Выделение сообществ

Анализ социальных сетей
Лекция 4

Михаил Пожидаев

Институт прикладной математики и компьютерных наук ТГУ

17 ноября 2026 г.

Сообщества в сети

Плотные группы взаимосвязанных участников

Сообществом называют группу вершин, связи внутри которой в среднем плотнее или значимее, чем связи этой группы с остальной сетью. В социальной сети сообщество может соответствовать группе друзей, коллег, однокурсников, участников общего проекта или пользователей с близкими интересами. Выделение сообществ позволяет перейти от анализа отдельных вершин и рёбер к анализу структуры сети на уровне групп.

Свободная формулировка

Универсального строгого определения сообщества не существует. Выбор определения зависит от типа сети, исследовательского вопроса и используемого алгоритма.

Зачем выделять сообщества

Исследовательские и прикладные задачи

Выделение сообществ помогает обнаруживать скрытую социальную организацию сети.

1. Анализировать группы пользователей с интенсивным взаимодействием.
2. Выявлять аудитории с разными интересами или предпочтениями.
3. Исследовать распространение информации, мнений и инноваций.
4. Находить границы между организациями, регионами или профессиональными группами.
5. Сегментировать сеть для рекомендательных систем и таргетированных коммуникаций.

Интерпретация

Сообщество не обязательно является формальной группой. Оно может отражать только устойчивый паттерн наблюдаемых связей.

Пусть множество вершин C рассматривается как возможное сообщество. Обозначим через m_{in} число рёбер между вершинами из C, а через m_{out} — число рёбер между C и остальной сетью. Группа является хорошим кандидатом на сообщество, если m_{in} относительно велико, а m_{out} относительно мало.

Пример

Группа студентов одного учебного потока может активно общаться внутри себя. При этом число связей с другими потоками будет существенно меньше.

Ограничение

Плотность внутренних связей сама по себе не всегда достаточна. В разреженной сети сообщество может быть менее плотным, чем в среднем, но всё равно иметь отчётливую структуру.

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

В задаче разбиения каждая вершина относится ровно к одному сообществу. Пусть сеть состоит из n вершин, а c_i — номер сообщества вершины i. Тогда разбиение можно представить как набор непересекающихся групп:

$$C_1, C_2, \ldots, C_k$$

Каждая вершина входит в одно из сообществ $C_1, \ldots, C_k$.

Интерпретация

Такое представление удобно для сегментации пользователей. Однако оно не учитывает ситуации, в которых человек одновременно принадлежит нескольким социальным кругам.

Перекрывающиеся сообщества

Одна вершина в нескольких группах

Перекрывающимися сообществами называют сообщества, которые могут иметь общие вершины. В социальной сети один пользователь может одновременно входить в сообщество семьи, коллег, соседей и участников спортивной команды. В этом случае для вершины задаётся не единственный номер сообщества, а множество групп, к которым она принадлежит.

Пример

Преподаватель может быть связан с коллегами по кафедре, со студентами учебной группы и с участниками научного проекта. Эти сообщества способны пересекаться через одну и ту же вершину.

Особенность

Алгоритмы с перекрывающимися сообществами обычно сложнее алгоритмов обычного разбиения сети.

Модулярность показывает, насколько наблюдаемое число внутригрупповых связей превышает ожидаемое число таких связей в случайной сети. Для неориентированной невзвешенной сети модулярность определяется как:

$$Q = \frac{1}{2m} \sum_{i,j \in N} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

Обозначения

- A_{ij} — элемент матрицы смежности;
- k_i и k_j — степени вершин i и j;
- m — число рёбер сети;
- $\delta(c_i, c_j) = 1$, если вершины входят в одно сообщество.

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

Смысл модулярности

Сравнение с нулевой моделью

Слагаемое A_{ij} показывает, существует ли наблюдаемая связь между вершинами i и j. Слагаемое $\frac{k_i k_j}{2m}$ показывает ожидаемую вероятность связи в случайной сети с такими же степенями вершин. Модулярность возрастает, если внутри сообществ оказывается больше рёбер, чем можно было бы ожидать случайно.

Интерпретация

Большое значение Q свидетельствует о выраженной модульной структуре. Значение, близкое к нулю, означает, что разбиение мало отличается от случайного распределения связей.

Осторожно

Модулярность следует сравнивать между сетями и разбиениями с осторожностью. На её величину влияют размер сети, плотность и выбранная нулевая модель.

Метод Гирвана—Ньюмана

Удаление связей между сообществами

Метод Гирвана—Ньюмана основан на предположении, что рёбра между сообществами часто лежат на большом числе кратчайших путей. Для каждого ребра рассчитывается **центральность посредничества ребра**. Затем ребро с наибольшей центральностью удаляется из сети. После каждого удаления центральности пересчитываются заново. Постепенно сеть распадается на отдельные группы.

Интерпретация

Ребро между двумя дружескими компаниями часто служит мостом. Через него проходят многие кратчайшие пути между участниками разных компаний.

Центральность ребра

Основа метода Гирвана—Ньюмана

Центральность посредничества ребра e равна числу кратчайших путей, проходящих через это ребро. Рёбра с высоким значением b_e часто соединяют разные части сети.

$$b_e = \sum_{s \neq t} \frac{\sigma_{st}(e)}{\sigma_{st}}$$

Обозначения

- σ_{st} — число кратчайших путей между вершинами s и t;
- $\sigma_{st}(e)$ — число таких путей, проходящих через ребро e.

Недостаток

Метод требует многократного пересчёта кратчайших путей. Поэтому он плохо масштабируется на очень большие сети.

Иерархическая структура

Дендрограмма сообществ

Последовательное удаление рёбер в методе Гирвана—Ньюмана создаёт иерархию разбиений. Эту иерархию можно представить в виде дендрограммы. На верхнем уровне все вершины принадлежат одной сети. На нижних уровнях сеть разделяется на всё более мелкие группы.

Выбор разбиения

В качестве итогового можно выбрать уровень дендрограммы, на котором модулярность достигает наибольшего значения.

Интерпретация

Иерархический подход полезен, если социальные группы имеют вложенную структуру. Например, учебная группа может входить в поток, факультет и университет.

Метод Лувена

Быстрая оптимизация модулярности

Метод Лувена ищет разбиение сети с высокой модулярностью.

1. На первом шаге каждая вершина образует собственное сообщество.
2. Затем вершины поочерёдно переносятся в соседние сообщества, если такой перенос увеличивает модулярность.
3. После стабилизации разбиения каждое найденное сообщество сжимается в одну вершину.
4. Процесс повторяется для укрупнённой сети.

Преимущество

Метод Лувена эффективен для крупных сетей. Поэтому он часто применяется для анализа социальных платформ и графов взаимодействий.

При переносе вершины i в сообщество C оценивается прирост модулярности ΔQ. Вершина перемещается только в том случае, если существует сообщество с положительным значением ΔQ. Для невзвешенной сети прирост можно вычислять по локальным характеристикам вершины и сообщества.

$$\Delta Q = \left[\frac{\Sigma_{in} + 2k_{i,in}}{2m} - \left(\frac{\Sigma_{tot} + k_i}{2m}\right)^2\right] - \left[\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right]$$

Обозначения

Здесь Σ_{in} — суммарный вес внутренних рёбер сообщества, а Σ_{tot} — суммарная степень его вершин.

Метод распространения меток

Сообщества через локальное согласование

В **методе распространения меток** каждой вершине сначала присваивается уникальная метка. Затем вершина принимает метку, наиболее распространённую среди её соседей. Процесс продолжается до тех пор, пока метки большинства вершин не перестанут изменяться. Вершины с одинаковой меткой образуют сообщество.

Преимущество

Метод очень быстро работает на крупных сетях. Он не требует заранее задавать число сообществ.

Недостаток

Результат может зависеть от порядка обновления вершин. При повторном запуске на одной сети могут получаться разные разбиения.

Кликой называется множество вершин, каждая пара которых соединена ребром. В методе клик сообщество строится из цепочки соседних k-клик. Две k-клики считаются соседними, если они имеют ровно $k-1$ общих вершин. Объединение достижимых друг из друга соседних клик образует сообщество.

Пример

При $k=3$ сообщества строятся из пересекающихся треугольников. Один пользователь может участвовать в нескольких таких цепочках и принадлежать нескольким сообществам.

CC BY-NC 4.0

Выбор числа сообществ

Параметры и критерии качества

Некоторые алгоритмы определяют число сообществ автоматически. Другие методы требуют заранее задать количество групп k или параметр масштаба. При выборе разбиения полезно учитывать несколько критериев.

1. Значение модулярности или другой целевой функции.
2. Устойчивость результата при повторных запусках.
3. Содержательную интерпретируемость найденных групп.
4. Согласованность с известными атрибутами участников.
5. Размеры сообществ и их внутреннюю связность.

Критерий

Наилучшее математическое разбиение не обязательно является наиболее полезным для исследовательской задачи.

Проблема разрешения

Малые сообщества могут быть незаметны

Оптимизация модулярности может не выделять небольшие, но реально существующие сообщества. Это явление называют **проблемой разрешения модулярности**. Два малых сообщества могут быть объединены алгоритмом, если такое объединение повышает итоговое значение Q. Размер различимых сообществ зависит от размера всей сети и количества её рёбер.

Интерпретация

Сообщество из нескольких тесно связанных пользователей может потеряться внутри очень большой социальной сети.

На практике

Для проверки результатов используют несколько алгоритмов, параметры масштаба и анализ локальных подграфов.

Сообщества и атрибуты

Структура связей не равна социальным признакам

Сообщество выделяется на основании структуры рёбер. Атрибуты вершин описывают свойства участников сети: возраст, регион, профессию, язык или интересы. Участники одного структурного сообщества могут иметь сходные атрибуты. Однако это сходство **не является обязательным**.

Пример

В сети обсуждений сообщество может объединять пользователей с разными профессиями. Их связывает не общий социальный статус, а интерес к одной теме.

Вывод

Структурное выделение сообществ и классификация пользователей по атрибутам являются разными задачами.

Межсообщественные связи

Мосты и пограничные участники

Некоторые вершины имеют связи сразу с несколькими сообществами. Такие участники могут передавать информацию между группами и снижать фрагментацию сети.
Для их поиска используют центральности, коэффициент участия и анализ межгрупповых рёбер.

Интерпретация

Пользователь, состоящий в нескольких профессиональных сообществах, может быть посредником при распространении новых идей.

Пример

Журналист связывает экспертное сообщество, государственные организации и широкую аудиторию. Его связи могут соединять группы, которые слабо взаимодействуют напрямую.

Динамика сообществ

Изменение групп во времени

В реальных социальных сетях сообщества не остаются неизменными. Группы могут возникать, расширяться, разделяться, объединяться и исчезать. Для динамического анализа сеть разбивается на временные срезы. Затем сообщества из соседних периодов сопоставляются по составу вершин и структуре связей.

Вопросы исследования

- Какие группы сохраняются дольше других.
- Какие участники переходят между сообществами.
- Какие события приводят к слиянию или распаду групп.

Организационный сетевой анализ

Реальные коммуникации вместо формальной оргструктуры

Organizational Network Analysis (ONA) — применение сетевых методов к взаимодействиям внутри организации.

ONA помогает выявить:

- неформальных лидеров;
- перегруженных посредников;
- изолированные команды;
- барьеры между подразделениями;
- риск потери экспертизы при уходе сотрудников.

Источники данных:

- добровольные опросы;
- данные о совместной работе;
- календари и системы задач;
- корпоративные мессенджеры;
- журналы взаимодействий.

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.