



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

# Алгоритмы кластеризации в моделях случайных графов

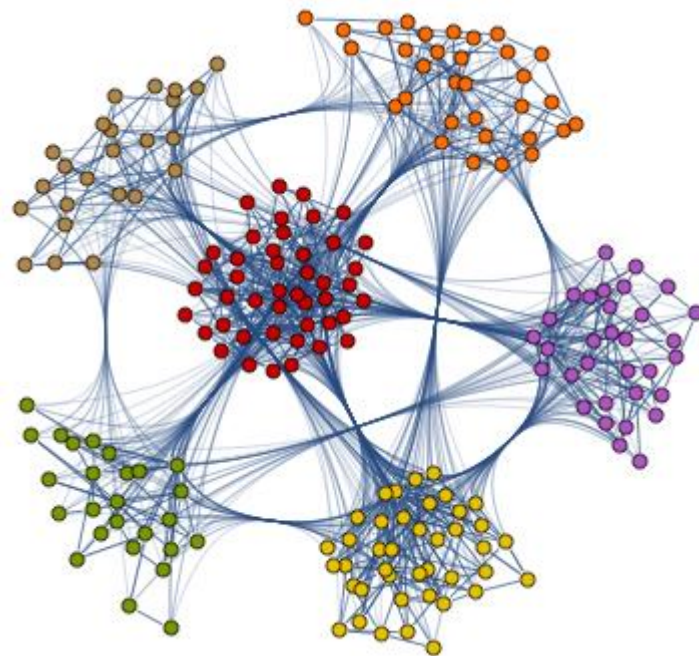
Аспирант первого года обучения Казаков М.А.

Утвержденная тема диссертации: Алгоритмы кластеризации в моделях случайных графов

Научный руководитель: проф., д.ф.-м.н. Калягин В.А.

# Задача кластеризации на графах

Разделить исходный граф на достаточно плотные подграфы, имеющие относительно малое количество связей между собой.



# Задача кластеризации на графах

## Вариации:

- Веса ребер
- Ориентация ребер
- Атрибуты вершин
- Динамические сети
  
- Пересекающиеся сообщества
- Нечеткая кластеризация

- Divisive algorithms
- Modularity-based methods
- Spectral algorithms
- Dynamic algorithms
- Methods based on statistical inference
- ...

Fortunato, Santo. "Community detection in graphs." *Physics reports* 486.3 (2010): 75-174.



# Benchmarks

- Real-life
- Generated

# Benchmarks: generated

- Stochastic block model
- Degree-corrected block model

Karrer, Brian, and Mark EJ Newman. "Stochastic blockmodels and community structure in networks." *Physical Review E* 83.1 (2011): 016107.

- LFR

Lancichinetti, Andrea, Santo Fortunato, and Filippo Radicchi. "Benchmark graphs for testing community detection algorithms." *Physical review E* 78.4 (2008): 046110.

# Stochastic block model

- The number  $n$  of vertices
- A partition of the vertex set  $V = \{1, \dots, n\}$  into disjoint subsets  $C_1, \dots, C_r$ , called *communities*
- A symmetric  $r \times r$  matrix  $P$  of inter-community edge probabilities

$$P_{ij} = \Pr\{\text{edge}(u, v)\}, \quad u \in C_i \text{ and } v \in C_j$$

# Erdős–Rényi model

- $|V| = n$  vertices
- $r = 1$  communities
- $P = p$



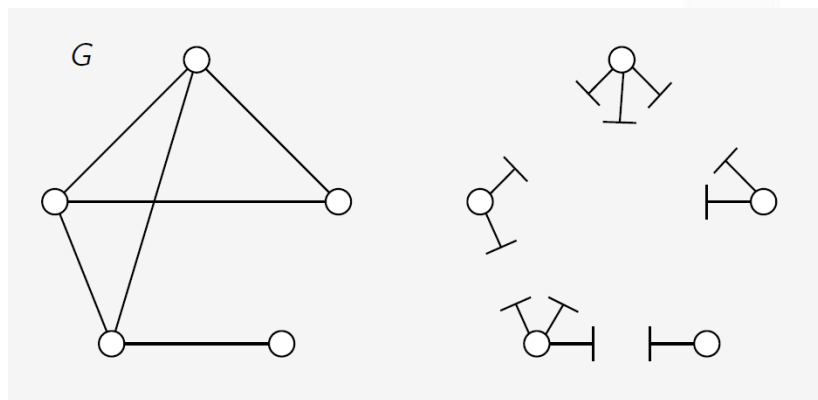
# Planted partition model

- $|V| = n$  vertices,  $r = 2$  communities
- $|C_1| = |C_2| = n/2$
- $P = \begin{pmatrix} p_{in} & p_{out} \\ p_{out} & p_{in} \end{pmatrix}$

# Modularity

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(C_i, C_j)$$

Null model:



$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j)$$

# Detectability of community structure

- $A$  – adjacency matrix  $n \times n$ , where  $A_{ij} = 1$  if there is an edge  $(i, j)$  and 0 otherwise

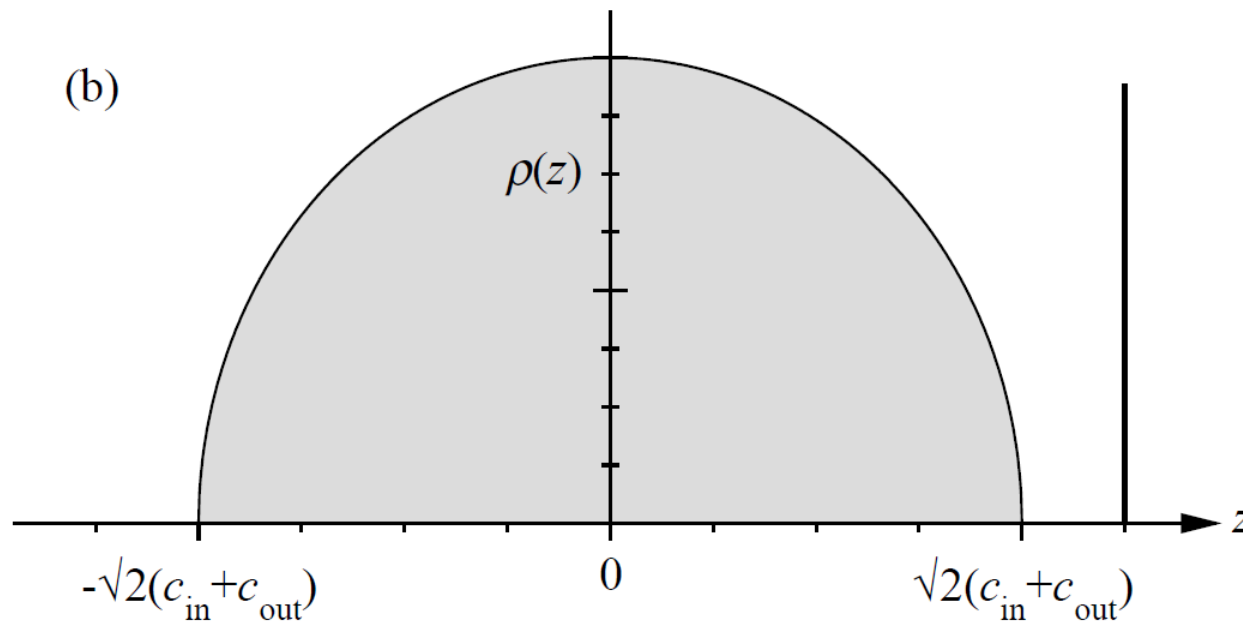
- $\langle A \rangle = \begin{pmatrix} p_{in} & p_{out} \\ p_{out} & p_{in} \end{pmatrix}$

- $\langle A \rangle = \frac{1}{2}(c_{in} + c_{out})\mathbf{1}\mathbf{1}^T + \frac{1}{2}(c_{in} - c_{out})\mathbf{u}\mathbf{u}^T$ ,  $c_{in} = np_{in}$ ,  $c_{out} = np_{out}$

$$\mathbf{1} = \underbrace{(1, 1, 1, \dots)}_n / \sqrt{n}, \quad \mathbf{u} = \underbrace{(1, 1, \dots, 1)}_{n/2}, \underbrace{(-1, -1, \dots, -1)}_{n/2} / \sqrt{n}$$

$$\mathbf{B} = \frac{1}{2}(c_{in} - c_{out}) \mathbf{u}\mathbf{u}^T + \mathbf{X} = \mathbf{A} - \frac{1}{2}(c_{in} + c_{out}) \mathbf{1}\mathbf{1}^T$$

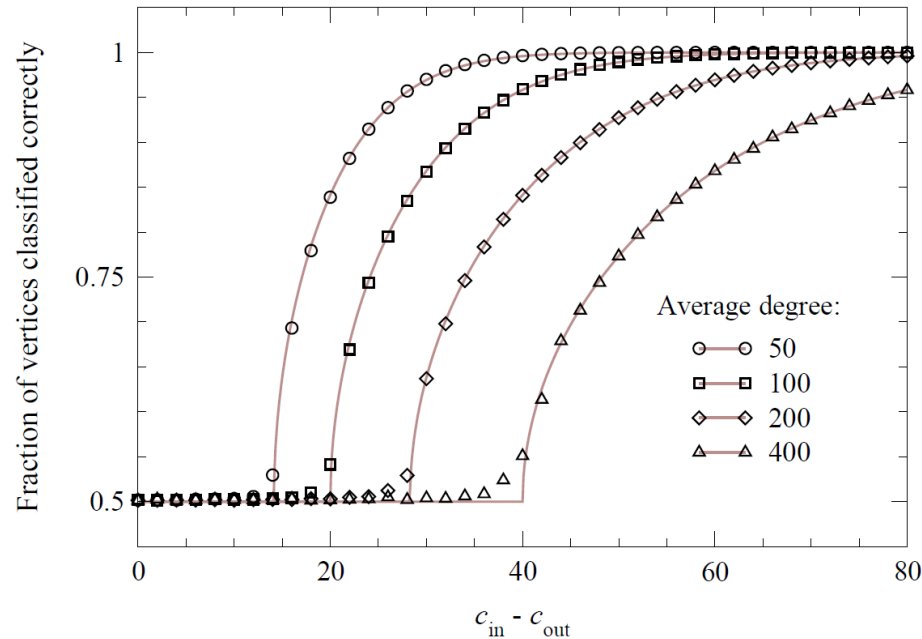
# Detectability of community structure



$$c_{in} - c_{out} = \sqrt{2(c_{in} + c_{out})}$$

$$c_{in} - c_{out} = \sqrt{q[c_{in} + (q - 1)c_{out}]}$$

# Detectability of community structure



$$\frac{1}{2} [1 + \operatorname{erf} \sqrt{\alpha^2 / 2(1 - \alpha^2)}]$$

$$\alpha^2 = \frac{(c_{\text{in}} - c_{\text{out}})^2 - 2(c_{\text{in}} + c_{\text{out}})}{(c_{\text{in}} - c_{\text{out}})^2}, \operatorname{erf} x - \text{Gaussian error function}$$

# Detectability of community structure

Nadakuditi, Raj Rao, and Mark EJ Newman. "Graph spectra and the detectability of community structure in networks." *Physical review letters* 108.18 (2012): 188701.

Decelle, Aurelien, et al. "Inference and phase transitions in the detection of modules in sparse networks." *Physical Review Letters* 107.6 (2011): 065701.

Другие алгоритмы:

- Min Cut
- Normalized Cut

Другие модели:

- Degree-corrected block model
- LFR
- ERGMs



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

Спасибо  
за внимание!