Detecting planted structures in random graphs

February 17 2021

Pete Linforth. Credit: Pixabay

Many complex systems can be modeled as irregular networks, with hubs and communities. Researcher Kay Martin Bogerd of the Department of Mathematics and Computer Science has investigated how existing

community detection methods can be extended to a setting of inhomogeneous random graphs. The results offer new insights in the working of extremely large networks, such as the internet, social media or the brain. Bogerd defended his thesis on 17 February 2021.

Many complex systems can be modeled as a network, which is simply a collection of objects and the relations between them. The objects are typically referred to as vertices and the connections between them are called edges. Networks have become a useful and flexible representation for a wide variety of systems in different areas, such as social, biological, technological, and communication sciences.

Networks representing real systems are usually not regular and inhomogeneous, with various parts of the network having radically different structure. For instance, many networks contain a few important vertices, often referred to as hubs, that have many more connections than a typical vertex. Another example is a network that consists of several communities, densely connected groups of vertices, with significantly more internal connections than external connections.

Community detection

The work in this thesis is centered around the analysis of community detection methods for inhomogeneous networks, as well the detection of other types of anomalies such as botnets. We study how existing community detection methods can be extended to a setting of inhomogeneous random graphs. Random graphs are network models that are used to test whether there is an actual structure present in networks or whether it can also be attributed to random noise.

This led to new insights on the properties of the random graphs we study, and we show how community detection methods can be extended to the inhomogeneous setting in an optimal manner. The insights from

this project also sparked the interest to consider a related project about the detection of botnets.

Lastly, we consider dynamically growing networks using the preferential attachment model. Below I will elaborate more on each of these projects, and the publications that originated from them.

Cliques in graphs

One of the main questions we answer in this thesis is: ``What is the smallest community that can theoretically be detected in an already inhomogeneous graph?'''. The initial work on this project resulted in novel insights about inhomogeneous graphs and, in particular, about the behavior of cliques in these graphs. Remarkably, the size of the largest clique is almost always the same, even in rather inhomogeneous random graphs.

We also extended these results to quasi-cliques and show that also the quasi-clique is highly concentrated in dense inhomogeneous random graphs. We were also able answer our initial question and characterized the smallest community that can theoretically be detected in an inhomogeneous graph. We have done this by proposing and analyzing a scan test and showing that this scan test is optimal in the sense that it is not possible to detect any community that cannot also be detected by our scan test.

Masked botnets

Communities are typically modeled as more densely connected subgraphs within a graph, but one is sometimes also interested in subgraphs that are different in other ways. For example, one could be interested in an anomaly such as a botnet that tries to mask its presence

by not making too many connections.

However, the connectivity structure or underlying geometry of a botnet is often still rather different, and this can be exploited to detect the presence of such a botnet. We formalized this idea and introduced two tests that can both detect such a botnet. Furthermore, we also show that these tests are optimal in an asymptotic sense.

Dynamic networks

Many networks are dynamic and change over time, with some rules concerning the evolution or growth of the graph. For this we consider the preferential attachment model and study what happens when the attachment function changes after some time. In particular, we investigated when it is possible to detect that the attachment function has changed. We show that this is indeed possible, even when there are only few vertices with using the alternative attachment function.

  More information: Detecting planted structures in random graphs. research.tue.nl/nl/publication … res-in-random-graphs

Provided by Eindhoven University of Technology