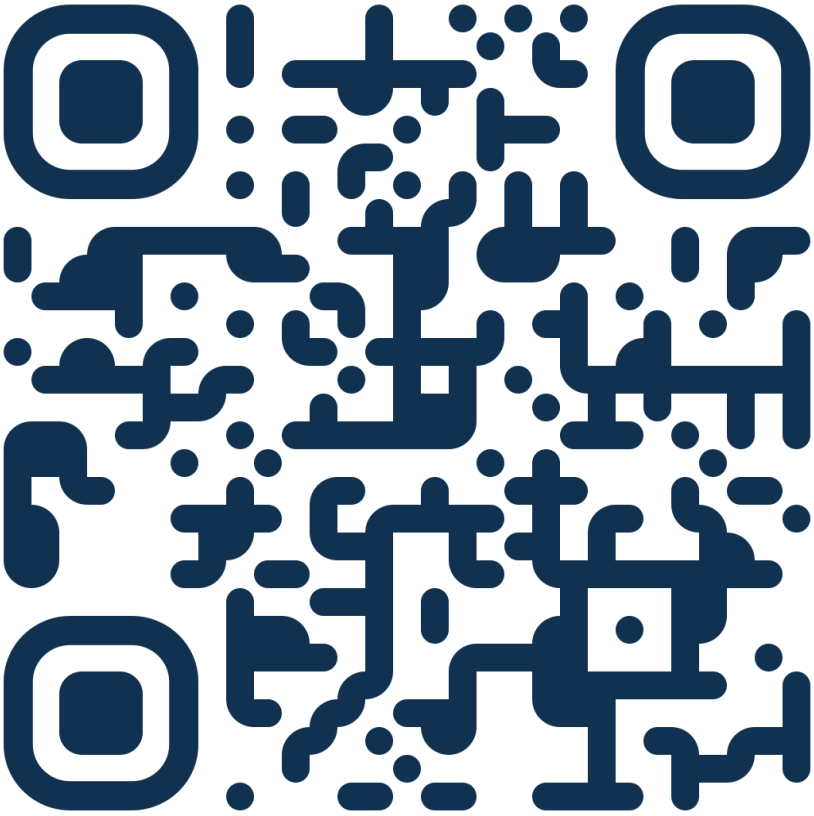
Машинное обучение в семантическом и сетевом анализе

Лекция 1: Базовые понятия теории сетей, сетевые модели и свойства реальных графов

Никита Владимирович Блохин

Обзор курса

Ссылка на материалы курса



Что мы будем делать?

> О чем этот курс?
>
> Мы пройдем путь от классической **теории сложных сетей** до современных методов **глубокого обучения на графах**.

Курс разделен на два смысловых блока:

1. Классический анализ

- Основные понятия теории сетей

- Сетевые модели
- Сообщества и кластеризация
- Случайные блуждания и распространение информации

2. Graph ML & DL

- Векторные представления узлов и графов
- Задачи машинного обучения на графах
- Графовые нейросети

Сети как объект исследования

Сети и графы

Сеть (Network)

Обычно употребляется по отношению к реальной системе.

- **Примеры:** социальная сеть, компьютерная сеть, дорожная сеть
- **Термины:**
  - сеть (network)
  - узел (node)
  - связь (link)

Граф (Graph)

Математическое представление сети (модель).

- **Примеры:** веб-граф, граф знаний
- **Термины:**
  - граф (graph)
  - вершина (vertex)
  - ребро (edge)

Ключевые определения

Формально граф G — это пара множеств (V, E), где:

- V — множество **вершин** (nodes);
- E — множество **ребер** (edges), соединяющих вершины.

> **Соглашение**
>
> В рамках данного курса мы будем использовать термины **«Граф»** и **«Сеть»** как **синонимы**.
> То же касается пар:
> - Вершина ↔ Узел
> - Ребро ↔ Связь

Сети повсюду

Интернет и веб-граф

- Физические каналы связи
- Страницы и гиперссылки

Коммуникационные сети

- Телефонные звонки
- Email-переписка
- Сообщения в мессенджерах

Социальные сети * Сети знакомств * Граф “дружбы” * Взаимодействия: лайки, репосты, комментарии

Netflix / Amazon / Spotify

- Товары — Покупки
- Зрители — Фильмы

Семантические сети

- Смысловые связи слов (WordNet)
- Базы знаний (Knowledge Graphs)

Инфраструктура

- Транспорт (авиа, ж/д, авто)
- Энергетика (ЛЭП)

Биология

- Взаимодействие белков
- Нейронные сети мозга

Промежуточные выводы

Сети — это универсальный язык для описания систем в природе, обществе и технике.

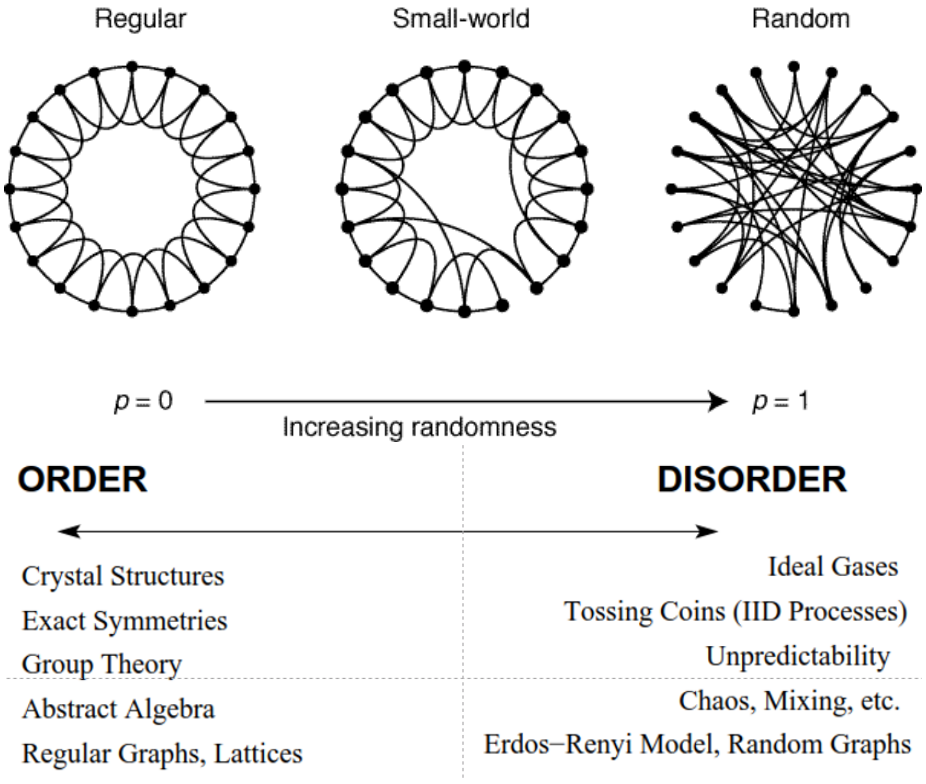
1. **Объект исследования:**
   - Реальные системы (Интернет, Facebook, мозг).
   - Гетерогенные данные (разные типы узлов и связей).
2. **Модель:**
   - Граф $G = (V, E)$.
3. **Терминология:**
   - Мы используем слова **граф** и **сеть** как синонимы.
4. **Разнообразие:**
   - От социальных взаимодействий до биологических процессов.

Сложные системы

Что такое “сложность”?

Есть математический аппарат для “крайностей”:

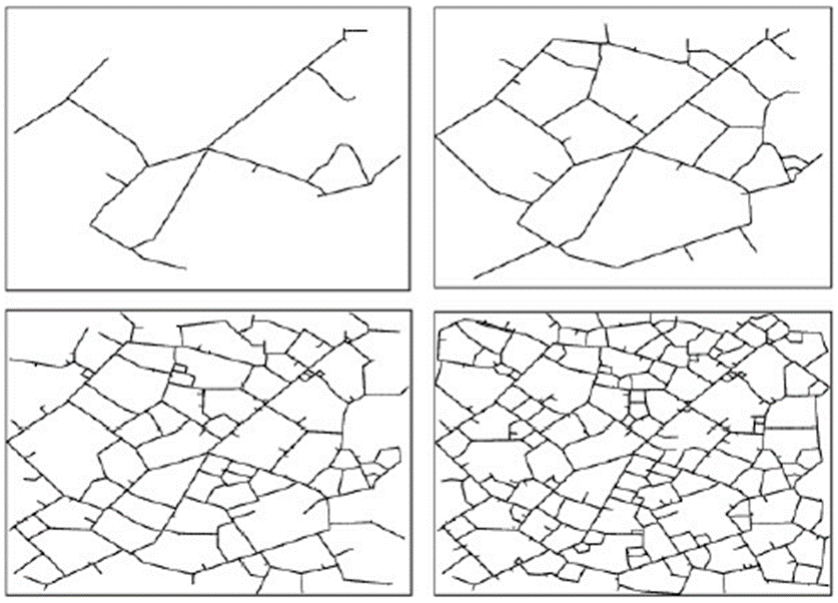
1. **Регулярные решетки:** полный порядок.
2. **Случайные графы:** полный хаос. **Сложные сети** находятся посередине. У них нетривиальная топология (кластеризация, хабы, сообщества).



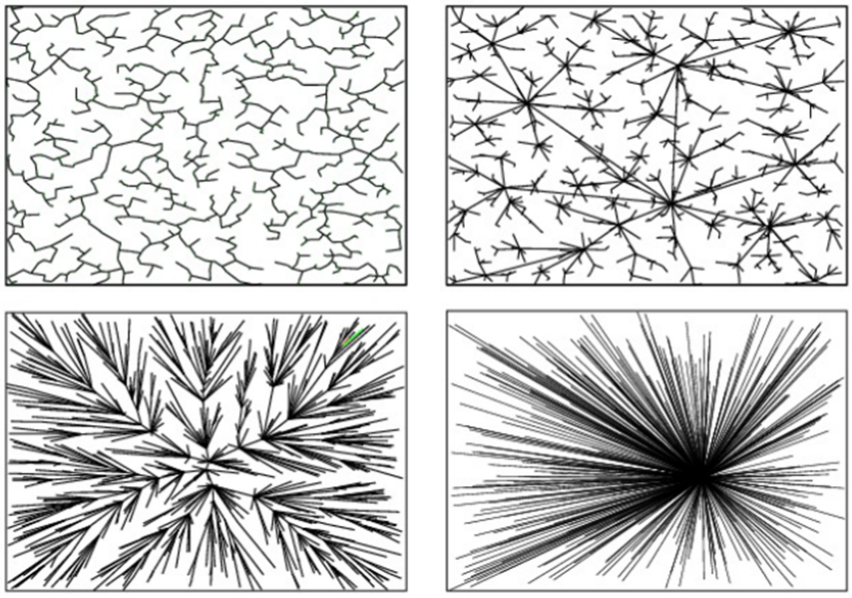
Универсальность

Почему одна теория описывает и биологию, и интернет?

> **i** Гипотеза универсальности
>
> 1. Сложные сети — продукт процесса **роста**.
> 2. Правила роста (например, “богатый становится богаче”) универсальны для разных доменов.
> 3. Схожие правила роста → схожая топология.



Рост дорожной сети

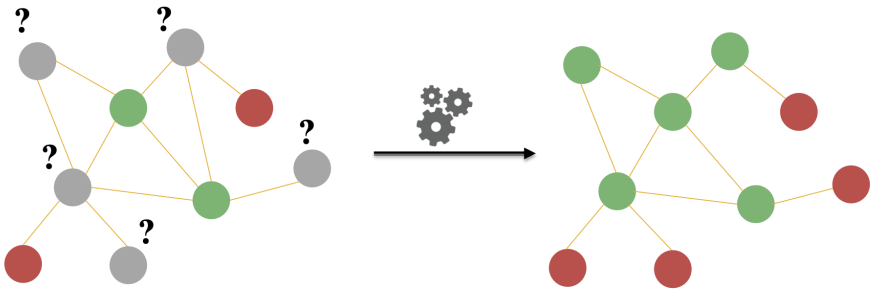


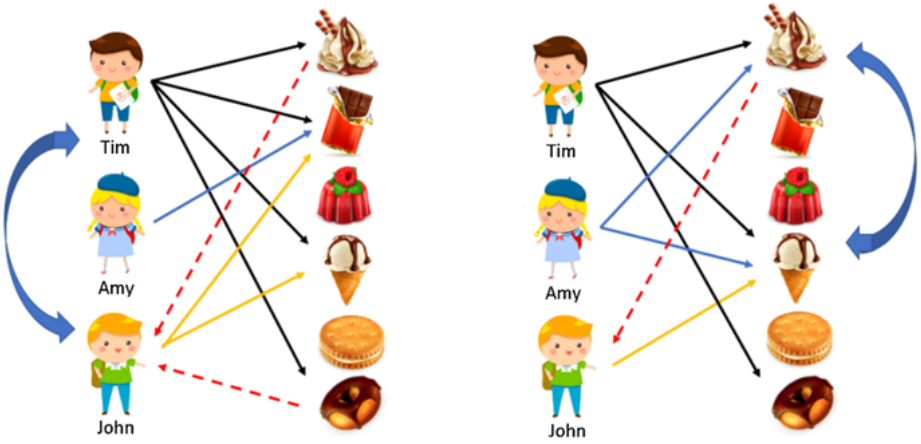
Алгоритмический рост

Постановка задач на графах и специфика данных

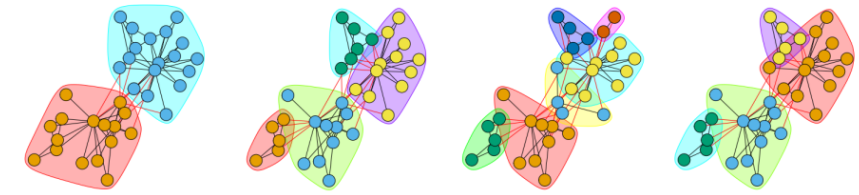
Уровни задач

1. **Node Level.** Пример: классификация узлов
2. **Edge Level.** Пример: предсказание связей
3. **Sub-graph Level.** Пример: выделение сообществ
4. **Graph Level.** Пример: классификация графов



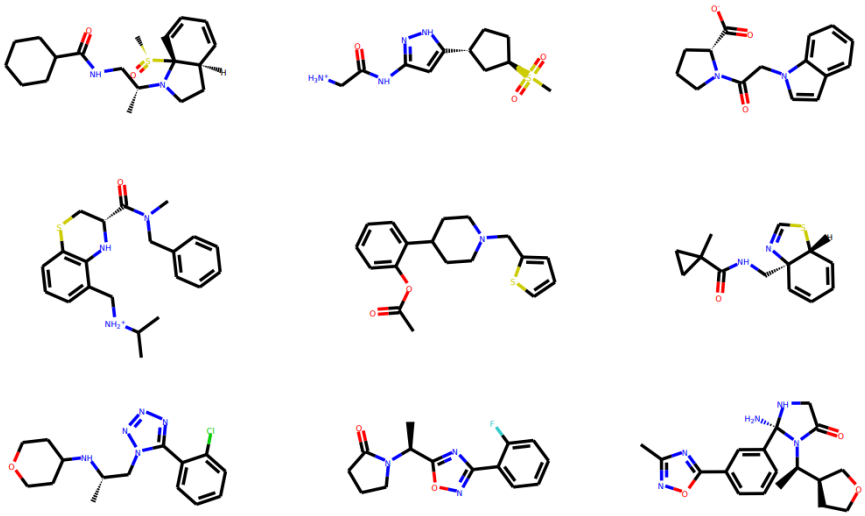


(a) User-based filtering **(b)** Item-based filtering



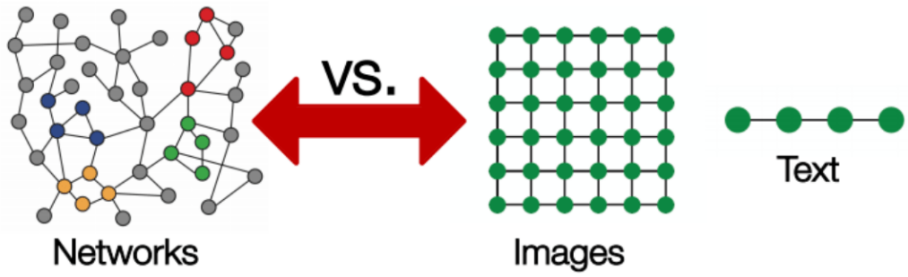
a) Ground Truth b) Louvain c) Girvan-Newman d) Walktrap

Modularity: 0.37 Modularity: 0.45 Similarity: 0.77 Modularity: 0.34 Similarity: 0.68 Modularity: 0.44 Similarity: 0.79



В чем отличие от классического ML

- **Произвольный размер:** Размер графа зависит от того, как система развивалась.
- **Сложная топология:** Нет линейного порядка, нет сетки
- **Проблема изоморфизма:** Один и тот же граф можно представить по-разному.
- **Динамика:** Графы часто меняются во времени.



Граф свойств (Property Graph)

> **Property Graph** — это **типовая модель данных** для большинства прикладных задач.

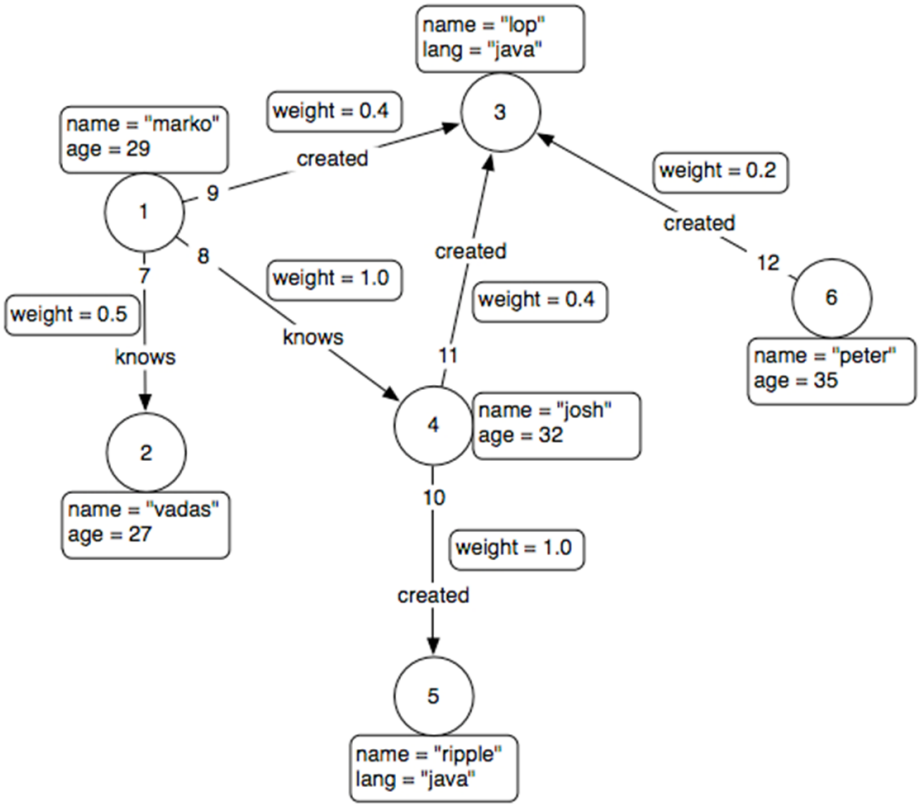
Структура модели

Узлы:

- Уникальный идентификатор.
- Множество входящих и исходящих связей.
- Набор свойств (Attributes): пары `key: value`.

Связи:

- Уникальный идентификатор.
- Стартовый и конечный узел (направление).
- Тип связи: например, `FRIEND`, `BOUGHT`.
- Набор свойств: например, `weight`, `date`.



Ключевые выводы

1. **Сети универсальны:** единый мат. аппарат описывает биологию, социум и технологии.
2. **Сложность:** мы изучаем системы с нетривиальной топологией (не случайные, не регулярные).

3. **Данные:** сам по себе граф - нетривиальная структура, а модель графа свойств достаточно сложна для обработки ввиду наличия атрибутов

Далее мы рассмотрим **математические модели**, объясняющие устройство типовых реальных графов.

Зачем нужны модели сетей?

Использование моделей

1. Эталон для сравнения

Имея сеть, мы можем сравнивать ее свойства с аналогичными свойствами случайной сети такого же размера.

2. Понимание механизмов

Данные — это следствие. Модель описывает *процесс* (например, “богатый становится богаче”), который привел к таким данным.

3. Прогнозирование

Как поведет себя сеть при атаке на узлы? Как сеть будет развивать во времени?

Что отличает реальные сети?

Большинство сетей (социальные, технологические, биологические) имеют общие свойства, которых нет у случайных графов.

Понимание этих свойств дает возможность строить более адекватные модели для конкретной прикладной задачи.

1. Наличие хабов.
2. Малый диаметр (эффект “тесного мира”).
3. Высокая кластеризация (“друзья моих друзей — мои друзья”).
4. Наличие компоненты большого размера

Степень узла

Распределение степеней узлов

Степень узла k_i — количество связей (соседей) у узла i.

Пусть N — общее количество узлов, а n_k — количество узлов со степенью k.

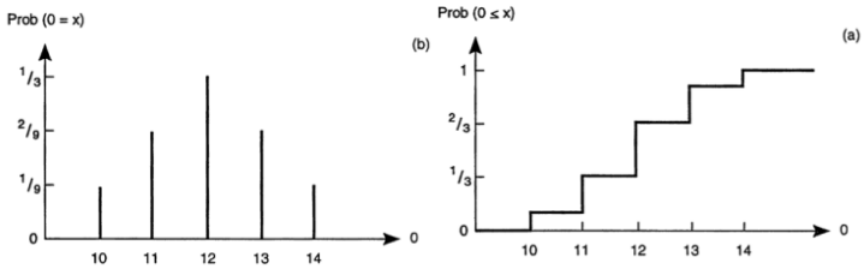
P_k - вероятность того, что случайно выбранный узел имеет степень k:

$$P(k) = \frac{n_k}{N}$$

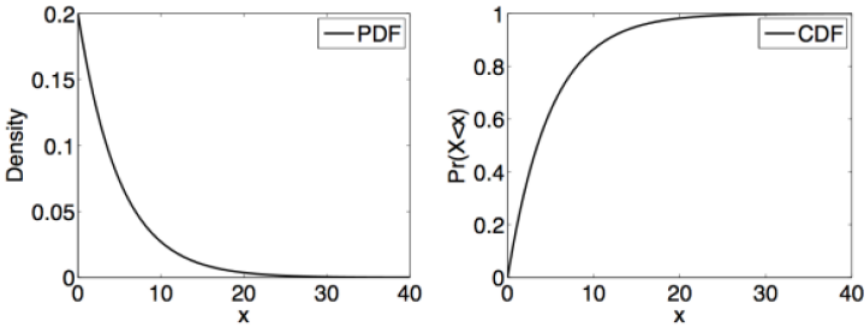
По определению $\sum_k P(k) = 1$

> **Уведомление**
>
> Использование вероятности $P(k)$ вместо абсолютного количества n_k позволяет сравнивать топологию сетей **разного размера** .

Дискретная случайная величина:



Непрерывная случайная величина:



Степени в ориентированных графах

В направленных сетях (Web, Twitter) связи имеют направление. Степень разделяется на две компоненты.

Определения:

- In-degree k_i^{in}: Число входящих связей (популярность).
- Out-degree k_i^{out}: Число исходящих связей (активность).

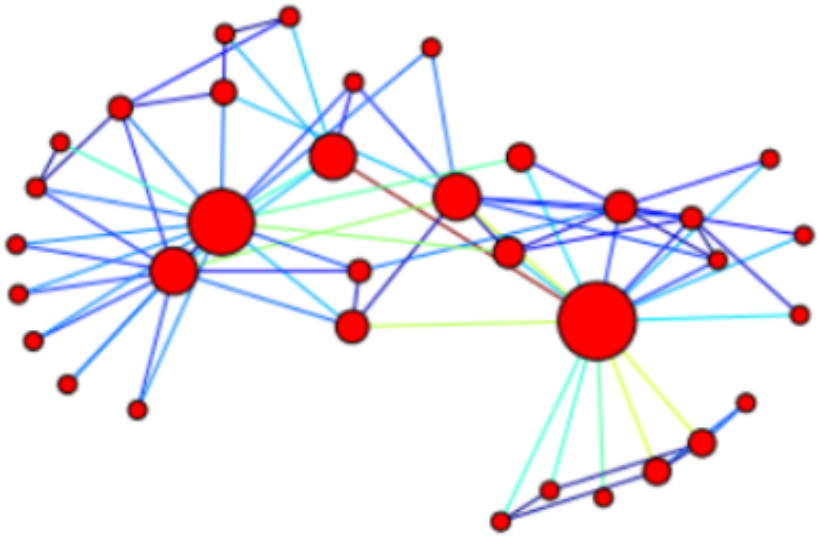
Баланс связей: Сумма входящих степеней равна сумме исходящих и равна общему числу связей L:

$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out}$$

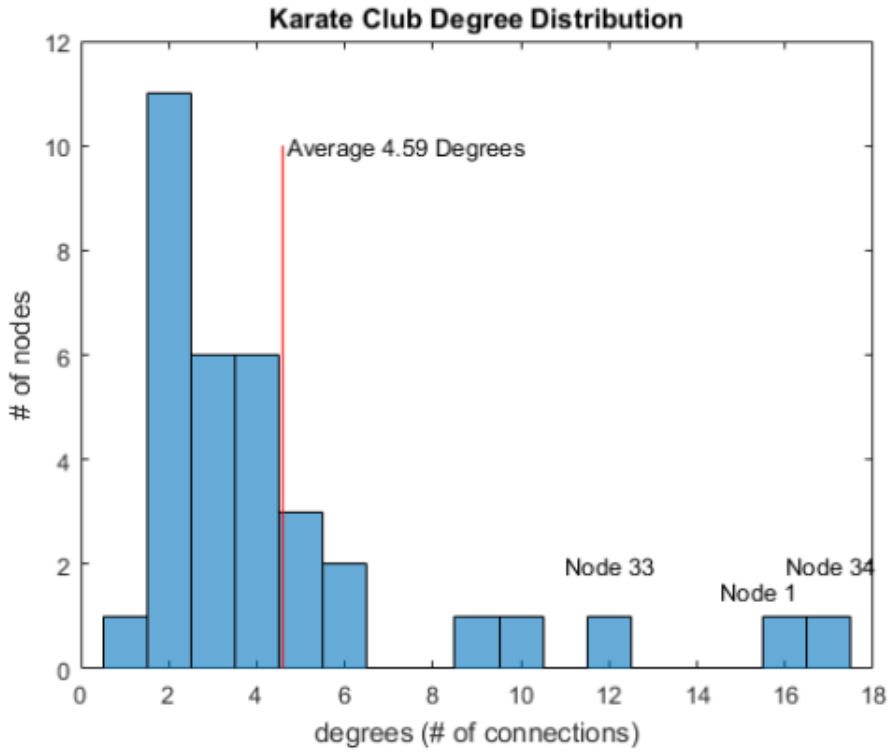
В таком случае анализируем две независимые функции: $P(k^{in})$ и $P(k^{out})$. Часто они имеют совершенно разную форму (например, в Twitter у многих мало читателей, но они читают многих).

Пример: Zachary Karate Club

Рассмотрим классический граф социального взаимодействия (34 узла, 78 связей).



Визуализация структуры



Распределение степеней (P_k)

> Инсайт
>
> Визуализация позволяет интуитивно находить **важные узлы** (хабы) — тех, у кого больше всего связей.
> Степень узла (k) — это простейшая метрика **центральности**. Более сложные метрики важности мы разберем в следующей лекции.

Степенной закон распределения (Power Law)

Многие реальные сети (Интернет, цитирование, связи белков) подчиняются степенному закону распределения.

$$P(k) = Ck^{-\gamma}$$

Если прологарифмировать обе части: $\ln P(k) = -\gamma \ln k + \ln C$

В лог-лог координатах график превращается в **прямую линию** с наклоном $-\gamma$.

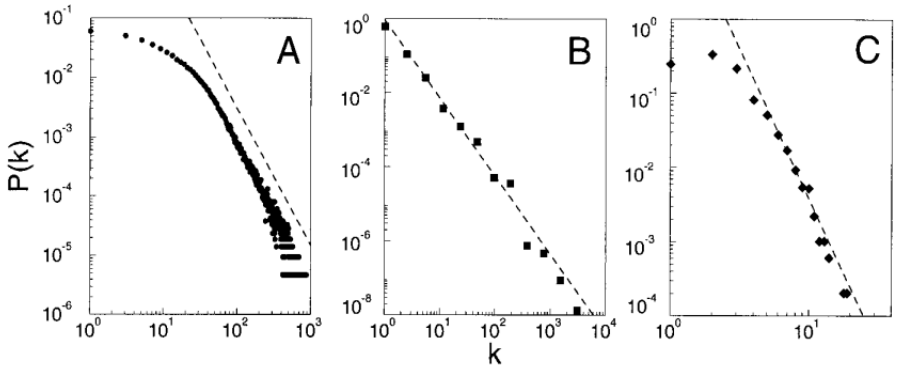


Fig. 1. The distribution function of connectivities for various large networks. **(A)** Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. **(B)** WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). **(C)** Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

Распределение степеней узлов

> **Свойство безмасштабности (Scale-free)**
>
> Такое распределение обладает специфичным свойством: есть много узлов с малой степенью и тяжелый хвост из редких, но гигантских хабов.

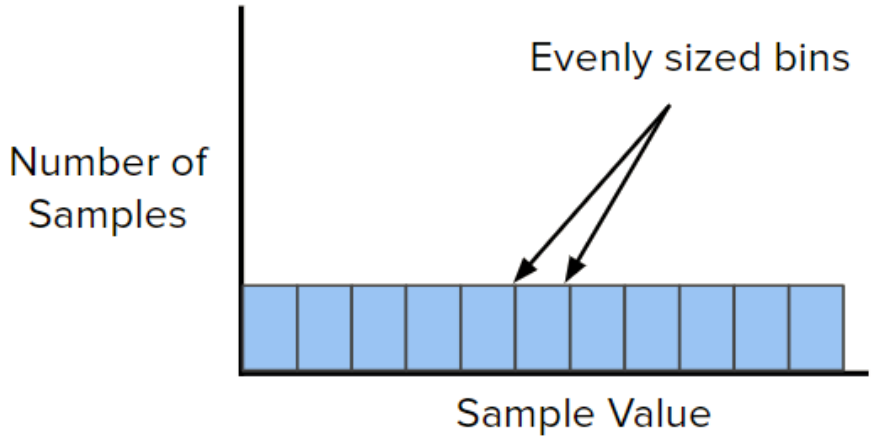
> **Уведомление**
>
> Сети с хабами устойчивы к случайным сбоям (вероятность отказа хаба мала), но критически уязвимы к направленным атакам на хабы. Хабы - это узлы, которые могут распространить информацию очень быстро по всей сети.

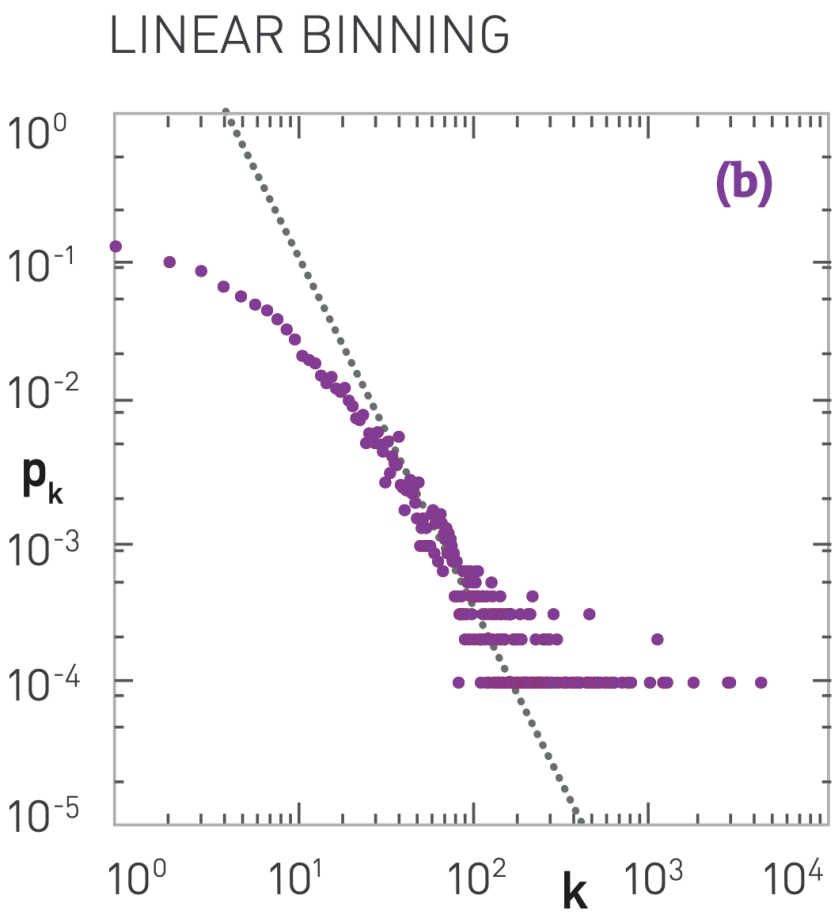
Проблемы визуализации и оценки

Из-за дискретности распределения степеней узлов без выбора специального представления распределения не дает достаточно информации о форме хвоста распределения

Проблема

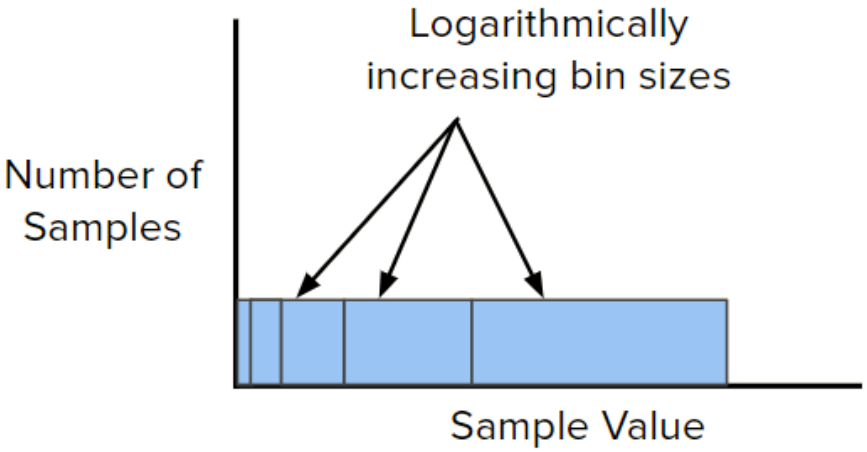
На хвосте данных мало (редкие события). При обычном (линейном) разбиении на бины мы получаем сильный шум в правой части графика.

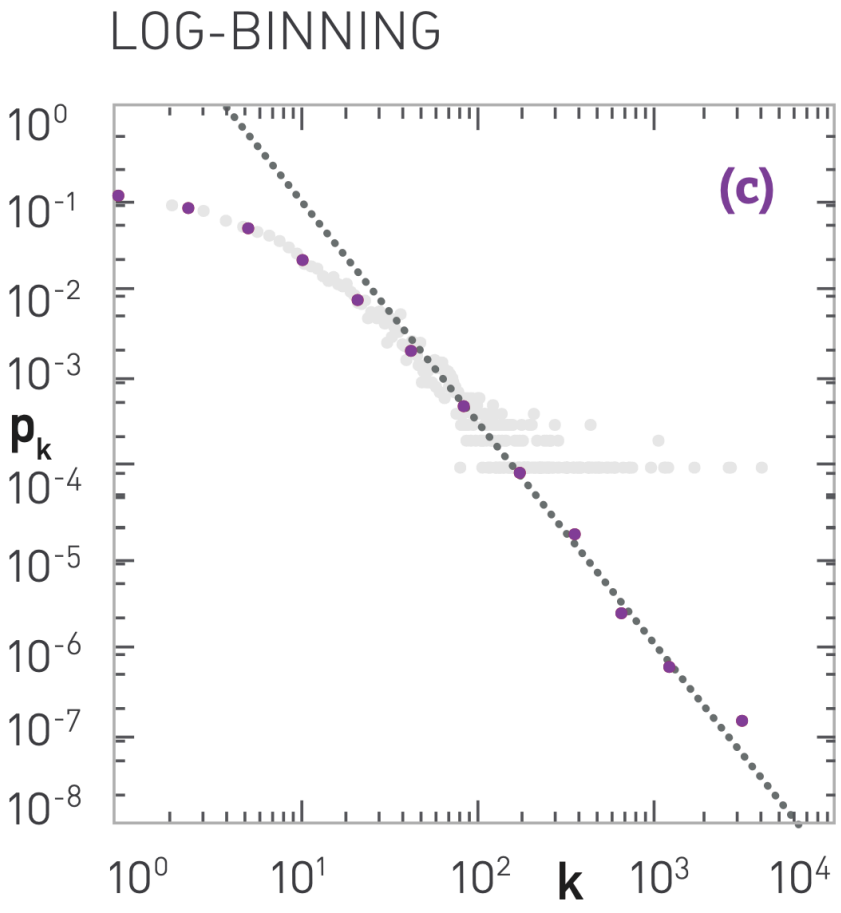

Evenly sized bins
Number of Samples
Sample Value



Log-Binning

Ширина корзины растет экспоненциально $(1, 2, 4, 8...)$. Это усредняет шум в хвосте и восстанавливает прямую линию.


Logarithmically
increasing bin sizes
Number of
Samples
Sample Value

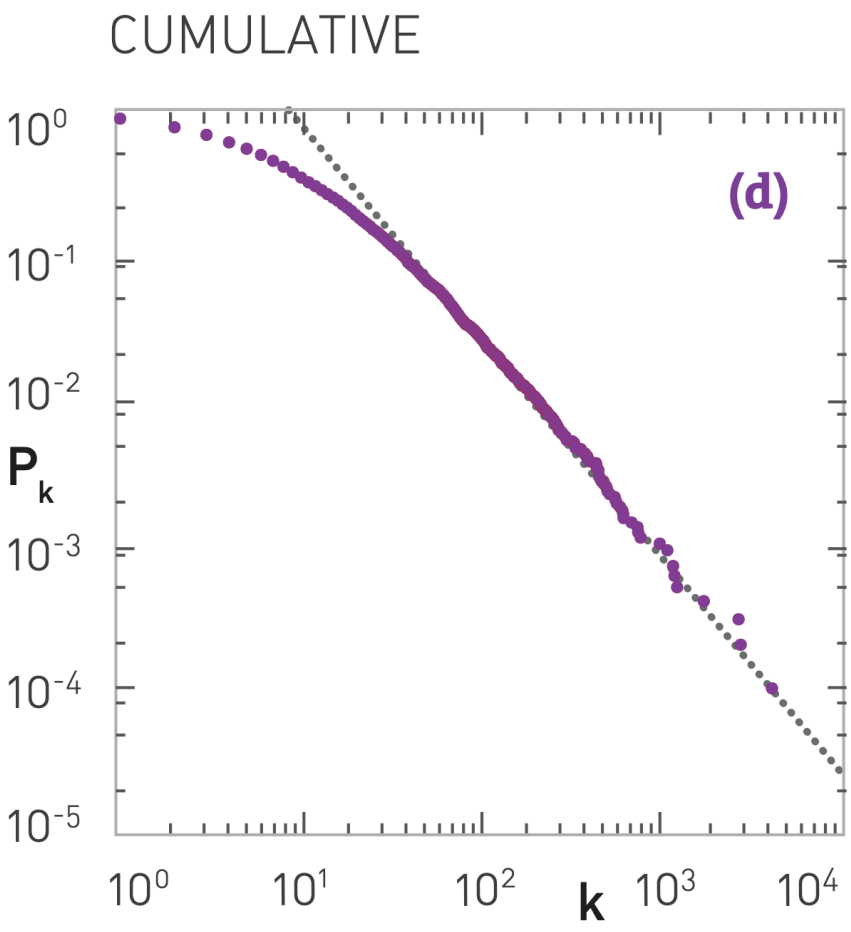

LOG-BINNING
(c)
10^{0}
10^{-1}
10^{-2}
10^{-3}
p_k
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{0}
10^{1}
10^{2}
k
10^{3}
10^{4}

CCDF

Еще один вариант - использовать CCDF (Complementary Cumulative Distribution Function):

$$P(X > k) = \sum_{x>k} P(x)$$

- Убирает шум без потери данных (биннинг не нужен).
- Для степенного закона распределения график CCDF тоже прямая линия в log-log координатах (наклон $\gamma - 1$).



Диаметр графа

Пути и расстояния

Введем следующие понятия:

Путь p_{ij} - последовательность узлов, где каждый следующий связан с предыдущим.

Длина пути l - количество ребер в пути (для невзвешенного графа) или сумма весов (для взвешенного).

Кратчайший путь d_{ij} - путь с минимальной длиной между i и j:

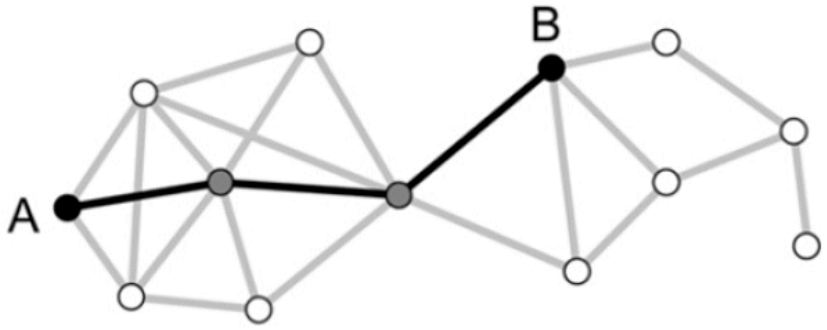
$$d_{ij} = \min\{l(p_{ij})\}$$

Средний путь $\langle l \rangle$: усредненное расстояние между всеми парами узлов.

$$\langle l \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$$

Диаметр D: максимальное расстояние в сети (длина самого длинного кратчайшего пути)

$$D = \max_{i,j} d_{ij}$$



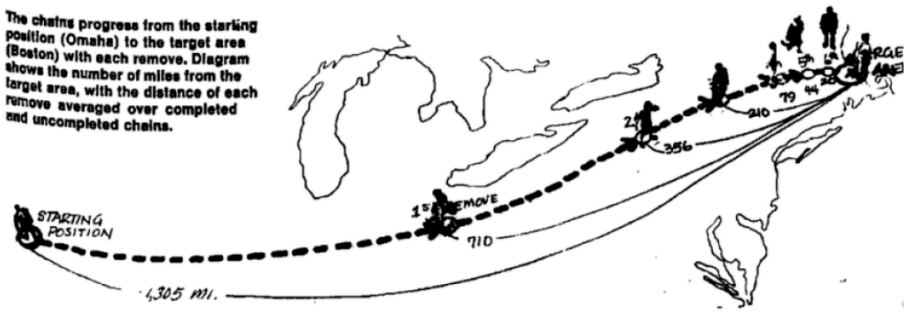
Жирным выделен кратчайший путь между A и B

Феномен “Тесного мира” (Small World)

В 1967 году Стэнли Мильграм провел эксперимент, ставший легендой социологии.

Суть эксперимента:

- Случайным людям в США (Омаха, Уичито) выдали письма.
- Цель: Доставить письмо целевой персоне в Бостоне.
- Правило: Передавать письмо можно *только* лично знакомым людям, которые (по мнению отправителя) ближе к цели.



> ! Важное уведомление
>
> Люди использовали только локальную информацию, но находили короткие глобальные пути!

Результат: Письма дошли в среднем за **~5.2 шага**. Так родилась теория **“6 рукопожатий”**.

В цифровую эпоху эксперимент повторили на миллионах пользователей в различных соц. сетях. Оказалось, что с ростом сети диаметр не растет (или растет очень медленно).

Почему важно понимать диаметр графа?

Малый диаметр означает, что любой узел “дотягивается” до всей сети за несколько шагов.

Если алгоритм собирает информацию от соседей (например, усредняет признаки или распространяет метки), то всего за **4-6 итераций** в область видимости одного узла попадает **весь граф**.

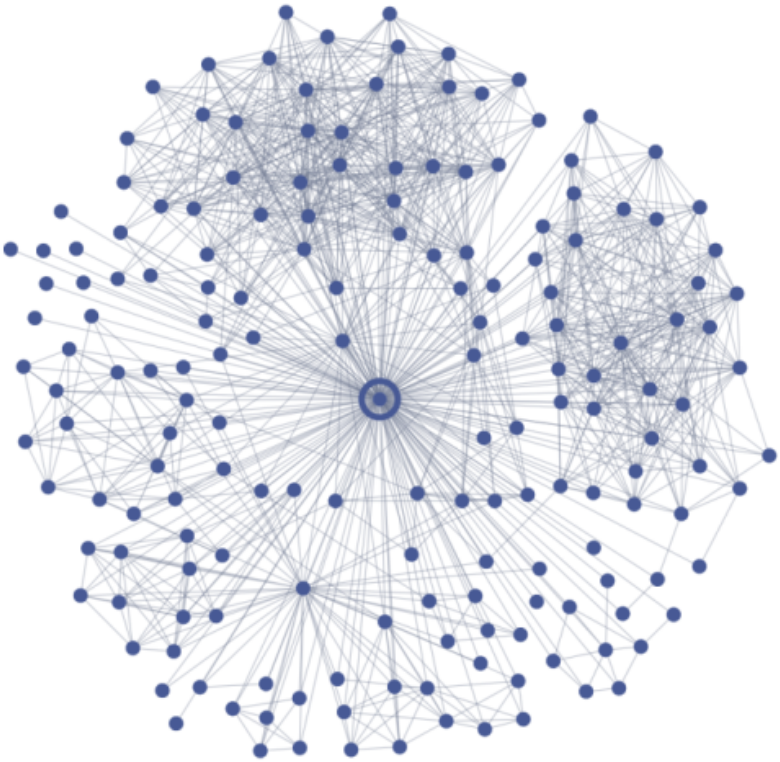
Узел перестает “видеть” свое уникальное сообщество и получает просто “среднее значение по всей сети” (проблема **Over-smoothing**).

Коэффициенты кластеризации

Эго-сети

Эго-сеть (Ego-network) — узел, его соседи и все связи между ними.

- в социальных сетях эго-сети плотные (высокая кластеризация)
- современные GNN обучаются, агрегируя информацию именно из эго-сетей



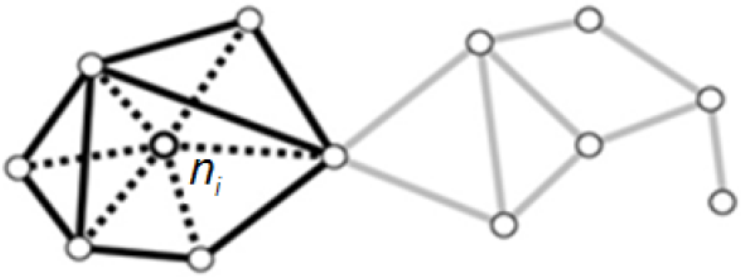
Визуализация эго-сети (Facebook)

Локальный коэффициент кластеризации

Пусть L_i — количество связей между соседями узла i, k_i — степень узла (количество соседей).

Локальный коэффициент кластеризации C_i - это доля существующих связей между соседями узла i от максимально возможного их числа. Отражает степень сгруппированности узлов (“друзья моих друзей — мои друзья”).

$$C_i = \frac{2 \cdot L_i}{k_i(k_i - 1)}$$



Средний к-т кластеризации C_{AVG}: среднее арифметическое локальных коэффициентов всех узлов.

$$C_{AVG} = \frac{1}{N} \sum_{i=1}^{N} C_i$$

Транзитивность T - отношение общего числа замкнутых треугольников к числу всех возможных "триад" (цепочек из 3 узлов: $i - j - k$).

$$T = \frac{3 \times N_{\triangle}}{N_3}$$

Сильные и слабые связи

Кластеризация тесно связана со структурой сообществ и распространением информации.

Гипотеза "Силы слабых связей" (1973):

- **сильные связи:** находятся внутри плотных кластеров (высокий C_i). Хороши для локального доверия.
- **слабые связи:** соединяют разные кластеры (мосты). Критически важны для распространения новой информации и доступа к ресурсам.

В ML-задачах это влияет на **Link Prediction**: предсказать связь внутри кластера легко (много общих соседей), между кластерами — сложно.

Промежуточные выводы

Реальные данные имеют уникальный "отпечаток", который определяется:

1. **Распределением степеней узлов:** В реальных сетях часто есть **хабы**. Среднее значение степени неинформативно.
2. **Свойствами малого мира:** Диаметр многих реальных сетей очень мал ($\sim \ln N$).
3. **Сильной локальной кластеризацией:** Например, социальные сети обладают высоким коэффициентом кластеризации.

> **Задача следующей части**
>
> Теперь нам нужны математические модели, которые умеют генерировать графы с этими тремя свойствами.

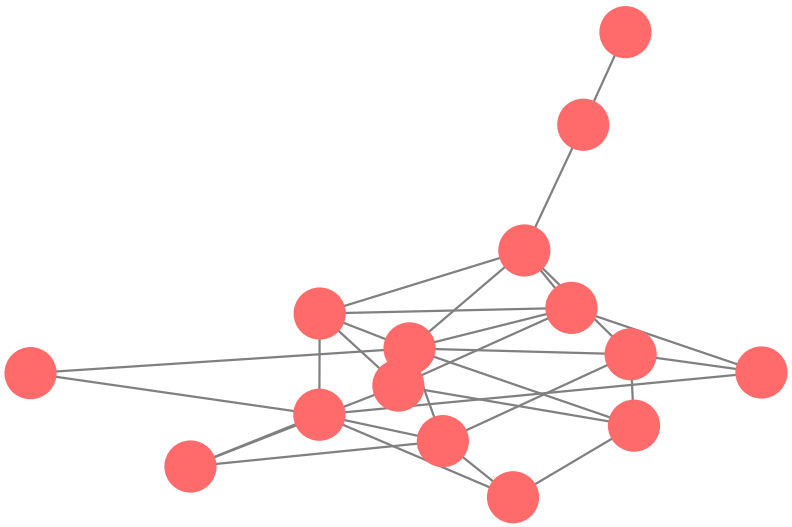
Модель Эрдеша-Реньи

Определение и свойства

Граф из N узлов, где каждая пара соединяется ребром с вероятностью p.

Свойства:

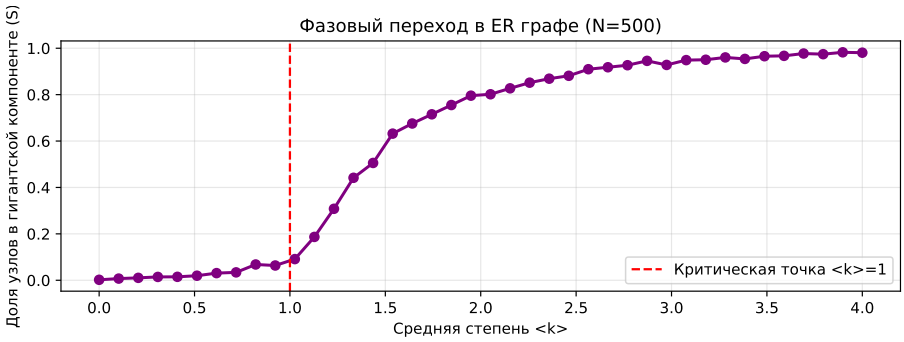
- Ожидаемое число ребер: $\langle L \rangle = p \cdot \frac{N(N-1)}{2}$
- **Средняя степень узла:** $\langle k \rangle = \frac{2\langle L \rangle}{N} \approx pN$
- **Распределение степеней узлов**:
  - Биномиальное: $P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$
  - При $N \to \infty$ и фиксированном $\langle k \rangle$ стремится к пуассоновскому $P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$
- **Диаметр** $D = logn$ при $\langle k \rangle > 1$
- **Коэффициент кластеризации** $C = p = \frac{\langle k \rangle}{n-1}$



Фазовый переход

При изменении $\langle k \rangle$ структура сети меняется скачкообразно

Режим	Параметр $\langle k \rangle$	Структура
Субкритический	$\langle k \rangle < 1$	Мелкие разрозненные острова ($O(\ln N)$).
Критический	$\langle k \rangle = 1$	Точка нестабильности. Компоненты $\sim N^{2/3}$.
Сверхкритический	$\langle k \rangle > 1$	**Гигантская компонента** ($\sim N$).



Применение и ограничения

Сильные стороны

- понятная математика, выводятся достаточно простые формулы для многих статистик
- фазовый переход (рождение гигантской компоненты) хорошо описывает реальность

Слабые стороны::

- распределение степеней биноминальное (стремится к Пуассоновскому), плохо соответствует реальности (не степенное, нет хабов)
- коэффициент кластеризации - константа и зависит от размера сети

Зачем она нужна:

- как нулевая гипотеза для многих алгоритмов
- для бенчмаркинга алгоритмов

Модель Уотса-Строгатса

Определение

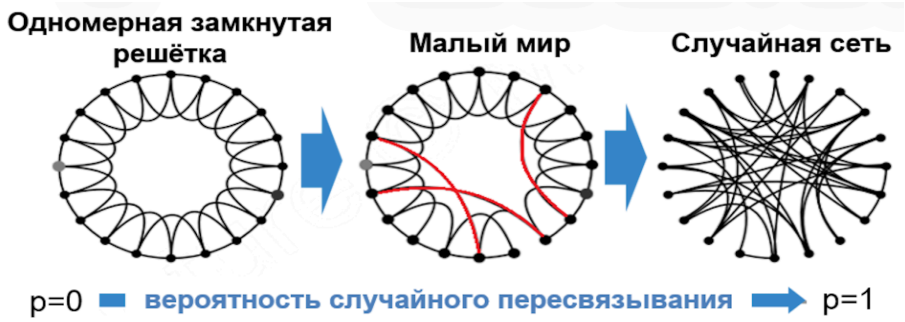
Интуиция: ER-модель объясняет “малый мир”, но теряет кластеризацию. Решетка имеет кластеризацию, но огромный диаметр. Как получить и то, и

другое? **Алгоритм пересвязывания:**

1. возьмем кольцевую решетку (каждый соединен с K соседями).
2. проходим по каждому ребру и с вероятностью p **пересвязываем** (rewire) один конец со случайным узлом.

Результат:

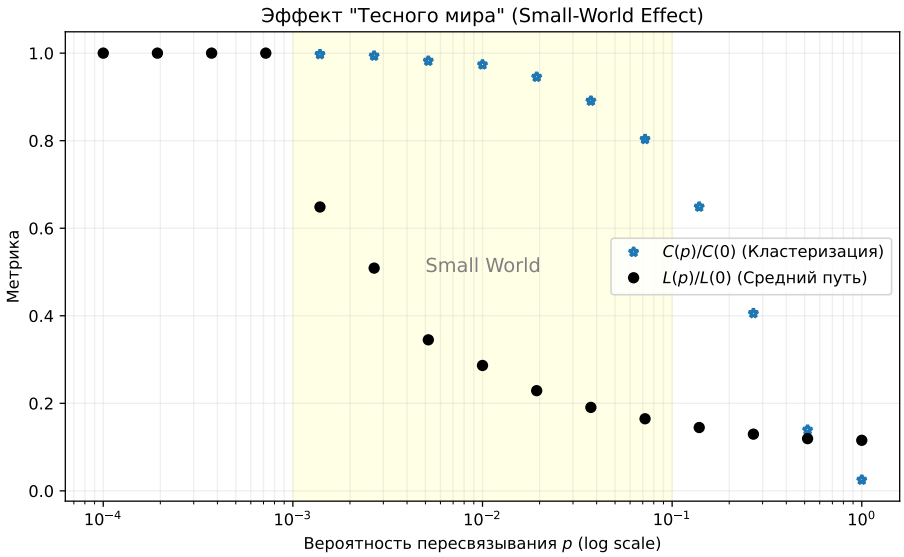
- $p = 0$: регулярная решетка (порядок).
- $p = 1$: случайный граф (хаос).
- $0 < p \ll 1$**:** - малый мир



Эволюция свойств

Что происходит при малом p:

1. Диаметр $L(p)$ быстро падает: достаточно всего пары случайных “мостов” через весь граф, чтобы сократить путь.
2. Кластеризация $C(p)$ остается высокой, локальную структура соседей почти не разрушается.



Применение и ограничения

Сильные стороны:

- в этом модели сосуществуют высокая кластеризация и малый диаметр.
- подходит для моделирования реальных систем: электросетей, социальных групп и пр..

Слабые стороны:

- распределение все еще не степенное (нет хабов)
- не объясняет рост (размер графа N фиксирован, в реальности сети растут)

Зачем она нужна:

- Используется как бенчмарк для на графовых задачах, где важна структура, а не только соседи.
- Хороший тест на проблему пересглаживания: из-за малого диаметра сигнал быстро “размывается”.

Модель растущего случайного графа

Гипотеза роста

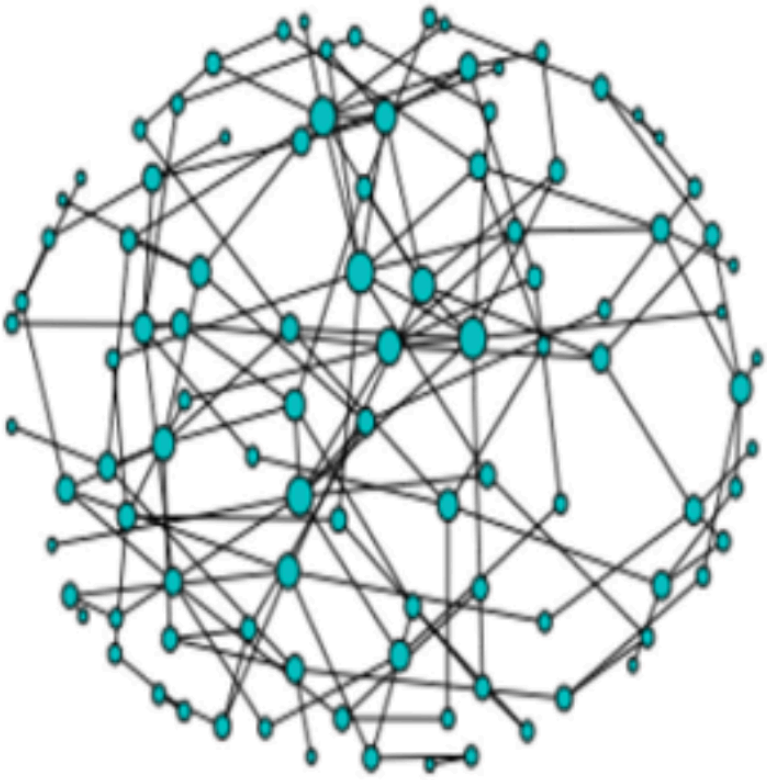
Предыдущие модели (ER, WS) были статичными: число узлов N фиксировано. Но реальные сети (Интернет, Цитаты) растут во времени.

Вопрос: может быть, хабы появляются просто потому, что старые узлы успевают набрать больше связей, чем новые?

Самая простая динамическая модель - **растущий случайный граф**.

Алгоритм

1. **Старт ($t = 0$):** Небольшое ядро из n_0 узлов.
2. **Рост:** На каждом шаге добавляем новый узел.
3. **Связывание (равновероятное присоединение):** Новый узел связывается с m уже существующими узлами.
   - Выбор соседа — **равновероятный** (случайный).
   - Вероятность присоединения к узлу i: $\Pi(i) = \frac{1}{N(t)}$



Модель случайного роста сети

Свойства модели

Результат: Старые узлы действительно имеют чуть большую степень (просто потому, что они дольше существуют).

Однако распределение степеней оказывается экспоненциальным, а не степенным: $P(k) \sim e^{-\frac{k}{m}}$

К чему это приводит:

- типичная степень узла находится в диапазоне k до $2k$; узлы со степенью, намного превышающей $2k$, исключительно редки
- распределение имеет экспоненциально убывающие хвосты (больше похоже на модель ER в этом смысле)

> ❗ Вывод
>
> Одного фактора **роста** недостаточно для формирования сложной топологии. Нужен механизм, который делает популярные узлы еще популярнее.

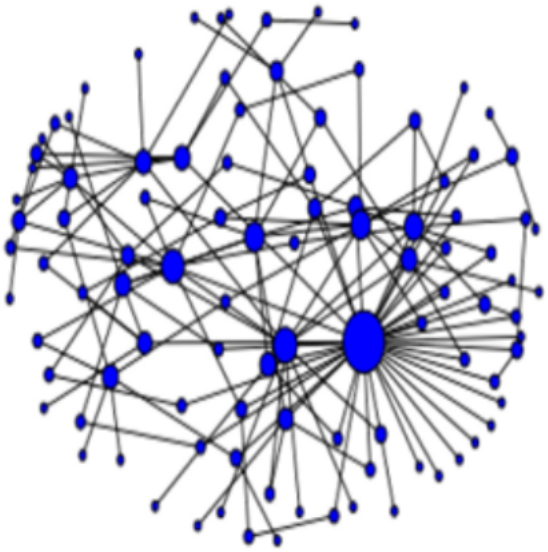
Модель Барабаши-Альберта

Алгоритм

Если заменить схему создания связей, результат меняется кардинально.

Предпочтительное присоединение (Preferential Attachment): Вероятность связи с узлом i зависит от его текущей степени k_i: $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$

Результат: Распределение степеней становится **степенным** (power law): $P(k) \sim k^{-3}$



❗ Необходимые условия

Для возникновения степенного закона требуются **оба** условия одновременно:

1. **Рост:** В сети постоянно добавляются новые узлы ($N \to \infty$).
2. **Предпочтение:** Новые узлы присоединяются к существующим пропорционально степени.

Ключевые свойства модели

1. **Стационарность:** распределение $P(k)$ не зависит от времени. Граф из 100 узлов и из 1 млн узлов имеет одинаковую структуру (наклон прямой).
2. **Отсутствие масштаба (Scale-free):** распределение выглядит "похожим на себя" при любом масштабировании логарифмических осей.
3. **Экстремальные Хабы:** степенной закон допускает узлы с гигантской степенью $k_{max} \sim \sqrt{N}$. (для сравнения: в ER-графе $k_{max} \sim \ln N$).
4. **Локализация в малых степенях:** несмотря на хабы, подавляющее большинство узлов имеют минимальную степень (равную m).
5. **Критический порог для гигантской компоненты**: в отличие от случайных графов (где есть критический порог образования гигантской компоненты), в моделях предпочтительного присоединения существует с самого начала или образуется очень плавно без резкого перехода.

Применение и ограничения

Сильные стороны:

- объясняет происхождение "тяжелых хвостов" и хабов в реальных данных.
- учитывает эволюцию (рост) сети.

Слабые стороны:

- коэффициент кластеризации $C \to 0$, многие реальные сети имеют высокую кластеризацию
- чистое предпочтительное присоединение слишком нереалистично

Где используется:

- базовая проверка гипотезы "является ли моя сеть безмасштабной?".
- моделирование атак на инфраструктуру (например, что будет, если убить главные хабы).

Конфигурационная модель

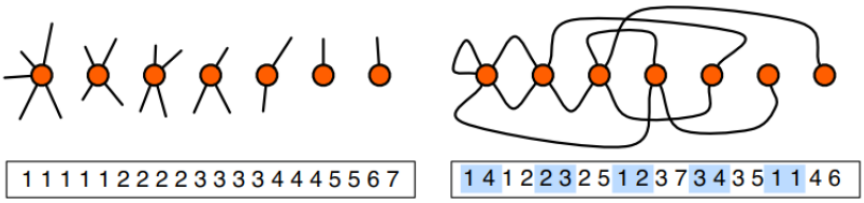
Алгоритм

Постановка задачи: мы хотим взять реальный граф с его специфическим распределением степеней и создать новый граф, который будет иметь такое же распределение, но узлы будут связаны по-другому.

Алгоритм:

1. Берем последовательность степеней $D = \{k_1, k_2, \dots, k_N\}$.
2. Рисуем у каждого узла i ровно k_i “половинок ребер” (заглушек, stubs).
3. Случайно выбираем пару заглушек и соединяем их.
4. Повторяем, пока заглушки не кончатся.

Результат: Случайный граф, который **в точности** сохраняет степени исходного.



Процесс случайного соединения заглушек

Свойства

1. могут возникать петли и кратные ребра (при больших N их число пренебрежимо мало).
2. коэффициент кластеризации $C \to 0$ (алгоритм уничтожает треугольники и сообщества).
3. диаметр масштабируется как у случайных графов $D \sim \ln N$
4. вероятность того, что узел i и узел j будут связаны случайно:

$$P_{ij} = \frac{k_i \cdot k_j}{2L}$$

(Где $2L$ — суммарное число заглушек).

Применение и ограничения

Преимущества:

- воспроизводит любое заданное распределение степеней
- позволяет отделить влияние степеней узлов от других структурных свойств
- позволяет точно рассчитать критические пороги (например, точку распада сети)

Недостатки:

- генерирует петли и кратные ребра
- уничтожает кластеры
- не объясняет процесс роста

Применение

- главное применение: поиск сообществ

Итоги лекции

Ключевые выводы

1. Обсудили, что **сети универсальны:** единый мат. аппарат описывает биологию, социум и технологии. Мы изучаем системы с нетривиальной топологией (не случайные, не регулярные).
2. Мы рассмотрели несколько основных свойств сетей (распределения степеней, диаметр, коэффициент кластеризация), которые позволяют формально описать структуру графа.
3. Рассмотрели различные модели. Каждая модель обладает своими особенностями. Нет одной модели, которая описывала бы все возможные ситуации.
4. Понимая, на какую модель похож реальный граф, мы можем **перенести некоторые свойства модели** на реальные данные и использовать это для анализа и прогнозирования.

На следующей лекции мы научимся автоматически находить сообщества в этих сетях.

Источники и литература

Источники и литература

Основная литература

- **Кочкаров А.А., Макрушин С.В., Каменчук В.Е, Блохин Н.В.**. Экспериментальная теория графов и алгоритмы анализа сетевых моделей
- **Макрушин С.В.** Курс лекций “Теория сложных сетей в экономике”.

Дополнительные ресурсы

- **Документация NetworkX:** networkx.org — официальный справочник и туториалы.
- **Stanford CS224W:** *Machine Learning with Graphs* (Jure Leskovec) — web.stanford.edu/class/cs224w/

- **Social Network Analysis:** Авторский курс Леонида Жукова (Leonid Zhukov).
- The average distances in random graphs with given expected degrees
- Erdős–Rényi Model for Network Formation
- Random (Erdős-Rényi) networks, degree distribution, clustering, Watts-Strogratz network
- Finding the diameter in real-world graphs experimentally turning a lower bound into an upper bound
- Network Science – The Scale□Free Property (Barabási)
- Watts–Strogatz model (Wikipedia)
- Epidemic spreading on scale□free networks
- Network Formation: Dynamic Models and Preferential Attachment