SAND2016-1534C

Counter-Adversarial Community Detection: Initial Investigations

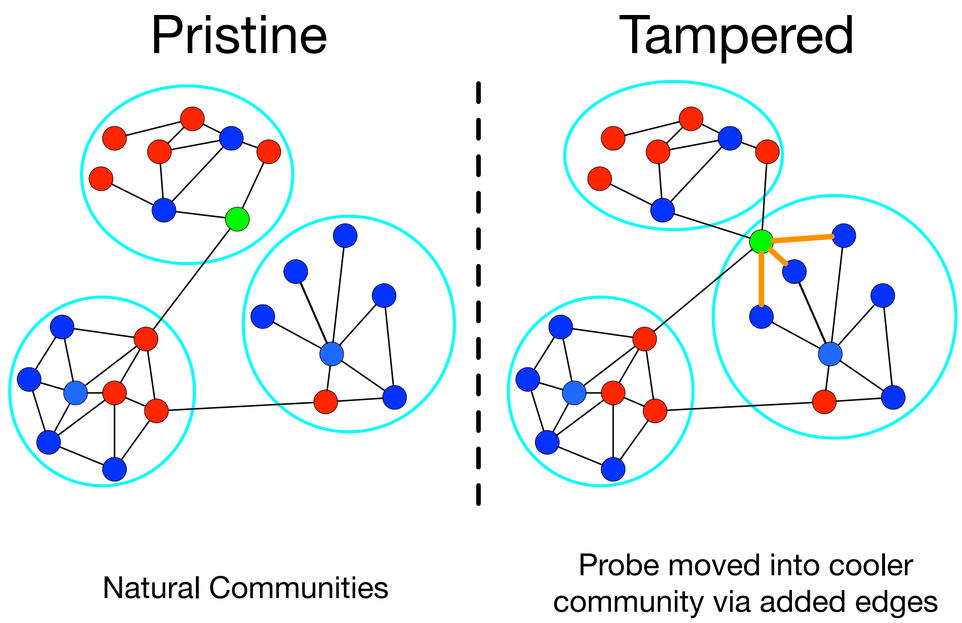
Authors: Philip Kegelmeyer (wpk@sandia.gov), Jeremy Wendt (jdwendt@sandia.gov), Ali Pinar (apinar@sandia.gov)

Assume every node has a "temperature": hot (1), cold (-1), or unknown (0). A community's temperature is the average of its nodes.

The adversary's goal is to move a specific "probe" node from a hot, obvious community into a cooler community.

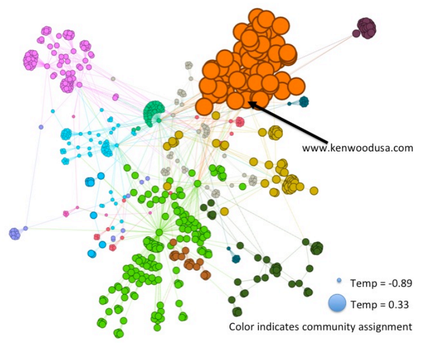
The adversary's only method, here, is to add edges between the probe node and other nodes.

But which edges, in what order? We abstract an "attack" as a method for generating a sorted order of all of the nodes in a graph, to be linked to one at a time.



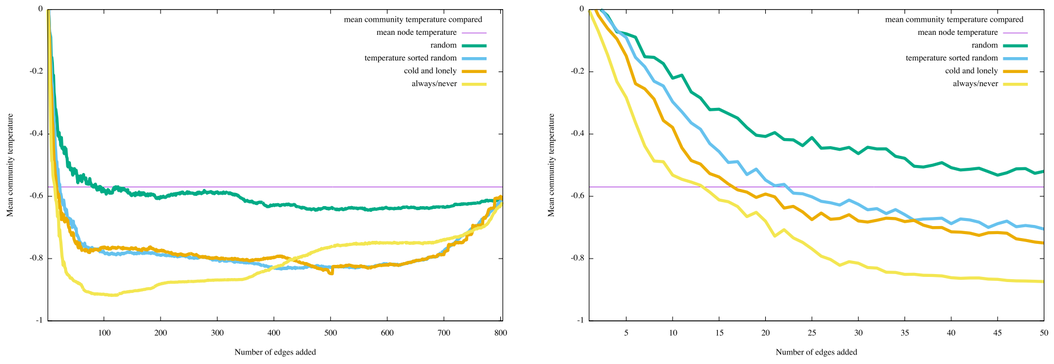
Natural Communities

Probe moved into cooler community via added edges

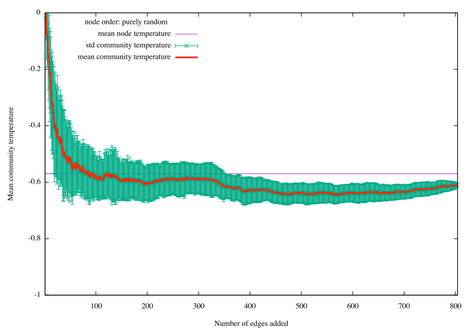


Data description: a web graph formed by combining the two hop ego networks of HTML links from a variety of ham radio related home pages. 804 nodes, 1137 edges, roughly 17 communities as detected by Louvain.

Tentative conclusions: smart attacks are indeed better than random attacks, and paying attention to “stable structure" is the most effective so far.

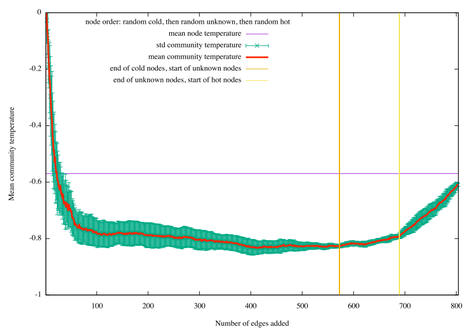


Attack plots: number of edges added on x, temperature of probe node's community on y. 100 runs each, with mean and std of temperature plotted. An effective attack goes to -1 (cold) quickly and stays low as long as possible.

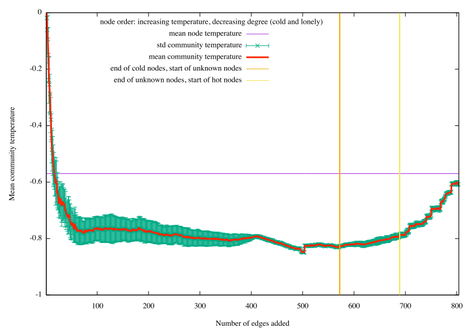


A random attack, for calibration.

Stable structure analysis: Louvain has a (mildly) random outcome. Define an “always” community as a group of nodes that were always in the same community across many Louvain runs. Then link to all the nodes in the coldest always community, then the next coldest, and only at the end to link to the less attached cold, unknown, and hot nodes.

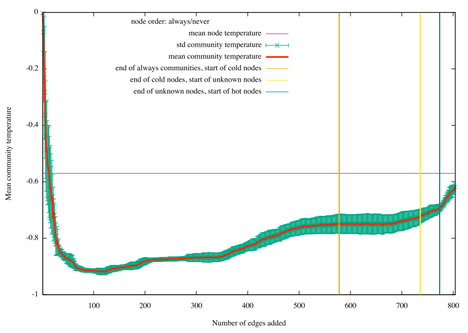


A less silly temperature-stratified random attack



The "cold and lonely" attack: link to cold nodes in increasing degree, then unknown nodes, then hot.

"Always" communities as reflected by node similarity.



The "always” attack.

This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-AC04-94AL85000.