

# Методы выявления сообществ на графах

Шестаков Андрей



ВЫСШАЯ ШКОЛА ЭКОНОМИКИ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Научно-учебная группа «Методы анализа и визуализации веб-корпусов»  
**Факультет Компьютерных Наук НИУ ВШЭ**

- 1 Задача выявления сообществ
- 2 Методы выявления сообществ
- 3 Плотные подграфы в сетях
- 4 Поиск и интерпретация мотивов

- 1 **Задача выявления сообществ**
- 2 Методы выявления сообществ
- 3 Плотные подграфы в сетях
- 4 Поиск и интерпретация мотивов

# Задача выявления сообществ

## Community Structure

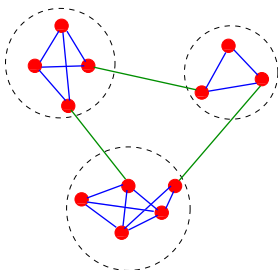


Рис. 1: Простой пример сети со структурой сообществ

- Реальные сети так же часто обладают свойством Community Structure как и другими важными свойствами (small-world & scale-free networks)
- Первое упоминание свойства Community Structure в статье [Girvan and Newman, 2002]

- 1 Задача выявления сообществ
- 2 Методы выявления сообществ**
- 3 Плотные подграфы в сетях
- 4 Поиск и интерпретация мотивов

# Задача выявления сообществ

## Подходы к решению задачи

- Разбиение графов
  - Разрезы
- Иерархическая кластеризация сети
  - Агломеративная и дивизивная кластеризация
- Применение канонических алгоритмов кластеризации
- Спектральная кластеризация
  - Матрица Лапласа  $\mathbf{L} = \mathbf{D} - \mathbf{A}$
- Стохастические алгоритмы
  - Матрица переходов  $\mathbf{T} = \mathbf{AD}^{-1}$
- Оптимизация модулярности
  - $Q = \frac{1}{2m} \sum_{ij} \left( a_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j)$

- 1 Задача выявления сообществ
- 2 Методы выявления сообществ
- 3 Плотные подграфы в сетях
- 4 Поиск и интерпретация мотивов

# Плотные подграфы в сетях

Пусть дан граф  $G = (V, E)$

## **n-clique**

**n-clique** – это максимальное подмножество вершин  $S \subseteq V$ , такое что:

$$\forall u, v \in S \quad d_G(u, v) \leq n$$

## **n-clan**

**n-clan** – это такая **n-clique**, что диаметр подграфа  $\text{diam}(G[S]) \leq n$

## **n-club**

**n-club** – это максимальный подграф  $G'$ , т.ч.  $\text{diam}(G') \leq n$



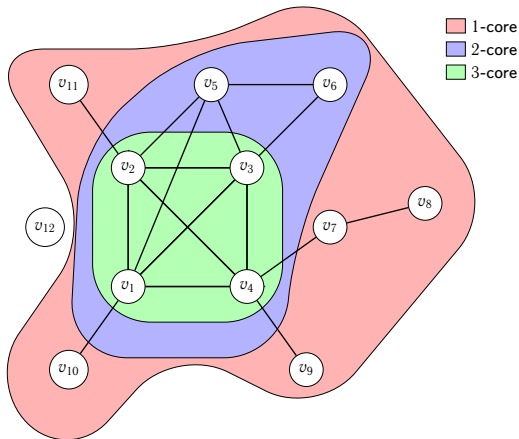
# Плотные подграфы в сетях

## k-core

k-core – максимальный подграф  $G_k$ , такой что все его вершины связаны хотя бы с  $k$  вершинами из  $V_k$

## k-plex

k-plex – это максимальный подграф  $G_k$ , такой что все его вершины связаны хотя бы с  $n - k$  вершинами из  $V_k$



**Рис. 2:** *k*-cores and *k*-plexes example  
 1-plexes  $\sim$  cliques  
 2-plex:  $\{v_1, v_2, v_3, v_4, v_5\}$

- 1 Задача выявления сообществ
- 2 Методы выявления сообществ
- 3 Плотные подграфы в сетях
- 4 Поиск и интерпретация мотивов**

# Плотные подграфы в сетях

Нет слайдов =(