



УДК 001.893.530.1

О. І. Мриглод, Р. Кенна, Ю. В. Головач, Б. Берш

Про порівняння екстенсивної та інтенсивної мір ефективності наукових груп

(Представлено академіком НАН України В. М. Локтевим)

Піднімаються проблеми, пов'язані із оцінюванням наукової ефективності. Розглядаються екстенсивні та інтенсивні міри для оцінювання загальної сили та усередненого показника якості роботи наукових колективів. Адже, з одного боку, досі немає однозначних критеріїв та підходів для аналізу наукової ефективності. З іншого боку, при формалізації процедури оцінювання важливо врахувати відмінності різних галузей науки. Обговорюються результати перевірки кореляції між експертними оцінками та показниками на основі цитованості для британських наукових груп, що працюють у різних дисциплінах.

Однією з характерних рис сучасної науки є те, що кількісні методи досліджень, які виникли і сформувалися в природничих науках, все більше поширюються на науки суспільні і гуманітарні [1]. Більше того, предметом досліджень стає сама наука, що породило порівняно нову галузь — наукометрію [2]. Доступні нині великі обсяги наукометричних даних зручно аналізувати за допомогою методів та інструментарію інших дисциплін, таких як наука про складні системи [3], теорія випадкових графів [4], соціофізика [5] тощо. У задачах соціофізики для розгляду складних соціальних систем застосовуються не лише методи, а й понятійний апарат, що виник при вивченні фізичних систем (переважно, об'єктів статистичної фізики). Зокрема, при аналізі наукових груп для розгляду їх структури використовується теорія складних мереж [6], а при дослідженні ефективності їх роботи застосовуються методи і концепції теорії фазових переходів [7]. У нашій роботі мова йтиме про ефективність роботи наукових колективів та методи її оцінювання. Ключовим спостереженням, на якому базується наш аналіз, є відмінність між інтенсивними та екстенсивними характеристиками цієї ефективності.

Хоча питання про оцінювання наукового доробку досліджується здавна, проте як тривіальні методи простого підрахунку (статей, грантів тощо), так і сучасний кількісний аналіз наукометричних даних наразі не дають змоги вивести узагальнений показник, який би

© О. І. Мриглод, Р. Кенна, Ю. В. Головач, Б. Берш, 2014

однозначно та коректно відображав різні аспекти наукової роботи. Більше того, навіть найновіші методи автоматичного збору даних страждають від багатьох недоліків, а бази даних є неповними. Безумовно, існує ще перевірений часом метод експертної оцінки, який і досі вважається найбільш надійним (див., наприклад [8]). Прикладом є незалежне рецензування рукописів у наукових виданнях. Проте і цей метод піддається критиці через, скажімо, вплив суб'єктивних факторів на кінцевий висновок [9]. Крім того, проведення масштабної експертної процедури вимагає значного часу та затрат. Таким чином, жоден із можливих підходів до оцінювання наукової ефективності не може вважатися досконалим, проблема балансування між надійністю та простотою неодноразово піднімається як у наукометричній спільноті, так і на рівні приватних чи державних організацій, що займаються рейтингуванням та/або управлінням у сфері науки. Існує думка, що найкращим вирішенням є комбіноване використання наукометричних показників та рецензування [10].

Нижче ми розглядаємо проблему взаємної кореляції наукометричних показників та експертних оцінок. Зокрема, ставиться питання про те, чи можна передбачити експертні висновки про роботу наукових груп на основі даних про цитованість їх публікацій, а отже, чи взаємозамінними є ці два підходи. Продовжуючи наші дослідження [11–13], у цій роботі ми акцентуємо увагу на особливостях окремих галузей науки.

Постановка задачі та деякі означення. Для досліджень наукової якості та впливу використовуємо інтенсивні та екстенсивні міри — концепції, запозичені із статистичної фізики. У цьому контексті термін “екстенсивний” використовується для опису загальних характеристик системи, що залежать від її розміру, тоді як “інтенсивні” значення відповідають усередненим показникам. Надалі, ведучи мову про оцінювання наукової групи, будемо використовувати два типи показників: *відносні* (інтенсивні), що можуть розглядатися як коефіцієнт ефективності групи, тобто усереднені за кількістю її членів N , та *абсолютні* (екстенсивні), що характеризують ефективність групи в цілому і є пропорційними до її розміру N . Обидві характеристики дослідницьких колективів є важливими та широко використовуються на практиці: відносні значення ефективності лягають в основу рейтингів та дають змогу порівняти між собою установи або групи різного розміру; розподілення коштів або ресурсів відбувається пропорційно до абсолютних значень. Очевидно, що ці два типи показників є взаємозалежними і можуть бути одержані шляхом множення/ділення на N .

У фізиці екстенсивні та інтенсивні величини, відповідно, позначаються великими та малими літерами. У своїй роботі ми також слідуємо цій традиції. Використовуючи термінологію із попередніх робіт [11–15], відносний показник на основі експертної оцінки ефективності наукового колективу називатимемо *якістю* s , а загальну, абсолютну оцінку якості — його *силою* S (від англ. strength)

$$S = sN, \quad (1)$$

де N — кількість членів групи. Відповідно, відносний показник, що базується на основі підрахунку одержаних цитувань, називатимемо *середнім впливом* групи i (від англ. impact), тоді як *загальний вплив* I буде відповідати абсолютному значенню

$$I = iN. \quad (2)$$

Вибір таких найменувань змінних пояснюється тим, що рівень цитованості тієї чи іншої наукової публікації інтерпретується як вплив на подальші наукові дослідження, ступінь

використання опублікованих результатів в інших роботах. З іншого боку, більш широке значення слова “якість” скоріше співвідноситься із експертною оцінкою, яка враховує різні аспекти роботи.

Метою роботи є порівняння між собою відносних (s та i) та абсолютних (S та I) показників. Ідеально скорельовані значення мали б лягти на пряму лінію і означали б, що результати експертного оцінювання можуть бути легко передбачені на основі даних про цитування і навпаки. Більше того, у такому дослідженні можна враховувати розміри груп, шукаючи кореляцію не для всіх, а для окремо взятих великих, середніх або малих груп — ці та інші питання, пов’язані із розміром групи та залежністю від нього ефективності групи, досліджувалися у роботах [14, 15].

Використані дані. Очевидно, що одержані результати будуть залежати від кожної окремої вибірки даних, адже у різних країнах по-різному організовано наукові дослідження та неоднаково відбувається розвиток галузей науки. В даному випадку дослідження проводилося на основі даних про ефективність наукових груп у вищих навчальних закладах Великобританії. Це обумовлено тим, що саме там існує багаторічна традиція проведення централізованої процедури оцінювання ефективності наукових та освітніх закладів — Research Assessment Exercise (RAE) (офіційна сторінка RAE 2008 року: <http://www.rae.ac.uk/>). Починаючи з 1986 р., кожні три–п’ять років організовується масштабна робота з оцінювання та рейтингування навчальних закладів (відсоток дослідних інститутів є малим, натомість основна кількість досліджень здійснюється саме у ВНЗ). Згідно із правилами, кожна установа, що претендує на одержання фінансової підтримки від держави, подає відомості про своїх працівників, які відносяться до різних тематичних груп або так званих модулів — units of assessment (UOA). У поданні вказується інформація про вибрані чотири (або менше) публікації кожного науковця, отримані гранти, присвоєні нагороди, кількість студентів та аспірантів, взаємозв’язки із виробництвом, доступні технічні ресурси тощо. Залучені до RAE експерти аналізують кожне подання і формують профіль якості для всіх наукових груп, визначаючи, яка частка їх результатів відноситься до кожного із п’яти рівнів якості: 4* — найвищий рівень якості у світовому масштабі; 3* — високий міжнародний рівень якості; 2* — міжнародний рівень; 1* — національний рівень; Unclassified — низький рівень якості або невідповідність опублікованим критеріям RAE. Таким чином, кожній науковій групі присвоюється власний профіль якості, що складається із набору п’яти цифр. Наприклад, результуючий профіль для групи з біології (UOA 14) Університету Кембриджа з $N = 213,69^1$: 20 (4*); 40 (3*); 30 (2*); 5 (1*); 5 (Unclassified). Оприлюднені профілі якості RAE використовуються для розрахунку так званої формули розподілу коштів, які щорічно виділяє Спеціальна рада Англії з питань фінансування вищої освіти — “Higher Education Funding Council for England” (HEFCE). Зокрема, безпосередньо після завершення попередньої процедури RAE у 2008 р. формула мала вигляд:

$$s = p_{4*} + \frac{3}{7}p_{3*} + \frac{1}{7}p_{2*}, \quad (3)$$

де p_{n*} — це значення із загального профіля якості, що припадає на рівень n . Саме результуюче значення s використовується нами як відносний показник *якості* наукової групи. Відповідно, значення *сили* легко розраховується множенням на N , яке вказується на

¹ N може набувати дробових значень, якщо частина членів групи працюють не на повну ставку — тоді їх враховують лише частково.

веб-сторінці RAE разом із відповідним профілем. Наприклад, для згаданої вище біологічної групи Університету Кембриджа значення s буде приблизно дорівнювати 41,429, і, відповідно, $S = s \cdot 213,69 \approx 8852,87$.

Для того щоб одержати показник ефективності для тієї ж вибірки наукових груп, але вже на основі даних про цитованість, було використано дані приватної компанії Evidence, що ввійшла до складу Thomson Reuters Research Analytics (офіційна сторінка: <http://www.evidence.co.uk>). Використовуючи базу даних Web of Science, для кожної групи, що зробила подання до RAE 2008 р., компанією Evidence були розраховані значення так званого нормалізованого впливу цитувань — *normalised citation impact* (NCI). Особливістю цього показника є нетривіальна процедура нормування, що дає змогу компенсувати відмінності у цитованості для різних галузей науки, різних видань або навіть різних проміжків часу. Отже, *середній вплив* групи i визначається нормованими значеннями цитованості статей, що підсумовуються та усереднюються на кількість членів групи. Відповідно, значення *загального впливу* I легко одержується з (2).

Порівняння відносних (s та i) та абсолютних (S та I) показників було здійснено для наукових груп Великої Британії із семи галузей науки: біологія (UOA 14), хімія (UOA 18), фізика² (UOA 19), інженерія (UOA 28), географія та науки про навколишнє середовище (UOA 32), соціологія (UOA 41), історія (UOA 62).

Обговорення результатів. Наведені у табл. 1 результати (див. також [11–13]) свідчать про те, що між відносними показниками ефективності наукових груп кореляція існує, проте лише помірна: для жодної із проаналізованих дисциплін коефіцієнт лінійної кореляції Пірсона r не перевищує 0,7. Між рейтингами наукових груп, що побудовані на основі s та i , також існує лише незначна кореляція з коефіцієнтом Спірмена $\rho \approx 0,59$ в кращому випадку [12].

Тоді як відносні показники ефективності наукових груп s та i корелюють не дуже сильно, їх абсолютні відповідники S та I виявляються добре скорельованими, як це також чітко видно із табл. 1. Це означає, що з великою долею ймовірності *сила* наукової групи відповідає її *загальному впливу*. Очевидно, що рейтингування за абсолютними значеннями позбавлене сенсу, проте можна оцінити пропорційний внесок кожного колективу.

В табл. 1 також можна простежити відмінності у значеннях коефіцієнтів кореляції між абсолютними показниками для груп з різних дисциплін: хоча у всіх випадках вони є високими, проте найкраще S та I корелюють для природничих галузей науки ($r \geq 0,95$), таких як біологія, хімія та фізика. Дещо нижчим ($r \approx 0,87$) коефіцієнт кореляції є для історії та соціології як гуманітарних дисциплін, а проміжних значень ($r \approx 0,94$) набуває для інженерії та географії. Останнє може пояснюватись тим, що в інтерпретації RAE географія є сильно міждисциплінарною наукою, як, зрештою, й інженерія. Щоб перевірити це спостереження, на рис. 1 і 2 наведено графіки залежності як відносних, так і абсолютних показників окремо для груп із точних, гуманітарних та міждисциплінарних галузей. З рис. 1 бачимо, що дійсно між s та i існує однакова позитивна тенденція, проте рівень скорельованості є невисоким. Добра ж скорельованість абсолютних показників S та I для всіх груп дисциплін продемонстрована візуально на рис. 2.

Таким чином, одержані у роботі результати ще раз підтверджують, що усереднені показники впливу наукових груп (інтенсивні величини), які базуються на підрахунку цитувань,

²В інтерпретації RAE “фізика”, в основному, включає експериментальні дослідження, тоді як теоретичні дослідження часто аналізуються разом із прикладною математикою.

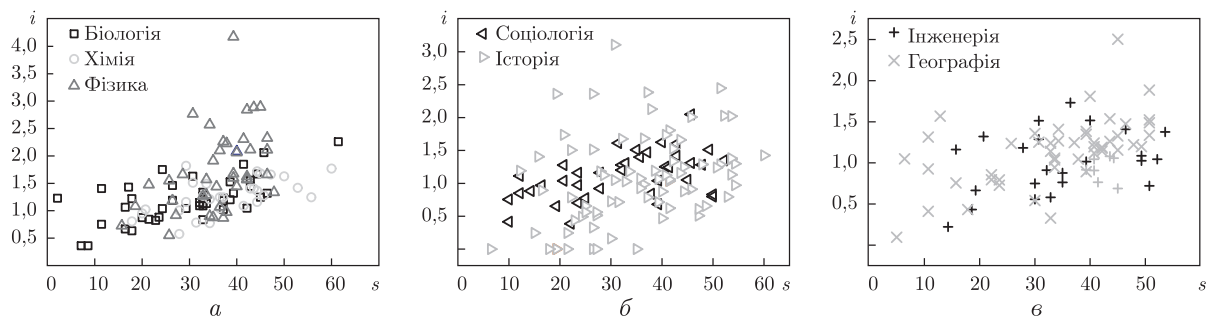


Рис. 1. Залежність між відносними показниками s та i для наукових груп в галузях: a — точних; b — гуманітарних та v — міждисциплінарних наук. Дані про різні дисципліни наведені різними символами

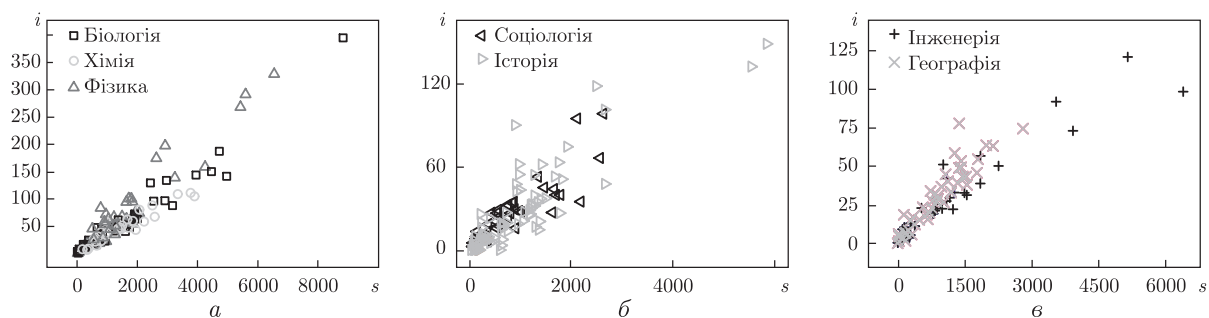


Рис. 2. Залежність між абсолютними показниками S та I для наукових груп в галузях: a — точних; b — гуманітарних та v — міждисциплінарних наук. Дані для різних дисциплін наведені різними символами

не можуть добре відтворити експертні оцінки якості. Відповідно, можна стверджувати, що результати рейтингування установ на основі нормалізованого підрахунку цитувань будуть відрізнятися від результатів рейтингування на основі висновків рецензентів. З іншого боку, при бажанні оцінити пропорційні внески від наукових груп різного розміру (екстенсивні величини) можна з високою долею впевненості використати дані про їх цитованість. Цей висновок справедливий для всіх ділянок науки, проте найкраще працює у випадку саме точних наук, для яких статті у періодичних виданнях є традиційною формою публікації наукових результатів. Значна, проте все ж слабша кореляція між абсолютними оцінками спостерігається для сектора гуманітарних наук, які, по-перше, найслабше представлені

Таблиця 1. Значення коефіцієнтів кореляції між відносними (s та i) та абсолютними (S та I) показниками ефективності для наукових колективів Великої Британії з різних дисциплін

Дисципліна	Коефіцієнт Спірмена ρ для порівняння рейтингів (лише на основі s та i)	Коефіцієнт лінійної кореляції Пірсона r			
		для груп з відповідної дисципліни		для декількох дисциплін (узагальнені дані)	
		s vs. i	S vs. I	s vs. i	S vs. I
Біологія (див. [11])	0,57	0,64	0,97	0,48	0,94
Хімія	0,59	0,58	0,95		
Фізика	0,52	0,46	0,96		
Соціологія	0,50	0,51	0,87	0,36	0,87
Історія	0,38	0,35	0,87		
Інженерія	0,25	0,30	0,94	0,45	0,91
Географія	0,58	0,57	0,93		

у базі даних Web of Science, а, по-друге, мають значну частку публікацій у підручниках, монографіях або інших специфічних формах видань (ноти для музики, методичні матеріали для педагогіки тощо).

Дослідження проведено в рамках проектів “Динаміка і кооперативні явища в складних фізичних і біологічних середовищах” (7-ма Рамкова угода, FP7-PEOPLE, IRSES project N269139) та “Статистична фізика у різноманітних реалізаціях” (7-ма Рамкова угода, FP7-PEOPLE, IRSES project N295302).

1. *Lazer D. et al.* Computational social science // *Science*. – 2009. – **323**. – P. 721–723; *Castellano C., Fortunato S., Loreto V.* Statistical physics of social dynamics // *Rev. Mod. Phys.* – 2009. – **81**. – P. 591–646.
2. *Добров Г. М.* Наука о науке. – Киев: Наук. думка, 1989. – 304 с.
3. *Encyclopedia of complexity and systems science* / Ed. by R. A. Meyers]. – Berlin: Springer, 2009. – 10370 p.
4. *Diestel R.* Graph theory. – Berlin: Springer, 2005. – 410 p.
5. *Galam S.* Sociophysics: A Physicist’s modeling of psycho-political phenomena. – New York: Springer. – 2012. – 439 p.
6. *Головач Ю., фон Фербер К., Олемської О., Головач Т., Мриглод О. та ін.* Складні мережі // *Журн. фіз. досліджень*. – 2006. – **10**. – С. 247–291.
7. *Order, disorder and criticality. Advanced problems of phase transition theory* / Yu. Holovatch (ed). – Singapore: World Scientific, 2004. – Vol. 2. – 2007. – Vol. 3. – 2012. – 248 p.
8. *van Raan A. F. J.* Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods // *Scientometrics*. – 2005. – **62**, No 1. – P. 133–143.
9. *Bornmann L.* The Hawthorne effect in journal peer review // *Ibid.* – 2012. – **91**. – P. 857–862.
10. *De Bellis N.* Bibliometrics and citation analysis: from the science citation index to cybermetrics. – Lanham: Scarecrow Press, 2009.
11. *Mryglod O., Kenna R., Holovatch Yu., Berche B.* Absolute and specific measures of research group excellence // *Scientometrics*. – 2013. – **95**, No 1. – P. 115–127.
12. *Mryglod O., Kenna R., Holovatch Yu., Berche B.* Comparison of a citation-based indicator and peer review for absolute and specific measures of research-group excellence // *Ibid.* – 2013. – **97**, No 3. – P. 767–777.
13. *Мриглод О., Кенна Р., Головач Ю., Берш Б.* Про вимірювання наукової ефективності // *Вісн. НАН України, прийнято до друку*. – Препринт ІФКС НАН України № 13. – 02U: <http://www.icmp.lviv.ua/sites/default/files/preprints/pdf/1302U.pdf>).
14. *Kenna R., Berche B.* The extensive nature of group quality // *Europhys. Lett.* – 2010. – **90**. – P. 58002.
15. *Kenna R., Berche B.* Critical mass and the dependency of research quality on group size // *Scientometrics*. – 2010. – **86**, No 2. – P. 527–540.

*Інститут фізики конденсованих систем
НАН України, Львів
Університет Ковентрі, Велика Британія
Інститут фізики конденсованих систем
НАН України, Львів
Університет Лотарингії, Нансі, Франція*

Надійшло до редакції 04.09.2013

О. И. Мрыглод, Р. Кенна, Ю. В. Головач, Б. Бэрш

О сравнении экстенсивных и интенсивных мер эффективности научных групп

Поднимаются проблемы, связанные с оценением научной эффективности. Рассматриваются экстенсивные и интенсивные меры для оценивания общей силы и усредненного показателя качества работы научных коллективов. Ведь, с одной стороны, до сих пор не существует однозначных критериев и подходов для анализа научной эффективности. С другой стороны, при формализации процедуры оценивания важно учитывать отличия разных отраслей науки. Обсуждаются результаты проверки корреляций между экспертными оценками и показателями на основе цитируемости для британских научных групп, которые работают в разных дисциплинах.

O. I. Mryglod, R. Kenna, Yu. V. Holovatch, B. Berche

On the comparison of extensive and intensive measures of the efficiency of scientific groups

The problem of scientific efficiency evaluation is discussed. On the one hand, there is no simple but reliable way to assess the individual or collective scientific performance. It is a challenge to decide which approach, i. e. peer-review or citation-based indicators, should be used for this purpose. On the other hand, differences between the disciplines should be taken into account for the assessment. We analyze correlations between values of citation-based impact indicators and peer-review scores for British scientific groups in several academic disciplines, from natural to social sciences and humanities.