GRAIN: Exact Graph Reconstruction from Gradients


Maria Drencheva


Ivo Petrov

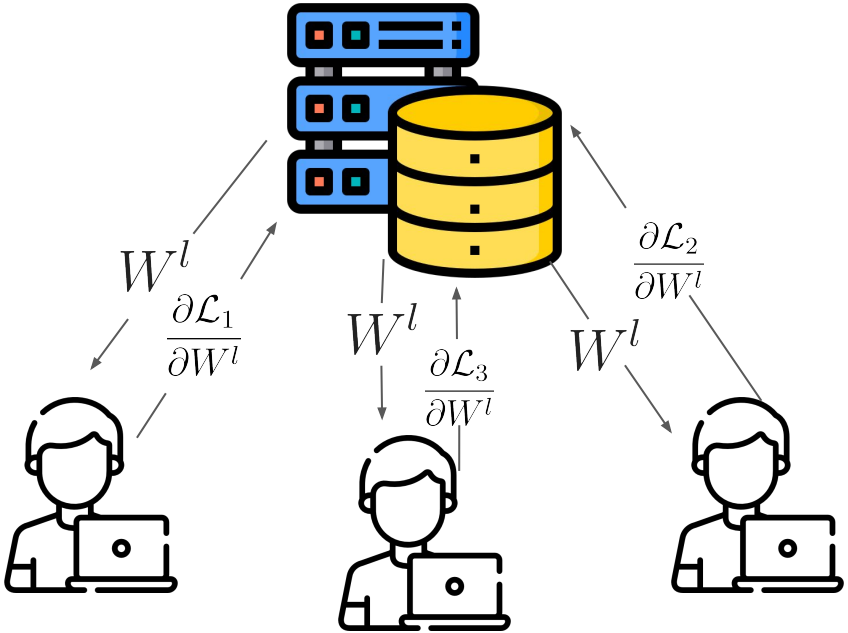

Maximilian Baader


Dimitar I. Dimitrov


Martin Vechev

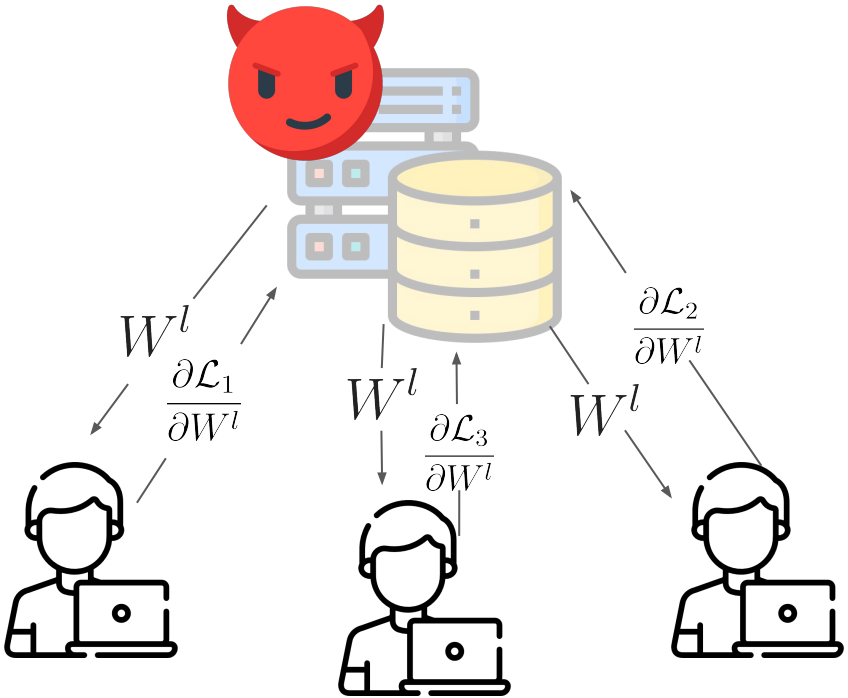
INSAIT SRILAB ETH zürich

Federated Learning and Gradient Inversion



Federated learning enables training a model across multiple clients without sharing their raw data with a central server; instead, they only share **gradient updates.**

Federated Learning and Gradient Inversion



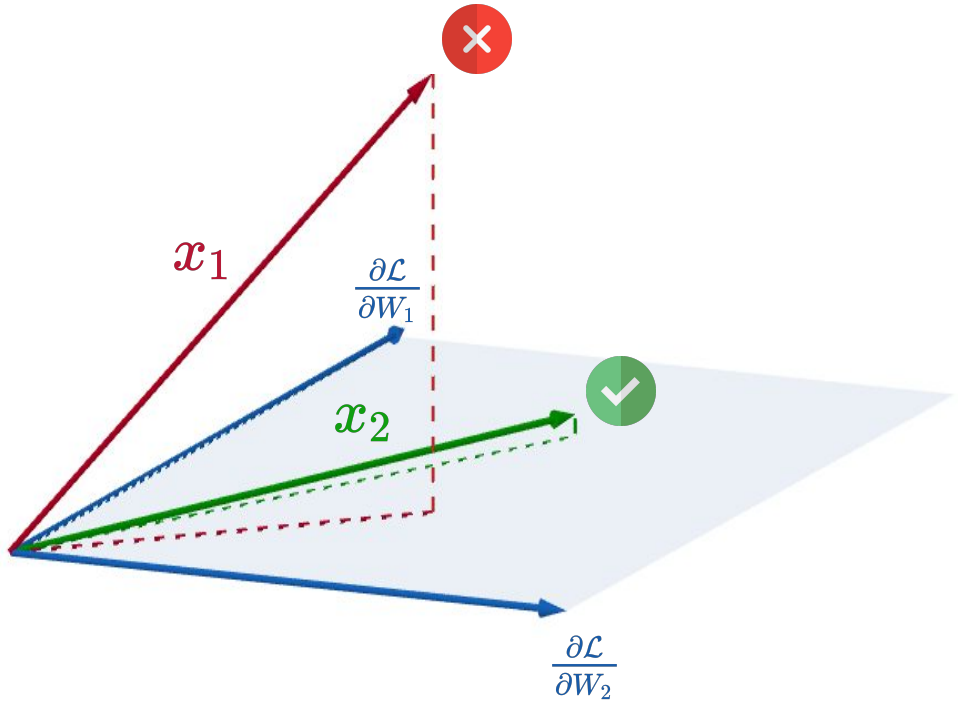
Federated learning enables training a model across multiple clients without sharing their raw data with a central server; instead, they only share **gradient updates.**

However, **Gradient Inversion Attacks** have revealed privacy risks in Federated Learning, as client data can sometimes be reconstructed from the shared gradient updates.

Spancheck filtering of inputs of linear layers

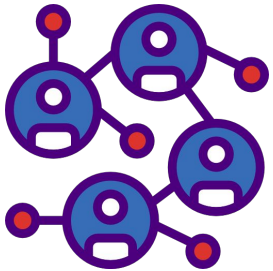
$$\mathbf{Y} = \mathbf{X}\mathbf{W}$$

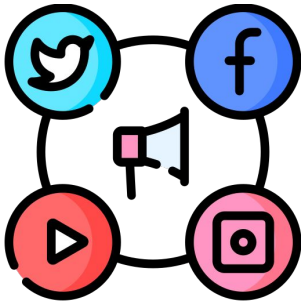
$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}} = \mathbf{X}^T \frac{\partial \mathcal{L}}{\partial \mathbf{Y}}$$



[1] Petrov et. al. in “DAGER: Exact Gradient inversion for Large Language Models”

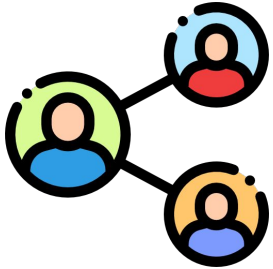
Graph Neural Networks in Federated Learning

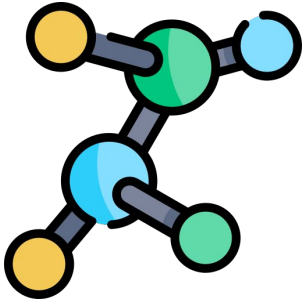




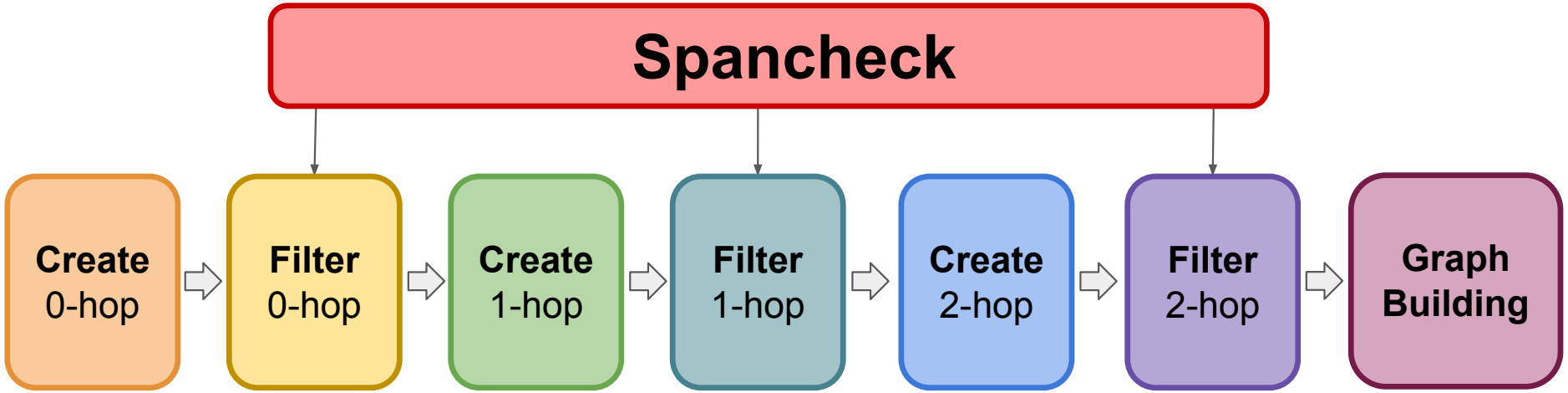
Graph Neural Networks allow for models to be trained on graph data, such as molecules or social and citation networks.

$$X^{l+1} = \mathrm{ReLU}(A^l X^l W^l)$$

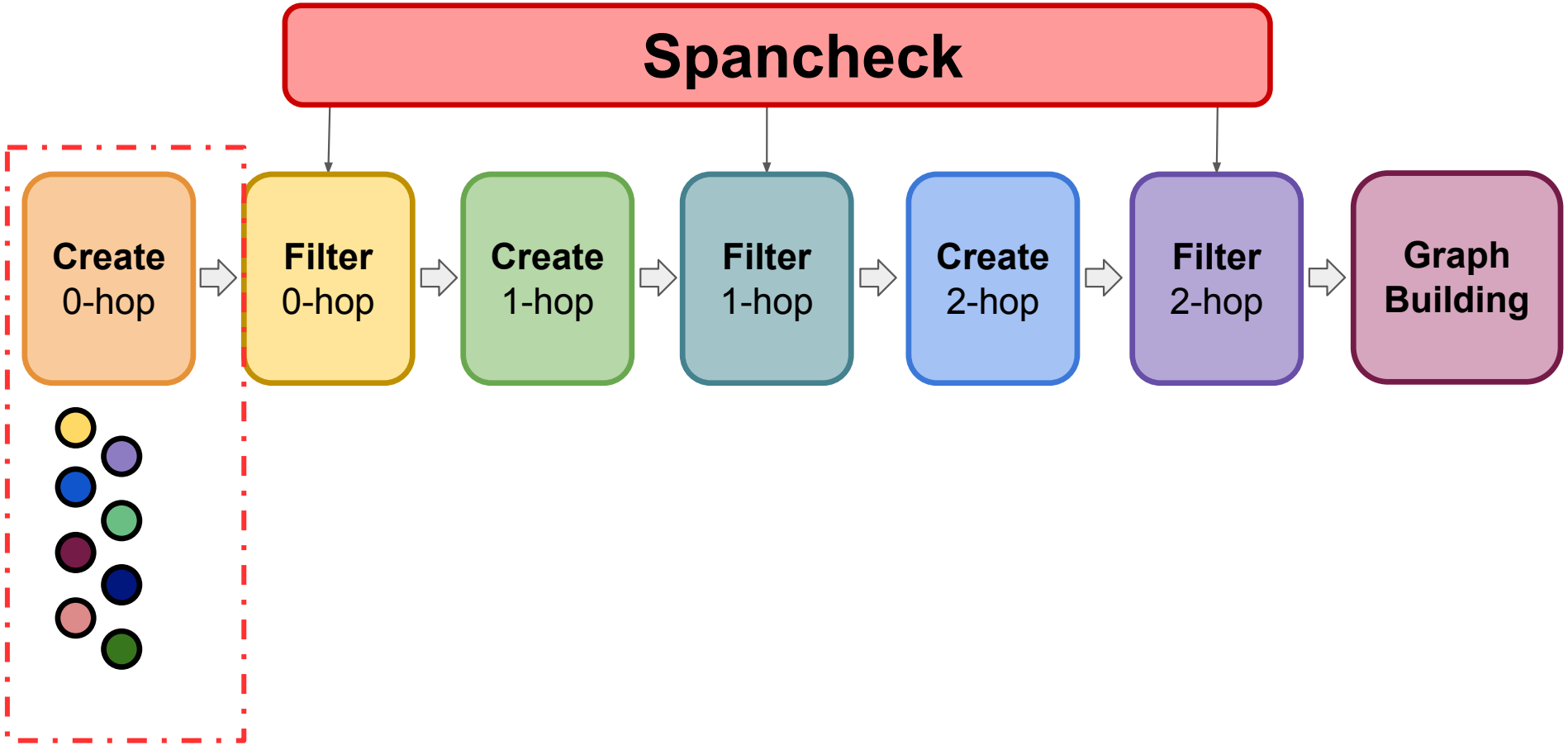




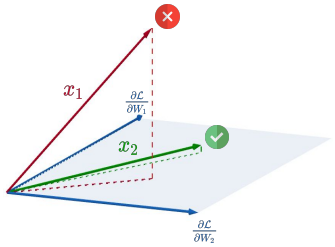
GRAIN Overview



GRAIN Overview

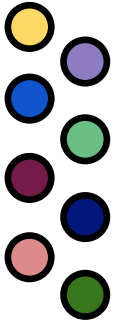


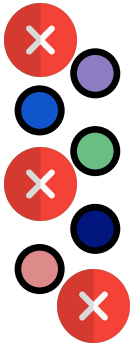
GRAIN Overview



Spancheck

Create 0-hop → **Filter** 0-hop → **Create** 1-hop → **Filter** 1-hop → **Create** 2-hop → **Filter** 2-hop → **Graph Building**





Single-node reconstruction

Spancheck - applied on the gradient of the loss w.r.t the weight matrix W^0 of the first GNN layer

Single-node reconstruction

Spancheck - applied on the gradient of the loss w.r.t the weight matrix W^0 of the first GNN layer

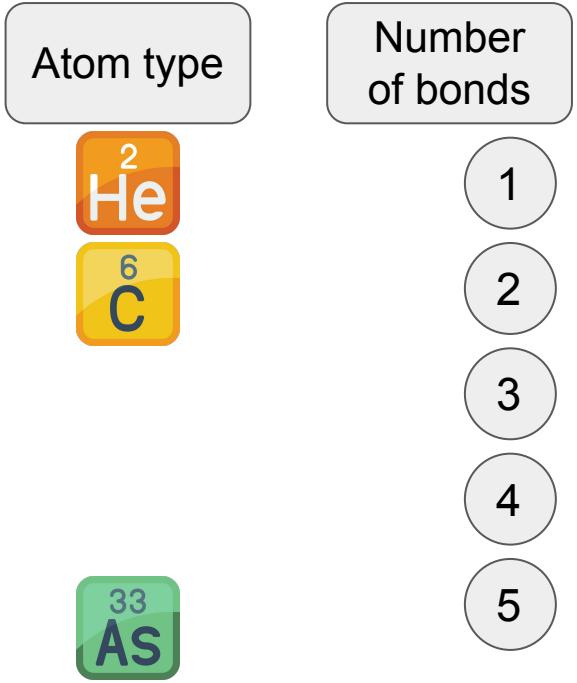
Single-node reconstruction

Spancheck - applied on the gradient of the loss w.r.t the weight matrix W^0 of the first GNN layer

Atom type

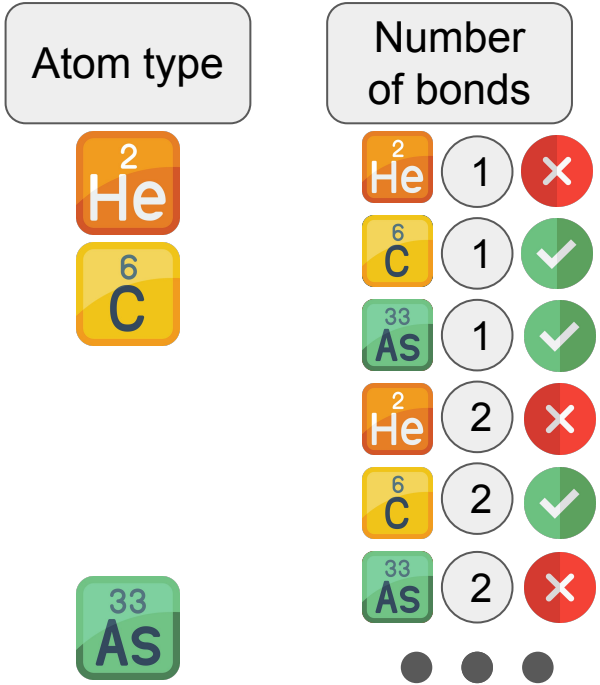
Single-node reconstruction

Spancheck - applied on the gradient of the loss w.r.t the weight matrix W^0 of the first GNN layer



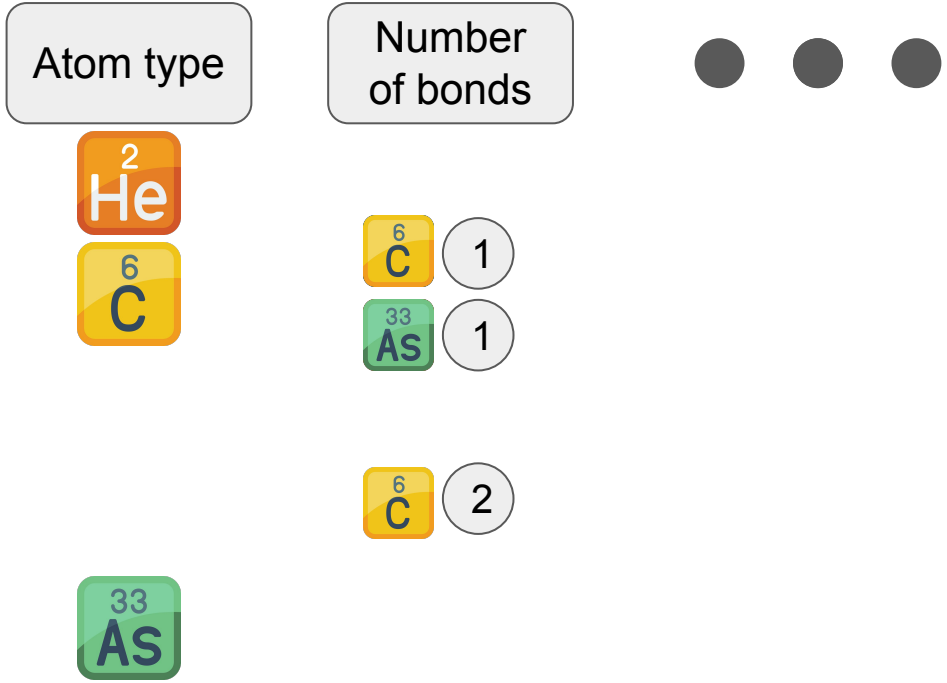
Single-node reconstruction

Spancheck - applied on the gradient of the loss w.r.t the weight matrix W^0 of the first GNN layer

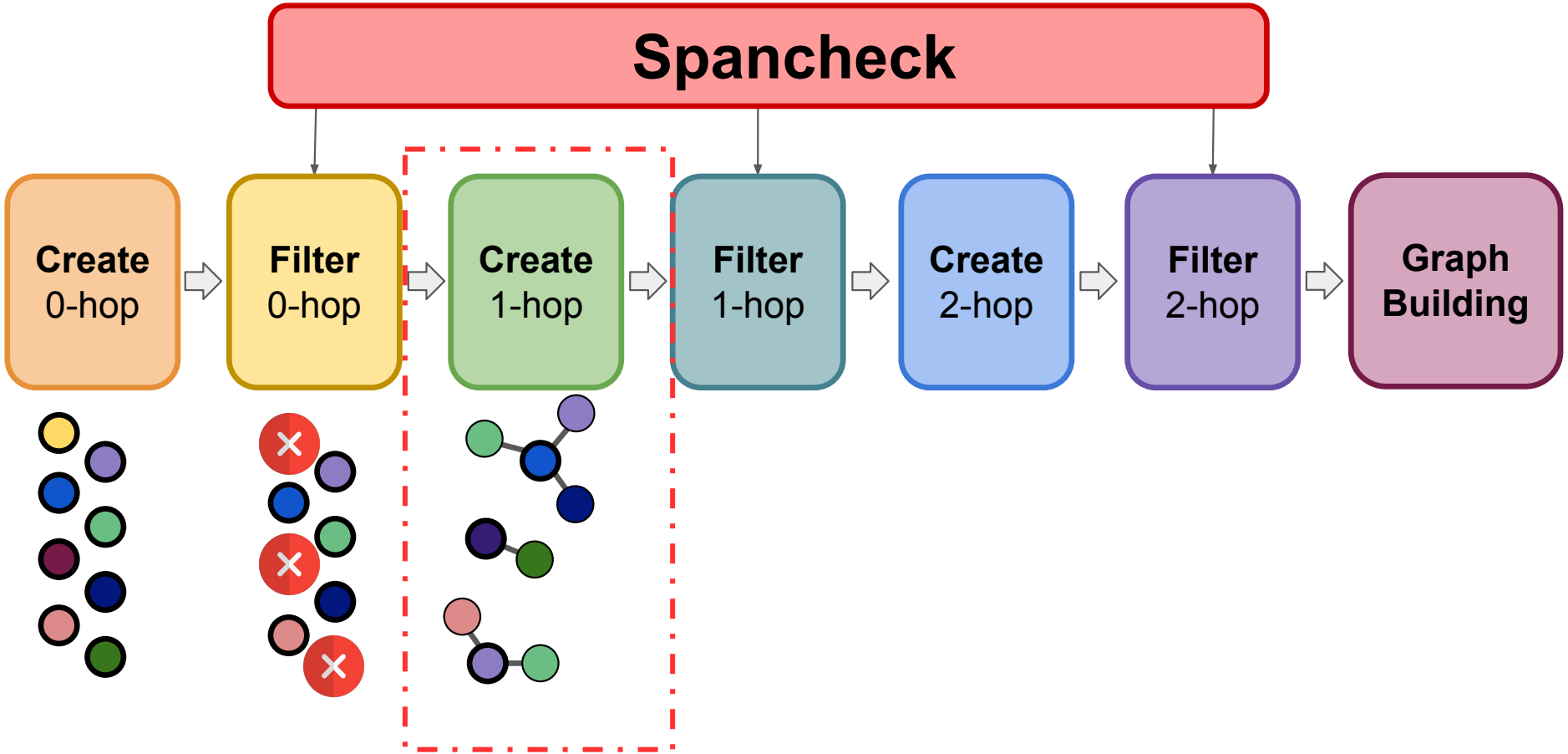


Single-node reconstruction

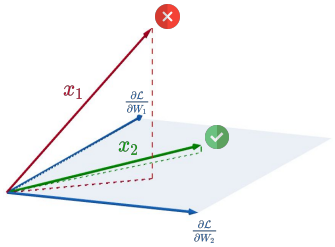
Spancheck - applied on the gradient of the loss w.r.t the weight matrix W^0 of the first GNN layer

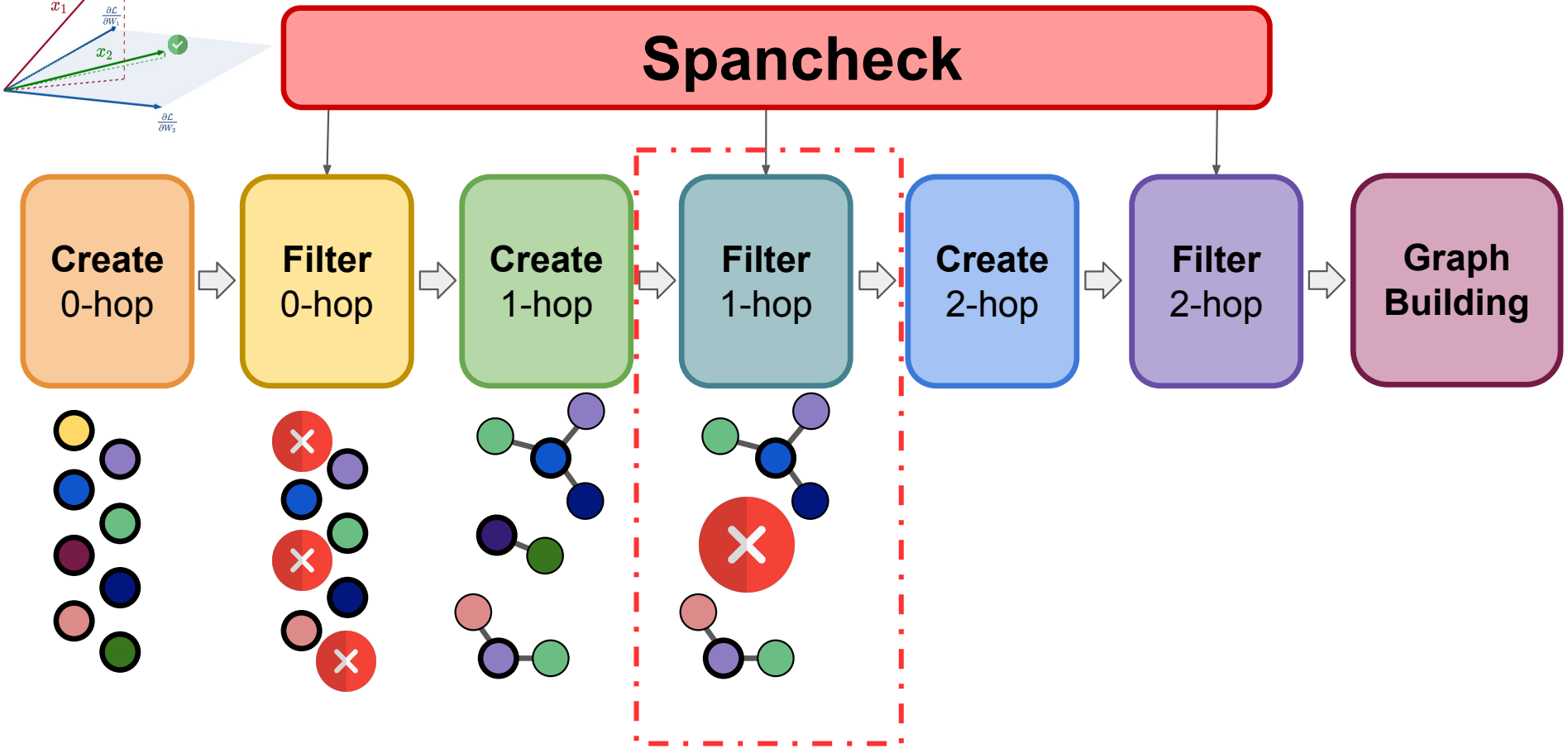


GRAIN Overview

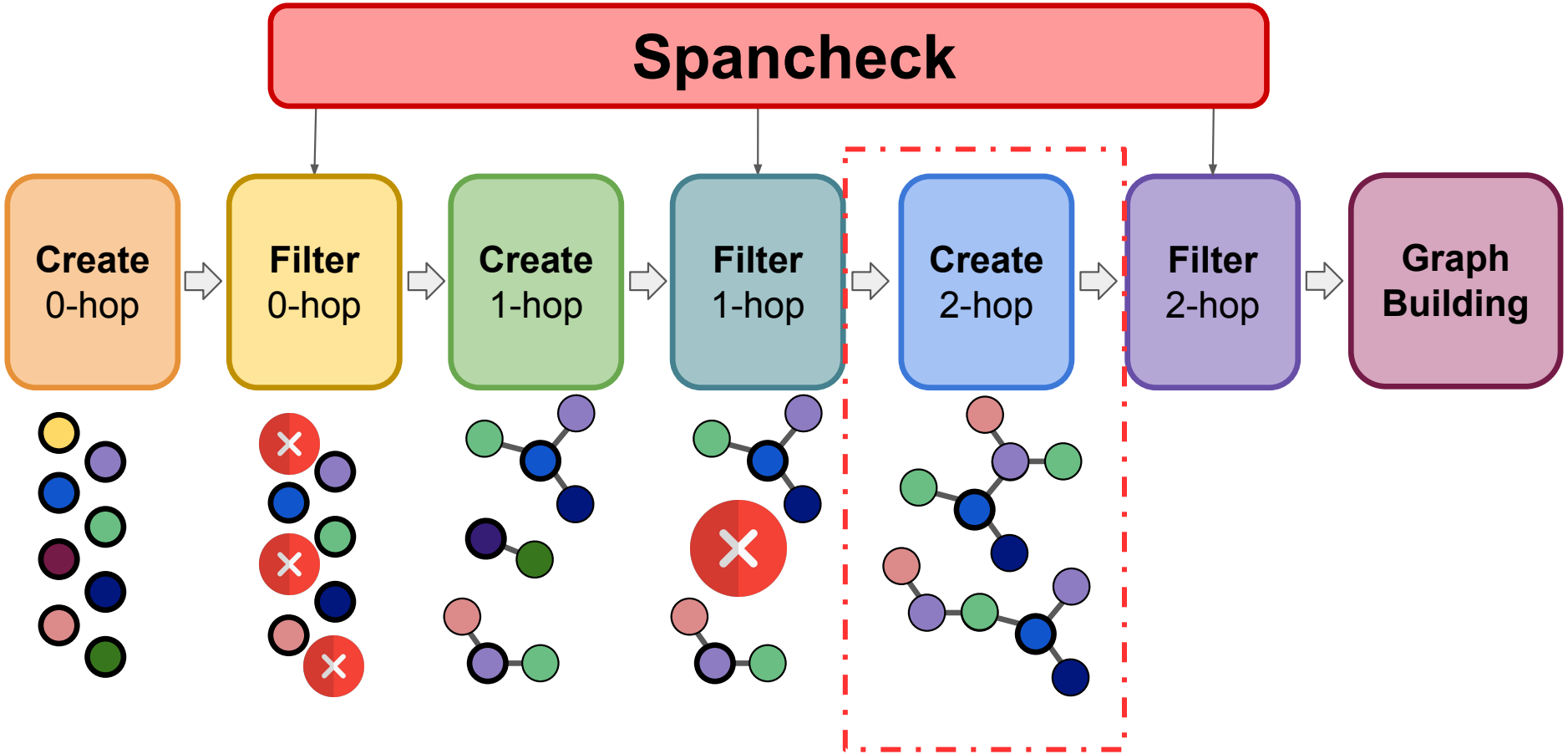


GRAIN Overview

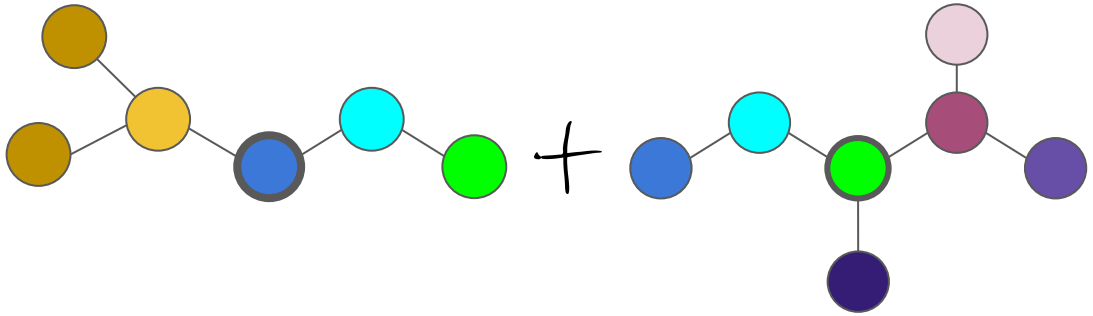




GRAIN Overview

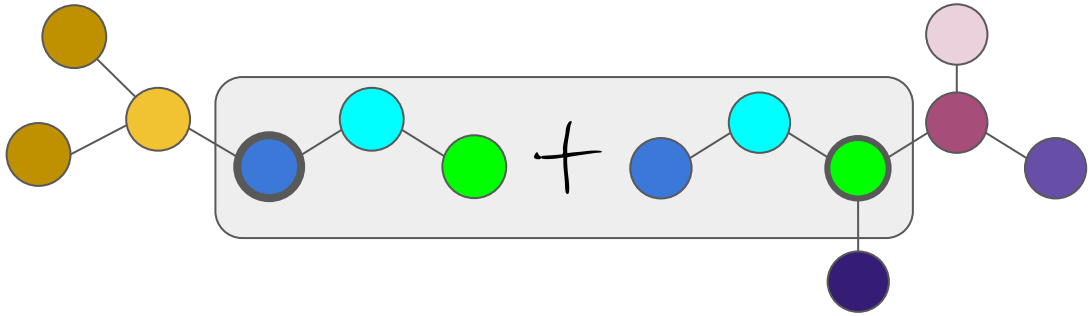


Building Algorithm - Graph gluing



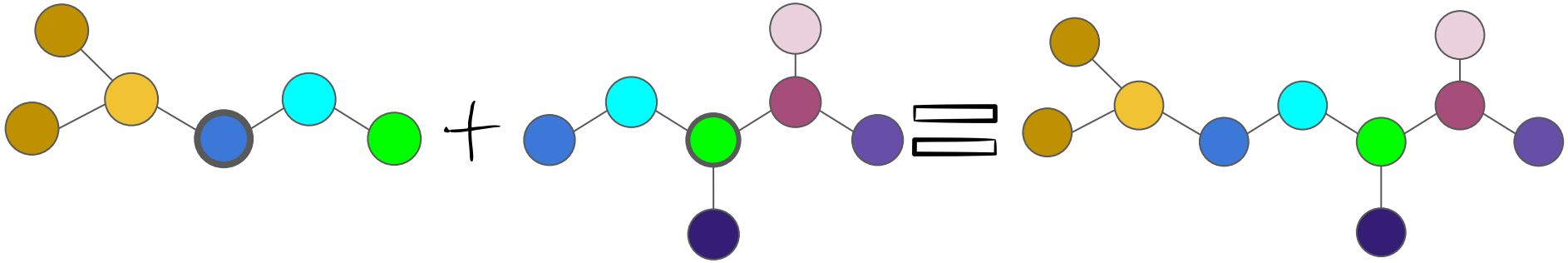
Building Algorithm - Graph gluing

Gluing building blocks into larger graphs

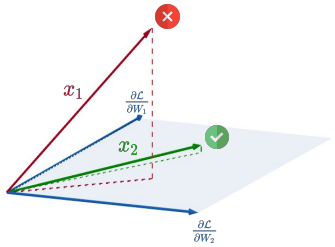


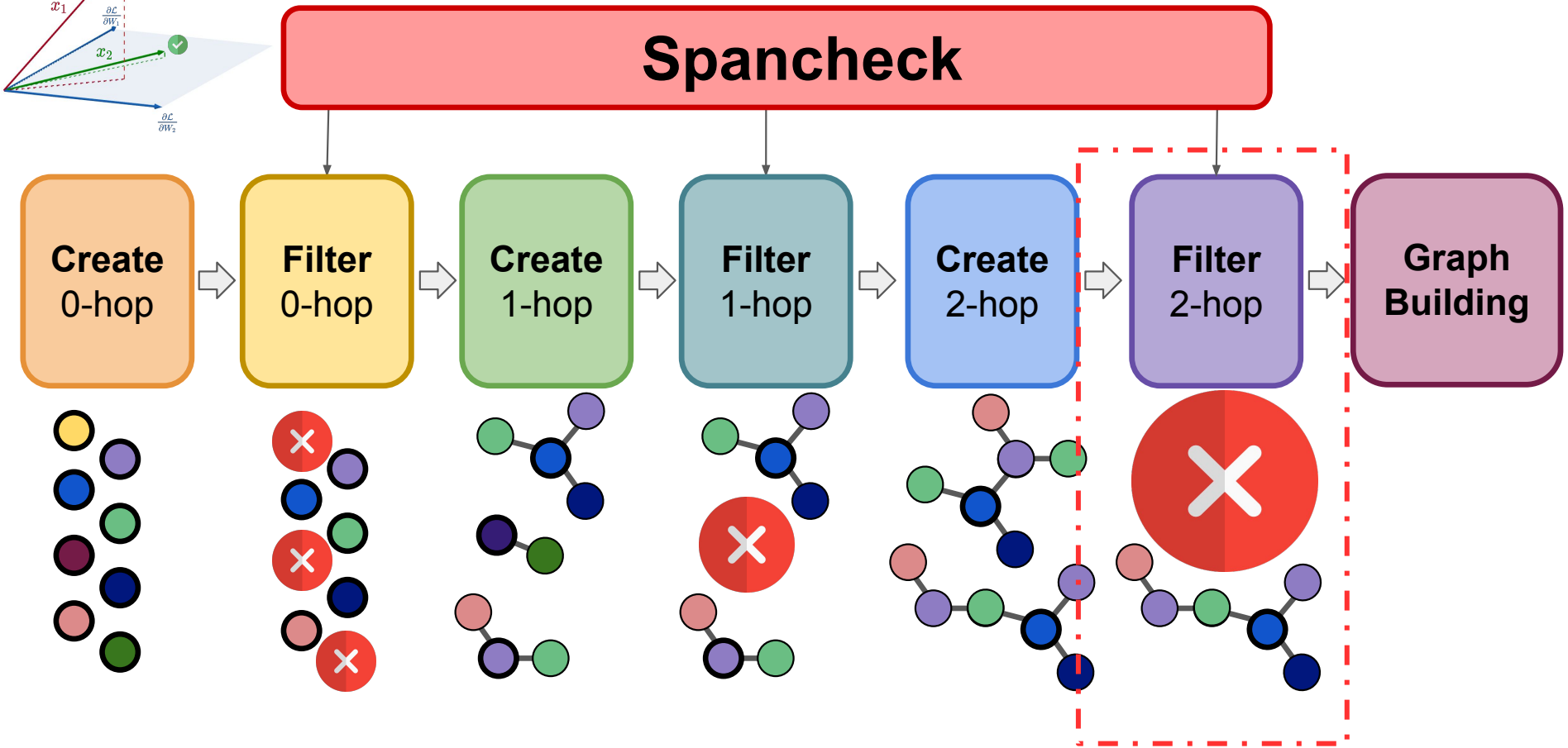
Building Algorithm - Graph gluing

Gluing building blocks into larger graphs

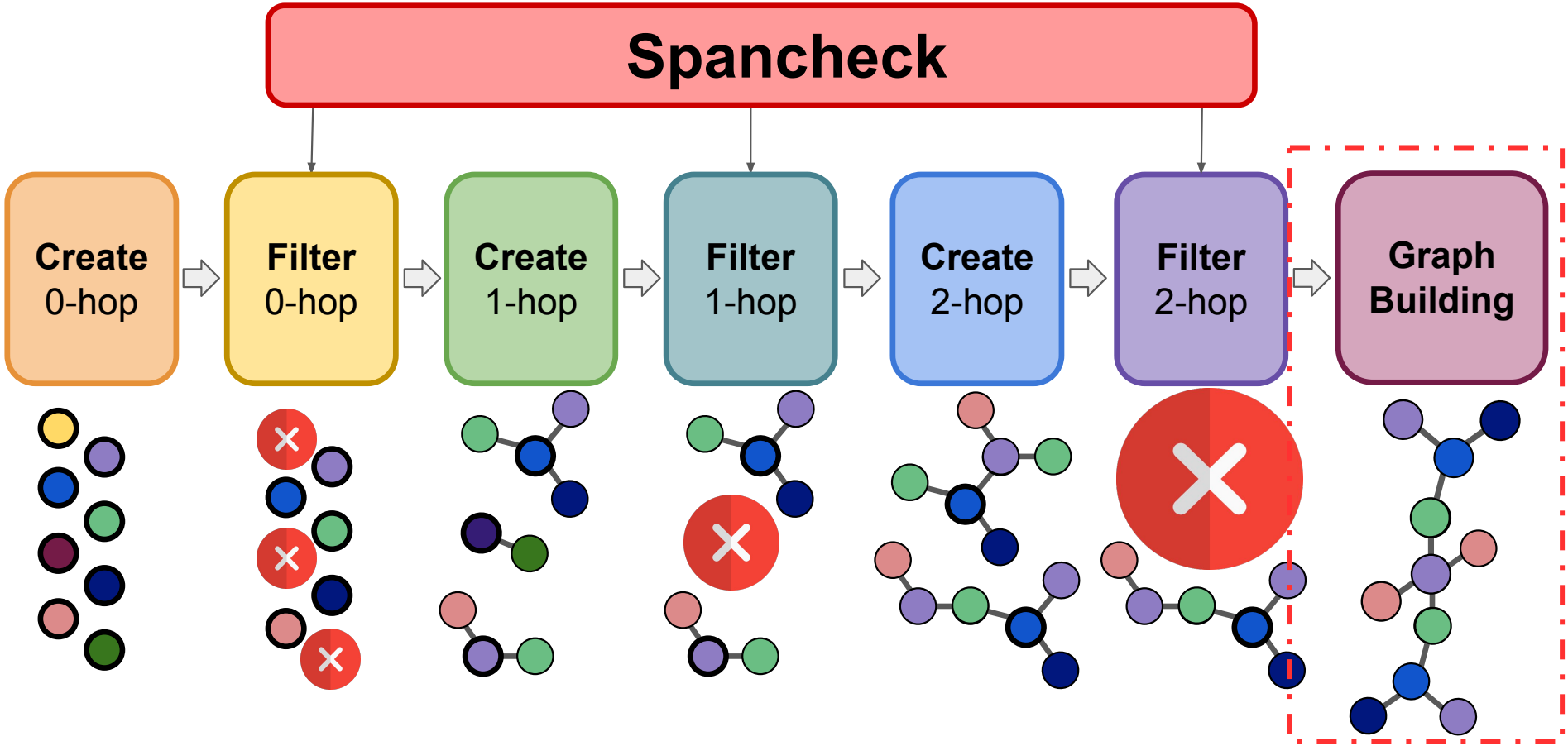


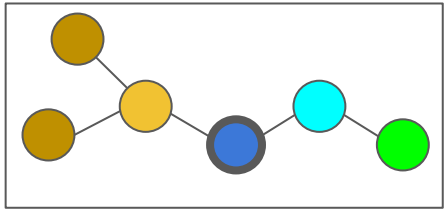
GRAIN Overview

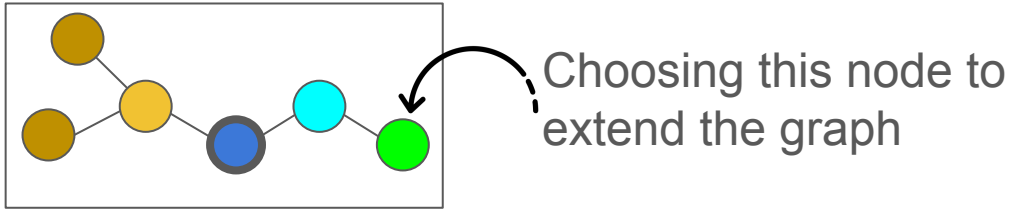


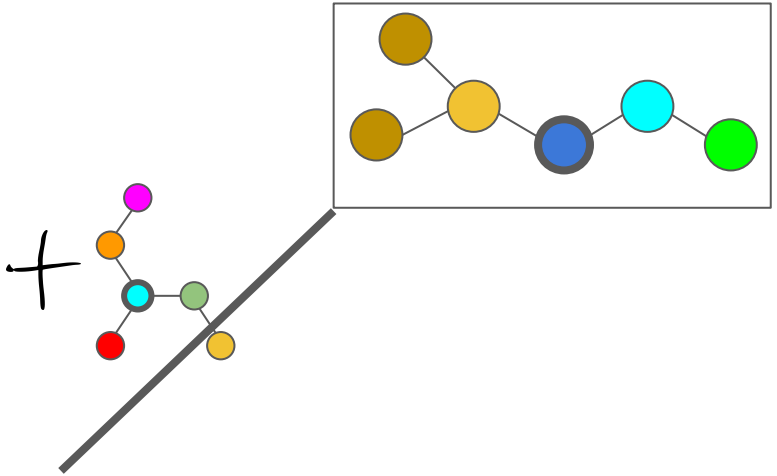


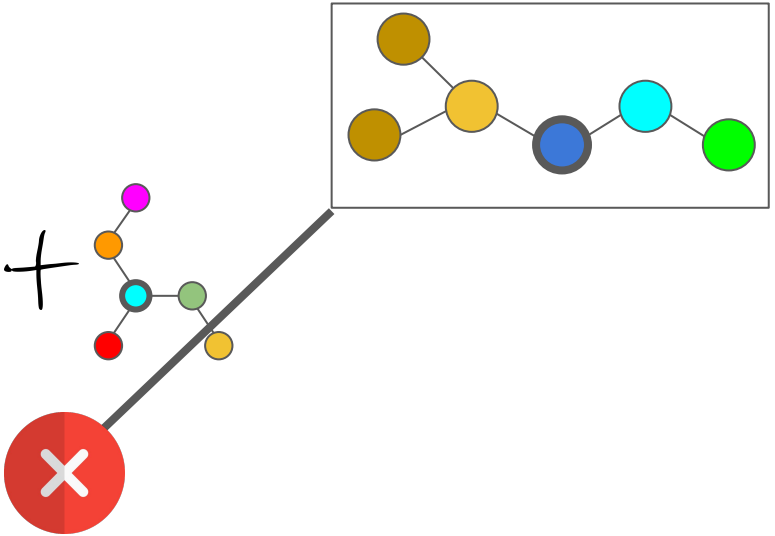
GRAIN Overview

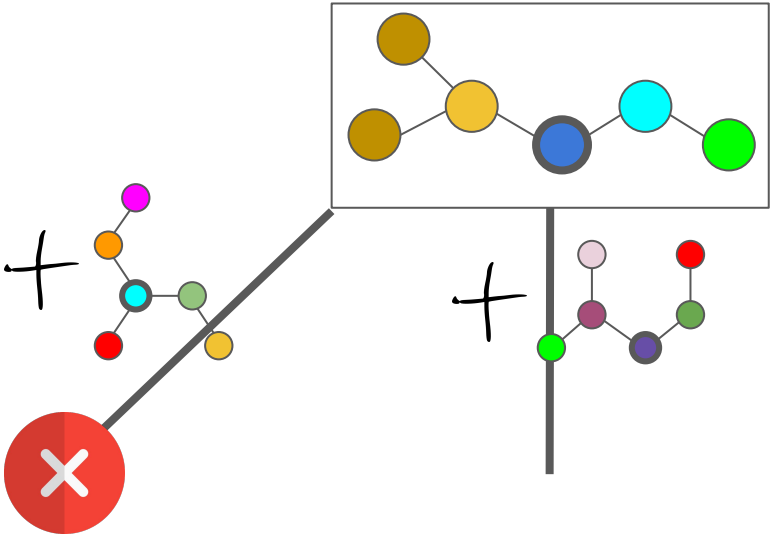


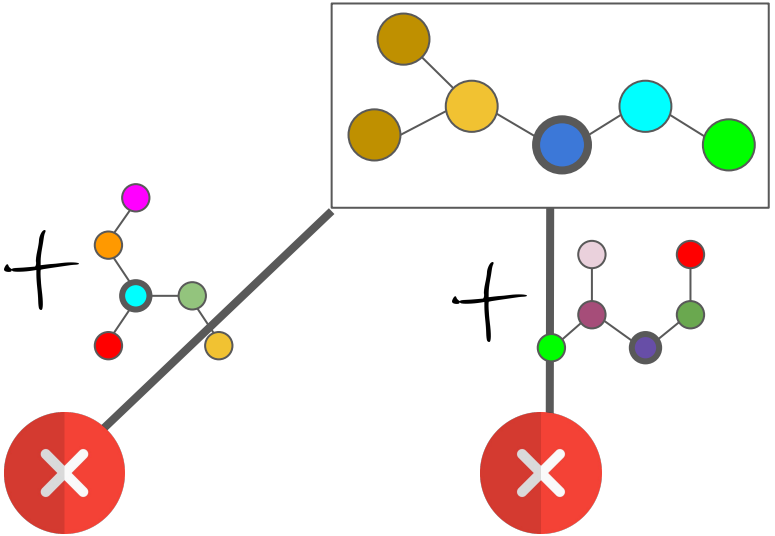


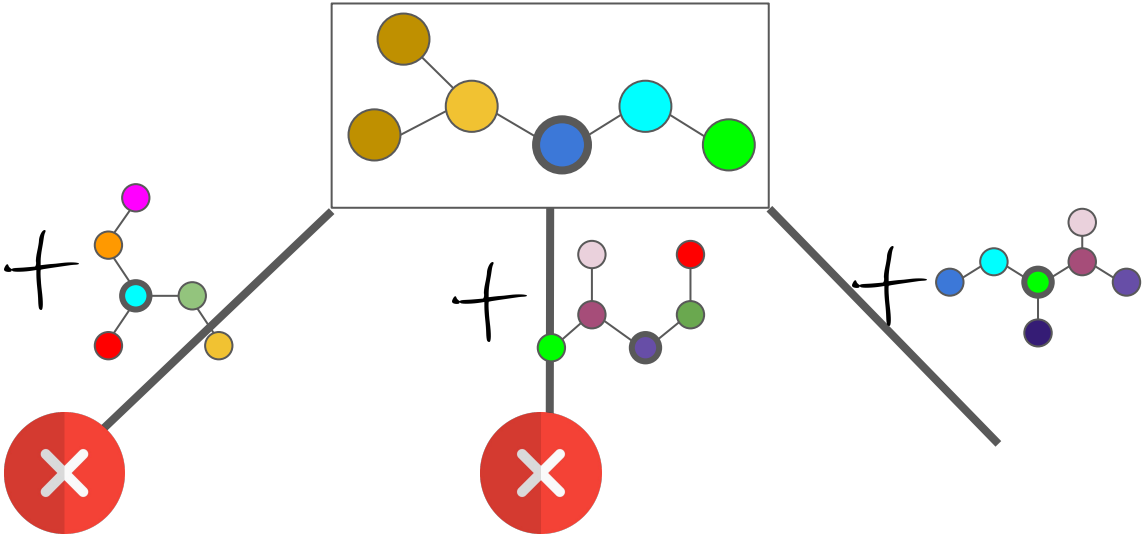

Choosing this node to
extend the graph

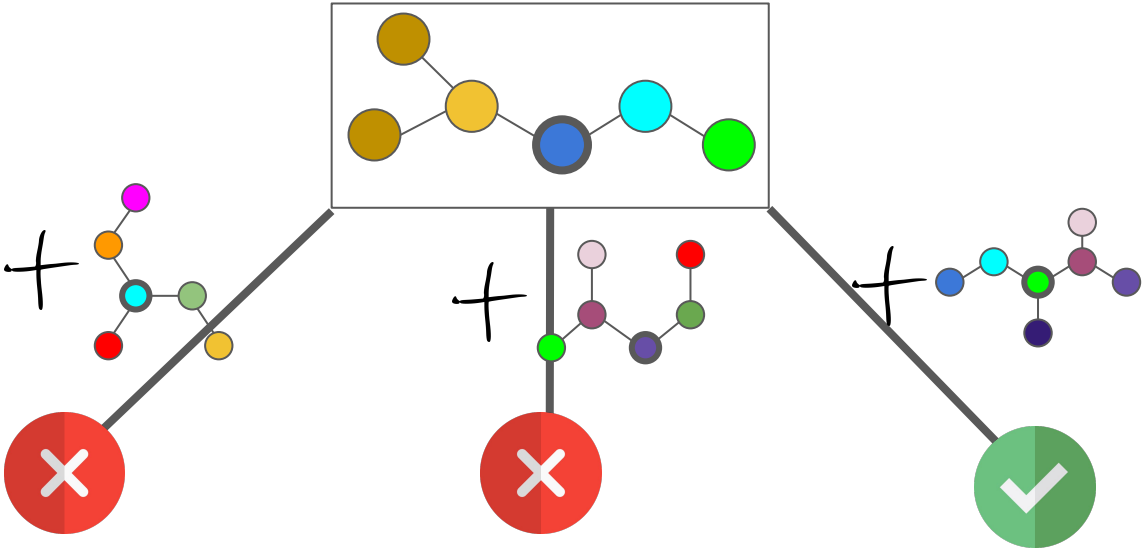


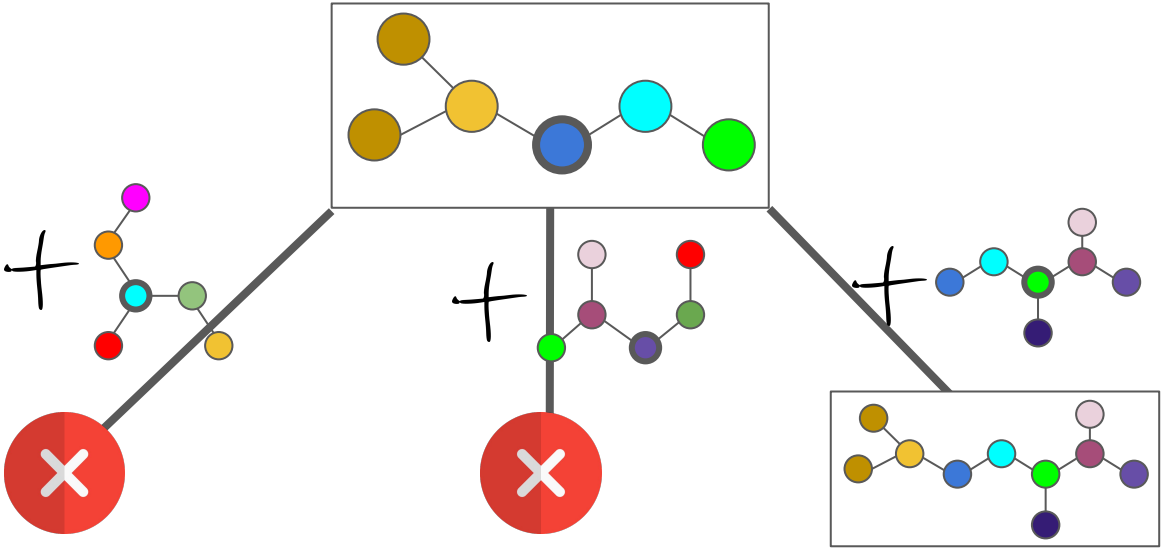


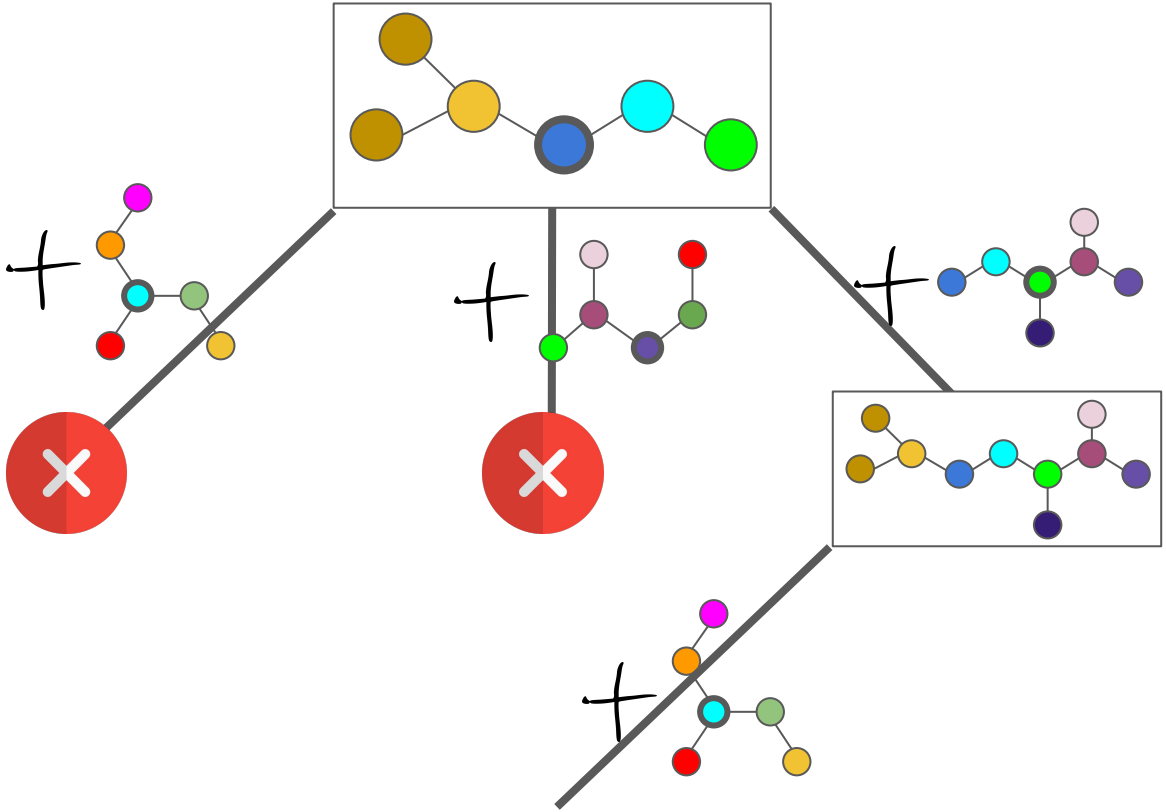


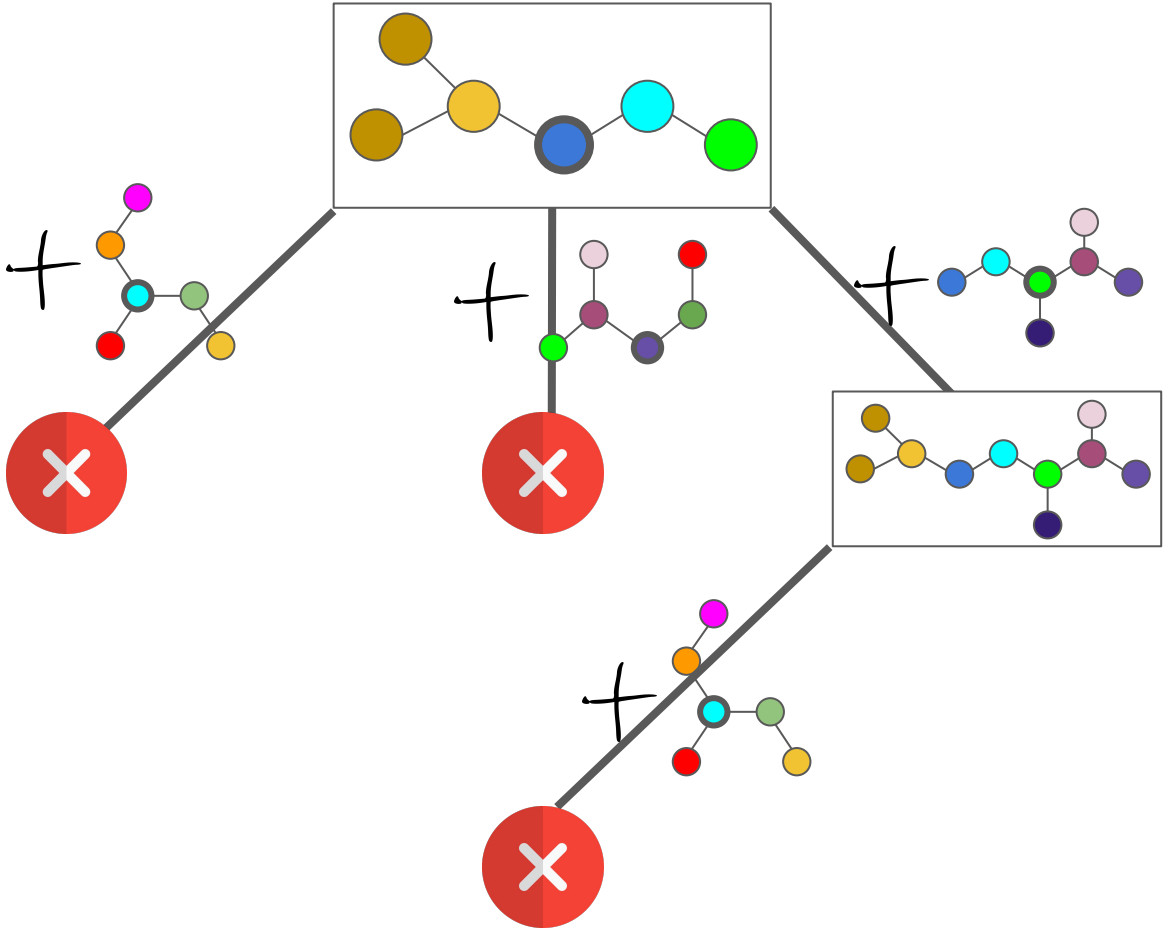


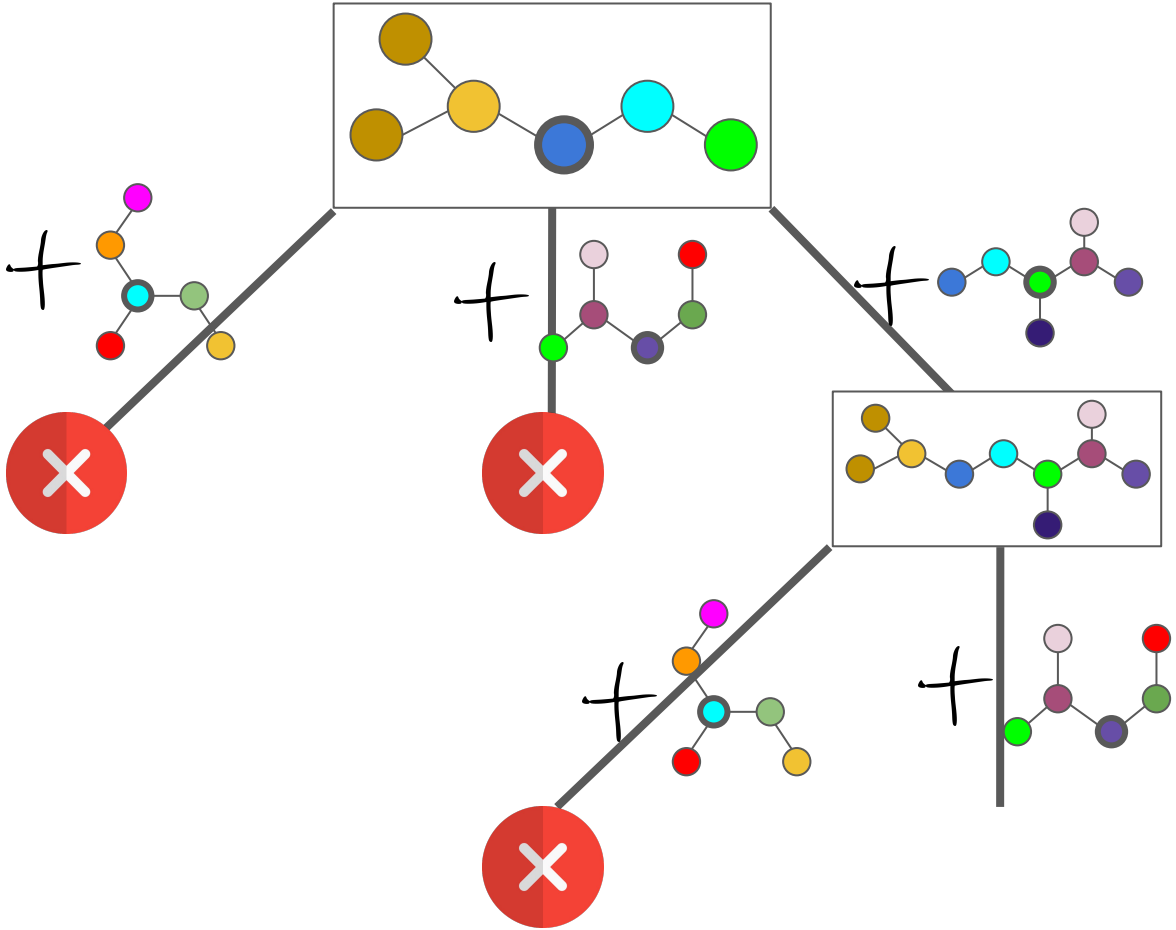


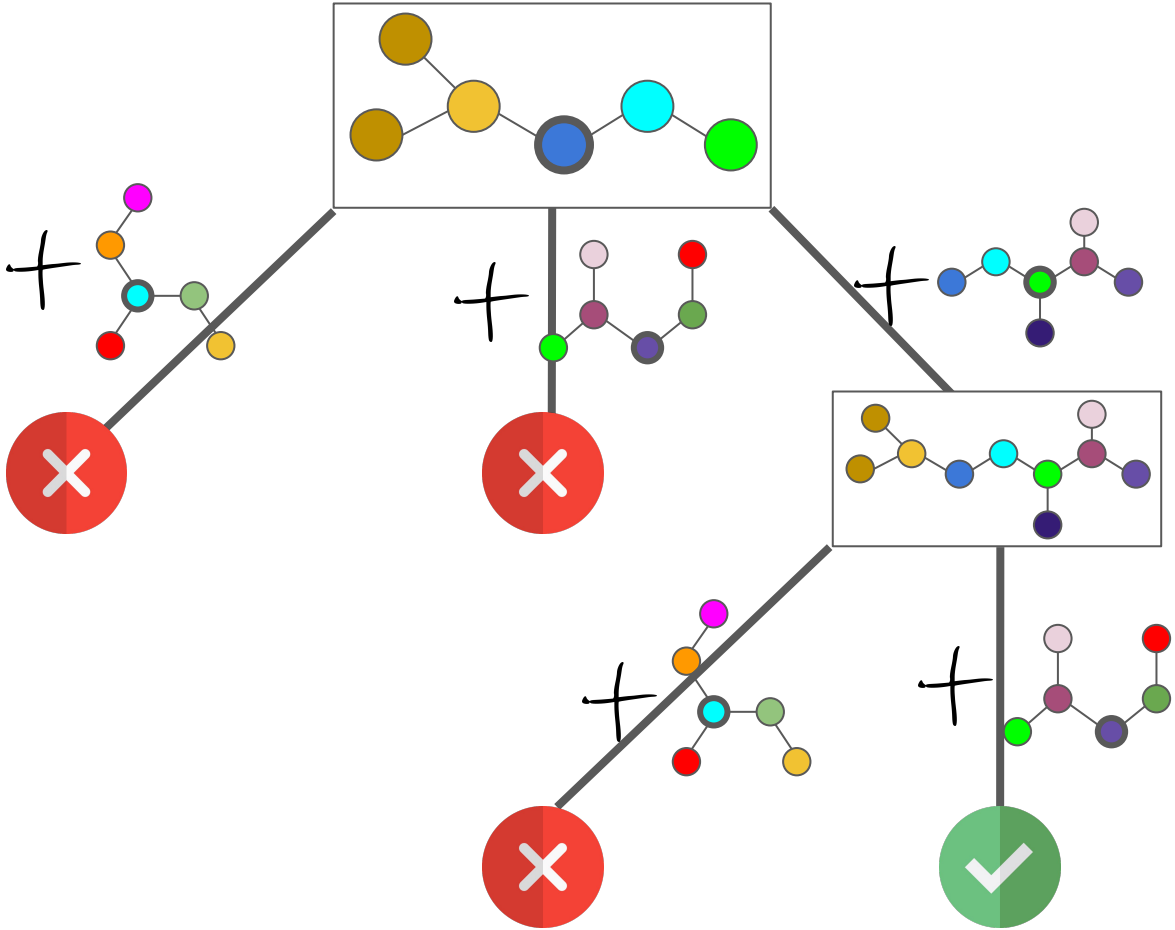


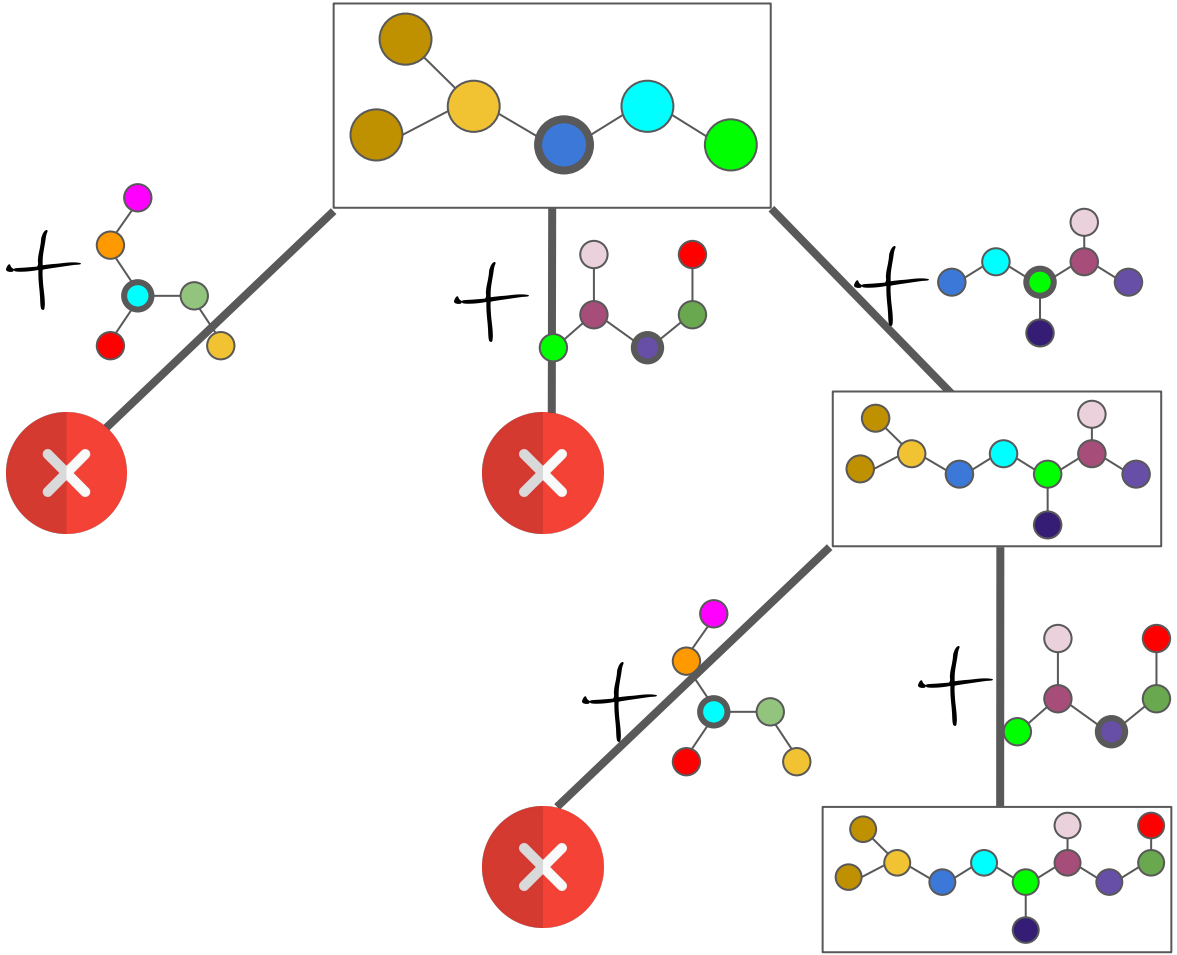












Evaluation - GCNs vs GATs

Baselines: DLG [1] and TabLeak [2] and the same with the adjacency matrix given (+A)

		GCN					GAT				
		GSM-0	GSM-1	GSM-2	FULL	Time [h]	GSM-0	GSM-1	GSM-2	FULL	Time [h]
Tox21	GRAIN	$\mathbf{86.9^{+4.2}_{-5.7}}$	$\mathbf{83.9^{+5.2}_{-6.9}}$	$\mathbf{82.6^{+5.7}_{-7.4}}$	$\mathbf{68.0 \pm 1.7}$	14.3	$92.9^{+3.8}_{-5.8}$	$\mathbf{90.7^{+5.0}_{-7.1}}$	$\mathbf{89.9^{+5.8}_{-7.2}}$	$\mathbf{75.0 \pm 1.8}$	10.8
	DLG	$31.8^{+4.5}_{-4.3}$	$20.3^{+5.5}_{-4.8}$	$22.8^{+6.6}_{-5.6}$	1.0 ± 0.2	3.3	96.0 ± 0.32	$9.3^{+4.4}_{-4.9}$	$6.5^{+3.9}_{-4.1}$	2.0 ± 0.3	**4.2**
	DLG $+A$	$54.7^{+3.9}_{-4.2}$	$60.1^{+4.6}_{-5.2}$	$76.7^{+3.6}_{-4.8}$	1.0 ± 0.2	**3.1**	$\mathbf{96.5 \pm 0.34}$	$69.7^{+4.1}_{-4.2}$	$81.3^{+3.4}_{-3.6}$	2.0 ± 0.3	4.5
	TabLeak	$25.1^{+5.1}_{-4.3}$	$12.4^{+5.5}_{-4.3}$	$10.8^{+5.6}_{-3.9}$	1.0 ± 0.2	13.1	$73.7^{+2.6}_{-2.0}$	$7.2^{+5.2}_{-4.9}$	10.0 ± 4.8	1.0 ± 0.2	6.0
	TabLeak $+A$	$55.6^{+3.9}_{-3.9}$	$57.7^{+4.1}_{-4.6}$	$73.8^{+2.8}_{-3.5}$	1.0 ± 0.2	12.3	$75.1^{+2.5}_{-1.9}$	$74.9^{+2.1}_{-1.9}$	$84.2^{+1.5}_{-1.3}$	1.0 ± 0.2	6.0

[1] Zhu et. al. in “Deep Leakage from Gradients”
[2] Vero et. al. in “TabLeak: Tabular Data Leakage in Federated Learning”

Evaluation – GRAIN on different domains

			GAT				
			GSM-0	GSM-1	GSM-2	FULL	Min/Rec
Citation Network	CiteSeer	GRAIN	$\mathbf{79.3^{+4.7}_{-6.3}}$	$\mathbf{69.1^{+6.1}_{-6.4}}$	$\mathbf{69.6^{+6.2}_{-6.0}}$	$\mathbf{61.0 \pm 1.6}$	**0.8**
		DLG	$67.7^{+3.9}_{-3.7}$	$0.0^{+0.0}_{-0.0}$	$0.0^{+0.0}_{-0.0}$	0.0 ± 0.0	31.0
		DLG $+A$	$67.7^{+4.0}_{-3.7}$	$0.0^{+0.0}_{-0.0}$	$0.0^{+0.0}_{-0.0}$	0.0 ± 0.0	27.7
		TabLeak	$67.7^{+3.9}_{-3.8}$	$0.0^{+0.0}_{-0.0}$	$0.0^{+0.0}_{-0.0}$	0.0 ± 0.0	153.0
		TabLeak $+A$	$67.7^{+4.0}_{-3.7}$	$0.0^{+0.0}_{-0.0}$	$0.0^{+0.0}_{-0.0}$	0.0 ± 0.0	148.7
Social Network	Pokec	GRAIN	$\mathbf{97.2^{+1.6}_{-1.9}}$	$\mathbf{93.5^{+3.4}_{-4.2}}$	$\mathbf{96.3^{+1.9}_{-2.3}}$	$\mathbf{79.0 \pm 1.8}$	**0.2**
		DLG	$44.7^{+2.3}_{-2.3}$	$2.2^{+3.1}_{-2.2}$	$0.0^{+0.0}_{-0.0}$	0.0 ± 0.0	26.3
		DLG $+A$	$57.4^{+3.7}_{-3.9}$	$69.5^{+3.6}_{-4.0}$	$88.6^{+2.0}_{-2.1}$	0.0 ± 0.0	21.6
		TabLeak	$50.8^{+12.4}_{-8.9}$	$13.9^{+13.5}_{-12.3}$	$7.9^{+11.9}_{-7.9}$	0.0 ± 0.0	204.5
		TabLeak $+A$	$52.6^{+3.3}_{-3.3}$	$68.1^{+4.1}_{-3.9}$	$82.7^{+4.0}_{-4.9}$	0.0 ± 0.0	254.5
Chemical dataset	Tox21	GRAIN	$92.9^{+3.8}_{-5.8}$	$\mathbf{90.7^{+5.0}_{-7.1}}$	$\mathbf{89.9^{+5.8}_{-7.2}}$	$\mathbf{75.0 \pm 1.8}$	10.8
		DLG	96.0 ± 0.32	$9.3^{+4.4}_{-4.9}$	$6.5^{+3.9}_{-4.1}$	2.0 ± 0.3	**4.2**
		DLG $+A$	$\mathbf{96.5 \pm 0.34}$	$69.7^{+4.1}_{-4.2}$	$81.3^{+3.4}_{-3.6}$	2.0 ± 0.3	4.5
		TabLeak	$73.7^{+2.6}_{-2.0}$	$7.2^{+5.2}_{-4.9}$	10.0 ± 4.8	1.0 ± 0.2	6.0
		TabLeak $+A$	$75.1^{+2.5}_{-1.9}$	$74.9^{+2.1}_{-1.9}$	$84.2^{+1.5}_{-1.3}$	1.0 ± 0.2	6.0

Evaluation - Model Width and Depth

		GSM-0	GSM-1	GSM-2	FULL
$L = 2$, $d' = 300$ (default)	GRAIN	$\mathbf{86.9^{+4.2}_{-5.7}}$	$\mathbf{83.9^{+5.2}_{-6.9}}$	$\mathbf{82.6^{+5.7}_{-7.4}}$	$\mathbf{68.0 \pm 1.7}$
	DLG	$31.8^{+4.5}_{-4.3}$	$20.3^{+5.5}_{-4.8}$	$22.8^{+6.6}_{-5.6}$	1.0 ± 0.2
	DLG $+A$	$54.7^{+3.9}_{-4.2}$	$60.1^{+4.6}_{-5.2}$	$76.7^{+3.6}_{-4.8}$	1.0 ± 0.2
	TabLeak	$25.1^{+5.1}_{-4.3}$	$12.4^{+5.5}_{-4.3}$	$10.8^{+5.6}_{-3.9}$	1.0 ± 0.2
	TabLeak $+A$	$55.6^{+3.9}_{-3.9}$	$57.7^{+4.1}_{-4.6}$	$73.8^{+2.8}_{-3.5}$	1.0 ± 0.2
$L = 3$, $d' = 300$	GRAIN	$\mathbf{82.5^{+5.7}_{-7.7}}$	$\mathbf{80.7^{+6.3}_{-7.7}}$	$\mathbf{80.4^{+6.2}_{-7.8}}$	$\mathbf{63.0 \pm 1.6}$
	DLG	$20.3^{+4.3}_{-3.4}$	$7.8^{+5.1}_{-3.3}$	$8.2^{+5.3}_{-3.4}$	1.0 ± 0.2
	DLG $+A$	$43.0^{+3.7}_{-3.6}$	$48.0^{+4.3}_{-4.5}$	$66.0^{+3.7}_{-4.6}$	1.0 ± 0.2
	TabLeak	$16.5^{+3.8}_{-2.9}$	$8.8^{+4.4}_{-3.1}$	$8.0^{+4.3}_{-3.0}$	1.0 ± 0.2
	TabLeak $+A$	$47.5^{+4.0}_{-4.2}$	$48.1^{+4.8}_{-5.0}$	$62.9^{+4.3}_{-4.4}$	1.0 ± 0.2
$L = 2$, $d' = 200$	GRAIN	$\mathbf{84.6^{+4.6}_{-6.4}}$	$\mathbf{81.4^{+5.8}_{-6.9}}$	$\mathbf{80.5^{+5.9}_{-7.2}}$	$\mathbf{62.0 \pm 1.6}$
	DLG	$30.8^{+4.5}_{-4.1}$	$18.9^{+5.8}_{-4.9}$	$22.2^{+6.7}_{-5.4}$	1.0 ± 0.2
	DLG $+A$	$50.3^{+4.2}_{-4.2}$	$53.4^{+5.3}_{-5.9}$	$68.7^{+4.9}_{-6.1}$	3.0 ± 0.4
	TabLeak	$22.1^{+4.8}_{-3.7}$	$10.3^{+5.3}_{-3.6}$	$8.9^{+5.5}_{-3.6}$	1.0 ± 0.2
	TabLeak $+A$	$55.0^{+4.8}_{-5.0}$	$62.1^{+4.9}_{-5.9}$	$76.7^{+3.6}_{-4.7}$	1.0 ± 0.2

Evaluation - Miscellaneous Settings

	GSM-0	GSM-1	GSM-2	FULL
Default	$86.9^{+4.2}_{-5.7}$	$83.9^{+5.2}_{-6.9}$	$82.6^{+5.7}_{-7.4}$	$\mathbf{68.0 \pm 1.7}$
σ = GELU	$82.0^{+5.3}_{-6.7}$	$79.1^{+6.0}_{-7.4}$	$78.4^{+6.2}_{-8.0}$	61.0 ± 1.6
Pre-trained	$73.5^{+6.4}_{-7.4}$	$70.0^{+7.3}_{-7.7}$	$68.6^{+7.6}_{-8.3}$	49.0 ± 1.4
Node Class.	$\mathbf{88.0^{+3.8}_{-5.4}}$	$\mathbf{85.5^{+4.6}_{-6.5}}$	$\mathbf{84.9^{+5.0}_{-6.6}}$	66.0 ± 1.6

Further details can be found in the paper.

OpenReview

Homepage

Code