Master Thesis

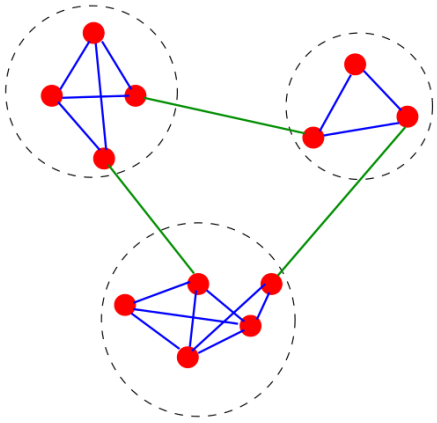
Chair for Theoretical Information Technology | RWTH AACHEN UNIVERSITY

Stochastic Block Model for Overlapping Community Detection
Fundamental Limits and Performance Analysis

Research Area
Community Detection, Clustering, Convex Optimization

Keywords
Maximum likelihood, Semidefinite programming, random graph

Description
Understanding the structure and dynamics of complex systems such as social networks and biological systems is a recurring research problem. For instance, consider a network of nodes, each one with an ensemble of properties. Nodes sharing a specific property can be organized as members of a community and it is often desirable to find these communities. The community detection aims at finding different communities based on observations about them. Most of the time, the main available observation is the interconnections of nodes in the network, which is done based on certain criteria. When the network is modeled as a graph, the community detection solves the problem of finding a vector of latent labels $\mathbf{l}_k$ at each node k based on observation of the adjacency matrix A. The stochastic block model (SBM) is a common model for community structure in complex networks. At its simplest form, SBM consists of n nodes and two non-overlapping communities where two nodes are connected with probability p or q, if respectively they belong to the same community or not. For this simple model, if $p = \frac{a \log n}{n}$ and $q = \frac{b \log n}{n}$, then it has been proved that the exact recovery of communities is possible if and only if $\sqrt{a} - \sqrt{b} \geq \sqrt{2}$. Moreover a semidefinite programming relaxation of maximum likelihood can be shown to approach this limit.



However, in general complex networks, the communities are not disjoint and might as well overlap. Classical SBM fails at recovering those communities for instance by announcing the overlapped part as a new community. The model has been generalized for overlapping community detection and one can mention for example mixed membership SBM and degree corrected SBM. Their fundamental limits as well as their comprehensive performance evaluation are still under active research.

Goal
The goal of this thesis is to study the limits and performance of overlapping community detection algorithms using stochastic block models. It is important to investigate theoretical recovery guarantees for stochastic block models and to set up numerical analysis for comparison and evaluation of proposed algorithms.

Requirements
- Strong background in probability theory and convex optimization
- Python programming

Contact

- Arash Behboodi ✉ behboodi@ti.rwth-aachen.de ☎ +49 241 80 27710