

Анализ социальных сетей, лекция 1. Основные понятия и характеристики

Михаил Пожидаев

31 октября 2024 г.

Сети

Что такое сеть?

Математически сеть определяется парой (N, E) , состоящей из множества вершин N и множества ребер E (или в случае направленных связей дуг), что соответствует понятию графа. Может представлять структурные связи, отношения или потоки информации. Наиболее распространённые примеры:

- ▶ семантические сети;
- ▶ социальные сети;
- ▶ коммуникационные сети;
- ▶ метаболические сети.

Цели

Зачем проводят исследования сетей?

1. Структурный анализ:
 - ▶ для определения «влиятельности» узлов;
 - ▶ для оценки устойчивости или поиска уязвимостей в сетях.
2. Подсчёт характеристик:
 - ▶ для прогнозирования эволюции сети;
 - ▶ для сравнения сетей между собой.

Понятия

Ключевые термины в сетевом анализе

1. Степень узла (порядок, валентность) — количество рёбер, связанных с узлом.
2. Диаметр сети — максимальной кратчайший путь по всем парам узлов в сети.
3. Компонента связности — множество узлов, между которыми существует путь.
4. Клика — множество попарно смежных узлов.

Транзитивность

Определение и назначение понятия транзитивности

Свойство транзитивности подразумевает существование связи (a, c) при существовании связей (a, b) и (b, c) .

- ▶ Триада — тройка вершин a , b и c с существующими связями (a, b) и (b, c) .
- ▶ Закрытая триада — триада с выполнением свойства транзитивности (учитывается порядок).

Коэффициент кластеризации

Коэффициент кластеризации — отношение количества замкнутых триад к количеству всех возможных триад.

$$c = \frac{t * 3}{q},$$

где t — число треугольников (замкнутых триад без учёта порядка) и q — число триад.

Для трёх попарно смежных узлов:

- ▶ шесть замкнутых триад;
- ▶ шесть незамкнутых триад;
- ▶ коэффициент кластеризации равен 1.

В жизни

Коэффициенты кластеризации для реальных сетей

1. Сеть совместного участия актёров в фильмах: 0,2.
2. Сеть соавторства учёных биологов: 0,09.
3. Сеть обмена электронной почтой: 0,16.
4. Физическая сеть Интернета: 0,01.

Для отдельной вершины

Локальный коэффициент кластеризации

$$c_i = \frac{p_i}{q_i},$$

где p_i — количество между собой смежных узлов и смежных с i , q_i — количество смежных с i узлов.

В некоторых случаях вычисляется средний коэффициент локальной кластеризации как среднее арифметическое c_i по всем i .

Влиятельность

Значение $c_i < c$ является индикатором «так называемой структурной дыры».

Меры центральности

Оценка «влиятельности» узлов

Локальный коэффициент кластеризации и «структурные дыры» являются простейшими мерами центральности. Они характеризуют значимость узла в передаче информации между другими узлами сети.

Промежуточность узла — количество всех кратчайших путей в сети между всеми парами связанных узлов, проходящих через некоторый узел i . Чаще всего заметно коррелирует с локальным коэффициентом кластеризации: из большой промежуточности часто следует низкая кластеризация.

Хитрый пиар!

В социальных сетях люди с высоким c_i не очень ценны для «дружбы».

Маленький мир

Правило шести «рукопожатий»

Два любых человека на планете знакомы через цепочку «рукопожатий» длиной не более шести. Теория была выдвинута в 1969 году американскими социальными психологами Стэнли Милгрэмом и Джефффри Трэверсом. При проведении социального эксперимента с рассылкой писем гражданами Америки выяснилось, что медианная длина цепочки равняется 5,2.

$$d = \frac{\ln n}{\ln k},$$

где d — диаметр сети, k — средняя степень узлов, а n — количество узлов.

Модели

Назначение и примеры

Проблема!

В некоторых сетях невозможно произвести анализ в силу недоступности информации о всех её частях.

В такой ситуации анализ фактической существующей сети заменяют анализом её математической модели, которую предстоит найти и обосновать её применимость.

Закон Ципфа

Вероятность узла иметь степень k равна $k^{-\alpha}$, где α — положительная константа.

Спасибо за внимание!

Всё о курсе: <https://marigostra.ru/materials/networks.html>

E-mail: msp@luwrain.org

Канал в Телеграм: <https://t.me/MarigostraRu>