

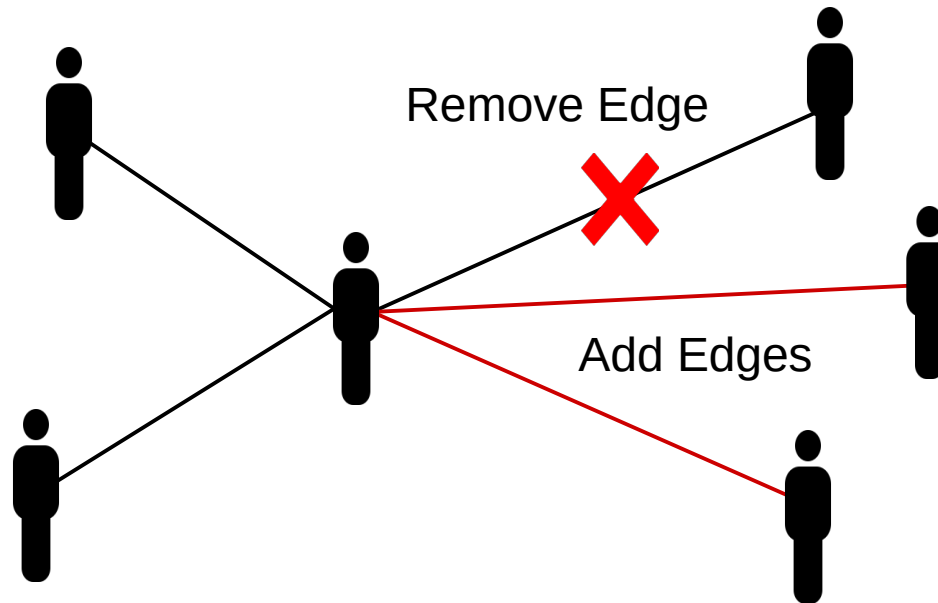
Adversarial Model Extraction of Graph Neural Networks

David DeFazio and Arti Ramesh
SUNY Binghamton



Problem: Can we steal an already trained GNN just by perturbing the underlying graph?

Approach:



Results: Our preliminary results have achieved fidelity of over 80% on citation networks.