI ethics guide

A book published by Springer Nature includes dozens of questionable citations, including references to journals that do not exist

Harris, The Times 2025

JUNO NEWS

CRA billed taxpayers over $18 million for AI chatbot that spouted inaccurate info

The Canada Revenue Agency blew nearly $20 million of taxpayer money on an "AI" chatbot that repeatedly gave out the wrong information to Canadians trying to file their taxes.

Patrick, Juno News 2025

404

ICE

ICE's Facial Recognition App Misidentified a Woman. Twice

JOSEPH COX · JAN 19, 2026 AT 9:28 AM

Cox, 404 Media 2026

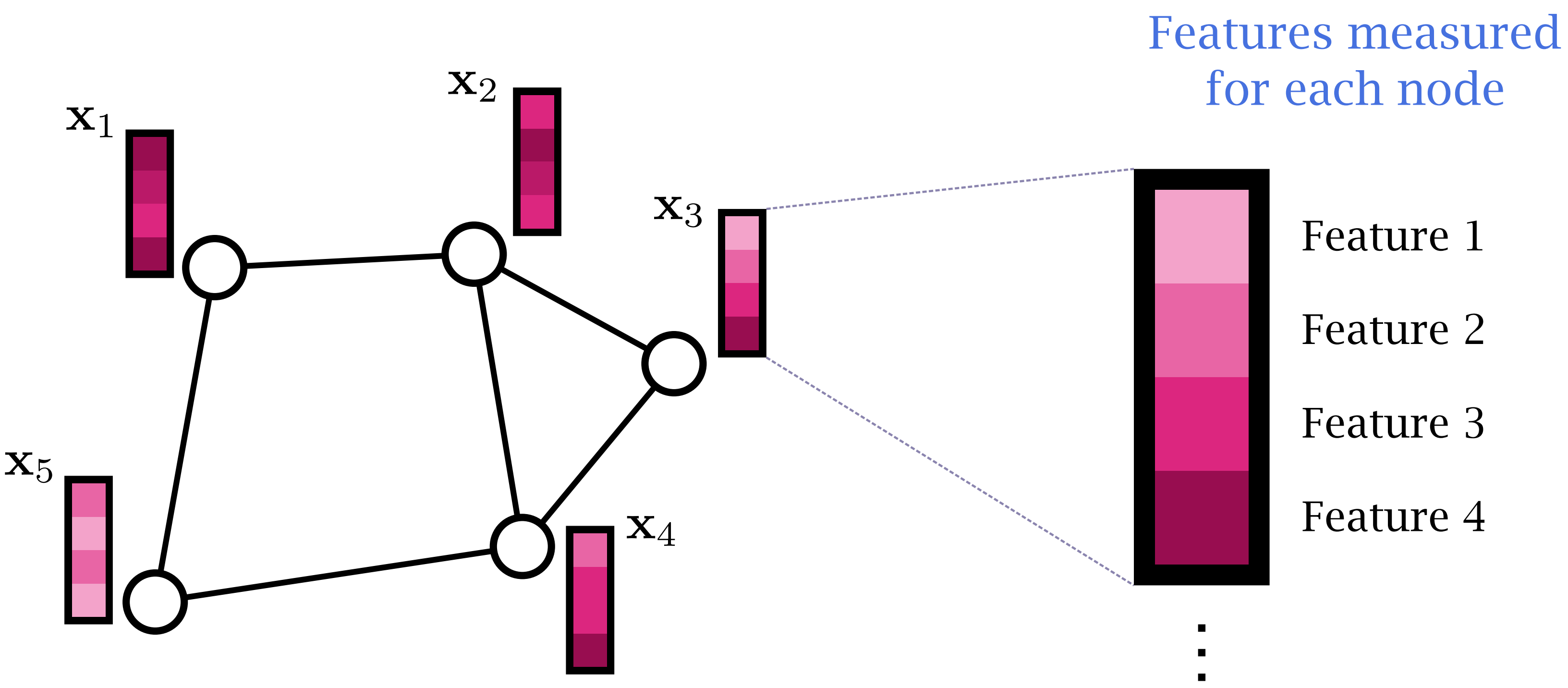
UK report reveals bias within medical tools and devices

Experts say action needed as report finds minority ethnic people, women and those from deprived backgrounds at risk of poorer healthcare

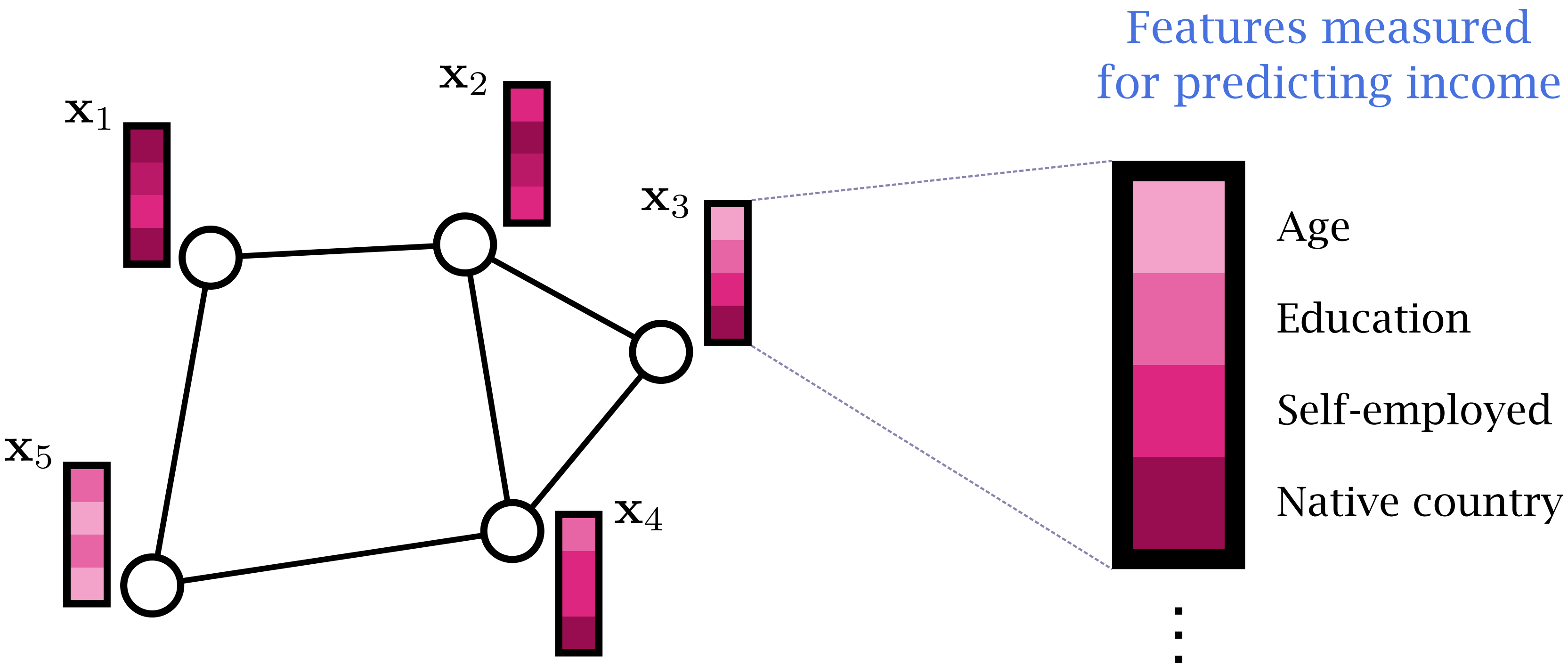
Davis, The Guardian 2024

As machine learning advances,
interpretability plummets

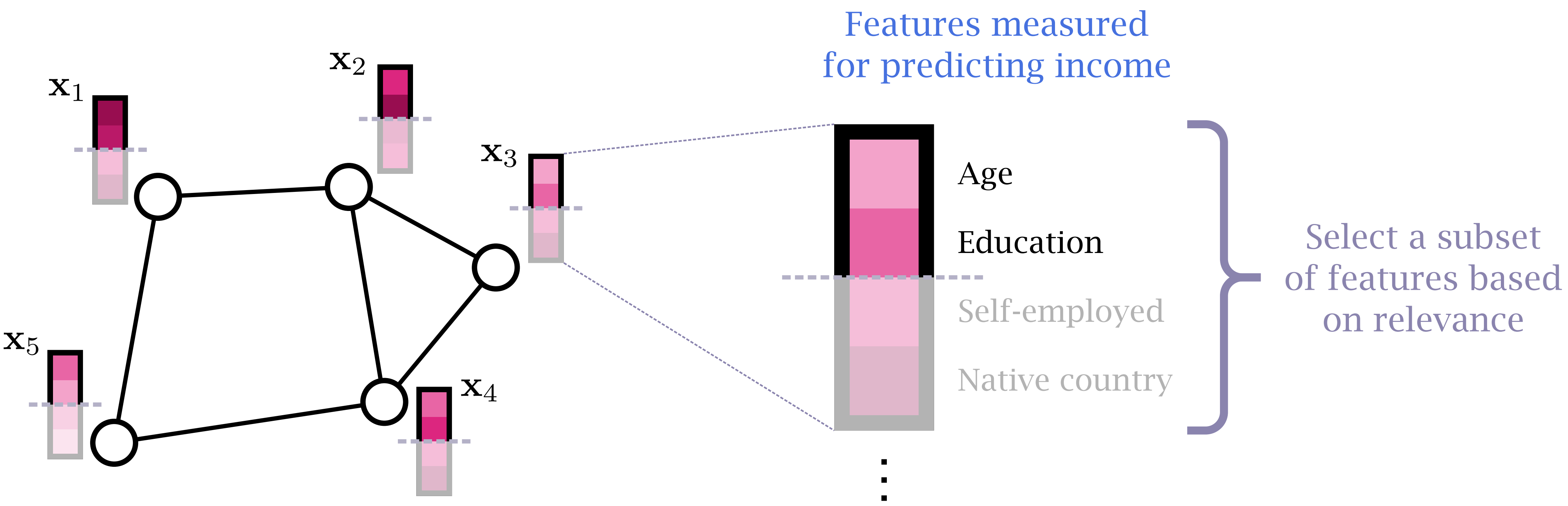
Feature selection reduces complexity by eliminating unnecessary variables for training



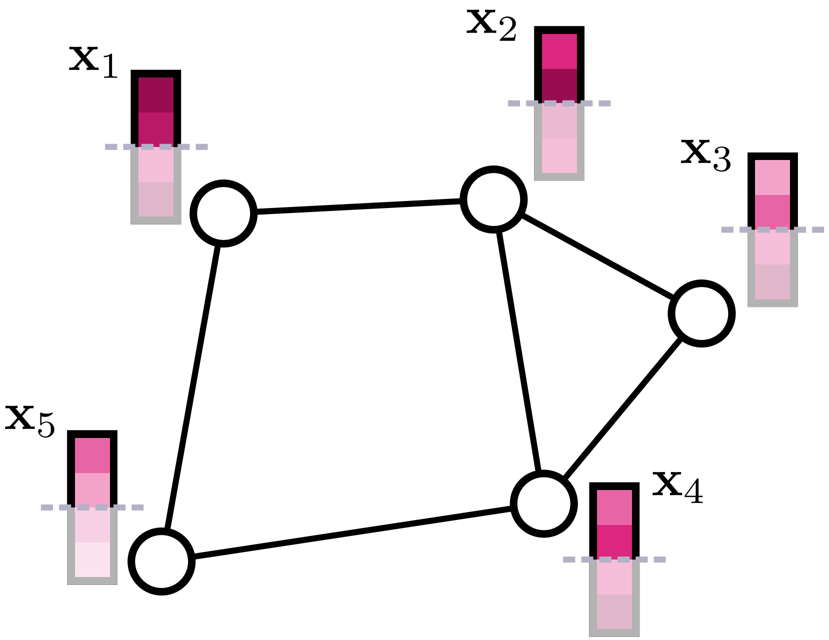
Feature selection reduces complexity by eliminating unnecessary variables for training



Feature selection reduces complexity by eliminating unnecessary variables for training



Feature selection reduces complexity by eliminating unnecessary variables for training



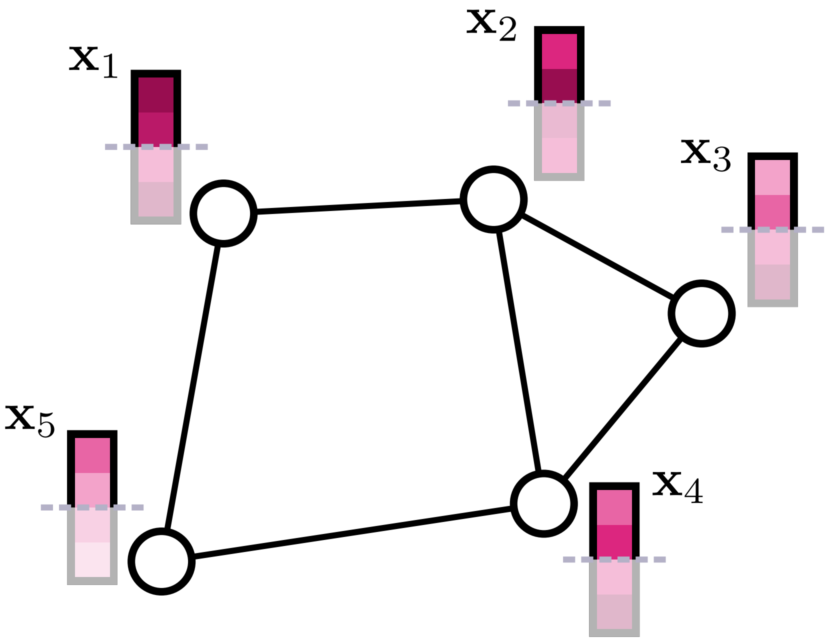
Interpretability: Identify which features exploited

Economy: Eliminate unnecessary or harmful variables

Rigor: Many well-established statistical methods

Li et al., "Feature selection: A data perspective", ACM Comput Surv 2017
Yang et al., "Feature selection for MLP neural network: The use of random permutation of probabilistic outputs", IEEE TNNLS 2009
Altmann et al., "Permutation importance: A corrected feature importance measure", Bioinformatics 2010

Feature selection reduces complexity by eliminating unnecessary variables for training



Interpretability: Identify which features exploited

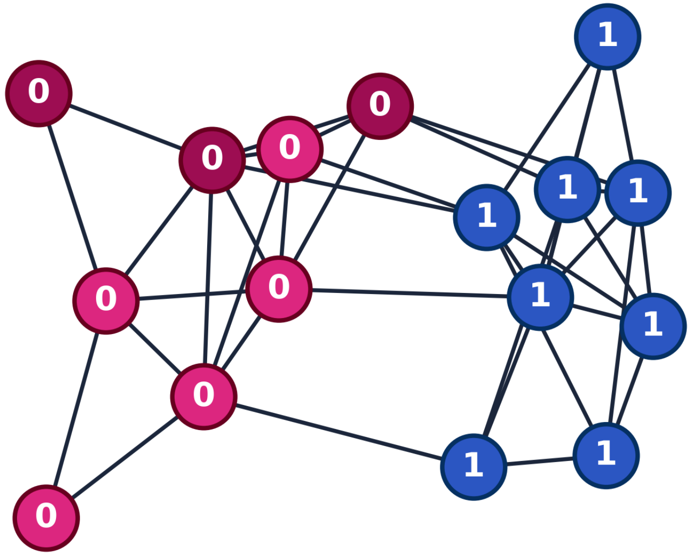
Economy: Eliminate unnecessary or harmful variables

Rigor: Many well-established statistical methods

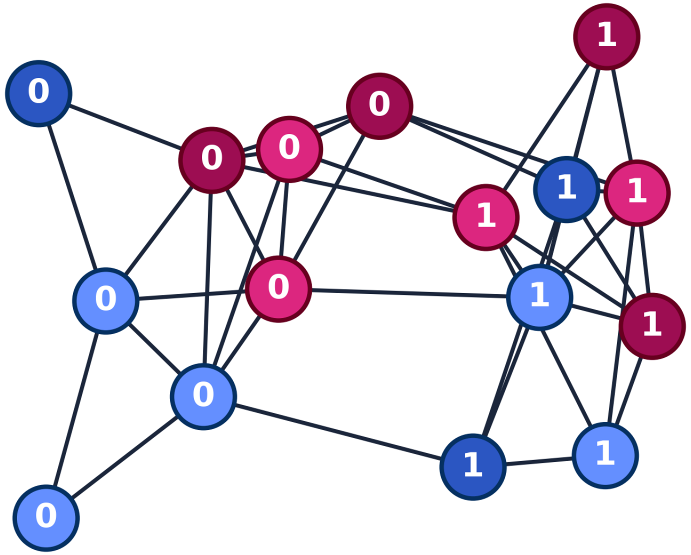
How is feature importance affected by underlying graph connections?

Li et al., "Feature selection: A data perspective", ACM Comput Surv 2017
Yang et al., "Feature selection for MLP neural network: The use of random permutation of probabilistic outputs", IEEE TNNLS 2009
Altmann et al., "Permutation importance: A corrected feature importance measure", Bioinformatics 2010

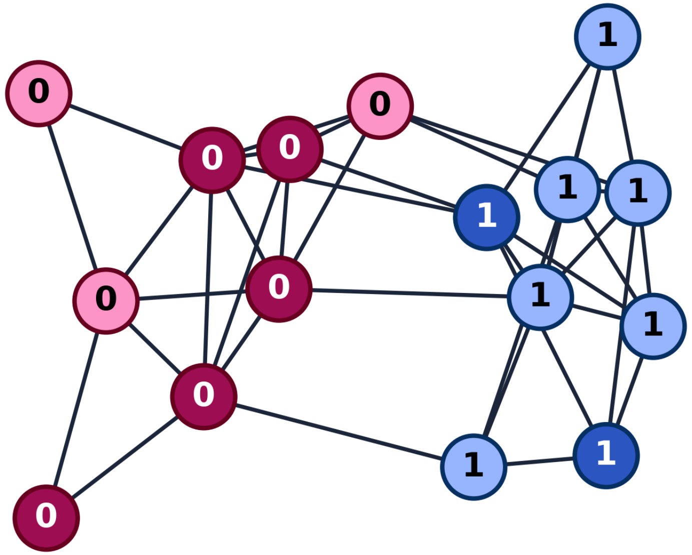
Node feature selection requires accounting for graph structure



Homophilic node features



Heterophilic node features



Homophilic node features with high within-class variance

How is feature importance affected by underlying graph connections?

nav@rice.edu

Navarro et al., "Adaptive node feature selection for graph neural networks", ICML 2026

Node classification on real-world data reveals feature and structure relevance

How does node classification performance change under edge and feature perturbations?

Setting	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
$\text{GNN}(\mathbf{X}, \mathbf{A}; \boldsymbol{\Theta})$	85.83 ±0.46	74.39 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
$\text{MLP}(\mathbf{X}; \boldsymbol{\Theta})$	74.21 ±1.40	70.02 ±1.39	88.65 ±0.41	88.73 ±0.73	81.63 ±0.75	78.92 ±5.51	82.70 ±2.16	83.60 ±6.62
$\text{GNN}(\mathbf{X}, \tilde{\mathbf{A}}; \boldsymbol{\Theta})$	36.97 ±1.80	35.13 ±2.41	67.07 ±0.80	30.52 ±5.77	37.64 ±0.45	67.57 ±7.83	70.27 ±7.05	78.40 ±6.37
$\text{GNN}(\mathbf{W}, \mathbf{A}; \boldsymbol{\Theta})$	82.92 ±1.54	67.37 ±1.61	76.23 ±0.53	89.92 ±0.58	85.49 ±0.44	49.19 ±7.91	55.68 ±2.76	47.20 ±7.00
$\text{GNN}(\tilde{\mathbf{X}}, \mathbf{A}; \boldsymbol{\Theta})$	76.53 ±1.12	63.34 ±1.44	47.42 ±3.25	66.76 ±8.58	51.09 ±8.84	43.78 ±8.95	52.97 ±6.07	46.80 ±6.76

Navarro et al., “Adaptive node feature selection for graph neural networks”, ICML 2026

Node classification on real-world data reveals feature and structure relevance

How does node classification performance change under edge and feature perturbations?

Setting	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
$\mathrm{GNN}(\mathbf{X}, \mathbf{A}; \mathbf{\Theta})$	85.83 ±0.46	74.39 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
$\mathrm{MLP}(\mathbf{X}; \mathbf{\Theta})$	74.21 ±1.40	70.02 ±1.39	88.65 ±0.41	88.73 ±0.73	81.63 ±0.75	78.92 ±5.51	82.70 ±2.16	83.60 ±6.62
$\mathrm{GNN}(\mathbf{X}, \tilde{\mathbf{A}}; \mathbf{\Theta})$	36.97 ±1.80	35.13 ±2.41	67.07 ±0.80	30.52 ±5.77	37.64 ±0.45	67.57 ±7.83	70.27 ±7.05	78.40 ±6.37
$\mathrm{GNN}(\mathbf{W}, \mathbf{A}; \mathbf{\Theta})$	82.92 ±1.54	67.37 ±1.61	76.23 ±0.53	89.92 ±0.58	85.49 ±0.44	49.19 ±7.91	55.68 ±2.76	47.20 ±7.00
$\mathrm{GNN}(\tilde{\mathbf{X}}, \mathbf{A}; \mathbf{\Theta})$	76.53 ±1.12	63.34 ±1.44	47.42 ±3.25	66.76 ±8.58	51.09 ±8.84	43.78 ±8.95	52.97 ±6.07	46.80 ±6.76

Citation networks
(trained with GCN)

Graphs with homophilic
labels and features

Node classification on real-world data reveals feature and structure relevance

How does node classification performance change under edge and feature perturbations?

Setting	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
$\mathrm{GNN}(\mathbf{X}, \mathbf{A}; \mathbf{\Theta})$	85.83 ±0.46	74.39 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
$\mathrm{MLP}(\mathbf{X}; \mathbf{\Theta})$	74.21 ±1.40	70.02 ±1.39	88.65 ±0.41	88.73 ±0.73	81.63 ±0.75	78.92 ±5.51	82.70 ±2.16	83.60 ±6.62
$\mathrm{GNN}(\mathbf{X}, \tilde{\mathbf{A}}; \mathbf{\Theta})$	36.97 ±1.80	35.13 ±2.41	67.07 ±0.80	30.52 ±5.77	37.64 ±0.45	67.57 ±7.83	70.27 ±7.05	78.40 ±6.37
$\mathrm{GNN}(\mathbf{W}, \mathbf{A}; \mathbf{\Theta})$	82.92 ±1.54	67.37 ±1.61	76.23 ±0.53	89.92 ±0.58	85.49 ±0.44	49.19 ±7.91	55.68 ±2.76	47.20 ±7.00
$\mathrm{GNN}(\tilde{\mathbf{X}}, \mathbf{A}; \mathbf{\Theta})$	76.53 ±1.12	63.34 ±1.44	47.42 ±3.25	66.76 ±8.58	51.09 ±8.84	43.78 ±8.95	52.97 ±6.07	46.80 ±6.76

Citation networks
(trained with GCN)
Graphs with homophilic labels and features

University social networks
(trained with TAGCN)
Graphs with heterophilic labels and features

Navarro et al., “Adaptive node feature selection for graph neural networks”, ICML 2026

Node classification on real-world data reveals feature and structure relevance

How does node classification performance change under edge and feature perturbations?

Setting	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
$\mathrm{GNN}(\mathbf{X}, \mathbf{A}; \boldsymbol{\Theta})$	85.83 ±0.46	74.39 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
$\mathrm{MLP}(\mathbf{X}; \boldsymbol{\Theta})$	74.21 ±1.40	70.02 ±1.39	88.65 ±0.41	88.73 ±0.73	81.63 ±0.75	78.92 ±5.51	82.70 ±2.16	83.60 ±6.62
$\mathrm{GNN}(\mathbf{X}, \tilde{\mathbf{A}}; \boldsymbol{\Theta})$	36.97 ±1.80	35.13 ±2.41	67.07 ±0.80	30.52 ±5.77	37.64 ±0.45	67.57 ±7.83	70.27 ±7.05	78.40 ±6.37
$\mathrm{GNN}(\mathbf{W}, \mathbf{A}; \boldsymbol{\Theta})$	82.92 ±1.54	67.37 ±1.61	76.23 ±0.53	89.92 ±0.58	85.49 ±0.44	49.19 ±7.91	55.68 ±2.76	47.20 ±7.00
$\mathrm{GNN}(\tilde{\mathbf{X}}, \mathbf{A}; \boldsymbol{\Theta})$	76.53 ±1.12	63.34 ±1.44	47.42 ±3.25	66.76 ±8.58	51.09 ±8.84	43.78 ±8.95	52.97 ±6.07	46.80 ±6.76

Citation networks
(trained with GCN)
Graphs with homophilic labels and features

Co-purchase networks
(trained with GIN)
Graphs with homophilic labels but some heterophilic features

University social networks
(trained with TAGCN)
Graphs with heterophilic labels and features

Navarro et al., "Adaptive node feature selection for graph neural networks", ICML 2026

Node classification on real-world data reveals feature and structure relevance

How does node classification performance change under edge and feature perturbations?

Setting	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
$\mathrm{GNN}(\mathbf{X}, \mathbf{A}; \mathbf{\Theta})$	85.83 ±0.46	74.39 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
$\mathrm{MLP}(\mathbf{X}; \mathbf{\Theta})$	74.21 ±1.40	70.02 ±1.39	88.65 ±0.41	88.73 ±0.73	81.63 ±0.75	78.92 ±5.51	82.70 ±2.16	83.60 ±6.62
$\mathrm{GNN}(\mathbf{X}, \tilde{\mathbf{A}}; \mathbf{\Theta})$	36.97 ±1.80	35.13 ±2.41	67.07 ±0.80	30.52 ±5.77	37.64 ±0.45	67.57 ±7.83	70.27 ±7.05	78.40 ±6.37
$\mathrm{GNN}(\mathbf{W}, \mathbf{A}; \mathbf{\Theta})$	82.92 ±1.54	67.37 ±1.61	76.23 ±0.53	89.92 ±0.58	85.49 ±0.44	49.19 ±7.91	55.68 ±2.76	47.20 ±7.00
$\mathrm{GNN}(\tilde{\mathbf{X}}, \mathbf{A}; \mathbf{\Theta})$	76.53 ±1.12	63.34 ±1.44	47.42 ±3.25	66.76 ±8.58	51.09 ±8.84	43.78 ±8.95	52.97 ±6.07	46.80 ±6.76

Citation networks
(trained with GCN)
Graphs with homophilic labels and features

Co-purchase networks
(trained with GIN)
Graphs with homophilic labels but some heterophilic features

University social networks
(trained with TAGCN)
Graphs with heterophilic labels and features

Navarro et al., “Adaptive node feature selection for graph neural networks”, ICML 2026

Node classification on real-world data reveals feature and structure relevance

How does node classification performance change under edge and feature perturbations?

	Setting	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
	$\mathrm{GNN}(\mathbf{X}, \mathbf{A}; \mathbf{\Theta})$	85.83 ±0.46	74.39 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
Perturb graph	$\mathrm{MLP}(\mathbf{X}; \mathbf{\Theta})$	74.21 ±1.40	70.02 ±1.39	88.65 ±0.41	88.73 ±0.73	81.63 ±0.75	78.92 ±5.51	82.70 ±2.16	83.60 ±6.62
Perturb graph	$\mathrm{GNN}(\mathbf{X}, \tilde{\mathbf{A}}; \mathbf{\Theta})$	36.97 ±1.80	35.13 ±2.41	67.07 ±0.80	30.52 ±5.77	37.64 ±0.45	67.57 ±7.83	70.27 ±7.05	78.40 ±6.37
	$\mathrm{GNN}(\mathbf{W}, \mathbf{A}; \mathbf{\Theta})$	82.92 ±1.54	67.37 ±1.61	76.23 ±0.53	89.92 ±0.58	85.49 ±0.44	49.19 ±7.91	55.68 ±2.76	47.20 ±7.00
	$\mathrm{GNN}(\tilde{\mathbf{X}}, \mathbf{A}; \mathbf{\Theta})$	76.53 ±1.12	63.34 ±1.44	47.42 ±3.25	66.76 ±8.58	51.09 ±8.84	43.78 ±8.95	52.97 ±6.07	46.80 ±6.76

Citation networks (Cora, CiteSeer, PubMed)
(trained with GCN)
Graphs with homophilic labels and features

Co-purchase networks (Photo, Computers)
(trained with GIN)
Graphs with homophilic labels but some heterophilic features

University social networks (Cornell, Texas, Wisconsin)
(trained with TAGCN)
Graphs with heterophilic labels and features

Navarro et al., “Adaptive node feature selection for graph neural networks”, ICML 2026

Node classification on real-world data reveals feature and structure relevance

How does node classification performance change under edge and feature perturbations?

	Setting	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
	$\mathrm{GNN}(\mathbf{X}, \mathbf{A}; \mathbf{\Theta})$	85.83 ±0.46	74.39 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
Perturb graph	$\mathrm{MLP}(\mathbf{X}; \mathbf{\Theta})$	74.21 ±1.40	70.02 ±1.39	88.65 ±0.41	88.73 ±0.73	81.63 ±0.75	78.92 ±5.51	82.70 ±2.16	83.60 ±6.62
	$\mathrm{GNN}(\mathbf{X}, \tilde{\mathbf{A}}; \mathbf{\Theta})$	36.97 ±1.80	35.13 ±2.41	67.07 ±0.80	30.52 ±5.77	37.64 ±0.45	67.57 ±7.83	70.27 ±7.05	78.40 ±6.37
Perturb features	$\mathrm{GNN}(\mathbf{W}, \mathbf{A}; \mathbf{\Theta})$	82.92 ±1.54	67.37 ±1.61	76.23 ±0.53	89.92 ±0.58	85.49 ±0.44	49.19 ±7.91	55.68 ±2.76	47.20 ±7.00
	$\mathrm{GNN}(\tilde{\mathbf{X}}, \mathbf{A}; \mathbf{\Theta})$	76.53 ±1.12	63.34 ±1.44	47.42 ±3.25	66.76 ±8.58	51.09 ±8.84	43.78 ±8.95	52.97 ±6.07	46.80 ±6.76

Citation networks
(trained with GCN)
Graphs with homophilic labels and features

Co-purchase networks
(trained with GIN)
Graphs with homophilic labels but some heterophilic features

University social networks
(trained with TAGCN)
Graphs with heterophilic labels and features

Navarro et al., "Adaptive node feature selection for graph neural networks", ICML 2026

Node classification on real-world data reveals feature and structure relevance

How does node classification performance change under edge and feature perturbations?

	Setting	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
	$\mathrm{GNN}(\mathbf{X}, \mathbf{A}; \mathbf{\Theta})$	85.83 ±0.46	74.39 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
Perturb graph	$\mathrm{MLP}(\mathbf{X}; \mathbf{\Theta})$	74.21 ±1.40	70.02 ±1.39	88.65 ±0.41	88.73 ±0.73	81.63 ±0.75	78.92 ±5.51	82.70 ±2.16	83.60 ±6.62
Perturb graph	$\mathrm{GNN}(\mathbf{X}, \tilde{\mathbf{A}}; \mathbf{\Theta})$	36.97 ±1.80	35.13 ±2.41	67.07 ±0.80	30.52 ±5.77	37.64 ±0.45	67.57 ±7.83	70.27 ±7.05	78.40 ±6.37
Perturb features	$\mathrm{GNN}(\mathbf{W}, \mathbf{A}; \mathbf{\Theta})$	82.92 ±1.54	67.37 ±1.61	76.23 ±0.53	89.92 ±0.58	85.49 ±0.44	49.19 ±7.91	55.68 ±2.76	47.20 ±7.00
Perturb features	$\mathrm{GNN}(\tilde{\mathbf{X}}, \mathbf{A}; \mathbf{\Theta})$	76.53 ±1.12	63.34 ±1.44	47.42 ±3.25	66.76 ±8.58	51.09 ±8.84	43.78 ±8.95	52.97 ±6.07	46.80 ±6.76

Citation networks
(trained with GCN)
Graphs with homophilic labels and features

Co-purchase networks
(trained with GIN)
Graphs with homophilic labels but some heterophilic features

University social networks
(trained with TAGCN)
Graphs with heterophilic labels and features

Perturbations suggest how features and graph jointly influence node-level predictions

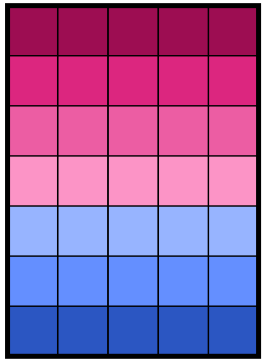
nav@rice.edu

Navarro et al., “Adaptive node feature selection for graph neural networks”, ICML 2026

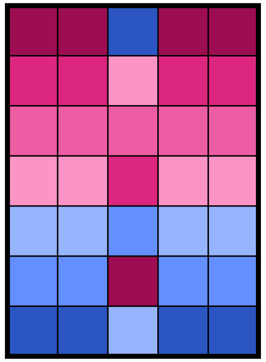
We measure feature importance through permutation testing

$$\delta_m\left(\mathbf{y}, \mathbf{X}, \tilde{\mathbf{X}}^{(m)}\right) = \mathrm{Acc}\Big(\mathbf{y}, f\left(\mathbf{X}, \mathbf{A}; \boldsymbol{\Theta}\right)\Big) - \mathrm{Acc}\Big(\mathbf{y}, f\left(\tilde{\mathbf{X}}^{(m)}, \mathbf{A}; \boldsymbol{\Theta}\right)\Big)$$

Node features $\mathbf{X}$



Permuted features $\tilde{\mathbf{X}}^{(3)}$

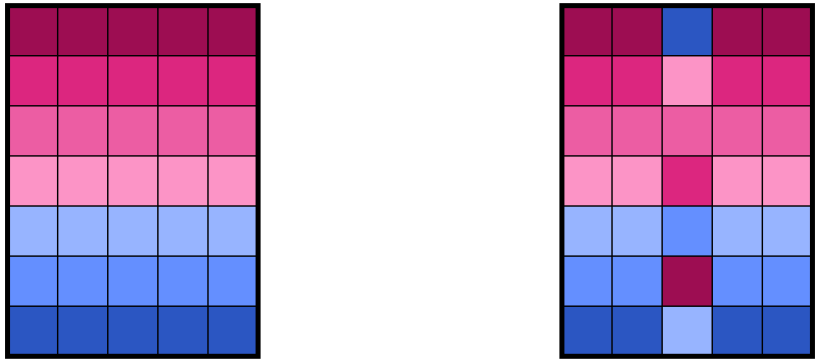


We measure feature importance through permutation testing

$$\delta_m\left(\mathbf{y}, \mathbf{X}, \tilde{\mathbf{X}}^{(m)}\right) = \mathrm{Acc}\Big(\mathbf{y}, f\left(\mathbf{X}, \mathbf{A}; \boldsymbol{\Theta}\right)\Big) - \mathrm{Acc}\Big(\mathbf{y}, f\left(\tilde{\mathbf{X}}^{(m)}, \mathbf{A}; \boldsymbol{\Theta}\right)\Big)$$

- $\tilde{\mathbf{X}}^{(m)}$ as $\mathbf{X}$ with permuted feature m
- Decouples feature m from remaining features and node labels $\mathbf{y}$
- $\mathrm{Acc}(\mathbf{y}, \mathbf{Z})$ measures the accuracy of embeddings $\hat{\mathbf{y}} = g(\mathbf{Z})$ for classifier g

Node features $\mathbf{X}$ Permuted features $\tilde{\mathbf{X}}^{(3)}$

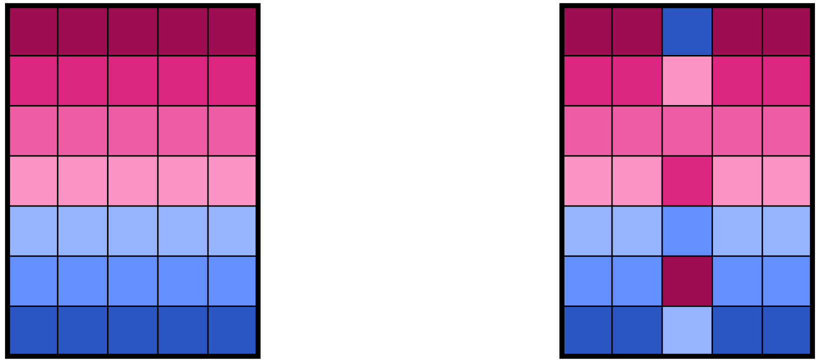


We measure feature importance through permutation testing

$$\delta_m\left(\mathbf{y}, \mathbf{X}, \tilde{\mathbf{X}}^{(m)}\right) = \mathrm{Acc}\Big(\mathbf{y}, f\left(\mathbf{X}, \mathbf{A}; \boldsymbol{\Theta}\right)\Big) - \mathrm{Acc}\Big(\mathbf{y}, f\left(\tilde{\mathbf{X}}^{(m)}, \mathbf{A}; \boldsymbol{\Theta}\right)\Big)$$

- $\tilde{\mathbf{X}}^{(m)}$ as $\mathbf{X}$ with permuted feature m
- Decouples feature m from remaining features and node labels $\mathbf{y}$
- $\mathrm{Acc}(\mathbf{y}, \mathbf{Z})$ measures the accuracy of embeddings $\hat{\mathbf{y}} = g(\mathbf{Z})$ for classifier g

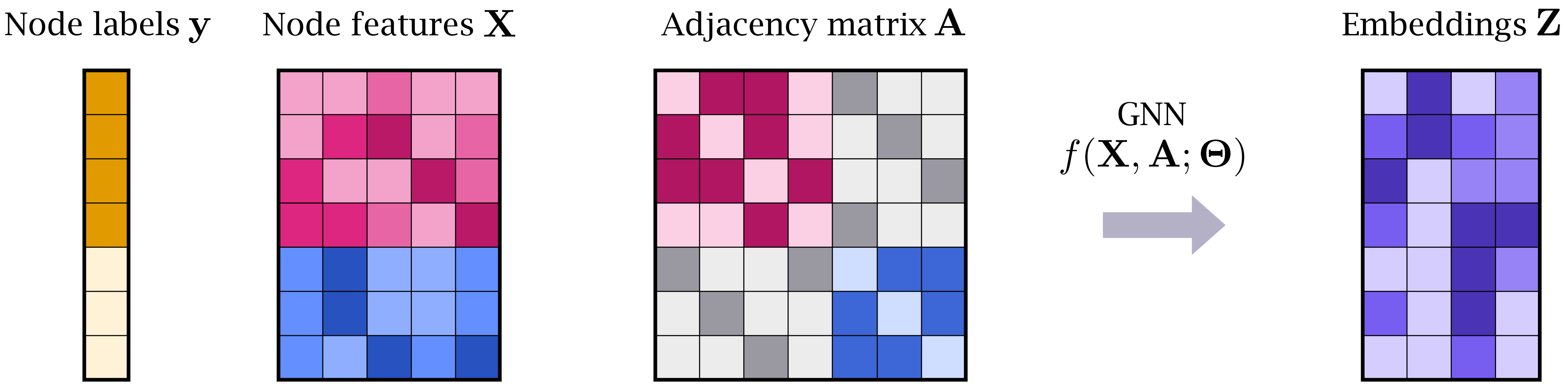
Node features $\mathbf{X}$ Permuted features $\tilde{\mathbf{X}}^{(3)}$



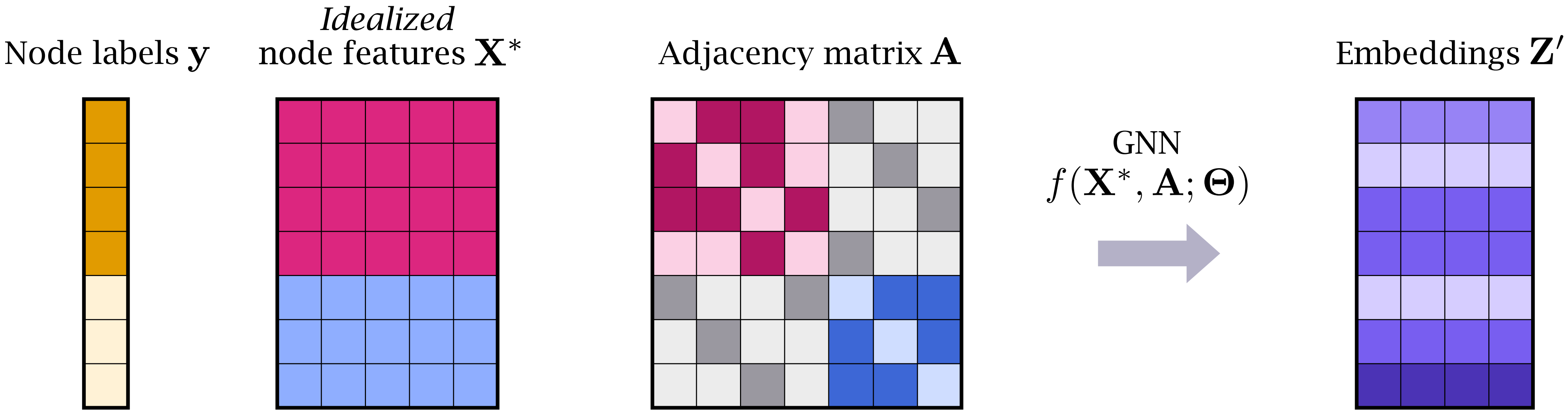
Change in accuracy approximates effect of feature on model performance

Navarro et al., "Adaptive node feature selection for graph neural networks", ICML 2026

How do node feature permutations affect GNN performance?



How do node feature permutations affect GNN performance?



$$\mathbf{X}^* = \underset{\mathbf{U}}{\operatorname{argmin}} \|\mathbf{U} - \mathbf{X}\|_F^2 \quad \text{s.t.} \quad \mathbf{U}_{i,:} = \mathbf{U}_{j,:} \quad \forall\, i, j \;\text{ s.t. } y_i = y_j$$

Features identical
within class

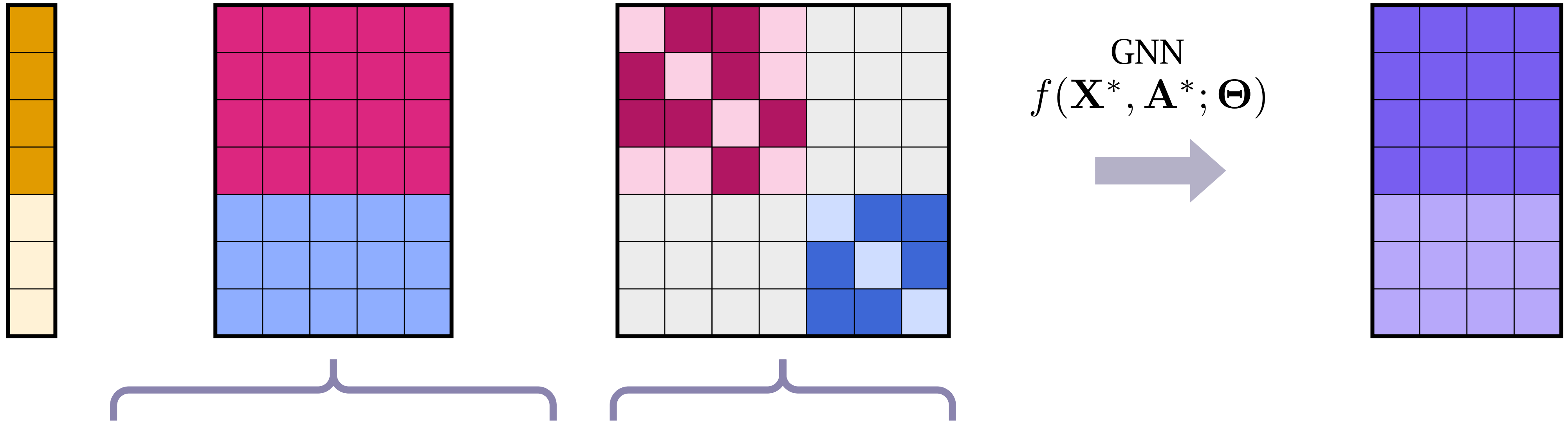
How do node feature permutations affect GNN performance?

Node labels $\mathbf{y}$

Idealized node features $\mathbf{X}^*$

Idealized adjacency matrix $\mathbf{A}^*$

Idealized embeddings $\mathbf{Z}^*$



$$\mathbf{X}^* = \underset{\mathbf{U}}{\operatorname{argmin}} \|\mathbf{U} - \mathbf{X}\|_F^2 \quad \text{s.t.} \quad \mathbf{U}_{i,:} = \mathbf{U}_{j,:} \quad \forall\, i, j \ \text{ s.t. } y_i = y_j$$

Features identical within class

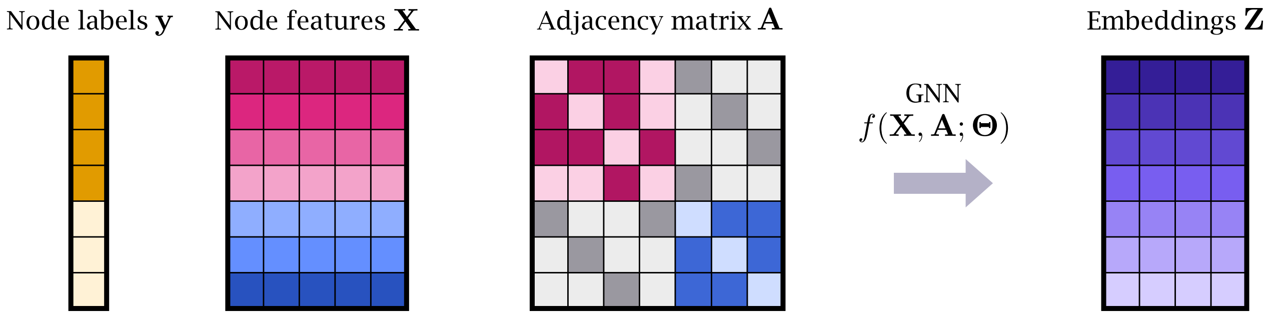
$$A^*_{ij} = \begin{cases} A_{ij}, & y_i = y_j \\ 0, & y_i \neq y_j \end{cases}$$

No cross-class edges

How do node feature permutations affect GNN performance?

Performance of GCN in the absence of feature permutations

Theorem For a GCN $f(\cdot,\cdot;\mathbf{\Theta})$, the error between $\mathbf{Z} = f(\mathbf{X},\mathbf{A};\mathbf{\Theta})$ and $\mathbf{Z}^* = f(\mathbf{X}^*,\mathbf{A}^*;\mathbf{\Theta})$ is bounded by



Navarro et al., “Adaptive node feature selection for graph neural networks”, ICML 2026

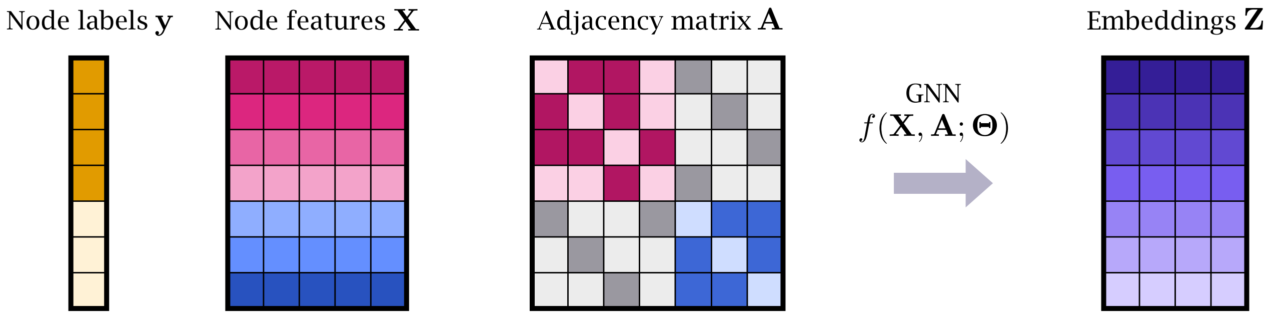
How do node feature permutations affect GNN performance?

Performance of GCN in the absence of feature permutations

Theorem For a GCN $f(\cdot,\cdot;\mathbf{\Theta})$, the error between $\mathbf{Z}=f(\mathbf{X},\mathbf{A};\mathbf{\Theta})$ and $\mathbf{Z}^*=f(\mathbf{X}^*,\mathbf{A}^*;\mathbf{\Theta})$ is bounded by

$$\|\mathbf{Z}^*-\mathbf{Z}\|_F \lesssim (1+\sqrt{N})\|\mathbf{A}-\mathbf{A}^*\|_F\|\mathbf{X}\|_F+\sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{\mathbb{I}\{y_i=y_j\}}{N_c}-\frac{A_{ij}}{d_i+1}\right|\cdot\|\mathbf{X}_{i,:}-\mathbf{X}_{j,:}\|_2$$

for labels $\mathbf{y}$, number of nodes N, degree vector $\mathbf{d}=\mathbf{A}\mathbf{1}$, and number of nodes N_c in class c.

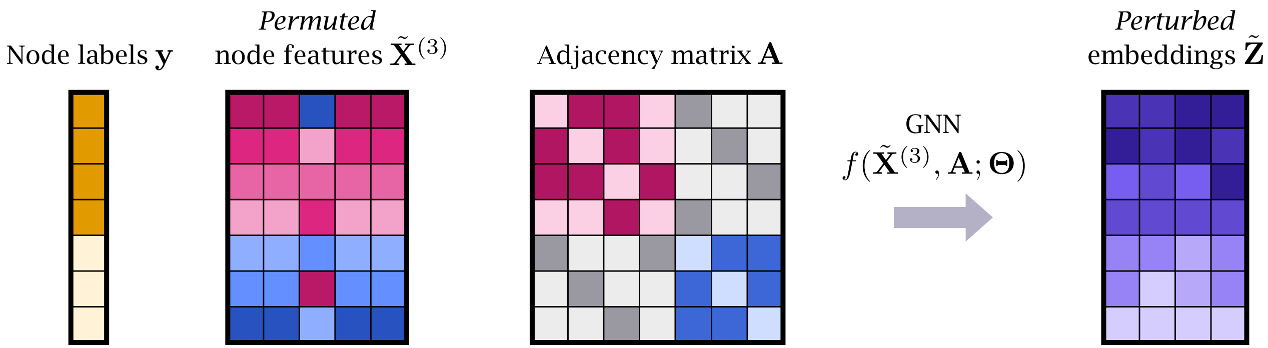


GCN performance depends on label homophily and node feature similarity

Navarro et al., "Adaptive node feature selection for graph neural networks", ICML 2026

Performance of GCN with permuted features

Theorem For a GCN $f(\cdot,\cdot;\mathbf{\Theta})$, the error between the embeddings $\tilde{\mathbf{Z}} = f(\tilde{\mathbf{X}},\mathbf{A};\mathbf{\Theta})$ for features $\tilde{\mathbf{X}}$ as $\mathbf{X}$ with randomly permuted rows and the idealized embeddings $\tilde{\mathbf{Z}}^* = f(\tilde{\mathbf{X}}^*,\mathbf{A}^*;\mathbf{\Theta})$ is bounded by



Navarro et al., "Adaptive node feature selection for graph neural networks", ICML 2026

How do node feature permutations affect GNN performance?

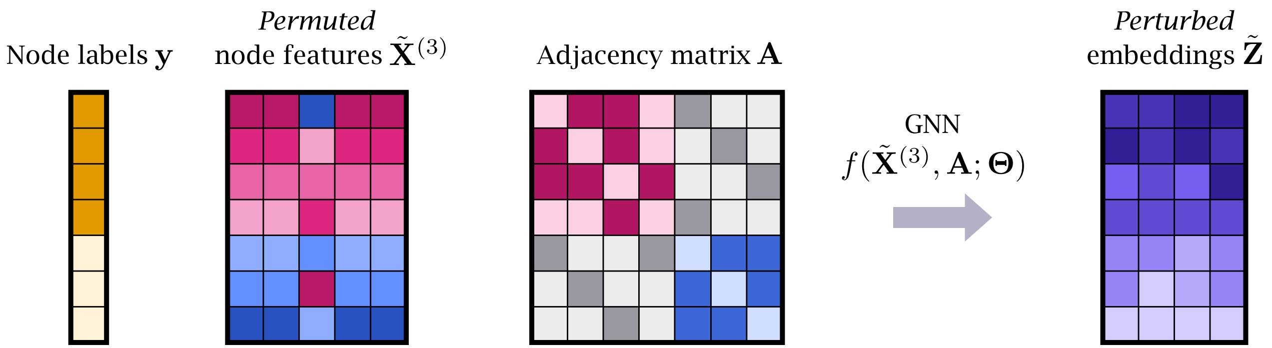
Performance of GCN with permuted features

Theorem For a GCN $f(\cdot,\cdot;\mathbf{\Theta})$, the error between the embeddings $\tilde{\mathbf{Z}} = f(\tilde{\mathbf{X}},\mathbf{A};\mathbf{\Theta})$ for features $\tilde{\mathbf{X}}$ as $\mathbf{X}$ with randomly permuted rows and the idealized embeddings $\tilde{\mathbf{Z}}^* = f(\tilde{\mathbf{X}}^*,\mathbf{A}^*;\mathbf{\Theta})$ is bounded by

$$\|\tilde{\mathbf{Z}}^* - \tilde{\mathbf{Z}}\|_F \lesssim (1+\sqrt{N})\|\mathbf{A}-\mathbf{A}^*\|_F\|\mathbf{X}\|_F + \sqrt{\gamma}\sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{\mathbb{I}\{y_i=y_j\}}{N_c} - \frac{A_{ij}}{d_i+1}\right|$$

$$\text{where} \qquad \gamma = \tfrac{2}{N-1}\left(\|\mathbf{X}\|_F^2 - \tfrac{1}{N}\|\mathbf{X}^\top\mathbf{1}\|_2^2\right) + \alpha t M\sqrt{N}$$

with probability at least $e^{-t^2/4}$ for labels $\mathbf{y}$, number of nodes N, degree vector $\mathbf{d}=\mathbf{A1}$, number of nodes N_c in class c, and $\alpha = \max_{m\in[M],k,\ell\in[N]}(X_{km}-X_{lm})^2$.



Error with permuted features depends on node feature variance and label homophily

Navarro et al., "Adaptive node feature selection for graph neural networks", ICML 2026

Node permutation testing for feature selection while preserving performance

Method	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
All features	85.83 ±0.46	74.38 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
NPT	79.19 ±2.45	69.35 ±1.49	87.11 ±0.75	93.59 ±0.79	90.09 ±0.51	69.73 ±6.26	72.97 ±9.21	73.20 ±6.88
NPT-gaussian	78.86 ±1.42	68.48 ±1.23	86.69 ±0.72	93.49 ±0.67	89.85 ±0.17	65.41 ±9.76	68.65 ±8.48	68.80 ±8.26
NPT-mask	76.05 ±1.08	68.12 ±1.69	86.11 ±0.82	93.48 ±0.59	89.94 ±0.22	63.24 ±4.39	64.86 ±5.14	72.40 ±7.31
TFI	72.73 ±5.47	65.77 ±2.04	83.80 ±0.92	93.02 ±0.68	90.09 ±0.23	61.62 ±7.13	61.08 ±4.39	52.40 ±3.44
MI	66.83 ±3.68	63.79 ±1.02	85.96 ±1.00	93.56 ±0.61	90.33 ±0.32	63.78 ±5.82	65.41 ±7.33	69.60 ±5.99
h_{attr}	39.96 ±1.00	22.59 ±1.01	78.85 ±0.21	93.53 ±0.46	90.09 ±0.46	55.14 ±5.01	58.38 ±5.01	45.60 ±6.62
h_{Euc}	32.77 ±1.66	22.62 ±1.03	74.37 ±0.60	93.59 ±0.50	89.19 ±0.51	52.43 ±4.05	57.30 ±3.15	44.00 ±7.48
h_{GE}	31.44 ±1.35	22.47 ±1.01	70.52 ±0.55	93.41 ±0.46	89.24 ±0.25	52.43 ±4.05	57.30 ±3.15	44.00 ±7.48
Random	39.76 ±1.22	34.39 ±4.07	70.71 ±1.08	91.99 ±0.49	88.27 ±0.26	56.65 ±4.69	58.49 ±3.23	57.04 ±4.49

Perturbation testing (Ours)

Citation networks
(trained with GCN)

Co-purchase networks
(trained with GIN)

University social networks
(trained with TAGCN)

Node permutation testing for feature selection while preserving performance

	Method	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
	All features	85.83 ±0.46	74.38 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
Perturbation testing (Ours)	NPT	79.19 ±2.45	69.35 ±1.49	87.11 ±0.75	93.59 ±0.79	90.09 ±0.51	69.73 ±6.26	72.97 ±9.21	73.20 ±6.88
	NPT-gaussian	78.86 ±1.42	68.48 ±1.23	86.69 ±0.72	93.49 ±0.67	89.85 ±0.17	65.41 ±9.76	68.65 ±8.48	68.80 ±8.26
	NPT-mask	76.05 ±1.08	68.12 ±1.69	86.11 ±0.82	93.48 ±0.59	89.94 ±0.22	63.24 ±4.39	64.86 ±5.14	72.40 ±7.31
Mutual information	TFI	72.73 ±5.47	65.77 ±2.04	83.80 ±0.92	93.02 ±0.68	90.09 ±0.23	61.62 ±7.13	61.08 ±4.39	52.40 ±3.44
	MI	66.83 ±3.68	63.79 ±1.02	85.96 ±1.00	93.56 ±0.61	90.33 ±0.32	63.78 ±5.82	65.41 ±7.33	69.60 ±5.99
	h_{attr}	39.96 ±1.00	22.59 ±1.01	78.85 ±0.21	93.53 ±0.46	90.09 ±0.46	55.14 ±5.01	58.38 ±5.01	45.60 ±6.62
	h_{Euc}	32.77 ±1.66	22.62 ±1.03	74.37 ±0.60	93.59 ±0.50	89.19 ±0.51	52.43 ±4.05	57.30 ±3.15	44.00 ±7.48
	h_{GE}	31.44 ±1.35	22.47 ±1.01	70.52 ±0.55	93.41 ±0.46	89.24 ±0.25	52.43 ±4.05	57.30 ±3.15	44.00 ±7.48
	Random	39.76 ±1.22	34.39 ±4.07	70.71 ±1.08	91.99 ±0.49	88.27 ±0.26	56.65 ±4.69	58.49 ±3.23	57.04 ±4.49

Citation networks (trained with GCN) — Cora, CiteSeer, PubMed

Co-purchase networks (trained with GIN) — Photo, Computers

University social networks (trained with TAGCN) — Cornell, Texas, Wisconsin

Navarro et al., “Adaptive node feature selection for graph neural networks”, ICML 2026

Node permutation testing for feature selection while preserving performance

	Method	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
	All features	85.83 ±0.46	74.38 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
Perturbation testing (Ours)	NPT	79.19 ±2.45	69.35 ±1.49	87.11 ±0.75	93.59 ±0.79	90.09 ±0.51	69.73 ±6.26	72.97 ±9.21	73.20 ±6.88
	NPT-gaussian	78.86 ±1.42	68.48 ±1.23	86.69 ±0.72	93.49 ±0.67	89.85 ±0.17	65.41 ±9.76	68.65 ±8.48	68.80 ±8.26
	NPT-mask	76.05 ±1.08	68.12 ±1.69	86.11 ±0.82	93.48 ±0.59	89.94 ±0.22	63.24 ±4.39	64.86 ±5.14	72.40 ±7.31
Mutual information	TFI	72.73 ±5.47	65.77 ±2.04	83.80 ±0.92	93.02 ±0.68	90.09 ±0.23	61.62 ±7.13	61.08 ±4.39	52.40 ±3.44
	MI	66.83 ±3.68	63.79 ±1.02	85.96 ±1.00	93.56 ±0.61	90.33 ±0.32	63.78 ±5.82	65.41 ±7.33	69.60 ±5.99
Homophily metrics	h_{attr}	39.96 ±1.00	22.59 ±1.01	78.85 ±0.21	93.53 ±0.46	90.09 ±0.46	55.14 ±5.01	58.38 ±5.01	45.60 ±6.62
	h_{Euc}	32.77 ±1.66	22.62 ±1.03	74.37 ±0.60	93.59 ±0.50	89.19 ±0.51	52.43 ±4.05	57.30 ±3.15	44.00 ±7.48
	h_{GE}	31.44 ±1.35	22.47 ±1.01	70.52 ±0.55	93.41 ±0.46	89.24 ±0.25	52.43 ±4.05	57.30 ±3.15	44.00 ±7.48
	Random	39.76 ±1.22	34.39 ±4.07	70.71 ±1.08	91.99 ±0.49	88.27 ±0.26	56.65 ±4.69	58.49 ±3.23	57.04 ±4.49

Citation networks (trained with GCN) — Cora, CiteSeer, PubMed

Co-purchase networks (trained with GIN) — Photo, Computers

University social networks (trained with TAGCN) — Cornell, Texas, Wisconsin

Navarro et al., “Adaptive node feature selection for graph neural networks”, ICML 2026

Node permutation testing for feature selection while preserving performance

	Method	Cora	CiteSeer	PubMed	Photo	Computers	Cornell	Texas	Wisconsin
	All features	85.83 ±0.46	74.38 ±1.09	88.85 ±0.42	94.04 ±0.69	90.58 ±0.79	74.59 ±7.76	82.70 ±4.05	82.80 ±3.25
Perturbation testing (Ours)	NPT	79.19 ±2.45	69.35 ±1.49	87.11 ±0.75	93.59 ±0.79	90.09 ±0.51	69.73 ±6.26	72.97 ±9.21	73.20 ±6.88
	NPT-gaussian	78.86 ±1.42	68.48 ±1.23	86.69 ±0.72	93.49 ±0.67	89.85 ±0.17	65.41 ±9.76	68.65 ±8.48	68.80 ±8.26
	NPT-mask	76.05 ±1.08	68.12 ±1.69	86.11 ±0.82	93.48 ±0.59	89.94 ±0.22	63.24 ±4.39	64.86 ±5.14	72.40 ±7.31
Mutual information	TFI	72.73 ±5.47	65.77 ±2.04	83.80 ±0.92	93.02 ±0.68	90.09 ±0.23	61.62 ±7.13	61.08 ±4.39	52.40 ±3.44
	MI	66.83 ±3.68	63.79 ±1.02	85.96 ±1.00	93.56 ±0.61	90.33 ±0.32	63.78 ±5.82	65.41 ±7.33	69.60 ±5.99
Homophily metrics	h_{attr}	39.96 ±1.00	22.59 ±1.01	78.85 ±0.21	93.53 ±0.46	90.09 ±0.46	55.14 ±5.01	58.38 ±5.01	45.60 ±6.62
	h_{Euc}	32.77 ±1.66	22.62 ±1.03	74.37 ±0.60	93.59 ±0.50	89.19 ±0.51	52.43 ±4.05	57.30 ±3.15	44.00 ±7.48
	h_{GE}	31.44 ±1.35	22.47 ±1.01	70.52 ±0.55	93.41 ±0.46	89.24 ±0.25	52.43 ±4.05	57.30 ±3.15	44.00 ±7.48
	Random	39.76 ±1.22	34.39 ±4.07	70.71 ±1.08	91.99 ±0.49	88.27 ±0.26	56.65 ±4.69	58.49 ±3.23	57.04 ±4.49

Citation networks
(trained with GCN)

Co-purchase networks
(trained with GIN)

University social networks
(trained with TAGCN)

Permutation-based feature selection adapts to task competitively with or superior to other metrics

nav@rice.edu

Navarro et al., "Adaptive node feature selection for graph neural networks", ICML 2026

We propose adaptive node feature selection via permutation testing during GNN training

Measure feature importance δ_m every T epochs

- Update $\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta} - \lambda \nabla_{\boldsymbol{\Theta}} \mathcal{L}(\mathbf{y}_{\text{train}}, f(\hat{\mathbf{X}}, \mathbf{A}; \boldsymbol{\Theta}))$ with pruned features $\hat{\mathbf{X}}$

We propose adaptive node feature selection via permutation testing during GNN training

Measure feature importance δ_m every T epochs

- Update $\mathbf{\Theta} \leftarrow \mathbf{\Theta} - \lambda \nabla_{\mathbf{\Theta}} \mathcal{L}(\mathbf{y}_{\text{train}}, f(\hat{\mathbf{X}}, \mathbf{A}; \mathbf{\Theta}))$ with pruned features $\hat{\mathbf{X}}$
- Every T epochs:
  - Measure δ_m for every $m = 1, \ldots, M$
  - Update $\hat{\mathbf{X}}$ by pruning features with δ_m below r-quantile $\delta^{(r)}$

Periodically evaluate feature importance and eliminate unnecessary features

We propose adaptive node feature selection via permutation testing during GNN training

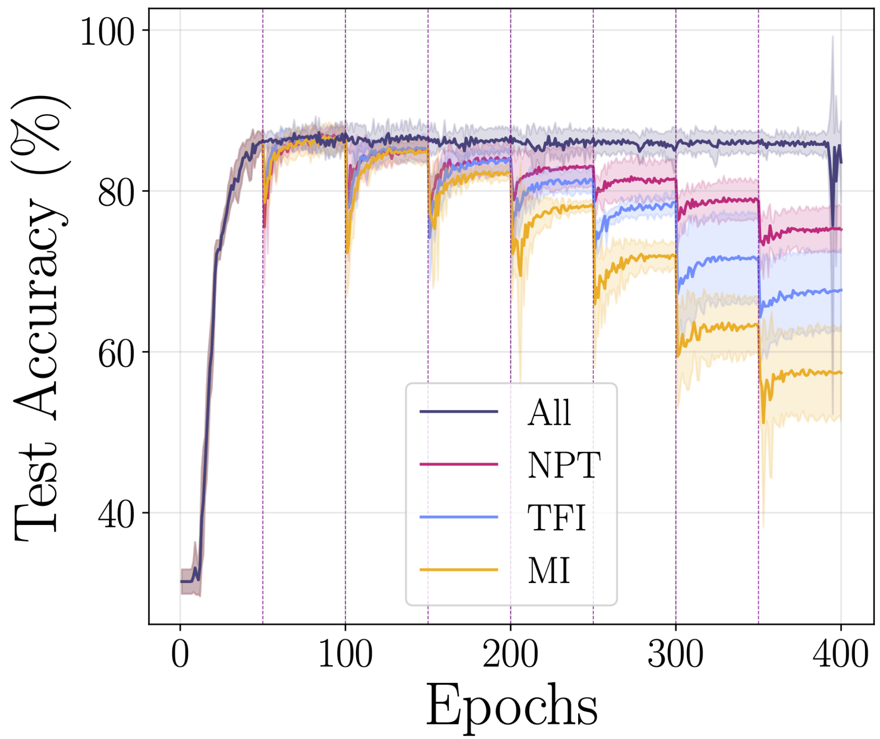
Measure feature importance δ_m every T epochs

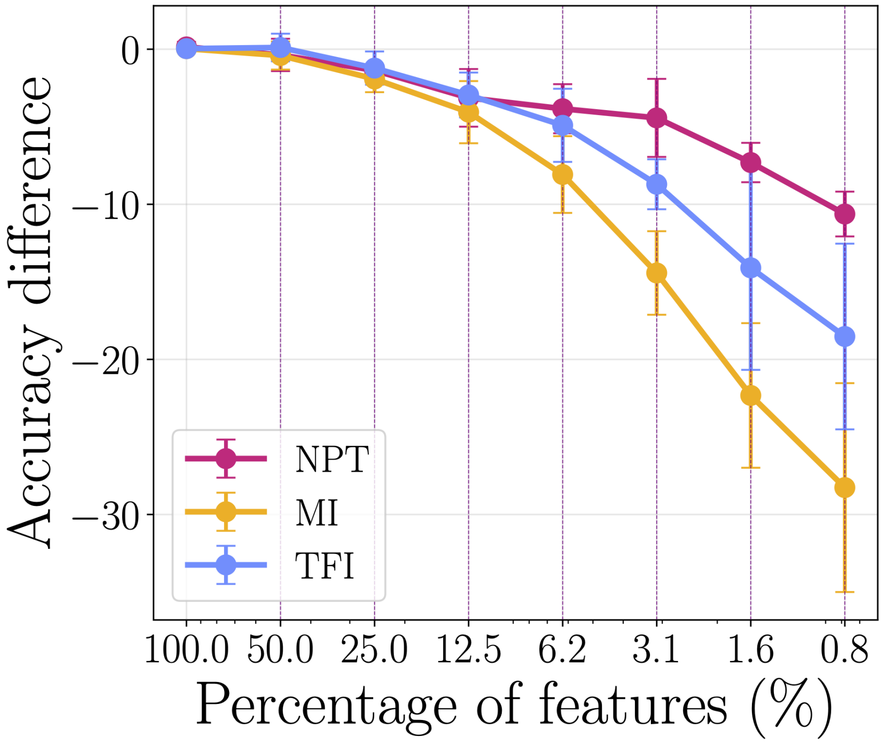
- Update $\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta} - \lambda \nabla_{\boldsymbol{\Theta}} \mathcal{L}(\mathbf{y}_{\text{train}}, f(\hat{\mathbf{X}}, \mathbf{A}; \boldsymbol{\Theta}))$ with pruned features $\hat{\mathbf{X}}$
- Every T epochs:
  - Measure δ_m for every $m = 1, \ldots, M$
  - Update $\hat{\mathbf{X}}$ by pruning features with δ_m below r-quantile $\delta^{(r)}$

- Adapt feature importance to learning task without graph data assumptions
- No restrictions on model architecture
- Flexible to any feature importance score with prior information

Periodically evaluate feature importance and eliminate unnecessary features

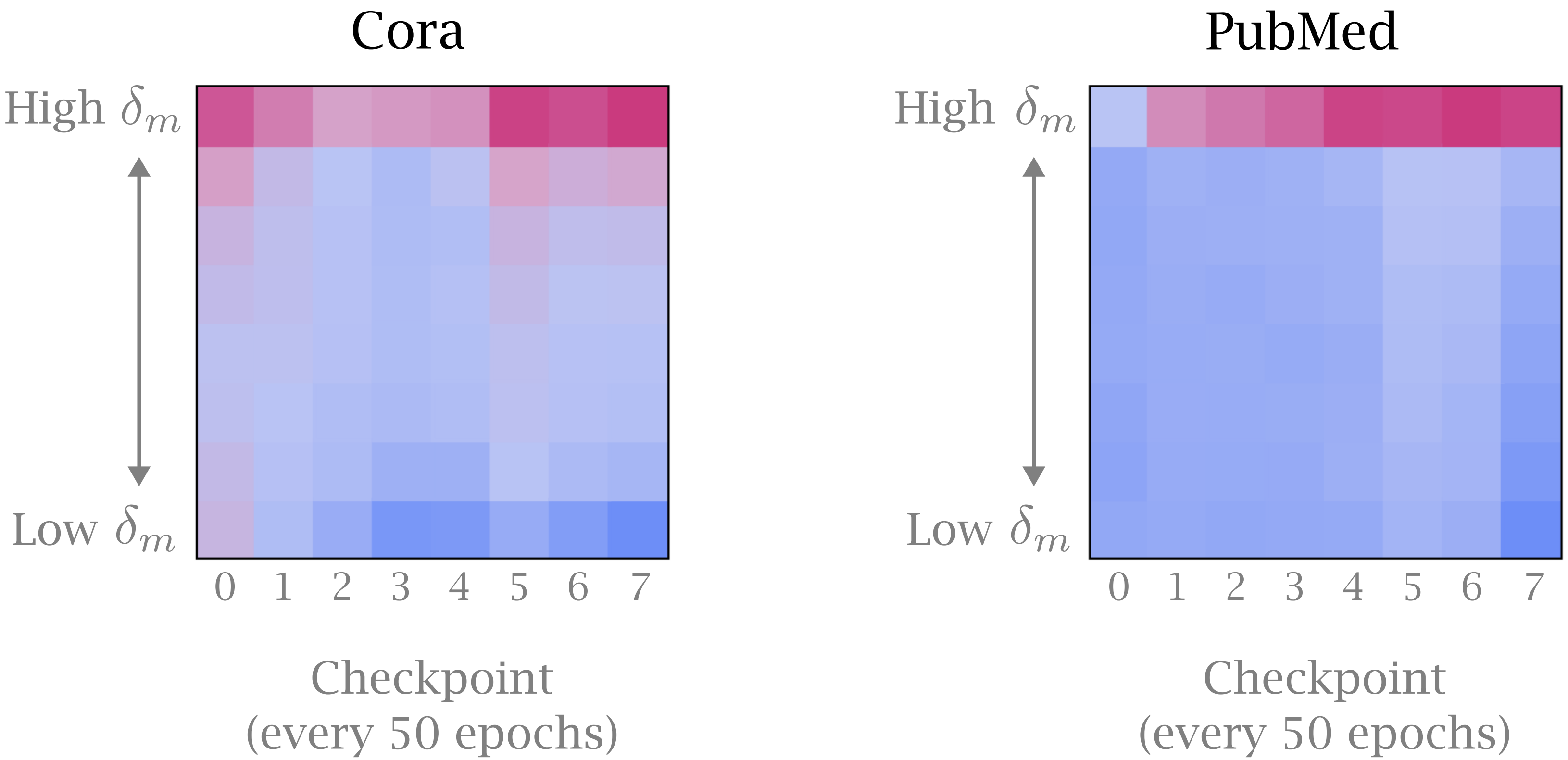
Node permutation testing for feature selection while preserving performance





Permutation-based feature importance maintains highest GCN accuracy as features are eliminated

Relative feature importance is identified quickly during GNN training



Feature importance sufficiently informative for adaptive node feature selection

Navarro et al., "Adaptive node feature selection for graph neural networks", ICML 2026

Adaptive learning can reduce GNN complexity while preserving performance

Node feature selection: Permutation-based scores for feature selection during training

- Exploited well-established statistical metric
- Theoretically showed that permutations return info unique to GNNs
- Empirically found our approach rivals other, data-specific metrics, regardless of dataset

Adaptive learning can reduce GNN complexity while preserving performance

Node feature selection: Permutation-based scores for feature selection during training

- Exploited well-established statistical metric
- Theoretically showed that permutations return info unique to GNNs
- Empirically found our approach rivals other, data-specific metrics, regardless of dataset

Future directions

- Feature selection for link prediction and graph classificationc
- Analysis of conditional permutation tests to account for complex collinearities
- Graph-specific perturbations and feature importance metrics