Сетевое программирование

Лекция 5

Метрики и характеристики сети. Степень узла. Распределение степеней

Усіпов Н.М., PhD · КазНУ имени аль-Фараби

Цель и план лекции

Цель

Изучение основных количественных характеристик сети, распределения степеней узлов и физического смысла этих показателей для анализа телекоммуникационных сетей.

1. Введение
2. Степень узла и распределение степеней
3. Основные глобальные метрики
4. Дополнительные характеристики структуры
5. Сравнение с базовой моделью
6. Особенности вычислений
7. Коррелированность, иерархия и эффективность
8. Парадокс дружбы и выборка узлов

1 Введение

- Визуальный осмотр эффективен лишь для графов из нескольких десятков узлов. Для описания сетей большого размера применяются количественные метрики.
- Метрики делятся на локальные, описывающие отдельный узел или ребро, и глобальные, описывающие сеть в целом.

2 Степень узла и распределение степеней

- Степень k_i узла i равна числу его соединений, а средняя степень сети вычисляется по формуле:

$$\langle k \rangle = 2|E| / |V|.$$

- Средняя степень даёт общее представление о плотности соединений, однако не показывает, как связи распределены между узлами.

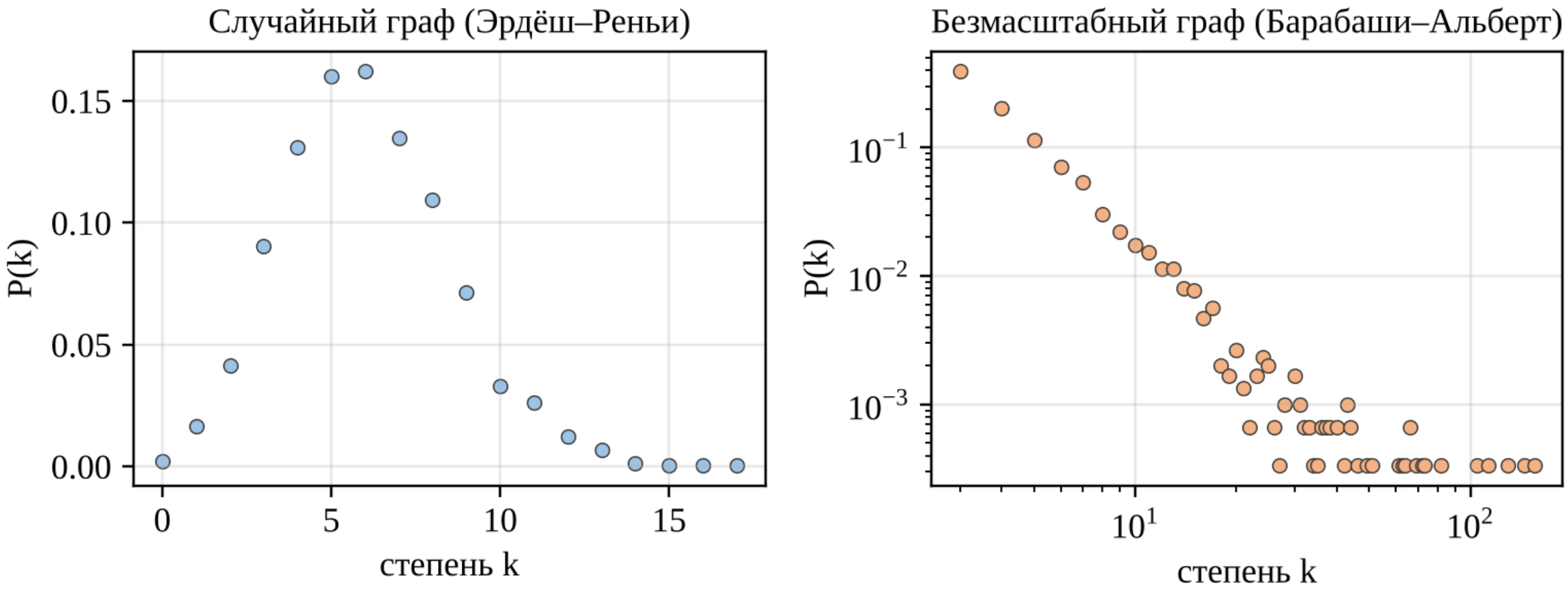
$$P(k) = n_k / |V|,$$

- где n_k – количество узлов со степенью k.

2 Степень узла и распределение степеней (продолжение)

- В случайном графе Эрдёша–Реньи степени группируются около среднего значения, и узлы с очень большой степенью практически отсутствуют.

2 Степень узла и распределение степеней (продолжение)



Распределения степеней случайной и безмасштабной сетей

2 Степень узла и распределение степеней (продолжение)

- Сеть называют безмасштабной, если её распределение степеней следует степенному закону:

$$P(k) \sim k^{-\gamma}.$$

- В логарифмических координатах такая зависимость имеет вид прямой линии. Следует проявлять осторожность: реальные данные редко описываются степенным законом на всём диапазоне, а прямая в двойном логарифмическом масштабе ещё не является доказательством этого закона.

3 Основные глобальные метрики

Плотность

Отношение числа существующих рёбер к максимально возможному. Для неориентированного графа она равна $2|E| / (|V|(|V|-1))$.

Расстояние и эксцентриситет

Расстояние d(u, v) – длина кратчайшего пути между узлами. Эксцентриситет узла равен наибольшему расстоянию от него до других узлов.

Диаметр

Наибольшее расстояние между любой парой узлов сети. Диаметр определяет наихудшее число переходов при передаче и, следовательно, влияет на предельную задержку.

Радиус и центр

Радиус – минимальный эксцентриситет. Центром сети называют множество узлов, на которых этот минимум достигается; такие узлы удобны для размещения общих серверов.

Средняя длина пути

Среднее расстояние по всем парам узлов. Эта величина характеризует типичное число переходов.

Коэффициент кластеризации

Для узла i со степенью k_i он показывает, какая доля пар его соседей соединена между собой:

$$C_i = 2T_i / (k_i(k_i-1)),$$

3 Основные глобальные метрики (продолжение)

Ассортативность

Коэффициент корреляции степеней узлов, соединённых ребром. При положительном значении узлы с большой степенью склонны соединяться между собой, при отрицательном – с узлами малой степени.

4 Дополнительные характеристики структуры

Транзитивность

Отношение утроенного числа треугольников к числу связных троек вершин. В отличие от среднего коэффициента кластеризации, этот показатель определяется по сети в целом и в большей степени учитывает узлы с большой степенью.

Средняя степень соседей

Показывает, с какими узлами обычно соединён узел заданной степени. Поэтому она дополняет коэффициент ассортативности.

k-ядро

Максимальный подграф, в котором степень каждого узла не меньше k. Ядра с большими значениями k выделяют плотную центральную часть сети.

Размер компонент

Для несвязной сети важны количество компонент связности и доля узлов в наибольшей из них.

4 Дополнительные характеристики структуры (продолжение)

Стоимостные метрики

Суммарная длина линий или суммарная стоимость каналов вычисляется по значениям атрибутов рёбер и необходима при сравнении вариантов проекта.

5 Сравнение с базовой моделью

- Отдельное значение метрики само по себе мало информативно. Чтобы определить, является ли кластеризация некоторой сети высокой, её сравнивают с кластеризацией случайных сетей с такими же числом узлов и степенями.

5 Усреднение метрик по нескольким реализациям

Листинг 5.1.

```
import networkx as nx
import statistics as st

def stats(make, runs=20):
    cl, pl = [], []
    for seed in range(runs):
        G = make(seed)
        H = G.subgraph(max(nx.connected_components(G), key=len))
        cl.append(nx.average_clustering(H))
        pl.append(nx.average_shortest_path_length(H))
    return st.mean(cl), st.stdev(cl), st.mean(pl)

n, k = 200, 6
er = lambda s: nx.gnp_random_graph(n, k / (n - 1), seed=s)
ws = lambda s: nx.watts_strogatz_graph(n, k, 0.1, seed=s)
for name, f in (("ER", er), ("WS", ws)):
    c, sd, l = stats(f)
    print(f"{name}: C = {c:.3f} +- {sd:.3f}, L = {l:.2f}")

G = nx.karate_club_graph()
print("Наибольшее k:", max(nx.core_number(G).values()))
print("Узлов в этом ядре:", len(nx.k_core(G)))
print("Транзитивность:", round(nx.transitivity(G), 3))
```

5 Результат: Усреднение метрик по нескольким реализациям

```
ER: C = 0.030 +- 0.006, L = 3.12
WS: C = 0.445 +- 0.013, L = 4.42
Наибольшее k: 4
Узлов в этом ядре: 10
Транзитивность: 0.256
```

Приведённые значения показывают, что при одинаковых размере и средней степени кластеризация модели Уоттса–Строгаца более чем на порядок выше, а средняя длина пути увеличивается лишь умеренно.

6 Особенности вычислений

- Диаметр и средняя длина пути определены только для связного графа: в противном случае NetworkX сообщит об ошибке.

6 Вычисление основных метрик сети

Листинг 5.2.

```
import networkx as nx
from collections import Counter

G = nx.barabasi_albert_graph(500, 2, seed=7)

degrees = [d for _, d in G.degree()]
print("Средняя степень:", round(sum(degrees) / len(G), 2))
print("Максимальная степень:", max(degrees))
print("Плотность:", round(nx.density(G), 4))
print("Связность:", nx.is_connected(G))
print("Диаметр:", nx.diameter(G))
print("Ср. длина пути:", round(nx.average_shortest_path_length(G), 3))
print("Кластеризация:", round(nx.average_clustering(G), 3))
r = nx.degree_assortativity_coefficient(G)
print("Ассортативность:", round(r, 3))

hist = Counter(degrees)
print("P(k) при k=2,3,4:",
      [round(hist[k] / len(G), 3) for k in (2, 3, 4)])
```

6 Результат: Вычисление основных метрик сети

```
Средняя степень: 3.98
Максимальная степень: 47
Плотность: 0.008
Связность: True
Диаметр: 7
Ср. длина пути: 3.815
Кластеризация: 0.024
Ассортативность: -0.148
P(k) при k=2,3,4: [0.502, 0.21, 0.106]
```

Для инженерной оценки сети важны не отдельные значения, а их сочетание. Например, небольшой диаметр при наличии нескольких узлов очень большой степени свидетельствует о короткой передаче, но одновременно указывает на зависимость от этих узлов.

7 Коррелированность, иерархия и эффективность

- При описании сетей часто используют дополнительные характеристики, которые отражают особенности устройства сети.
- Эффективность является удобной альтернативой средней длине пути. Для всей сети её определяют как среднее значение обратных расстояний:

$$E = (1 / (n(n-1))) \sum_{u \neq v} 1 / d(u, v).$$

- Несвязные пары в этой формуле дают нулевой вклад, поэтому показатель определён и для несвязных сетей.
- Важным понятием является зависимость кластеризации от степени C(k). Если значение C(k) убывает с ростом k, это может указывать на иерархическую организацию: малые узлы входят в плотные группы, а концентраторы связывают разные группы между собой.

7 Коррелированность, иерархия и эффективность (продолжение)

- Средняя степень соседей $k_{nn}(k)$ помогает понять характер соединений. В магистральных сетях часто наблюдается дисассортативность, когда крупные узлы соединяются с множеством малых.

8 Парадокс дружбы и выборка узлов

- Выбранный случайным образом сосед имеет большую ожидаемую степень, чем случайно выбранный узел.

$$\langle k_n \rangle = \langle k^2 \rangle / \langle k \rangle,$$

- что не меньше, чем ⟨k⟩, причём разность определяется дисперсией степеней. Это свойство называют парадоксом дружбы.
- Для степенного распределения с показателем $2 < \gamma < 3$ величина $\langle k^2 \rangle$ при увеличении размера сети неограниченно возрастает.

9 Оценка степенного закона

- Часто показатель степенного распределения оценивают линейной регрессией по логарифмам значений гистограммы, однако такой метод даёт смещённые результаты.

$$\gamma = 1 + n \left[\sum \ln(k_i / k_{min}) \right]^{-1}.$$

- Для целых значений используется приближение с заменой k_{min} на $k_{min} - 0{,}5$. Величина k_{min} подбирается на основе критерия Колмогорова–Смирнова.
- На практике следует помнить, что большое число параметров не всегда улучшает описание.

10 Разбор примера

1. Определим показатели графа с рёбрами (1,2), (2,3), (3,4), (4,1), (2,4), (4,5) из лекции 2.
2. Коэффициенты кластеризации узлов таковы.
3. Соседи узла 1 (2 и 4) связаны, поэтому $C_1 = 1$.
4. Узел 2 имеет соседей 1, 3 и 4; из трёх пар связаны две (1–4 и 3–4), следовательно, $C_2 = 2/3$.
5. Аналогично $C_3 = 1$.
6. Узел 4 имеет четырёх соседей, из шести пар связаны две (1–2 и 2–3), поэтому $C_4 = 1/3$.
7. Для узла 5 со степенью 1 принимается $C_5 = 0$.

10 Разбор примера (продолжение)

8 Среднее значение равно (1 + 2/3 + 1 + 1/3 + 0)/5 = 0,6.

9 Расстояния между несмежными парами (1,3), (1,5), (2,5) и (3,5) равны 2, поэтому диаметр равен 2.

10 Средняя длина пути равна (6·1 + 4·2)/10 = 1,4.

11 Вопросы для обсуждения на занятии

Интерпретация

Малый диаметр при очень высокой максимальной степени – это достоинство или недостаток сети связи?

Выбор показателя

Какие три показателя вы включили бы в ежемесячный отчёт о состоянии сети?

Выборка

Почему обследование сети по соседям может завышать среднюю степень?

Контрольные вопросы

1. Чем локальные метрики отличаются от глобальных?
2. Что такое распределение степеней и почему средней степени недостаточно?
3. Какая сеть называется безмасштабной? Как проверяется степенной закон?
4. Что показывает диаметр сети и как он связан с задержкой передачи?
5. Дайте определение коэффициента кластеризации и объясните его смысл для сети связи.
6. Что такое центр сети и в каких инженерных задачах он может быть использован?
7. Какие условия необходимо проверить перед вычислением диаметра в NetworkX?
8. Как можно оценить метрики сети очень большого размера?

Список литературы

1. Newman M. Networks. – 2nd ed. – Oxford: Oxford University Press, 2018. – 780 p.
2. Barabási A.-L. Network Science. – Cambridge: Cambridge University Press, 2016.
3. Clauset A., Shalizi C.R., Newman M.E.J. Power-law distributions in empirical data // SIAM Review. – 2009. – Vol. 51, No. 4. – P. 661–703.
4. Newman M.E.J. Assortative mixing in networks // Physical Review Letters. – 2002. – Vol. 89. – Art. 208701.
5. Watts D.J., Strogatz S.H. Collective dynamics of ‘small-world’ networks // Nature. – 1998. – Vol. 393. – P. 440–442.
6. Zachary W.W. An information flow model for conflict and fission in small groups // Journal of Anthropological Research. – 1977. – Vol. 33, No. 4. – P. 452–473.
7. NetworkX Documentation [Электронный ресурс]. – URL: https://networkx.org/documentation/stable/