

**Кучанський Олександр Юрійович**

Кандидат технічних наук, доцент, доцент кафедри інформаційних систем та технологій,

[orcid.org/0000-0003-1277-8031](https://orcid.org/0000-0003-1277-8031)

Київський національний університет імені Тараса Шевченка, Київ

**СКЛАДОВІ АНАЛІЗУ НАУКОВИХ МЕРЕЖ**

**Анотація.** Дослідження наукових просторів є складною задачею, яка має враховувати динаміку розвитку їх компонентів, а також враховувати численні зміни, які виникають внаслідок зростання швидкості продукування наукової інформації та її поширення. Аналіз об'єднань суб'єктів та об'єктів наукового простору, зокрема мереж наукової співпраці, мереж цитування та спільного цитування, є актуальною задачею дослідження, оскільки є в основі коректного оцінювання показників науково-дослідної діяльності загалом. В роботі наведено огляд задач, які виникають в межах аналізу наукових мереж: дослідження мереж наукової співпраці, дослідження мереж цитування, спільного цитування та бібліографічних з'єднань, ідентифікація напрямів наукових досліджень та аналіз тематик, побудова наукових мереж на основі спільних ключових слів, дослідження гетерогенних наукових мереж тощо. Глобальність поширення та динаміка розвитку наукових мереж визначає підходи до їх аналізу. Дослідження наукових мереж дає змогу зрозуміти структуру взаємодії суб'єктів та об'єктів наукових мереж (науковців, закладів вищої освіти та їх структурних підрозділів), а також провести ґрунтовне оцінювання їх науково-дослідної діяльності. Також наведено візуалізацію мережі наукової співпраці, що утворилась в межах виконання чотирьох бюджетних науково-дослідних робіт і одного проекту програми Erasmus+KA2 за період з 2015 по 2020 роки. Візуалізація виконана з використанням силового алгоритму Фрухтермана – Рейнгольда. Також на основі алгоритму візуалізації Yifan Hu зображена мережа цитування публікацій науковців за цими ж проектами. Результати оцінювання можуть бути використані як власне науковцями та освітніми установами для моніторингу динаміки науково-дослідної діяльності, так і державними органами, зокрема Міністерством освіти і науки України, а також органами місцевого самоврядування для фінансового стимулювання певних напрямів наукових досліджень, які мають позитивну динаміку та перспективи розвитку.

**Ключові слова:** аналіз наукових мереж; мережа наукової співпраці; мережа цитування; наукометрія; науковий простір

**Вступ**

Наукова мережа – це об'єднання суб'єктів наукової діяльності, тобто науковців, які об'єднані певними правилами створення і функціонування членів в межах мережі. Однорідність суб'єктів наукової діяльності не виключають наявність у них специфічних ідентифікаторів, що допомагає досліджувати їх окремо.

Наукові мережі можуть поєднуватись між собою, утворюючи ієрархію. Наукові мережі (як об'єкти) можуть розглядатися як частини наукового простору. Простором є деяка множина структурно упорядкованих об'єктів та їх ідентифікаторів, яка подається моделлю, що відображає логічну структуру групування і упорядкування ідентифікаторів об'єктів такого простору [1].

Дослідження наукових просторів є складною задачею, що має враховувати динаміку розвитку їх компонентів, а також враховувати численні зміни, які виникають внаслідок зростання швидкості продукування наукової інформації та її поширення.

Оскільки наукова інформація поширюється шляхом публікування наукових робіт, то для формування і оцінювання розвитку взаємодії між компонентами наукових просторів можна скористатися даними про цитування наукових публікацій, аналізом спільних ключових слів публікацій тощо. Відповідний математичний та алгоритмічний апарат, що допомагає дослідити наукову діяльність в різних вимірах визначається аналізом наукових мереж.

Основними завданнями, які вирішуються в межах аналізу наукових мереж є:

1. Дослідження принципів створення і функціонування, а також оцінювання результатів співпраці в межах мереж співавторів або мереж наукової співпраці.
2. Побудова і дослідження мереж цитування.
3. Дослідження мереж спільного цитування та бібліографічних з'єднань [2].
4. Ідентифікація напрямів наукових досліджень і аналіз тематик, а також побудова наукових мереж на основі спільних ключових слів.
5. Дослідження гетерогенних наукових мереж.

Детально про деякі задачі оцінювання науково-дослідної діяльності суб'єктів та об'єктів наукового простору описано в роботах [3 – 9].

### Мета статті

Мета – розгляд основних задач, які виникають в межах аналізу наукових мереж, що можуть бути використані для оцінювання науково-дослідної діяльності суб'єктів та об'єктів наукових просторів.

### Виклад основного матеріалу

#### Мережі наукової співпраці

Мережа наукової співпраці – це соціальна мережа, вузли якої відображають науковців, а зв'язки відображають співавторство спільних наукових публікацій [10; 11]. Публікування спільних наукових праць є основним видом наукової взаємодії, що активно досліджується і може бути оцінена.

В роботі [12] було проведено кластерний та частотний аналізи мережі для виявлення мікроструктури мережі наукової співпраці щодо певних аспектів науковців; виявлення напрямку досліджень в рамках локальних підмереж мережі наукової співпраці; ідентифікація центру співпраці.

Згідно з дослідженнями, які опубліковані в роботі [13], мережа наукової співпраці є асортативною, тобто хаби мережі поєднуються з іншими хабами, які концентрують навколо себе вузли, а вузли нижчого рівня поєднуються з відповідними вузлами нижчого рівня. Асортативність не структурна, а створюється деяким механізмом, який регулює еволюцію мережі.

На рис. 1 зображена мережа наукової співпраці науковців, які є виконавцями чотирьох бюджетних науково-дослідних робіт та одного проєкту програми Erasmus+KA2 за період з 2015 по 2020 роки: «Розробка методів аналізу якості науково-дослідної роботи вчених, ЗВО МОН України та окремих структурних підрозділів», Київський національний університет імені Тараса Шевченка, № ДР 0119U100187 (2019 – 2020), керівник – А.О. Білощицький, відповідальний виконавець – О.Ю. Кучанський; «Розробка комбінованих методів ідентифікації неповних дублікатів та виявлення повноти висвітлення наукових результатів дисертаційних досліджень, опублікованих автором», Київський національний університет будівництва і архітектури, № ДР 0119U002579 (2019 – 2020), керівник – П.П. Лізунов, відповідальний виконавець – О.Ю. Кучанський; «Методологічні основи створення інформаційного середовища управління науковими дослідженнями структурних одиниць ВНЗ МОН України», Київський національний університет будівництва і архітектури, № ДР 0115U000330 (2015 – 2017), керівник –

А.О. Білощицький; «Розробка методів пошуку запозичень контенту в електронних документах та інтеграція згенерованих документів в систему МОН України», Київський національний університет будівництва і архітектури, № ДР 0116U000845 (2016 – 2017), керівник – П.П. Лізунов; проєкт програми Erasmus+KA2 «GameHub: співпраця університетів та підприємств в ігровій індустрії в Україні» 561728-EPP-1-2015-1-ES-EPPKA2-SVHE-JP.

Дані про наукові публікації авторів були зібрані з наукових звітів за темами з врахуванням індексації у Google Scholar. Мережа наукової співпраці побудована у програмному забезпеченні Gephi 0.9.2 з використанням силового алгоритму візуалізації Фрухтермана – Рейнгольда (Fruchterman-Reingold). Ціль алгоритму Фрухтермана – Рейнгольда, що було застосовано для візуалізації мережі, полягає, як і всіх силових алгоритмів візуалізації графів, у розміщенні графа у двовимірному та тривимірному просторах так, щоб ребра мали більш-менш однакову довжину і мали менше перетинів [14].

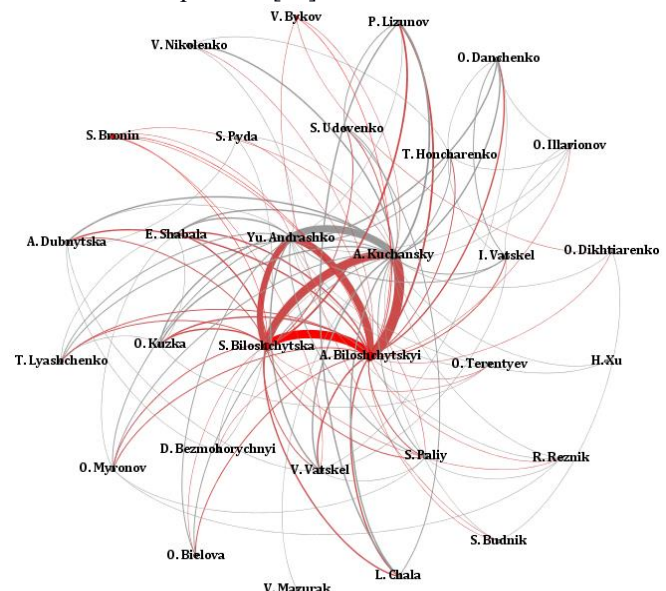


Рисунок 1 – Мережа наукової співпраці. Вузли мережі – науковці, зв'язки – співавторство наукових публікацій

Важливою задачею дослідження мережі наукової співпраці є визначення підмереж, які включають вузли, які є найбільш поєднаними. Для таких підмереж визначається хаб, тобто вузол, кількість зв'язків якого значно перевищує середню кількість в мережі. Властивими характеристиками кожного вузла мережі наукової співпраці є здатність поєднуватися з іншими вузлами (придатність) та здатність до централізації, тобто до перетворення у хаб [13]. Для встановлення центральності використовуються такі чинники: ступінь центральності, ступінь посередництва, гармонічна центральність, ступінь впливовості тощо. Ці чинники допомагають визначити роль кожного вузла в мережі.

Окрім мережних хабів, які гарантують узгодженість, спрямованість мережі можуть активувати інші вузли мережі, важливим є встановлення вузлів брокерів, які поєднують різні реляційні острови, тобто забезпечують зв'язок між підмережами і закривають отвори в мережі. Також деякі вузли можуть бути одночасно брокерами і хабами [15].

Розглянемо основні параметри графів наукових мереж, які необхідно розраховувати для проведення всебічного аналізу мережі.

Степінь вершини або валентність вершини – це кількість ребер графа, які інцидентні цій вершині. У мережі наукової співпраці, чим більший степінь вузла, тим більше співавторів у такого суб'єкта наукової діяльності. Середній степінь графа визначається як середня кількість ребер графа.

Метричний простір, визначений на множині вершин за допомогою відстаней на графі, називається метрикою графа. Множина вершин (неорієнтованого графа) і функція відстаней утворюють метричний простір тоді, коли граф є зв'язним. Ексцентриситет вершини визначається як найбільша відстань від даної вершини до будь-якої іншої вершини. Радіус графа – це мінімальний ексцентриситет будь-якої з вершин графа. Діаметр графа – це максимальний ексцентриситет будь-якої з вершин графа. Центральна вершина графа – це вершина, ексцентриситет якої дорівнює радіусу графа [16].

Степінь близькості вузла визначається як обернена величина суми довжин найкоротших шляхів між вузлом та всіма іншими вузлами графа. Чим більше центральний вузол, тим ближче він до всіх інших вузлів [17]:

$$C(u) = \left( \sum_v d(v, u) \right)^{-1},$$

де  $C(u)$  – степінь близькості вузла;  $d(v, u)$  – відстань між вершинами  $u$  і  $v$ .

Якщо граф не сильно зв'язаний, то дуже поширена ідея використання суми обернених величин до відстаней, замість оберненої величини до суми відстаней:

$$H(u) = \sum_{v \neq u} (d(v, u))^{-1},$$

де  $H(u)$  – гармонічний степінь близькості вузла, причому  $(d(v, u))^{-1} = 0$ , якщо шляху  $(v, u)$  немає [18].

Степінь посередництва – це міра центральності в графі, заснована на найкоротших шляхах. Для будь-якої пари вершин в зв'язному графі існує щонайменше один (найкоротший) шлях між вершинами, для якого кількість ребер або сума ваг

цих ребер мінімальна. Степінь посередництва для кожної вершини дорівнює числу цих найкоротших шляхів, які проходять через вершину. Степінь посередництва вершини  $u$   $g(u)$  задається за формулою:

$$g(u) = \sum_{v \neq u \neq s} \frac{\sigma_{vs}(u)}{\sigma_{vs}},$$

де  $\sigma_{vs}$  – загальна кількість найкоротших шляхів з  $v$  в  $s$ ;  $\sigma_{vs}(u)$  – кількість цих шляхів, які проходять через вершину  $u$ .

Степінь впливовості – це міра впливу вузла в мережі. Відносні величини цього показника призначаються всім вузлам на основі концепції, що зв'язок з вузлом високого ступеня впливовості вкладає більше в показник розглянутого вузла, ніж аналогічний зв'язок з вузлом низького ступеня впливовості. Високий степінь впливовості означає, що вузол пов'язаний з багатьма вузлами, які мають високі ступені впливовості. Степінь впливовості є мірою впливу, що вузол здійснює на мережу. Якщо вузол пов'язаний з багатьма вузлами, які також мають високі показники впливовості, то вузол буде мати високий степінь впливовості. Якщо  $a_{u,v}$  – елементи бінарної матриці суміжності деякого графа  $G$ ,  $x_v$ ,  $x_u$  – відносні показники впливовості вершин  $v$  і  $u$  відповідно, то

$$x_v = \frac{1}{\lambda} \sum_{u \in G} a_{u,v} x_u,$$

де  $\lambda$  – константа, що є власним вектором для даного виразу [19].

Для графа мережі наукової співпраці також розраховують модулярність, що визначає міру структурності мережі. Міра модулярності була розроблена для вимірювання сили розбиття мережі на модулі (групи, кластери або наукові спільноти). Мережі з високою модулярністю мають щільні зв'язки між вузлами всередині модулів, але слабкі зв'язки між вузлами в різних модулях.

Модулярність часто використовується в оптимізації методів для розпізнавання структури наукової спільноти в мережах. Однак було показано, що для модулярності є проблема меж роздільної здатності, тому цей підхід не може розрізнити малі спільноти. Модулярність визначається як частка ребер від загального числа ребер, які потрапляють в ці групи, мінус очікувана частка ребер, які потрапили б в ті ж групи, якби вони були розподілені випадково. Значення модулярності лежить в інтервалі  $[-1, 1]$  [20].

Коефіцієнт кластеризації є мірою ступеня того, як тісно вузли в графі мають тенденцію групуватися разом. Наявні дані свідчать про те, що в більшості реальних мереж, і, зокрема в соціальних мережах,

вузли, як правило, створюють тісно пов'язані групи, що характеризуються відносно високою щільністю зв'язків. Ця ймовірність більше ніж середня ймовірність випадкового зв'язку між двома вузлами [21].

Локальний коефіцієнт кластеризації з вершиною (вузлом) в графі визначає наскільки близько його сусіди повинні бути згруповані (повний граф). Нехай граф  $G = (V, E)$  складається з множини вершин  $V$  та ребер  $E$ . Ребро  $e_{ij}$  з'єднує вершину  $v_i$  з вершиною  $v_j$ . Окіл  $N_i$  вершини  $v_i$  визначається множиною таких вершин графа  $G$ , що з нею зв'язані:

$$N_i = \{v_j \mid e_{ij} \in E \vee e_{ji} \in E\},$$

тоді локальний коефіцієнт кластеризації для орієнтованих графів визначається за формулою [21]:

$$C_i = \frac{\text{card}(\{e_{jk} \mid v_k, v_j \in N_i, e_{jk} \in E\})}{\text{card}(N_i)(\text{card}(N_i) - 1)},$$

тобто визначається як відношення кількості зв'язків в околі вершини до кількості зв'язків, які взагалі б могли існувати між вершинами в цьому околі.

Для неорієнтованих графів локальний коефіцієнт кластеризації визначається за формулою:

$$C_i = \frac{2 \cdot \text{card}(\{e_{jk} \mid v_k, v_j \in N_i, e_{jk} \in E\})}{\text{card}(N_i)(\text{card}(N_i) - 1)}.$$

Загалом коефіцієнт кластеризації визначається за формулою:

$$C = \frac{\lambda_G(v)}{\tau_G(v)},$$

де  $\lambda_G(v)$  – кількість трикутників на множині вершин в околі вершини  $v$ , або число підграфів графа  $G$  з трьома ребрами і трьома вершинами, одна з яких  $v$ .  $\tau_G(v)$  – число трійок в околі  $v$ , що не обов'язково індуковані, тобто для неорієнтованого графа:

$$\tau_G(v) = \frac{\text{card}(N_i)(\text{card}(N_i) - 1)}{2}.$$

Також для встановлення щільності зв'язку визначається мережевий коефіцієнт кластеризації, що розраховується як середнє значення локальних коефіцієнтів кластеризації для всіх вершин  $n$  графа  $G$  [21; 22]:

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i,$$

де  $C_i$  – локальний коефіцієнт кластеризації вершини  $v_i$  графа  $G$ .

### Мережі цитування наукових публікацій

Окрім мереж наукової співпраці, будують та досліджують мережі цитування публікацій

науковців. Мережа цитування включає вузли, що відповідають науковим публікаціям та зв'язки між ними, що відображають цитування одних наукових публікацій у інших. В роботі [23] було сформульовано означення мережі цитування так: коли документ  $d_i$  цитує документ  $d_j$  ми можемо показати це стрілкою, що з'єднує вузол  $d_i$  з вузлом  $d_j$ , причому відображення всіх цитувань документів таким чином буде візуалізуватися у мережу цитат або граф цитування.

Відомі загальні властивості мережі цитування:

1. Мережа цитування представляється ациклічним орієтованим графом.

2. Всі дуги цитування в графі спрямовані назад в часі, тобто цитування може відбуватися тільки тих публікацій, які були опублікованими в минулому.

3. Вага цитування між співавторами в мережі наукової співпраці може зростати, якщо автори будуть продовжувати наукову співпрацю в подальшому.

4. Вузли та дуги, які додаються до мережі в ній залишаються та не можуть в подальшому видалятися.

5. Зміни мережі відбуваються здебільшого в передній частині мережі. Мережа, яка вже сформована, в цілому має статичний характер.

Для мережі цитування, окрім характеристик, які наведені вище, розраховують документ-хаб та авторитетний документ. Авторитетний документ – це документ, який має більшу питому вагу серед документів певної тематики, тобто більше число документів посиляються на певний документ. Хаб-документ – це документ, що містить багато посилань на авторитетні документи. Знаходження документу хабу та авторитетного документу можна виконати за допомогою ітеративного алгоритму HITS (Hyperlink Induced Topic Search) [24].

Ще одним алгоритмом, який може використовуватися при дослідженні мережі цитування є алгоритм PageRank. PageRank загалом являє собою сімейство алгоритмів оцінювання важливості вебсторінок за допомогою розв'язання систем лінійних алгебраїчних рівнянь. Для кожної сторінки обчислюється деяке дійсне число. Чим більше це число – тим «важливіша» сторінка [25]. Значення PageRank для сторінки  $A$  обчислюється за правилом: нехай  $T_1, T_2, \dots, T_n$  – сторінки, що посиляються (цитують) сторінку  $A$ . Якщо  $C(T_i)$  – кількість посилань, які виходять зі сторінки  $T_i$ , то

$$\text{PR}(A) = d \sum_{i=1}^n \frac{\text{PR}(T_i)}{C(T_i)} + (1-d),$$

де  $\text{PR}(A)$  – значення PageRank для сторінки  $A$ ;  $d$  – коефіцієнт демпінгу,  $d \in [0, 1]$ . На практиці  $d = 0.85$ .



В роботі [3] розглянуто метод PR-q на основі PageRank, що допомагає розраховувати оцінку науково-дослідної діяльності науковців або суб'єктів наукового простору, враховуючи їх мережу цитування.

Аналіз мережі цитування необхідний передусім для виявлення наукових спільнот, що є одним з основних завдань мережевого аналізу [26]. Також побудована мережа цитування є потужним інструментом для наукометричного оцінювання наукової діяльності окремих суб'єктів наукового простору [3].

Отже, для мережі цитування розраховують загалом ті ж показники, що і для мережі наукової співпраці з поправкою на орієнтованість графа мережі. Важливе значення для встановлення щільності зв'язку між публікаціями науковців, зокрема тих, які працюють в межах спільних проєктів, є розрахунок коефіцієнта кластеризації.

На рис. 2 зображена мережа цитування публікацій науковців, які є виконавцями бюджетних науково-дослідних тем № ДР 0119U100187, № ДР 0119U002579, № ДР 0116U000845 та проєкту програми Erasmus+KA2 561728-EPP-1-2015-1-ES-EPPKA2-CBHE-JP за період з 2015 по 2020 роки.

Мережа цитування публікацій науковців побудована у програмному забезпеченні Gephi 0.9.2 з використанням силового алгоритму візуалізації Yifan Hu [27]. Вершинами графа є наукові публікації виконавців, а також і статей інших науковців, які ці публікації цитують.

### Мережі спільного цитування та бібліографічних з'єднань

Мережі спільного цитування використовуються для встановлення семантичної міри подібності для документів, які належать мережі цитування. Спільне цитування визначається як частота, з якою два документи цитуються разом іншими документами [2]. Якщо хоча б в одному документі цитуються два спільних документи, ці документи, як вважається, цитуються разом. Чим більше спільних цитат отримують два документи, тим вищий їх ступінь спорідненості і тим більше шансів на те, що вони семантично пов'язані.

Документи поєднуються бібліографічно, якщо вони цитують один або декілька спільних документів. На рис. 3 показано цей принцип: документи А і В вважаються пов'язаними між собою, оскільки вони цитують документи С, D і E.

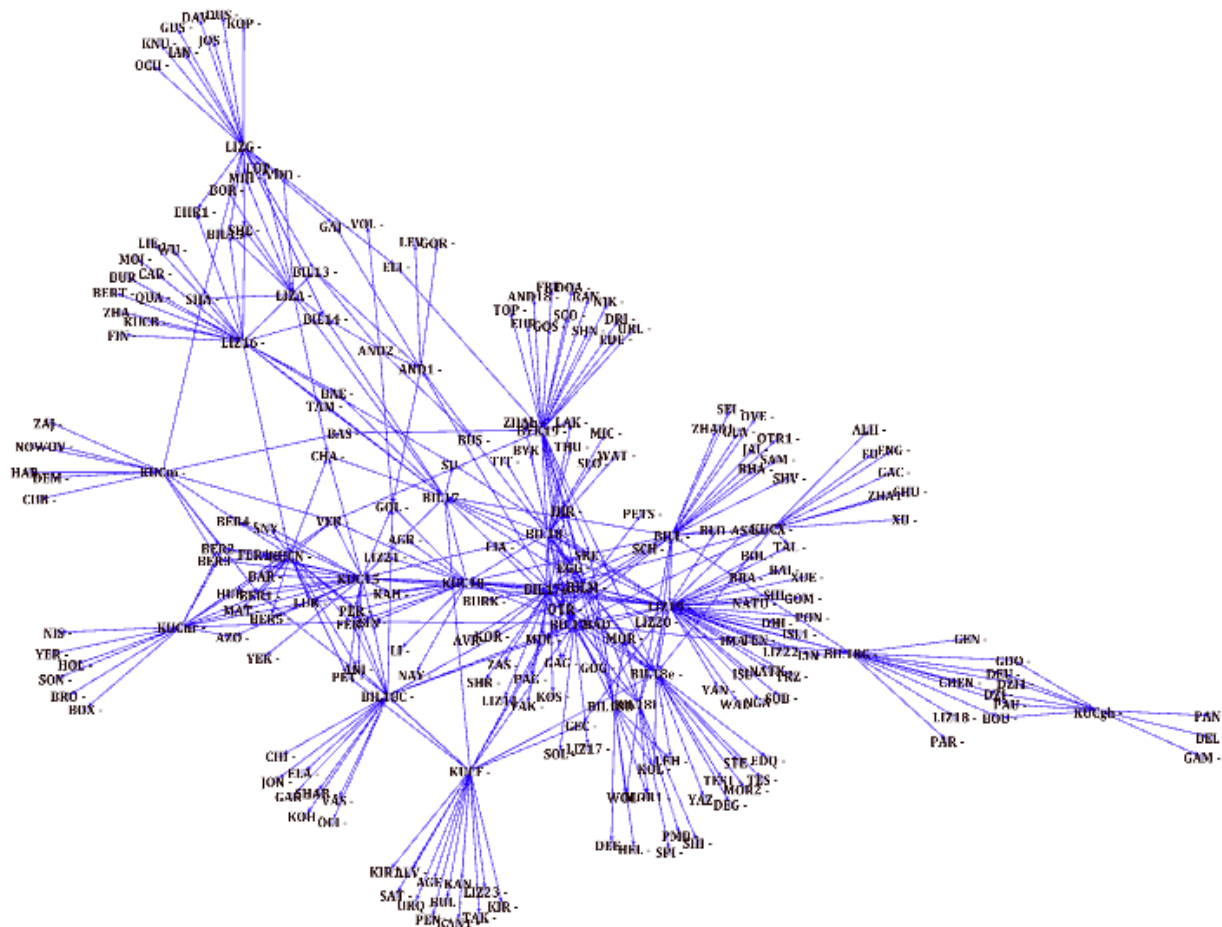


Рисунок 2 – Мережа цитування публікацій науковців

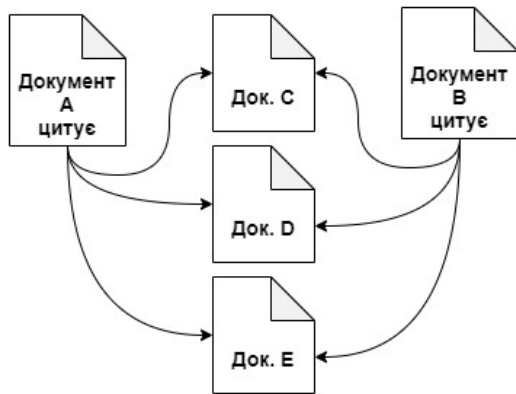


Рисунок 3 – Бібліографічне з'єднання

На противагу цьому два документи спільно цитуються, якщо принаймні одна публікація цитує обидва. Такий підхід проілюстровано на рис. 4. Документи А і В пов'язані між собою, оскільки вони цитуються у документах С, D і Е. Чим більше спільних цитувань отримують два документи, тим більше вони пов'язані [28].

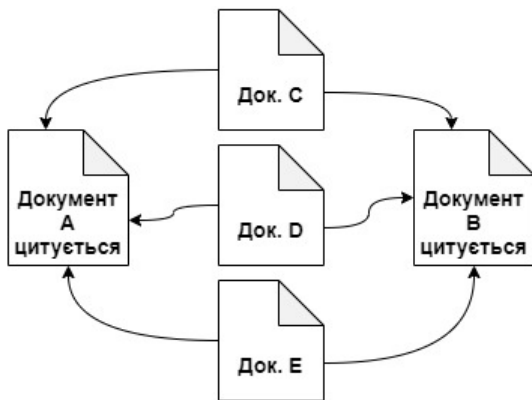


Рисунок 4 – Аналіз спільного цитування

Хоча обидва підходи можна використовувати для виявлення подібних робіт, вони необхідні для різних цілей. Бібліографічне з'єднання є здебільшого ретроспективним аналізом, на відміну від спільного цитування [29]. Однак обидва підходи часто дають незадовільні результати, оскільки вони використовують лише посилання в кінці документа, не аналізуючи при цьому сукупність цитат. Для виправлення цього недоліку було створено індекс близькості цитування. Замість того, щоб просто використовувати бібліографію, в аналізі близькості цитування, інформація, отримана від близькості цитат один до одного в повнотекстовій публікації, використовується для обчислення індексу близькості цитування (CPI) в три етапи:

1. Документ аналізується з використанням евристик для опрацювання цитат, включаючи їх позицію в документі.
2. Встановлюється відповідність між цитатами і відповідними пунктами бібліографії.
3. На третьому кроці досліджується близькість між кожною парою цитат. Основне

припущення полягає в тому, що чим ближче розташовані цитати одна до одної в тексті публікації, тим більше ймовірність того, що вони пов'язані. Наприклад, якщо в одному реченні подано дві цитати, то ймовірність того, що вони дуже схожі, вище (CPI = 1), ніж якщо б вони були лише в тому ж абзаці (CPI = 1/2), параграфі (CPI = 1/4), спільному збірнику праць або монографії (CPI = 1/8), спільному журналі, але в різних випусках (CPI = 1/16). Оцінки, які вказані вище є емпіричними і описані в роботі [2]. На рис. 5 зображено принцип розрахунку індексу близькості між двома цитатами.



Рисунок 5 – Схема розрахунку індексу близькості цитування

Недоліком такого підходу до розрахунку близькості є те, що для різних наукових напрямів, видань тощо, доцільно враховувати різні ваги, що впливає на розрахунок. Проте результати аналізу близькості цитування можна покращити, оцінивши якомога більше джерел.

### Ідентифікація напрямів наукових досліджень

Розвиток науки та технологій в останні кілька десятиліть засвідчує зростання ступеня міждисциплінарності. В роботі [30] проведено дослідження гіпотези щодо підвищення рівня проникнення одних дисциплін в інші. Зокрема досліджувався ступінь міждисциплінарності за період з 1975 по 2005 роки у шести наукових галузях, на основі розрахунку бібліометричних показників, оцінювання інтеграції та різноманітності Рао-Стірлінга. Дослідження підтверджують посилення міждисциплінарних зв'язків з часом, що змінює розуміння меж напрямів наукових досліджень та дає загальні орієнтири для майбутніх досліджень характеристик та взаємозв'язків між дисциплінами або напрямками наукових досліджень.

Існує загальноприйнята думка, що чіткий дисциплінарний поділ напрямів наукових досліджень, який складається в консервативних академічних колах, не сприяє популяризації науки та технологій і ускладнює процес науково-технічного прогресу. Для активізації зусиль з міждисциплінарних

досліджень, зокрема щодо вивчення процесів інтеграції суб'єктів та об'єктів наукової діяльності, Національною академією наук США, Національною технічною академією та Інститутом медицини США було запроваджено ініціативу National Academies Keck Futures Initiative ("NAKFI") [31]. Важливою задачею дослідження за цією ініціативою є виявлення меж дисциплін або напрямів наукових досліджень та динаміки їх розвитку, використовуючи різні показники, в тому числі показники цитування публікацій, які відповідають конкретному напрямку.

Задача ідентифікації напрямів наукових досліджень тісно пов'язана із задачею розрахунку ступеня близькості або подібності документів, які відображають наукові праці. Подібність наукових праць визначається на основі аналізу тексту з ідентифікацією спільних ключових слів та тематики. В роботі [30] пропонуються такі шляхи ідентифікації напрямів досліджень:

1. Розрахунок індексу інтеграції або дисциплінарної різноманітності.

2. Побудова матриці подібності для кожної тематичної категорії та проведення групування елементів тематичної категорії.

4. Генерування мап для візуалізації зв'язку в межах напрямів.

Для того щоб ефективно оцінити якість науково-дослідної роботи в межах університету, регіону або держави, необхідно насамперед визначити зв'язки між всіма науковими публікаціями за допомогою цитувань між цими публікаціями. Ідентифіковані зв'язки між науковими публікаціями відкривають можливості для кластеризації цих публікацій за науковими напрямками. Кожен науковий напрям включатиме в такому випадку певну кількість авторів публікацій, які належать до конкретного напрямку або кластеру. Отже, можна поставити у відповідність конкретному автору певну кількість напрямів досліджень, тобто провести ідентифікацію напрямів досліджень авторів. Потенціал наукового напрямку визначатиметься оцінкою результатів науково-дослідної діяльності науковців, які до цього напрямку належать. По суті кластери або напрями наукових досліджень можна розглядати як динамічні об'єкти, які мають історію розвитку та передбачуваний потенціал.

Завдання побудови методу ідентифікації напрямів досліджень науковців та кластерний аналіз наукових публікацій є актуальними для виявлення перспективних напрямів наукових досліджень, які формуються або активно розробляються в науковому просторі. Це дасть можливість більш ефективно здійснювати управління фінансовими та організаційними складовими розвитку конкретних наукових напрямів на рівні держави. Також це допоможе підтримувати напрями, які є визначальними в стратегії науково-технічного

розвитку, залучати міжнародні гранти, формувати консорціуми для проведення досліджень тощо.

Під ідентифікацією напрямів досліджень науковців будемо вважати процес встановлення відповідності між конкретним науковцем та науковими напрямками. Задача ідентифікації напрямів досліджень науковців активно розглядається в науковому середовищі. Наприклад, в роботі [32] така задача розв'язується шляхом побудови методів визначення напрямів наукової діяльності на основі аналізу ключових слів публікацій. Проте в роботі [32] автори виокремлюють публікації, які містять спільні ключові слова і словосполучення, але не враховують інші можливі зв'язки між цими публікаціями: цитування, подібність контенту статей тощо. В роботі [33] запропоновано методи використання бібліометричних ознак наукових робіт для кластеризації авторів цих робіт за напрямками наукових досліджень. Також в роботі [33] розглядається задача визначення нових напрямів наукових досліджень шляхом побудови графа спільного цитування між публікаціями науковців різних країн. В роботі [34] описана технологія класифікації об'єктів адаптованим методом класифікації, в якому використовується спеціальна функція відстані, яка дає змогу враховувати декілька ознак об'єкта. Така технологія допомагає здійснювати параметричне керування розбиттям відповідно до важливості ознак. Цей підхід має важливе значення для класифікації різноманітних соціальних явищ. Метод, який описано в роботі [34], може бути використано і для задачі розподілу науковців за напрямками досліджень.

Задачею, яка пов'язана із задачею ідентифікації наукових напрямів, є виявлення колективів авторів, які працюють в конкретному науковому напрямі. Метод виявлення колективів науковців на основі аналізу подібності текстів наукових робіт розглянуто в роботі [35]. Проте в роботі [35] не враховано, що до певного наукового напрямку можуть належати одразу кілька колективів авторів.

Для проведення кластерного аналізу публікацій зручно використовувати граф цитування. В роботі [36] розглядаються методи інтелектуального аналізу даних, які представляються за допомогою графів, зокрема кластеризації графа цитування. Задача кластеризації графа досить широко досліджена та існує багато методів розв'язання цієї задачі. Одним з алгоритмів розв'язання задачі кластеризації графа є алгоритм Louvain, який описано в роботі [37]. Цей алгоритм реалізує метод максимізації модулярності графа та може бути застосований для швидкої кластеризації графів великої розмірності. Пропонований алгоритм може бути корисним для кластеризації графів цитувань, оскільки ці графи

характеризуються значною кількістю вершин і дуг. При проведенні процедури кластеризації графа необхідно враховувати проблему недостатньої стійкості кластерів, структура яких змінюється із плином часу.

В роботі [38] запропоновано метод визначення кумулятивних ядер, використання яких сприяє підвищенню стійкості кластеризації. В роботі [39] також розглядаються питання стійкості кластеризації графів та запропоновано інший метод для вирішення цієї проблеми, а саме: знаходження стійких ансамблів шляхом здійснення послідовних кластеризацій графа. В роботі [40] запропоновано метод кластеризації даних BIRCH (Balanced Interval Reducing and Clustering using Hierarchies), який дає змогу мінімізувати кількість звертань до бази даних. Його застосування є доцільним при опрацюванні даних великого обсягу, зокрема у випадку, коли база даних містить значну кількість публікацій, які необхідно кластеризувати.

Побудова напрямів наукової діяльності, в межах яких співпрацюють науковці на основі кластеризації публікацій та ідентифікації напрямів досліджень науковців, є важливою задачею для наукових та освітніх установ. Ця задача може бути використана для проведення оцінювання діяльності наукових установ та прогнозування перспектив розвитку установ з урахуванням наявних ресурсів. В роботі [41] розглядається проблема побудови таких моделей. В роботі [42] побудована параметрична модель, що забезпечує прогнозування та оцінювання функціонування наукових та освітніх установ на основі перерозподілу наявних ресурсів. В роботі [3] наведено огляд традиційних методів оцінювання результатів наукової діяльності науковців на основі аналізу цитування публікацій цих науковців. Також в роботі [3] пропонуються скалярні та векторні методи оцінювання результатів наукової діяльності. Ці методи можуть бути застосовані для оцінювання результатів наукової діяльності науковців, що працюють в певному науковому напрямі та відповідного наукового напрямку в цілому.

Для кластеризації публікацій з метою виявлення предметних наукових просторів може бути використано аналіз тематик. Для цього може бути застосовано метод латентного розміщення Діріхле (LDA). Латентне розміщення Діріхле представляється ймовірнісною тематичною моделлю, що використовується для моделювання дискретних наборів даних, шляхом вивчення зв'язків між словами, темами та документами. LDA розглядає кожен документ як множину тем і опрацьовує її, порівнюючи зі зразком [43; 44]. Такий підхід подібний до латентно-семантичного аналізу (pLSA),

але різницею є те, що тематики, як вважається, мають розподіл Діріхле. Також для аналізу тематик використовується метод розкладу невід'ємних матриць (NMF), що є оптимізаційним методом з лінійної алгебри і дає змогу зменшити розмірність даних, якщо тематики представляються у вигляді матриць [44].

Ще одним напрямом дослідження є вивчення еволюції дослідницької галузі. Внаслідок проведеного в роботі [45] аналізу предметних галузей дослідження отримано такі емпіричні характеристики:

1. Еволюція дослідницької галузі визначається, зазвичай, кількома дисциплінами (3 – 5), які генерують понад 80% документів (концентрація наукової продукції).

2. Еволюція предметних напрямів дослідження залежить від критичного шляху кожної дисципліни, що в нього включається: вони можуть бути материнськими дисциплінами, що виникли в галузі дослідження, або новими дисциплінами, що з'явилися внаслідок еволюції.

3. Еволюція також може бути зумовлена новою дисципліною, що виникла внаслідок зміни спеціалізації в галузі прикладних або фундаментальних наук та/або конвергенції між дисциплінами.

4. Еволюція конкретних предметних напрямів може бути зумовлена під впливом отриманих результатів як з прикладної, так і з фундаментальної складової науки.

Отримані результати допомагають зрозуміти деякі особливості динаміки науки в цілому, а також окремих предметних наукових просторів. Зокрема, в роботі [46] досліджується динаміка зростання цитування наукових публікацій, що є важливим показником оцінювання мережі цитування.

## Висновки

За останні десятиліття відбуваються важливі зміни в географії науки та наукової співпраці. Наукові мережі або мережі наукової співпраці, які раніше функціонували локально в межах університетів та наукових інститутів, а також розповсюджувались на одну або кілька країн і були більш статичні, зараз охоплюють практично весь світ. Швидкість створення нової інформації і використання засобів створення мережевої співпраці науковців сприяють динамічному розвитку та урізноманітненню наукових мереж.

В роботі розглянуто актуальні задачі аналізу наукових мереж, які є в основі коректного розрахунку показників науково-дослідної діяльності суб'єктів та об'єктів наукового простору, зокрема закладів вищої освіти та окремих науковців.



## Список літератури

1. Кремень В.Г., Биков В.Ю. Категорії «простір» і «середовище»: особливості модельного подання та освітнього застосування. *Теорія і практика управління соціальними системами*. 2013. № 2. С. 3–16.
2. Gipp B., Beel J. Citation Proximity Analysis (CPA) – A New Approach for Identifying Related Work Based on Co-Citation Analysis. *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09), Rio de Janeiro, Brazil*. 2009. С. 571 – 575.
3. Biloshchytskyi A., Kuchansky A., Andrashko Yu., Biloshchytska S., Kuzka O., Terentyev O. Evaluation methods of the results of scientific research activity of scientists based on the analysis of publication citations. *Eastern-European Journal of Enterprise Technologies*. 2017. 3(2 (87)). P. 4 – 10. doi: 10.15587/1729-4061.2017.103651
4. Kuchansky A., Andrashko Yu., Biloshchytskyi A., Danchenko O., Ilarionov O., Vatskel I., Honcharenko T. The method for evaluation of educational environment subjects' performance based on the calculation of volumes of M-simplexes. *Eastern-European Journal of Enterprise Technologies*. 2018. 2(4(92)). P. 15 – 25. doi: 10.15587/1729-4061.2018.126287
5. Lizunov P., Biloshchytskyi A., Kuchansky A., Andrashko Yu., Biloshchytska S. Improvement of the method for scientific publications clustering based on n-gram analysis and fuzzy method for selecting research partners. *Eastern-European Journal of Enterprise Technologies*. 4(4(100)). 2019. P. 6–14. doi: <https://doi.org/10.15587/1729-4061.2019.175139>
6. Biloshchytskyi A., Myronov O., Reznik R., Kuchansky A., Andrashko Yu., Paliy S., et al. A method to evaluate the scientific activity quality of HEIs based on a scientometric subjects presentation model. *Eastern-European Journal of Enterprise Technologies*. 2017. Vol. 6, Issue 2 (90). – P. 16–22. doi: 10.15587/1729-4061.2017.118377
7. Biloshchytskyi A., Kuchansky A., Andrashko Yu., Biloshchytska S., Dubnytska A., Vatskel V. The Method of the Scientific Directions Potential Forecasting in Infocommunication Systems of an Assessment of the Research Activity Results. 2017 *IEEE International Conference «Problems of Infocommunications. Science and Technology» (PIC S&T)*. 2017. P. 69–72. doi: 10.1109/INFOCOMMST.2017.8246352
8. Білоцицький А.О., Кучанський О.Ю., Андрашко Ю.В., Білоцицька С.В., Кузка О.І. Концептуальна модель інформаційної технології оцінювання результатів науково-дослідної роботи. *Управління розвитком складних систем*. 2017. № 30. С. 163 – 168.
9. Biloshchytskyi A., Kuchansky A., Andrashko Yu., Paliy S., Biloshchytska S., Bronin S., et al. Development of technical component of the methodology for project-vector management of educational environment. *Eastern-European Journal of Enterprise Technologies*. 2018. Vol. 2/2 (92). P 4–13. doi: 10.15587/1729-4061.2018.126301
10. Moed H., Glanzel W., Schmoch U. *Handbook of Quantitative Science and Technology Research*. Springer Netherlands. 2004. 800 p. ISBN 978-1-4020-2755-0.
11. Ying Ding. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics*. 2011. 5 (1). С. 187 – 203. doi: 10.1016/j.joi.2010.10.008
12. Hou H., Kretschmer H., Liu Z. The Structure of Scientific Collaboration Networks in Scientometrics. *Scientometrics*. 2008. 75. С. 189 – 202.
13. Barabási A.-L. *The New Science of Networks*. Perseus Books Group. 2002, 288 p.
14. Kobourov S.G. *Spring Embedders and Force-Directed Graph Drawing Algorithms*. Biocode. 2012. URL: <https://arxiv.org/abs/1201.3011>
15. Long J.C., Cunningham F.C., Carswell P., Braithwaite J. Who are the key players in a new translational research network? *BMC Health Serv Res*. 2013. 13:338. doi: 10.1186/1472-6963-13-338
16. Øystein O. *Theory of graphs*. Colloquium Publications, American Mathematical Society. 1967. 104 p.
17. Bavelas A. Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*. 1950. 22. С. 725 – 730.
18. Marchiori M., Latora V. Harmony in the small-world. *Physica A: Statistical Mechanics and its Applications*. 2000. 285 (3 – 4). P. 539 – 546. doi: 10.1016/s0378-4371(00)00311-3
19. Newman M.E. *Networks: An Introduction*. Published to Oxford Scholarship. 2010. 784 p. doi: 10.1093/acprof:oso/9780199206650.003.0006
20. Brandes U., Delling D., Gaertler M., Gorke R., Hoefer M., Nikoloski Z., Wagner, D. On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*. 20(2). 2008. P. 172–188. doi: 10.1109/TKDE.2007.190689.
21. Watts D.J., Strogatz S. Collective dynamics of 'small-world' networks. *Nature*. 393. 1998. – P. 440 – 442.
22. Kemper A. *Valuation of Network Effects in Software Markets: A Complex Networks Approach (Contributions to Management Science) 2010 Edition, Kindle Edition*. 2010. – 330 p.
23. Egghe L., Rousseau R. *Introduction to Infometrics*. Amsterdam: Elsevier. 1990. – 450 p.
24. Manning C.D., Raghavan P., Schütze H. *Introduction to Information Retrieval*. Cambridge University Press. Retrieved 2008. – 506 p.
25. Leskovec J., Rajaraman A., Ullman J.D. *Mining of Massive Datasets 2nd Edition, Kindle Edition*. – 2014. – 480 p.
26. Yudhoatmojo S.B., Samuar M.A. Community detection on citation network of dblp data sample set using linkrank algorithm. *Procedia Computer Science*. 2017. 124. – P. 29 – 37.
27. Hu Y. Efficient and high quality forcedirected graph drawing. *Mathematica Journal*. 2005. 10. – P. 37 – 71.
28. Small H. "Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 1973, 24, 265 – 269.

29. Garfield E. *From Bibliographic Coupling to Co-Citation Analysis Via Algorithmic Historio-Bibliography. A citationist's tribute to Belver C. Griffith, Presented at Drexel University, Philadelphia, PA. 2001.* URL: <https://garfield.library.upenn.edu/papers/drexelbelvergriffith92001.pdf>
30. Porter A.L., Rafols I. *Is science becoming more interdisciplinary? Measuring and mapping six research fields over time.* *Scientometrics.* 2009. 81(3). – P. 719 – 745.
31. National Academies of Sciences Engineering Medicine. *Facilitating Interdisciplinary Research.* URL: <https://www.nap.edu/catalog/11153/facilitating-interdisciplinary-research>
32. Bhattacharya S., Basu P. *Mapping a research area as the micro level using co-word analysis.* *Scientometrics.* 1998. 43(3). – P. 359 – 372. doi:10.1007/BF02457404
33. Glänzel W. *Bibliometric methods for detecting and analysing emerging research topics.* *El profesional de la informacion.* 2012. 21(2). – P. 194 – 201.
34. Mulesa O., Geche F., Batyuk A. *Information technology for determining structure of social group based on fuzzy c-means.* *Proceedings of the International Conference on Computer Sciences and Information Technologies (CSIT).* 2015. – P. 60 – 62. doi: 10.1109/STC-CSIT.2015.7325431
35. Shvets A., Devyatkin D., Sochenkov I., Tikhomirov I., Popov K., Yarygin K. *Detection of current research directions based on full-text clustering.* *Proceedings of the Science and Information Conference (SAI).* 2015. – P. 152 – 156. doi: 10.1109/SAI.2015.7237186
36. Samatova N., Hendrix W., Jenkins J., Padmanabhan K., Chakraborty A. *Practical Graph Mining with R Chapman and Hall/CRC.* – 2013. – 495 p.
37. Blondel V., Guillaume J., Lambriotte R., Lefebvre E. *Fast unfolding of communities in large networks.* *Journal of Statistical Mechanics: Theory and Experiment.* 2008, P10008. doi:10.1088/1742-5468/2008/10/P10008
38. Seifi M., Guillaume J. *Community codes in evolving networks.* *2012 International World Wide Web Conference (WWW).* 2012. – P. 1173 – 1180. doi:10.1145/2187980.2188258
39. Ovelgönne M., Geyer-Schulz A. *An ensemble learning strategy for graph clustering.* *Contemporary Mathematics.* 2013. – 588. P. 187–205. doi:10.1090/conm/588/11701
40. Zhang T., Ramakrishnan R., Livny M. *BIRCH: an efficient data clustering method for very large databases.* *Proceedings of the 1996 ACM SIGMOD international conference on Management of data.* 1996. 25(2). – P. 103–114. doi:10.1145/233269.233324
41. Otradsкая T., Gogunskii V. *Development process models for evaluation of performance of the educational establishments.* *Eastern-European Journal of Enterprise Technologies.* – 2016. – 3(81). P. 12 – 21.
42. Otradsкая T., Gogunskii V., Antoschuk S., Kolesnikov O. *Development of parametric model of prediction and evaluation of the quality level of educational institutions.* *Eastern-European Journal of Enterprise Technologies.* – 2016. – 5(3(83)). – P. 12 – 21. doi:10.15587/1729-4061.2016.80790
43. Blei D.M., Ng A.Y., Jordan M.I. *Latent Dirichlet Allocation.* *Journal of Machine Learning Research,* 2003, 3, 993 – 1022.
44. Pawar R.S., Sobhgol S., Durand G.C., Pinnecke M., Broneske D., Saake G. *Codd's World: Topics and their Evolution in the Database Community Publication Graph.* *Grundlagen von Datenbanken.* 2018. – P. 74 – 81.
45. Coccia M. *General properties of the evolution of research fields: a scientometric study of human microbiome evolutionary robotics and astrobiology.* *Scientometrics.* 2018. 117(2). – P. 1265 – 1283. doi: 10.1007/s11192-018-2902-8
46. Higham K. W., Governale M., Jaffe B., Zulicke U. *Unraveling the dynamics of growth, aging and inflation for citations to scientific articles from specific research fields.* *Journal of Informetrics.* 2017. 11. – P. 1190 – 1200.

Стаття надійшла до редакції 28.01.2020

#### Kuchansky Alexander

PhD (Eng.), Associate Professor, Department of Information Systems and Technologies, [orcid.org/0000-0003-1277-8031](https://orcid.org/0000-0003-1277-8031)  
Taras Shevchenko National University of Kyiv, Kyiv

### COMPONENTS OF SCIENTIFIC NETWORK ANALYSIS

**Abstract.** *Scientific spaces research is a complex task that must take into account the dynamics of the development of their components, as well as account for the numerous changes that result from the increase in the speed of information production and its dissemination. The analysis of the associations of subjects and objects of the scientific space, including networks of scientific cooperation, citation networks and joint citation, is an urgent task of the research since it is the basis for the correct evaluation of their R&D indicators. The paper gives an overview of the problems arising in the analysis of scientific networks: research of scientific cooperation networks, research of citation networks, joint citation and bibliographic connections, identification of directions of scientific research and analysis of topics, construction of scientific networks based on common keywords, research of heterogeneous scientific networks. The globalization of the dissemination and the dynamics of the scientific network's development determine the approaches to their analysis. Study of scientific networks allows to understand the structure of interaction of subjects and objects of scientific networks (scientists, HEI and their structural divisions), as well as to conduct a thorough evaluation of their research activities. The paper presents a visualization of the scientific cooperation network within the*

framework of the implementation of four budget research topics and one Erasmus + KA2 project for the period from 2015 to 2020. The visualization performed using the power algorithm of the Fruchterman-Rheingold viewing. The citation network of scientists's publications based on the Yifan Hu visualization algorithm was presented. The results of the evaluation can be used both by scientists and educational institutions to monitor the dynamics of research activities. In particular, the Ministry of Education and Science of Ukraine, as well as by local self-government bodies, to financially stimulate specific areas of research that have positive dynamics and development prospects.

**Keywords:** scientific network analysis; scientific collaboration network; citation network; scientometrics; scientific space

### References

1. Kremen V., Bykov, V. (2013). Space and environment categories: Features of model presentation and educational use. *The theory and practice of social systems management*, 2, 3 – 16. [In Ukrainian]
2. Gipp, B. & Beel, J. (2009). Citation Proximity Analysis (CPA) – A New Approach for Identifying Related Work Based on Co-Citation Analysis. *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, Rio de Janeiro, Brazil, 571 – 575.
3. Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., Biloshchytska, S., Kuzka, O. & Terentyev, O. (2017). Evaluation methods of the results of scientific research activity of scientists based on the analysis of publication citations. *Eastern-European Journal of Enterprise Technologies*, 3(2 (87)), 4–10. doi: 10.15587/1729-4061.2017.103651
4. Kuchansky, A., Andrashko, Yu., Biloshchytskyi, A., Danchenko, O., Ilarionov, O., Vatskel, I. & Honcharenko, T. (2018). The method for evaluation of educational environment subjects' performance based on the calculation of volumes of M-simplexes. *Eastern-European Journal of Enterprise Technologies*, 2(4(92)), 15– 5. doi: 10.15587/1729-4061.2018.126287
5. Lizunov, P., Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu. & Biloshchytska, S. (2019). Improvement of the method for scientific publications clustering based on n-gram analysis and fuzzy method for selecting research partners. *Eastern-European Journal of Enterprise Technologies*. 4(4(100)), 6–14. doi: <https://doi.org/10.15587/1729-4061.2019.175139>
6. Biloshchytskyi, A., Myronov, O., Reznik, R., Kuchansky, A., Andrashko, Yu., Paliy, S., et al. (2017). A method to evaluate the scientific activity quality of HEIs based on a scientometric subjects presentation model. *Eastern-European Journal of Enterprise Technologies*, 6, 2 (90), 16–22. doi: 10.15587/1729-4061.2017.118377
7. Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., Biloshchytska, S., Dubnytska, A., Vatskel, V. (2017). The Method of the Scientific Directions Potential Forecasting in Infocommunication Systems of an Assessment of the Research Activity Result. *2017 IEEE International Conference «Problems of Infocommunications. Science and Technology» (PIC S&T)*, 69–72. doi: 10.1109/INFOCOMMST.2017.8246352
8. Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., Biloshchytska, S. & Kuzka, O. (2017). The conceptual model of information technology evaluated the results of the research work. *Management of Development of Complex Systems*, 30, 163–168. [In Ukrainian]
9. Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., Paliy, S., Biloshchytska, S., Bronin, S. et al. (2018). Development of technical component of the methodology for project-vector management of educational environment. *Eastern-European Journal of Enterprise Technologies*, 2/2 (92), 4–13. doi: 10.15587/1729-4061.2018.126301
10. Moed, H., Glanzel, W. & Schmoch, U. (2004). *Handbook of Quantitative Science and Technology Research*. Springer Netherlands, 800. ISBN 978-1-4020-2755-0.
11. Ying, Ding. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics*, 5 (1), 187–203. doi: 10.1016/j.joi.2010.10.008
12. Hou, H., Kretschmer, H. & Liu, Z. (2008). The Structure of Scientific Collaboration Networks in Scientometrics. *Scientometrics*, 75, 189-202.
13. Barabási, A.-L. (2002). *The New Science of Networks*. Perseus Books Group, 288.
14. Kobourov, S.G. (2012). *Spring Embedders and Force-Directed Graph Drawing Algorithms*. Biocode, URL: <https://arxiv.org/abs/1201.3011>
15. Long, J.C., Cunningham, F.C., Carswell, P. & Braithwaite, J. (2013). Who are the key players in a new translational research network? *BMC Health Serv Res*, 13, 338. doi: 10.1186/1472-6963-13-338
16. Øystein, O. (1967). *Theory of graphs*. Colloquium Publications, American Mathematical Society, 104.
17. Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, 22, 725 – 730.
18. Marchiori, M. & Latora, V. (2000). Harmony in the small-world. *Physica A: Statistical Mechanics and its Applications*, 285 (3–4), 539–546. doi: 10.1016/s0378-4371(00)00311-3
19. Newman, M.E. (2010). *Networks: An Introduction*. Published to Oxford Scholarship, 784. doi: 10.1093/acprof:oso/9780199206650.003.0006
20. Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefler, M., Nikoloski, Z. & Wagner, D. (2008). On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20(2), 172–188. doi: 10.1109/TKDE.2007.190689.
21. Watts, D.J. & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393, 440 – 442.
22. Kemper, A. (2010). *Valuation of Network Effects in Software Markets: A Complex Networks Approach (Contributions to Management Science) 2010 Edition*, Kindle Edition, 330.
23. Egghe, L. & Rousseau, R. (1990). *Introduction to Infometrics*. Amsterdam: Elsevier, 450.

24. Manning, C.D., Raghavan, P. & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. Retrieved, 506.
25. Leskovec, J., Rajaraman, A. & Ullman, J.D. (2014). *Mining of Massive Datasets 2nd Edition, Kindle Edition*, 480.
26. Yudhoatmojo, S.B. & Samuar, M.A. (2017). Community detection on citation network of dblp data sample set using linkrank algorithm. *Procedia Computer Science*, 124, 29 – 37.
27. Hu, Y. (2005). Efficient and high quality forcedirected graph drawing. *Mathematica Journal*, 10, 37 – 71.
28. Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265 – 269.
29. Garfield, E. (2001). *From Bibliographic Coupling to Co-Citation Analysis Via Algorithmic Historio-Bibliography*. A citationist's tribute to Belver C. Griffith, Presented at Drexel University, Philadelphia, PA, URL: <https://garfield.library.upenn.edu/papers/drexelbelvergriffith92001.pdf>
30. Porter, A.L. & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3), 719 – 745.
31. National Academies of Sciences Engineering Medicine. *Facilitating Interdisciplinary Research*. URL: <https://www.nap.edu/catalog/11153/facilitating-interdisciplinary-research>
32. Bhattacharya, S. & Basu, P. (1998). Mapping a research area as the micro level using co-word analysis. *Scientometrics*, 43(3), 359–372. doi:10.1007/BF02457404
33. Glänzel, W. (2012). *Bibliometric methods for detecting and analysing emerging research topics*. *El profesional de la informacion*, 21(2), 194 – 201.
34. Mulesa, O., Geche, F. & Batyuk, A. (2015). Information technology for determining structure of social group based on fuzzy c-means. *Proceedings of the International Conference on Computer Sciences and Information Technologies (CSIT)*, 60–62. doi: 10.1109/STC-CSIT.2015.7325431
35. Shvets, A., Devyatkin, D., Sochenkov, I., Tikhomirov, I., Popov, K. & Yarygin, K. (2015). Detection of current research directions based on full-text clustering. *Proceedings of the Science and Information Conference (SAI)*, 152–156. doi: 10.1109/SAI.2015.7237186
36. Samatova, N., Hendrix, W., Jenkins, J., Padmanabhan, K. & Chakraborty, A. (2013). *Practical Graph Mining with R* Chapman and Hall/CRC, 495.
37. Blondel, V., Guillaume, J., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, P10008. doi:10.1088/1742-5468/2008/10/P10008
38. Seifi, M. & Guillaume, J. (2012). Community codes in evolving networks. *2012 International World Wide Web Conference (WWW)*, 1173–1180. doi:10.1145/2187980.2188258
39. Ovelgönne, M. & Geyer-Schulz, A. (2013). An ensemble learning strategy for graph clustering. *Contemporary Mathematics*, 588, 187–205. doi:10.1090/conm/588/11701
40. Zhang, T., Ramakrishnan, R. & Livny, M. (1996). BIRCH: an efficient data clustering method for very large databases. *Proceedings of the 1996 ACM SIGMOD international conference on Management of data*, 25 (2), 103 – 114. doi:10.1145/233269.233324
41. Otradsкая, T. & Gogunskii, V. (2016). Development process models for evaluation of performance of the educational establishments. *Eastern-European Journal of Enterprise Technologies*, 3(81), 12–21. doi:10.15587/1729-4061.2016.66562
42. Otradsкая, T., Gogunskii, V., Antoschuk, S. & Kolesnikov, O. (2016). Development of parametric model of prediction and evaluation of the quality level of educational institutions. *Eastern-European Journal of Enterprise Technologies*, 5(3(83)), 12 – 21. doi:10.15587/1729-4061.2016.80790
43. Blei, D.M., Ng, A.Y. & Jordan, M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993 – 1022.
44. Pawar, R.S., Sobhghol, S., Durand, G.C., Pinnecke, M., Broneske, D. & Saake, G. (2018). *Codd's World: Topics and their Evolution in the Database Community Publication Graph*. *Grundlagen von Datenbanken*, 74 – 81.
45. Coccia, M. (2018). General properties of the evolution of research fields: a scientometric study of human microbiome evolutionary robotics and astrobiology. *Scientometrics*, 117(2), 1265-1283. doi: 10.1007/s11192-018-2902-8
46. Higham, K. W., Governale, M., Jaffe, B. & Zulicke, U. (2017). Unraveling the dynamics of growth, aging and inflation for citations to scientific articles from specific research fields. *Journal of Informetrics*, 11, 1190 – 1200.

#### Посилання на публікацію

- APA Kuchansky, Alexander. (2020). Components of scientific network analysis. *Management of development of complex systems*, 41, 115 – 126, [dx.doi.org/10.32347/2412-9933.2020.41.115-126](https://doi.org/10.32347/2412-9933.2020.41.115-126).
- ДСТУ Кучанський О.Ю. Складові аналізу наукових мереж [Текст] / О.Ю. Кучанський // Управління розвитком складних систем. – № 41. – 2020. – С. 115 – 126, [dx.doi.org/10.32347/2412-9933.2020.41.115-126](https://doi.org/10.32347/2412-9933.2020.41.115-126).