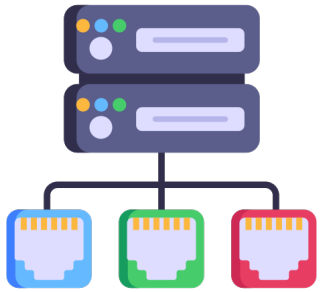
IEEE INFOCOM

CiDer: A Black-box Approach to Classify Node with Certified Robustness Guarantees

Xiaoyu Liang, Haohua Du, Wen Ma, Ye Tian, and Xiaoya Xu

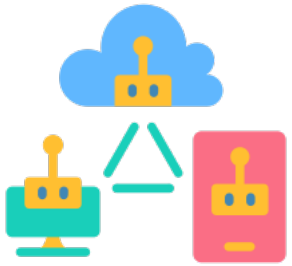
May 21

Graph-structured data

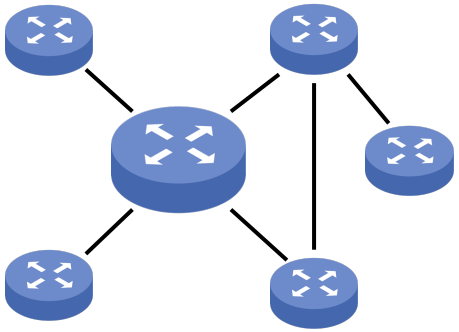


Router network

IoT



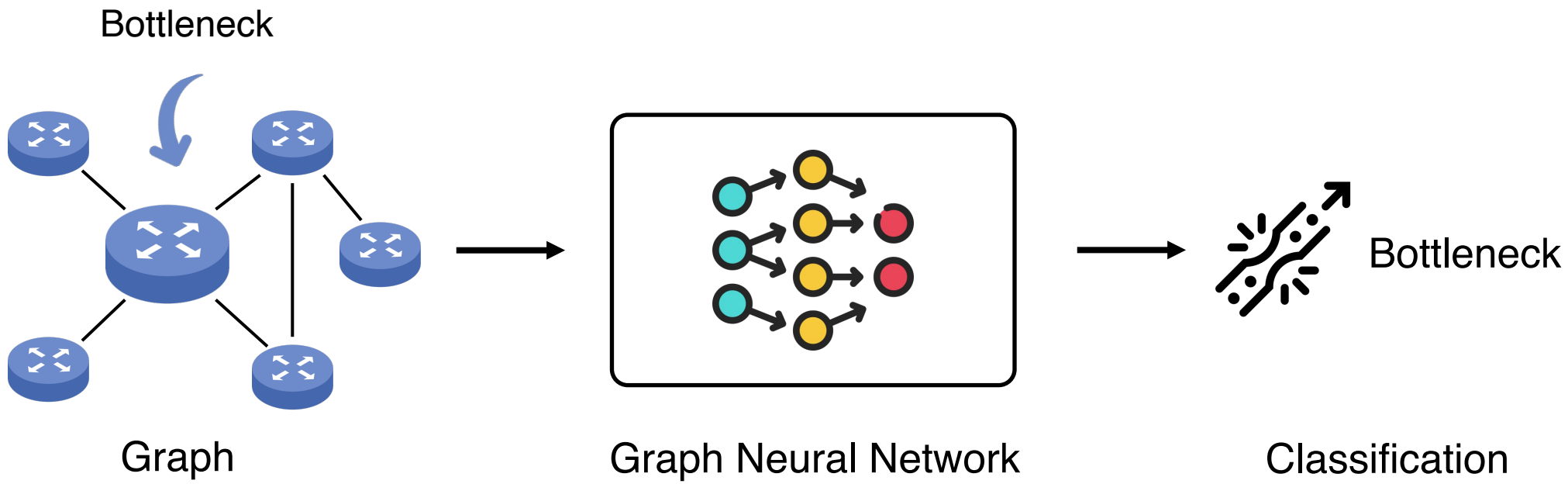
Edge computing



Graph

Real-world relationships are often abstracted into **graphs.**

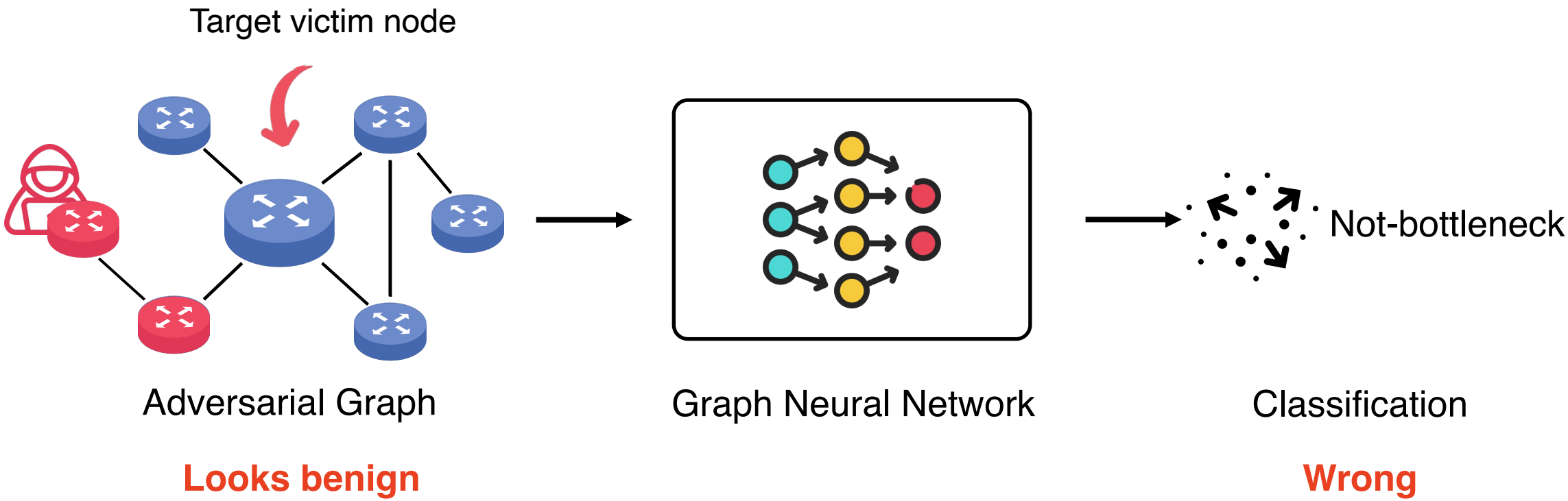
Graph Node Classification



Node classification

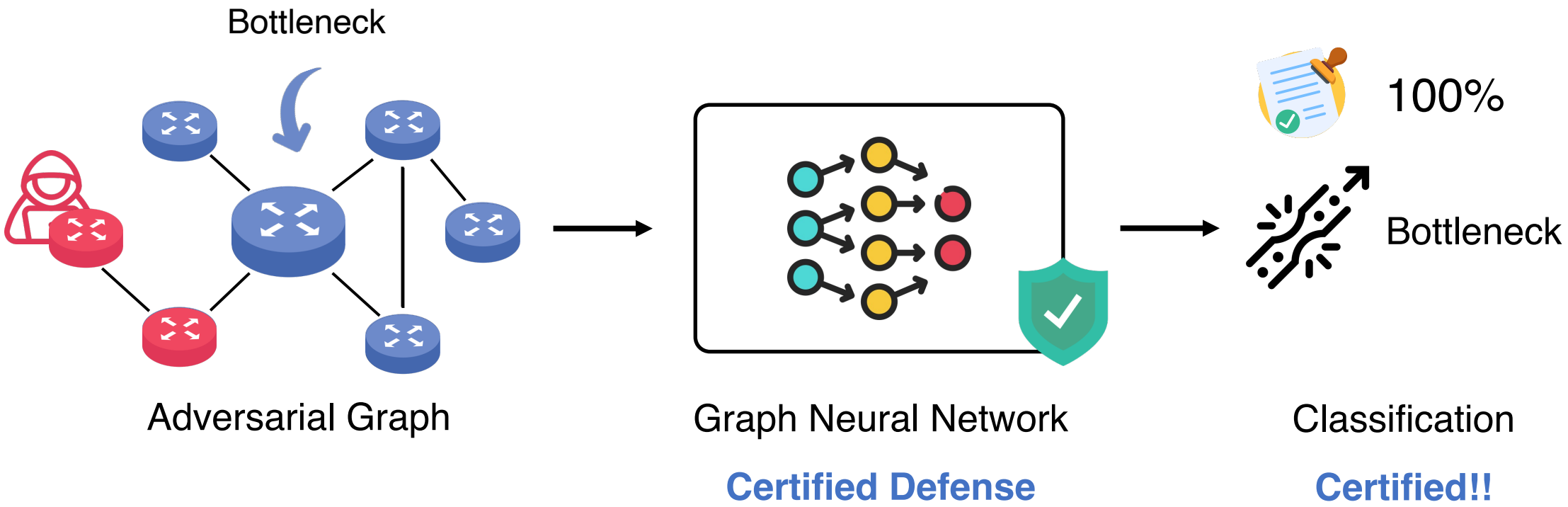
Graph neural network (**GNN**) is used for **node classification**.

Graph Evasion Attack



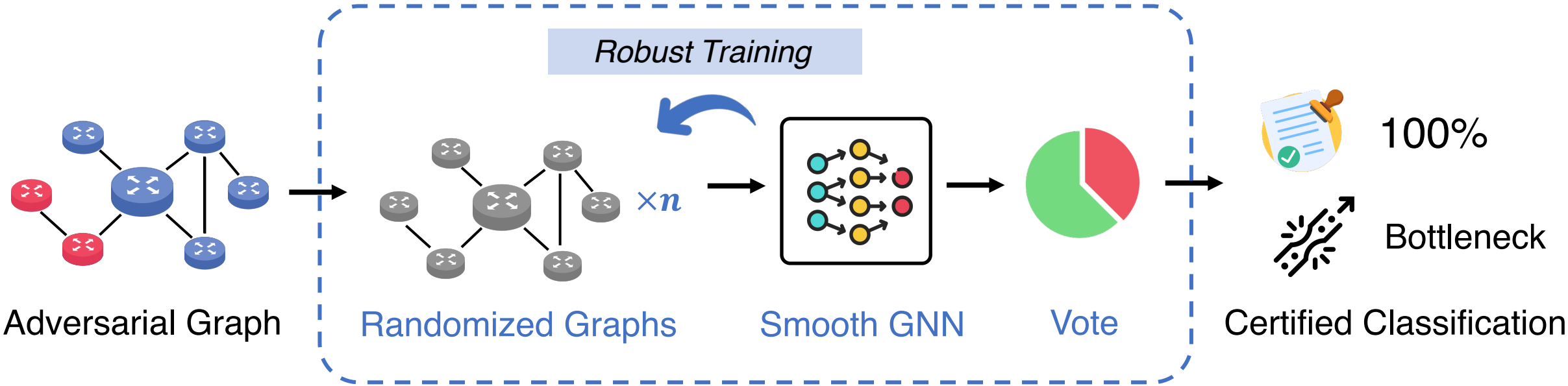
Attackers on the graph network could lead to misclassification by GNN.

Robustness Certificate



Robustness certificate gives provable robustness guarantee of the classification.

Randomized Smoothing is successful, but



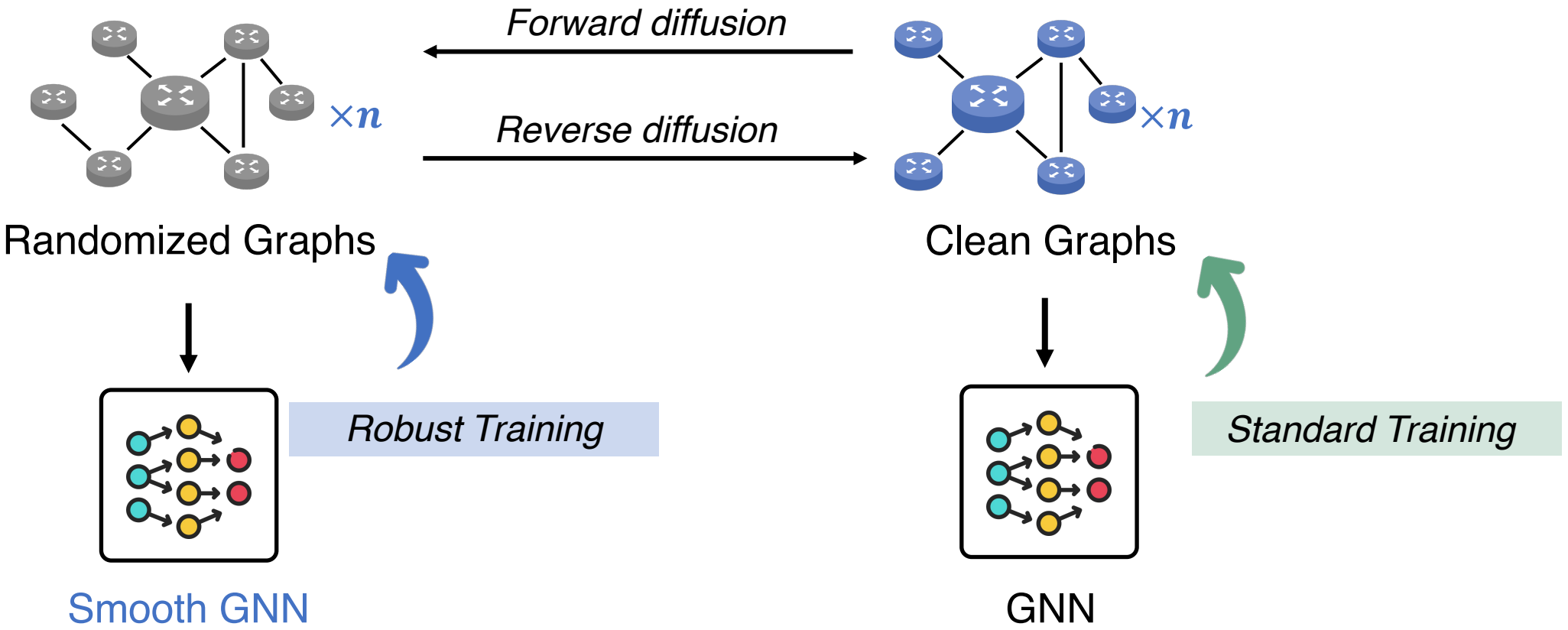
Randomized Smoothing

(Cohen *et al.* 2019, Bojchevski *et al.* 2020)

1. Need **fine-tuning** on the noisy data, and 2. **Accuracy drop** on clean samples.

Cohen *et al.*, "Certified Adversarial Robustness via Randomized Smoothing", in ICML 2019.
Bojchevski *et al*, "Efficient Robustness Certificates for Discrete Data: Sparsity-Aware Randomized Smoothing for Graphs, Images and More", in ICML 2020.

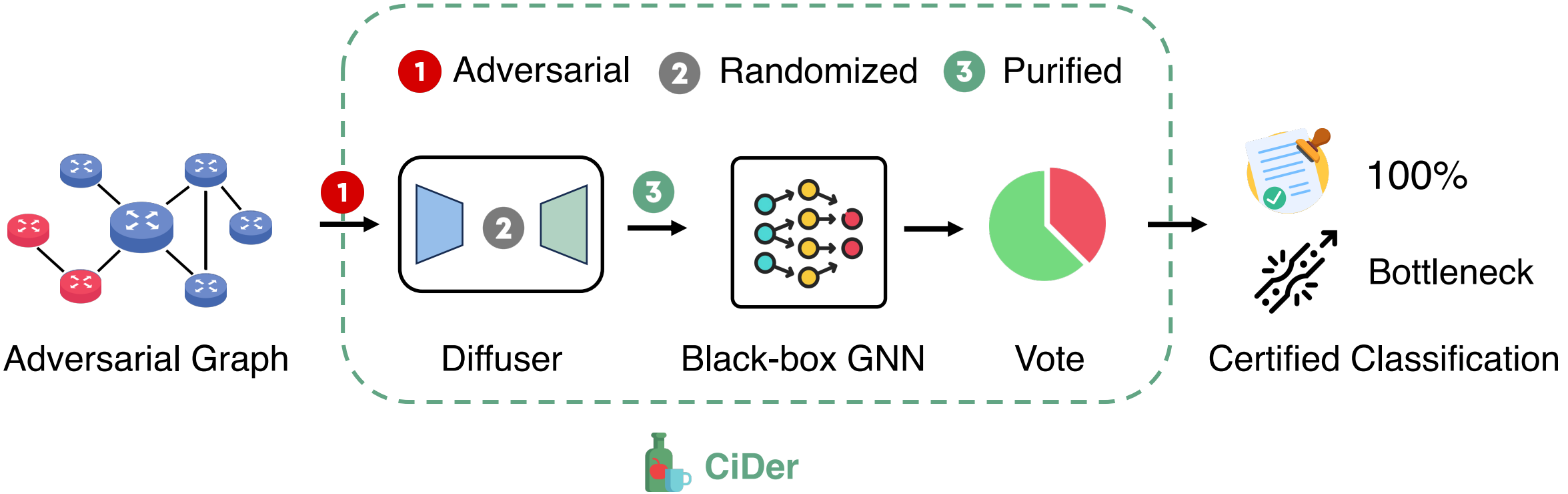
How to certify node classification with black-box GNNs?



Use **diffusion model** as a denoiser to purify the graph while maintain randomness.

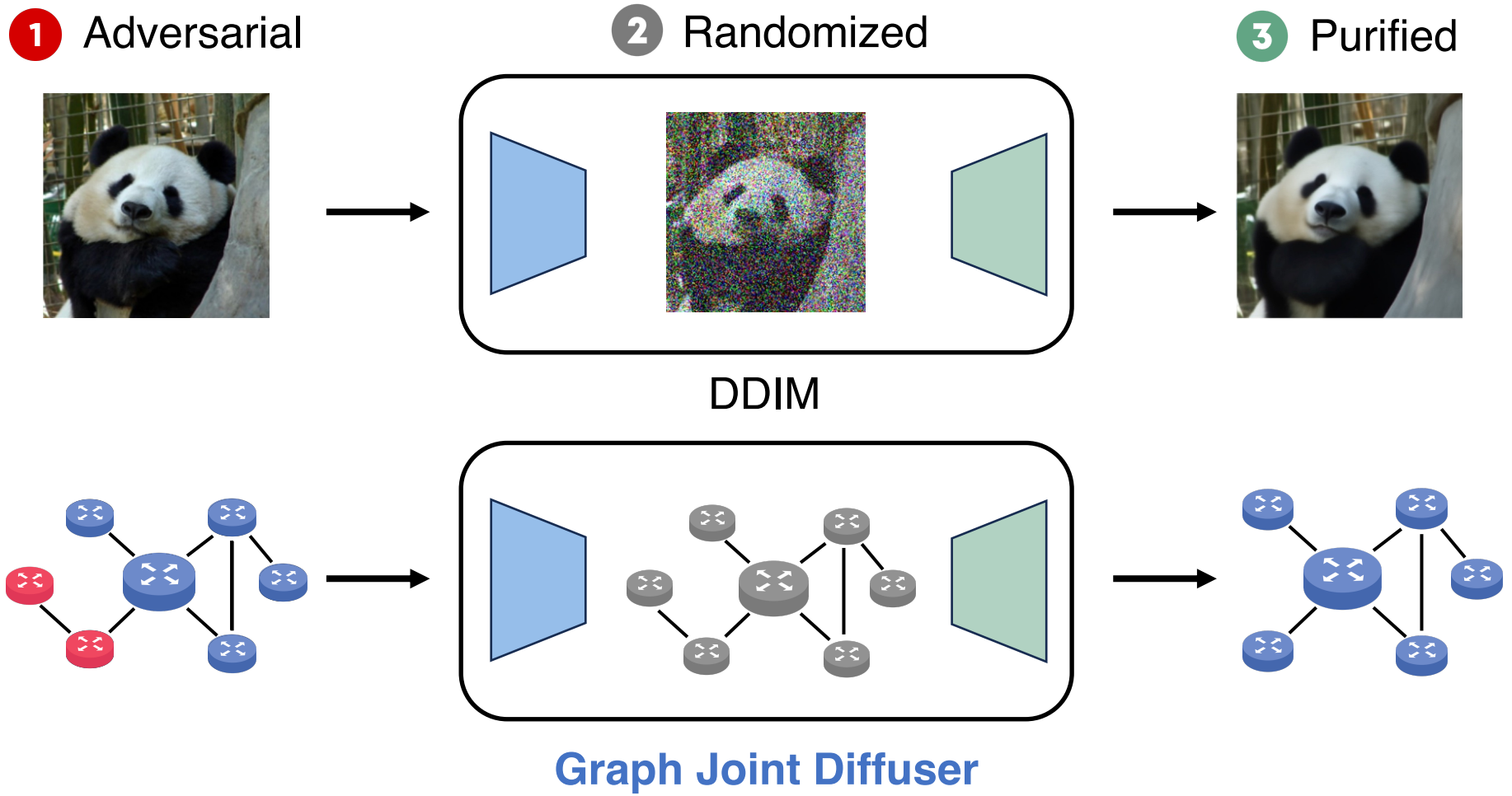
Our Idea: Graph diffusion denoised smoothing

Graph Diffusion Denoiser: Leverage the *denoising* capability of discrete diffusion models (Austin *et al.* 2021) to select the *core features* of different node types and generate the certificate with RS.



Austin *et al.*, "Structured Denoising Diffusion Models in Discrete State-Spaces", in NeurIPS 2021.

Diffusion denoiser works on images, but



- Euclidean
- Pixel-wise
- Fixed-size

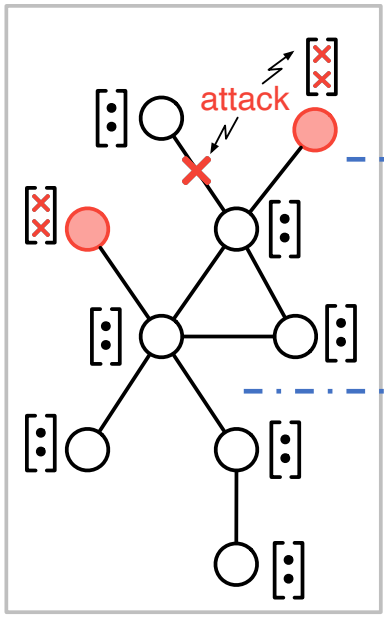
- Non-Euclidean
- Node-edge
- Large graph

Carlini *et al.*, "(Certified!!) Adversarial Robustness for Free!", in ICLR 2023.
Zhang *et al.*, "DiffSmooth: Certifiably Robust Learning via Diffusion Models and Local Smoothing", in USENIX Security 2023.

Subgraph Extractor

Graph modification attack

benign attacked

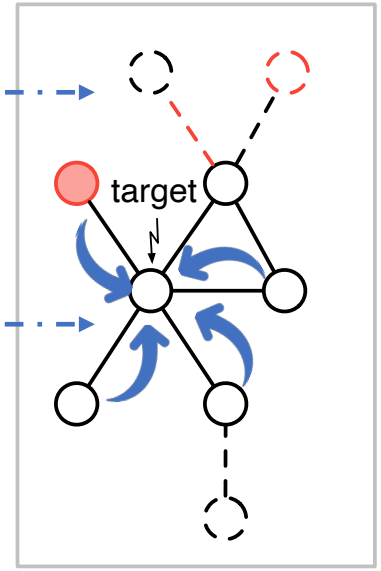


adversarial graph

$\tilde{G} = (\tilde{X}, \tilde{A})$

1 $X' = \{u \mid d(u, v) \leq L\}$

2 $A' = \{(u, w) \mid u, w \in X'\}$



subgraph $G' = (X', A')$ of target v

Large attributed graph

Efficiency gain

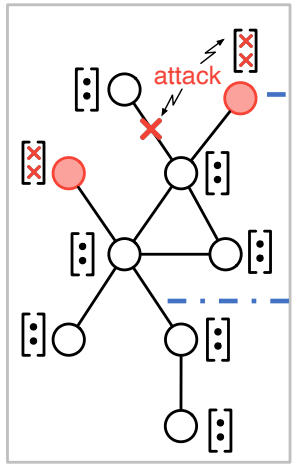
Subgraph for MPNN

Graph Joint Diffuser

diffusion
randomization

Graph modification attack

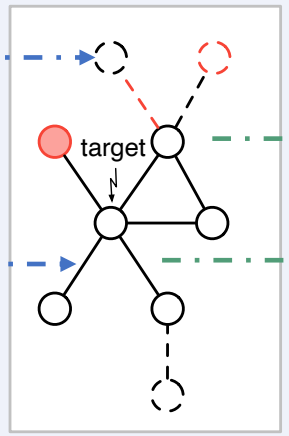
benign attacked



adversarial graph
$\tilde{G} = (\tilde{X}, \tilde{A})$

1 Subgraph Extractor

1. $X' = \{u \mid d(u, v) \leq L\}$
2. $A' = \{(u, w) \mid u, w \in X'\}$

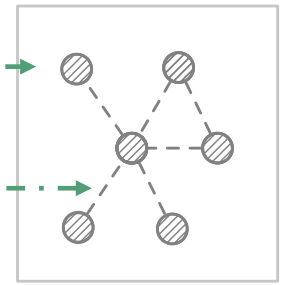


subgraph $G' = (X', A')$ of target v

Non-Euclidean graph structure

2. $X_t \sim X' \oplus \epsilon_{t_x}^X$

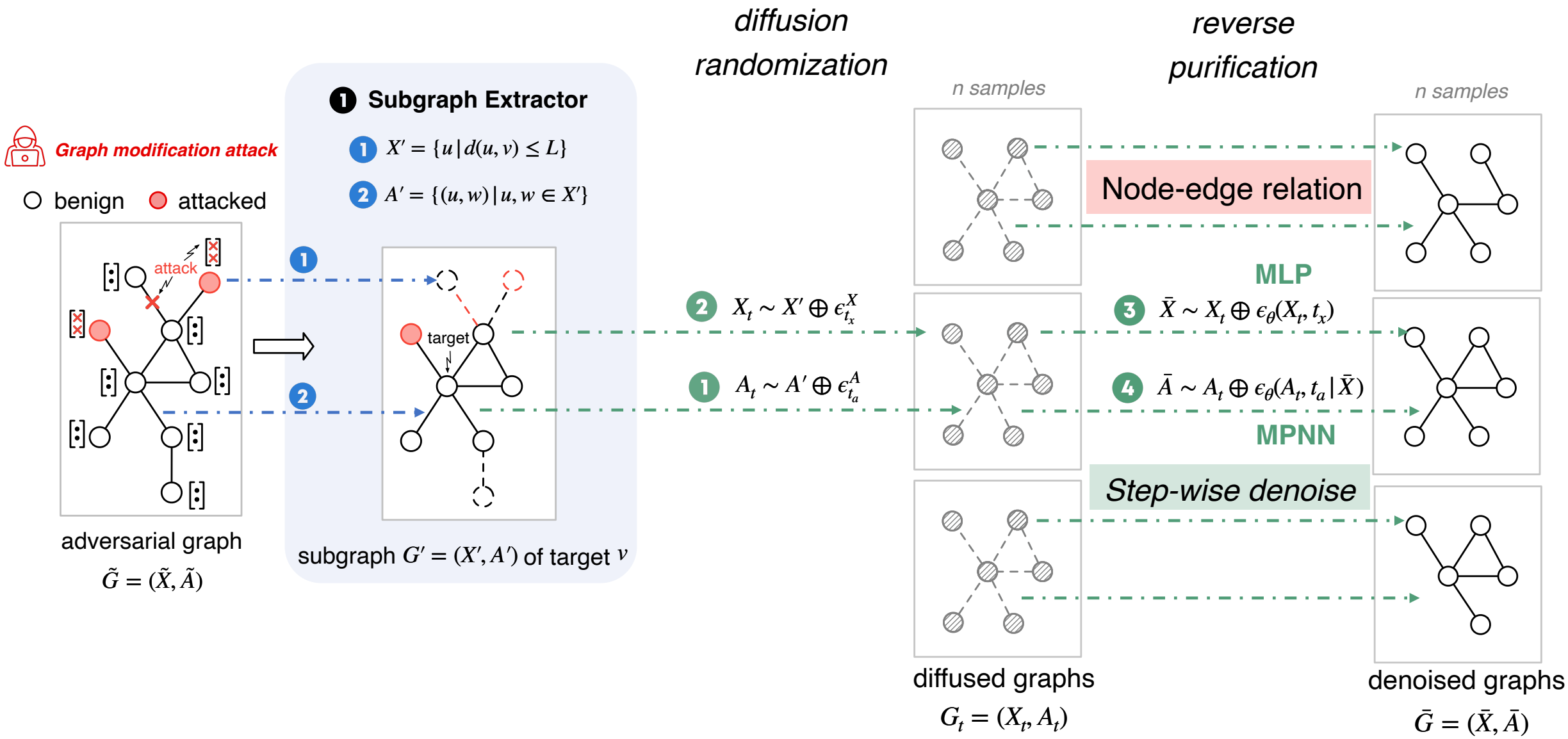
1. $A_t \sim A' \oplus \epsilon_{t_a}^A$



$$\epsilon_{t_z}^Z \sim \mathrm{Ber}\left(p = \left(\overline{\beta_{t_z}} m_1^Z\right)^{1-Z} \left(\overline{\beta_{t_z}} m_0^Z\right)^{Z}\right) \text{ with } Z \in \{X, A\}$$

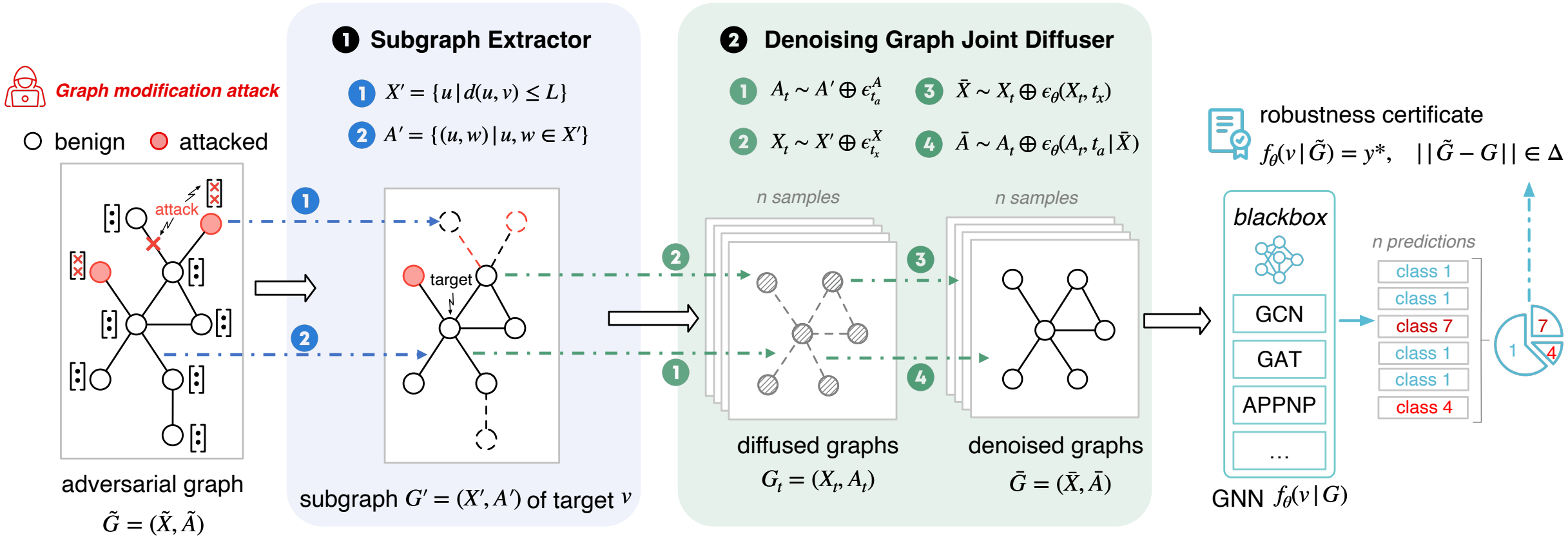
Data-dependent

Graph Joint Diffuser

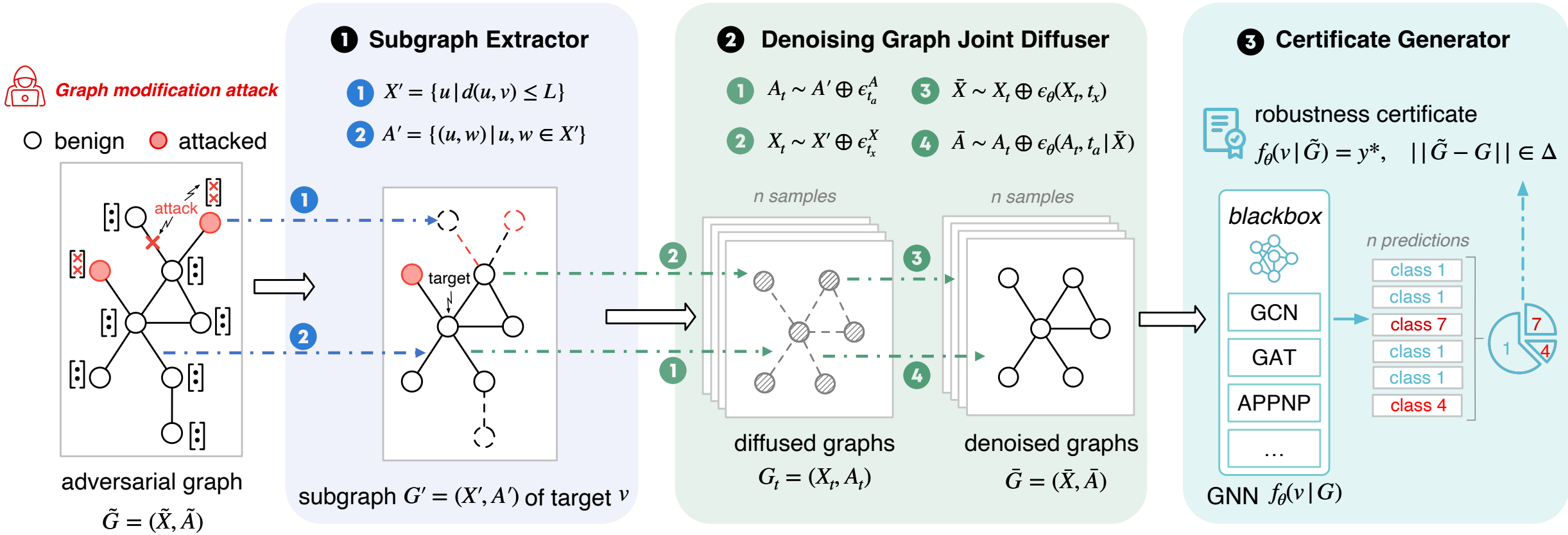


Certificate Generator

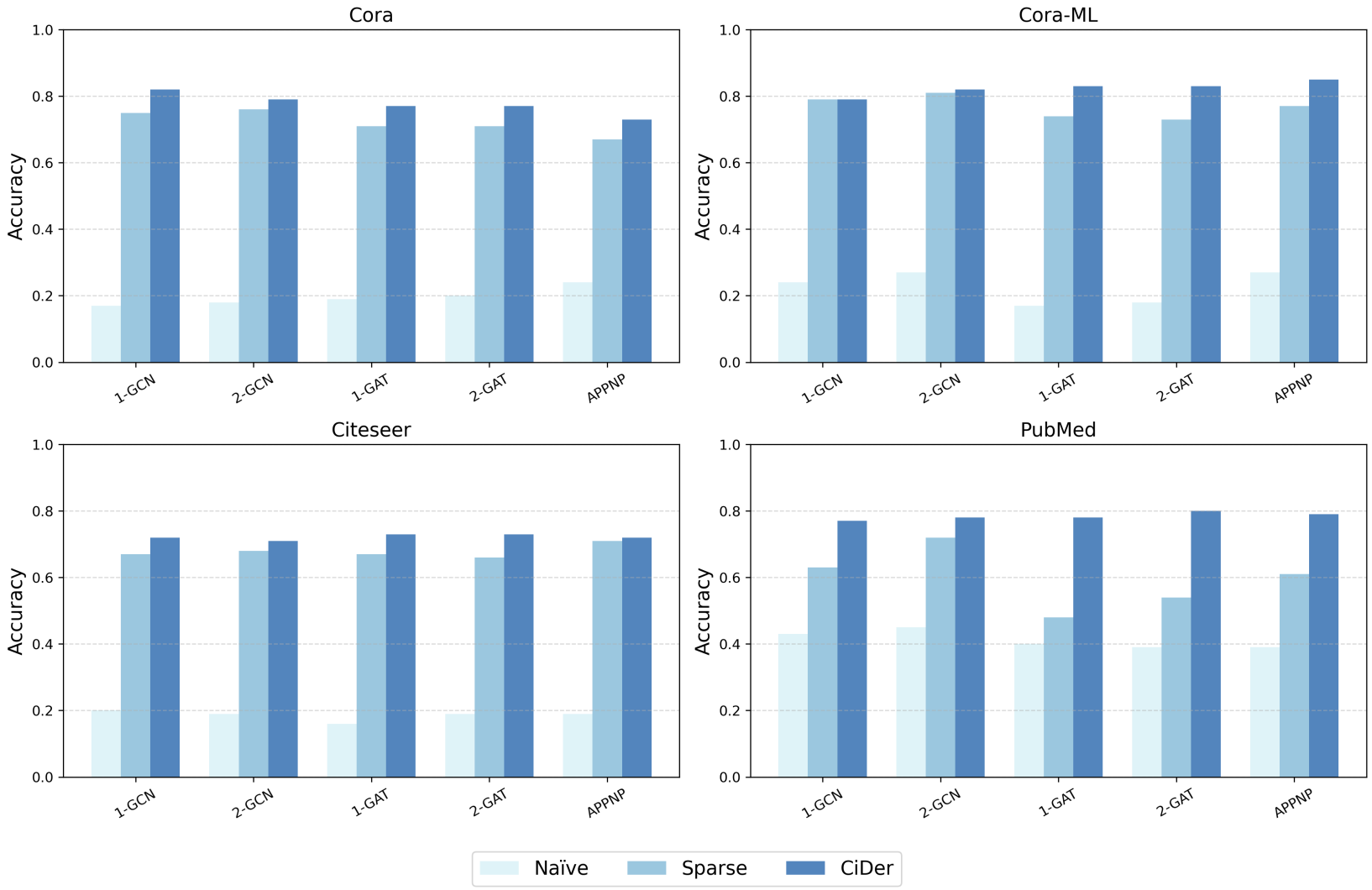
diffusion timestep = randomization parameter



Certified robustness via Diffusion model on graph



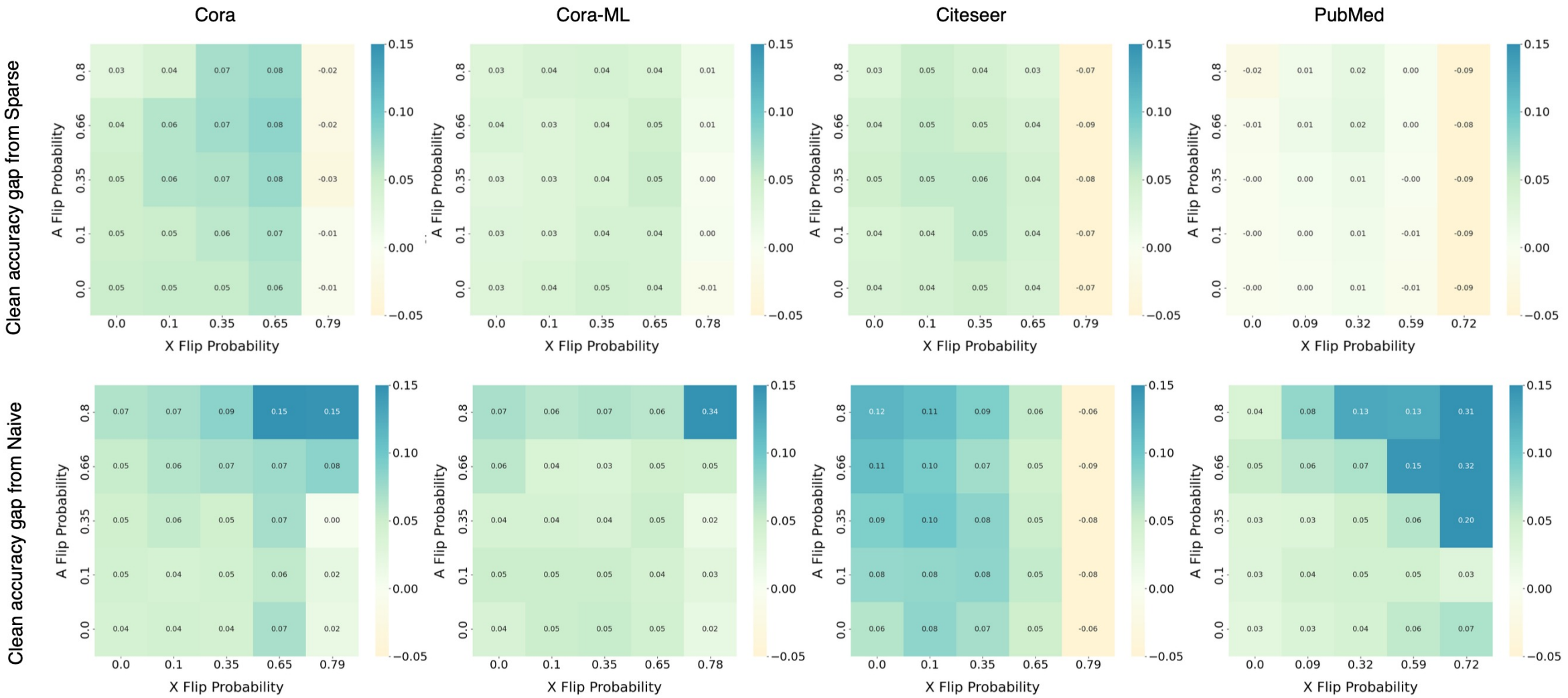
Evaluation – Clean Accuracy



Greater gain on large dataset

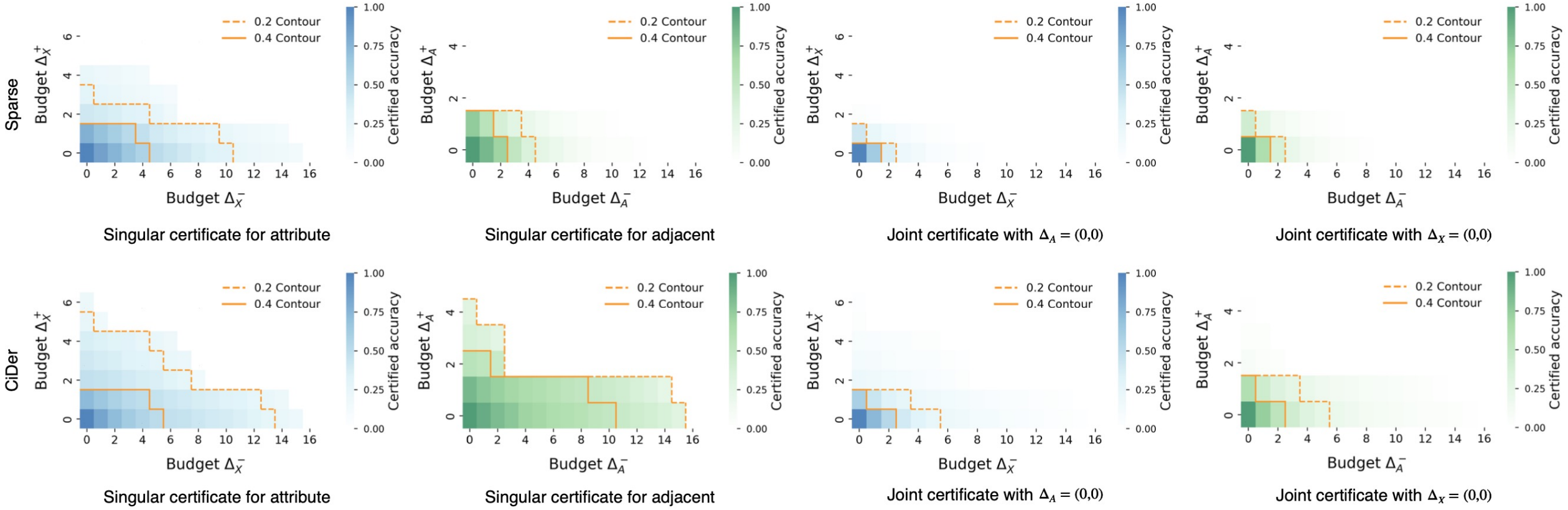
Higher accuracy across datasets, GNN architectures, and GNN layers.

Evaluation – Clean Accuracy



Higher accuracy across different perturbations and even than un-attacked GNN.

Evaluation – Certified Accuracy



Higher certified accuracy and **larger certifiable budget** under structure perturbation.

Summary

- **Black-box Certification**: Black-box adversary, black-box GNN.
- **Graph Joint Diffuser**: Adversarial graph purification with discrete diffusion model.
- **High Accuracy:** CiDer can provide robustness certificates with high accuracy.

- **Code**: https://github.com/xiaoyuu-liang/CiDer

IEEE INFOCOM

Thank you!

Xiaoyu Liang

- xiaoyuliang@buaa.edu.cn
- Looking for a PhD position in Spring/Fall 2026.

Haohua Du, Wen Ma, Ye Tian, and Xiaoya Xu

May 21