

## Правова інформатика

УДК 004.622:004.822

ЛАНДЕ Д.В., доктор технічних наук,

АНДРУЩЕНКО В.Б., аспірант,

Інститут проблем реєстрації інформації НАН України

### ПОБУДОВА МЕРЕЖ СПІВАВТОРСТВА ФАХІВЦІВ З ЮРИСПРУДЕНЦІЇ ЗА ДАНИМИ СЕРВІСУ GOOGLE SCHOLAR CITATIONS

**Анотація.** У роботі наводиться алгоритм побудови мережі співавторства вчених в області юриспруденції. На основі зондування наукометричних мереж пропонується методика побудови мереж співавторства – моделей співробітництва вчених. Як така мережа в роботі розглядається мережа понять, відповідних тегам сервісу Google Scholar Citations. Показано, що кластери в мережах співавторства можуть розглядатися як основа для виявлення наукових шкіл.

**Ключові слова:** мережа співавторства, наукометрія, Google Scholar Citations, предметна область, граф зв'язків, зондування мереж.

**Аннотация.** В работе приводится алгоритм построения сети соавторства ученых в области юриспруденции. Предлагается методика построения сетей соавторства – моделей сотрудничества ученых на основе зондирования наукометрических сетей. Как такая сеть в работе рассматривается сеть понятий, соответствующих тегам сервиса Google Scholar Citations. Показано, что кластеры в сетях соавторства могут рассматриваться как основа для выявления научных школ.

**Ключевые слова:** сеть соавторства, наукометрия, Google Scholar Citations, предметная область, граф связей, зондирование сетей.

**Summary.** The algorithm of creation of the network of a co-authorship of scientists in the field of law regulated by their scientific interests is given in work. The network of a co-authorship is formed on the basis of exploration of the Google Scholar Citations service. It is shown that clusters in networks of a co-authorship can be considered as a basis for identification of schools of sciences.

**Keywords:** co-authorship network, scientometrics, Google Scholar Citations, subject area, domain model, network sensing, information network.

**Постановка проблеми.** На сьогоднішній день заходи щодо підтримки науки висувають жорсткі вимоги до досягнень науковця або наукової спільноти. Наукометричні показники є одним із чинників, що визначають впливовість науковця чи наукового колективу у дзеркалі світової та вітчизняної науки. Також наукометричні показники є одним із параметрів пошуку впливових публікацій та відповідно достовірних та актуальних результатів досліджень за тією чи іншою тематикою. Із започаткуванням систем європейських та світових програм грантової підтримки наукових проєктів, що передбачають участь у конкурсах міжнародних колаборацій, є необхідним доступ до інформації, яка може стати вирішальною при прийнятті рішення щодо надання гранту.

З огляду на наявність великої кількості наукометричних, реферативних та препринт ресурсів в глобальних інформаційних мережах можна зауважити, що не вся інформація є: доступною, достовірною, достатньою. Саме тому значної необхідності

набуває не створення нових ресурсів, а розширення можливостей існуючих наукометричних сервісів. Розширення можливостей передбачає побудову сервісів, що дозволяють будувати мережу предметної області та мережі співавторів за рахунок доступної інформації.

Як наслідок розвитку сервісів наукової інформації, з'явилися нові можливості оцінки наукової інформації і вивчення закономірностей наукового взаємодії [1]. Основним інструментом вивчення закономірностей наукового співробітництва є мережі співавторів, за допомогою яких можна отримати не тільки наукометричні оцінки, але і визначати експертів для вирішення складних завдань. Одним з великих сервісів наукової інформації є Google Scholar Citation, який дозволяє створювати вченим профілі, що містять відповідну бібліографічну інформацію, а також здійснювати пошук публікацій з усього світу.

Вивченню мереж співавторів, так само як і сервісу Google Scholar Citation (<http://scholar.google.com/citations>), присвячена велика кількість робіт, що підтверджує актуальність проведених досліджень [2]. Серед них методи побудови мереж співавторів, визначення значущих вузлів, структури мережі, дослідження цитування в Google Scholar, а так само відповідних корпусів ін. [3].

**Метою статті** є визначення структури наукометричного сервісу Google Scholar Citations та його можливостей. Зазначене передбачає дослідити спільне та різне у способі представлення та механізмі організації пошуку інформації, і можливість застосування унікального алгоритму для розширення існуючих можливостей систем, а також здійснити опис теоретичних принципів і методології автоматизованого формування мереж співавторства, зокрема, галузі юриспруденції шляхом зондування великої інформаційної мережі.

**Виклад основних положень.** Для досягнення вказаної вище мети розроблено спеціальний алгоритм сканування ресурсів сервісу Google Scholar Citations який передбачає отримання репрезентативного набору співавторів як основи (вузлів) майбутньої мережі. Під зондуванням мереж будемо розуміти вибірку невеликого обсягу найважливішого змісту з великих мереж, які з технологічних причин не підлягають повному скануванню [4; 5].

Очевидно, мережа співавторства може мати досить великі розміри, якщо її не обмежувати певною тематикою. Ця властивість значно ускладнює сприйняття сформованої мережі і призводить до такого ефекту, як дрейф тематики. Для подолання цього ефекту застосовується тематична фільтрація, тобто використовуються дескриптори, які приписуються авторам наукометричної мережі, що визначають їх тематичну спрямованість. Відповідно до цих дескрипторів і визначають розмір сформованих мереж співавторства, а також динаміку їх зростання. Крім того, розпізнавання кластерів в таких мережах може розглядатися як основа для виявлення наукових шкіл, експертних груп тощо [6].

**Методика досліджень.**

До розгляду було взято систему, що є доступною в глобальній мережі і не передбачає передплати: Google Scholar Citations (Google). Для систем було застосовано спеціальні теги з юридичної проблематики для пошуку. З огляду на застосування тегів, було визначено представлення інформації наукометричною системою, і, в той же час, зауважені обмеження у доступі інформації.

Авторами запропоновано алгоритм, який адаптувався до реальної мережі співавторів з юриспруденції сервісу Google Scholar Citations [7]. Цей алгоритм складається з наступних кроків:

Крок 1. Обирається перший автор (вузол), з якого починається зондування.

Крок 2. Експертним шляхом визначається невеликий перелік базових тегів-дескрипторів, що відповідають найважливішим поняттям, наприклад, теги law, judicial, criminology тощо у прикладі, що розглядається нижче.

Крок 3. Відкривається сторінка веб-сервісу, що відповідає обраному автору. До створюваної мережі додаються всі співавтори, що містяться на сторінці обраного учасника. Формуються ребра-зв'язки до цих вузлів (співавторів) з вихідного вузла (автора).

Крок 4. Із списку вузлів мережі, що формується, випадковим чином обирається той, на сторінку якого планується перейти для подальшого аналізу. Цей вузол також повинен задовольняти тематиці обраної предметної області (його теги входять до складу дескрипторів, визначених на кроці 2, і не входять до складу тих вузлів, до сторінок яких вже був здійснений перехід).

Крок 5. Якщо такий вузол-автор обрано, то здійснюється перехід до кроку 3.

Крок 6. Якщо такого вузла не існує, то вважається, що мережу зондування побудовано.

Відповідно до цього алгоритму процес зондування мережі, починаючи з певного вузла, припиняється при “зациклюванні”, тобто коли відповідно до алгоритму має відбуватися перехід до вже пройденого вузла, а також при відхиленні сусідніх вузлів від основної тематики (визначається з урахуванням лексичного складу тегів).

*Отримані результати.*

Побудована відповідно до наведеного алгоритму мережа співавторів за базовим дескриптором law при досить широкому списку дескрипторів-обмежень (judicial, criminology, politic, social, sociology, legal) і обмеженні на кількість сканованих вузлів у 1000. За допомогою програмного засобу Gephi отримана візуалізація даної мережі співавторів (Рис. 1).

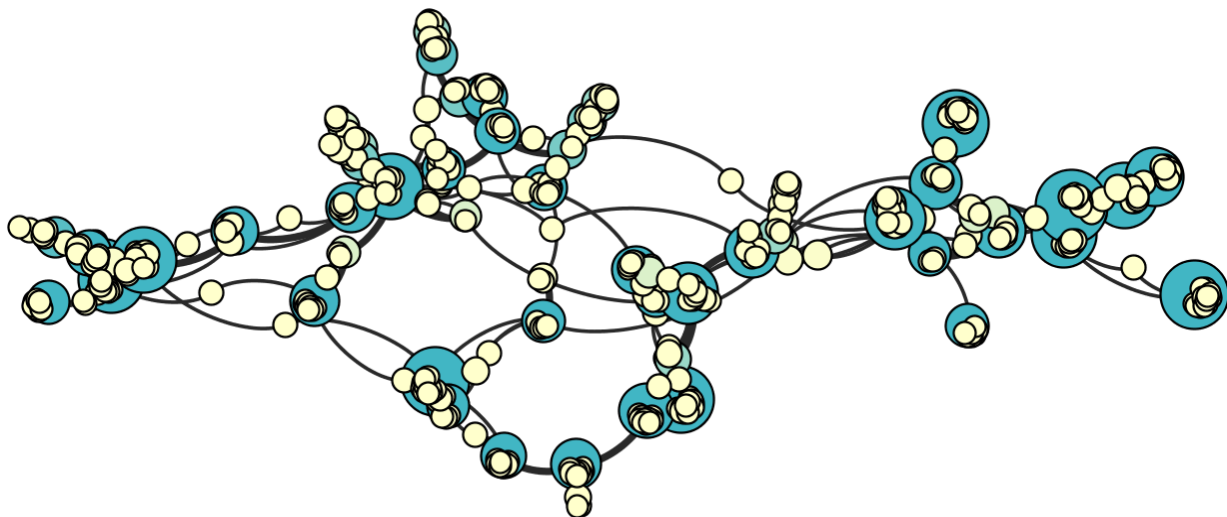


Рис. 1. Фрагмент мережі співавторів, яка побудована з урахуванням дескрипторів широкої тематики

У розглянутому на Рис. 2 випадку наведений вище алгоритм застосовується до групи науковців, які мають відповідно до даних Google Scholar Citations найбільше цитування.

Застосування методів кластерного аналізу дозволяє виявляти найбільш тісно пов'язані між собою групи вчених-співавторів, наукових шкіл, експертних груп. У цьому випадку під науковою школою будемо розуміти неформальний творчий колектив дослідників різних поколінь, об'єднаних загальною програмою і стилем дослідницької

роботи, які діють під керівництвом визнаного лідера. На Рис. 3 показано приклад процесу виявлення кластерів шляхом покрокового видалення найменш вагомих ребер з мережі співавторства, побудованої за тематикою юриспруденції.

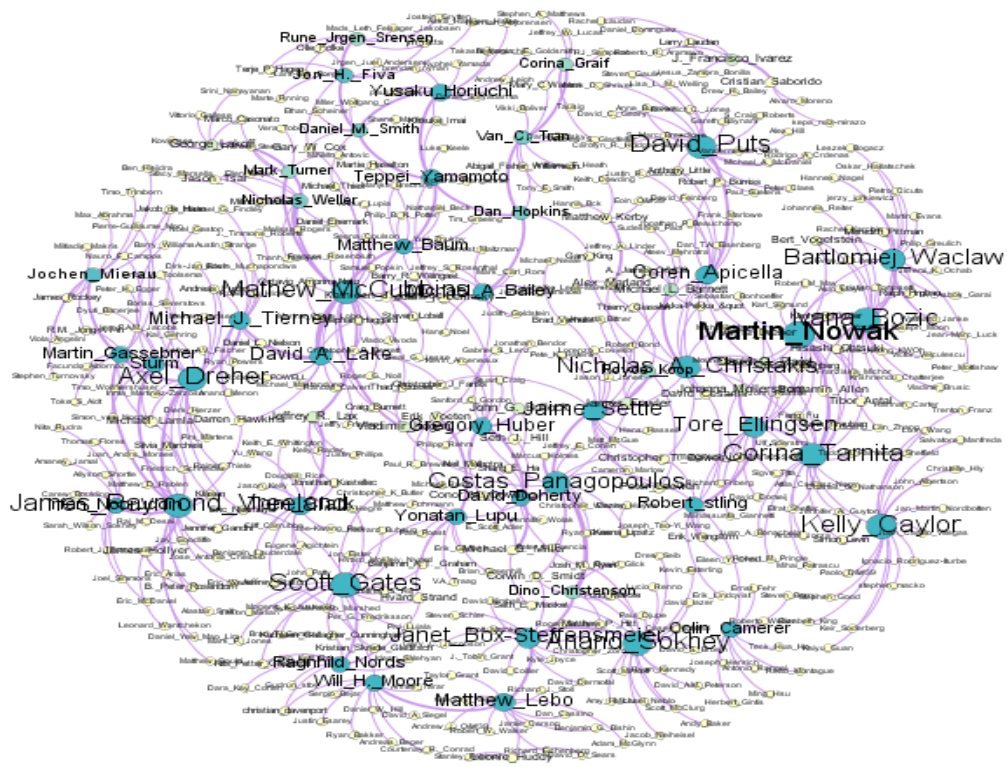


Рис. 2. Мережа співавторів за заданим тегом

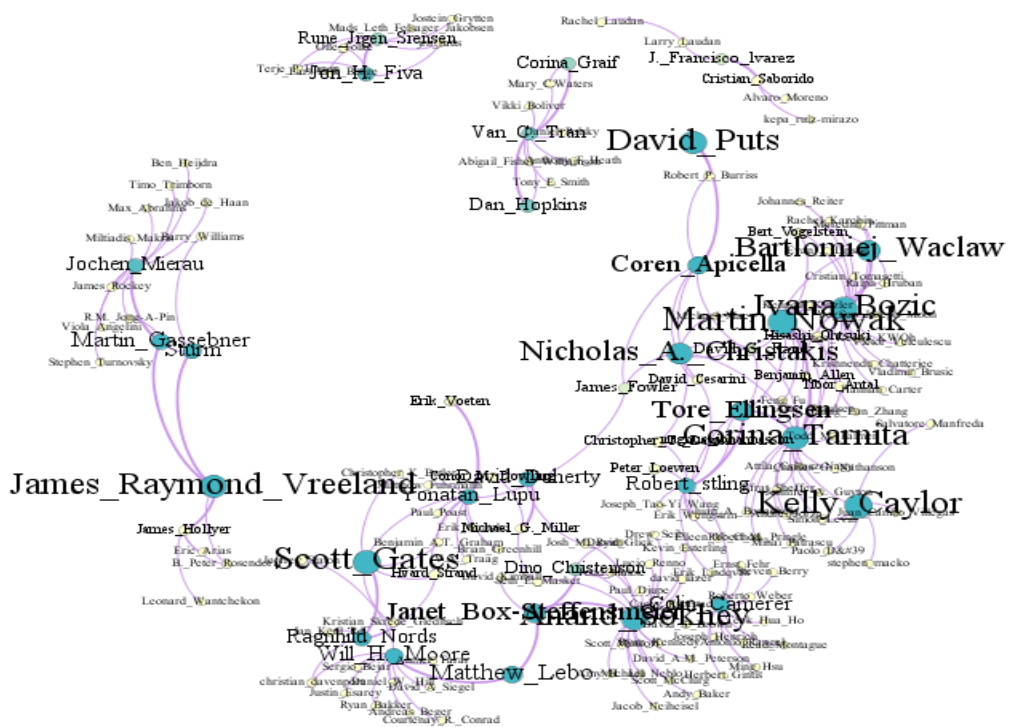


Рис. 3. Кластеризація вихідної мережі шляхом видалення ребер засобами програми Gephi

Слід відзначити принципову відмінність запропонованої моделі автоматичного формування мереж співавторства від існуючих, що базуються на особистій участі експертів при виборі конкретних вузлів і зв'язків. В даному випадку дослідник для побудови мережі використовує лише крупіці знань, представлених у вигляді набору базових тегів. Надалі програма використовує знання, закладені співавторами, теги позначені як головні для них. Тобто експертна середа в цьому випадку істотно розширюється.

Можна зауважити, що система Google Scholar Citations є зручною щодо доступу до інформації, не передбачає створення власного профілю користувача для доступу до інформації, доступ є необмежений, порівняно із зареєстрованими у системах користувачами. І в той же час кожен зареєстрований користувач отримує додаткові пріоритети у вигляді можливості збереження отриманої і скорегованої інформації. При цьому пошук в системі Google Scholar є спрощеним, на відміну, наприклад, від сервісу Science Direct, що пропонує користувачеві одразу кілька параметрів для здійснення пошуку. Можливості уточнення пошуку за тегом є більш орієнтованими на користувача і дозволяють за рахунок чітких уточнень швидше отримати необхідну інформацію.

### **Висновки.**

У роботі представлено методіку побудови мереж співавторства – моделей співробітництва вчених на основі зондування наукометричних мереж. Як така мережа в роботі розглядається мережа понять, що відповідають тегам сервісу Google Scholar Citations. Запропоновано і реалізовано підхід до формування мереж співавторства в рамках предметної області, обмежувальними елементами якого є деякі маркери знань (теги), задалегідь задані вченими – учасниками проекту Google Scholar Citations.

Модель застосовувалася для юридичної науки в рамках сервісу Google Scholar Citations, але запропонований підхід можна використовувати і для інших наукових областей, або для інших наукометричних масивів.

Враховуючи дослідження запропонованого алгоритму для системи Google Scholar Citations, постає питання застосування цього алгоритму для інших наукометричних систем, що потребує подальшого проведення порівняльного аналізу ресурсів.

### **Використана література**

1. Ortega J. How is an academic social site populated ? A demographic study of Google Scholar Citations population // *Scientometrics*. – 2015. – 104. – P. 1-18.
2. Liu J., Li Y., Ruan Z., Fu G., Chen X., Sadiq, Deng Y. A new method to construct co-author networks // *Physica A*. – 2015. – 419. – P. 29-39.
3. Brezina V. Use of Google Scholar in corpus-driven EAP research // *Journal of English for Academic Purposes*. – 2012. – 11. – P. 319-331.
4. Ланде Д.В. Створення термінологічної моделі предметної області шляхом зондування Google Scholar Citations // *Правова інформатика*. – 2015. – № 2(45). – С. 3-8.
5. Ландэ Д.В. Построение модели предметной области путем зондирования сервиса Google Scholar Citations // *Онтология проектирования*. – 2015. – Т 5. – № 3 (17). – С. 328-335.
6. Ландэ Д.В., Горбов И.В., Балагура И.В. Характеристики сети соавторов медицинских наук // *Клиническая информатика и телемедицина*. – 2013. – Т. 9. – Вып. 10. – С. 141-144.
7. Ландэ Д.В., Балагура И.В., Андрущенко В.Б. Построение сетей соавторства по данным сервиса Google Scholar Citations : материалы VI междунар. науч.-техн. конф. [“Открытые семантические технологии проектирования интеллектуальных систем” (OSTIS-2016)], (Минск, 18-20 февраля 2016 года). – Минск: БГУИР, 2016. – С. 233-237.

~~~~~ \* \* \* ~~~~~