

# Методы анализа социальных сетей. Базовые понятия

Николай Ильич Базенков, к.т.н.

Институт проблем управления им. В.А. Трапезникова  
Российской академии наук

16 сентября 2015 г.



# Содержание курса

## Лекция 1 Базовые понятия

- Базовые понятия теории графов
- Графы социальных отношений

## Лекция 2 Структурные свойства социальных сетей

## Лекция 3 Модели формирования сетей

## Лекция 4 Сообщества в социальных сетях

# Сети и графы

Граф или сеть – формальное представление структуры отношений (дружбы, коммуникации, управления и др.)

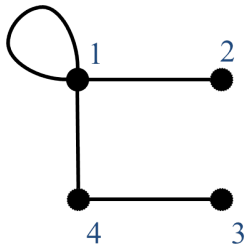
Граф обозначается как  $G = (N, E)$ , где

- $N = \{1, \dots, n\}$  – множество вершин,  $|N| = n$
- $E$  – множество ребер. Ребро  $(i, j) \in E$  обозначает, что узлы  $i$  и  $j$  связаны интересующим нас отношением.

# Неориентированные графы

Неориентированный граф – ребра заданы без учета ориентации.

Симметричные отношения: Дружба в Facebook и Vk, соавторство и др.



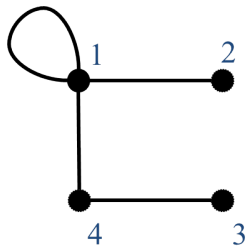
- $N = \{1, 2, 3, 4\}$
- $E = \{(1, 1), (1, 2), (1, 4), (3, 4)\}$

# Представление графов

Список ребер:  $N = \{1, 2, 3\}$ ,  $E = \{(1, 2), (2, 3)\}$

Матрица смежности (adjacency matrix):  $A = \|a_{ij}\|_{n \times n}$

$$a_{ij} = \begin{cases} 1, & \text{если } (i, j) \in E; \\ 0, & \text{если } (i, j) \notin E. \end{cases} \quad (1)$$



$$A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

# Степени

**Степень узла** – количество его ребер в графе  $G = (N, E)$

$$d_i(G) = |\{(i, j) : (i, j) \in E\}| = \quad (2)$$

$$= \sum_{j \in N} a_{ij} \quad (3)$$

**Соседи узла** – узлы, связанные с ним ребром

$$N_i(G) = |\{j : (i, j) \in E\}| \quad (4)$$

## Соотношения количества ребер и степеней

Общее количество ребер  $|E| = m$  в графе  $G = (N, E)$

$$2m = \sum_{i \in N} d_i \quad (5)$$

Средняя степень  $\langle d \rangle$

$$\langle d \rangle = \frac{1}{n} \sum_{i \in N} d_i \quad (6)$$

Соотношение

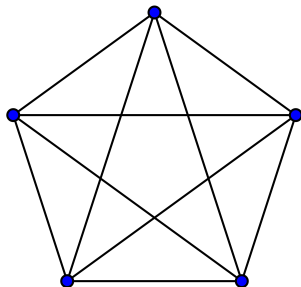
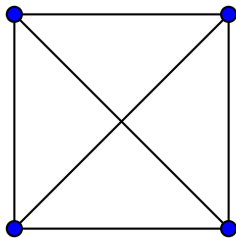
$$\langle d \rangle = \frac{2m}{n} \quad (7)$$

# Полные графы

Полный граф – любые две вершины связаны ребром. Обозначается  $K_n$ .

Количество ребер в полном графе

$$|E| = \binom{n}{2} = \frac{n(n-1)}{2} \quad (8)$$





# Плотность графа

**Плотный граф.** Граф  $G = (N, E)$ ,  $|N| = n$  называется плотным, если

$$|E| \sim n^2 \quad (9)$$

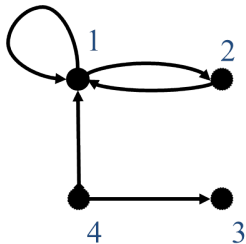
**Разреженный граф.** Граф  $G = (N, E)$ ,  $|N| = n$  называется разреженным, если

$$|E| \sim n \quad (10)$$

# Ориентированные графы

**Ориентированный граф** – ребра имеют ориентацию:  $(i, j)$  и  $(j, i)$  это разные ребра

**Несимметричные отношения** – цитирование статей, комментирование постов



$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

# Входящие и исходящие степени

**Входящая степень** – количество входящих ребер узла в орграфе  
 $G = (N, E)$

$$d_i^{in}(G) = |\{(j, i) : (j, i) \in E\}| = \quad (11)$$

$$= \sum_{j \in N} a_{ji} \quad (12)$$

**Исходящая степень** – количество исходящих ребер узла в орграфе  
 $G = (N, E)$

$$d_i^{out}(G) = |\{(i, j) : (i, j) \in E\}| = \quad (13)$$

$$= \sum_{j \in N} a_{ij} \quad (14)$$

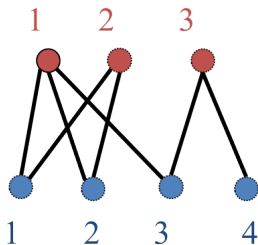
**Петли** – входят в оба показателя

# Двудольные графы

**Двудольный граф** – вершины разделены на два непересекающихся множества  $N = N_1 \cup N_2$  и ребра связывают только элементы из разных множеств

**Примеры:** статьи – авторы, мужчины – женщины, сотрудники – организации

**Матрица инцидентности**  $B = ||b_{ij}||$ , где  $i \in N_1$  и  $j \in N_2$



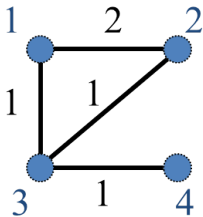
$$B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

# Взвешенные графы

**Взвешенный граф** – ребрам приписаны веса  $w : E \rightarrow R$

**Интенсивность и знак отношений** – количество общих статей, эмоциональная окраска и частота комментирования и др.

Вес = количество общих статей  
(см предыдущий рисунок)

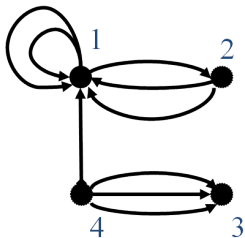


$$A = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

# Мультиграфы

**Мультиграф** – между вершинами может быть несколько ребер

**Множественные события** – ссылки между веб-страницами, комментарии к постам. Практически любое сложное взаимодействие порождает мультиграфы



$$A = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 \end{pmatrix}$$

# Резюме

**Неориентированные графы** – много однозначно интерпретируемых показателей

**Ориентированные графы** – почти все показатели хорошо обобщаются.

**Взвешенные графы** – некоторые показатели теряют смысл. Часто веса игнорируются.

**Двудольные графы** – можно рассматривать проекции в виде взвешенных графов

**Мультиграфы** – сводятся к взвешенным

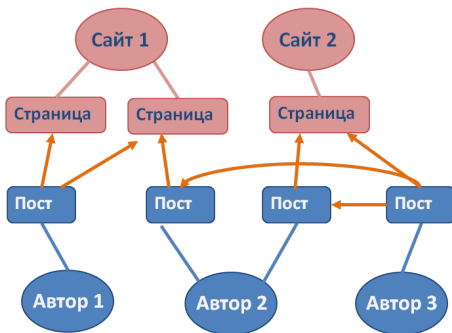
# Онлайновые социальные сети

**Формальные связи** – указаны пользователями в явном виде  
(дружба, принадлежность группе)

**Информационные связи** – возникают в результате коммуникации  
(цитирование, комментирование, лайки и др.)



## Сети цитирования



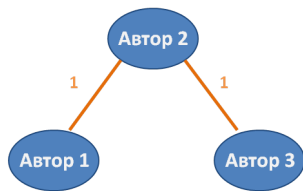
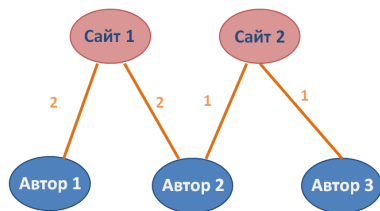
# Проекции двудольных графов

Исходный граф –  $G^0 = (N_1 \cup N_2, E)$ ,  $B_{ij}$  – матрица инцидентности

Новое множество узлов – одна из “долей”. Пусть  $N_1$

Вес связи – количество цепочек длины два или “общих” узлов

$$w_{ij} = \sum_{k \in N_2} b_{ik} b_{kj} \quad (15)$$



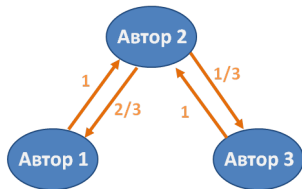
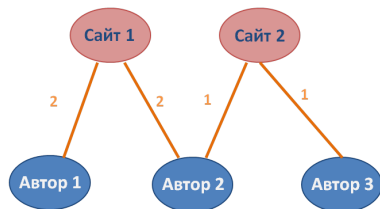
## Оценка “сходства интересов”

Исходный граф –  $G^0 = (N_1 \cup N_2, E)$ ,  $B_{ij}$  – матрица инцидентности

Множество узлов – одна из “долей”. Пусть  $N_1$

Вес связи – относительная доля “однонаправленных” ребер

$$w_{ij} = \frac{1}{d_i(G^0)} \sum_{k \in N_2} b_{ik} b_{kj} \quad (16)$$



# Заключение

**Онлайновые социальные сети** – сложная структура объектов и отношений

**Графы социальных сетей** – упрощенное представление, отражающее интересующие нас аспекты сети

**Показатели** – теоретико-графовые метрики, отражающие ключевые характеристики сети и отдельных узлов