

DOI: 10.63527/1607-8829-2025-3-71-91

Ланде Д.^{1,2} (<https://orcid.org/0000-0003-3945-1178>),
Снарський А.^{1,2} (<https://orcid.org/0000-0002-4468-4542>),
Манко Д.² (<https://orcid.org/0000-0003-1848-2952>),
Лінчевський І.¹ (<https://orcid.org/0000-0002-2896-9580>),
Федотов В.¹ (<https://orcid.org/0000-0001-7821-0989>)

¹Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”;
²Інститут проблем реєстрації інформації НАН України
Автор-кореспондент: Снарський А., e-mail: asnarskii@gmail.com

Мережевий аналіз співавторства та напрямки досліджень у термоелектриці

У даній статті представлено методологію побудови розширеної мережі співавторства в галузі термоелектрики шляхом інтеграції прямих зв'язків співавторства з тематичною близькістю, отриманою з профілів дослідників у Google Scholar. Мережа будується шляхом цілеспрямованого зондування з використанням релевантних ключових слів (наприклад, «термоелектрика»), а ваги зв'язків визначаються як кількістю спільних публікацій, так і спільними дослідницькими інтересами. Такий підхід дозволяє ідентифікувати не лише безпосередніх співробітників, а й потенційних міждисциплінарних партнерів, розкриваючи приховану структуру наукової спільноти. Крім того, мережа дескрипторів будується на основі спільної появи ключових слів у профілях авторів, формуючи онтологічну карту галузі. Тематичні кластери, що представляють основні напрямки досліджень, ідентифікуються за допомогою алгоритму Лувена. Вперше в цьому контексті для інтерпретації вмісту кластера використовуються моделі великих мов (LLM) шляхом створення змістовних, зрозумілих для людини міток зі списків дескрипторів. Це дозволяє швидко, об'єктивно та масштабовано ідентифікувати наукові тенденції без залежності від експертних анотацій. Аналіз показує, що розширена мережа демонструє вищу щільність, ніж традиційна мережа співавторства, що підкреслює важливість тематичних зв'язків. Показники центральності (ступінь та проміжність) допомагають визначити ключових учасників та структурні зв'язки в межах галузі. Запропонований підхід підтримує аналіз наукових спільнот, виявлення дослідницьких шкіл та прогнозування співпраці.

Цитування: Ланде Д., Снарський А., Манко Д., Лінчевський І., Федотов В. (2025). Мережевий аналіз співавторства та напрямки досліджень у термоелектриці. Термоелектрика, (3), 72–93. <https://doi.org/10.63527/1607-8829-2025-3-71-91>

Ключові слова: узагальнена мережа співавторства, предметна область, LLM, наукометричний сервіс, мережеве зондування, тематичні дескриптори, термоелектрика.

1. Вступ

В результаті розвитку наукових інформаційних систем з'явилися нові можливості для оцінки рівня вчених, наукових шкіл та вивчення закономірностей наукової взаємодії [1]. Наразі актуальним є завдання вибору експертних груп та прогнозування [2] співпраці вчених у різних галузях, зокрема в термоелектриці. Розглядаючи взаємозв'язок між спільними науковими інтересами різних вчених та їх співавторством, можна формувати мережі, які можна використовувати для вирішення цього завдання.

У галузі теорії складних мереж вивчаються структури, властивості та еволюція мереж, що складаються з численних взаємодіючих елементів – вузлів та ребер. Вивчення складних мереж – це широка галузь, яка охоплює як вивчення білкових структур, так і телекомунікаційні мережі та Інтернет. Серед іншого, ця теорія може допомогти виміряти науковий потенціал, визначити наукових лідерів та школи, оскільки мережевий аналіз дозволяє візуалізувати та аналізувати зв'язки між вченими та їхніми роботами, а також зрозуміти, як ці зв'язки формують структуру наукової спільноти. Зокрема, теорія складних мереж може допомогти в таких аспектах:

- вимірювання наукового потенціалу: мережевий аналіз може показати, які вчені та наукові групи є найактивнішими та найвпливовішими в певній науковій галузі;
- визначення лідерів: аналіз мережі співавторства може допомогти визначити лідерів у науковій спільноті; вчені, які є найбільш впливовими, можуть мати велику кількість співавторів і часто цитуватися в публікаціях;
- вивчення наукових шкіл: мережевий аналіз також може допомогти у вивченні наукових шкіл та визначенні їхнього впливу в науковій спільноті.

Наразі в галузі дослідження складних мереж працює безліч науковців, зокрема найвпливовіші та інші. Серед них Альберт-Ласло Барабаші (Albert-László Barabási) [3–6], який досліджує загальні питання теорії складних мереж як науки, а також питання впливу та взаємодії в складних мережах. Дункан Воттс (Duncan Watts) [7–9] у своїх роботах також обговорює «нову» науку про мережі, моделі мереж малого світу та їх властивості. Марк Ньюмен (Mark Newman) [10–12] зосереджується на фізиці складних мереж, їх структурі та функціональності. Стівен Строгац (Steven Strogatz) [13–15] працює в галузі складних систем, динаміки мереж та їх застосування в біології та соціології.

Мережевий аналіз може бути корисним інструментом у наукометрії, зокрема для аналізу зв'язків між вченими та їхніми роботами. Наприклад, мережі можна використовувати для аналізу співавторства між вченими та університетами. Мережі цитування можуть допомогти дослідникам визначити найчастіше цитовані роботи та вчених у певній галузі, а також виявити нові тенденції в науці та техніці.

Ця стаття присвячена задачі побудови та вивчення мереж співавторства. Такі мережі будуються з урахуванням спільних публікацій різних авторів. Кожен вузол у мережі представляє автора, а ребра між вузлами показують наявність спільних робіт. Для побудови мереж співавторства використовуються дані про спільні публікації, які можна зібрати з різних джерел, таких як наукові бази даних (наприклад, Google Scholar, Scopus), монографії, журнали, конференції.

Такі мережі можна використовувати для:

- вивчення структури наукових спільнот, визначення ключових гравців та виявлення найпродуктивніших авторів і наукових груп;
- визначення наукового впливу авторів (індекси співпраці можна використовувати для визначення того, які автори є найвпливовішими в певній галузі);
- пошука потенційних партнерів для наукової співпраці;
- аналізу наукових тенденцій, відстеження популярності різних тем у певній галузі та визначення груп авторів з науковими інтересами та популярними темами в межах цих груп;
- вивчення динаміки наукових спільнот, змін у структурах спільнот та появи нових гравців у певних галузях.

Мережі співавторства пов'язані з мережевими моделями предметних областей, оскільки вони визначаються науковою взаємодією між вченими. На основі мереж співавторства можна ідентифікувати групи вчених, які працюють у певній галузі зі схожими інтересами. Поєднання мереж співавторства та мережових моделей предметних областей може допомогти в розумінні наукових дисциплін та їх розвитку, а також в оцінці наукової продуктивності вчених.

Мережі співавторства досліджувалися фахівцями з прикладної математики, наукометрії, соціології. Наразі існує окремий напрямок під назвою Scholarly Data Mining [16], який присвячений поглибленому аналізу наукової комунікації, включаючи відносини співавторства. У роботі [17] розглядаються проблеми, що виникають на різних етапах аналізу співпраці, пов'язані зі збором даних, встановленням меж мережі, визначенням реляційної матриці даних, аналізом даних та інтерпретацією результатів. Гаутам Ахуджа (Gautam Ahuja) досліджував зв'язок між структурою мереж співавторства та інноваціями в організаціях у роботі [18]. М.Е.Дж. Ньюмен (M.E.J. Newman) заклав основу для сучасного дослідження мереж співавторства в роботі [19], а в статті [20] досліджувалися підходи до побудови термінологічних мереж як онтологічної моделі предметної області, зокрема пропонувалися нові правила визначення синтаксичних та семантичних зв'язків між термінами в тексті, а також напрямків цих зв'язків у неорієнтованих та спрямованих мережах, побудованих на основі тематичного текстового корпусу. У статті [2] досліджувалися питання, пов'язані з прогнозуванням потенційної наукової співпраці між вченими на основі аналізу наукових публікацій.

Так звані мережі співавторства вже стали традиційним інструментом для вивчення закономірностей науково-дослідної співпраці, які дозволяють отримувати не лише

наукометричні оцінки, а й виявляти експертів для вирішення складних завдань. Найбільші науково-інформаційні сервіси дозволяють дослідникам створювати свої профілі, що містять відповідну наукометричну інформацію. Значна кількість досліджень присвячена вивченню мереж співавторства, а також наукометричний сервіс Google Scholar (<http://scholar.google.com/citations>), що підтверджує актуальність цієї роботи.

Багато сучасних наукометричних сервісів базуються на методах формування мереж цитування, співавторства, визначення значущих вузлів, структури мережі та вивчення відповідних колекцій документів. Зокрема, у роботі [22] представлено метод оцінки важливості вузлів у мережі співавторства на основі вдосконаленого алгоритму PageRank. У статті також запропоновано схему оцінки внеску кожного автора в роботу. У роботі [23] аналізується мережа співавторства з метою пошуку міждисциплінарних наукових спільнот, тоді як у роботі [24] досліджується мережа тематичних потоків – Topic Flow Network (TFN), яка будується з використанням інформації про кожного автора та анотації статті.

Метою цієї роботи є представлення нового підходу до побудови мережі зв'язків між науковцями шляхом цілеспрямованого зондування доступних наукометричних сервісів, формування та подальшого дослідження узагальненої мережі співпраці науковців з урахуванням зв'язків їхнього співавторства та змістовних кореляцій їхніх дослідницьких напрямків.

Термін «зондування» стосується вибору невеликої вибірки контенту з великих мереж, які неможливо повністю просканувати з технологічних причин [25]. Багато сучасних досліджень наукометричних мереж використовують механізми їх зондування, після чого робляться висновки про топологію таких мереж. Однак у [26] було показано, що цей підхід може бути недосконалим. Зображення наукометричних мереж, отримані в результаті моніторингу, можуть суттєво відрізнитися та лише частково відобразити властивості таких мереж. Це може статися тому, що властивості цих зображень сильно залежать від алгоритмів, що використовуються для моніторингу. Отже, зондування мережі повинно задовольняти умову повноти вилучення інформації про об'єкти певного типу. У цьому дослідженні мінімальна кількість цитувань авторів наукових публікацій слугує обмеженням для зондування. Вчені з кількістю цитувань нижче певного порогу не розглядаються. Таким чином, фактично виконується повне сканування для визначеного набору дескрипторів та цього параметра набору вузлів, і відповідно розглядається мережа, утворена ними.

Також, очевидно, мережа співпраці може стати досить великою, якщо вона не обмежується певною темою. У нашому дослідженні темою є термoeлектрика. Відхилення від цієї теми та вивчення співавторів, які мають до неї лише опосередковане відношення, ускладнює сприйняття сформованої мережі та призводить до ефекту «дрейфу теми». Для подолання цього ефекту використовується фільтрація тем, тобто дескриптори, що присвоюються авторам у наукометричній мережі, використовуються для визначення їхньої тематичної орієнтації. Таким чином, розмір та топологія узагальненої мережі співавторства залежать від тематичної орієнтації (формально

визначеної тегами-дескрипторами, що присвоюються вченим, та граничним значенням цитування). Слід зазначити, що виявлення кластерів у таких мережах можна розглядати як основу для подальшої ідентифікації наукових шкіл, експертних груп. У цьому випадку під науковою школою розуміється творчий колектив дослідників, об'єднаний спільною галуззю досліджень та має визнаного лідера.

2. Математичні позначення

Розглянемо формально умови задачі, а саме, нехай A – набір авторів, A_i – автор з індексом i , а P_i – профіль автора A_i . Нехай D позначає набір всіх існуючих дескрипторів. Нас цікавлять дескриптори, що входять до профілю автора. Для спрощення будемо вважати, що профіль – це набір дескрипторів P_i , а $d_j \in D$ – дескриптор з індексом j . Нехай \hat{d}_j^i позначає індикатор присутності дескриптора з індексом j у автора з індексом i :

$$\hat{d}_j^i = \begin{cases} 1, & d_j \in P_i, \\ 0, & d_j \notin P_i. \end{cases} \quad (1)$$

Вектор $\bar{A}_i = (\hat{d}_1^i, \hat{d}_2^i, \dots, \hat{d}_{|D|}^i)$ присвоюється автору з індексом i .

Скалярний добуток відповідних векторів будемо розглядати як тематичну близькість інтересів авторів з індексами i та k :

$$Sim(A_i, A_k) = (\bar{A}_i, \bar{A}_k). \quad (2)$$

Зв'язок співавторства між авторами з індексами i та k позначається як $Co(A_i, A_k) \in \{0,1\}$.

Відповідно, у цих позначеннях зв'язок в узагальненій мережі співавторства вчених між авторами з індексами i та k дорівнює

$$Link(A_i, A_k) = Sim(A_i, A_k) + C \cdot Co(A_i, A_k), \quad (3)$$

де C – константа, яку обирають експерти.

Набір всіх можливих значень $Co(A_i, A_k)$ утворює матрицю простого співавторства. Таким чином, матриця, що відповідає мережі вчених, є комбінацією мережі тематичних інтересів та мережі співавторства. В результаті матриця мережі вчених є щільнішою.

3. Алгоритм

Алгоритм зондування довідкової мережі наукометричного інформаційного сервісу та подальшого формування мережі вчених адаптовано до реальної мережі співавторів сервісу (у цьому випадку Google Scholar) наступним чином (Рисунок 1):

Крок 1: Дескриптори (ключові слова або теги, зазначені в профілях авторів) вибираються як основа для зондування (спочатку вибирається один дескриптор, у нашому випадку очевидний – «термоелектрика», рисунок 2).

Крок 2. За допомогою інструментів наукометричного сервісу відбираються всі автори, які присвоїли собі обраний дескриптор/дескриптори (теги). В результаті цього відбору автори сортуються у порядку спадання на основі кількості цитувань. Для побудови мережі методом зондування розглядаються автори зі значеннями цитувань не менше заданого порогу τ (наприклад, $\tau = 5000$).

Крок 3. Переглядається список дескрипторів, призначених авторам, визначеним на кроці 2. Вибираються дескриптори, що відповідають основній темі. Цей процес може бути виконаний експертом або автоматично, наприклад, у цьому випадку, за наявності фрагмента слова «термоел». Зокрема, у цьому випадку сторінки авторів під першим дескриптором містять дескриптори, що відповідають основній темі, такі як термоелектричний, термоелектрика.

Крок 4. Для кожного з авторів, визначених на кроці 2, розглядаються їхні співавтори зі значеннями цитування не меншими за поріг τ . В якості вузлів мережі розглядаються лише ті вчені, чий дескриптор близький до основної теми термоелектрики. Ці автори також розглядаються як вузли майбутньої мережі вчених. Також розглядаються відповідні дескриптори, серед яких знайдено такі дескриптори, як термоелектричний, термоелектрика, термоелектричні матеріали, топологічна термоелектрика, фототермоелектричний зв'язок, термоелектричні властивості, термоелектричні матеріали, термоелектричний генератор, термоелектричні пристрої, іонний термоелектрик тощо.

Крок 5. Для всіх вибраних дескрипторів вибираються автори, які призначили собі ці дескриптори. Якщо список авторів зі значеннями цитувань, більшими за τ для всіх вибраних дескрипторів, вичерпано, процес завершується. В іншому випадку алгоритм повертається до кроку 2.

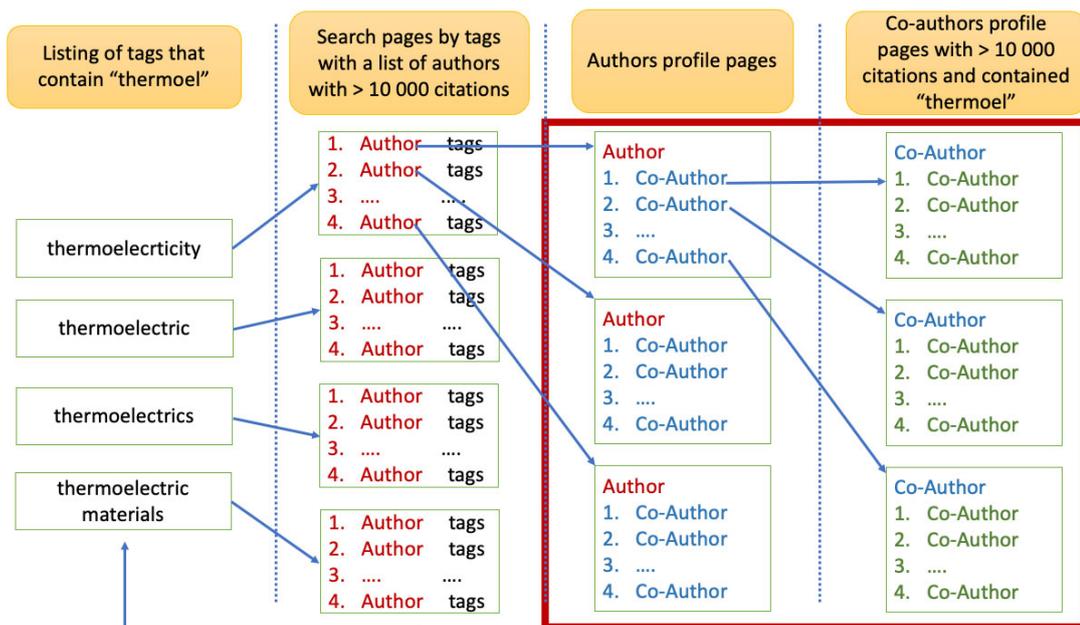


Рис. 1. Розширений алгоритм розпізнавання послуг цитування Google Scholar

Список тегів, що містять «thermoel» Пошук сторінок за тегами зі списком авторів з більш ніж 10 000 цитатами Сторінки профілів авторів Сторінки профілів співавторів з більш ніж 10 000 цитатами, що містили «thermoel»

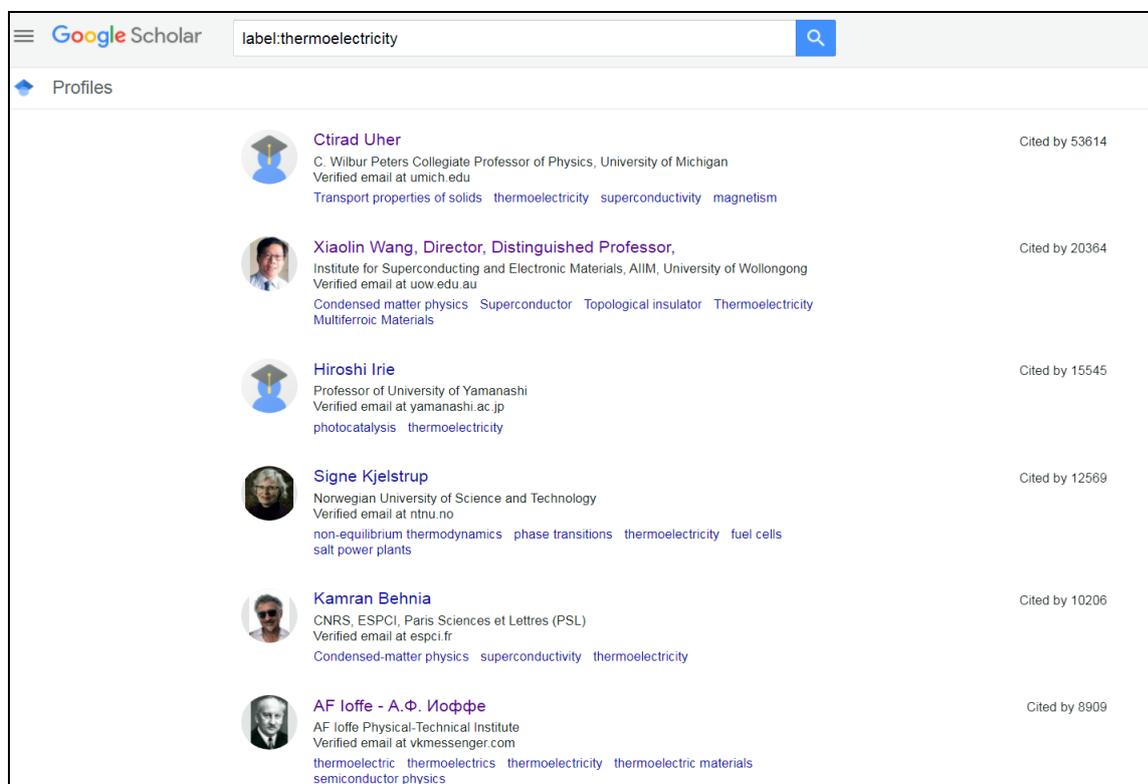


Рис. 2. Фрагмент результатів пошуку для дескриптора “thermoelectricity”. Цитовано 53614...

Очевидно, що представлений алгоритм сходиться, оскільки кількість науковців, охоплених наукометричним сервісом, обмежена.

Вагове значення зв'язків між авторськими вузлами в мережі відповідає кількості спільних дескрипторів, призначених їм. Крім того, якщо між авторами є співавторство, до ваги відповідного зв'язку додається певна константа, див. формулу (3).

Таким чином, представлений алгоритм дозволяє сформувати матрицю узагальненого співавторства. Також можливо розглянути просту матрицю співавторства, не враховуючи перший член у формулі (3), тобто не враховуючи зв'язки між дескрипторами. Такі мережі досліджувалися в багатьох роботах і становлять певний інтерес. Властивості такої мережі будуть розглянуті нижче.

4. Мережа дескрипторів як модель предметної області

Як проміжний результат, свого роду подвійну допоміжну мережу, що формується під час побудови мережі співавторства, можна розглядати як мережу, де вузлами є дескриптори. Зв'язки в такій мережі можна визначити, наприклад, наступним чином. Нехай A_j – набір дескрипторів для автора i . Якщо \hat{a}_j^i – індикатор наявності дескриптора j для автора i , то вага зв'язку між дескрипторами з індексами j та k дорівнює $Link(A_j, A_k)$

$$Link(A_i, A_k) = \begin{cases} 1, \exists i: \hat{d}_j^i \in A_i, \hat{d}_k^i \in A_i, j \neq k, \\ 0, otherwise. \end{cases} \quad (4)$$

Ця мережа також може мати зважені ребра $Link(A_j, A_k)$, якщо розглядати їх як кількість спільних входжень дескрипторів з індексами j та k у наборах дескрипторів різних авторів.

Слід зазначити, що дескриптори, які не пов'язані безпосередньо з термоелектрикою, також можуть бути включені до цієї сформованої мережі як вузли. Якщо такі дескриптори мають низьку вагу, їх можна ігнорувати, але якщо вони мають значну вагу, слід дослідити їхні глибокі зв'язки з основною темою, оскільки вони можуть визначати нові напрямки в рамках цієї теми.

Очевидно, що цю мережу можна кластеризувати за допомогою методів кластерного аналізу, де окремі кластери відповідають різним підтемам у межах основної теми. Таким чином, сформовану мережу можна розглядати як модель предметної області.

На рисунку 3 показано центральний фрагмент сформованої мережі дескрипторів, що відповідають темі термоелектрики. У таблиці 1 наведено 30 найпопулярніших дескрипторів, пов'язаних з термоелектрикою, з найвищими ступенями в даній мережі. Розподіл ступенів вузлів у такій мережі відповідає степеневому закону, а його графік у логарифмічному масштабі представлено на рисунку 4 (де абсциса відповідає номерам дескрипторів, відсортованим за частотою, а ордината відповідає ступеням відповідних вузлів у мережі).

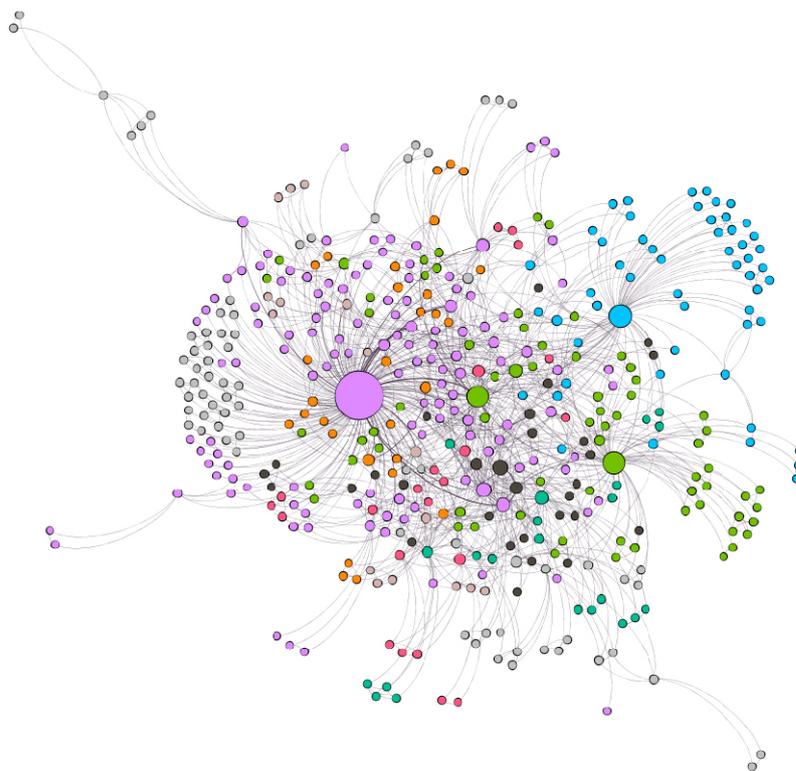


Рис. 3. Фрагмент мережі дескрипторів

Таблиця 1

Дескриптори з найбільшими ступенями

Ранг	Дескриптор	Ступінь вузла
1	thermoelectrics (термоелектрика)	211
2	thermoelectric (термоелектричний)	74
3	thermoelectric materials (термоелектричні матеріали)	72
4	materials science (матеріалознавство)	69
5	thermoelectricity (термоелектрика)_	40
6	condensed matter physics (фізика конденсованого стану)	35
7	magnetism (магнетизм)	33
8	energy materials (енергетичні матеріали)	27
9	heat transfer (теплопередача)	27
10	nanomaterials (наноматеріали)	26
11	superconductivity(надпровідність)	24
12	energy storage (накопичення енергії)	22
13	photovoltaics (фотоелектричні системи)	20
14	nanotechnology (нанотехнологія)	19
15	thin films (тонкі плівки)	18
16	organic electronics (органічна електроніка)	18
17	optoelectronics (оптоелектроніка)	17
18	batteries (батареї)	17
19	computational materials science (обчислювальне матеріалознавство)	17
20	solid state chemistry (хімія твердого тіла)	17
21	solar cells (сонячні елементи)	16
22	graphene (графен)	16
23	quantum materials (квантові матеріали)	15
24	chemistry (хімія)	15
25	spintronics (спінтроніка)	14
26	material science (матеріалознавство)	14
27	materials physics (фізика матеріалів)	14
28	catalysis (каталіз)	14
29	nanowires (нанодроти)	14
30	fuel cells (паливні елементи)	13

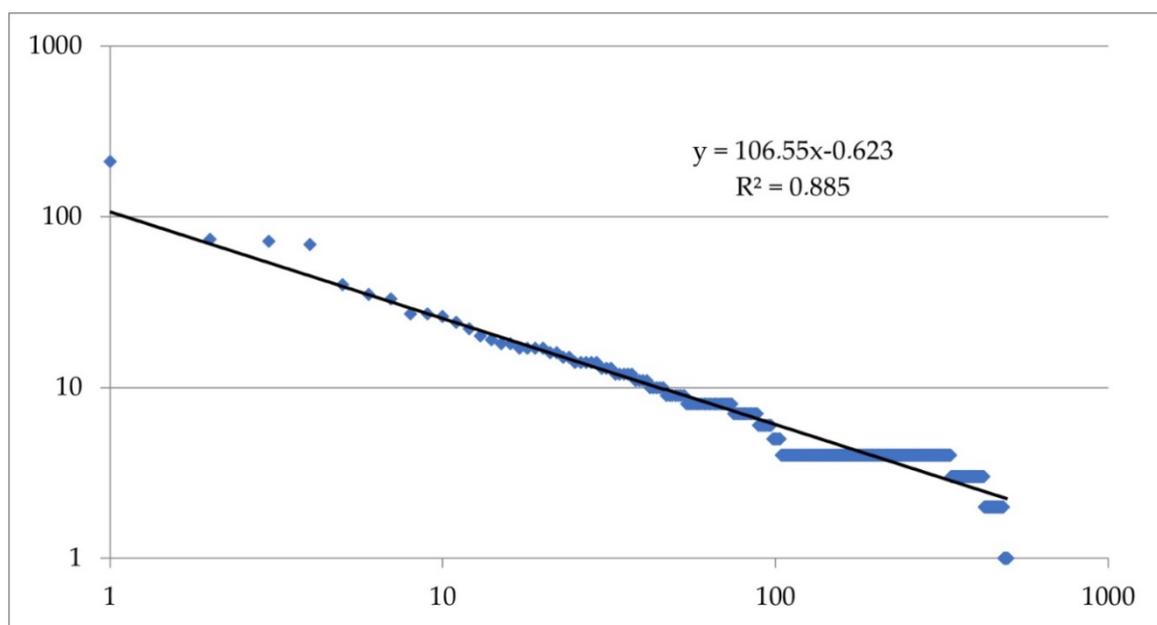


Рис. 4. Розподіл ранжування ступенів вузлів дескрипторної мережі

Мережу було кластеризовано за допомогою алгоритму модульності (алгоритм Лувена [26]), який базується на максимальній модульності – міри, що визначає, настільки щільно пов'язані вузли в кластері порівняно з випадковою мережею.

Формальне визначення міри модульності задається формулою:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \gamma \frac{k_i k_j}{2m} \right] \delta(c_i, c_j), \quad (5)$$

де A_{ij} – елемент матриці суміжності графа (1 якщо між вершинами i і j є ребро, в іншому випадку 0); k_i, k_j – ступені вершин i і j (кількість ребер, інцидентних кожній вершині); m – загальна кількість ребер у графі; c_i, c_j – спільноти, до яких належать вершини i і j ; $\delta(c_i, c_j)$ – функція, що дорівнює 1, якщо $c_i = c_j$, і 0 в іншому випадку; γ – параметр роздільної здатності (гамма), який контролює вагу розподілу випадкових з'єднань. Зазвичай $\gamma > 0$

У нашому випадку було використано значення, яке дозволило ідентифікувати кілька великих кластерів у межах термінологічної мережі. Аналіз цих кластерів було виконано за допомогою моделі великих мов (LLM) [27] за такою командою:

Було проведено кластерний аналіз концепцій у галузі термоелектрики. Нижче наведено список концепцій, що належать до одного з кластерів. Як би ви назвали цей кластер? Ось концепції з нього: термоелектричні матеріали, фізика конденсованих середовищ, накопичення енергії, фотоелектричні елементи, нанотехнології тощо.

Виконання аналогічних підказок дозволило визначити такі основні напрямки дослідження для визначених класів (кластерів) модульності, що узгоджуються з експертними оцінками авторів:

1. Матеріалознавство та фундаментальні дослідження в термоелектриці: інтеграція нанотехнологій, енергетики та передових матеріалів.
2. Наноматеріали та інноваційні технології: фотокаталіз, енергетичні застосування та вирощування кристалів.
3. Енергетичні матеріали та обчислювальне матеріалознавство: теорія конденсованих середовищ, наноматеріали та машинне навчання.
4. Електрохімічні системи та нові функціональні матеріали: батареї, каталізатори та енергетичні технології.
5. Термічні процеси та нанотехнології: теплопередача, термоелектрика та квантові матеріали.
6. Термоелектрика та нанотехнології: енергетичні матеріали, фотоніка та інноваційні застосування.
7. Матеріалознавство та фундаментальні дослідження: термоелектрика, магнетизм та передові нанотехнології.
8. Органічна електроніка та нанотехнології: сонячна енергія, обчислювальне моделювання та застосування.

5. Проста мережа співавторства

Мережа співавторства становить значний інтерес, оскільки дозволяє ідентифікувати наукові групи та школи. На рисунку 5 показано фрагмент мережі вчених у галузі термоелектрики, побудованої за допомогою алгоритму з порогом цитування $\tau = 5000$ та набором дескрипторів. Як бачимо, мережа вчених, яка містить 278 вузлів, має високу зв'язність, 1 зв'язний компонент та чітко визначені кластери, які були ідентифіковані за класами модульності в середовищі програмного забезпечення Gephi (Gephi.org) [27].

Як видно з рисунка, автори згруповані в кілька щільних підграфів, кожен з яких близький до повного графа. Зв'язки між групами розріджені. Окрім кількох великих груп, існує велика кількість дуже малих груп. Зверніть увагу, що наявність одного вузла (автора), не пов'язаного з іншими, не обов'язково означає, що автор не має співавторів – можливо, кількість цитувань їхніх співавторів нижча за встановлене нами порогове значення (5000), тому вони не включені до мережі. Включення таких співавторів призведе до появи нових зв'язків, але поділ мережі на основні групи (школи) залишиться незмінним. Зрештою, слід зазначити, що поділ на групи базувався на критерії модульності – мірі структури мереж, яка використовується для їх поділу на групи (модулі, спільноти, кліки) зі значно щільнішими зв'язками всередині них, ніж між окремими групами.

Мережа співавторства дозволяє аналізувати окремі вузли (авторів), включаючи визначення їхнього ранжованого розподілу. Нижче, у таблиці 2, представлено перших

30 авторів у мережі співавторства, відсортованих за ступенем (кількість посилань, що відповідають вузлу автора) та за проміжністю, характеристикою центральності вузла, що визначається кількістю шляхів між будь-якими двома вузлами, що проходять через даний вузол відносно всіх можливих шляхів. Як видно з таблиці, список перших 30 авторів, відсортованих за ступенем та проміжністю, збігається лише частково. Це пояснюється тим, що на перший впливає кількість публікацій (враховується у всій мережі), тоді як другий враховує такі шляхи, як «співавтор мого співавтора» (непрямі зв'язки між авторами). На рисунку 6 показано ранжований розподіл вузлів у простій мережі співавторства, відсортований за ступенем. Як видно, цей розподіл відповідає логарифмічній функції.

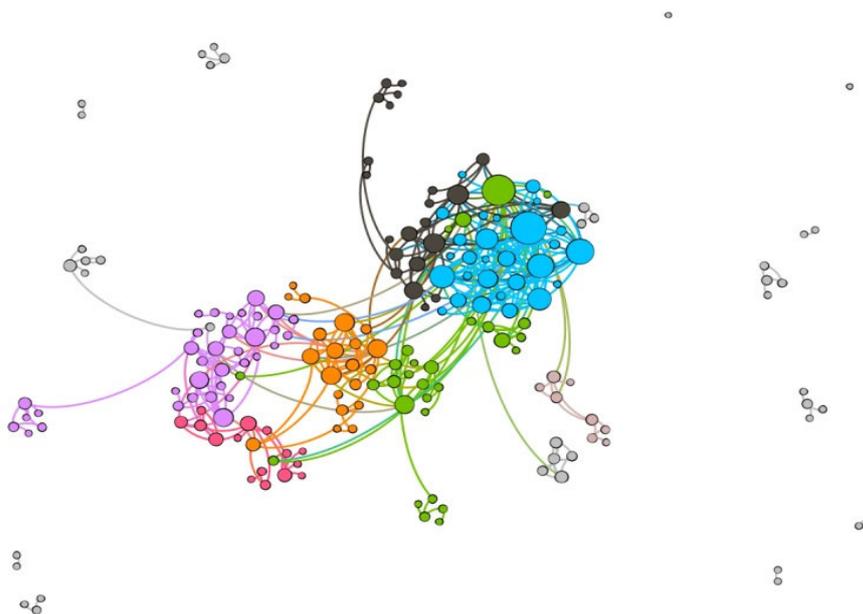


Рис. 5. Фрагмент простої мережі співавторства

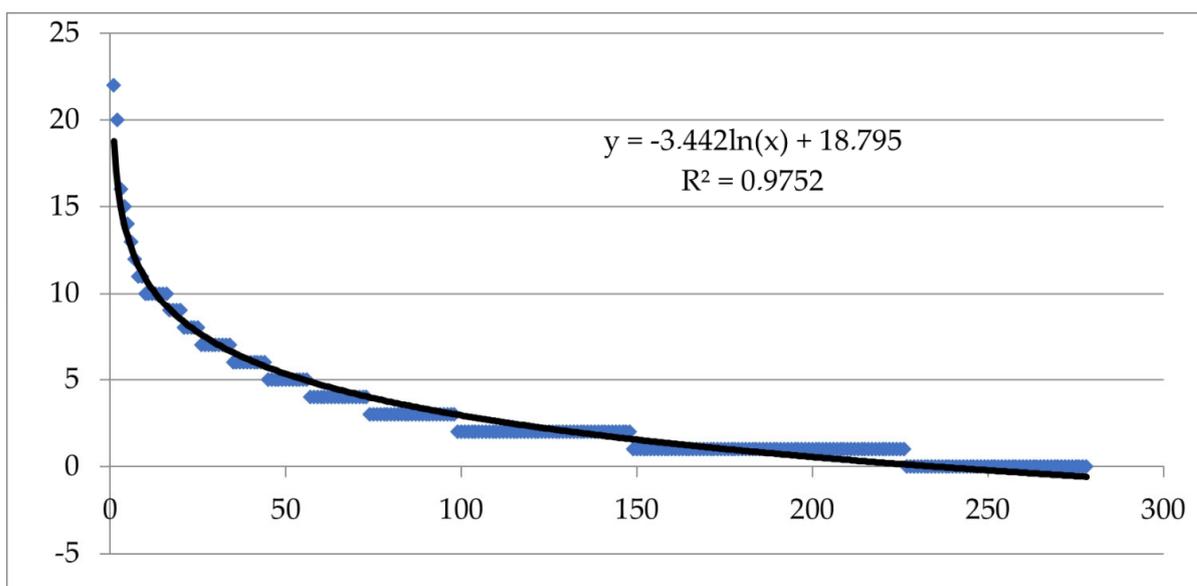


Рис. 6. Розподіл рангів авторів простої мережі співавторства за рівнем вузлів

Таблиця 2

Рейтинг автора простої мережі співавторства

Прізвище	Ступінь вузла	Прізвище	Проміжність вузла
Kanatzidis	22	G. Jeffrey Snyder	0.1128
G. Jeffrey Snyder	20	Kanatzidis	0.0660
Gangjian Tan (谭刚健)	16	Weishu liu	0.0548
Ctirad Uher	15	David J. Singh	0.0479
Li-Dong Zhao	14	Li-Dong Zhao	0.0374
Jiaqing HE	13	Wang Heng	0.0363
Chris Wolverton	12	Gang Chen	0.0363
Terry Tritt	11	Keivan Esfarjani	0.0359
Xianli Su (苏贤礼)	11	Ronggui Yang 杨荣贵	0.0357
Zhifeng Ren	10	Li Shi	0.0354
David J. Singh	10	Zachary M. Gibbs	0.0351
Zachary M. Gibbs	10	Olivier Delaire	0.0347
Gang Chen	10	Daryoosh Vashae	0.0325
Brian Sales	10	Mildred S. Dresselhaus	0.0317
Andrew F. May	10	Zhifeng Ren	0.0289
Timothy Hogan	10	Andrew F. May	0.0265
Jian He	9	Ali Shakouri	0.0257
Xinfeng Tang (唐新峰)	9	Terry Tritt	0.0250
Kanishka Biswas	9	S. Joseph Poon	0.0230
Keivan Esfarjani	9	Marisol Martin-Gonzalez	0.0218
Olivier Delaire	8	Georgy Samsonidze	0.0216
Vinayak Dravid	8	Eric Toberer	0.0207
Di Wu	8	A Majumdar	0.0194
Jie Ma	8	Chris Wolverton	0.0191
Zihang Liu 刘紫航	8	Gangjian Tan (谭刚健)	0.0187
A Majumdar	7	Jie Ma	0.0181
Wang Heng	7	Boris Kozinsky	0.0175
Jihui Yang	7	anke weidenkaff	0.0174
Shanyu Wang (王善禹)	7	Maria Ibáñez	0.0174
David N Seidman	7	David R G Mitchell	0.0174

6. Узагальнена мережа співавторства

На рисунку 7 показано мережу, сформовану за описаним вище алгоритмом на основі критеріїв співавторства та тематичної близькості для одного й того ж набору авторів та зв'язків, розрахованих за виразом (3).

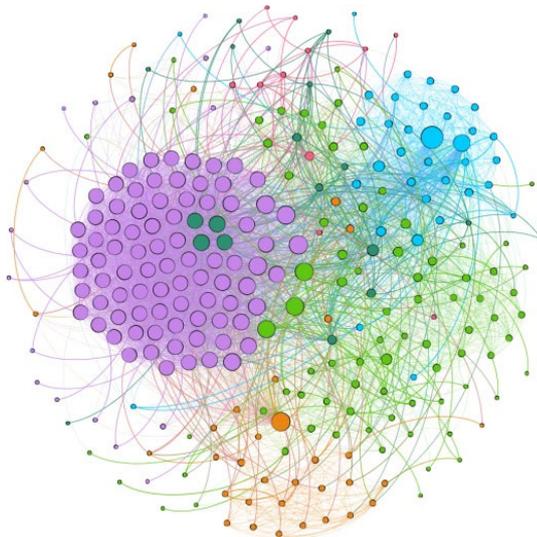


Рис. 6. Узагальнена мережа співавторства

Як видно, кластери в цій мережі не так чітко визначені, але є очевидна «зона ядра» – практично повністю зв'язний підграф, а також численні «листя» – поява яких пояснюється другим компонентом згаданої вище формули. Очевидно, що ця мережа набагато щільніша порівняно з простою мережею співавторства, але метод кластеризації на основі модульності дозволяє розділити її на окремі групи (які візуально менш очевидні).

Водночас, розподіл рангів ступенів авторів також якісно змінюється. Графік (Рисунок 7) чітко показує поділ авторів на частини – з великими рангами (від 1 до приблизно 100, Рисунок 8) та меншими (від приблизно 100, Рисунок 9). Нанесені лінії тренду показують, що ці дві частини мають принципово різний тип тренду. Для рангів на початку шкали розподіл має степеневий закон:

$$S(A \leq 97) \approx 137 \cdot A^{-0.1} \quad (6)$$

а для більших значень рангів

$$S(A \geq 98) \approx -45 \cdot \ln A + 251 \quad (7)$$

Якісно отриманий поділ рангового розподілу ступенів може бути пов'язаний з тим, що ми маємо справу з перекриттям двох мереж – мережі співавторства та мережі кореляцій між дескрипторами. Як було показано раніше, проста мережа співавторства також мала логарифмічний розподіл рангів.

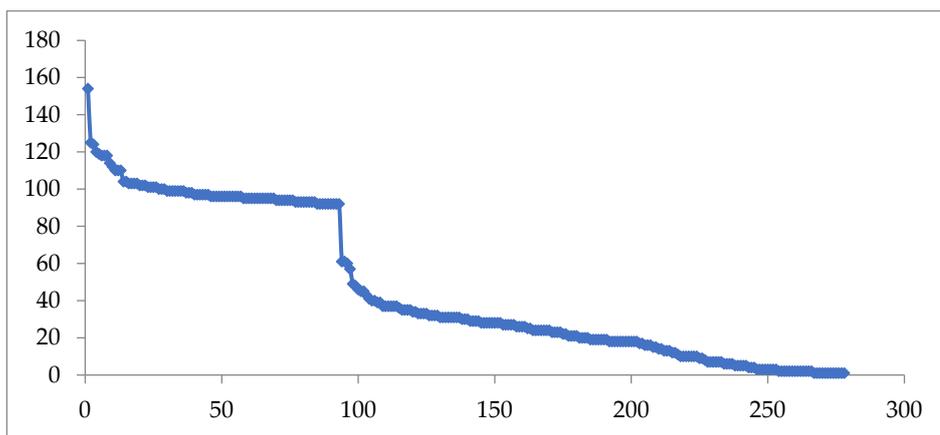


Рис. 7. Ранжований розподіл авторів за ступенями для всіх авторів

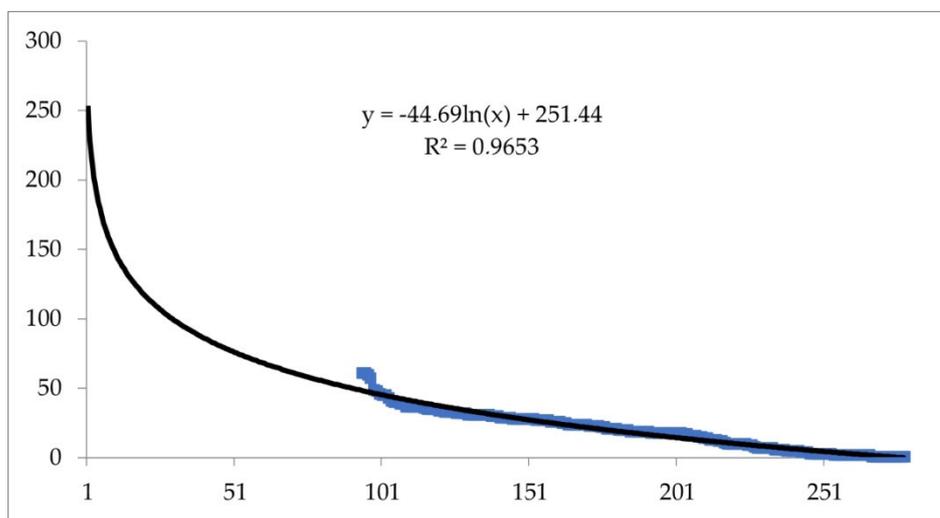


Рис. 8. Ранжований розподіл авторів за ступенями для авторів з рангом від 1 до 97

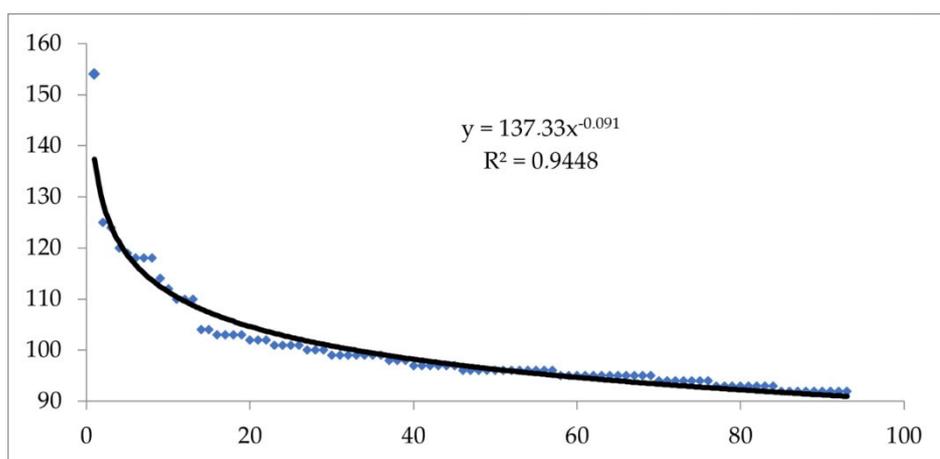


Рис. 9. Ранжований розподіл авторів за ступенями для авторів з рангом від 98 до кінця

У таблиці 3 наведено список авторів із найвищим рейтингом згідно з індексом Гірша (h -index) від Google.

Таблиця 3

Список авторів, ранжованих за індексом Гірша (*h*-індекс)

Прізвище	Індекс Гірша (Google)
Gerbrand Ceder	159
Lidong Chen	107
Brian Sales	97
Jiaqing HE	82
Kornelius Nielsch	74
Shuo Chen (陈硕)	73
Xinfeng Tang (唐新峰)	71
Natalio Mingo	69
Terry Tritt	67
Andreu Cabot	62
Robert A Taylor	61
Armin Feldhoff	59
Kanishka Biswas	58
Eric Toberer	57
Jian He	55
Weishu liu	55
Georgy Samsonidze	51
Subhendra D Mahanti	49
Gerhard Jakob	48
Anders E.C. Palmqvist	47
Choongho Yu	46
Devashi Adroja	46
Nuo YANG	45
Andrew F. May	43
Jorge O. Sofo	42
Gangjian Tan (谭刚健)	40
Wang Heng	38
Graeme R. Blake	37
Matthieu Verstraete	35
Michitaka Ohtaki	34

У таблиці 4 наведено перших 30 авторів узагальної мережі співавторства, відсортованих за рівнями центральності авторів. Як видно з таблиці, список перших 30 авторів, відсортованих за ступенем вузла та проміжністю, збігається лише частково.

Таблиця 4

30 авторів з найвищим рейтингом в узагальненій мережі співавторства

Прізвище	Ступінь вузла	Прізвище	Проміжність вузла
A.F. Ioffe	154	A.F. Ioffe	0.094
Andreu Cabot	125	Michitaka Ohtaki	0.0321
Matthieu Verstraete	124	Brian Sales	0.0217
Weon Ho Shin	120	Terry Tritt	0.0212
Michitaka Ohtaki	119	Kanishka Biswas	0.0209
Andrew F. May	118	Weon Ho Shin	0.0208
Georgy Samsonidze	118	Andreu Cabot	0.0204
Jie Ma	118	Jorge O. Sofo	0.0201
Brian Sales	114	Matthieu Verstraete	0.0184
Jian He	112	Kornelius Nielsch	0.016
Natalio Mingo	110	Andrew F. May	0.0155
Jorge O. Sofo	110	Kamran Behnia	0.0151
Gerhard Jakob	110	Anders E.C. Palmqvist	0.0149
Armin Feldhoff	104	Rachel Segalman	0.0143
Graeme R. Blake	104	Gerhard Jakob	0.0142
Jiaqing HE	103	Armin Feldhoff	0.0142
Kanishka Biswas	103	Boris Kozinsky	0.0141
Boris Kozinsky	103	Marisol Martin-Gonzalez	0.0141
Subhendra D Mahanti	103	Jiaqing HE	0.0135
Shuo Chen (陈硕)	102	G. Jeffrey Snyder	0.013
Robert A Taylor	102	Jian He	0.0127
Anders E.C. Palmqvist	102	Nuo YANG	0.0126
Eric Toberer	101	Devashi Adroja	0.0123
Di Wu	101	Georgy Samsonidze	0.0122
Choongho Yu	101	Rama Venkatasubramanian	0.01178
Nuo YANG	101	Jie Ma	0.0115
Zachary M. Gibbs	100	Boris Kozinsky	0.0115
Gerbrand Ceder	100	anke weidenkaff	0.0112
Weishu liu	100	Maria Ibáñez	0.0111
Xinfeng Tang (唐新峰)	99	David R G Mitchell	0.011

Цікаво дослідити розподіл ступеня центральності авторів на основі їхньої проміжної центральності (рис. 10), яка відповідає експоненціальній функції.

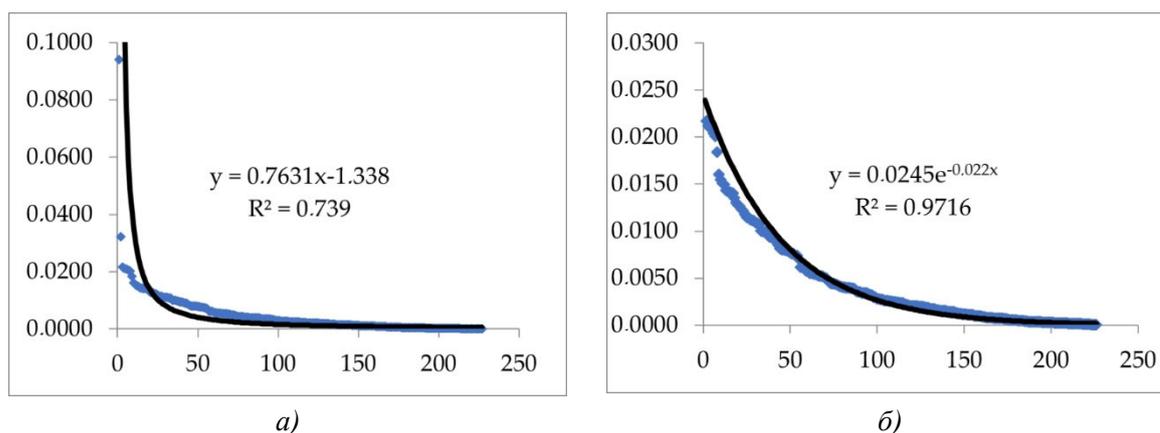


Рис. 10. Розподіл рангів ступенів вузлів за параметром центральності «проміжності»:
а) Усі автори – лінія тренду експоненціальна з показником -0.023 та низькою значущістю $R^2 = 0$, б) Без першого автора – розподіл експоненціальний з майже таким самим показником -0.022 , але з високою значущістю $R^2 = 0.97$

Як видно на рисунку 10а, один з авторів має значно вище значення проміжності порівняно з іншими авторами. Цим автором є один із засновників наукового напрямку та технічних застосувань термоелектрики – А.Ф. Іоффе. Той факт, що тенденція розподілу (чорна плавна лінія на рисунках 10 та 11) набагато краще відповідає розподілу без першого унікального автора, також вказує на його особливе становище. Слід зазначити, що саме проміжність мережі за тегами та співавторством дозволяє визначити найважливішу позицію автора в рамках заданої теми. Його виняткова позиція призвела до підвищення достовірності наших емпіричних даних до 0.97.

7. Висновки

Запропоновано та реалізовано підхід до формування мережі вчених у предметній області термоелектрики. Алгоритм формування цієї мережі обмежений маркерами знань (дескрипторами), які заздалегідь визначені вченими як учасниками наукометричного сервісу у своїх профілях.

Слід зазначити, що запропонована модель автоматичного формування мереж науковців принципово відрізняється від існуючих, які спираються на безпосередню участь експертів у виборі авторів. У запропонованому алгоритмі для побудови узагальненої мережі співавторства використовуються як відносини авторства, так і кореляція змісту дескрипторів, призначених авторам. Таким чином, програма сканування мережі використовує знання, вбудовані самими авторами, що значно розширює експертне середовище.

Як проміжний результат, основні сучасні напрямки досліджень у галузі термоелектрики були визначені за допомогою LLM, на основі змісту дескрипторів та мовної моделі, побудованої для цієї предметної області.

Слід зазначити, що списки вчених, які відповідають найбільшим вузлам двох згаданих мереж, відрізняються. Крім того, можна побачити, що індекси співавторства

вчених, які відповідають найбільшим вузлам, ранжовані за індексом Гірша, не збігаються та більше відповідають досліджуваній предметній області.

Таким чином, запропонована мережа має кілька важливих переваг для аналізу:

- малий діаметр графіка та середня довжина шляху, що може призвести до формування експертних груп науковців, які не є безпосередніми співавторами;
- обмежена кількість кластерів науковців, які чітко відповідають дескрипторам, тобто темам.

Зрештою, врахування не лише одного критерію співавторства збільшує варіативність рішень та дозволяє регулювати баланс між кластеризацією та тематичною подібністю. На додаток до запропонованої мережі, можна розглянути пов'язану мережу, вузлами якої є дескриптори, а зв'язки визначаються кількістю авторів, призначених відповідним парам дескрипторів. Таку мережу можна розглядати як модель основної предметної області. Результати дослідження забезпечують наукову основу для автоматизації та прискорення процесу відбору компетентних експертів для вирішення різних питань у галузі термоелектрики. Хоча модель застосовувалася в рамках сервісу Google Scholar, запропонований підхід також може бути використаний і для інших наукометричних сервісів.

Дослідження було підтримано грантом Національного фонду досліджень України (реєстраційний номер проекту 2023.04/0087).

Інформація про авторів

Ланде Д. – Доктор фізико-математичних наук, професор, завідувач кафедри інформаційної безпеки.

Снарський А. – Доктор фізико-математичних наук, професор кафедри загальної фізики.

Манко Д. – Доктор, старший науковий співробітник з комп'ютерної інженерії.

Лінчевський І. – Доктор фізико-математичних наук, професор кафедри загальної фізики.

Федотов В. – Старший викладач кафедри загальної фізики.

Література

1. Ortega J.L. (2015). How is an academic social site populated? A demographic study of Google Scholar Citations population. *Scientometrics*, 104(1), 1–18. <https://doi.org/10.1007/s11192-015-1593-7>
2. Lande D., Fu M., Guo W., Balagura I., Gorbov I., & Yang H. (2020). Link prediction of scientific collaboration networks based on information retrieval. *World Wide Web*, 23(4), 2239–2257. <https://doi.org/10.1007/s11280-019-00768-9>
3. Barabási A.-L. (216 C.E.). *Network science* (1st edition). Cambridge University Press.
4. Meng X., Varol O. and Barabási A.L. (2024). Hidden citations obscure true impact in science. *PNAS nexus*, 3(5), 155. <https://doi.org/10.1093/pnasnexus/pgae155>

5. Liu Y.-Y., Slotine J.-J., & Barabási A.-L. (2013). Observability of complex systems. *Proceedings of the National Academy of Sciences*, 110(7), 2460–2465. <https://doi.org/10.1073/pnas.1215508110>
6. Barzel B., & Barabási A.-L. (2013). Universality in network dynamics. *Nature Physics*, 9(10), 673–681. <https://doi.org/10.1038/nphys2741>
7. Watts D.J. (1999a). Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology*, 105(2), 493–527. <https://doi.org/10.1086/210318>
8. Watts D.J. (2004). The “New” Science of Networks. *Annual Review of Sociology*, 30(1), 243–270. <https://doi.org/10.1146/annurev.soc.30.020404.104342>
9. Almaatouq A., Alsobay M., Yin M. and Watts D.J., 2024. The effects of group composition and dynamics on collective performance. *Topics in cognitive science*, 16(2), 302–321. <https://doi.org/10.1111/tops.12706>
10. Newman M.E.J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2), 167–256. <https://doi.org/10.1137/S003614450342480>
11. Newman M. (2008). The physics of networks. *Physics Today*, 61(11), 33–38. <https://doi.org/10.1063/1.3027989>
12. Kirkley A., Cantwell G.T., & Newman M.E.J. (2019). Balance in signed networks. *Physical Review E*, 99(1), 012320. <https://doi.org/10.1103/PhysRevE.99.012320>
13. Strogatz S.H. (2001). Exploring complex networks. *Nature*, 410(6825), 268–276. <https://doi.org/10.1038/35065725>
14. Newman M.E.J., Watts D.J., & Strogatz S.H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(suppl_1), 2566–2572. <https://doi.org/10.1073/pnas.012582999>
15. Arbesman S., Strogatz S., & Vitevitch M. (2010). Comparative Analysis of Networks of Phonologically Similar Words in English and Spanish. *Entropy*, 12(3), 327–337. <https://doi.org/10.3390/e12030327>
16. Dridi A., Gaber M.M., Azad R.M.A., & Bhogal J. (2021). Scholarly data mining: A systematic review of its applications. *WIREs Data Mining and Knowledge Discovery*, 11(2). <https://doi.org/10.1002/widm.1395>
17. De Stefano D., Giordano G., & Vitale M.P. (2011). Issues in the analysis of co-authorship networks. *Quality & Quantity*, 45(5), 1091–1107. <https://doi.org/10.1007/s11135-011-9493-2>
18. Ahuja G. (2000). Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. *Administrative Science Quarterly*, 45(3), 425–455. <https://doi.org/10.2307/2667105>
19. Newman M.E.J. (2001). From the Cover: The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404–409. <https://doi.org/10.1073/pnas.021544898>
20. Dmytro Lande, Oleh Dmytrenko, Oksana Radziievska. Determining the Directions of Links in Undirected Networks of Terms. Selected Papers of the XIX International Scientific

- and Practical Conference "Information Technologies and Security" (ITS 2019). *CEUR Workshop Proceedings* (ceur-ws.org). – Vol-2577. – pp 132–145. ISSN 1613-0073.
21. Liu J., Li Y., Ruan Z., Fu G., Chen X., Sadiq R., & Deng Y. (2015). A new method to construct co-author networks. *Physica A: Statistical Mechanics and Its Applications*, 419, 29–39. <https://doi.org/10.1016/j.physa.2014.10.006>
 22. Ullah M., Shahid A., Din I. ud, Roman M., Assam M., Fayaz M., Ghadi Y., & Aljuaid H. (2022). Analyzing Interdisciplinary Research Using Co-Authorship Networks. *Complexity*, 2022, 1–13. <https://doi.org/10.1155/2022/2524491>
 23. Schäfermeier B., Hirth J., & Hanika T. (2022). Research topic flows in co-authorship networks. *Scientometrics*. <https://doi.org/10.1007/s11192-022-04529-w>
 24. Lande D. (2015). A Domain Model Created on the Basis of Google Scholar Citations. In *CEUR Workshop Proceedings (Ceur-Ws. Org) Vol-1536*, 57–61.
 25. Lande D., & Dmytrenko O. (2022). *Research of Topological Properties of Network Reflections Obtained Using Different Algorithms for Scanning Initial Networks* (pp. 347–361). https://doi.org/10.1007/978-3-030-89902-8_26
 26. Cai G. (2024, January). Research on the Application of Louvain's Algorithm in Traditional Cultural Media Communication. In *2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC)* (pp. 1–5). IEEE. DOI: <https://doi.org/10.1109/ICOCWC60930.2024.10470776>
 27. Lande D., Strashnoy L. *Advanced Semantic Networking based on the large language models: Monograph*. Kyiv: *Engineering*, 2025. 258 p. ISBN 978-617-8180-02-7
 28. Cherven K. (2015). *Mastering Gephi Network Visualization*. Packt Publishing.

Submitted: 14.08.2025

D. Lande^{1,2} (<https://orcid.org/0000-0003-3945-1178>),
A. Snarskii^{1,2} (<https://orcid.org/0000-0002-4468-4542>),
D. Manko² (<https://orcid.org/0000-0003-1848-2952>),
I. Linchevskiy¹ (<https://orcid.org/0000-0002-2896-9580>),
V. Fedotov¹ (<https://orcid.org/0000-0001-7821-0989>)

¹National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”;

²Institute for Information Recording, NAS of Ukraine

Network Analysis of Co-authorship and Research Directions in Thermoelectrics

This paper presents a methodology for constructing an extended co-authorship network in the field of thermoelectrics by integrating direct co-authorship links with thematic proximity derived from researchers' Google Scholar profiles. The network is built through targeted probing using relevant keywords (e.g., "thermoelectrics"), and link weights are determined by both the number of joint publications and shared research interests. This approach enables the identification of not only direct collaborators but also potential interdisciplinary partners, revealing the latent structure of the scientific community.

Additionally, a descriptor network is constructed based on the co-occurrence of keywords across author profiles, forming an ontological map of the field. Thematic clusters are identified using the Louvain algorithm, representing core research directions. For the first time in this context, Large Language Models (LLMs) are employed to interpret cluster content by generating meaningful, human-readable labels from lists of descriptors. This allows for fast, objective, and scalable identification of scientific trends without relying on expert annotation. The analysis shows that the extended network exhibits higher density than the traditional co-authorship network, highlighting the significance of thematic connections. Centrality measures (degree and betweenness) help identify key contributors and structural bridges within the field. The proposed approach supports the analysis of scientific communities, detection of research schools, and collaboration forecasting.

Keywords: generalized co-authorship network, subject domain, LLM, scientometric service, network sounding, topic descriptors, thermoelectricity.

Надійшла до редакції 14.08.2025