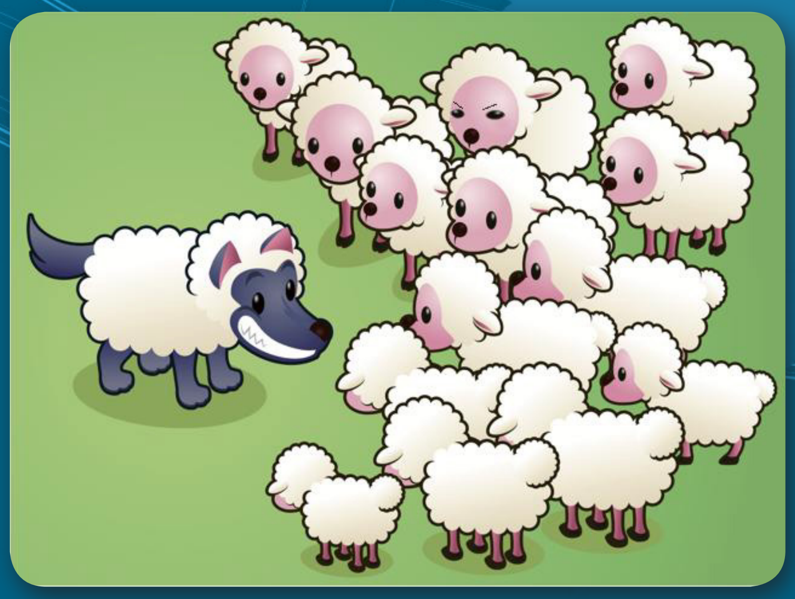
The Center for Cyber Defenders

Expanding computer security knowledge

CAGA: Counter Adversarial Graph Analytics

Dalton Cole, Missouri University of Science & Technology
Sean Flaherty, Texas Tech University

Project Mentor: Jeremy Wendt, Org. 5853



Problem Statement:

There are a variety of techniques to label unknown nodes in a graph. Our goal is to learn how to attack those labeling algorithms and how to defend against those attacks.

Objectives:

- Add a variety of metrics to better determine attack avenues.
- Create a variety of attacking algorithms.
- Create algorithms to defend against the attacking algorithms.

2004 Election Blog Dataset

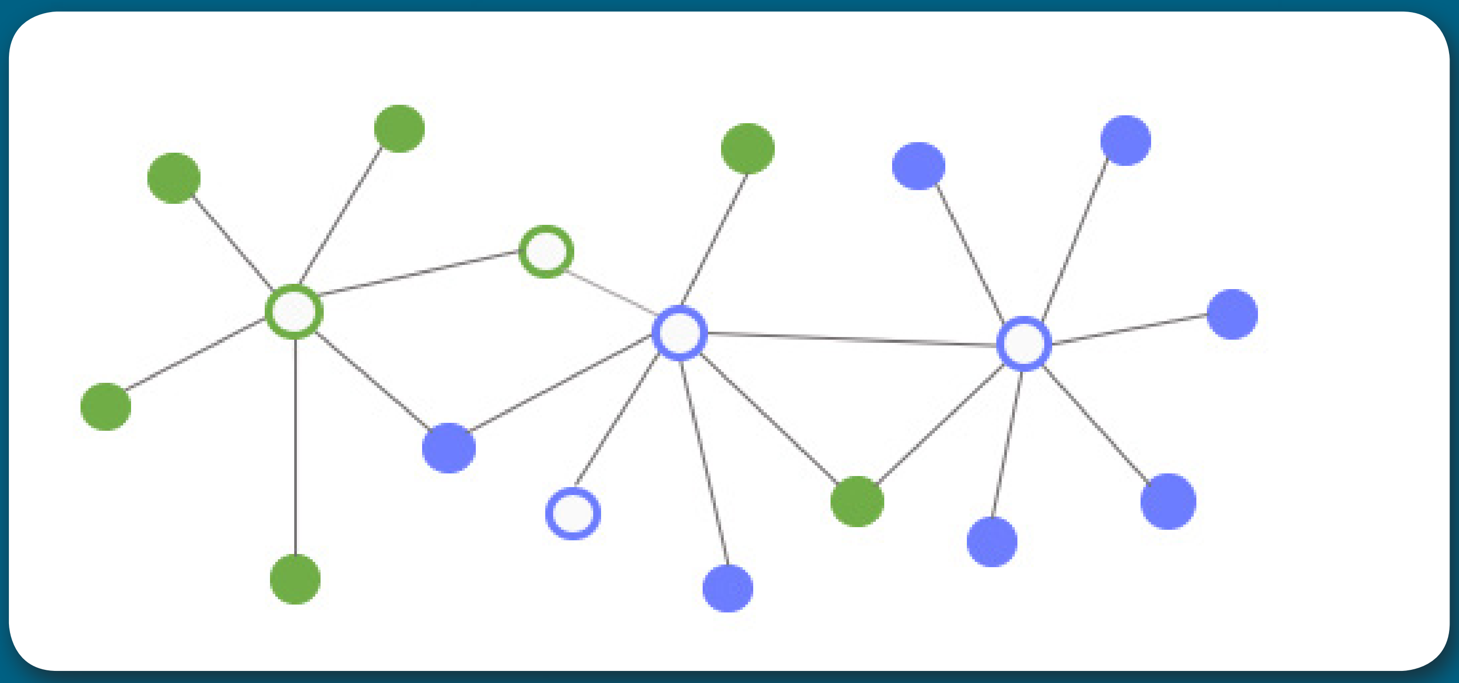
Examples:

- Political affiliation of one blog can be easily inferred by blogs linked to it.
- Data miners can determine private information about social media users based on their friends' public information.
- Search engine algorithms can be exploited to increase page visibility in results.

Approach:

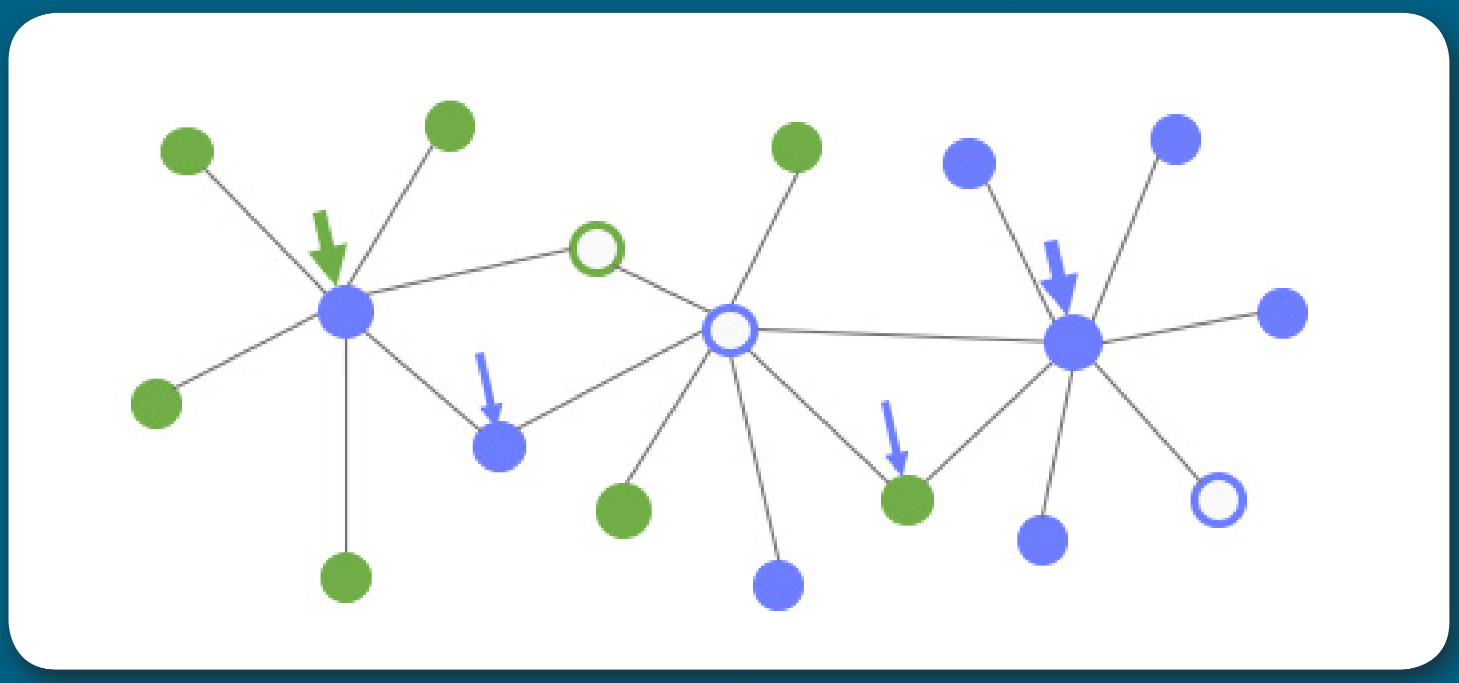
- Algorithms:
  - WVRN: Uses one-hop majority vote to guess the label of a node.
  - LINK: Uses the adjacency matrix as input to a machine learning algorithm such as SVM or decision tree.
- Attack Heuristics:
  - Weak-willed: Sort nodes based on low conversion cost and connects to another class.
  - Whitewash: Similar to Weak-willed, connects low-cost nodes to highest cost majority class node.
  - Random Whitewash: Connect low-cost nodes to random majority class nodes.
  - Information Gain Attack: Increase information entropy for a decision tree, splitting on high information-gain nodes.
  - Clique Attack: Connects nodes to cliques in the graph to convert them.

WVRN



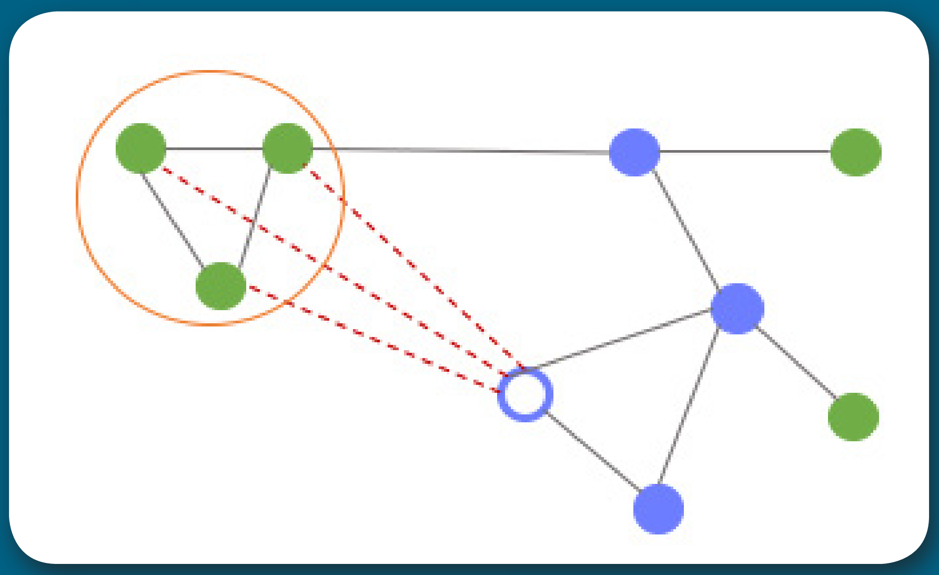
The uncolored nodes are classified by the majority color of neighboring nodes. If a node has no majority, the algorithm is repeated until each node is colored.

LINK



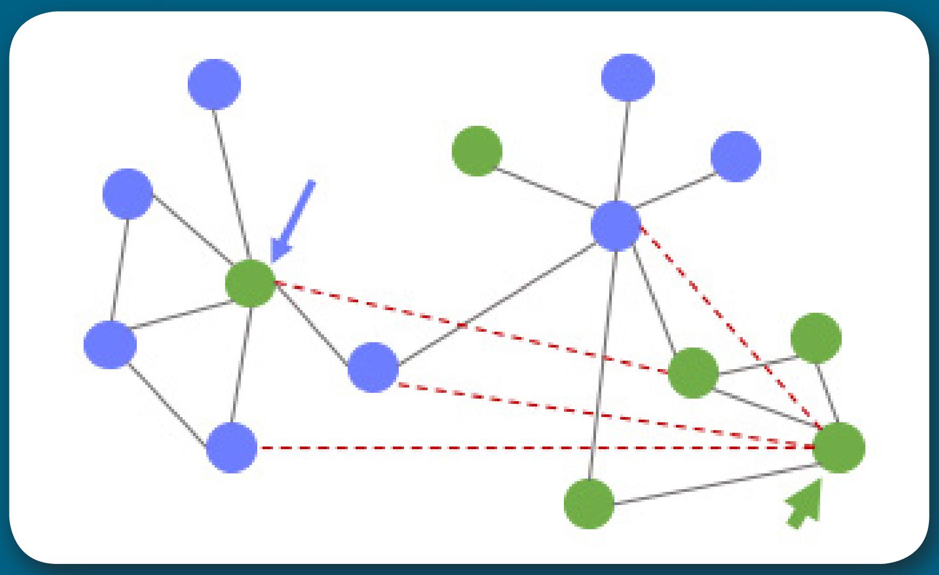
Arrows indicate which color "witness" a node is, i.e., nodes linked to a node with a green arrow are likely to be green, and LINK would likely classify it as such.

Clique Attack



The attack algorithm looks for cliques of the opposite color in the graph. It then precedes to connect the unlabeled node to each node in the clique. Theoretically, the clique should be highly representative of the clique's color in LINK.

Information Gain Attack



The algorithm calculates an information gain value for each node, and, in ascending order, links low-information gain nodes to high-information gain nodes. High-information gain nodes and their witness colors are shown with arrows.

Results:

- Weak-willed and Whitewash both decrease accuracy of WVRN.
- Weak-willed and Whitewash were found to increase accuracy of LINK.
- Random Whitewash is currently the most effective attack on LINK.

Impact and Benefits:

- Analyze attack vectors adversaries could employ to popular graph labeling methods.
- Improve data privacy by analyzing how neighboring nodes affect a local node.

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2018-7541D.

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration