Структурные характеристики сетей

Анализ социальных сетей
Лекция 3

Михаил Пожидаев

Институт прикладной математики и компьютерных наук ТГУ

10 ноября 2026 г.

Компоненты связности

Связные части неориентированной сети

Компонента связности — максимальное множество вершин, в котором из каждой вершины можно попасть в любую другую по некоторому пути. В неориентированной сети любые две вершины из одной компоненты связности достижимы друг из друга. Сеть называется **связной**, если она состоит ровно из одной компоненты связности.

Примеры

- Сеть с 12 вершинами может состоять из одной компоненты размера 9, одной компоненты размера 2 и одной изолированной вершины.
- Изолированная вершина образует компоненту связности размера 1.

Гигантская компонента

Наибольшая связная часть сети

Гигантской компонентой называют компоненту связности, размер которой сопоставим с размером всей сети. Обычно гигантская компонента содержит существенно больше вершин, чем каждая из остальных компонент.

Числовая величина

Пусть N — число всех вершин, а N_G — число вершин в наибольшей компоненте. Тогда долю сети, охваченную гигантской компонентой, можно определить как $g = \frac{N_G}{N}$.

Интерпретация

Если g близко к единице, большинство участников сети объединено путями взаимодействия. Если g мало, сеть фрагментирована на множество относительно независимых частей. В социальной сети гигантская компонента может соответствовать основной массе пользователей, между которыми возможна передача информации через цепочки знакомств.

Связность в ориентированной сети

Направление рёбер имеет значение

В ориентированной сети путь из вершины i в вершину j не обязательно означает существование пути из j в i.

Сильная компонента связности — максимальное множество вершин, в котором каждая вершина достижима из каждой другой с учётом направлений рёбер. **Слабая компонента связности** — компонента, получаемая после игнорирования направлений всех рёбер сети.

Пример

Если существуют дуги $A \to B$ и $B \to C$, то A достижима из C только после игнорирования направлений.

Каждая сильная компонента целиком содержится в некоторой слабой компоненте.

Сильные компоненты ориентированной сети можно заменить отдельными вершинами. Между двумя такими вершинами проводится дуга, если в исходной сети есть хотя бы одно ребро между соответствующими сильными компонентами. Полученный граф называется **конденсацией** ориентированной сети.

Интерпретация

Конденсация показывает направления достижимости между устойчивыми группами взаимосвязанных вершин.

Пример

В сети подписок сильная компонента может описывать группу аккаунтов, каждый из которых достижим из другого по цепочке подписок.

CC BY-NC 4.0 | Институт прикладной математики и компьютерных наук ТГУ

Расстоянием d_{ij} между вершинами i и j называется длина кратчайшего пути между ними. В невзвешенной сети длина пути равна количеству рёбер в нём. В ориентированной сети расстояние d_{ij} может отличаться от d_{ji}.
Средняя длина кратчайшего пути определяется как:

$$l = \frac{1}{n(n-1)} \sum_{\substack{i,j \in N \\ i \neq j}} d_{ij}$$

Связность

Формула применима к связной неориентированной сети или к сильно связной ориентированной сети. Для несвязных сетей обычно рассчитывают среднюю длину пути внутри компонент либо только для пар достижимых вершин.

CC BY-NC 4.0 | Институт прикладной математики и компьютерных наук ТГУ

Диаметр сети

Наибольшее кратчайшее расстояние

Диаметр сети — наибольшее расстояние между парой вершин:

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

Диаметр характеризует максимальное число шагов, необходимое для передачи информации между наиболее удалёнными вершинами сети.

Пример

Если самая удалённая пара вершин соединена кратчайшим путём из семи рёбер, диаметр сети равен 7.

Связность

У несвязной сети диаметр не определён или считается бесконечным, поскольку между вершинами разных компонент пути не существует. Поэтому на практике часто вычисляют диаметр гигантской компоненты.

Маленький мир

Правило шести «рукопожатий»

Теория маленького мира

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 Джеффри Трэверсом. При проведении социального эксперимента с рассылкой писем гражданами Америки выяснилось, что медианная длина цепочки равняется 5,2.

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

Обозначения

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

Эксцентриситет и радиус

Положение вершины относительно сети

Эксцентриситет вершины — максимальное расстояние от неё до остальных вершин сети:

$$e_i = \max_{j \in N} d_{ij}$$

Радиус сети — наименьший эксцентриситет среди всех вершин:

$$R = \min_{i \in N} e_i$$

Диаметр сети равен наибольшему эксцентриситету:

$$D = \max_{i \in N} e_i$$

Центр сети

Вершины с эксцентриситетом, равным радиусу, образуют центр сети.

Мера близости

Анализ расстояний между вершинами

$$l_i = \frac{\sum\limits_{j \in N} d_{ij}}{n_i}$$

Обозначения

- d_{ij} — расстояние между вершинами i и j;
- n_i — количество вершин, достижимых из вершины i.

Степень близости вершины

$$c_i = \frac{1}{l_i}$$

Недостаток!

Применение степени близости затруднено, если в сети более одной компоненты связности.

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

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

Здесь t — число треугольников, то есть замкнутых триад без учёта порядка, а q — число триад.

Для трёх попарно смежных узлов:

- шесть замкнутых триад;
- шесть незамкнутых триад;
- коэффициент кластеризации равен единице.

CC BY-NC 4.0 |

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.

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

Обозначения

- p_i — количество пар узлов, смежных между собой и смежных с i;
- q_i — количество пар узлов, смежных с i.

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

Влиятельность

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

CC BY-NC 4.0

Ассортативное смешивание

Связи узлов по назначенным свойствам

Назначим

Каждому узлу сопоставим некоторое свойство, значение которого может быть качественным, то есть категориальным или дискретным, либо количественным, то есть числовым или непрерывным. Определение Сеть называется ассортативной, если вероятность появления связи выше для узлов, у которых значение некоторого свойства совпадает. Социальные сети чаще всего ассортативны по возрасту или региону проживания пользователя.

Корреляция

Для количественного случая ассортативность сети можно рассматривать как корреляцию значений свойства на разных концах рёбер.

Пусть задано m групп $g_1, g_2, \ldots, g_m$, и все узлы из N принадлежат к одной из них. Примем $\delta_{ij} = 1$, если $g_i = g_j$, и 0 в противном случае. Оценка ассортативности выражается как:

$$q_1 = \frac{1}{2} \sum_{i,j \in N} A_{ij} \delta_{ij}$$

Полезно сравнивать ассортативность со средней ассортативностью:

$$q_2 = \frac{1}{2} \sum_{i,j \in N} p_{ij} \delta_{ij}$$

Здесь p_{ij} — вероятность существования ребра между i и j. Удобно нормировать значения на количество рёбер m.

Модулярность сети

$$q = \frac{q_1 - q_2}{m}$$

Спасибо за внимание!

Всё о курсе:
https://marigostra.ru/materials/sna.html

Электронная почта:
msp@luwrain.org

Telegram:
@Marigostra



Материал можно распространять с соблюдением требований Creative Commons Attribution-NonCommercial 4.0 International License. При обнаружении любых ошибок и неточностей в материале просьба сообщить о них автору.