

Anna Małgorzata Kamińska
Instytut Bibliotekoznawstwa i Informatyki Naukowej
Uniwersytet Śląski w Katowicach
anna.kaminska@us.edu.pl

Gry wojenne. Ewaluacja ewaluacji nauki¹

Streszczenie: Artykuł przedstawia subiektywne spostrzeżenia autorki na temat funkcji ewaluacji współczesnej nauki realizowanej m.in. przez ocenę danych bibliograficznych. Po krótkim wstępie przypominającym rolę nauki oraz rolę jej ewaluacji przedstawione zostały aktualne problemy związane z oceną produktywności środowisk naukowych. Część z nich nie wynika z niskiej jakości danych bibliograficznych, ale raczej z braku powszechnej akceptacji obecnie stosowanych metod ilościowych wynikającej chociażby z nieintuicyjnego i nieprzewidywalnego zachowania się niektórych wskaźników oraz braku powszechnego przekonania, że odzwierciedlają one wymiar pozytywnego wpływu danych dokonań naukowych na dziedzinę życia codziennego. Powoduje to włączanie metod eksperckich jako składowych metod złożonych, co skutkuje dodatkowym zmniejszeniem przejrzystości systemu oceny. Autorka, posiłkując się ilustrowanym przykładem, stawia hipotezę, że dalsze poszukiwanie metod ilościowej oceny wkładu naukowego bazujących na zastanym, tradycyjnym modelu danych bibliograficznych może stanowić ślepią uliczkę, na wyjście z której daje szansę rozbudowa tego modelu na przykład w stronę możliwości uwzględniania zależności wagowych.

Słowa kluczowe: dane bibliograficzne, bibliometria, ewaluacja nauki, bibliometryczny model danych, Impact Factor, indeks Hirscha, ocena dorobku naukowego

Wprowadzenie

Dane bibliograficzne, wskazując na treści zawarte w opisywanych przez nie obiektach, odgrywają kluczową rolę dla rozwoju dyscyplin naukowych. Mimo ciągłego postępu technologicznego artykuły naukowe publikowane bądź to w formie drukowanej, bądź elektronicznej, będąc podstawową formą dokumentowania wyników badań, wciąż stanowią fundament komunikacji naukowej. Elektronizacja informacji bibliograficznych polegająca na ich cyfrowym opracowaniu jako strukturalnego zbioru danych, a następnie zdeponowaniu w bibliograficznych bazach danych, umożliwia z założenia szybkie i relewantne przeszukiwanie zgromadzonych zasobów. Korzyści z budowania takich usług są dla nauki ogromne. Dzięki nim badacze mają możliwość wyszukania źródeł zawierających informacje o interesujących ich obszarach problemowych. Zapoznanie się z tymi źródłami może stanowić inspirację naukową, służyć uzupełnieniu niezbędnej wiedzy do dalszych badań w kierunkach komplementarnych, a czasami pomóc zawczasu podjąć decyzję o zaprzestaniu pewnych badań w przypadku odkrycia, że dany obszar został już gruntownie zbadany wcześniej.

Aby móc prawidłowo pełnić wyżej opisane funkcje, zgromadzone dane bibliograficzne powinny być prawdziwe, kompletne i spójne. Implementowanie nawet najbardziej nowoczesnych usług informacyjno-wyszukiwawczych w oparciu o dane

¹ Treść niniejszego opracowania została przedstawiona 23 marca 2018 r. na „3. Warsztacie Naukometrycznym – Badania nad produktywnością naukowców” zorganizowanym przez Scholarly Communication Research Group – Instytut Filozofii Uniwersytetu im. Adama Mickiewicza w Poznaniu.

niespełniające tych kryteriów będzie prowadziło do poważnego upośledzenia funkcji bibliograficznych baz danych. Jak wynika z badań prowadzonych w ostatnim czasie nad jakością danych gromadzonych w krajowych bibliograficznych bazach danych, niektóre systemy wciąż doświadczają problemów natury podstawowej².

Trzeba tutaj zauważyć, że chociaż w powyżej opisanych scenariuszach przejawia się „konsumpcyjna” rola informacji bibliograficznych, to naukowcy dokumentując wyniki swoich badań publikowanymi pracami, są zarazem „producentami” tychże informacji.

Granice nauki, rola nauki i rola jej ewaluacji

Efekty „produkcji” informacji stanowią, w Polsce i wielu innych krajach, jedną ze składowych oceny tzw. produktywności³ poszczególnych naukowców i instytucji ich zatrudniających. Ocena ta, mając często bezpośredni wpływ na rangę naukową i poziom finansowania poszczególnych ośrodków, stanowi narzędzie stymulacji rozwoju nauki. Stymulacja owa, rozumiana literalnie jako pobudzenie do działania lub rozwoju, powinna w pierwszej kolejności wyływać z przekonań dotyczących roli nauki we współczesnym świecie oraz tego, co powinna ona oferować społeczeństwom zamieszkującym bliższe i dalsze rejony jej uprawiania. Dopiero mając tak zdefiniowane cele, możliwe jest w drugiej kolejności skupianie się na metodach oceniających produktywność i skuteczności osiągania tych celów, a równocześnie pozwalających na lepszą aktywizację tejże produktywności.

Wraz z postępowaniem cywilizacyjnym klasyfikacje dyscyplin naukowych podlegały różnym przemianom, jednak niezależnie od nich zauważyć można różnorodność funkcji pełnionych przez poszczególne gałęzie nauki mające pośrednio lub bezpośrednio zaspokajać różnorakie potrzeby uczestników życia społecznego. Potrzeby te z pewnością dywersyfikują takie czynniki jak stan zdrowia, zamożność czy światopogląd poszczególnych jednostek czy społeczeństw. Dlatego nauka w służbie człowieka choć powinna wychodzić w pierwszej kolejności naprzeciw najbardziej bieżącym jego potrzebom, od tych najbardziej podstawowych poczynając, nie powinna się jednak do nich ograniczać. Jak pisze Michał Heller w swojej książce *Granice nauki*⁴: *Jeśli bowiem nauka nie ma stać się tylko rzemiosłem do produkowania wyników, które szybko można zamienić na praktyczne zastosowania, lecz ma pozostać wierną swemu powołaniu rozumienia świata (zresztą najlepsze zastosowania wynikają z głębokiego rozumienia), to nie może cofnąć się z drogi „powrotu do początku”. Ostatecznie przecież wszystko redukuje się do cząstek elementarnych i tego wszystkiego, z czego powstały, a to*

² KAMIŃSKA, A.M. ProBIT – prospektywna metoda tworzenia trawersowalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych. *Zagadnienia Informacji Naukowej* 2017, nr 1, s. 66–82. ISSN 0324-8194; KAMIŃSKA, A.M. Miary podobieństw łańcuchów znakowych a deduplikacja rekordów w bibliograficznych bazach danych. *Przeгляд Библиотечный* 2017, nr 4, s. 477–495. ISSN 0033-202X.

³ Słowo „produktywność” używane jest najczęściej w kontekście ewaluacji nauki, choć pojawiają się także takie określenia jak wkład naukowy czy dokonania naukowe.

⁴ HELLER, M. *Granice nauki*. Kraków: Copernicus Center Press, 2014, s.137. ISBN 978-83-788-6049-5.

automatycznie przenosi nas do ery kwantowej kosmologii. Późniejszy Wszechświat nie mógł rozpocząć swojej ewolucji bez odpowiednich warunków początkowych, a one ustaliły się w wyniku działania kwantowych efektów grawitacji (...).

Wypowiedź ta zwraca naszą uwagę na różne aspekty dyscyplin naukowych. Pierwszy to taki, że podstawowym celem ich uprawiania nie musi być wytwarzanie produktów gotowych do użycia przez końcowego konsumenta (w niniejszym artykule określenie produkt badań naukowych rozumiane jest w o wiele szerszym znaczeniu⁵), ale dochodzenie do prawdy, lepsze poznanie zjawisk, które może stać się podstawą dostarczania przez przemysł lepszych produktów ich użytkownikom. Drugi dotyczy przeplatania się (interdyscyplinarności) dziedzin i dyscyplin naukowych tworzących „łańcuch dostaw”, gdzie kolejne ogniwo korzysta z dorobku poprzedniego i często tylko ostatniej z dziedzin tego łańcucha przypisuje się sukces za jakość uzyskanego dzieła, w które wkład miała przecież całość zrealizowanych badań. Trzeci zaś aspekt wskazuje, że nie wszystkie z prowadzonych badań powinny mieć na celu jedynie ulepszanie produktów wykorzystywanych w życiu codziennym.

Nauka jako obszar nieukierunkowany na masową produkcję i wdrażanie produktów końcowych ma ograniczone możliwości samofinansowania i dlatego w większości państw dotowana jest z budżetu. Biorąc pod uwagę zmiany w czasie oczekiwań co do kierunków realizowanych badań (wynikających np. z przyjętej aktualnie polityki) oraz istnienie wielu ośrodków naukowych z nierzadko powielającymi się obszarami badawczymi, racjonalizacja ponoszonych nakładów narzuca konieczność organizowania rynku nauki i kontrolowania jego działań i ich skutków w każdym z ośrodków. Jednym z narzędzi jest tutaj właśnie ewaluacja nauki, której metody powinny umożliwiać porównanie produktywności poszczególnych jej dziedzin celem stworzenia warunków zdrowej konkurencyjności realizowanych prac badawczych.

W zależności od zasobności budżetu i oceny pilności bieżących oczekiwań w stosunku do wykonawców badań naukowych miary ewaluacyjne powinny promować ściśle podążanie za wytyczonymi kierunkami badań lub przeciwnie, tolerować meandrowanie w kierunkach pobocznych. Dodatkowym uzasadnieniem jest tu fakt, że nauka podlega zjawiskom rozwoju ewolucyjnego. Oznacza to, że wraz z rozszerzeniem frontów badawczych zwiększa się prawdopodobieństwo odkryć przełomowych, ale i również potencjalnie zwiększa się ilość środków utraconych na badania nieprzynoszące oczekiwanych rezultatów.

Niezależnie jednak od zastosowanych metod ewaluacji powinny być one transparentne i znane wszystkim podmiotom, których one dotyczą w całym okresie realizowanych badań.

⁵ Produktem naukowym w szerszym rozumieniu może być na przykład odkrycie prawdy historycznej rzutującej na kształtowanie się światopoglądu społecznego czy też odkrycie nowego typu patogenu, choć droga do jego zwalczania może być jeszcze bardzo daleka.

Błędne koło ewaluacji nauki

Implementacja koncepcji miar przedstawionych w poprzednim rozdziale nie jest zadaniem trywialnym, a weryfikacja jakości zamierzonych skutków jest możliwa dopiero po pewnym czasie od momentu ich obowiązywania. Często retrospektywna refleksja podpowiada, że nie do końca osiągnięto planowane cele lub nawet pogorszone funkcjonowanie w pewnych obszarach. Skłania to do opracowywania nowych miar lub powrotu do starszych metod ewaluacji nauki. Ciekawe spostrzeżenia na temat rozwoju metod ewaluacji przedstawił Miloš Jovanović podczas *15th International Society of Scientometrics and Informetrics Conference*⁶, według którego rozwój tych metod uwięziony został w błędnym kole składającym się z czterech kolejno po sobie następujących faz (il.1.):

1. Stosowanie metod subiektywnych z nieprzejrzystymi kryteriami oceny; wkład naukowców czy instytucji oceniany jest przez wąską grupę ekspertów, co często prowadzi do niezgody i sprzyja tworzeniu konfliktów.
2. Wprowadzenie obiektywnie ilościowo mierzalnych kryteriów, według których możliwe jest ilościowe porównywanie dorobku poszczególnych badaczy czy całych instytucji; często okazuje się, że środowiska naukowe adaptują się do zaproponowanych miar i zwiększają swoją produktywność, jednak nie przekłada się to na wzrost poziomu jakości uprawianej nauki.
3. Celem podniesienia jakości badań naukowych zaproponowane wcześniej miary ilościowe próbuje się modyfikować tak, aby bardziej odzwierciedlały jakość prowadzonych badań, a nie tylko możliwości produkcji artefaktów naukowych; próby te często nie przynoszą oczekiwanych rezultatów.
4. Rezygnacja z mierzalnych i transparentnych funkcji celu i powrót do metod stosowanych na początku.

⁶ JOVANOVIĆ, M.M. The Vicious Circle of Evaluation Transparency – An Ignition Paper. W: *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference*. Istanbul: Bogazici University Printhouse, 2015, s. 646–647. ISBN 978-975-518-381-7.



Il. 1. Cykl ewaluacji nauki

Źródło: JOVANOVIĆ, M.M. The Vicious Circle of Evaluation Transparency – An Ignition Paper. W: *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference*. Istanbul: Bogazici University Printhouse, 2015, s. 646.

Być może obraz przedstawiony przez autora przytoczonej pracy jest nieco przejaśkrawiony. Być może wnioski wyciągane z braku oczekiwanej skuteczności kolejno wdrażanych metod pozwolą jednak je udoskonalić. Być może kluczem jest rozwijanie metod hybrydowych łączących w sobie elementy ilościowe i jakościowe. Być może pozwoli to przerwać zamknięty cykl i stworzyć metody skutecznie motywujące środowiska naukowe do ciągłego podnoszenia jakości swoich badań. Aby to osiągnąć ważne jest częste spoglądanie wstecz i trzeźwa ocena skutków zmian wywołanych kolejno testowanymi na żywym organizmie metodami.

O doborze jednostek miar dla gruszek i jabłek

Przyglądając się propozycji modelu rozwoju metod ewaluacji nauki przedstawionej w rozdziale poprzednim, wydaje się, że krajowym metodom najbliższej obecnie do jej trzeciej fazy. Prosto mierzalne kryteria, jak liczba cytowań czy liczba publikacji, wzbogacane są o składowe takie jak IF (ang. *Impact Factor*) czy wskaźnik Hirscha (ang. *h-index*). Podejścia te budzą wątpliwości w środowiskach naukowych co najmniej z dwóch powodów. Pierwszy to wykorzystywanie wartości poszczególnych miar do oceny badań naukowych w ramach dziedzin, w których charakterystyki tych miar są zupełnie różne. Pisali już o tym np. Zbigniew Błocki i Karol Życzkowski w artykule *Czy można porównywać jabłka i gruszki? O danych bibliometrycznych w różnych dziedzinach nauki*⁷.

⁷ BŁOCKI, Z., ŻYCZKOWSKI, K. Czy można porównywać jabłka i gruszki? O danych bibliometrycznych w różnych dziedzinach nauki. *Nauka* 2013, nr 2, s.37–46. ISSN 1231-8515.

Drugi to wdrażanie miar, których zachowania są niezgodne chociażby z intuicyjnymi oczekiwaniami wobec obiektywnych wskaźników ewaluacji. Przykładem może być najbardziej obecnie rozpowszechniony wskaźnik Hirscha. W poniższej tabeli przedstawiono charakterystykę dorobku dwóch naukowców w danym okresie. Autor X jest autorem 9 publikacji cytowanych 12 razy oraz 3 cytowanych 4 razy, natomiast Autor Y 7 publikacji cytowanych 15 razy oraz 3 cytowanych 5 razy. Ich wskaźniki Hirscha wynoszą więc odpowiednio 9 oraz 7, co oznacza że dorobek Autora X jest oceniony znacznie wyżej.

Tab. 1. Charakterystyka dorobku dwóch naukowców

	Autor X	Autor Y
	9*12	7*15
	3*4	3*5
H	9	7
artykułów	12	10
cytowań	120	120

Źródło: opracowanie własne.

Po pewnym czasie tych dwóch autorów podwoiło swoje dorobki. Ich charakterystykę przedstawiono w tabeli 2.

Tab. 2. Charakterystyka dorobku dwóch naukowców

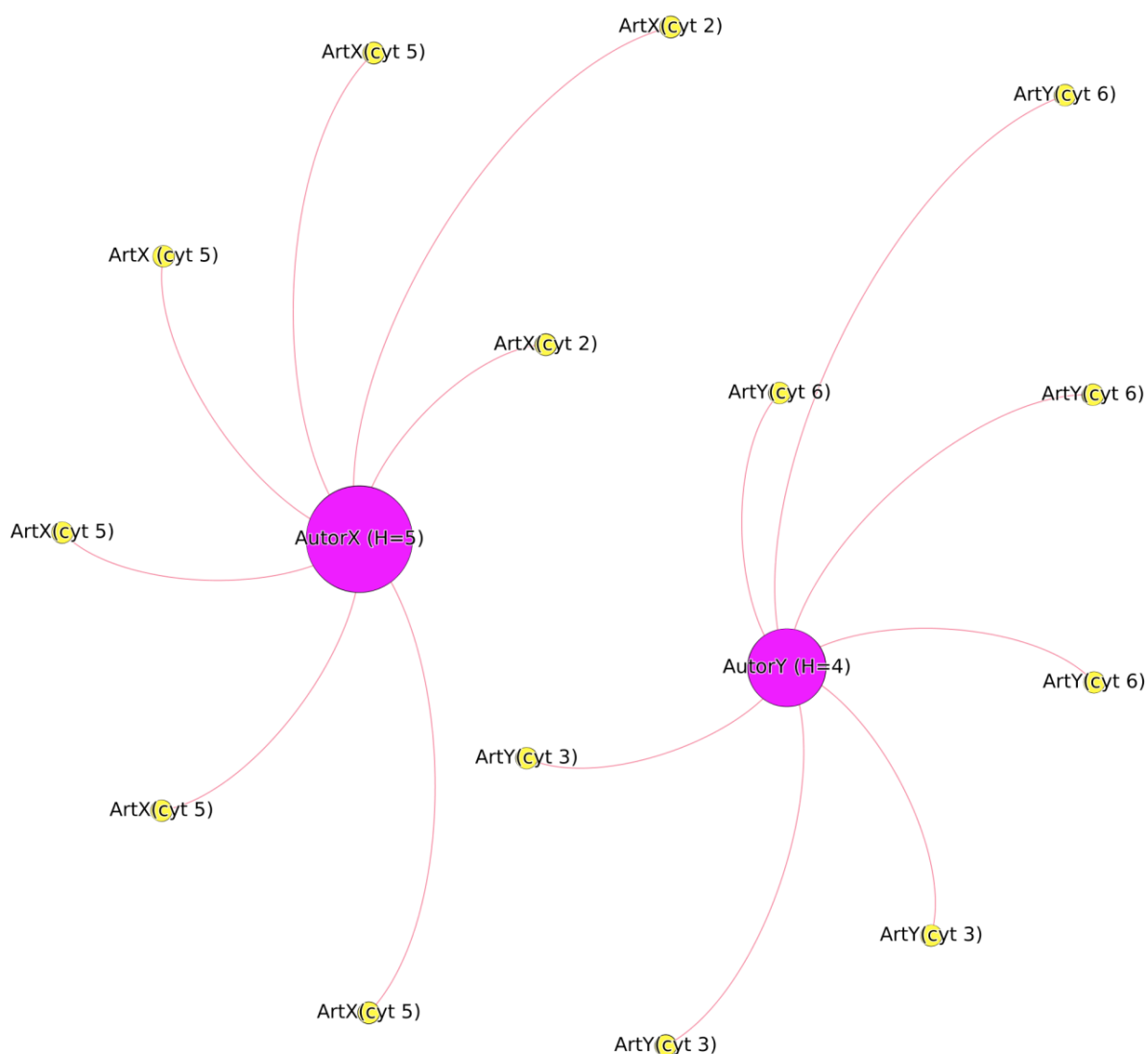
	Autor X	Autor Y
	18*12	14*15
	6*4	6*5
H	12	14
artykułów	24	20
cytowań	240	240

Źródło: opracowanie własne.

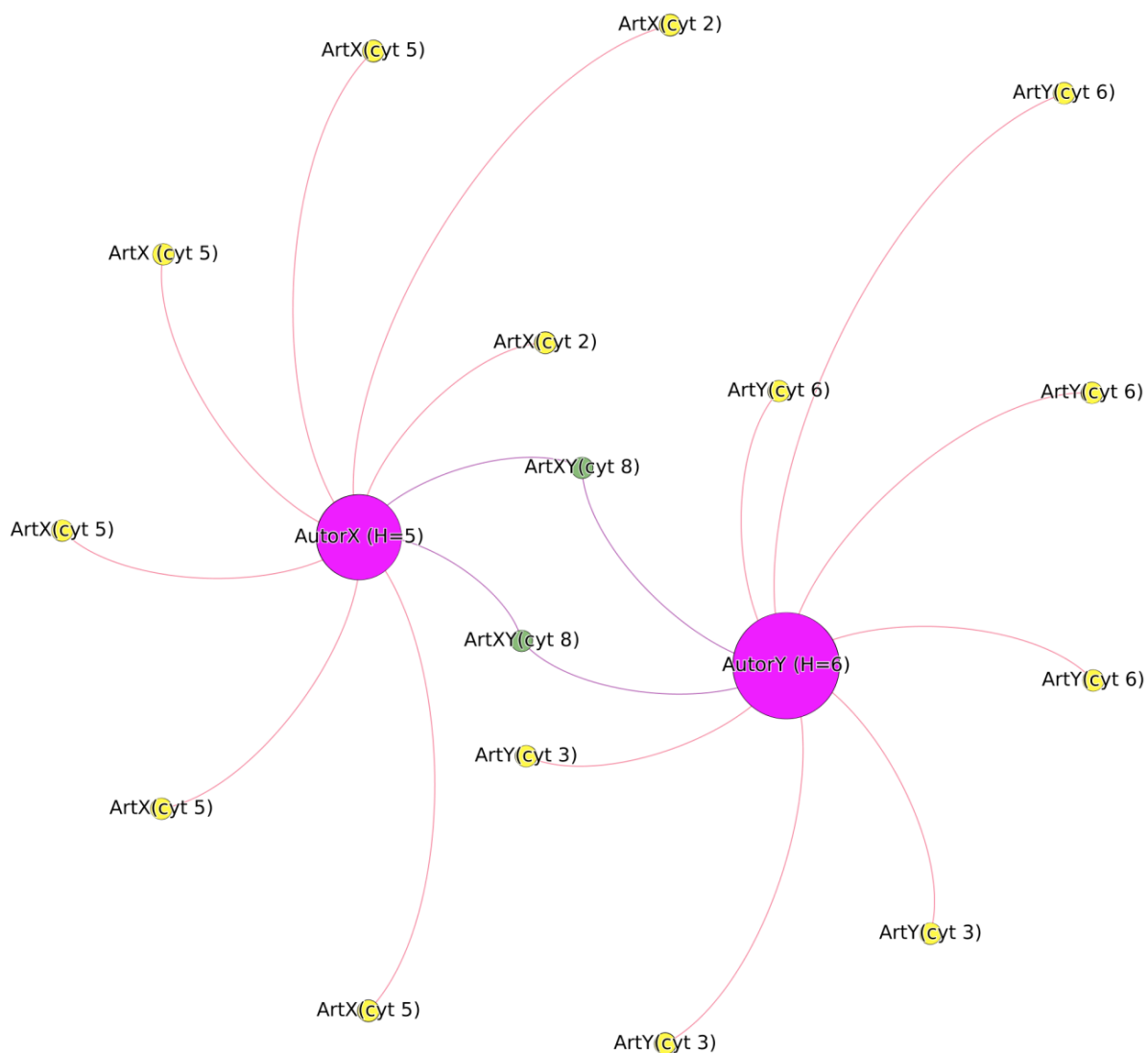
Widzimy, że teraz to Autor Y wyprzedza Autora X o dwa punkty wskaźnika Hirscha. Mimo że Autor X wcześniej oceniany był lepiej i w kolejnym okresie podwoił swój dorobek, to ostatecznie oceniony jest gorzej od Autora Y, wcześniej ocenionego gorzej, który w kolejnym okresie nie zmienił swojej produktywności i wypracował dorobek dokładnie taki jak poprzednio. Wynika stąd, że miara ta narusza oczekiwanie niezmienności kolejności rankingu naukowców w przypadku poczynienia przez każdego z nich tych samych postępów względnych, czyli zwielokrotnienia indywidualnego dorobku o taki sam mnożnik.

Drugi przykład obrazują dwie kolejne ilustracje (ill. 2., Il. 3.). Pierwsza przedstawia dorobek dwóch naukowców, którzy pracowali niezależnie od siebie. Autor X wypracował 5 publikacji, z których każda cytowana była 5 razy oraz 2 publikacje cytowane

2 razy, natomiast Autor Y wypracował 4 publikacje, z których każda cytowana była 6 razy oraz 3 publikacje cytowane 3 razy. Ich wskaźniki Hirscha wynoszą więc odpowiednio 5 i 4, co wskazuje na lepszą ocenę dorobku Autora X. Następnie autorzy ci nawiązali współpracę, co zaowocowało 2 wspólnymi publikacjami, z których każda cytowana była 8 razy.



II. 2. Przykładowy dorobek dwóch autorów pracujących niezależnie
Źródło: opracowanie własne.



II. 3. Skumulowany dorobek dwóch autorów po rozpoczęciu współpracy
Źródło: opracowanie własne.

Poskutkowało to odwróceniem kolejności rankingu tych naukowców, ponieważ wskaźnik Hirscha dla pierwszego z nich się nie zmienił, natomiast dla drugiego zwiększył o dwa punkty. Mimo że Autor X wcześniej oceniany był lepiej od Autora Y, a nowy dorobek obydwu z nich był taki sam i dotyczył jedynie wspólnie wypracowanych publikacji, to kolejność w rankingu zmieniła się. Wynika stąd, że miara ta narusza oczekiwanie niezmienności kolejności w rankingu naukowców w przypadku poczynienia przez każdego z nich tych samych postępów bezwzględnych.

Warto zauważyć, że wyżej opisane niespójności dotyczą również propozycji modyfikacji klasycznego wskaźnika Hirscha, takich jak np. g-index⁸, h(2)-index⁹ czy w-index¹⁰, a szerszą dyskusję na ten temat prowadzą Ludo Waltman i Nees Jan van Eck¹¹.

Wysoką popularność wskaźnik Hirscha w zastosowaniach ewaluacyjnych zawdzięcza prostocie wyliczenia miar dla poszczególnych badaczy. Operuje on na relacjach autorstwa i cytowania¹², a do jego wyznaczenia potrzebna jest jedynie znajomość publikacji cytujących artykuły danego naukowca. Jednak oprócz wyżej przywołanych przykładów niekonsekwentnego zachowania tego wskaźnika oraz ryzyk wynikających ze „wspólnego zliczania gruszek i jabłek” opisanych w pracy Zbigniewa Błockiego i Karola Życzkowskiego warto zwrócić uwagę na jednostki miar wykorzystywane przez poszczególne wskaźniki ewaluacyjne. Sumowanie liczby wystąpień (czyli liczby gruszek czy jabłek) artykułów cytujących jest metodą najprostszą i nie uwzględnia w żaden sposób „ważności” artykułów cytujących. Zatem jeśli gruszki czy jabłka w danym zbiorze owoców są zróżnicowane, to czy nie właściwszym sposobem oceny ich wartości, oprócz podziału całości na osobny zbiór jabłek i osobny zbiór gruszek, byłoby uwzględnienie chociażby wagi owoców w ramach każdego ze zbiorów, a nie tylko określenie ich liczebności?

Gry wojenne

Spostrzeżenie powyższe zdaje się znajdować swoje uzasadnienie zwłaszcza wobec faktu, że liczba pozycji bibliografii załącznikowej artykułów naukowych nie jest niczym ograniczana. Naukowcy biorą tym samym w posiadanie narzędzie „oddawania nieograniczonej liczby głosów” na członków swojej społeczności, a głosy te są tak samo ważne bez względu na doniosłość i wkład w rozwój nauki publikacji, która je zawiera. Widać więc, że wdrażanie ilościowych metod ewaluacji otwiera arenę współzawodnictwa naukowego, którego skuteczność jest tym większa, im bardziej uwzględnia ono specyfikę reguł rządzących tą areną. Wynika z tego, że wprowadzanie nowych metod może prowadzić do zmiany stylu pracy naukowców niekoniecznie ukierunkowanego na zwiększanie wkładu badawczego, ale raczej na zwiększanie wartości wskaźników ewaluacyjnych. Zachowania charakteryzujące się ciągłą adaptacją do reguł zmieniających się w obrębie społeczności są zrozumiałe, a nawet całkowicie naturalne i możliwe do zaobserwowania nie tylko

⁸ EGGHE, L. Theory and practise of the g-index. *Scientometrics* 2006, nr 1, s. 131–152. ISSN 0138-9130.

⁹ KOSMULSKI, M. A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter* [online]. 2006, nr 2. [Dostęp 18.03.2018]. ISSN 1998-5460. Dostępny w: <http://sci-2s.ugr.es/sites/default/files/files/TematicWebSites/hindex/kosmulski2006.pdf>.

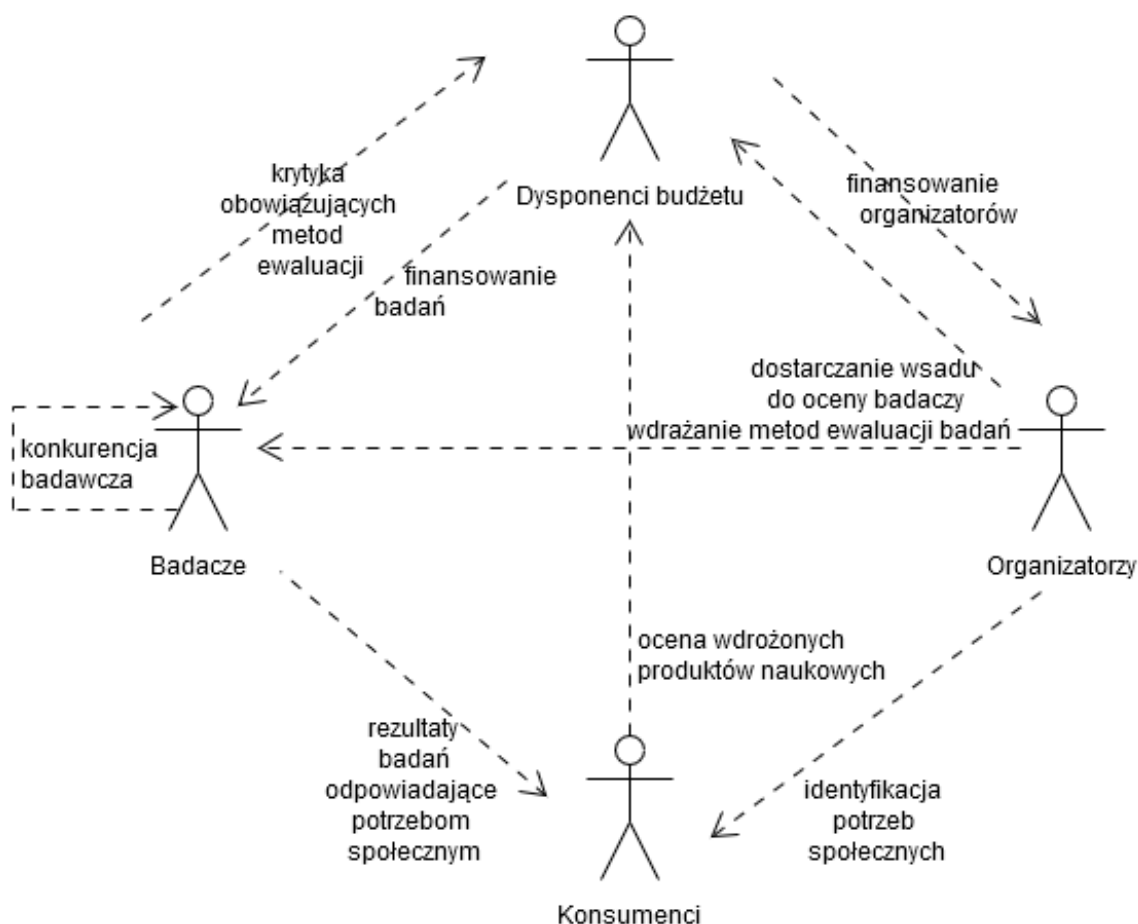
¹⁰ WU, Q. The w-index: A measure to assess scientific impact by focusing on widely cited papers. *Journal of the American Society for Information Science and Technology* [online]. 2010, nr 3, s. 609–614. [Dostęp 18.03.2018]. Dostępny w doi: 10.1002/asi.21276.

¹¹ WALTMAN, L., VAN ECK, N. The Inconsistency of the H-index. *Journal of the American Society for Information Science and Technology* [online]. 2012, nr 2, s. 406–415. [Dostęp 18.03.2018]. Dostępny w doi: 10.1002/asi.21678.

¹² Więcej informacji na temat typów relacji bibliometrycznego modelu danych znaleźć można w: KAMIŃSKA, A.M. Zastosowanie struktur grafowych do analiz bibliometrycznych i webometrycznych. Modele i metody. *Nowa Biblioteka. Usługi, Technologie Informacyjne i Media* (w druku).

w środowiskach naukowych, a liczne przykłady ich alegorycznego zobrazowania znaleźć można zarówno w literaturze, jak i sztuce filmowej (np. *Gry wojenne*¹³).

Jeszcze przed projektowaniem miar ewaluacji jakichkolwiek oddziaływań w jakichkolwiek społecznościach warto zidentyfikować udziałowców tych oddziaływań, czyli klasy podmiotów powiązane między sobą poszczególnymi typami zależności. W przypadku rynku badań naukowych szkic takich oddziaływań przedstawiono na Il. 4.



Il. 4. Szkic oddziaływań pomiędzy uczestnikami rynku nauki
Źródło: opracowanie własne.

Jak zauważono wcześniej, badania naukowe finansowane są najczęściej w systemie budżetowym. Oznacza to konieczność efektywnego alokowania ograniczonych środków celem maksymalizacji jakości badań naukowych. Z potrzeby podnoszenia efektywności tej alokacji dysponenci budżetu powołują organ odpowiedzialny za organizację rynku nauki, którego rolą jest wdrażanie skutecznych metod ewaluacji nauki uwzględniających ograniczenia budżetu oraz zidentyfikowane potrzeby społeczne.

¹³ STABLEFORD, B. *Gry wojenne*. Bydgoszcz: Express Books, 1990. ISBN 83-8517-504-0; BATHAM, J. *Gry wojenne* [film]. USA: United Artists, Sherwood Productions, 1983.

Jakość realizowanych badań naukowych przekłada się pośrednio na zaspokajanie potrzeb społecznych. Jest to najślabsze ogniwo tego obiegu zależności, gdyż naukowcy mają ograniczone możliwości wdrażania wyników swoich badań w postaci gotowych produktów. Konsumenci odczuwają jakość prowadzonych badań jako jedną ze składowych zaspokajających ich potrzeby, a ich ocena ma bezpośredni wpływ na dysponentów budżetu. Ci zaś w ramach działań operacyjnych finansują środowiska badawcze w zależności od uzyskanych przez nie wskaźników ewaluacji, zaś w ramach działań strategicznych zgłaszają oczekiwania zmian do organizatorów rynku nauki. Naukowcy w ramach działań operacyjnych maksymalizują swoją produktywność zgodnie z obowiązującymi metodami ewaluacji, zaś w ramach działań długoplanowych zgłaszają uwagi dotyczące niedoskonałości obowiązujących wskaźników.

Naukowcy mają prawo postępować w sposób zapewniający maksymalizację wartości obowiązujących wskaźników, gdyż z założenia są one wyrazem oczekiwań dysponentów finansującego ich budżetu, a dodatkowo zależy od nich ich dalsza kariera zawodowa. To metody powinny być wypracowane w sposób umożliwiający mierzenie jakości pośredniego lub bezpośredniego wkładu badawczego na całość społeczności, przy zachowaniu oglądu całości oddziaływań występujących na rynku nauki.

Przy projektowaniu metod ewaluacji warto każdej z ich składowych przyjrzeć się przez pryzmat wyżej przedstawionego modelu oddziaływań, a w szczególności celów ostatecznych, jakim mają służyć badania naukowe, czyli zaspokajaniu potrzeb społeczeństwa. Przykładem może być tutaj premiowanie publikowania w wydawnictwach z określonym progowym wskaźnikiem odrzuceń artykułów. Faworyzowanie takich wydawnictw zwiększa bowiem popyt na publikowanie w nich, co zwiększa liczbę artykułów tam zgłaszanych. Założenia wydawnicze możliwości publikacji stałej liczby artykułów w określonej jednostce czasu powodują wtedy konieczność odrzucania jeszcze większej liczby artykułów, co z kolei zwiększa współczynnik odrzuceń i czyni wydawnictwa jeszcze bardziej „prestżowym”. Artykuły przyjęte do publikacji przez takie wydawnictwa muszą nieraz czekać na ukazanie się drukiem wiele miesięcy, co wpływa hamująco na ogólny rozwój badań naukowych i czas przekuwania ich wyników na produkty użyteczne dla końcowych konsumentów.

Mędrca szkiełko i oko

Mimo posługiwania się stosunkowo prostym aparatem analitycznym (a może właśnie dlatego) organizatorzy rynku nauki stoją przed niełatwym zadaniem interpretacji wyników swoich analiz. Nawet tak proste operacje jak obliczanie udziału procentowego poszczególnych rozłącznych podzbiorów w całym badanym zbiorze mogą rodzić wątpliwości co do ostatecznego sformułowania wniosków. W tabeli 3. przedstawiono syntetyczny przykład badawczy¹⁴ mający na celu określenie cytowalności prac w dziedzinach nauk ścisłych i humanistycznych. Praca jest zliczana, jeśli była cytowana przynajmniej jeden raz; do analizy przyjęto równoliczne zbiory badawcze wynoszące 350 prac dla każdej z dziedzin.

¹⁴ Przykład nie jest oparty na rzeczywistych danych.

Tab. 3. Cytowalność prac w naukach ścisłych i humanistycznych

	Nauki ścisłe	Nauki humanistyczne
Łącznie	78% (273/350)	83% (289/350)

Źródło: opracowanie własne.

Analiza porównawcza wykazała, że więcej (o 5%) cytowanych prac wystąpiło w ramach nauk humanistycznych. Można by się więc pokusić o sformułowanie konkluzji: „więcej prac cytowanych jest w ramach nauk humanistycznych” lub też: „w naukach ścisłych więcej prac nie ma dalszego wpływu na rozwój nauki”.

Przyjrzyjmy się jednak dokładniej mierzonym elementom. Wiemy o nich nieco więcej (choć na początku wiedza ta zdawała nam się nie mieć wpływu na wyniki naszych analiz), np. potrafimy je zakwalifikować do grupy wydawnictw zwartych lub ciągłych. Wyniki analiz z podziałem na te kategorie przedstawiono w kolejnej tabeli.

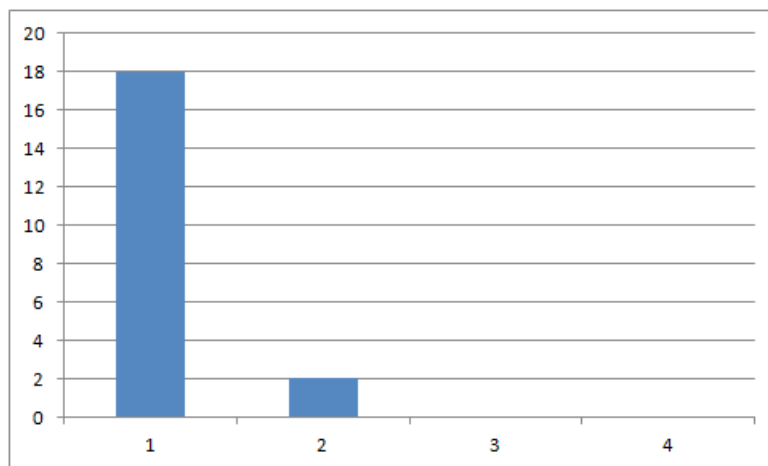
Tab. 4. Cytowalność prac w naukach ścisłych i humanistycznych z podziałem na typ wydawnictwa

	Nauki ścisłe	Nauki humanistyczne
Wydawnictwa zwarte	Grupa 1 93% (81/87)	Grupa 2 87% (234/270)
Wydawnictwa ciągłe	Grupa 3 73% (192/263)	Grupa 4 69% (55/80)
Łącznie	78% (273/350)	83% (289/350)

Źródło: opracowanie własne.

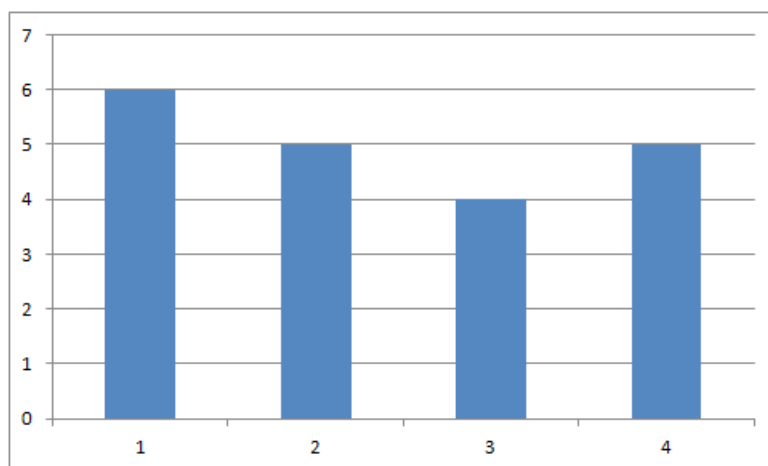
Chociaż łączny wynik oczywiście nie zmienił się, warto przyjrzeć się bliżej danym cząstkowym, z których wynika, że w każdej z obydwu podgrup to dziedzina nauk ścisłych wychodzi na zdecydowane prowadzenie. Można by się więc pokusić o sformułowanie konkluzji: „w obrębie każdej z kategorii więcej prac cytowanych jest w ramach nauk ścisłych” lub „w obrębie każdej z kategorii w naukach humanistycznych więcej prac nie ma dalszego wpływu na rozwój nauki”. Choć wnioski te wydają się sprzeczne, to przecież zostały wywiedzione z bezbłędnie przeprowadzonych obliczeń i dopiero bliższe przyjrzenie się zawartości tabeli pozwala zrozumieć przyczynę tej sytuacji. Dlatego zawsze warto próbować lepiej zrozumieć naturę analizowanego problemu oraz sięgać głębiej w analizowane grupy danych (ang. *drill down*), gdyż może zdarzać się, że w miarę przechodzenia przez kolejne ich agregacje wyniki liczbowe sugerować będą wnioski przeciwne do wysnuwanych poprzednio. Efekt działania kilku grup danych odwrotny do efektu uzyskanego poprzez połączenie tych grup znany jest już od połowy XX wieku i nazywany paradoksem Simpsona.

Drugi z przykładów dotyczyć będzie równie elementarnych obliczeń – wyciągnięcia wartości średnich z liczby cytowań zbioru artykułów. W taki właśnie sposób oblicza się wartość czynnika IF dla wybranych czasopism, a jego wartość przesądza o atrakcyjności i prestiżu danego czasopisma. Syntetyczny i elementarny przykład przedstawiono na poniższym wykresie.



Wykr. 1. Liczba cytowań 4 artykułów w przykładowym czasopiśmie
Źródło: opracowanie własne.

Widzimy tutaj 4 artykuły, z których jeden cytowany był 18 razy, drugi raz, a pozostałe ani razu. Średnio każdy z artykułów cytowany był 5 razy. Identycznie jak w sytuacji przedstawionej kolejnym wykresie.



Wykr. 2. Liczba cytowań 4 artykułów w innym czasopiśmie
Źródło: opracowanie własne.

Mimo tych samych wartości średnich już na pierwszy rzut oka zobaczyć można, że zbiory wartości w tych dwóch przypadkach mają zupełnie inny charakter. Rozrzut wartości pierwszego jest prawie o rząd wielkości większy od rozrzutu wartości drugiego. Oznacza to, że co prawda wartość oczekiwana liczby cytowań artykułu opublikowanego w obydwu czasopismach jest taka sama, to jednak prawdopodobieństwo za-

cytowania danego artykułu przynajmniej raz jest dla nich odmienna. Warto tutaj zwrócić uwagę, że w poszczególnych dziedzinach nauki średnia liczba cytowań jest bardzo zmienna, więc jedną z przyczyn dużego rozrzutu cytowań danego czasopisma może być np. jego interdyscyplinarność.

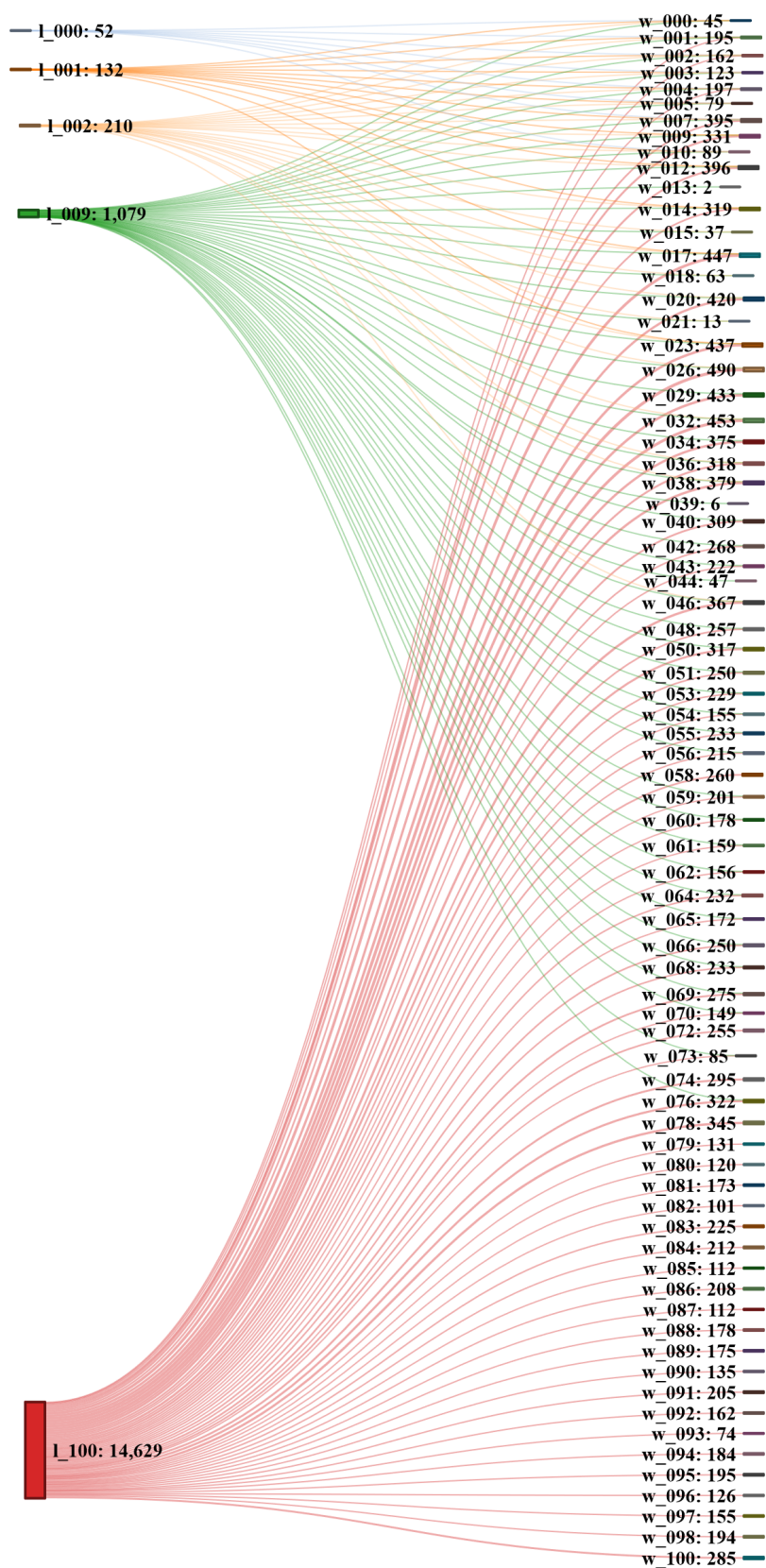
Te same dane a różne wyniki?

Poszukiwania rozwiązań problemów identyfikowanych w obszarach ewaluacji nauki metodami bibliometrycznymi zwykle dotyczą prób opracowywania nowych wskaźników ewaluacyjnych bazujących na zastanym i niezmiennym od samego początku powstania metod bibliograficznym modelu danych. Rodzi się pytanie, dlaczego mimo wieloletnich już poszukiwań doskonalszych metod oraz ogromnego jeszcze potencjału metod algebraicznych nie powstały dotąd narzędzia, których wyniki mogłyby być powszechnie akceptowalne. Być może powodują to ograniczenia samego modelu danych, które zauważalne stają się dopiero na przykład przy wagowym, a nie tylko ilościowym, postrzeganiu relacji autorstwa czy cytowań.

Na naukowych portalach społecznościowych obserwować można wiele ożywionych dyskusji na temat dobrych praktyk dotyczących deklarowania wkładu autorskiego za pomocą kolejności¹⁵ [sic!] opisu atrybucji autorskiej czy sposobów cytowania wykorzystywanych źródeł¹⁶ w ramach danej publikacji oraz tzw. cytowań negatywnych, czyli zaprzeczających tezom stawianym w cytowanej pracy. Z tak znacznego zainteresowania tymi tematami wynika nie tylko duża świadomość specyfiki badań naukowych, ale przede wszystkim potrzeba uczynienia ich bardziej transparentnymi. Droga ku temu nie może jednak prowadzić metodami półśrodków polegających na zachowywaniu nieformalnych konwencji w formatowaniu danych bibliograficznych rozumianych jedynie przez grono badaczy, które się na nie umówiło, gdyż i tak nie zostaną one uwzględnione przez powszechnie stosowane metody ewaluacyjne.

¹⁵ LINTO, G. T. *How do authors decide upon first, second, third and last author?* [online]. [Dostęp 18.03.2018]. Dostępny w: https://www.researchgate.net/post/How_do_authors_decide_upon_first_second_third_and_last_author.

¹⁶ AKHILESH, G. *What is the importance of citation?* [online]. [Dostęp 18.03.2018]. Dostępny w: https://www.researchgate.net/post/What_is_the_importance_of_citation.



II. 5. Porównanie jednostkowego (ilościowego) i ważonego rankingu publikacji naukowych na podstawie cytowań bazy CYTBIN

Źródło: opracowanie własne.

Wydaje się, że wprowadzenie wagowego atrybutu relacji autorstwa (czyli możliwości wyrażenia procentowego wkładu autorskiego) oraz wagowego atrybutu relacji cytowania (czyli możliwości wyrażenia wielkości wpływu jednostki cytowanej na daną publikację) mogłoby otworzyć nowe kierunki bardziej transparentnej ewaluacji nauki z większą możliwością uczestniczenia w niej samych zainteresowanych, czyli naukowców. Podobnie jak w przypadku autorstwa, tak i w przypadku cytowań należałoby założyć, że całkowity wkład prac cytowanych zamykać się musi w 100%, co uczyniłoby relację cytowania zasobem ograniczonym, a przez to bardziej racjonalnie wykorzystywanym. Dopuszczenie możliwości przypisania wkładu o zerowej wartości rozwiązałoby również potrzebę możliwości negatywnego cytowania artykułów, bądź zamieszczania spisów literatury uzupełniającej.

Uwzględnianie atrybutów wagowych nawet w przypadku elementarnych metod bibliometrycznych spowodować może znaczne przetasowania w rankingach produktywności czy wkładu naukowego, ale na skutek tych przetasowań (czyli zmian wartości miar ewaluacji) największy wpływ będą sami badacze. Na ilustracji 5. zobrazowano porównanie tradycyjnego rankingu cytowań artykułów z bazy CYTBIN do rankingu uwzględniającego wartości wagowe. Ponieważ jednak w bibliograficznych bazach danych wartości wagowe relacji cytowań nie są rejestrowane, to na potrzeby przedstawionego porównania przyjęto równomierny rozkład wag w ramach pozycji bibliografii załącznikowych poszczególnych artykułów. Oznacza to na przykład, że jeśli artykuł A cytuje artykuł X oraz Y i artykuł B cytuje X, Y, Z to artykuł X oraz artykuł Y osiągną sumaryczne wartości (5/6), zaś artykuł Z wartość (1/3). Porównanie zostało skonstruowane z wykorzystaniem rankingu percentylowego i wartości poszczególnych jego „wierszy” obrazują udział procentowy wyników lepszych lub równych danemu wierszowi – im więc wartość jest mniejsza, tym wyższa pozycja w rankingu. Po lewej stronie umieszczono wyniki uzyskane tradycyjną metodą sumowania liczby cytowań. Zaobserwować tu można dość małą granulację uzyskanych wyników, gdyż artykuły cytowane do dwóch razy, stanowiące większość w analizowanym korpusie danych, znajdują się w ostatnim percentylu. Warto zwrócić uwagę na wysokości prostokątów reprezentujących poszczególne przedziały rankingu, gdyż świadczą one o liczebności zbioru elementów (suma wysokości prostokątów z lewej strony równa się sumie wysokości prostokątów po stronie prawej). Dużo większe zróżnicowanie rankingu zaobserwować można po stronie prawej, obrazującej wagowe zliczanie cytowań – wynika ono z tego, że choć większość artykułów cytowana była do dwóch razy, to jednak liczba pozycji bibliografii załącznikowych je zawierająca była różnorodna.

Szczegółowa analiza konkretnych pozycji przedstawionego rankingu, choć oczywiście jest możliwa do przeprowadzenia, nie jest celem niniejszych rozważań, a przedstawiony przykład został przywołany jedynie w celu zilustrowania rozmiaru różnic wyników dwóch metod mających te same zastosowania i operujących na tych samych danych. Linie łączące poszczególne grupy rankingów przedstawionych po stronie lewej i prawej wskazują, że zdarzają się nawet jednostki bibliograficzne z ostatniej pozycji rankingu ilościowego, które potrafią znaleźć się wyżej w rankingu wagowym, niż niektóre jednostki z pierwszej pozycji rankingu ilościowego. Analizując zjawiska największych spadków rankingu wagowego w stosunku do ilościowego, zaob-

serwowano częste ich występowanie w przypadku wydawnictw zwartych takich jak słowniki czy kodeksy, które są cytowane często, ale jako literatura uzupełniająca, a więc w spisach bibliograficznych o znacznej liczbie pozycji (a tym samym stosunkowo niewielkim wkładzie wagowym). Zachowanie w takich przypadkach zaproponowanego w niniejszym artykule rankingu wagowego wydaje się więc być bardziej zgodne z intuicyjnym rozumieniem uzupełniającego, a niekoniecznie przewodniego wkładu takich pozycji w rozwój nauki.

Podsumowanie

Obserwując tempo postępu technologicznego, które zawdzięczamy w dużej mierze intensyfikacji badań naukowych, zastanawiać może fakt zastoju metod ewaluacji produktywności naukowej przy jednoczesnej krytyce metod obecnie stosowanych. O ile jakość i kompletność danych bibliograficznych gromadzonych w krajowych bazach danych może stanowić pewną przeszkodę w rozwoju badań nad tymi metodami, to komercyjne instytucje oferujące usługi wyliczania wskaźników bibliometrycznych skupiają się przede wszystkim nad rozwojem ilościowym (pozyskiwaniem coraz to nowych źródeł danych), a w mniejszym stopniu jakościowym (rozwijaniem metod operujących na gromadzonych danych), nie zauważając głosów zwątpienia w możliwość bezpośredniej interpretacji dostarczanych wskaźników. Wobec przytoczonych w artykule przykładów niespójności zachowania niektórych wskaźników czy braku możliwości uwzględniania przy ich wyliczaniu wielkości wkładu poszczególnych autorów czy cytowanych jednostek bibliograficznych trudno przejść obojętnie. Wydaje się, że rozbudowa modelu danych bibliograficznych na potrzeby analiz bibliometrycznych o brakujące wyżej wymienione informacje mogłaby pozwolić na wyjście metod ewaluacji nauki z błędnego koła. Oczywiście rozbudowa modelu jest tylko jednym z czynników poprawy transparentności ewaluacji. Choć w każdej dziedzinie nauki publikowanie wyników badań należy do dobrej praktyki i stanowi wspólny, zarówno komunikacyjny, jak i bibliometryczny, mianownik działalności naukowej, to przekazywanie wyników tych badań na produkt społecznie konsumowalny może już przebiegać bardzo różnie, a przecież wartość badań naukowych powinna być mierzona również jako pochodna i tego produktu. Wydaje się, że wypracowanie sposobów śledzenia wstecz (ang. *traceability*) naukowego łańcucha dostaw (badań) wprost od produktów, które dzięki nim powstały, pozwoliłoby przesunąć bibliometrycznym metodom ewaluacji punkt pomiaru z „produktywności publikacyjnej” na „produktywność użytkową”¹⁷.

Bibliografia:

1. AKHILESH, G. *What is the importance of citation?* [online]. [Dostęp 18.03.2018]. Dostępny w: https://www.researchgate.net/post/What_is_the_importance_of_citation.
2. BADHAM J. *Gry wojenne* [film]. USA: United Artists, Sherwood Productions, 1983.
3. BŁOCKI, Z., ŻYCKOWSKI, K. Czy można porównywać jabłka i gruszki? O danych bibliometrycznych w różnych dziedzinach nauki. *Nauka* 2013, nr 2, s. 37–46. ISSN 1231-8515.
4. EGGHE, L. Theory and practise of the g-index. *Scientometrics* 2006, nr 1, s. 131–152. ISSN 0138-9130.

¹⁷ Rozumianą oczywiście szerzej niż materializacje w postaci przedmiotów codziennego użytku.

5. HELLER, M. *Granice nauki*. Kraków: Copernicus Center Press, 2014. ISBN 978-83-788-6049-5.
6. JOVANOVIĆ, M.M. The Vicious Circle of Evaluation Transparency – An Ignition Paper. W: *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference*. Istanbul: Bogazici University Printhouse, 2015, s. 646–647. ISBN 978-975-518-381-7.
7. KAMIŃSKA, A.M. Miary podobieństw łańcuchów znakowych a deduplikacja rekordów w bibliograficznych bazach danych. *Przegląd Biblioteczny* 2017, nr 4, s. 477–495. ISSN 0033-202X.
8. KAMIŃSKA, A.M. ProBIT – prospektywna metoda tworzenia trawersowalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych. *Zagadnienia Informacji Naukowej* 2017, nr 1, s. 66–82. ISSN 0324-8194.
9. KAMIŃSKA, A.M. Zastosowanie struktur grafowych do analiz bibliometrycznych i webometrycznych. Modele i metody. *Nowa Biblioteka. Usługi, Technologie Informacyjne i Media* (w druku).
10. KOSMULSKI, M. A new Hirsch-type index safes time and works equally well as the original h-index. *ISSI Newsletter* [online]. 2006, nr 2. [Dostęp 18.03.2018]. ISSN 1998-5460. Dostępny w: <http://sci2s.ugr.es/sites/default/files/files/TematicWebSites/hindex/kosmulski2006.pdf>.
11. LINTO, G. T. *How do authors decide upon first, second, third and last author?* [online]. [Dostęp 18.03.2018]. Dostępny w: https://www.researchgate.net/post/How_do_authors_decide_upon_first_second_third_and_last_author.
12. STABLEFORD, B. *Gry wojenne*. Bydgoszcz: Express Books, 1990. ISBN 83-8517-504-0.
13. WALTMAN, L., VAN ECK, N. The Inconsistency of the H-index. *Journal of the American Society for Information Science and Technology* [online]. 2012, nr 2, s. 406–415. [Dostęp 18.03.2018]. Dostępny w doi: 10.1002/asi.21678.
14. WU, Q. The w-index: A measure to assess scientific impact by focusing on widely cited papers. *Journal of the American Society for Information Science and Technology* [online]. 2010, nr 3, s. 609–614. [Dostęp 18.03.2018]. Dostępny w doi: 10.1002/asi.21276.