

УДК 004.67+519.83

ЛАНДЕ Д.В., доктор технічних наук, старший науковий співробітник
БАЛОГУРА І.В., інженер Інституту проблем реєстрації інформації НАН України

ДОСЛІДЖЕННЯ МЕРЕЖ СПІВАВТОРСТВА У ПРАВОВІЙ НАУЦІ ПО БАЗІ ДАНИХ “УКРАЇНІКА НАУКОВА”

Анотація. В статті представлено деякі напрямки дослідження комунікаційних зв'язків співавторства в різних напрямках правової науки в Україні на базі застосування методології складних мереж. Запропоновано використання засобів оцінки монопольного становища в окремих напрямках правової науки, досліджено феномен “клубу багатих” для мереж співавторства. Показано, що для досліджуваних мереж спостерігається феномен “клубу багатих”, а для вузлів цих мереж виконується закономірність Парето-Лоткі.

Ключові слова: мережа співавторів, правова наука, феномен “клубу багатих”, реферативна база даних “Україніка наукова”, формат UKRMARC.

Аннотация. В статье представлены некоторые направления исследования коммуникационных связей соавторства в различных направлениях правовой науки в Украине на базе применения методологии сложных сетей. Предложено использование средств оценки монопольного положения в отдельных направлениях правовой науки, исследован феномен “клуба богатых” для сетей соавторства. Показано, что для исследуемых сетей наблюдается феномен “клуба богатых”, а для узлов этих сетей выполняется закономерность Парето-Лотки.

Ключевые слова: сеть соавторов, правовая наука, феномен “клуба богатых”, реферативная база данных “Украиника научная”, формат UKRMARC.

Summary. In the article some directions of research of communication connections of collaborators are presented in different directions of legal science in Ukraine on the basis of application of the methodology of complex networks. Proposed the use of facilities of estimation of monopolistic position is offered in separate directions of legal science, the “club rich” phenomenon for the co-authorship networks. It is shown that for the investigated networks there is the “club rich” phenomenon, and for the nodes of these networks conformity to law of Pareto-Lotka is executed.

Keywords: collaborators network, legal science, “rich club” phenomenon, the “Ukrainica naukova” database, UKRMARC format.

Постановка проблеми. Важливою та необхідною частиною наукового прогресу в правових науках є взаємодія науковців і практиків різних напрямів, яка виявляється, серед іншого, у мережах співавторства. Сумісне написання наукових праць відображує роботу, що націлена на досягнення єдиної цілі або розподілення зусиль та знань окремих науковців [1].

При вивченні наукової співпраці найчастіше використовують методи бібліо- та наукометрії, експертне оцінювання. У цьому сенсі мережі співавторства надають, можливо, найкраще комплексне бачення цілісного наукового процесу [2 – 4].

Зазначимо, що властивості наукового процесу в Україні значно відрізняються від зарубіжних країн, тому існує необхідність дослідити мережі співавторства в українських наукових виданнях та з їх допомогою визначити основні тенденції наукової співпраці в нашій країні.

Метою статті є дослідження проблеми комунікаційних зв'язків співавторства в різних напрямках правової науки в Україні.

Виклад основних положень. Для дослідження зв'язків співавторства у галузі правової науки було обрано наукометричну реферативну базу даних “Україніка наукова”, що входить до Системи реферування української наукової літератури [5]. В ній містяться реферати статей наукових періодичних видань та збірників, монографій, праць наукових конференцій, авторефератів дисертацій, довідників та словників, підручників для вищих навчальних закладів, що видається в Україні. Аналізувалися дані, що знаходились у базі даних за станом на листопад 2012 року, що становило понад 430 000 записів, серед яких до правової у відповідності з рубрикаторм Національної бібліотеки України ім. Вернадського (НБУВ) відносяться 28600 документів з рубрики “Держава і право. Юридичні науки”.

Реферативна база даних “Україніка наукова” є корпоративним проектом Інституту проблем реєстрації інформації НАН України та Національної бібліотеки України ім. Вернадського, крім них активну участь у реферуванні беруть Національна наукова медична бібліотека України та Державна науково-педагогічна бібліотека України ім. В.О. Сухомлинського. В рамках загального проекту також випускається реферативний журнал “Джерело” [6].

Відповідно з вимогами до можливостей обміну інформації в Україні та на міждержавному рівні, реалізовано необхідність використання комунікативного бібліотечного формату [7]. Дані з реферативної бази даних “Україніка наукова” представлені у форматі УКРМАРК, що є цифровим форматом представлення бібліографічних даних та українською версією формату UNIMARK [8] та зберігає його структуру [9].

Автоматизована обробка даних. Для проведення наукометричних досліджень на основі реферативної бази даних “Україніка наукова” було розроблено програмний комплекс фільтрації та аналізу даних, який забезпечує виділення тематичних фрагментів вихідної бази даних, окремих записів і полів. При аналізі файлу з даними реферативної бази даних станом на листопад 2012 року за правилами формату УКРМАРК було використано дані полів 70X, що позначають авторів. Для визначення тематики використовувалося поле 686 формату УКРМАРК, що містить індекс рубрикатора НБУВ.

Також було розроблено спеціальне програмне забезпечення для візуалізації мереж співавторства.

Ранговий розподіл кількості робіт, що приходяться на одного автора. У 1926 р. А. Дж. Лотка дослідив розподіл кількості наукових співробітників у залежності від числа опублікованих ними наукових робіт [10] – степеневу функцію:

$$y = f(x) = Cx^{-n}.$$

З іншого боку, розподіл кількості наукових робіт, що відповідають авторам відповідає виразу:

$$x = Dy = Dx^{-1/n}; \quad D = 1/C.$$

Було досліджено декілька фрагментів бази даних “Україніка наукова”, зокрема за напрямками правові науки. Було отримано розподіл кількості наукових робіт за авторами та експериментально підтверджено їх відповідність вищенаведеній закономірності (Рис. 1) із степеневими показниками 0,71.

Рис. 1 – Графік розподілу кількістю статей, що приходяться на автора, для галузі правових наук

У разі рівномірного розподілу присутності праць автора кожна група об’єктів зустрічається в базі даних кількість разів, пропорційну своїй чисельності. Такий випадок описується кривою рівності (line of perfect equality), прямою, що сполучає початок координат і точку (1;1). У разі повної нерівності (коли лише один об’єкт представлений в інформаційному просторі) крива (line of perfect inequality) спочатку “прилипає” до осі абсцис, а потім з точки (1;0) “злітає” до точки (1;1). Крива Лоренця поміщена між кривими рівності і нерівності (Рис. 2).

Рис. 2. Крива Лоренця

Крива Лоренця, індекси Джині, Гувера. Для незалежної цифрової оцінки значення авторів в мережі співавторства пропонуються підходи, що застосовуються в економічній науці при визначенні монополізму на ринках. Зупинимось на змісті зазначених індексів детальніше, маючи на увазі той факт, що там, де в “чистій” економіці застосовують термін “прибутки”, будемо розуміти рівень присутності праць автора в базі даних “Україніка наукова”.

Крива Лоренця [11] – це графічне зображення функції кумулятивного розподілу. Вона була запропонована американським економістом Максом Отто Лоренцом в 1905 році як показник нерівності в доходах населення. Крива Лоренця – це представлення функції кумулятивного розподілу, в якому акумулюються долі медіаприсутності. У прямокутній системі координат крива Лоренця є опуклою вниз і проходить під діагоналлю одиничного квадрата, розташованого в I координатних чвертях.

З кривої Лоренця можна вивести кількісні показники нерівності, наприклад коефіцієнт Джині та індекс Гувера (який ще називають індексом Робін Гуда).

Коефіцієнт Джині [12] – це статистичний показник, що свідчить про ступінь розшарування об’єктів по відношенню до деякої ознаки, що вивчається, наприклад, по рівню присутності праць автора в базі даних. Коефіцієнт Джині розраховується як відношення площі фігури, що утворюється кривою Лоренця і кривою рівності, до площі трикутника, утвореного кривими рівності і нерівності. У разі повної рівності коефіцієнт буде рівний 0; у разі повної нерівності він буде рівний 1. Коефіцієнт розраховується за формулою:

$$G = \frac{\sum_{i=1}^N \sum_{j=1}^N |y_i - y_j|}{2N^2 \bar{y}}$$

де: G – коефіцієнт Джині N – число авторів; y_k – частка ознак об’єкту (публікацій автора) k ; \bar{y} – середнє арифметичне значення кількості публікацій одного автора.

Індекс Гувера (Hoover index) [13], також відомий як індекс *Робін Гуда* (Robin Hood index), – це ще один показник нерівності по параметрах, що має зв’язок з кривою Лоренця. Він рівний тій частці параметра (кількості публікацій) в сукупності об’єктів (авторів), яку необхідно перерозподілити для досягнення рівності. Графічно він представимо як найдовший вертикальний відрізок, що сполучає фактичну криву Лоренця з лінією рівності (бісектрисою I координатних чвертей). Індекс Гувера також належить напіввідкритому інтервалу [0;1]. Для розглянутих фрагментів бази даних “Україніка наукова” були визначені індекси порівняння, значення яких наведено у Таблиці 1.

Таблиця 1. Індеси порівняння для деяких галузей правової науки

Галузь науки	Індекс Гувера	Коефіцієнт Джині
Адміністративне право	0.288	0.384
Цивільне і торгове право	0.320	0.413
Кримінальне право	0.323	0.419

Розподіл кількості наукових робіт за співавторами

У відповідності з [14] розподіл кількості наукових робіт, що відповідають кількості співавторів описується формулою:

$$y = f(x) = Cx^{-n} e^{x/x_c},$$

яка може розглядатися як узагальнення закономірності Лоткі-Парето для наукової співпраці вчених. Саме такому рівнянню задовольняють розподіли кількості наукових робіт за кількістю співавторів за названими вище напрямками – фрагментами бази даних “Україніка наукова” (Рис. 3).

Клуб багатих. Степеневий розподіл значень показників “клубу багатих”.

Отримані мережеві структури досліджувались на наявність феномену “клубу багатих”. Даний феномен спостерігається в складних мережах та проявляється у якості скупчення великої кількості зв’язків в окремих вузлах, тобто чим більша степінь вершини, тим більше зв’язків вона в подальшому буде отримувати.

Рис. 3. Розподіл кількості наукових робіт за співавторами у деяких напрямках правової науки у логарифмічній шкалі

Коефіцієнт “клубу багатих”, за визначенням авторів роботи [16], $\varphi(k)$ рівний сумі кількості наявних ребер графу E_k розділених на максимальну можливу кількість ребер, що можуть поєднувати k вершин:

$$\varphi(k) = \frac{2E_k}{k(k-1)}.$$

Апроксимаційні рівняння для значень коефіцієнтів “клубу багатих” для розглянутих вище напрямків правової науки наведено у таблиці 2. На Рис. 4 наведено графік розподілу значень показників “клубу багатих” для галузі правових наук у подвійній логарифмічній шкалі.

Рис. 4 – Степеневий розподіл значень показників “клубу багатих” для всієї галузі правових наук

Показано, що для мереж співавторства для розглянутих галузей науки спостерігається феномен “клубу багатих”, тобто автори, які мають найбільшу кількість робіт пишуть роботи переважно з такими ж солідними авторами.

Таблиця 2. Апроксимаційні рівняння для значень коефіцієнтів «клубу багатих»

Напрямок	Рівняння	Точність апроксимації
Цивільне і торгове право	$y = 1.5324x^{-0.8581}$	$R^2 = 0.9292$

Кримінальне право	$y = 0.6944x^{-0.7605}$	$R^2 = 0.9537$
Адміністративне право	$y = 0.2149x^{-0.6187}$	$R^2 = 0.95$

Діаграма мережі співавторів і усереднений коефіцієнт клубу багатих. Для визначення відношення між “багатими” та “бідними” вершинами було проведено вдосконалення вищенаведеного виразу – усереднений коефіцієнт “клубу багатих” RCC (Rich Club Coefficient) [17]. Запропоновано використовувати відношення коефіцієнту “клубу багатих” вершин з найбільшими степенями до аналогічного коефіцієнту вершин з найменшими степенями, також у чисельниках членів відношення додатково було використано одиницю, що виключає випадки ділення на 0:

$$RCC = \frac{2(E_{k_v} + 1)}{k(k-1)} : \frac{2(E_{k_n} + 1)}{k(k-1)} = (E_{k_v} + 1) / (E_{k_n} + 1),$$

де E_{k_v} – кількість ребер, що поєднують k вузлів з найбільшими степенями, E_{k_n} – кількість ребер, що поєднують k вузлів з найменшими степенями k – проміжок спостереження, що обирається, в нашому випадку $k = N / 5$ (N – кількість вузлів мережі співавторів).

Для побудови мережі наукової взаємодії будується мережева структура [17], в якій вузлам відповідають автори наукових публікацій, а ребра позначають наявність спільних публікацій між ними. При чому чим товща лінія ребра у цьому графі, тим більше спільних статей було написано авторами. Розміри вузлів також відповідають кількості спільних статей автора. Для наочності відображення на рис. 1 і 2 відображається лише фрагмент мережі співавторства, до якої входять 150 найбільш продуктивних авторів (Рис. 5).

У ході досліджень було створено програмний модуль, що виконує обчислення усередненого коефіцієнту «клубу багатих» на основі створених мереж співавторів. Для кожної галузі науки було побудовано окремі графи та розраховано усереднений коефіцієнт “клубу багатих”. Значення RCC для деяких напрямків правової науки наведено у Таблиці 3.

Таблиця 3. Значення RCC для окремих напрямків правової науки

Напрямок науки	RCC
Цивільне і торгове право	12,56
Кримінальне право	10,33
Адміністративне право	4.69

Рис. 5 – Фрагменти мережі співавторства (напрямок – цивільне і торгове право)

Висновки.

Аналізуючи мережі співавторів наукових робіт, що містяться в реферативній базі даних, можна зробити висновок, що вона не є суцільною, а має вигляд скупчень за окремими напрямками. Детально розглянуто деякі напрямки правової науки, серед них при найменшому впливі феномену “клубу багатих” найбільш комунікативно себе проявили автори, що досліджують адміністративне право, тобто виявлено саму розгалужену сітку взаємодії.

Разом з цим, визначено, що для всіх напрямків правової науки спостерігається феномен “клубу багатих”. Науковці, що мають найбільшу кількість робіт у співавторстві публікуються разом з такими ж продуктивними авторами. Проте значення вдосконаленого коефіцієнту “клубу багатих” для різних напрямків відрізняються.

За даними реферативної бази даних “Україніка наукова” правові науки мають найбільший коефіцієнт “клубу багатих” серед усіх галузей, при цьому наукові роботи з даної тематики відрізняються найменшою у середньому кількістю співавторів.

Для різних напрямків правової науки рівень “феномену “клубу багатих” значно різний, що може пояснюватися декількома причинами, наприклад: деякі напрямки фактично об’єднують лише формально пов’язані розділи знань, в межах яких працюють слабо пов’язані одна з однією групи учених; у деяких галузях науки активно працюють конкуруючі наукові школи; в окремих випадках еліта настільки сконцентрована, що не приймає у своє коло молоде поповнення .

Використана література

1. Alireza A. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks / Abbasi Alireza, Hossain Liaquat, Loet Leydesdorff / *Journal of Informetrics*. – 2012. – Vol. 6, № 3. – P. 403-412.
2. Liaquat H. The social networks of collaborative process / Hossain Liaquat, Daniel Fazio / *The Journal of High Technology Management Research*. – 2009. – Vol. 20, № 2. – P. 119-130.
3. Ding Y. Community detection: Topological vs. topical / Ying Ding / *Journal of Informetrics*. – 2011. – Vol. 5, № 4. – P. 498-514.
4. Chung K.H. On the relation between intellectual collaboration and intellectual output: Evidence from the finance academe / Kee H. Chung, Raymond A.K. Cox, Kenneth A. Kim / *The Quarterly Review of Economics and Finance*. – 2009. – Vol. 49, № 3. – P. 893-916.
5. Реферативна база даних “Україніка наукова”. – Режим доступу : <http://nbuv.gov.ua/db/ref.html>.
6. Система реферування української наукової літератури. – Режим доступу : <http://nbuv.gov.ua/db/ri.html>.
7. Значення видання українського реферативного журналу “Джерело” / [Крючин А.А. , Костенко Л.Й., Мініна Н.М., Балагура І.В., Овсієнко Л.М.]. – Режим доступу : http://www.nas.gov.ua/publications/books/serii/academy/1102010/Documents/2012_6/07_Kruchin.pdf
8. Воройский Ф. С. Основы проектирования автоматизированных библиотечно-информационных систем / Ф. С. Воройский. – М. : Физматлит, 2002. – 384 с.
9. УкрМарк. Національний формат представлення бібліографічних даних (проект). – Режим доступу : <http://nbuv.gov.ua/library/ukrmark.html>.
10. Лобанова Э.Ш. Долгий путь эволюций форматов : от MARC I до MARC 21 / Э.Ш. Лобанова / *Библиотека*. – 2003. – № 9. – С. 55-57.
11. Lotka A. J. The Frequency Distribution of Scientific Productivity // *Journal of the Washington Academy of Sciences* 16, 1926. – P. 317 – 323.
12. Lorenz M.O. Methods of measuring the concentration of wealth / *Publications of the American Statistical Association*. – 1905. – Vol. 9, №. 70, 9 (70). – P. 209-219.
13. Gastwirth J.L. The Estimation of the Lorenz Curve and Gini Index / *The Review of Economics and Statistics*,. – 1972. – Vol. 54, № 3, 54 (3). – P. 306-316.
14. Hoover E.M. jr. The Measurement of Industrial Localization / *Review of Economics and Statistics*. – 1936, № 18. – P. 162-171.
15. Newman M. E. J. The structure of scientific collaboration networks / *Proc. Natl. Acad. Sci. USA* 98, 2001. – P. 404-409.
16. Colizza V. Detecting rich-club ordering in complex networks / V. Colizza, A. Flammini, M.A. Serrano, A. Vespignani // *Nature Physics*. – 2006. – Vol. 2. – P. 110-115.

17. Ландэ Д.В., Балагура И.В. Сети соавторства по базе данных “Украиника научная” / MegaLing'2012. Горизонти прикладної лінгвістики та лінгвістичних технологій : тези доповідей міжнародної наукової конференції. 20 – 23 листопада 2012. – К. : Український мовно-інформаційний фонд НАН України. – Режим доступу : <http://megaling.ulif.org.ua/mova-misle-nnya-kommun-kats-ya/lande-d-v-balagura-v-seti-soavtorstva-po-baze-dannich-ukrainika-nauchnaya>

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