Course project
«Optimization approaches to community detection»

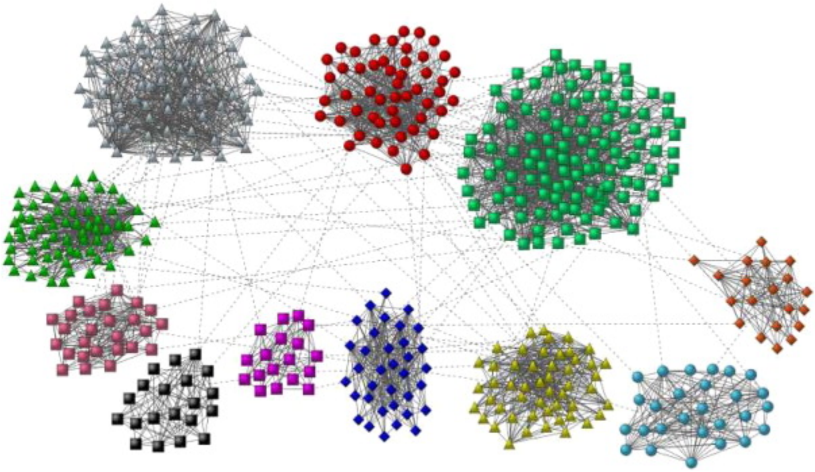
Marina Danilova, Alexander Podkopaev, Nikita Puchkin,
Igor Silin

Skolkovo institute of science and technology

December 16, 2016

Plan

Example



Notations

Assumption

We consider **undirected unweighted** graphs **without loops** with n nodes. The nodes are enumerated as $\{1, \ldots, n\}$.
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 $\{\mathcal{C}_1, \ldots, \mathcal{C}_k\}$.

Spectral method

Formulating an optimization problem

Spectral method

Modularity-based method

Formulating an optimization problem

Modularity-based method

Natural conjugate gradients method

Bayesian approach:

$$Z_i \sim \mathrm{Poly}(\pi), \quad i = \overline{1, n}$$
$$\pi \sim \mathrm{Dirichlet}(\alpha)$$
$$P_{ii} \sim \mathrm{Beta}(\beta), \quad i = \overline{1, k}$$

$p(Z, \pi, P|A)$ – true posterior

$\mathcal{Q} = \left\{ q : \; q(Z, \pi, P) = q(\pi)q(P)\prod_i q(Z_i) \right\}$ – feasible distributions

Formulating an optimization problem

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

Natural conjugate gradients method

- The problem can be reduced to unconstrained optimization:

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

- 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