DOI: 10.32347/2412-9933.2020.43.19-25
УДК 658.5:330.131.5

Бушуєв Сергій Дмитрович
Доктор технічних наук, професор, завідувач кафедри управління проєктами, *orcid.org/0000-0002-7815-8129*
Київський національний університет будівництва і архітектури, Київ
Трач Роман Володимирович
Кандидат економічних наук, доцент кафедри мостів і тунелів, опору матеріалів і будівельної механіки *orcid.org/0000-0001-6654-9870*
Національний університет водного господарства та природокористування, Рівне

МЕТОДИ КЛАСТЕРИЗАЦІЇ МЕРЕЖІ УЧАСНИКІВ РЕАЛІЗАЦІЇ ПРОЄКТУ

Анотація. *Одним з найбільш затребуваних завдань аналізу мережі учасників проєкту, що представлені у формі графа, є кластеризація множини вузлів, тобто виокремлення таких підмножин, в кожній з яких вузли пов'язані між собою більшою мірою, ніж з вузлами поза цією підмножиною. Для оцінювання якості алгоритму кластеризації використовується функціонал модулярності, яка являє собою скалярну величину на відрізку [-1, 1] та кількісно описує неформальне визначення структури спільнот. Розглянуто класифікацію методів, що використовують для кластеризації графів: жадібні алгоритми; методи переміщень; методи оптимізації, які застосовуються для визначення оптимальної кластеризації мереж. Проаналізовано два алгоритми для кластеризації графів, в основу яких покладено функціонал модулярності. Перший алгоритм – метод Girvan-Newman, який грунтується на аналізі графа за допомогою методів ієрархічної кластеризації і забезпечує виявлення природного поділу мереж на групи залежно від мір подібності або сили зв'язків між вузлами. Кожен крок алгоритму починається з обчислення значення посередництва для кожного ребра в графі, а потім ребро з найбільшим значенням цієї міри видаляється. Так мережа розбивається на незв'язні компоненти, кожна з яких своєю чергою, піддається тій же процедурі. Розбиття може проводитися допоки в графі не залишиться ребер або поки модулярність результуючого розбиття не досягне максимуму. Другий алгоритм – метод Louvain, який належить до жадібної агломеративної ієрархічної кластеризації і являє собою багатоетапну процедуру, яка передбачає локальну оптимізацію модулярності по відношенню до сусідів кожного вузла. Алгоритм Louvain можна розділити на два етапи. На першому етапі кожному вузлу мережі призначається окрема спільнота, потім для кожного вузла вивчаються варіанти зміни модулярності, при можливому переміщенні вузла між спільнотами. Вузол переміщається в ту спільноту, в якій значення модулярності є максимальним. Другий етап алгоритму полягає в тому, що на підставі поділів, отриманих після виконання локальної оптимізації, відбувається побудова нового графа, вузлами якої є спільноти, знайдені на першому етапі. Наведено приклади застосування обох алгоритмів для кластеризації мережі дружби між людьми в «The karate club study of Zachary».*

Ключові слова: кластеризація; проєкт; алгоритм; графи; виокремлення спільнот; мережа

Постановка проблеми

Одним з найбільш затребуваних завдань аналізу мережі учасників проєкту, які візуально представлені у формі графа, є кластеризація множини вузлів, тобто виокремлення таких підмножин, в кожній з яких вузли пов'язані між собою більшою мірою, ніж з вузлами поза цією підмножиною [1; 2]. Кластеризація (виокремлення спільнот) найчастіше ґрунтується на обчисленні тих чи інших критеріїв оптимальності розбиття мережі (графа) на підмножини.

Кластеризація означає процес групування набору об'єктів в підмножини або кластери, при якому виконується умова, що об'єкти, які належать до одного кластеру більш «пов'язані», ніж ті, що належать до різних кластерів [3].

В основі багатьох методів і алгоритмів кластеризації лежать стандартні методи «розділяй і володарюй», тобто виконується певна кількість скорочень допоки поки задачу не вдасться розв'язати досить легко. Етап скорочення може настільки змінити вхідний граф, що «скорочений» варіант не обов'язково повинен бути субстанцією вхідної задачі. Етап відновлення зазвичай складається з двох частин. На першому кроці виявляють певні підструктури: мости, які розділяють кластери та щільні групи, що об'єднують частини кластерів.

Така ідентифікація може бути сформульована як гіпотеза, а утворені підструктури вважаються доказом її правильності. Після фази розпізнавання до поточного графа застосовується відповідна модифікація. Такими перетвореннями можуть бути довільні графові операції, такі як додавання і видалення ребер, а також згортання підграфів, тобто представлення підграфа новим мета-вузлом.

Аналіз останніх досліджень і публікацій

Отже, методи, що використовують для кластеризації графів більшість авторів розділяє на три групи:

– жадібні [4];

– переміщень [5];

– оптимізації, що застосовуються для знаходження оптимальної кластеризації [6].

1. Основою *жадібних методів* є послідовне застосування локально оптимального вибору. Такі методи ефективні в задачах, в самій природі яких закладено, що послідовність локально оптимальних виборів приводить до глобально оптимального рішення. Наприклад, алгоритм Дейкстри [7] (метод знаходження найкоротшого шляху в графі) є жадібним, тому що на кожному кроці шукають вузол з найменшою вагою, в якому ще не були, після чого оновлюють значення інших вузлів. На кожному кроці алгоритм робить вибір такого варіанта, який здається найкращим на цьому етапі. Вибір, зроблений в жадібному алгоритмі, може залежати від зроблених раніше виборів, але він ніяк не залежить від виборів на наступних кроках або від подальших рішень. Таку поведінку називають евристичною, тобто обраний метод не обов'язково гарантує знаходження оптимального рішення, але знайдений розв'язок приймається як допустимий в умовах вирішення конкретного завдання.

Більшість жадібних методів вписуються в таку структуру: почніть з тривіального й можливого для реалізації рішення та виконуйте операції оновлення для рекурсивного зниження витрат, поки подальша оптимізація виявиться неможливою. Наведена ітераційна схема також може бути виражена для ієрархічної кластеризації, яка являє собою процес ітераційного укрупнення. Жадібні методи, які в якості оновлень використовують операції *злиття або поділу* [8], природним способом формують ієрархію. Обмеження однієї з цих операцій гарантує сумісність кластерів, а отже, призводить до ієрархії.

При використанні *злиття (агломерації)* завдання полягає в тому, щоб обчислити схожість між парами вузлів, а потім в порожню мережу (n вузлів без ребер) додати ребра, починаючи з пар вузлів з найбільшою схожістю. Процедура може бути зупинена в будь-який момент, і результуючі компоненти в мережі вважаються спільнотами. Основна ідея цього процесу полягає в тому, щоб виконати найдешевшу операцію злиття. Агломераційні процеси, засновані на широкому розмаїтті мір схожості, були застосовані до різних мереж.

При застосуванні процесів *поділу* завдання полягає в пошуку на заданому графі найменш схожих пар пов'язаних вузлів, а потім видаленні ребра між ними. Повторюючи цей процес багаторазово, можна ділити мережу на все більш дрібні компоненти. Алгоритм можна зупинити на будь-якому етапі і прийняти сформовані на цьому етапі компоненти за мережеві спільноти. Подібно до процесу злиття, основна ідея цього процесу полягає у виконанні найдешевшої операції поділу. На відміну від злиття, модель поділу має додатковий ступінь свободи, оскільки кластери можуть бути скорочені декількома способами.

Перевагою жадібних методів є досить висока швидкість роботи, оскільки як рішення формується тільки один раз і на кожному етапі виконується лише вибір найкращого кроку без урахування результатів наступних кроків. Важливо відзначити, що в багатьох оптимізаційних задачах ефективність жадібних методів значно залежить від вибору початкового рішення та умов задачі, що розв'язується.

2. На відміну від жадібних методів, які діють глобально, *методи переміщень* працюють більш локально. Ці методи вибирають початкову кластеризацію й ітераційно модифікують її, поки не буде знайдено локальний оптимум. Зазвичай виконуються три операції:

– вузол переміщається між вже існуючими кластерами;

– вузол переміщається з одного кластера в новостворений кластер;

– два вузла міняються своїми місцями в кластерах.

Іноді допускаються більш складні операції, такі як миттєве видалення одного кластера і перепризначення вузлів в уже наявних кластерах.

Методи переміщень мають досить багато ступенів свободи, але вирішальне значення має вибір потенційних функцій. У методі переміщень є два підтипи для вибору функцій: базовий і стислий. Методи, засновані на базовому типі – це функції, які сильно залежать від типу операцій і часто використовуються для збереження певних властивостей. У разі, коли створення нового кластера при переміщеннях вузла є дуже дорогим, ймовірно, що число кластерів в остаточній кластеризації буде таким же, як і в початковій. Стислі функції об'єднують серію операцій в одну метаоперацію та оцінюють тільки її результат, що дає змогу

здійснювати певну кількість операцій безкоштовно. Стислі функції часто використовують у поєднанні зі стандартною функцією, яка має безліч локальних оптимумів. Ігноруючи ряд проміжних кроків, можна легше досягти глобального оптимуму.

3. Оптимізаційні методи, які застосовуються для знаходження оптимальної кластеризації засновані на ідеї, що задача кластеризації мережі може бути розв'язана, як результат загального процесу оптимізації. Вхідні дані можуть бути певним чином згенеровані з неявною структурою кластеризації. Завдання оптимізації полягає в тому, щоб «витягнути» кластеризацію, яка відносно близька до неявної. Альтернативно, автономна кластеризація є результатом невідомого процесу оптимізації. Відомо тільки, що цей процес враховує певні парадигми, такі як щільність всередині кластерів, розрідженість між кластерами.

Різноманітність методів для вирішення завдань оптимізації є дуже великою, тому розглянемо лише популярні останнім часом еволюційні методи. Ці методи добре зарекомендували себе на практиці і використовують закономірності та принципи, запозичені у природи. Еволюційні методи належать до так званих популяційних методів, оскільки використовують системи, що складаються з популяцій агентів. Як правило, під агентом розуміється деяка точка в множині пошуку рішень, а процес оптимізації полягає в переміщенні агентів в цій множині. При цьому методи еволюційної оптимізації передбачають створення на кожному кроці нових популяцій агентів за допомогою модифікації вхідної сукупності, а також з урахуванням досвіду, отриманого на попередніх ітераціях. Найбільш поширеними ітераційним операціями є «кросовер» і «мутація». «Кросовер» створює нового агента шляхом рекомбінації двох наявних, в той час як «мутація» змінює «старого» агента. Кожен агент має відповідати певним значенням придатності, які зазвичай виражаються функцією оптимізації. Після низки основних операцій на базі наявної популяції створюється нова, в якій агенти відбираються відповідно до їх придатності. Загальна проблема полягає в тому, щоб гарантувати здійсненність модифікованих рішень. Зазвичай це досягається за допомогою специфікації моделі. В контексті кластеризації модель може використовувати або розбиття, або відносини еквівалентності. Підходи, засновані на пошуку, використовують задану (неявну) топологію простору-агента і виконують випадкове блукання, починаючи з довільного агента. Межі кластеризації зазвичай являють собою набір кластерів, які виникають в результаті зсуву вузлів, злиття або поділу кластерів. Вибір меж також заснований на деякому значенні придатності агента, заданого функцією оптимізації. Пошук зазвичай зупиняється після певної кількості ітерацій або після знаходження локального оптимуму.

Мета статті

Метою статті є аналіз методів, що використовують для кластеризації (виокремлення спільнот) мережі учасників реалізації проєкту.

Виклад основного матеріалу

Наведені методи лежать в основі графових алгоритмів, які використовуються для вирішення завдань кластеризації – графів. Графові алгоритми кластеризації – сукупність алгоритмів, які спрямовані на упорядкування даних та створення ієрархії вкладених кластерів (спільнот). Поняття спільноти неоднозначне і пов'язане з класифікацією об'єктів за категоріями з метою запам'ятовування і пошуку інформації. Залежно від контексту воно може бути еквівалентне модулю, класу, групи, кластеру і т.д. На концептуальному рівні спільнотою називається така група вузлів, в яких внутрішньогрупові зв'язки набагато щільніші міжгрупових [9]. Використання алгоритмів кластеризації забезпечує поділ набору об'єктів на кластери, що не пересікаються, таким чином, що члени одного кластера дуже «схожі» між собою, тоді як об'єкти, що належать до різних кластерів, відносно неоднакові [10]. Для оцінювання якості алгоритму кластеризації використовується функціонал модулярності, який був запропонований Girvan і Newman [11].

Модулярність – це скалярна величина на відрізку [-1, 1], яка кількісно описує неформальне визначення структури спільнот та розраховується так:

$$Q = \frac{|E_{in}| - |E_{in-R}|}{|E|}, \qquad (1)$$

де $|E_{in}|$ – кількість зв'язків, що з'єднують вузли, які належать до однієї спільноти; $|E_{in-R}|$ – оцінюється як $|E_{in}|$, якщо зв'язки були випадковими.

Перевагою модулярності є те, що вона досить просто інтерпретується. Її значення дорівнює різниці між часткою ребер всередині спільноти і очікуваної частки зв'язків, якби ребра були розміщені випадково. Модулярність досить ефективно перераховується при невеликих змінах в кластерах. Недоліком модулярності є те, що її функціонал не є безперервним, і завдання його оптимізації – дискретне. Для пошуку глобального оптимуму використовують наближені схеми. Деякі з них дійсно оптимізують значення функціоналу, інші ж за значення модулярності вибирають найкраще

рішення з визначених, тобто без гарантій локальної оптимальності рішення.

Критерій оцінки якості виявлених спільнот – модулярність визначається на основі щільності зв'язків всередині спільноти в порівнянні зі зв'язками між спільнотами. Для зваженого графа модулярність виражається так [12]:

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

де A_{ij} – вага ребра між вузлами i, j; k_i, k_j – сума ваг ребер, що сполучені з вузлами i, j; δ – функція $\delta(u, v)$, яка дорівнює 1, якщо u = v, інакше = 0; c_i, c_j – спільноти вузлів i, j; m – напівсума ваг всіх ребер графа ($m = \frac{1}{2}\sum_{i,j\in} A_{ij}$).

Отже, модулярність дорівнює різниці між часткою ребер всередині спільноти при даному розбитті та часткою ребер, якби вони були випадково згенеровані. Саме тому вона показує вираженість спільнот (випадковий граф структура спільнот не має). Також слід відзначити, що модулярність дорівнює 1 для повного графа, в якому всі вузли поміщені в одну спільноту та дорівнює нулю для розбиття, при якому кожен вузол є окремою спільнотою. Для особливо невдалого розбиття модулярність може бути від'ємною.

Виграш модулярності ΔQ, отриманий шляхом переміщення вузла *i* до спільноти *C*, можна обчислити так:

$$Q = \left[\frac{\sum in + k_{i,in}}{2m} - \left(\frac{\sum tot + k_i}{2m}\right)^2\right] - \left[\frac{\sum in}{2m} - \left(\frac{\sum in}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right], \qquad (3)$$

де $\sum in$ – сума ваг всіх зв'язків всередині спільноти *C*, з якої переміщається вузол i; $\sum tot$ – сума ваг всіх зв'язків в спільноті *C*; k_i – сума ваг зв'язків вузлів i; $k_{i,in}$ – сума ваг зв'язків між *i* та іншими вузлами *C*; m – сума ваг всіх зв'язків в мережі.

З погляду на зростаючу популярність напряму аналізу мереж та значне різноманіття моделей кластеризації графів, дослідження буде зосереджене лише на алгоритмах, які найбільш адаптовані для вирішення раніше поставлених завдань. У 2001 році M. Girvan і M. Newman [13] представили алгоритм, який став одним з перших ієрархічних методів виокремлення спільнот для неорієнтованого і незваженого графа. Авторами використано підхід, який полягає в аналізі мережі за допомогою методів ієрархічної кластеризації. Ці методи спрямовані на виявлення природного поділу мереж на групи і засновані на різних мірах подібності або сили зв'язків між вузлами. Переваги методу:

1. Автори фокусуються не на видаленні ребер між парами вузлів з найменшою схожістю, а на пошуку ребер з найбільшим рівнем «посередництва» (betweenness). Найпростішим прикладом такого показника «посередництва» є міра, заснована на найкоротших (геодезичних) шляхах. Для цієї мети авторами була адаптована добре відома міра центральності вузла за посередництвом Фрімена (betweenness centrality) [14]. Значення посередництва розраховується як кількість найкоротших шляхів між усіма парами вузлів, що проходять через це ребро. Якщо між вузлами *n* найкоротших шляхів, то кожному ребру додається *1/n* до значення коефіцієнта. Чим більше ця величина, тим більш імовірно, що це ребро з'єднує вузли з різних спільнот.

2. Включення в алгоритм «кроку перерахунку». При виконанні стандартної кластеризації розподілу на основі міжграничних ребер, необхідно було розраховувати граничність між усіма ребрами в мережі, а потім ребра видалялися за порядком зменшення інтервалу. Однак в такому алгоритмі після видалення першого ребра в мережі значення посередництва для інших ребер більше не будуть відображати модифіковану мережу, що може призвести до небажаного результату. Для вирішення цієї проблеми автори пропонують перераховувати міру посередництва після видалення кожного ребра.

Отже, кожен крок алгоритму починається з обчислення значення посередництва для кожного ребра в графі, а потім ребро з найбільшим значенням цієї міри видаляється. Так мережа розбивається на незв'язні компоненти, кожна з яких своєю чергою піддається тій же процедурі. Розбиття може проводитися допоки в графі не залишиться ребер або поки модулярність результуючого розбиття не досягне максимуму.

Схема роботи алгоритму наведена на рис. 1.

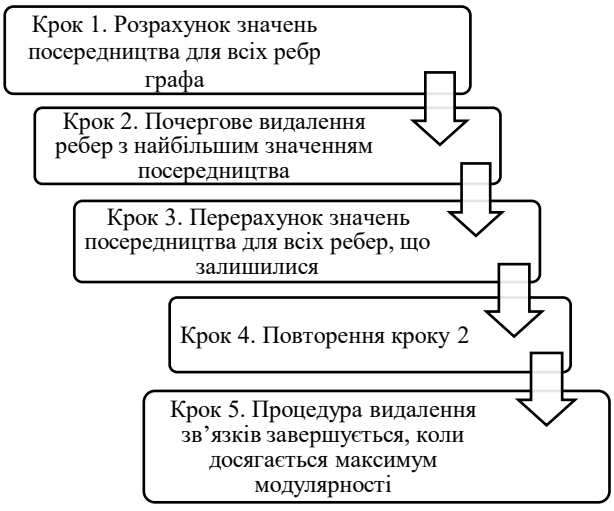

Рисунок 1 – Схема роботи алгоритму Girvan-Newman

Графічно процес виконання алгоритму можна представити у вигляді ієрархічного дерева або дендрограми (рис. 2).

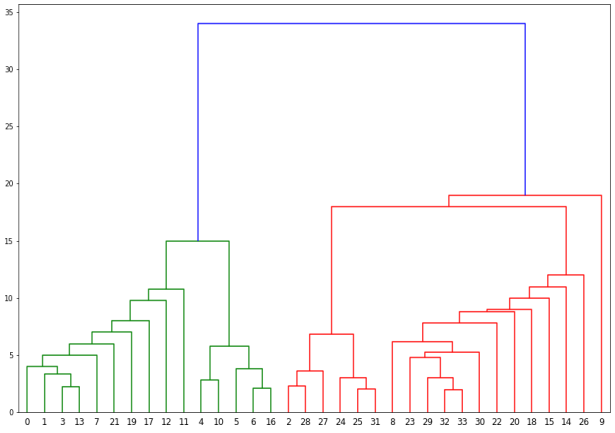

Рисунок 2 – Ієрархічне дерево, що ілюструє мережу дружби між людьми в «The karate club study of Zachary» [15]

Як альтернативні автори алгоритму пропонують використовувати методи «Випадкового блукання» (Random walks) і «Резисторних мереж» (Resistor networks). Найкоротший шлях між вузлами можна уявити в термінах сигналів, що проходять через мережу. Якщо сигнали переміщаються від джерела до місця призначення шляхами геодезичної мережі, і всі вузли посилають сигнали з однаковою сталою швидкістю, то посередництво є мірою швидкості, з якою сигнали проходять уздовж кожного ребра. Можна припустити, що сигнали не поширюються геодезичними шляхами, а замість цього просто виконують випадковий обхід мережі, поки не досягнуть свого місця призначення. Це дає змогу використовувати на ребрах ще одну міру – випадкове блукання. Для визначення випадкового блукання необхідно вирахувати, як часто в середньому випадкові блукання, що починаються з вузла *s*, будуть проходити по певному ребру від вузла *v* до вузла *w* (або навпаки), перш ніж знайти шлях до заданого цільового вузла *t*.

Наступна міра мотивується ідеями теорії елементарних ланцюгів. Автори розглядають схему, яка утворена за допомогою розміщення одиничного опору на кожному краю мережі одиничного джерела струму і приймача в конкретній парі вузлів. Результуючий потік струму в мережі буде проходити від джерела до приймача по множині шляхів, причому шлях з найменшим опором отримає найбільшу частку струму. Проміжний струм між ребром, який автори визначають як абсолютне значення струму уздовж ребра, підсумовується по всіх парах «джерело / приймач».

Іншим поширеним методом машинного навчання, призначеним для виявлення спільнот у великих мережах даних, є метод Louvain, який забезпечує позитивний баланс між універсальністю і продуктивністю [16]. Метод належить до жадібної агломеративної (злиття) ієрархічної кластеризації і заснований на максимізації модулярності [17]. Алгоритм розроблено в 2008 році колективом авторів: V. Blondel, J. Guillaume, R. Lambiotte і E. Lefebvre [18] і допомагає проводити оптимізацію неорієнтованих зважених графів. Алгоритм являє собою багатоетапну процедуру і передбачає локальну оптимізацію модулярності по відношенню до сусідів кожного вузла: процедура виконується ітераційно допоки триває зростання модулярності.

Алгоритм Louvain можна розділити на два етапи:

1. Етап локальної оптимізації. На першому етапі кожному вузлу мережі призначається окрема спільнота, а отже, при початковому розподілі в графі є стільки спільнот, скільки й вузлів.

Отже, спочатку для кожного вузла i відбувається ідентифікація всіх суміжних кластерів $N_c(i)$, тобто кластерів, що містять щонайменше один сусідній вузол j та $l(i) \neq l(j)$ (мітка кластера i, $l(i)$ відрізняється від мітки кластера j, $l(j)$.

Потім для кожного вузла i вивчаються варіанти зміни модулярності при можливому переміщенні вузла з однієї в іншу спільноту. Вузол i переміщається в ту спільноту, в якій значення модулярності є максимальним. Якщо позитивного результату (виграшу) від переміщення вузла немає – вузол залишається в своїй початковій спільноті.

2. Етап агрегації або укрупнення. Другий етап алгоритму полягає в тому, що на підставі поділів, отриманих після виконання локальної оптимізації, відбувається побудова нового графа G', вузлами якого є спільноти, знайдені на першому етапі. Кожен кластер C_l графа G, утворений на етапі локальної оптимізації, стає вузлом графа G'. Зв'язки між вузлами однієї і тієї ж спільноти приводять до створення контурів цієї спільноти в новій мережі.

Ваги ребер нового графа G' визначаються, як сума ваг зв'язків, які з'єднують між собою кластери графа G (міжкластерні ваги).

Послідовність етапів 1 і 2 називається прогоном. Кожен запуск приводить до створення нового кластерного розділу, який виникає в результаті об'єднання кластерів, отриманих при попередньому запуску. Це створює ієрархію вкладених розділів на різних рівнях агрегації. Прогони повторюються допоки стане неможливо поліпшити модулярність, тобто не буде переміщення вузла (локальна оптимізація) і агрегування вузлів з подальшою локальною оптимізацією.

Схема роботи алгоритму наведена на рис. 3.

Рисунок 3 – Схема роботи алгоритму Louvain

Приклад виконання алгоритму наведено на рис. 4. Алгоритм було застосовано до тієї самої мережі університетського клубу карате.

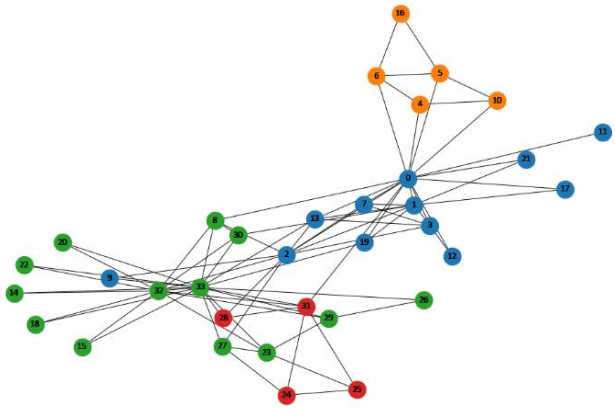

Рисунок 4 – Кластеризація мережі дружби між людьми в «The karate club study of Zachary»

Цей алгоритм має кілька переваг:

– етапи реалізації алгоритму інтуїтивно зрозумілі і прості;

– алгоритм є надзвичайно швидким, тобто комп'ютерне моделювання у великих спеціальних модульних мережах передбачає, що його складність є лінійною для типових і розріджених даних.

Простота і ефективність роблять алгоритм Louvain популярним рішенням для кластеризації мережі. Однак він має свої обмеження:

– ітеративний процес злиття кластерів може призвести до того, що важливі, але невеликі кластери можуть бути поглинуті;

– в мережах, які містять кластери, що перекриваються, визначити оптимальне кластерне рішення може бути складно, що призводить до безлічі можливих кластерних конфігурацій.

Висновок

Розглянуто класифікацію методів, що використовують для кластеризації графів: жадібні алгоритми; методи переміщень; методи оптимізації, які застосовуються для визначення оптимальної кластеризації мереж. Проаналізовано два алгоритми для кластеризації графів, в основу яких покладено функціонал модулярності. Перший алгоритм – метод Girvan-Newman, який полягає в аналізі графа за допомогою методів ієрархічної кластеризації та забезпечує виявлення природного поділу мереж на групи, залежно від мір подібності або сили зв'язків між вузлами.

Другий алгоритм – метод Louvain, який належить до жадібної агломеративної ієрархічної кластеризації і являє собою багатоетапну процедуру, яка передбачає локальну оптимізацію модулярності по відношенню до сусідів кожного вузла. Наведено приклади застосування обох алгоритмів для кластеризації мережі дружби між людьми в «The karate club study of Zachary».

Список літератури

1. *Garey, M.R. & Johnson, D.S. (1979). Computers and Intractability: A Guide to the Theory of NP-Completeness Freeman, San Francisco.*
2. *Scott, J. (2000). Social Network Analysis: A Handbook, 2nd ed. Sage Publications, London.*
3. *Everitt, B. S., Landau, S. & Leese, M. (2011). Cluster analysis. 5th ed. Arnold.*
4. *Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical review E, 69(2), 026113.*
5. *Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. In International symposium on computer and information sciences, 284-293, Springer, Berlin, Heidelberg.*
6. *Wu, F. & Huberman, B. A. (2004). Finding communities in linear time: a physics approach. The European Physical Journal B, 38(2), 331-338.*
7. *Dijkstra, E.W. (1959). A note on two problems in connexion with graphs. Numerische mathematik, 1(1), 269-271.*
8. *Scott, J. (1988). Social network analysis. Sociology, 22(1), 109-127.*
9. *Thomas, S.L.(2000). Ties that Bind: A Social Network Approach to Understanding Student Integration and Persistence. Journal of Higher Education, 71(5), 591 – 615.*

10. Xu, R. & Wunsch, D. (2005). Survey of clustering algorithms. IEEE Transactions on Neural Networks, 16(3):645–678.

11. Newman, M.E. (2003). Mixing patterns in networks. Physical Review E, 67(2), 026126.

12. Newman, M.E. (2004). Analysis of weighted networks. Physical review E, 70(5), 056131.

13. Girvan, M., & Newman, M.E. (2001). Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA, 99 (cond-mat/0112110), 8271-8276.

14. Leskovec, J., Lang, K.J. & Mahoney, M. (2010). Empirical comparison of algorithms for network community detection. Proceedings of the 19th international conference on World wide web. ACM, 631-640.

https://doi.org/10.1145/1772690.1772755

15. Zachary, W.W. (1977). An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 33, 452–473

16. Brath, R. & Jonker, D. (2015). Graph analysis and visualization: discovering business opportunity in linked data. John Wiley & Sons.

17. Tang, L., & Liu, H. (2010). Community detection and mining in social media. Synthesis lectures on data mining and knowledge discovery, 2(1), 1-137. https://doi.org/10.2200/S00298ED1V01Y201009DMK003

18. Blondel, V.D., Guillaume, J.L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 10, P10008.

Стаття надійшла до редколегії 05.09.2020

Bushuyev Sergiy

DSc (Eng.), Professor, Head of Project Management, *orcid.org/0000-0002-7815-8129*

Kyiv National University of Construction and Architecture, Kyiv

Trach Roman

PhD (Economics), Associate Professor of the Department of Bridges, Tunnels, Strength of Materials and Structural Mechanics, *orcid.org/0000-0001-6654-9870*

National University of Water and Environmental Engineering, Rivne

CLUSTERING METHODS OF PROJECT PARTICIPANTS NETWORK

Abstract. *One of the most popular tasks for analyzing the project participants network presented, in the graph form, is the clustering of set nodes. Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). Assess the quality of clustering algorithm, the modularity functional is used, which is a scalar value on the interval [-1, 1] and quantitatively describes an informal definition of the community structure. The classification of methods used for clustering graphs is considered: greedy; displacement methods; optimization methods that are used to find the optimal clustering of networks. Two graph clustering algorithms are analyzed, which are based on the modularity functional. The first algorithm is the Girvan-Newman method, which is based on the analysis of the graph using hierarchical clustering methods and provides detection of separation network of links into groups, depending on measures of similarity or strength of links between nodes. Each step of algorithm begins by calculating value of betweeness for each edge in the graph, and then edge with the highest value of this measure is deleted. The network is divided into disconnected components, each of which, in turn, undergoes the same procedure. A breakdown can be carried out until there are no edges in graph or until resulting modularity the split reaches a maximum. The second algorithm, Louvain method, refers to greedy agglomerative hierarchical clustering. The algorithm is a multi-stage procedure, which provides for local modularity optimization with relationship to the neighbors of each node. The Louvain algorithm can be divided into two stages. At the first stage, a separate community is assigned to each node of network, options for modularity change are studied for each node, with the possible node movement between communities. The node moves to community in which modularity value is maximum. The second algorithm stage is that, based on division obtained after performing local optimization, a new graph is constructed, nodes of which are the communities found in the first stage. Examples of using both algorithms for clustering a network of friendship between people in "The karate club study of Zachary" are given.*

Keywords: clustering, project, algorithm, graphs, community allocations, network

---

Посилання на публікацію

APA *Bushuyev, S. & Trach, R. (2020). Clustering methods of project participants network. Management of Development of Complex Systems, 43, 19 – 25; dx.doi.org\10.32347/2412-9933.2020.43.19-25.*

ДСТУ *Бушуєв С. Д. Методи кластеризації мережі учасників реалізації проєкту [Текст] / С. Д. Бушуєв, Р. В. Трач // Управління розвитком складних систем. – 2020. – № 43. – С. 19 – 25; dx.doi.org\10.32347/2412-9933.2020.43.19-25.*