

Course project

«Optimization approaches to community detection»

Marina Danilova, Alexander Podkopaev, Nikita Puchkin,
Igor Silin

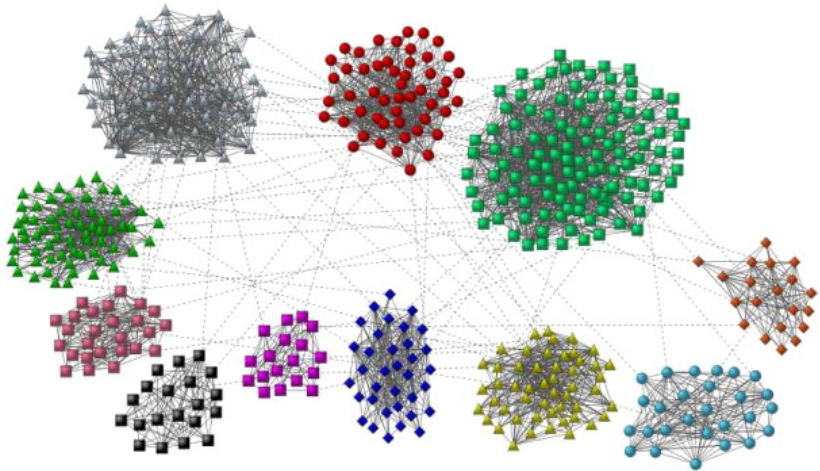
SKOLKOVO INSTITUTE OF SCIENCE AND TECHNOLOGY

December 16, 2016

Plan

- 1 Introduction to community detection
- 2 Algorithms
 - Spectral method
 - Modularity-based method
 - Natural conjugate gradients method
 - Semidefinite relaxations
- 3 Experimental results

Example



Notations

Assumption

We consider **undirected unweighted** graphs **without loops** with n nodes. The nodes are enumerated as $\{1, \dots, n\}$.
Graph is given by its $n \times n$ adjacency matrix A .

Goal of community detection

Find **partition** of nodes into **non-overlapping** clusters.
The number of clusters is k .
The clusters are denoted as $\{C_1, \dots, C_k\}$.

Spectral method

Formulating an optimization problem

Spectral method

Modularity-based method

Formulating an optimization problem

Modularity-based method

Natural conjugate gradients method

- Model parametrized by

$$z(i) \sim \text{Poly}(\pi), \quad i = \overline{1, n}$$

$P = \|p_{ij}\|_{i,j=\overline{1,k}}$ - probabilities of inter-cluster edges occurrence

- Bayesian approach:

$$\pi \sim \text{Dirichlet}(\alpha)$$

$$p_{ii} \sim \text{Beta}(\beta), \quad i = \overline{1, k}$$

$$p_{ij} \ll 1, \quad \forall i \neq j$$

- $p(z, \pi, P|A)$ - true posterior with observed adjacency matrix A
- \mathcal{Q} - family of feasible distributions

Natural conjugate gradients method

Formulating an optimization problem

$$\mathcal{L}(q) \equiv -\text{KL}(q \| p(Z, \pi, P|A)) \longrightarrow \max_{q \in Q}$$

- The problem can be reduced to unconstrained optimization:

$$\mathcal{L} = \mathcal{L}(\theta) \longrightarrow \max, \quad \theta - n \times (k - 1) \text{ matrix}$$

- Use conjugate gradients method in a statistical manifold
- Metrics is defined by a matrix

$$\mathcal{I}(\theta) - \text{Fischer information}$$

Semidefinite relaxations

Formulating an optimization problem

Semidefinite relaxations

