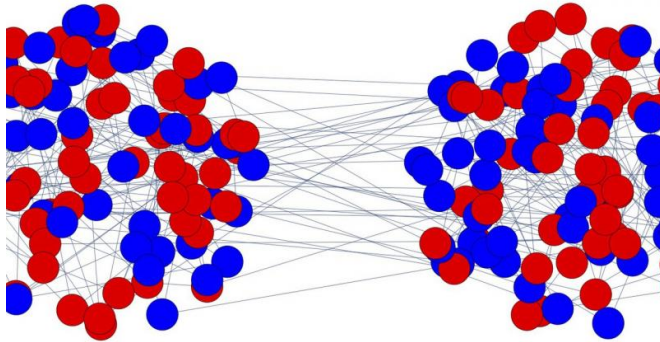


What algorithms can't tell us about community detection

28 July 2017



"Illusory" communities: two equally good ways to divide a network into communities that have nothing to do with each other. Credit: Santa Fe Institute

Many who study networks care about groups of interconnected nodes. These groups, called "communities" or "modules," represent real-world relationships like friend groups on Facebook, businesses in a supply chain, and even species within a food web. The challenge is to identify whether, and ultimately where, these structures exist within a mass of data.

In a recent paper, Jess Banks, a Ph.D. candidate in mathematics at UC Berkeley and a former Santa Fe Institute undergraduate intern, Robert Kleinberg, Associate Professor of Computer Science at Cornell, and SFI Professor Christopher Moore set out to test under what conditions a computer algorithm can verify the absence of [community structure](#) in a network. Without an algorithm that can do this, network scientists can't tell whether the communities they find are statistically significant—that is, they can't tell real communities from fake ones.

Banks posed the research as a thought experiment: "If I generate a random network with no community structure 'baked in,' will it have

communities by chance? If not, can an algorithm certify that it doesn't?"

After generating random networks with no real community structure, the researchers put one particular algorithm to the test—the simplest algorithm in a popular class called "the Sum of Squares hierarchy." They decided to investigate the algorithm's ability to verify the absence of dissasortative community structures, which, like competitive businesses, are characterized by a lack of connections with each other. In computer science, this corresponds to the classic Graph Coloring problem, where nodes connected by an edge are required to have different colors.

By studying the behavior of this algorithm, the researchers uncovered a blind spot. If a [network](#) is too sparse, with too few connections, the algorithm cannot tell whether or not it has communities. Using some clever mathematics, they proved that the [algorithm](#) can be fooled into thinking that communities exist even when they don't.

"If we care about doing good science, and honestly testing our hypotheses, then verifying the absence of [structure](#) in the data is just as important as being able to find it when it is there," Banks says.

"We're all looking for patterns in data," Moore adds. "But just like humans, our algorithms often find patterns that aren't really there. We need to understand the fundamental limits on our ability to tell whether patterns truly exist, so we'll know when we need more and better data before we can draw any conclusions."

Going forward, the researchers' method could be used to test other, more powerful algorithms in the same hierarchy.

More information: The Lovász Theta Function for Random Regular Graphs and Community Detection in the Hard Regime. *arXiv*.

arxiv.org/abs/1705.01194

Provided by Santa Fe Institute

APA citation: What algorithms can't tell us about community detection (2017, July 28) retrieved 20 January 2022 from <https://techxplore.com/news/2017-07-algorithms.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.