

ПОРІВНЯЛЬНИЙ АНАЛІЗ ОБЧИСЛЮВАЛЬНОЇ ЕФЕКТИВНОСТІ ЩІЛЬНОГО ТА РОЗРІДЖЕНОГО ПІДХОДІВ У СПЕКТРАЛЬНОМУ АНАЛІЗІ МАСШТАБНИХ ГРАФІВ- ШЛЯХІВ

Гусак Станіслав Леонідович

студент групи МАМм-1-25-14д,

Київського столичного університету імені Бориса Грінченка,

науковий керівник – к.ф.-м. н. Соломко В.О.

ВСТУП

Аналіз складних мережевих структур є важливим напрямом сучасних досліджень у галузі моделювання складних систем. Теорія графів та її спектральний розділ дозволяють точно описати зв'язки в мережі й дослідити її структуру через власні значення відповідних матриць. Зокрема, розріджені графи широко застосовуються при моделюванні обчислювальних, соціальних, транспортних та біологічних мереж, де критично важливою є оцінка надійності структур та зв'язності при зростанні розмірності системи.

У попередній роботі було розглянуто основи спектрального аналізу на прикладі базової топології графа-шляху малого порядку P_5 , де показано відповідність теоретичних розрахунків спектрів матриці суміжності A та матриці Лапласа L результатам, отриманим за допомогою стандартних функцій бібліотеки NumPy [1]. Однак, при переході від мікро-моделей до реальних масштабних мереж ($n \rightarrow \infty$) виникає складність в обчисленнях. Класичні спектральні методи, орієнтовані на обробку щільних (Dense) матриць, демонструють кубічну часову складність $O(n^3)$ та відповідні затрати оперативної пам'яті $O(n^2)$. При розмірностях $n > 1000$ це призводить до вичерпання системних ресурсів.

Постановка проблеми полягає у подоланні обчислювальних обмежень спектральних алгоритмів при аналізі великомасштабних розріджених графів. Окремою проблемою є обмеження спектрального аналізу через існування коспектральних графів [5; 6], проте для графів-шляхів P_n спектр однозначно визначає топологію. Порівняння Dense та Sparse підходів дозволяє встановити граничні межі застосування та подолати обмеження щільного методу при масштабуванні n .

Метою дослідження є розробка та експериментальне дослідження алгоритмів спектрального аналізу графів-шляхів P_n при масштабуванні кількості вершин $n \in [5, 10\,000]$, визначення часових меж обчислюваності щільного та розрідженого підходів, а також оцінка динаміки зниження алгебраїчної зв'язності (числа Фідлера μ_2) та щільності графа $d(n)$ при такому масштабуванні.

Для досягнення поставленої мети визначено такі завдання:

1. Сформулювати математичний опис матриці Лапласа L для графа-шляху довільного порядку P_n ;

2. Запрограмувати обчислення щільності графа $d(n)$ та алгебраїчної зв'язності (числа Фідлера μ_2);
3. Провести порівняння часових витрат щільного та розрідженого методів;
4. Проаналізувати граничні значення n .

Аналіз досліджень. Аналіз наукових публікацій свідчить про значний інтерес до спектральної теорії графів та алгоритмів її застосування. У фундаментальних працях М. Фідлера [2], Ф. Чанг [3], а також А. Брауера та В. Хаемерса [4] викладено класичні основи спектрального аналізу матриць Лапласа та суміжності. Дослідження коспектральних структур [5; 6] показують особливості спектральної ідентифікації мереж. Сучасні роботи [7; 8] присвячені аналізу спектральних властивостей складних алгебраїчних систем та розріджених мереж великої розмірності. Алгоритмічні аспекти реалізації розріджених алгебраїчних методів висвітлено у працях [9–12]. Втім, питання експериментального порівняння часової ефективності Dense та Sparse методів для графів P_n до $n=10\,000$ залишається відкритим. Питання ефективного застосування математичних методів та інформаційних технологій у наукових дослідженнях і моделюванні детально розглядаються у праці [13].

ОПИС ДОСЛІДЖЕННЯ, ЙОГО МЕТОДІВ І ЗАСОБІВ

Для дослідження обчислювальної ефективності алгоритмів спектрального аналізу графів великої розмірності у роботі побудовано математичну модель графа-шляху P_n та розроблено відповідне програмне забезпечення мовою Python у вигляді скрипта.

Математична модель та об'єкти дослідження. Граф-шлях P_n є неорієнтованим зв'язним графом, що складається з n вершин $\{v_1, v_2, \dots, v_n\}$ та $m = n - 1$ ребер, які утворюють єдиний лінійний ланцюг [1; 3; 7]. Основним об'єктом спектрального аналізу обрано матрицю Лапласа $L = D - A$, де D — діагональна матриця степенів вершин, A — матриця суміжності графа [3; 4].

Для графа P_n матриця Лапласа є симетричною тридіагональною матрицею розмірності $n \times n$:

$$L_{i,j} = \{1, \text{якщо } i = j \in \{1, n\}, 2, \text{якщо } i = j \in \{2, \dots, n-1\}, -1, \text{якщо } |i - j| = 1, 0, \text{в інших випадках.}$$

Важливою кількісною характеристикою топологічної розрідженості графа є його щільність $d(n)$, що визначається як відношення фактичної кількості ребер m до максимально можливої кількості ребер у повному графі K_n :

$$d(n) = \frac{2m}{n(n-1)} = \frac{2(n-1)}{n(n-1)} = \frac{2}{n}.$$

Зі зростанням порядку графа $n \rightarrow \infty$ щільність прямує до нуля ($d(n) \rightarrow 0$), що підтверджує приналежність P_n до класу високо розріджених графів [8]. На відміну від складних коспектральних графів [5; 6], для P_n спектр Лапласа однозначно характеризує топологічну структуру.

Ключовим показником алгебраїчної зв'язності та структурної стійкості мережі є число Фідлера μ_2 — друге найменше власне значення матриці Лапласа

L ($0 = \mu_1 < \mu_2 \leq \dots \leq \mu_n$) [2; 4]. Для графа-шляху P_n теоретичне значення числа Фідлера описується аналітичною формулою:

$$\mu_2(P_n) = 4\sin^2\left(\frac{\pi}{2n}\right).$$

При великих n асимптотика числа Фідлера має вигляд $\mu_2(P_n) \approx \frac{\pi^2}{n^2}$, що відображає стрімке зниження зв'язності ланцюгової мережі при збільшенні її довжини [2; 3].

Методи та алгоритми обчислення. У розробленому скрипті реалізовано паралельне обчислення спектральних характеристик двома принципово різними методами.

Перший підхід оснований на використанні щільного представлення матриці (Dense / NumPy), де матриця Лапласа L будується у вигляді повного двовимірного масиву `numpy.ndarray` розмірністю $n \times n$. Обчислення повного вектора власних значень здійснюється за допомогою процедури `numpy.linalg.eigvalsh`, яка базується на виклику процедур бібліотеки LAPACK [10]. Повна діагоналізація щільної симетричної матриці має теоретичну часову складність $O(n^3)$ та вимагає обсягу $O(n^2)$ оперативної пам'яті для збереження всіх n^2 елементів.

Слід звернути увагу на важливу методичну відмінність між підходами. Процедура `numpy.linalg.eigvalsh` обчислює абсолютно всі n власних значень матриці, тобто повний спектр. Проте для практичного визначення числа Фідлера такий розрахунок є надлишковим. Розріджений підхід у `scipy.sparse.linalg.eigsh` націлений на пошук лише трьох крайніх власних значень, чого повністю достатньо для розв'язання поставленої задачі. Таким чином, радикальна перевага розрідженого методу досягається не лише завдяки стиснутому формату збереження матриці, а й через відсутність витрат на розрахунок решти непотрібних власних значень.

Другий підхід базується на розрідженому матричному представленні (Sparse / SciPy), у якому матриця L зберігається у стиснутому рядковому форматі CSR (Compressed Sparse Row) за допомогою функції `scipy.sparse.diags` [12]. У цьому випадку в пам'яті фіксуються лише ненульові елементи трьох діагоналей ($3n - 2$ елементи), що скорочує просторову складність до $O(n)$ [9]. Пошук алгебраїчної зв'язності (числа Фідлера μ_2) реалізовано на основі ітераційного алгоритму Арнольді та Ланцоша з бібліотеки ARPACK за допомогою функції `scipy.sparse.linalg.eigsh` [11]. Для прискорення сходження алгоритму застосовано режим інверсії зі зсувом (Shift-Invert mode з параметром $\sigma = -10^{-5}$), що дозволяє швидко виділити $k = 3$ найменших за модулем власних значень. Теоретична часова складність обчислення крайніх спектральних елементів для тридіагональної розрідженої структури становить $O(k \cdot n)$, де $k \ll n$, що забезпечує значне обчислювальне прискорення при масштабуванні n .

Інструментарій у вигляді скрипта розроблено мовою Python з використанням бібліотек наукових обчислень NumPy та SciPy [12].

Вимірювання часу виконання чисельних процедур T_{Dense} та T_{Sparse} здійснювалося за допомогою системного таймера високої точності `time.perf_counter()`. Ефективність розрідженого підходу оцінювалася коефіцієнтом прискорення:

$$Speedup = \frac{T_{Dense}}{T_{Sparse}}.$$

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для оцінки ефективності розробленого програмного забезпечення було проведено серію обчислювальних експериментів із вимірюванням часу формування спектра та розрахунку числа Фідлера μ_2 для графів-шляхів P_n у діапазоні розмірностей $n \in [5, 10\,000]$. Отримані кількісні показники занесено до таблиці 1.

Таблиця 1.

Порівняльна обчислювальна ефективність алгоритмів спектрального аналізу P_n

n	Щільність ь d	Час Dense T_{Dense} (сек)	Час Sparse T_{Sparse} (сек)	Прискорення $Speedup$	Число Фідлера μ_2
5	0,400000	0,000117	0,000683	0,06	0,38196601
10	0,200000	0,000043	0,000511	0,08	0,09788697
50	0,040000	0,000218	0,000851	0,26	0,00394654
100	0,020000	0,000597	0,000760	0,79	0,00098688
200	0,010000	0,001782	0,001136	1,57	0,00024674
500	0,004000	0,032718	0,000968	33,79	0,00003948
1000	0,002000	0,115818	0,001428	81,08	0,00000987
2000	0,001000	0,435296	0,002844	153,05	0,00000247
5000	0,000400	5,779886	0,007433	777,57	0,00000039
10000	0,000200	39,662320	0,020779	1908,77	0,00000010

Аналіз обчислювальних даних свідчить про наявність чітко вираженої точки перелому. Для малих графів порядку $n < 200$ щільний метод на базі процедури `numpy.linalg.eigvalsh` демонструє менші часові затрати порівняно з розрідженим алгоритмом. Це пояснюється низькими витратами на ініціалізацію процедур бібліотеки LAPACK [10] для малих матриць та наявністю додаткових обчислень у стиснутому форматі CSR [12]. У даній області коефіцієнт прискорення є меншим за одиницю ($Speedup < 1$).

При досягненні розмірності $n \approx 200$ спостерігається «точка перелому» ($Speedup \approx 1,57 \times$). При подальшому збільшенні порядку графа $n > 200$ складність щільного методу $O(n^3)$ спричиняє стрімке зростання обчислювальних витрат. Зокрема, при $n = 1000$ час розрахунку щільним методом досягає 0,116 секунди, а при $n = 10\,000$ обчислення тривають понад 39 секунд.

Натомість розріджений алгоритм `scipy.sparse.linalg.eigsh` у режимі Shift-Invert демонструє майже лінійне масштабування часу обчислення [11; 12]. При $n = 10\,000$ розріджений алгоритм знаходить число Фідлера всього за 0,021 секунди, що забезпечує коефіцієнт прискорення $Speedup > 1900 \times$. Таке колосальне прискорення пояснюється як ефективністю розрідженого формату, так і цільовим спрощенням задачі: розрахунком лише трьох потрібних значень замість повного спектра з десяти тисяч значень, який вираховує щільний метод. Динаміку часових характеристик та зростання коефіцієнта прискорення ілюстровано на рисунку 1.

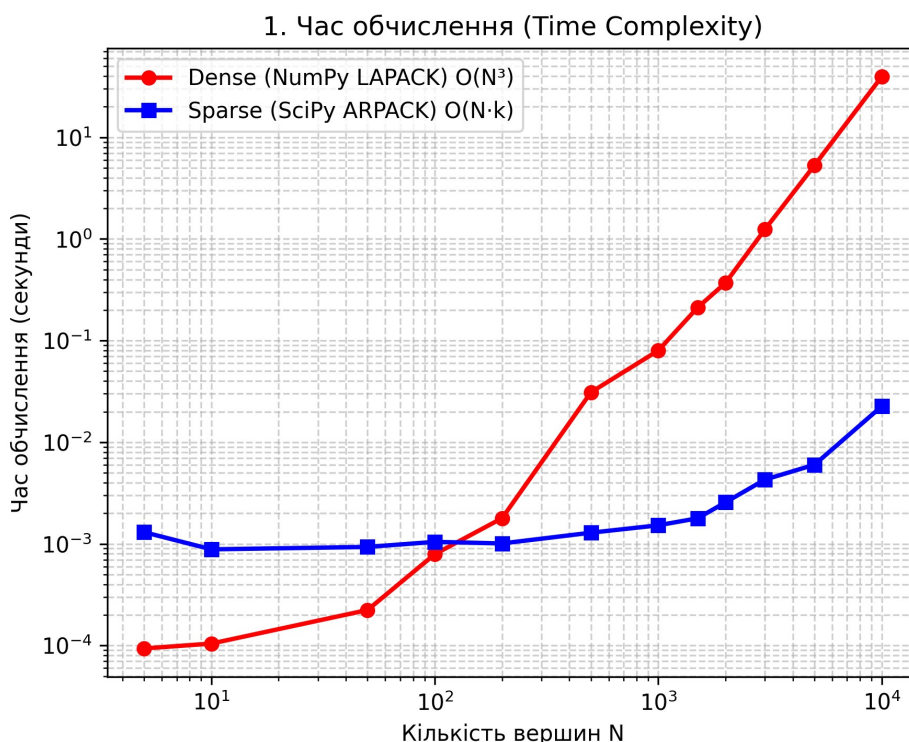


Рис. 1. Залежність часу обчислення $T(n)$ від розмірності графа n

Паралельно з часовими вимірюваннями зафіксовано закономірності зміни структурних та спектральних параметрів топології P_n . Зі зростанням порядку графа щільність $d(n) = \frac{2}{n}$ зменшується від 0,400000 ($n = 5$) до 0,000200 ($n = 10\,000$), що підтверджує належність великомасштабних графів-шляхів до класу розріджених систем [8; 9]. Значення числа Фідлера μ_2 зменшується відповідно до асимптотичної залежності $\mu_2 \approx \frac{\pi^2}{n^2}$ [2; 4]. Зменшення алгебраїчної зв'язності від 0,381966 ($n = 5$) до $9,87 \times 10^{-8}$ ($n = 10\,000$) відображає суттєву структурну вразливість довгих ланцюгових мереж до розривів та втрати зв'язності при виході з ладу навіть одного елемента [2; 3].

```

# 1. Обчислення щільності (d)
actual_edges = n - 1
max_edges = n * (n - 1) / 2
density = actual_edges / max_edges

# 2. Щільний метод (Dense / NumPy)
main_diag = np.full(n, 2.0)
main_diag[0] = 1.0
main_diag[-1] = 1.0
off_diag = np.full(n - 1, -1.0)
L_dense = np.diag(main_diag) + np.diag(off_diag, -1) + np.diag(off_diag, 1)

t0 = time.perf_counter()
vals_dense = np.linalg.eigvalsh(L_dense)
t_dense = time.perf_counter() - t0
fiedler_dense = vals_dense[1] if n > 1 else 0.0

# 3. Розріджений метод (Sparse / SciPy)
L_sparse = sp.diags([off_diag, main_diag, off_diag], [-1, 0, 1], format='csr')

t0 = time.perf_counter()
vals_sparse = spla.eigsh(L_sparse, k=min(3, n-1), sigma=-1e-5, which='LM', return_eigenvectors=False)
t_sparse = time.perf_counter() - t0
vals_sparse_sorted = np.sort(np.abs(vals_sparse))
fiedler_sparse = vals_sparse_sorted[1] if len(vals_sparse_sorted) > 1 else 0.0

# 4. Коефіцієнт прискорення
speedup = t_dense / t_sparse if t_sparse > 0 else 0.0

```

Рис. 2. Фрагменти коду, що виконує обчислення

ВИСНОВКИ

В рамках даного дослідження здійснено математичне та програмне оцінювання обчислювальної ефективності алгоритмів спектрального аналізу для графів-шляхів P_n великої розмірності при масштабуванні кількості вершин $n \in [5, 10\,000]$. На основі проведених експериментальних замірів та аналізу отримано такі основні результати:

Доведено обчислювальну перевагу розрідженого підходу: Експериментально зафіксовано, що для розріджених графів (де щільність прямує до нуля $d(n) = \frac{2}{n} \rightarrow 0$) використання розрідженого представлення у форматі CSR та ітераційного алгоритму Арнольді/Ланцоша (SciPy/ARPACK) забезпечує радикальне скорочення часових витрат [11; 12]. При $n = 10\,000$ розріджений метод обчислює число Фідлера μ_2 за $\sim 0,02$ секунди, забезпечуючи коефіцієнт прискорення $Speedup > 1900 \times$ порівняно зі щільним методом NumPy (LAPACK), який вимагає понад 39 секунд [10].

Визначено «точку перелому» застосовності алгоритмів: Встановлено, що для мікро-графів ($n < 100$) щільний метод є дещо швидшим ($Speedup < 1$) за рахунок відсутності накладних витрат на ініціалізацію ітераційного процесу. Точка паритету обчислювальної ефективності відповідає розмірності $n \approx 100 \dots 200$. При $n > 200$ розріджений підхід стає беззаперечно ефективнішим, а кубічна часова складність щільного методу $O(n^3)$ робить його непридатним для аналізу великомасштабних мережевих систем.

Оцінено динаміку алгебраїчної зв'язності: Підтверджено стрімке зменшення числа Фідлера ($\mu_2 \approx \frac{\pi^2}{n^2}$) при збільшенні довжини ланцюга [2; 4]. Для $P_{10\,000}$ значення μ_2 досягає порядку $9,87 \times 10^{-8}$, що математично вказує на

крайню вразливість і низьку структурну стійкість протяжних лінійних топологій до розривів.

ДЖЕРЕЛА

1. Гусак С. Л. Розробка та дослідження ефективних алгоритмів спектрального аналізу графів. Збірник матеріалів Всеукраїнської конференції молодих учених «Інформаційні технології». 2026. № 13. С. 197–202.
2. Fiedler M. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*. 1973. Vol. 23, No. 2. P. 298–305.
3. Chung F. R. K. Spectral Graph Theory. Providence, RI: American Mathematical Society, 1997. 207 p.
4. Brouwer A. E., Haemers W. H. Spectra of Graphs. New York: Springer, 2012. 250 p.
5. Грушка Д., Лебідь В. Побудова коспектральних графів відносно узагальненої матриці суміжності. *Могилянський математичний журнал*. 2018. Т. 1. С. 11–14.
6. Соколов В., Соломко В. Побудова пари коспектральних 5-регулярних графів, один з яких має досконале парування, а інший – ні. *Могилянський математичний журнал*. 2021. Т. 4. С. 24–27.
7. Hołubowski W., Oliynyk B., Solomko V. On the Characterization of the Unitary Cayley Graphs of the Upper Triangular Matrix Rings. *Symmetry*. 2024. Vol. 17(12), 2180.
8. Chung F. R. K., Lu L. Complex graphs and networks. Providence, RI: American Mathematical Society, 2006. 264 p.
9. Erciyes K. Algebraic graph algorithms. Cham: Springer, 2021. 300 p.
10. Anderson E., Bai Z., Bischof C. et al. LAPACK Users' Guide. 3rd ed. Philadelphia: SIAM, 1999. 407 p.
11. Lehoucq R. B., Sorensen D. C., Yang C. ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods. Philadelphia: SIAM, 1998. 142 p.
12. Virtanen P., Gommers R., Oliphant T. E. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. 2020. Vol. 17. P. 261–272.
13. Abramov, V., Astafieva, M., Boiko, M., Bodnenko, D., Bushma, A., Vember, V., Hlushak, O., Zhylytsov, O., Ilich, L., Kobets, N., Kovaliuk, T., Kuchakovska, H., Lytvyn, O., Lytvyn, P., Mashkina, I., Morze, N., Nosenko, T., Proshkin, V., Radchenko, S., & Yaskevych, V. (2021). Theoretical and practical aspects of the use of mathematical methods and information technology in education and science. <https://doi.org/10.28925/9720213284km>