UPPSALA UNIVERSITET

Community Detection with Stochastic Block Modeling (SBM) in Bipartite Graphs

Project Members:

Deepthi Hulithala Venkataramana,

Naeim Rashidfarokhi

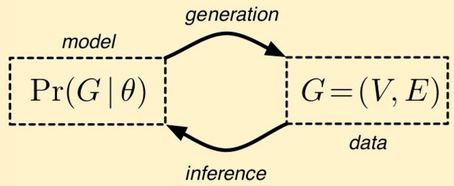
dehu1371@student.uu.se
nara5324@student.uu.se

Supervisor:

Davide Vega D'Aurelio
davide.vega@it.uu.se

Project in Computational Science
Fall term 2020

INTRODUCTION



Stochastic block model(SBM) as generative model for structure detection in network

★ Generation: having a set of parameters as θ, draw an instance network G from this distribution.
★ Inference: Having V number of vertices and a set of edge as E between them, find a set of parameters as θ which describes the network

THEORY & METHOD

Standard SBM

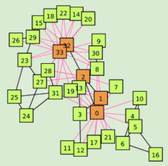
$$L(G|g) = \sum_{rs} \frac{m_{rs}}{2m} log \frac{\frac{m_{rs}}{2m}}{\frac{n_r n_s}{n^2}}$$

★ r, s = communities
★ k_r = sum of degrees of the vertices in community r
★ n_r = number of vertices in community r

Degree corrected SBM

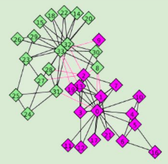
$$L(G|g) = \sum_{rs} \frac{m_{rs}}{2m} log \frac{\frac{m_{rs}}{2m}}{\frac{k_r}{2m}\frac{k_s}{2m}}$$

★ m = total number of edges in the network
★ m_{rs} = number of edges between communities r and s



Community detection using Standard SBM

In networks with substantial degree heterogeneity, the Standard SBM prefers to split networks into communities of high and low degree. The degree-corrected model correctly ignores divisions based solely on degree and hence is more sensitive to underlying structure.



Community detection using Degree corrected SBM

IMPLEMENTATION

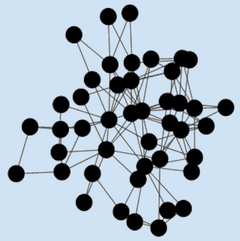
★ Brute Force Search
★ Local heuristic algorithm
  ○ proposed by B. Karrer & M. Newman.

★ Minimum description length principle
  ○ To infer the optimal number of communities from the data.

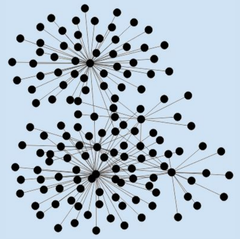
RESULT

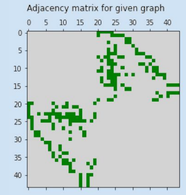
A small bipartite network of the affiliations among elite individuals and the corporate, museum, university boards, or social clubs to which they belonged, from 1962. (44 Nodes, 99 Edges)

A bipartite network of the memberships of notable people and organizations, from the American Revolution (1765-1783) between users and groups on YouTube, from YouTube network in 2007 (141 Nodes, 160 Edges)

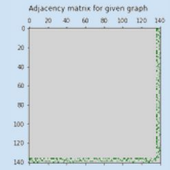


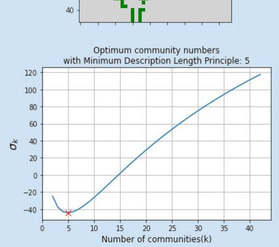
Bipartite Network



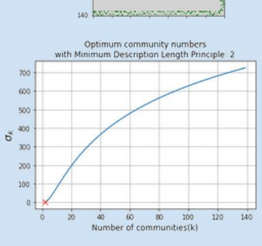


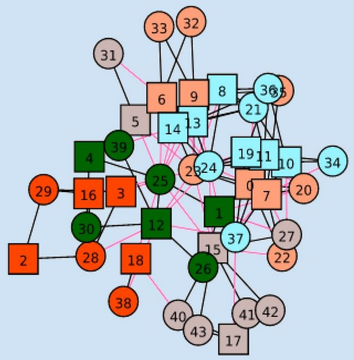
Adjacency Matrix



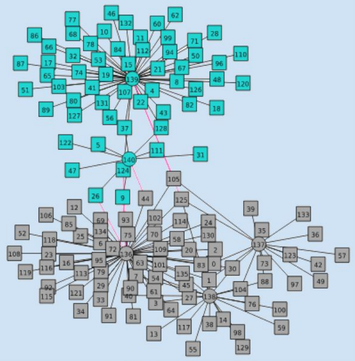


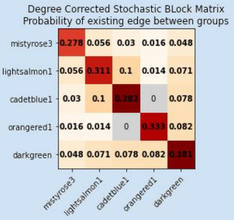
Number of communities using MDL





Communities detected





Stochastic Block Matrix