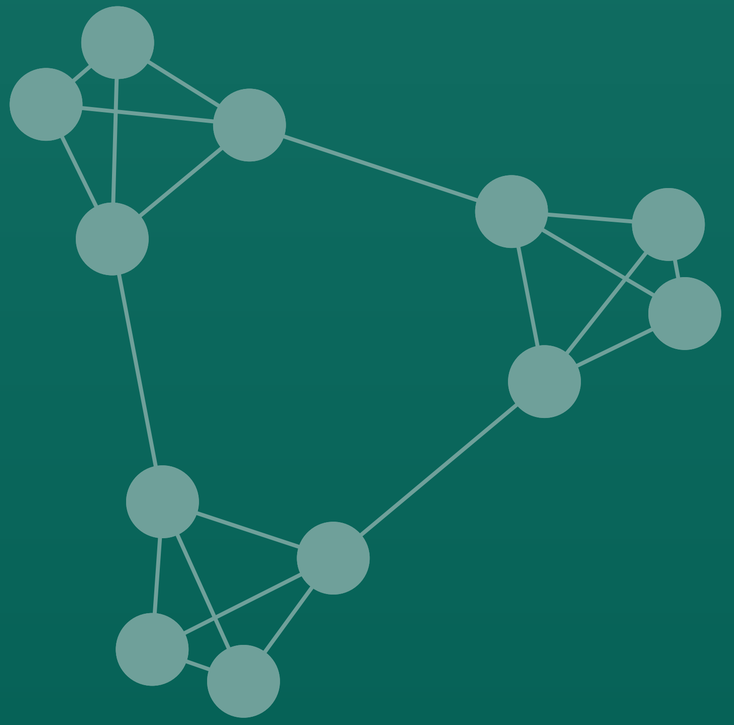
No Metric to Rule Them All



Toward Principled Evaluations
of Graph-Learning Datasets

The Fellowship

Corinna Coupette

Jeremy Wayland

Emily Simons

Bastian Rieck

Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

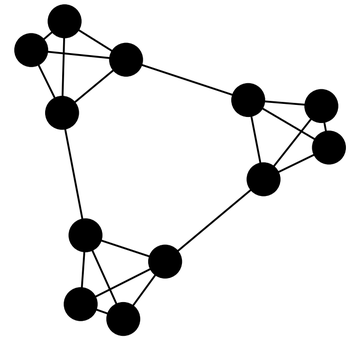
Graph Learning Benchmarks need a reality check.

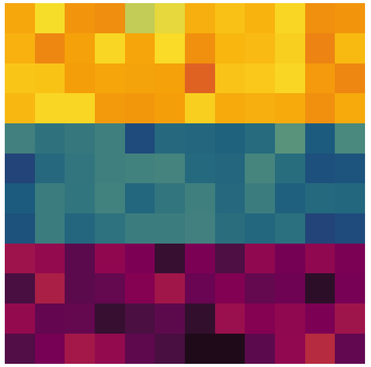
Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Guiding Questions

Q1: What characterizes a good graph-learning dataset?

Q2: How can we evaluate dataset quality in graph learning?

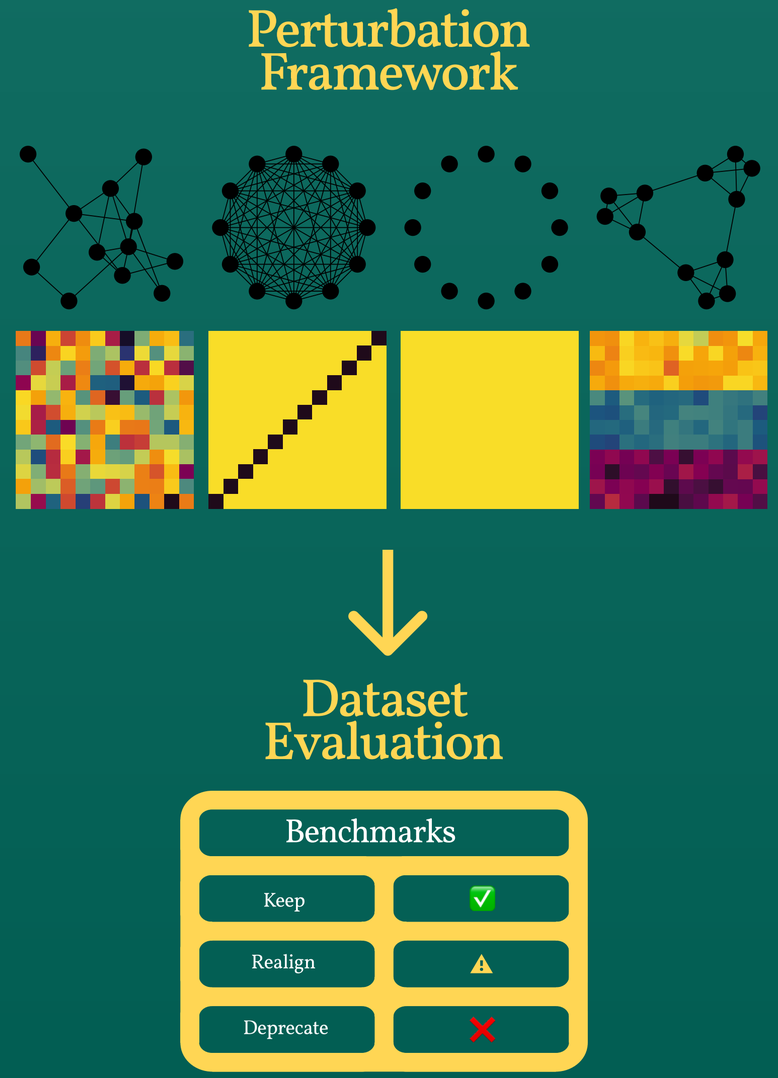




Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

RINGS

Relevant **I**nformation in **N**ode features and **G**raph **S**tructure



Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Guiding Principles

P1: *Task-Relevant Information*
→ Both structure and features should matter for a given task.

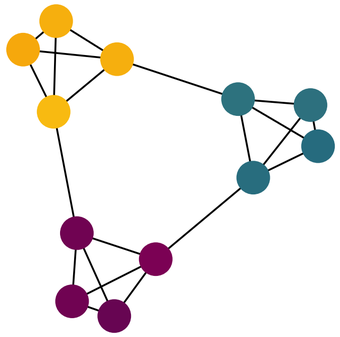
P2: *Complementary Information*
→ Structure and features should offer distinct views of the data.

Associated Metric

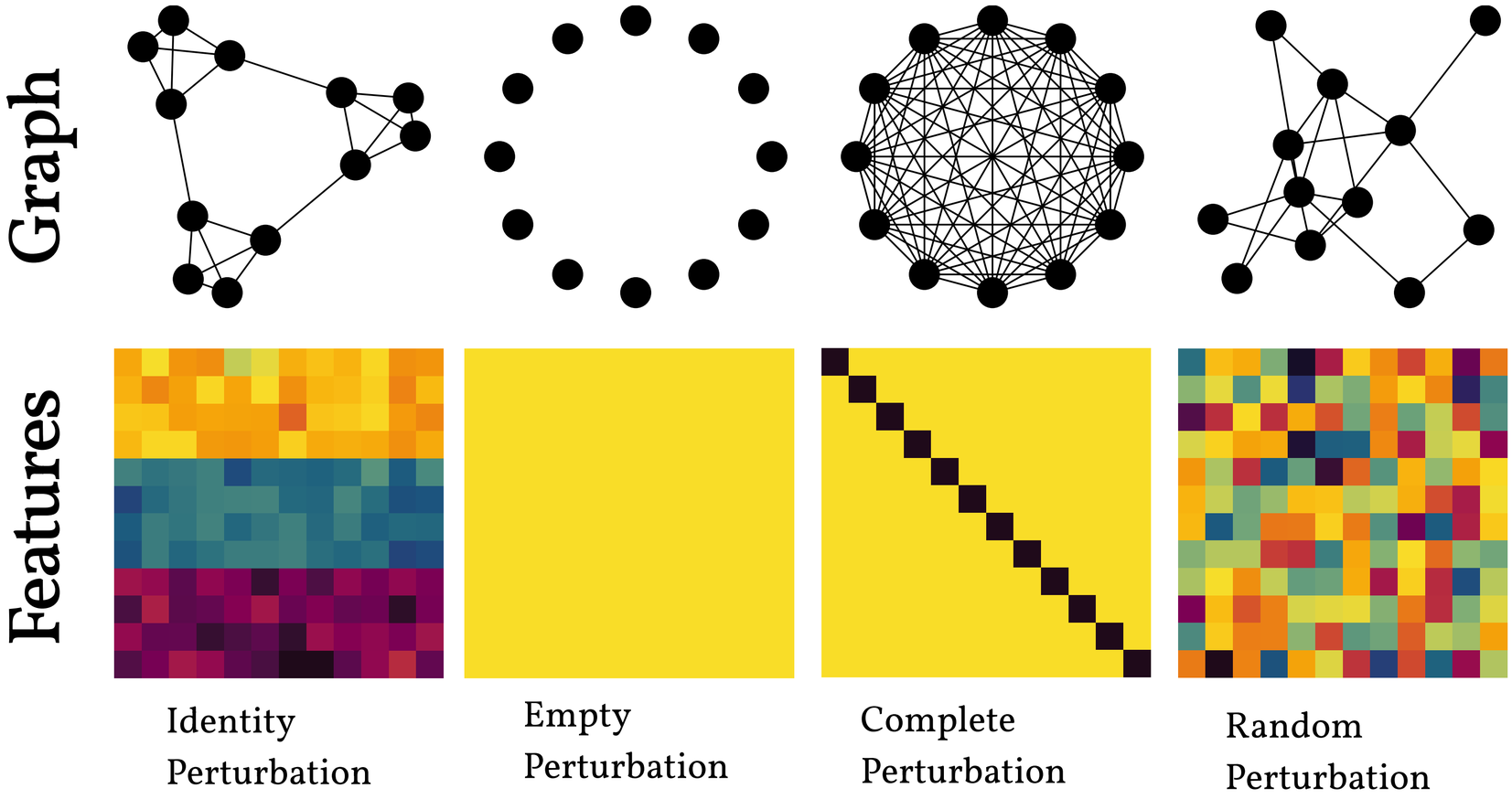
Performance separability

Mode complementarity

Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.



Mode Perturbations



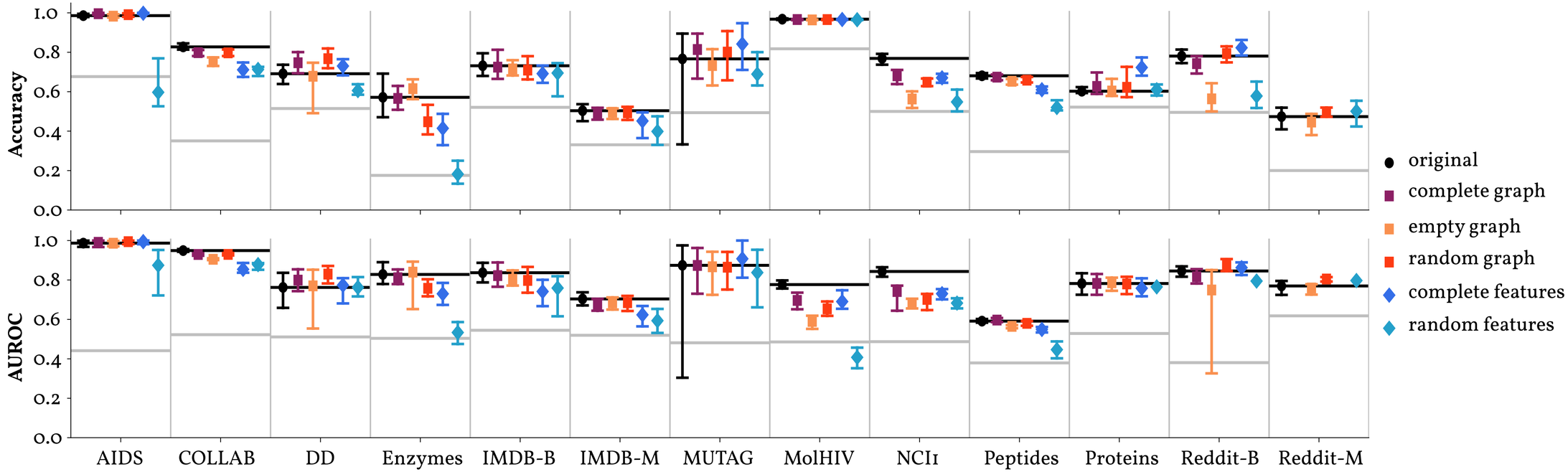
Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Performance Separability

A Task-Dependent Measure

Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Many popular datasets fail to statistically outperform their perturbed versions.



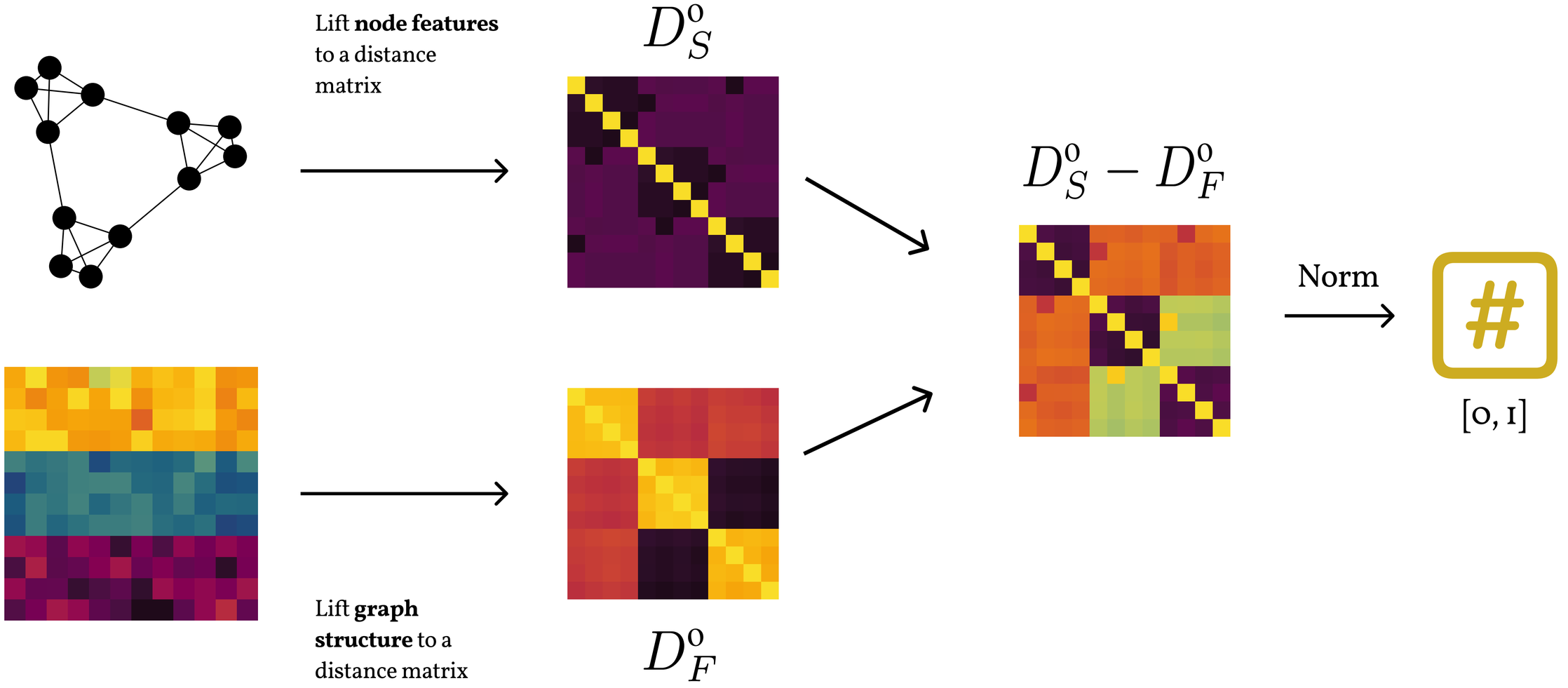
Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Mode Complementarity

A Task-Independent Measure

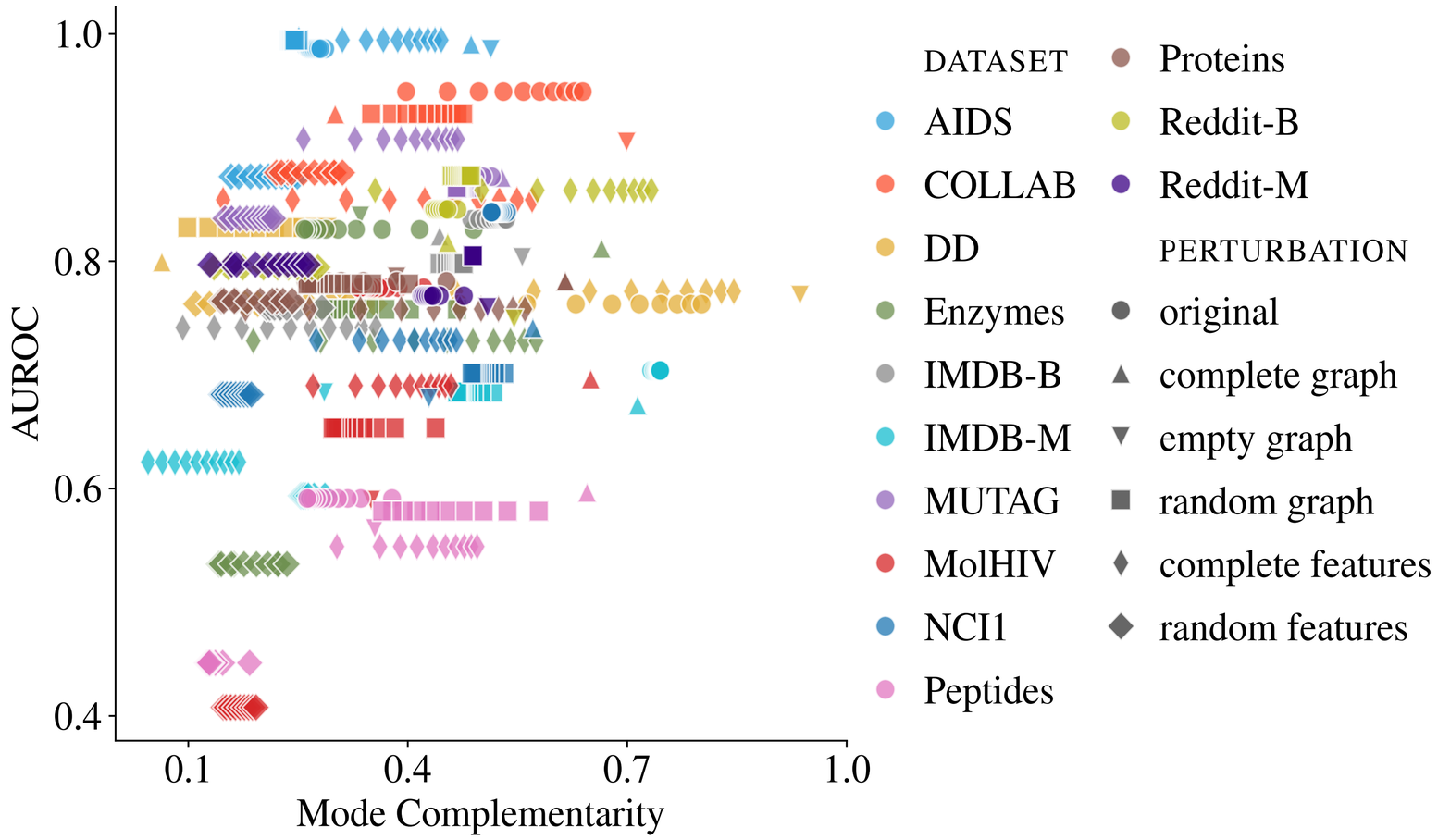
Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Calculating Mode Complementarity



Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Mode complementarity correlates with performance.



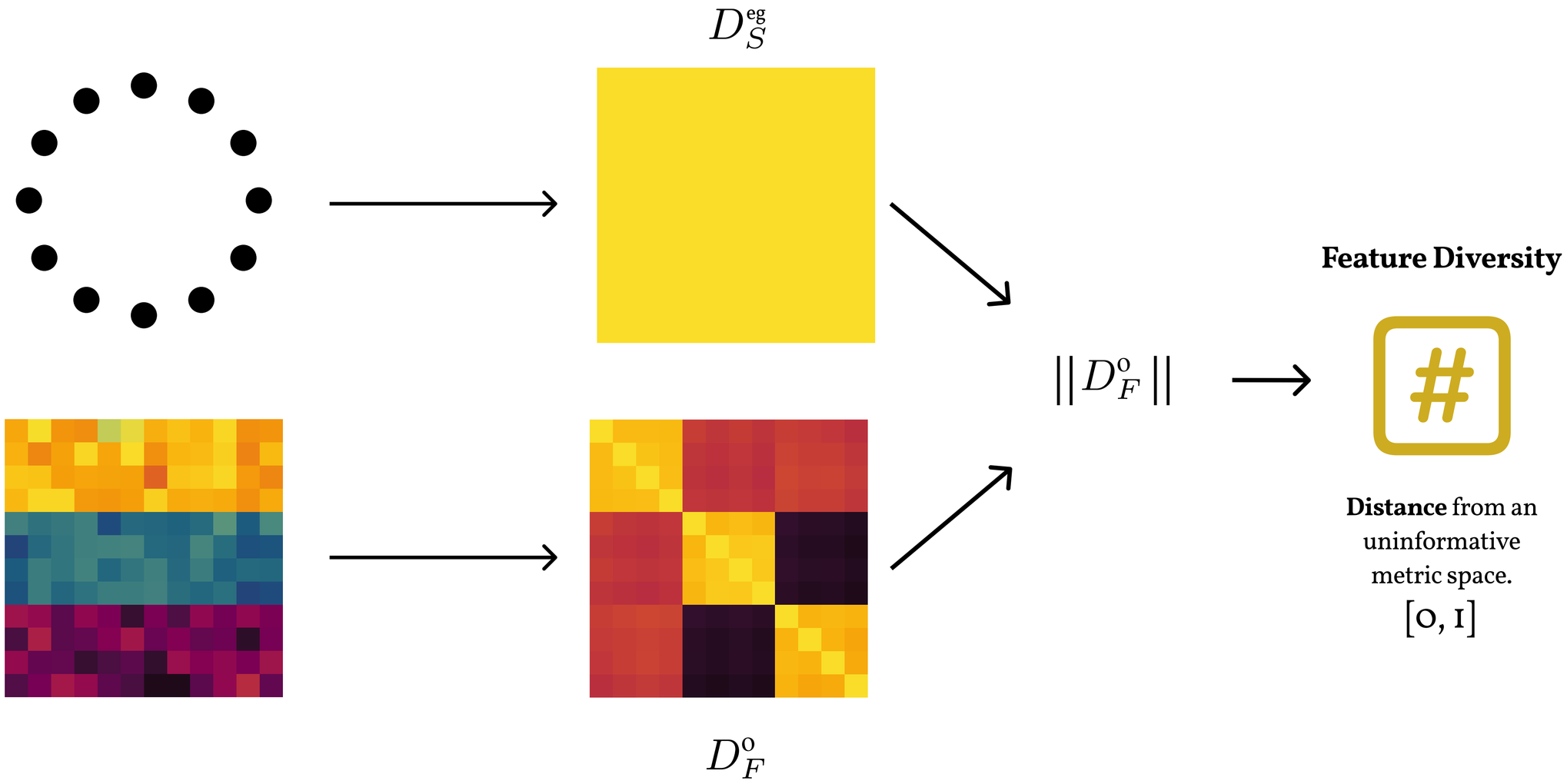
Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Mode Diversity

A Task-Independent Measure

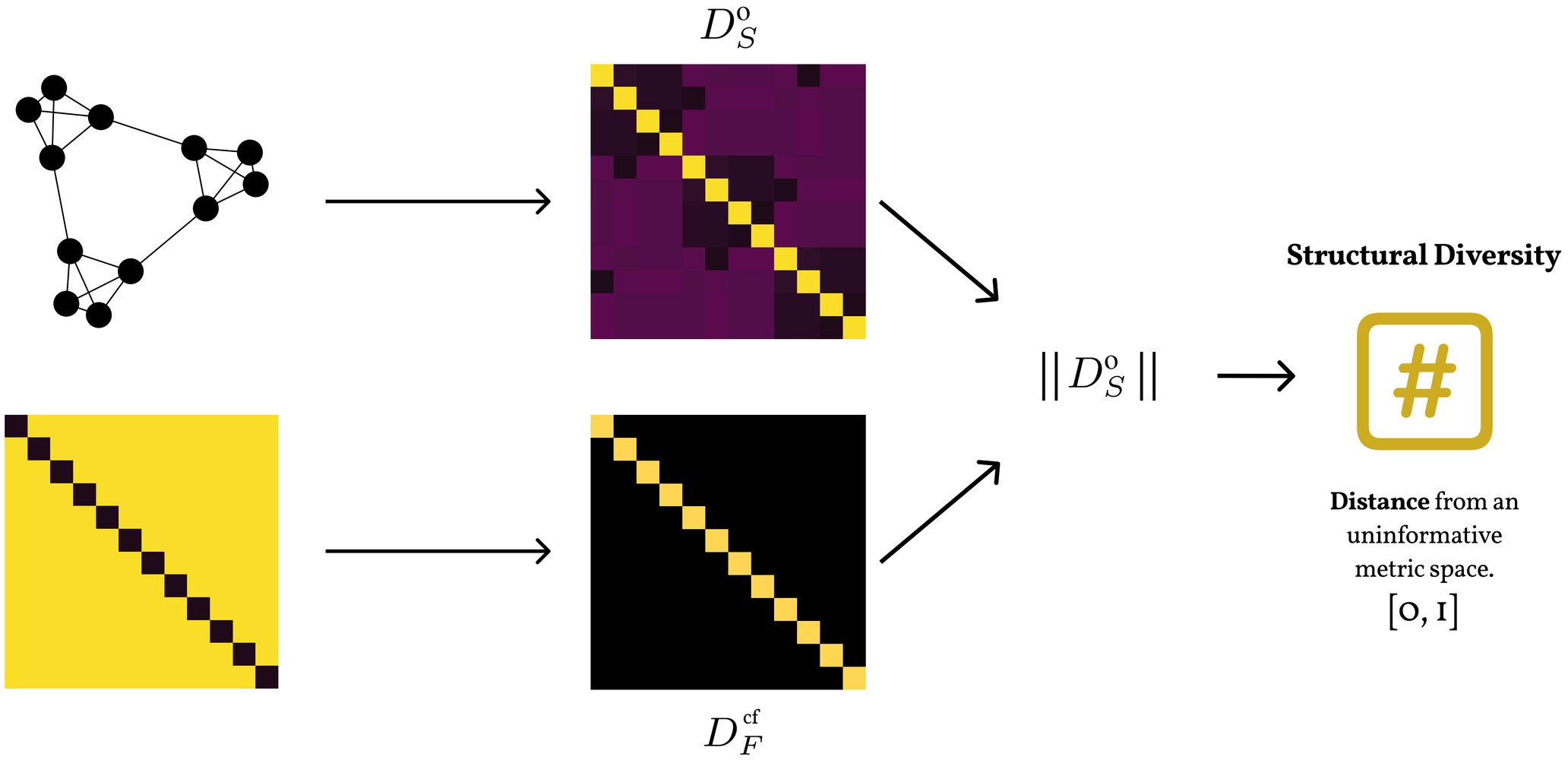
Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Calculating Feature Mode Diversity



Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Calculating Structural Mode Diversity



Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Final Evaluation

A Benchmark Taxonomy

Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

Result Overview

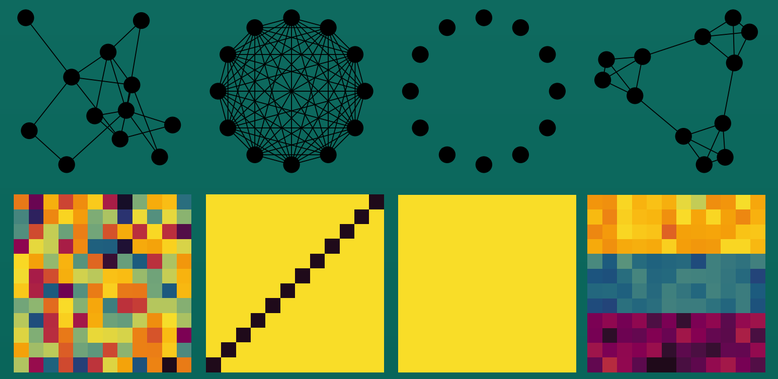
Dataset	Performance Separability (†)			Mode Diversity (‡)		Action
	Structure	Features	Overall	Structure (μ:σ)	Features (μ:σ)	
MolHIV	+	+	++	○:−−	++:++	Keep (†|‡)
NCI1	+	+	++	○:−−	++:++	
Peptides	○	+	+	++:−−	++:++	
AIDS	−	−	−−	○:−	++:○	Realign (†|‡)
DD	−	−	−−	++:−	−−:−−	
MUTAG	○	−	−	○:−−	++:○	
Reddit-{B,M}	−	−	−−	++:−−	++:○	
COLLAB	+	+	++	−:++	−:++	Deprecate (‡)
IMDB-B	○	○	○	−−:○	○:++	
IMDB-M	○	+	+	−−:○	−:++	
Enzymes	−	+	−	−:−	++:○	Deprecate (†|‡)
Proteins	−	−	−−	−:○	++:−	

Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.

RINGS

Relevant **I**nformation in **N**ode features and **G**raph **S**tructure

Perturbation Framework



Desiderata

Both modes should matter for a given task. → **Performance Separability**

Modes should offer distinct and interesting views of the data. → **Mode Complementarity & Mode Diversity**

Metrics

Dataset Evaluation

Benchmarks	
Keep	✅
Realign	⚠
Deprecate	❌

Coupette*, Wayland*, Simons, and Rieck. *No Metric to Rule Them All.* ICML 2025.