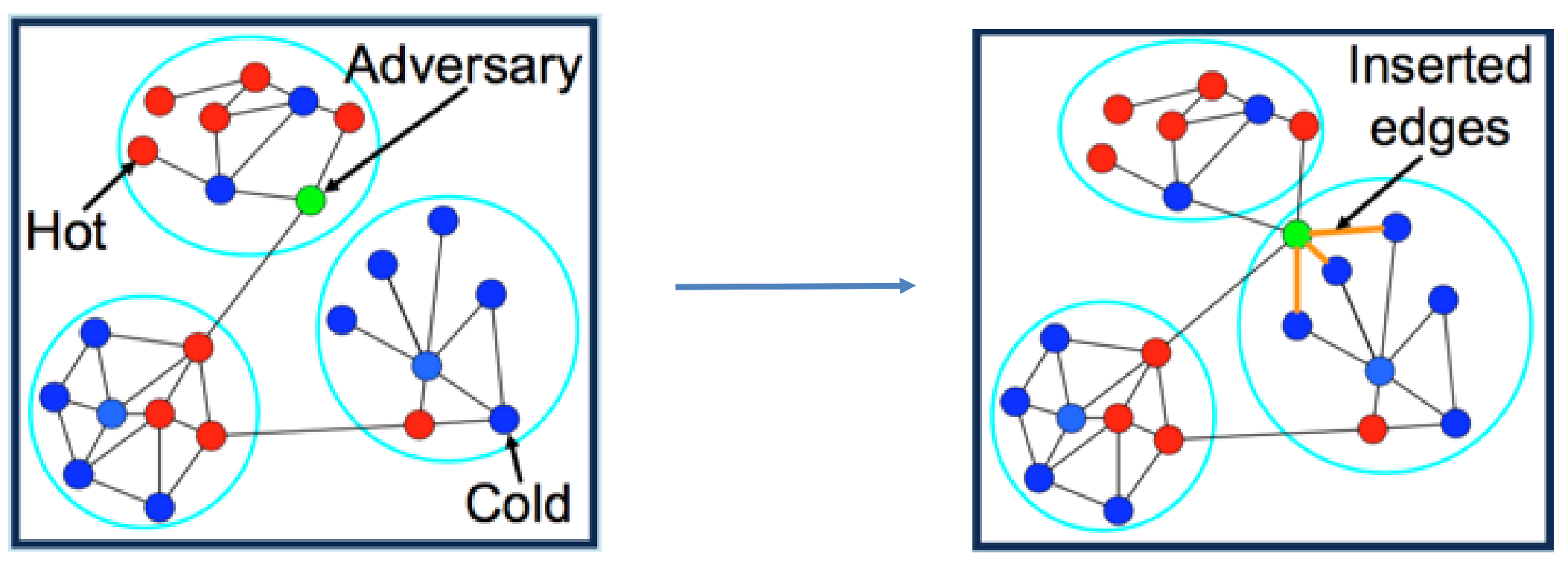
SAND2017-8515C

Generating Network Graphs for CAGA

Sandia National Laboratories
Clifford Anderson-Bergman, Philip Kegelmeyer, Ali Pinar, Jeremy Wendt
Sandia National Laboratories, California 94551

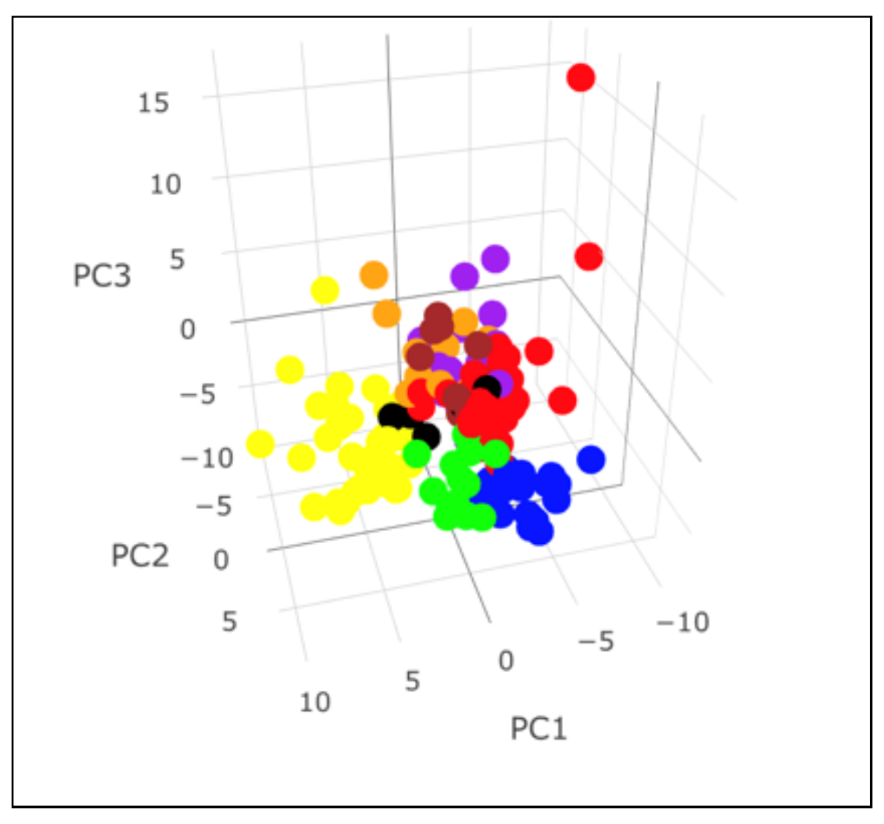
Problem

- Setting: Counter adversarial graph analytics (CAGA) project
  - Adversary can add fake network connections to deceive analyst
  - Goal: develop analytic tools that are robust to limited number of false connections
- Limited number of test datasets
  - Generating realistic graphs is an open research question
  - Project uses real graphs generated from mission problems
- Question: how sensitive are attack and defense strategies to small changes in graph?

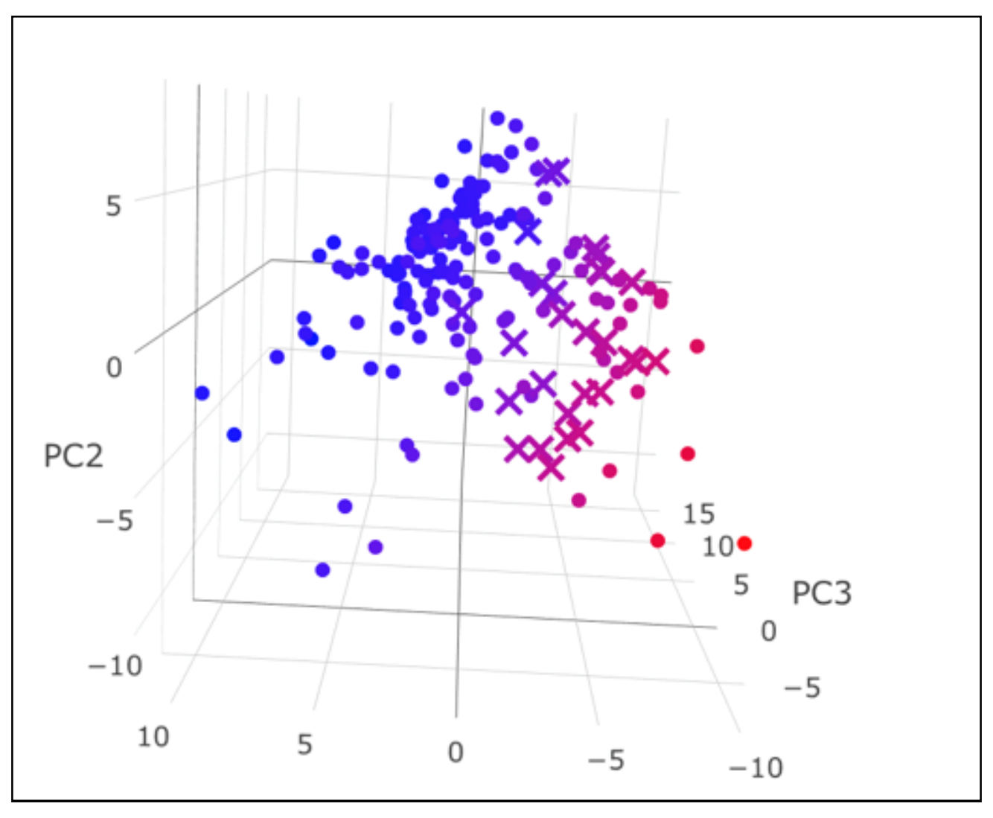


Approach

- Given a real graph, randomly permute connections such that overall graph structure is similar, but individual network connections have changed
- Can then reassess strategy on new synthetic graph; if ideal strategies change, they are not robust!
- Two proposed methods of permuting:
  - canacSBM: preserve node labels and expected degree counts across community and conditional on node labels, while randomly reassigning connections
  - XPCA: represent graph as low-rank factorization, regenerate graph from low-rank structure



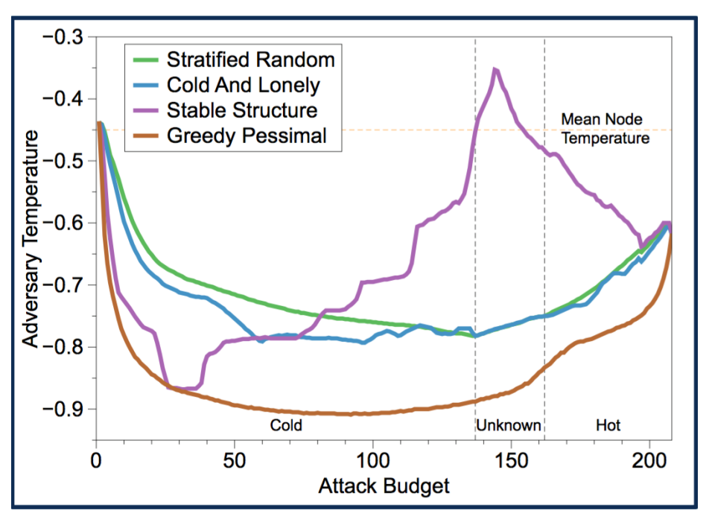
Low rank representation of graph, colored by clustering community



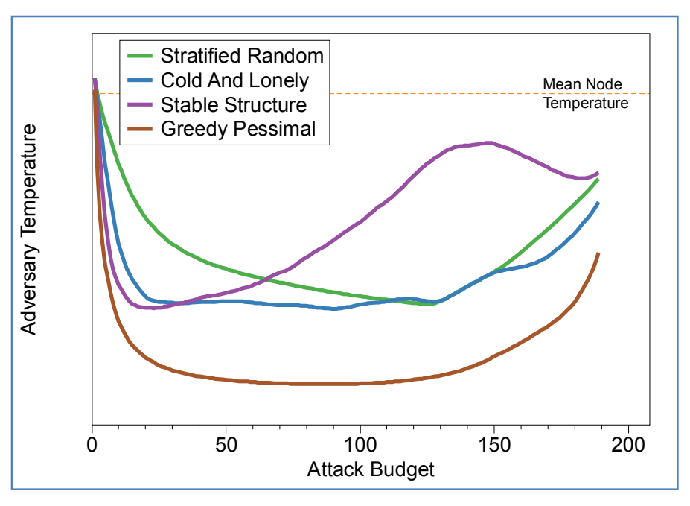
Low rank representation of graph, colored by estimated node label probability, shaped by observed node label

Results

- canacSBM:
  - Code implemented and able to generate synthetic graphs from example graph
  - In motivating example, ROI for attack strategies remained fairly constant, implying strategies robust to minor permutations of graphs



Attack ROI on original graph



Average attack ROI over many synthetic graphs

- XPCA
  - Code borrowed from MXD project
  - Appears to capture interesting aspects of network relations
  - Can directly use low rank factorization to generate new graph
  - Preliminary results only; further investigations required

Significance

- Malicious actors attempt to fool analysts by adding false connections on network
  - Example: twitter bots attempt to friend real users in real twitter communities
- Analysts need tools that will be robust to intentional deception
  - Main thrust of CAGA
- Real datasets required to test proposed methodology
- If new data sets are difficult to obtain, want to regenerate similar structure graphs to test reliability of methodology
  - Current sub-problem for CAGA
- Necessary step to test robustness of methodology before deploying to analysts in the field

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.

SAND 2017-xxxP