
Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych.

Analiza piśmiennictwa.

Część 1. Publikacje pionierskie, metoda powiązań bibliograficznych, metoda współcytowań i metoda współwystępowania specjalistycznej terminologii naukowej

Łukasz Opaliński

Biblioteka Politechniki Rzeszowskiej im. I. Łukasiewicza

Abstrakt

Cel/Teza: Celem pracy jest dokonanie krytycznego przeglądu literatury naukowej dotyczącej metod ilościowych stosowanych w obszarze bibliometrycznego prognozowania przyszłego rozwoju dyscyplin naukowych. Przez „przegląd krytyczny” rozumie się wskazanie i omówienie mankamentów prezentowanych w literaturze przedmiotu podejść do rozpatrywanego problemu.

Koncepcja/Metody badań: Przyjęta w pracy metoda polega na analizie literatury przedmiotu, stworzeniu typologii wykorzystywanych w ramach bibliometrii metod prognozowania i oceny rozwoju dyscyplin, podkreśleniu ograniczeń stosowanych w tym zakresie metod oraz wyciągnięciu wniosków końcowych. Postarano się również zaznaczyć obszary, w których przejawia się praktyczna stosowalność prognoz tempa i kierunku rozwoju dyscyplin nauki.

Wyniki i wnioski: Z uwagi na fakt, że niniejszy artykuł stanowi pierwszą z dwóch części, na jakie podzielono całość analizy piśmiennictwa, przedstawione w jego ramach wyniki mają jedynie wstępny charakter. Podkreślono w ich ramach wiodącą i wciąż aktualną rolę metod zapoczątkowanych w publikacjach pionierskich, wskazano, że towarzyszące im mankamenty i próby ich przewyżczenia stanowią integralną część ewolucji całego omawianego nurtu, a także zaznaczono istnienie silnego związku pomiędzy bibliometryczną metodologią zorientowaną prognostycznie a algorytmicznymi metodami grupowania publikacji w klastry dokumentów spójnych tematycznie, które reprezentują dyscyplinarne obszary problemowe.

Oryginalność/Wartość poznawcza: Oryginalność i wartość poznawcza pracy polega na podjęciu próby przybliżenia rozległości i stopnia zaawansowania badań prowadzonych nad przewidywaniem i oceną stanu rozwoju dyscyplin na świecie. Badania te są realizowane w polskim środowisku naukowym stosunkowo rzadko w związku z czym wydaje się, że przegląd przyjmowanej w ich ramach metodologii mógłby przyczynić się do wzbudzenia większego zainteresowania tą problematyką. Zaakcentowanie słabych stron poszczególnych metod jest w tym kontekście podstawą do dalszego modyfikowania i udoskonalania poszczególnych metod. W obrębie całego spektrum bibliometrii istnieje bowiem jeszcze wiele przestrzeni dla tego rodzaju przedsięwzięć, ponieważ obecnie o żadnej z metod nie można powiedzieć, że idealnie spełnia ona postawione przed nią zadanie i jest „ostatnim słowem” w ramach całego nurtu prognostycznego.

Słowa kluczowe

Bibliometria. Metody ilościowe w informatologii. Prognozowanie. Rozwój dyscyplin naukowych.

Otrzymano: 3 stycznia 2017. Zrecenzowano: 10 lipca 2017. Zaakceptowano: 5 sierpnia 2017.

1. Wprowadzenie

Za twórcę terminu „bibliometria” (ang. *bibliometrics*) powszechnie uznaje się Alana Pritcharda, który zaproponował go w miejsce używanej dotychczas nazwy „bibliografia statystyczna” (ang. *statistical bibliography*) (zob. np.: Diodato, 1994, 14; Nowak, 2006, 16; Skalska-Zlat, 1988, 259). Określenie „bibliografia statystyczna” zostało z kolei użyte po raz pierwszy przez Edwarda Wyndhama Hulme’a w 1922 r. podczas wykładów na University of Cambridge, wydanych rok później w książce pt. *Statistical bibliography in relation to the growth of modern civilization: two lectures delivered in the University of Cambridge in May, 1922* (Hertzfel, 1987, 150; Nowak, 2006, 16; Pindlowa, 1988, 305; Pritchard, 1969, 348). Niedawno Ronald Rousseau zakwestionował jednak pierwszeństwo Pritcharda w tym względzie i stwierdził, że termin „bibliometria” (fr. *bibliométrie*) pojawił się już wcześniej w książce pt. *Traité de Documentation* autorstwa belgijskiego prawnika i bibliografa, twórcy nauki o dokumentacji – Paula Otleta, wydanej w 1934 r. (Rousseau, 2014, 218). Nazwisko Otleta jest pomimo tego współcześnie kojarzone przede wszystkim z Uniwersalną Klasyfikacją Dziesiętną (UKD).

Pritchard zdefiniował bibliometrię jako „(...) zastosowanie metod matematycznych i statystycznych do książek i innych środków komunikacji [piśmienniczej]”¹ (Pritchard, 1969, 348–349). Z punktu widzenia tematyki niniejszego artykułu ważne wydaje się również uzupełnienie definicji bibliometrii Pritcharda o jeden z jej celów, jakim w opinii Pritcharda było „naświetlenie (...) natury i kierunków rozwoju dyscyplin [naukowych] (...) za pomocą zliczania i analizowania różnorodnych aspektów komunikacji piśmienniczej”² (Pritchard, 1969, 348). Tradycyjne zadania stojące przed bibliometrią to więc m. in. ilościowe badania struktury, koncentracji, rozproszenia czy przyrostu literatury naukowej w czasie, jej starzenia się, porównawcze oceny produktywności indywidualnych naukowców lub instytucji badawczych, ewaluacyjna (wartościująca) analiza cytowań lub ustalanie zależności pomiędzy czynnikami biorącymi udział w procesach komunikowania i transferu wiedzy (np. tzw. prawo Bradforda, prawo Lotki, prawo Zipfa). Poza ogółem tych zadań istnieje jeszcze jedna grupa zagadnień należących do przedmiotu bibliometrii, na którą składają się: ocena dynamiki zmian i prognozowanie dalszego rozwoju dyscyplin naukowych oraz wyodrębnianie nowo powstających i nie w pełni jeszcze ukształtowanych (ang. *emerging*) subdyscyplin (obszarów problemowych czy, inaczej mówiąc, wąskich specjalności dyscyplinarnych³)

¹ Oryg.: „(...) the application of mathematics and statistical methods to books and other media of communication” (Pritchard, 1969, 348–349).

² Oryg.: „(...) to shed light on (...) the nature and course of development of a discipline (...), by means of counting and analysing the various facets of written communication” (Pritchard, 1969, 348).

³ W tym miejscu nasuwa się konieczność doprecyzowania wykorzystywanej w artykule terminologii. Zgodnie z *Rozporządzeniem Ministra Nauki i Szkolnictwa Wyższego z dnia 8 sierpnia 2011 r. w sprawie obszarów wiedzy, dziedzin nauki i sztuki oraz dyscyplin naukowych i artystycznych* (Dz.U. 2011, nr 179,

(Skalska-Zlat, 1993, 20; Skalska-Zlat, 1999, 60). Z tą ostatnią funkcją działalności bibliometrycznej ściśle związane jest tzw. mapowanie (ang. *mapping*), czyli odwzorowywanie i wizualizowanie epistemologicznej struktury różnych dyscyplin naukowych, realizowane obecnie często za pomocą wyspecjalizowanych programów komputerowych, takich jak np. *CiteSpace* (zob. np.: Chen, 2006), *Pajek* (zob. np.: Liu & Lu, 2012, 529) lub *BibExcel* (zob. np.: Liu & Gui, 2016, 162). Wizualizacje te określa się jako „topologiczne mapy nauki” (ang. *topological maps of science*) (zob. np.: Garfield, 2004, 119). Analiza takiej mapy, będącej w rzeczywistości siatką cytowań, bazuje na założeniu, że właściwości strukturalne siatki są związane z jej właściwościami dynamicznymi. Można też, alternatywnie, tworzyć mapy w oparciu o publikacje wydane w określonych przedziałach czasowych (ang. *time slices*), uzyskując w ten sposób serię map i obserwować ich transformacje (zob. np.: Diodato, 1994, 37–38; Shibata et al., 2007, 876; Verbeek et al., 2002, 203; Zitt & Bassecoulard, 1994, 334–335). W szczególności badanie wieku zbioru powiązanych publikacji (tj. określenie średniej wieku publikacji wchodzących w skład tego zbioru) jest jednym ze sposobów na wczesne wykrywanie młodych, nie w pełni jeszcze ukształtowanych obszarów problemowych, jako „zaczątków” nowych subdyscyplin. Dodatkowym wskaźnikiem jest tutaj cytawalność takich zbiorów. Im jest ona wyższa, tym szersze zainteresowanie w środowisku naukowym wzbudza taki zbiór i tym bardziej prawdopodobne jest, że w przyszłości zbiór ten będzie się rozrastać i ewoluować w stronę samodzielnej subdyscypliny lub dyscypliny (Shibata et al., 2008, 765; Small, 2006, 596). Niektórzy badacze twierdzą przy tym, że okres największego zainteresowania i najintensywniejszych prac prowadzonych w ramach tego odgałęzienia bibliometrii przypada na ostatnie 15 lat (Ena et al., 2016, 1015; Wang & Ho, 2016, 482), chociaż początków realizacji takich zadań można dopatrywać się już w latach pięćdziesiątych XX w., zaraz po założeniu filadelfijskiego Institute for Scientific Information (Boyack & Klavans, 2014, 671). Obecnie można nawet spotkać się z opinią, że zadanie „mapowania” nauki, wizualizacja dyscyplin naukowych i rozpoznawanie – mówiąc słowami Dereka de Solli Price’a – „frontów badawczych” (ang. *research fronts*) jest naczelnym, pierwszoplanowym komponentem bibliometrii (Huang & Chang, 2015, 2042; Price, 1967, 114.; Wang & Ho, 2016, 482).

poz. 1065) terminem o najszerszym zakresie znaczeniowym jest „obszar wiedzy” (np. obszar nauk przyrodniczych). Terminem węższym jest „dziedzina nauki” (np. dziedzina nauk o Ziemi), a terminem najwęższym jest „dyscyplina” (np. geologia). Terminy takie jak „domena”, „subdyscyplina”, „specjalność” czy „obszar problemowy” nie zostały zdefiniowane w ramach ww. Rozporządzenia, w związku z czym autor niniejszego artykułu uznał, że należy dodatkowo wyjaśnić sens, w jakim używane są one w ramach niniejszego opracowania oraz, że pod warunkiem zachowania konsekwencji w posługiwaniu się nimi dopuszczalna jest pewna dowolność w ich zdefiniowaniu. Z tego względu zdecydowano, że „domena”, „subdyscyplina” i „specjalność” mogą być używane wymiennie i że wszystkie one desygnują struktury węższe niż dyscypliny w rozumieniu ww. Rozporządzenia MNiSW (np. za specjalności w ramach geologii można byłoby uznać paleoklimatologię lub sedimentologię). Określenie „obszar problemowy” charakteryzuje się natomiast jeszcze węższym znaczeniem, które odnosi się do zespołu hipotez lub propozycji teoretycznych związanych ściśle z jakimś jednym problemem rozpatrywanym w ramach specjalności. Jeśli specjalnością byłaby wspomniana paleoklimatologia, jednym z jej obszarów problemowych mogłyby być np. wyjaśnianie tektonicznych przyczyn zmian klimatycznych zachodzących w epoce wczesnego plejstocenu. Jeżeli taki zespół hipotez czy twierdzeń udałoby się potwierdzić doświadczalnie lub zbudować dla nich spójną konstrukcję teoretyczną wyjaśniającą je dostatecznie wiarygodnie i kompleksowo, obszar problemowy mógłby wtedy przekształcić się w nową specjalność (subdyscyplinę czy domenę).

Przewidywanie dalszego rozwoju dyscyplin naukowych i jego tempa oraz kluczowych kierunków tego rozwoju niesie ze sobą rozliczne korzyści praktyczne. Jest ono na tyle istotne, że istnieją nawet towarzystwa naukowe i cykliczne konferencje, które poświęca się wyłącznie tym zagadnieniom (Small et al., 2014, 1450).

Po pierwsze może być ono cenne z punktu widzenia samych naukowców. Śledzenie najnowszych tendencji dyscyplinarnych i identyfikacja najbardziej perspektywicznej, priorytetowej tematyki i pytań badawczych (ang. *hot topics*) pozwoliłyby naukowcom nadażyć za gwałtownym, wielopłaszczyznowym rozwojem dyscyplin i racjonalnie ukierunkować swoje wysiłki badawcze, a także ułatwiłoby im planowanie przyszłej współpracy naukowej. Ponieważ obecnie w erze „zalewu informacyjnego” (zob. np.: Mabe, 2010, 138; Meadows, 2000; Muraszkiewicz, 2014) do tych celów nie wystarcza już bieżące monitorowanie najważniejszych czasopism dyscyplinarnych (np. czasopism przeglądowych), powstaje potrzeba zaprojektowania efektywniejszych, zautomatyzowanych metod wykrywania aktualnie wiodącej problematyki (Wang et al., 2013, 717; zob. też: Takeda & Kajikawa, 2009, 544; Tu & Hsu, 2016, 2016–2017). Przyczyną, dla której uwagę aktywnych w swojej dyscyplinie badaczy mogą przykuwać nowe („młode”) i obiecujące kierunki badań, jest także możliwość zaistnienia w międzynarodowej społeczności naukowej poprzez np. nawiązanie współpracy z cenionymi badaczami i udział w zespołach publikujących wpływowe prace, mające duże szanse na pozyskanie wielu cytowań (Guo et al., 2011, 421). Publikowanie w ramach specjalności rozwojowej niesie bowiem ze sobą zdecydowanie większe perspektywy, niż publikowanie w obszarze, który społeczność naukowa uznała już za wyeksploatowany i mało progresywny, inaczej mówiąc za taki, który okres rozkwitu ma już za sobą i w którym panuje stagnacja (Sun et al., 2016, 750). Duże znaczenie ma też tutaj szybkość, z jaką poszczególni naukowcy będą w stanie pozyskać, przyswoić i wykorzystać wiedzę naukową rozwijającą się w młodej i dopiero formującej się specjalności. Im szybciej będą w stanie to osiągnąć i im bardziej zdystansują potencjalnych rywali, tym bardziej podniosą swoją konkurencyjność (ang. *competitive advantage*) i zyskają prestiż związany potencjalnie z posiadaniem statusu jednego z współtwórców czy pionierów nowo powstałej domeny (Shibata et al., 2008, 758).

W omawianym kontekście praktycznego znaczenia prognoz rozwoju dyscyplin naukowych wydaje się ponadto, że warto dodatkowo zaznaczyć, iż analizy tego rodzaju mogą okazać się nie bez znaczenia także dla bibliotekarzy, a zwłaszcza dla bibliotekarzy dziedzinowych. Hanning Guo i in. (2011) wskazali bowiem, że także i ta grupa zawodowa może odnieść korzyść z posługiwania się narzędziami ułatwiającymi tworzenie nowych kategorii lub haseł przedmiotowych w trakcie rzeczowego opracowywania zbiorów, uzupełniania istniejących języków informacyjno-wyszukiwawczych (np. tezaurusów) o nowe terminy albo wyodrębnianie specjalnych kolekcji zbiorów bibliotecznych, ukierunkowanych na potrzeby informacyjne badaczy prowadzących studia o profilu zgodnym z nowo powstającymi subdyscyplinami (Guo et al., 2011, 422).

Następnym obszarem, w którym zauważalne są praktyczne korzyści płynące z bibliometrycznego prognozowania przyszłego rozwoju dyscyplin naukowych, jest ogólnie rozumiana polityka naukowa, w skład której wchodzi administrowanie i zarządzanie działalnością naukową oraz kwestie finansowania naukowych przedsięwzięć z budżetu państwa (zob. np.: Braun et al., 2000, 24). Ponieważ badania naukowe są „zaczynem” (ang. *seeds*) innowacyjności przemysłowej, znaczenie prac badawczo-rozwojowych (ang. *R&D – Research and*

Development) jest uznawane za fundamentalne w procesie stymulowania innowacyjności i konkurencyjności narodowego przemysłu i technologii (Shibata et al., 2008, 758–759). W warunkach polskich prace badawczo-rozwojowe obejmują programy rządowe dotyczące naukowo-technologicznego rozwoju państwa, badania marketingowe i strategie rozwoju przedsiębiorstw. Ich celem w tej perspektywie jest proces doskonalenia i usprawniania techniki i technologii oraz kreowanie zrównoważonej polityki w zakresie rozwoju technologicznego (zob. np.: KPB, 2011; MNil, 2004). Istotną kwestią jest tutaj również racjonalna alokacja funduszy na cele naukowe, która byłaby zgodna z badawczymi priorytetami, a także możliwościami danego kraju lub instytucji (zob. np.: Huang & Chang, 2015, 2041–2042; Kajikawa et al., 2008, 772–773; Lee et al., 2012, 814–815; Tseng et al., 2009, 73–74). Finansowanie badań naukowych w racjonalny sposób przyczyniłoby się do dokonywania odkryć, które mogą znaleźć zastosowania komercyjne (np. w przemyśle) i zaowocować dalszymi odkryciami o naturze interdyscyplinarnej. Interdyscyplinarność jest natomiast postrzegana jako „siła napędowa” innowacyjności (Upham & Small, 2010, 15–17; zob. też: Sun et al., 2016, 751).

Jeszcze jedną sposobnością praktycznego wykorzystania wiedzy o rozwoju dyscyplin naukowych jest zastosowanie zbliżone do omówionych wyżej, ale usytuowane w sektorze prywatnym. Zarządzanie technologiami pozwala bowiem na poprawę efektywności wykorzystania technologicznego i finansowego kapitału, planowanie krótko- lub długoterminowych strategii produkcyjno-rynkowych, jak również przewidywanie cykli życiowych produktów technologicznych i racjonalizację dalszych inwestycji (Chen et al., 2012, 1099). Ponieważ wszelkie strategie inwestycyjne są nieodłącznie związane z ryzykiem, np. co do właściwego ulokowania środków finansowych czy długości cyklu życia produktu (czy popyt na taki produkt będzie trwał wystarczająco długo aby zwróciły się koszty jego produkcji, promocji i rozprowadzania?), pożądanym byłby dostęp do metod pozwalających na zredukowanie tego ryzyka i wskazanie takich obszarów, w których inwestowanie jest potencjalnie najbardziej opłacalne (Érdi et al., 2013, 226; Lee, 2008, 504).

Niejako przy okazji identyfikowania najważniejszych ścieżek historycznego rozwoju dyscyplin (złożonych ze szczególnie wpływowych publikacji) oraz budowania prognoz, pojawia się możliwość przewidzenia tego, które z artykułów mogą stać się w przyszłości tzw. publikacjami kluczowymi, centralnymi dla pewnych subdyscyplin (ang. *core papers*). Na tej samej podstawie możliwe stałoby się również konstruowanie zautomatyzowanych systemów rekomendujących publikacje relewantne do pewnej tematyki (ang. *recommendation systems*). Mogłyby być one pomocne tak dla samych autorów, jak i dla recenzentów oceniających bibliografie załącznikowe publikacji pod kątem adekwatności przywoływanych w nich pozycji (Shibata et al., 2012, 78–79).

2. Cel pracy

Celem pracy jest dokonanie krytycznego przeglądu literatury naukowej, dotyczącej metod ilościowych stosowanych w obszarze bibliometrycznego prognozowania dalszego rozwoju dyscyplin naukowych. W szczególności istotne wydaje się w tym kontekście wskazanie mankamentów poszczególnych podejść wykorzystywanych przez autorów zajmujących się tego rodzaju zagadnieniami. Mankamenty te sprawiają, że wśród wskazanych i omówionych

podejść pozostaje wciąż wiele miejsca na dalsze udoskonalenia, modyfikacje i eksperymenty związane z projektowaniem nowych metod służących realizacji celów prognostycznych.

Literatura przedmiotu dotycząca wyodrębniania nowo powstających specjalności w ramach dobrze ugruntowanych teoretycznie dyscyplin oraz prognozowania ich dalszego rozwoju jest nie tylko bardzo obszerna, ale też wysoce zróżnicowana pod względem stosowanych w tym zakresie metod i samego przedmiotu badań. Przedmiotem tym mogą być bowiem zarówno artykuły naukowe, patenty czy konferencje, jak i autorzy publikacji naukowych i sieć współpracy między nimi. Z uwagi na to autor niniejszej pracy zdecydował, że najkorzystniejsze z punktu widzenia czytelnika oraz z punktu widzenia przejrzystości przeglądu literatury, będzie podzielenie go na podgrupy publikacji pod pewnymi względami podobnych. Podział ten można określić mianem „typologii”, ponieważ nie jest on tak rygorystyczny, jak wymaga tego klasyfikacja, kategoryzacja czy systematyka. Typologia oznacza bowiem podzielenie elementów pewnego zbioru na podzbiory elementów pod pewnymi względami do siebie podobnych. Elementy te nie muszą być więc identyczne, ale musi łączyć je pewien, na ogół różny, stopień podobieństwa. Ponadto wyróżnione podzbiory (typy) nie muszą być rozłączne, ponieważ niektóre elementy mogą spełniać warunek podobieństwa w stosunku do więcej niż jednego elementu wzorcowego (Sosińska-Kalata, 2002, 19). Cechą typologii jest ponadto tzw. infinityzm, czyli akceptacja nieskończonej różnorodności poddawanych podziałowi elementów i dopuszczalność abstrahowania ich cech wspólnych bez konieczności definiowania ściśle jednolitego kryterium wyboru (Sosińska-Kalata, 2002, 26).

Wydzielono w sumie 11 typów publikacji mających pewne cechy wspólne. Należy przy tym wyjaśnić, że z uwagi na wspomnianą rozległość literatury przedmiotu niniejszy artykuł stanowi pierwszą z dwóch części, na jakie podzielono całość analizy piśmiennictwa. Każdy z kolejnych podrozdziałów (od 3 do 7) omawia jedną z metod, które w przeciwieństwie do metod demonstrowanych w części drugiej można określić mianem tradycyjnych (klasycznych, konwencjonalnych, podstawowych). Są nimi po pierwsze, metody wypracowane w publikacjach pionierskich, które zapoczątkowały jakąś technikę badań lub po raz pierwszy postawiły problem, który był następnie rozwijany, modyfikowany lub szeroko komentowany przez kolejnych badaczy. W dalszej kolejności omówiono bardziej szczegółowo każdą z tych metod, tj. wskazano w jaki sposób ewoluowały, jakie zastosowania zostały dla nich odnalezione oraz jakiego rodzaju wyników można oczekiwać w związku z zastosowaniem każdej z nich. W ramach tego omówienia starano się zachować porządek chronologiczny, w jakim wprowadzano do naukoznawstwa wszystkie metody. Za najwcześniejszą z nich można uznać metodę generowania powiązań bibliograficznych, za niemal równie wczesną – tzw. metodę historiograficzną (zob. podrozdz. 3). Następnie, po około dziesięciu latach, w obszarze bibliometrii pojawiła się metoda współcytowań, która sama z kolei wyprzedziła o około 13 lat metodę badania współwystąpień terminów naukowych Callona i in. (zob. podrozdz. 7). Ostatnią z metod tradycyjnych, której prognostyczny potencjał dostrzeżono jednak z pewnym opóźnieniem, była analiza tzw. głównych ścieżek rozwoju dyscyplin naukowych oparta w znacznej mierze na matematycznej teorii grafów (zob. podrozdz. 3). Pierwsze cztery z pięciu wymienionych (tj. wszystkie metody tradycyjne, podstawowe, z wyłączeniem szczegółów współczesnych zastosowań analizy ścieżek rozwoju dyscyplin) wraz ze wskazaniem na możliwość ich interpretacji w kategoriach ustalania stopnia podobieństwa znaczeniowego publikacji, wyczerpują część pierwszą niniejszego

artykułu. W części drugiej⁴ skupiono się na dalszych modyfikacjach metod podstawowych, na pracach porównujących przyniesione przez nie wyniki, jak również na pracach, które za swój przedmiot obrały pomijane wcześniej odgałęzienia naukowej komunikacji, a przede wszystkim dokumenty patentowe jako szczególny typ publikacji naukowych. W części drugiej rozwinięto także zasygnalizowane w części pierwszej zagadnienie prognostycznego ukierunkowania analiz ścieżek rozwoju dyscyplin naukowych oraz poświęcono więcej uwagi metodom zaczerpniętym wprost ze statystyki matematycznej, a zwłaszcza z tzw. statystycznej analizy dynamiki zjawisk masowych. Na zakończenie części drugiej podsumowano kilka wyjątkowo oryginalnych podejść metodologicznych do problemu prognozowania i oceny rozwoju dyscyplin naukowych, które nie zmieściły się w żadnej z grup omówionych dotychczas. Po ich skrótowym scharakteryzowaniu zaprezentowano wnioski, które wypływają z całokształtu analizy piśmiennictwa oscylującego wokół rozpatrywanego zagadnienia bibliometrycznego prognozowania rozwoju nauki.

Należy także dodać, że w ramach poniższego przeglądu skupiono się przede wszystkim na krytycznej analizie określonych relacji pomiędzy dokumentami, jako nośnikami informacji naukowej. Charakterystyka sieci współpracy między autorami takich dokumentów (ang. *co-authorship network*) została natomiast pominięta, przede wszystkim z uwagi na ograniczoną ilość miejsca oraz – jak się wydaje – nieco mniejszą wagę takiej analizy w zakresie rozpoznawania nowo wykształcających się frontów badawczych. W ramach obydwu części poniższego przeglądu postarano się również zaznaczyć wady, które towarzyszą większości opisywanych podejść.

3. Publikacje pionierskie

W swojej prekursorskiej pracy z 1963 r. Maxwell Kessler zaproponował metodę tzw. generowania powiązań bibliograficznych (ang. *bibliographic coupling*), która początkowo miała służyć jako narzędzie usprawniające proces wyszukiwania informacji (Kessler, 1963, 11). Jej głównym założeniem była możliwość stwierdzenia, że istnieje powiązanie treściowe (znaczeniowe, poznawcze) pomiędzy tymi dokumentami, które mają przynajmniej jedną wspólną pozycję bibliograficzną (odwołują się do co najmniej jednej, tej samej publikacji). Siła tego związku może być zmierzona za pomocą liczby takich wspólnych jednostek bibliograficznych (ang. *coupling units*), figurujących jednocześnie w bibliografiach łącznikowych dwóch publikacji (Kessler, 1963, 10; zob. też: Diodato, 1994, 12–13). Z czasem zastosowanie tej metody wykroczyło poza jej pierwotne przeznaczenie i zaczęło obejmować sposób tworzenia zbiorów (tzw. klastrów⁵ – ang. *clusters*) prac traktujących o zbliżonych zagadnieniach, a więc

⁴ Część druga artykułu zostanie opublikowana w następnym numerze ZIN [red.].

⁵ Angielski termin *data clustering* w znaczeniu „grupowanie danych” (ewentualnie też „gromadzenie danych”) pojawił się w literaturze naukowej po raz pierwszy w 1954 r. w tytule publikacji Anila Jaina dotyczącej analizy danych antropologicznych (Wierzchoń & Kłopotek, 2015, 21). W języku polskim termin *clustering* tłumaczy się jako „grupowanie” i to wyrażenie będzie odąd używane, jako odpowiednik terminu angielskiego, w niniejszym artykule. Rzeczownik *cluster* może być tłumaczony jako „zbiór”, „klastr”, „gromada”, „zbitka” lub „grono”, a w niniejszym artykule autor posługuje się dwoma pierwszymi słowami spośród pięciu wymienionych. W literaturze polskojęzycznej można także spotkać się z tłumaczeniem angielskiego *cluster* jako „gniazdo tematyczne”, które trafnie oddaje jego sens (W. Pindłowa, 1994, 42).

np. zbiorów prac wytyczających granice pewnego obszaru problemowego, włączając w to także fronty badawcze w rozumieniu Price'a (zob. np.: Huang, Chang, 2014, 1723; Jarneving, 2005, 246–248). Obecnie wykrywanie powiązań bibliograficznych jest jedną z najważniejszych metod, na których bazuje odwzorowywanie („mapowanie”) wewnętrznej, kognitywnej struktury dyscyplin naukowych i zachodzących wewnątrz niej relacji.

Rok po ukazaniu się artykułu Kesslera Eugene Garfield, Irving Sher i Richard Torpie wysunęli propozycję stworzenia tzw. historiografii nauki (ang. *historiography of science*). Była ona metodą graficznego obrazowania historii rozwoju danej dziedziny lub dyscypliny naukowej w postaci grafów⁶, w których punkty (tzw. wierzchołki grafu – ang. *nodes*) symbolizują prace naukowe, a łączące je linie – wiążące je ze sobą cytowania. Uzyskuje się dzięki temu poglądową genealogię danej dziedziny lub dyscypliny. Demonstruje ona „rodowód” dziedzin/dyscyplin, kierunki ich rozwoju i kluczowe (najczęściej cytowane) publikacje, które doprowadziły do stanu, w jakim obecnie znajduje się dziedzina/dyscyplina. Metoda uwidacznia bibliograficznych „poprzedników” i „potomków” (ang. *bibliographic antecedents and descendents*) każdego rozpatrywanego punktu (publikacji) połączonego z pozostałymi elementami w obrębie grafu (Garfield et al., 1964, 1–6; Garfield et al., 2003, 400; Garfield, 2004, 119–121). Metoda ta nie została wprawdzie wypracowana wyłącznie z myślą o przewidywaniu dalszych trendów w rozwoju dyscyplin, ale sam pomysł budowania graficznych reprezentacji danych segmentów nauki w oparciu o cytowania literatury, a zwłaszcza w połączeniu z generowaniem klastrów (jako reprezentantów frontów badawczych) jest wykorzystywany do celów prognostycznych (zob. np.: Takeda & Kajikawa, 2009, 547). Rozrysowanie dyscypliny jako „chmury” punktów (symbolizujących publikacje) wraz z łączącą je siecią linii reprezentujących albo bezpośrednio cytowania (ang. *direct citation*), albo też i inne – najczęściej oparte na cytowaniach – miary siły (lub „bliskości”) związków pomiędzy punktami, pomaga bowiem dostrzec np. kognitywne centra (ang. *hubs*) dyscyplin, ich rozmiar (w sensie liczby dedykowanych im publikacji), związki z dyscyplinami pokrewnymi, a przede wszystkim zmiany zachodzące w czasie w strukturze dyscypliny lub ewolucję czy przesunięcia wiodącej tematyki frontów badawczych. Warto dodać, że w literaturze przedmiotu wyróżniono pięć rodzajów czasowej ewolucji frontów badawczych, które dają się zaobserwować za pośrednictwem podejścia wywodzącego się z metody badania zmian struktury grafów w czasie. Są to: fronty nowo powstałe (ang. *emerging fronts*), fronty rozrastające się (ang. *growing fronts*), fronty stabilne (ang. *stable fronts*), fronty zmniejszające swój rozmiar, „kurczące się” (ang. *shrinking fronts*) i fronty zanikające (ang. *exiting fronts*) (Huang & Chang, 2014, 1722; Upham & Small, 2010, 16–19). Każdy taki typ, do którego zakwalifikuje się jakiś front, pozwala więc już na stawianie pewnych hipotez co do jego przyszłej postaci. Aktualnie dostępne jest narzędzie pozwalające na komputerowe przetwarzanie bardzo dużych zbiorów danych w celu wykreowania historiograficznej mapy danej dziedziny/dyscypliny, które nazwano *HistCite* (zob. np.: Garfield et al., 2002).

W 1973 r. Henry Small, równocześnie z Iriną Marshakovą Shaikevich (1973), zainicjowali badania nad nową metodą badania stopnia podobieństwa treści dokumentów, która daje

⁶ Z formalnego punktu widzenia graf to zbiór wierzchołków (punktów), które są połączone „krawędziami” (ang. *edges*) w taki sposób, że każda krawędź kończy się i zaczyna w którymś z wierzchołków. Ścisłą definicję i omówienie podstawowych zagadnień teorii grafów można znaleźć np. w podręczniku Robina Wilsona (2012).

możliwość „mapowania” rozwoju nauki i intelektualnej architektury dyscyplin naukowych (Marshakova Shaikevich, 2005, 1535; Small, 1973, 266). Metodę tę nazwano „współcytowaniem” (ang. *co-citation*) i zdefiniowano ją jako częstość, z jaką pewne dwa dokumenty są cytowane razem (w parze) przez inne, wydane później od nich publikacje naukowe. Siła kognitywnego związku par dokumentów jest równa (niekiedy znormalizowanej z uwagi np. na różnice w średniej długości bibliografii załącznikowych w różnych dyscyplinach) liczbie takich publikacji naukowych, które cytują jednocześnie obie prace składające się na skojarzoną ze sobą (właśnie poprzez współcytowanie) parę. W zbiorze dokumentów utworzonych z użyciem tej metody częstość współcytowań może być naturalnie różna i można ustalić dla niej pewne minimum (Small, 1973, 265–267; zob. też: Diodato, 1994, 42–47). Metoda ta jest niejako odwrotnością metody generowania powiązań bibliograficznych Kesslera. Small zaznaczył także, że w przeciwieństwie do metody generowania powiązań bibliograficznych, metoda współcytowań jest dynamiczna. To znaczy, że wzorce współcytowań i liczba publikacji cytujących jakieś wydane wcześniej pary innych prac może ulegać zmianom z biegiem czasu, co ma czynić metodę zliczania współcytowań bardziej odpowiednią w zakresie odwzorowywania relacji łączących różne dokumenty w dłuższych okresach. Ponadto związki wykryte przez obie te metody są zazwyczaj różne – nie ma pomiędzy nimi symetrii (Small, 1973, 265–267).

W publikacji Smalla i Belvera Griffith’a z 1974 r. po raz pierwszy zastosowano w praktyce (jako eksperyment) metodę tworzenia klastrów publikacji powiązanych ze sobą poprzez współcytowania. Small i Griffith dążyli do odtworzenia struktury takich specjalności naukowych, które cechowały się wtedy wysoką „aktywnością” (innymi słowy, były specjalnościami rozwijającymi się w szybkim tempie) (Small & Griffith, 1974, 17). Punktem wyjścia metody było wyselekcjonowanie grupy publikacji cytowanych w pierwszym kwartale 1972 r. powyżej pewnego minimum (ang. *initial threshold*) równego 10 cytowaniom. Publikacje te musiały być zarazem połączone relacją współcytowania również mającą różne progi (1, 3, 6 i 10 współcytowań) (Small & Griffith, 1974, 20, 22). Generowanie klastra odbywało się poprzez wybór pewnego dokumentu i dołączenie do niego wszystkich innych dokumentów, z którymi był on związany współcytowaniem. Po dodaniu tych nowych dokumentów, utworzony zbiór uzupełniono o prace, z którymi wszystkie publikacje nowo dodane były połączone tą samą relacją. Proces ten był kontynuowany aż do chwili, w której nie pozostał już żaden dokument mogący dołączyć do utworzonego klastra w oparciu o tę samą zasadę⁷. Według Smalla i Griffith’a jest to najprostsza metoda (algorytm) tworzenia klastrów, którą określono mianem „grupowania jednopięciowego” (ang. *single-link clustering*) z uwagi na fakt, że do włączenia w klaster wystarczy tylko jedno połączenie relacją cytowania z którymkolwiek dokumentem będącym już wewnątrz klastra (Small & Griffith, 1974, 23; zob. też: Wieruchoń & Kłopotek, 2015, 35–36). W efekcie zastosowania tej procedury otrzymano szereg klastrów, z których jeden wyróżniał się wyraźnie pod względem swojego dużego rozmiaru i tematycznej heterogeniczności zawartych w nim publikacji (należących głównie do dyscyplin biomedycznych). Okazało się bowiem, że publikacje opisujące pewne szeroko stosowane metody prowadzenia badań były szczególnie wysoko

⁷ Warto w tym miejscu podkreślić, że w tak skonstruowanym klastrze nie mamy do czynienia ze zbiorem publikacji, w którym każda praca jest powiązana (poprzez współcytowanie) z każdą inną, ale ze zbiorem, w którym każda publikacja jest powiązana z przynajmniej jedną inną pracą.

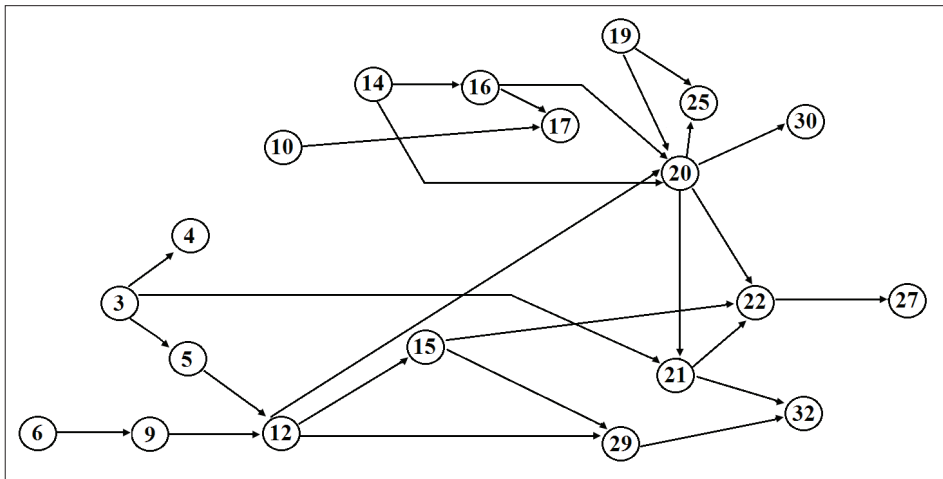
cytowane w obrębie różnych dyscyplin, co wywołało niekorzystną tendencję do łączenia ze sobą bardzo dużej liczby publikacji, które w istocie miały ze sobą niewiele wspólnego pod względem poruszanej w nich problematyki (Small & Griffith, 1974, 38; Small & Sweeney, 1985, 392). Drugim problemem był fakt, że publikacje wysoko współcytowane były zarazem także wysoko cytowane indywidualnie. Z tego powodu pojawiła się kolejna, niepożądana tendencja (ang. *bias*) do faworyzowania dyscyplin, w których wysokie liczby cytowań są normą (np. dyscypliny medyczne), w przeciwieństwie do dyscyplin mniejszych i bardziej wyspecjalizowanych (np. dyscyplin inżynierskich) (Small & Sweeney, 1985, 392). Sprawilo to, że ten szczególny klaster publikacji biomedycznych był wyjątkowo trudny do zinterpretowania z naukowca punktu widzenia. W związku z tym Small i Ernest Sweeney (1985) zdecydowali, że należy nieco zmodyfikować metodę tworzenia klastrów. Stwierdzili oni, że po pierwsze należy wprowadzić procedurę normalizacji liczby cytowań prac wysoko cytowanych podczas pierwszego etapu selekcji publikacji mogących potencjalnie stać się składnikami klastrów. Normalizacja przybrała postać tzw. liczenia ułamkowego (ang. *fractional citation counting*). Przedtem każde cytowanie było uważane za równe każdemu innemu. Obecnie dzięki zliczaniu ułamkowemu, cytowanie otrzymane od publikacji z bibliografią załącznikową zawierającą niewiele pozycji, posiadało większą wagę niż cytowanie otrzymane od publikacji z bardzo liczną bibliografią załącznikową, np. od artykułu przeglądowego (Small & Sweeney, 1985, 392–393). Druga zasadnicza modyfikacja pierwotnej metody to ustalenie górnego limitu wielkości klastra (tj. maksymalnej liczby dokumentów, które mogą wchodzić w jego skład), która zapobiegłaby formowaniu się amorficznych makro-klasterów (Small & Sweeney, 1985, 395–397). Pomimo wprowadzenia tych udoskonaleń, Loet Leydesdorff w 1987 r. podkreślił, że sama metoda grupowania jednopiętowego zawiera w sobie czynniki, które powodują, że w wyniku jej zastosowania zawsze otrzyma się przynajmniej jeden klaster heterogeniczny i trudny do sensownego zinterpretowania. Wynika to z faktu, że w ramach tej metody publikacje marginalne, słabo powiązane z publikacjami centralnymi, utworzą własny klaster leżący niejako na uboczu (Leydesdorff, 1987, 301). Ponadto Leydesdorff sformułował drugi (wynikający z pierwszego) zarzut wobec tej metody, według którego wygenerowane w jej ramach klastry zawsze będą wykazywać skłonność do pewnej niespójności (Leydesdorff, 1987, 302–304). W efekcie „trudno jest tak naprawdę określić, co uzyskujemy jako końcowy rezultat w wyniku generowania [tzw.] map nauki [przy użyciu tej metody]”⁸ (Leydesdorff, 1987, 303). Trzecim zarzutem Leydesdorffa było potraktowanie zbioru danych zawartych w *Science Citation Index* (SCI) w ramach metody grupowania jednopiętowego, jako zbioru („worka” – ang. *garbage can*) indywidualnych i w zasadzie przypadkowych połączeń dokumentów. Zamiast tego, w opinii Leydesdorffa, konceptualizacja danych pochodzących z SCI powinna polegać na przyjęciu, że zbiór ten jest wysoce ustrukturyzowaną siecią zhierarchizowanych relacji między pracami cytowanymi a cytującymi (Leydesdorff, 1987, 320).

W 1989 r. Norman Hummon i Patrick Doreian wysunęli własną propozycję identyfikacji głównego nurtu przepływu informacji (ang. *main path*) w ukierunkowanej (ang. *directed*) sieci cytowań obejmującej pewną dyscyplinę lub subdyscyplinę. Sieć ukierunkowana cechuje się tym, że występuje w niej rozróżnienie na publikacje cytujące i cytowane. Przepływ

⁸ Oryg.: „It becomes difficult to say what one is actually producing in the end in terms of ‘maps of science’” (Leydesdorff, 1987, 303).

informacji ma tu więc wymiar czasowy (z przeszłości w przyszłość). Identyfikacja głównego nurtu przepływu informacji zasadza się na identyfikacji najważniejszych par dokumentów cytowanych i cytujących (ang. *links*). Najważniejszymi są połączenia par wierzchołków o najwyższej tzw. wartości przejściowej (ang. *traversal count*). Innymi słowy są to te połączenia dwóch wierzchołków sieci, które występują najczęściej podczas obliczania największych „długości” (sekwencji kolejnych punktów w sieci) wszystkich możliwych „ścieżek” (ang. *search paths*) łączących te dwa punkty (Hummon & Doreian, 1989, 48–50). Wartość przejściowa określa, ile razy każda z par połączonych punktów pojawiła się na ścieżce łączącej wszystkie kolejne pary w całej sieci cytowań. Rozważmy następujący przykład. Istnieje pewna sieć cytowań o postaci przedstawionej na rysunku 1.

Rys. 1. Przykładowy fragment grafu ukierunkowanego reprezentującego siatkę cytowań publikacji. Źródło: (Hummon & Doreian, 1989, 41)



Strzałki reprezentują istniejące cytowania (przepływ wiedzy od dokumentu cytowanego do cytującego). Np. publikacja nr 3 jest cytowana przez publikacje nr 4, 5 i 21. Z całego grafu można wyodrębnić wszystkie możliwe „ścieżki” lub „podgrafy” (ang. *subgraphs*) leżące na drodze łączącej dowolne pary wierzchołków. Niech będą to np. wierzchołki nr 3 i 22. W tym przypadku istnieją 42 możliwe ścieżki⁹. Są nimi: 3–5; 3–12; 3–15; 3–20; 3–21; 3–22; 5–12; 5–15; 5–20; 5–21; 5–22; 12–15; 12–20; 12–21; 12–22; 15–20; 15–21; 15–22; 20–21; 20–22; 21–22; 5–3; 12–3; 15–3; 20–3; 21–3; 22–3; 12–5; 15–5; 20–5; 21–5; 22–5; 15–12; 20–12; 21–12; 22–12; 20–15; 21–15; 22–15; 21–20; 22–20; 22–21.

⁹ W grafie przedstawionym na rysunku 1 istnieje 7 wierzchołków, które posiadają połączenia prowadzące od wierzchołka nr 3 do wierzchołka nr 22. Są to wierzchołki nr: 3, 5, 12, 15, 20, 21 i 22. Tylko „przechodząc” przez nie można dotrzeć od wierzchołka nr 3 do wierzchołka nr 22. Np. wierzchołek nr 16 jest usytuowany całkowicie poza tymi możliwymi ścieżkami. Jeżeli natomiast wyróżniony w ten sposób „podgraf” ma dokładnie 7 wierzchołków, teoretycznie istnieją $7(7-1) = 7 \cdot 6 = 42$ możliwe połączenia między nimi – tj. każdy wierzchołek z każdym innym wierzchołkiem, oprócz samego siebie (zob. Hummon & Doreian, 1989, 49). Dodatkowo znaczenie ma tu kierunek połączenia (przepływu informacji). Połączenie „3–5” nie jest więc tym samym, co połączenie „5–3”.

Tylko 19 wyróżnionych połączeń, spośród wszystkich 42 możliwych, jest w rzeczywistości zrealizowanych w przykładowym grafie na rysunku 1. Np. połączenie „3–21” (leżące na drodze pomiędzy wierzchołkami nr 3 i nr 22) ma postać następującej ścieżki: 3–5–12–20–21. Kiedy rozpatrzylibyśmy wszystkie możliwe ścieżki dla powyższego przykładu dziewiętnastu par okazałoby się, że połączenie „5–12” występuje najczęściej, tj. pojawia się w największej liczbie połączeń, spośród wszystkich dziewiętnastu faktycznie zrealizowanych w tym konkretnym grafie. Istnieje ono np. w ścieżce łączącej parę „3–15”: 3–5–12–15, w ścieżce łączącej parę „3–22”: 3–5–12–15–22, w ścieżce łączącej wierzchołki „5–21”: 5–12–20–21, a ponadto jeszcze w siedmiu innych ścieżkach. Ogółem występuje więc ono w 10 ścieżkach, podczas gdy np. para „3–21” tylko w dwóch („3–21” i „3–22”). W związku z tym połączenie „5–12” ma najwyższą wartość przejściową (*traversal count*), równą 10 czyli jest połączeniem o najwyższej „sile” w przykładowym podgrafie „3–22” (zob. Hummon & Doreian, 1989, 50). Wartość tę można w analogiczny sposób ustalić dla każdej pary połączonych wierzchołków (dla wszystkich możliwych podgrafów danego grafu), a „główna ścieżka” całego grafu to taka, która odzwierciedla sieć najsilniejszych połączeń wewnątrz niego. Będąc przy każdym wierzchołku mamy bowiem do wyboru jeden lub więcej kolejnych wierzchołków różniących się wartością przejściową o znanej nam wartości. Tym sposobem identyfikuje się publikacje najistotniejsze (o największej sile wpływu) w całej sieci lub, inaczej mówiąc, najważniejszy strumień (ang. *main stream*) przepływu informacji. Należy dodatkowo podkreślić, że publikacje składające się na „główną ścieżkę” niekoniecznie muszą być tym samym publikacjami najwyżej cytowanymi w obrębie sieci (Liu & Lu, 2012, 529). Metoda ta sama w sobie nie została zaprojektowana z myślą o prognozowaniu rozwoju dyscyplin, pomimo tego wydaje się, że posiada ona znaczny potencjał w tym zakresie (zob. np.: Ma & Liu, 2016; Tu & Hsu, 2016).

4. Miary podobieństwa publikacji

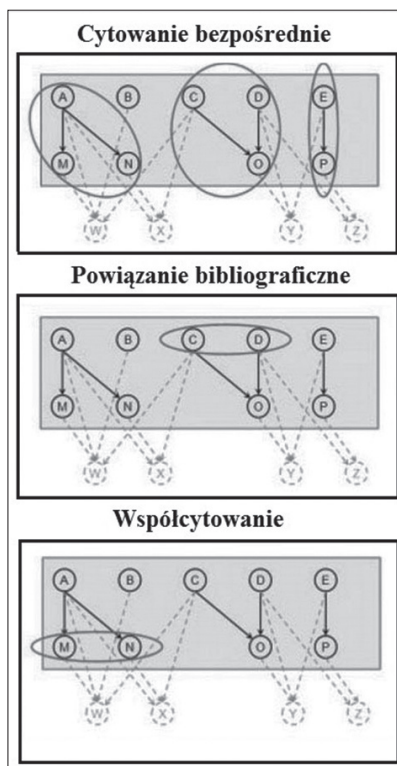
Jak wynika z powyższego przeglądu publikacji pionierskich, większość prób identyfikacji kierunków dalszego rozwoju dyscyplin naukowych opierała się na wyodrębnianiu podzbiorów (klastrow) publikacji naukowych należących do określonej dyscypliny, które stanowią reprezentacje jej subdyscyplin – węższych niż cała dyscyplina specjalności czy obszarów problemowych, które mogą ewoluować w swoim własnym tempie i kierunku. Metody oceny tempa tej ewolucji wymagają posłużenia się specjalnie zaprojektowanymi technikami analizy wyróżnionych już wcześniej klastrow. Ocenia się np. ich rozmiar, średni wiek publikacji wchodzących w ich skład, wiodącą tematykę, a przede wszystkim zmiany zachodzące we wspomnianych cechach (w szczególności w średnim wieku i rozmiarze) w miarę przesuwania się na osi czasu w kierunku przyszłości. Operacja wydzielenia klastrow w ramach dyscypliny jest więc podstawą dalszej analizy ukierunkowanej prognostycznie. Z samej definicji klastra wynika z kolei wprost, że stanowi on zgrupowanie elementów homogenicznych, chociaż niekoniecznie identycznych. Dwa dowolne elementy należące do jednego klastra powinny być jednak podobne do siebie w znacznie większym stopniu, niż do dowolnego elementu należącego do innego klastra (Wierzchoń & Kłopotek, 2015, 17). Powstaje w związku z tym zasadnicze pytanie dotyczące tego, w jaki sposób należy zdefiniować to podobieństwo oraz w jaki sposób posłużyć się zaakceptowaną już miarą

podobieństwa w trakcie samej operacji grupowania. Ponieważ odpowiedzi na te dwa pytania są od siebie niezależne, w efekcie badacze mają do czynienia z ogromną wielością różnorodnych algorytmów służących do wyróżniania klastrów (Wierzchoń & Kłopotek, 2015, 17), co znajduje swoje odbicie także w bibliometrycznych zastosowaniach tej operacji. Z ogółu stosowanych w tym zakresie miar podobieństwa trzy zostały już wspomniane w poprzednim podrozdziale. Były nimi siła powiązania bibliograficznego, siła relacji współcytowania i „siła” (w sensie liczebności) cytowań bezpośrednich. Czwartą miarą podobieństwa jest z kolei omówiona w kolejnym podrozdziale siła (intensywność) relacji współwystępowania specjalistycznych terminów/słów kluczowych (ang. *co-word analysis*). Wszystkie tego rodzaju miary pozwalają na kwalifikowanie publikacji do klastrów jako tworów względnie homogenicznych, stanowiących reprezentacje ewoluujących w czasie subdyscyplin lub specjalności dyscyplinarnych (zob. np.: Soós, 2014, 23). Trzy pierwsze miary w przystępny sposób omówili i przedstawili graficznie Kevin Boyack i Richard Klavans w 2010 r. (zob. Rysunek 2) (Boyack & Klavans, 2010, 2390). Cytowanie bezpośrednie pozwala stworzyć klaster obejmujący prace cytowane wraz z cytującymi. Na rysunku 2 są to np. publikacje „A”, „M” i „N” tworzące trójelementowy klaster {A, M, N}. Powiązane bibliograficznie występuje pomiędzy publikacjami „C” i „D” ponieważ obydwie cytują publikację „O”. W ten sposób powstał dwuelementowy klaster {C, D}. Publikacje „M” i „N” są z kolei połączone relacją współcytowania z uwagi na fakt, że obydwie zostały zacytowane przez dokument „A”. W tym przypadku powstał więc klaster {M, N} (Rys. 2).

W związku z niewielką liczbą prac omawiających relację podobieństwa w kontekście mapowania nauki wydaje się, że w ramach drugiej podgrupy wystarczy wspomnieć o przeglądzie tego rodzaju miar dokonany przez Annę Huang (2008) i zwrócić szczególną uwagę na tzw. podobieństwo kosinusowe (ang. *cosine similarity*) publikacji naukowych. Jest ono bowiem techniką często stosowaną w perspektywie podziału zbioru publikacji na klastry. W dużym skrócie polega ono na zamianie całości treści dokumentu, znormalizowanej w odpowiedni sposób (np. ujednotwiona się wyrażenia synonimiczne, eliminuje tzw. „słowa stopujące” – ang. *stop words* – które są wyrażeniami w rodzaju „i”, „lub”, „także” itd. oraz usuwa prefiksy i sufiksy pozostawiając jedynie semantyczne rdzenie słów) na wektory n -wymiarowe. Wektor ma postać ciągu kolejnych n wyrażen, np.: {*chemic, kinetic, fabric, polymer...*}. Jeżeli po normalizacji w dokumencie pozostaje np. 1000 unikalnych słów, utworzony zeń wektor ma 1000 wymiarów – każdy wymiar odpowiada każdemu, unikalnemu słowu/terminowi. W ten sposób można zbudować tzw. macierze (ang. *matrices*) pokazujące, które znormalizowane terminy i w jakiej liczbie występują w określonych, porównywanych dokumentach. Jeżeli dwa dokumenty wykorzystują wiele tych samych terminów to stopień ich podobieństwa będzie wysoki. Dzięki temu możliwe jest precyzyjne, ilościowe określenie stopnia semantycznej równoważności dwóch lub więcej publikacji naukowych. Podobieństwo kosinusowe przyjmuje wartości z przedziału od 0 do 1. Wartość „1” oznacza, że dokumenty są identyczne. Więcej informacji na ten temat można znaleźć np. w publikacjach Anny Huang (2008, 51–52), Mirosława Wierzchońa i Sławomira Kłopotka (2015, 31–32), Piotra Potiopy (2011, 416–417), Gerarda Saltona i Donny Bergmark (1979, 146–147), a także w poradniku Erica Garcii (2015).

Rys. 2. Graficzna reprezentacja powiązania hipotetycznych publikacji za pomocą trzech relacji wykorzystywanych w obrębie nurtu bibliometrycznego mapowania dyscyplin naukowych.

Źródło: opracowanie własne na podstawie (Boyack & Klavans, 2010, 2390)



5. Metoda współcytowań

Jak już zasygnalizowano w części czwartej, metoda współcytowań to pierwsza z czterech podstawowych metod wyróżniania względnie homogenicznych klastrów publikacji powiązanych ze sobą siecią cytowań, która daje w efekcie swego rodzaju ekwiwalenty frontów badawczych Price'a. Jej odmianę polegającą na analizie współcytowań autorów publikacji naukowych (zamiast samych publikacji) wykorzystał w praktyce Olle Persson (1994). Zarazem powtórzył on zarzut Leydesdorffa wobec metody grupowania jedno-połączeniowego (wytwarzanie zbyt dużych i heterogenicznych klastrów) oraz zauważył, że sama liczba współcytowań autorów jest zbyt prostą miarą podobieństwa i że należy stosować normalizację z uwagi na ogólną cytawalność poszczególnych autorów (Persson, 1994, 34). Persson zademonstrował następnie w jaki sposób normalizację tego rodzaju można wprowadzić do analizy współcytowań autorów artykułów ogłoszonych w czasopiśmie *Journal of the American Society for Information Science* w latach 1986–1990 oraz zaproponował ograniczenie, według którego algorytm nie pozwoli na uformowanie się klastra, który

zawierałby mniej niż trzech autorów. Algorytm eliminował ponadto z dalszego procesu przyłączania do innych klastrow takich par autorów, którzy zostali już wcześniej włączeni do odrębnego klastra. Modyfikacja ta miała zapobiegać powstawaniu klastrow heterogenicznych (Persson, 1994, 32, 34).

W tym samym roku Michael Zitt i Elise Bassecouard (1994) uzupełnili analizę klastrow publikacji współcytowanych jako reprezentacji obszarów problemowych o badanie średniego wieku klastrow celem identyfikacji ich trendów rozwojowych (Zitt & Bassecouard, 1994, 335). Zaznaczyli przy tym, że wadą tego podejścia jest konieczność zaakceptowania faktu, że daty ostatecznej publikacji tekstów są późniejsze niż daty nadesłania ich do redakcji czasopism, które z kolei są późniejsze niż daty prowadzenia badań i formułowania końcowych wniosków lub hipotez. Opóźnienie to (ang. *time lag*) wpływa negatywnie na możliwość oceny dynamiki zmian klastrow, która jest kluczowym elementem podczas identyfikowania bieżących trendów badawczych (Zitt & Bassecouard, 1994, 336). W publikacji Zitta i Bassecouard zwrócono ponadto uwagę na inną wadę metody analizy współcytowań polegającą na zdolności tej metody do jedynie wycinkowej klasyfikacji całości literatury cytowanej przez społeczność naukową. Inaczej mówiąc, te publikacje, które nie są współcytowane (lub są współcytowane w niewystarczającej liczbie) z innymi pracami, pozostają poza nawiasem stosowalności metody. Problematyczna w opinii obojga autorów była również arbitralność ustalania maksymalnej wielkości klastrow (jak uczynili to np. Small i Sweeney w 1985 r.). Według Zitta i Bassecouard pytanie o to, jak duży musi być klastrow, aby mógł on zostać uznany za samodzielną jednostkę, pozostawało otwarte (Zitt & Bassecouard, 1994, 337, 339). Oryginalna metoda szacowania trendu została ostatecznie oparta przez parę autorów o stosunek wieku publikacji tworzących (wchodzących w skład) klastrow do wieku publikacji cytujących dokumenty wchodzące w skład klastra. Trend badawczy dyscypliny konstytuowały w ich opinii tematy progresywne i zakodowane tekstualnie, tj. określone w pewien sposób pod względem tematyki odrębną metodą (Zitt & Bassecouard, 1994, 340). „Progresywne” oznacza tu bycie cytowanym przez publikacje wydane stosunkowo niedawno. Zarazem tematy te muszą być oparte na materiale „młodym” (tj. na klastrow publikacji o niskiej średniej arytmetycznej ich wieku) (Zitt & Bassecouard, 1994, 343–344, 348). Dopiero te dwa warunki spełnione łącznie pozwalają określić, które trendy czy obszary problemowe można uznać za „wschodzące”.

Henry Small (2006) również postawił pytanie o to, czy możliwe jest przewidzenie wzrostu pewnych obszarów badawczych i kierunku postępu w nauce, albo wyłonienia się całkowicie nowych dyscyplin spośród tych już istniejących (Small, 2006, 596). Jego metoda rozwiązania tego problemu opierała się na analizie klastrow publikacji współcytowanych i zarazem takich, które stanowiły górny 1% publikacji cytowanych indywidualnie najwyższej (ang. *top 1%*) w zachodzących na siebie przedziałach czasowych (np. przedziały postaci 2000–2003, 2002–2004 itd.). Small położył przy tym szczególny nacisk na „kontynuowalność” (ang. *continuity*), inaczej mówiąc ciągłość występowania w klastrow tych samych, najwyższej cytowanych publikacji w kolejnych przedziałach czasu. Występowanie takiej ciągłości nazwał on „łańcuchami klastrow” (ang. *cluster strings*) (Small, 2006, 597; Upham & Small, 2010, 19). Predykcji obszarów wzrostu służyła w tym ujęciu tzw. „aktualność klastra” (ang. *currency of cluster*) będąca średnim wiekiem publikacji najwyższej cytowanych wchodzących w jego skład. Publikacje te stanowiły centrum klastra, jego kognitywny „rdzeń”. Obserwując jak zachowywały się „łańcuchy” podczas przejść w coraz bliższe współczesności

przedziały czasu i mierząc ich aktualność, można było wychwycić transformacje jakim ulegał „łańcuch” (rozrost, przyłączanie publikacji z pokrewnych dyscyplin, rozgałęzianie się, zanikanie itd.). Aktualność mieściła się w przedziale od 0 do 1. Aktualność równa „1” oznaczała, że wszystkie publikacje najwyższej cytowane (cały *top 1%*) zostały wydane w ostatnim roku przedziału czasowego, który poddaje się analizie. Aktualność równa zeru oznaczała natomiast, że wszystkie te prace pochodziły z pierwszego roku wyznaczającego dolną granicę przedziału (Small, 2006, 607–608). Następnie zbadano związek rozmiaru klastrów z ich aktualnością. Rozmiar klastra był tu wskaźnikiem tego, jak ożywiona jest aktywność publikacyjna w danym obszarze tematycznym, jak wielu badaczy przyciąga jego tematyka i jaki jest jego ogólny potencjał badawczy. Small zaobserwował w tym względzie pewną korelację. Okazało się np., że z wysoką aktualnością z przedziału 0,8–1 związany był rozrost klastra o około 35%, a z aktualnością równą lub niższą niż 0,2 – spadek liczby artykułów bardzo wysoko cytowanych w klastrze (Small, 2006, 608). Small skonkludował jednak, że korelacja ta była zbyt słaba aby wyjaśnić rozrastanie się klastrów za pomocą samej jedynie aktualności. Poza aktualnością musiały więc istnieć inne, niezidentyfikowane czynniki, które wpływały na powiększanie się obszarów badawczych. Według Smalla należało wobec tego poszukiwać dalszych determinantów krótko – i długoterminowej dynamiki ewoluowania dyscyplin (Small, 2006, 609–610).

Ostatnią publikacją, o której należy powiedzieć w kontekście metody badania współcytowań jest artykuł Samuela Phineasa Uphama i Henry’ego Smalla (2010), w którym wspomniano o m. in. dwóch poważnych ograniczeniach metody w takiej postaci, jaką zaprezentował Small w 2006 r. Pierwszym była jej niezdolność do zidentyfikowania nowego frontu badawczego, jeżeli żaden z należących doń artykułów nie stanie się wysoko cytowany. Ograniczenie to dotyczy przede wszystkim takich dyscyplin nauki, w których normą są raczej niskie wskaźniki cytowań (np. matematyka, nauki społeczne i humanistyczne). Stąd metoda ta nigdy nie pozwoli na identyfikację nowo wyłaniającego się obszaru problemowego czy subdyscypliny natychmiast po jej uformowaniu się. Drugim ograniczeniem metody była niemożność zidentyfikowania wszystkich prac należących do nowo powstałego obszaru problemowego. Zamiast tego, przy jej użyciu możliwa będzie tylko detekcja samego istnienia takiego obszaru i dostarczenie pewnej próbki należących do niego prac wysoko cytowanych, a nie jego precyzyjne wyodrębnienie. Metoda ta nadała się więc najlepiej do „szybkiego skanowania” rozleglejszych obszarów nauki, a nie do dokładnego rozpoznawania i wyszczególnienia nowych prądów intelektualnych czy nowych tendencji dyscyplinarnych (Upham & Small, 2010, 18; zob. też: Van Den Besselaar & Heimeriks, 2006, 380). Oryginalność omówionego przez Smalla i Uphama podejścia prognostycznego sprowadzała się po pierwsze, do wyróżnienia nowych kategorii klastrów. Były nimi klastry publikacji wysoko cytowanych o zerowej „kontynuowalności”, tj. takie, które zaobserwowano po raz pierwszy dopiero w pewnym (którymś z kolei) przedziale czasowym ze wszystkich przedziałów objętych analizą (Upham & Small, 2010, 19). Ponadto do nowych kategorii klastrów należały klastry „rosnące”, „stabilne”, „kurczące się” i „zanikające” pod względem liczby składających się na nie dokumentów w kolejnych przedziałach czasu (Upham & Small, 2010, 21–22). Po drugie, wprowadzono nowy wskaźnik – tzw. endogeniczność klastra. Określała ona procentowo, jak wiele z dokumentów cytowanych (tworzących „rdzeń” klastra) było zarazem dokumentami cytującymi ten „rdzeń”. Wskaźnik ten informuje więc o tym, do jakiego stopnia klastr „bazuje sam na

sobie”, jak bardzo jest samowystarczalny. Front badawczy o wysokiej endogeniczności jest tworem rozwijającym się w oparciu o swoje własne podstawy, jest mało interdyscyplinarny i będzie prawdopodobnie ewoluował w kierunku samodzielnej subdyscypliny (Upham & Small, 2010, 20). Ogółem, prognoza wyłaniania się nowych subdyscyplin polegała więc ostatecznie na rozważeniu zależności między typem klastra („rosnący”, „stabilny” itd.), zmianami cytawalności publikacji tworzących klastry w kolejnych przedziałach czasu (tzw. „wchłanianie” frontu/klastra przez sąsiednie subdyscypliny), „natychmiastowością” klastra (tj. odsetkiem publikacji tak cytujących, jak i cytowanych w klastrze, które opublikowano w „górnym” brzegowym roku analizowanego przedziału czasowego) oraz stopniem jego endogeniczności (Upham & Small, 2010, 23–26). Podsumowując, wydaje się, że metoda Uphama i Smalla była udoskonaloną i bardziej wyrafinowaną formą metody zaproponowanej przez samego Smalla w 2006 r.

W związku z metodą badania współcytowań należy ponadto wyraźnie podkreślić, że jednym z jej ograniczeń, o którym nieczęsto wspomina się w literaturze zagranicznej, jest konieczność dysponowania indeksem cytowań (takim jak np. *Web of Science* lub *Scopus*). Indeks cytowań, który obejmowałby całość polskojęzycznej literatury naukowej ze wszystkich dziedzin nauki nie jest natomiast na chwilę obecną dostępny. Istniejące polskie indeksy cytowań są bowiem albo dopiero w fazie „embrionalnej” („POL-index”), albo dotyczą tylko wybranych dyscyplin (np. „Arton”, „CYTBIN”), albo prace nad rozwijaniem indeksu nie są już kontynuowane (np. „Indeks Cytowań Socjologii Polskiej”, „Indeks Cytowań Historiografii Mediów Polskich”) (Seweryn & Swoboda, 2013). Jedyną realną ewentualnością wydaje się tu „Google Scholar Citations”, chociaż i w tym przypadku istnieje szereg mankamentów i możliwych nieprawidłowości, do których prowadzi poleganie na tym źródle danych. Przegląd takich mankamentów przedstawili np. Alberto Martín-Martín i in. (2016, 31–47).

6. Metoda generowania powiązań bibliograficznych

Generowanie powiązań bibliograficznych posiada przynajmniej dwie zalety w porównaniu z metodą współcytowań. Po pierwsze, do realizacji badania z użyciem tej metody nie jest potrzebny indeks cytowań, a jedynie dostęp do bibliografii załącznikowych wybranego zbioru publikacji. Po drugie, w metodzie tej nie występuje problem dotyczący metody współcytowań, polegający na fakcie konieczności opóźnienia badania aż do czasu zakułowania się wystarczającej liczby cytowań (a zatem i współcytowań) wokół publikacji reprezentatywnych dla dyscypliny lub funkcjonujących w jej obrębie specjalności (zob. np.: Boyack & Klavans, 2010, 2391; Glänzel, 2012, 195; Glänzel & Thijs, 2012, 400–401; Shibata et al., 2008, 762; Shibata et al., 2009, 577). Ocenia się, że do czasu wystąpienia cytawalności maksymalnej musi minąć przynajmniej 2–3 lata, a niekiedy nawet więcej, w zależności od konkretnej, wybranej jako przedmiot badań dyscypliny (zob. np.: Adler et al., 2009, 5; Briotta Parolo et al., 2015, 736–738; Dorta-González & Dorta-González, 2013, 594–600; Glänzel & Schoepflin, 1995, 39; Glänzel & Schoepflin, 1999, 43; Moed et al., 1998, 394, 396, 399).

Kevin Boyack, Katy Börner i Richard Klavans (2009) wykorzystali metodę powiązań bibliograficznych w celu wizualizacji dynamiki rozwoju dyscyplin chemicznych pomiędzy rokiem 1974 a 2004. Wykorzystaną tu miarę podobieństwa między czasopismami (siłą powiązania bibliograficznego) znormalizowano przy pomocy podobieństwa kosinusowego

(zob. podrozdział 4; Boyack et al., 2009, 47, 51). Głównym przedmiotem analizy była mapa nauki za 2002 r. złożona z 671 wyodrębnionych klastrów oraz połączeń pomiędzy nimi. Mapę uzupełniono dodatkowo o dane dotyczące relacji wzajemnych cytowań czasopism. Relacja występowała między czasopismami A i B, jeżeli czasopismo A cytowało czasopismo B, a czasopismo B cytowało czasopismo A (Boyack et al., 2009, 48, 51). Dzięki temu wszystkie czasopisma nieobecne na mapie za rok 2002 mogły zostać do niej dołączone w oparciu o siłę związku z poszczególnymi klastrami z 2002 r. (ang. *journal-to-cluster-similarity*). To znaczy, że każde czasopismo, które pojawiło się w danych o cytowaniach *Journal Citation Reports* (JCR) w którymkolwiek roku z przedziału 1974–2004 zostało przypisane do jednego z 671 klastrów wykrytych w 2002 r. (Boyack et al., 2009, 51). Dzięki użyciu kategorii tematycznych JCR uzyskano następnie czasowo-tematyczny przekrój wszystkich 671 klastrów dla każdego kolejnego, pięcioletniego przedziału czasowego (Boyack et al., 2009, 52, 54). Pozwoliło to na zaobserwowanie, jak w następujących po sobie przedziałach czasowych zmieniały się ilościowo rozmiary poszczególnych, należących do dyscyplin chemicznych, specjalności oraz ich stopnie wzajemnego przenikania się. Przykładowo przed 30 laty bioinżynieria niemalże nie istniała samodzielnie poza obszarem biochemii, podczas gdy w 2004 r. odgrywała już znaczącą rolę w mikrobiologii. Odnotowano też istnienie przepływu wiedzy z bioinżynierii do chemii polimerów, chemii analitycznej i chemii ogólnej (Boyack et al., 2009, 55–59). Jako słabą stronę swojego podejścia metodologicznego Boyack i in. wskazali natomiast fakt, że przyjęcie czasopism jako jednostki analizy jest w stanie wskazać jedynie duże, silne trendy w badanych obszarach. Obniżenie poziomu agregacji danych do poziomu indywidualnych publikacji pozwoliłoby natomiast na identyfikację także tych mniej widocznych i węższych trendów, a w szczególności takich trendów badawczych, które usytuowane są „na przecięciu” (ang. *at intersection*) szerszych specjalności (Boyack et al., 2009, 59–60).

Z interesującym pomysłem związanym z udoskonaleniem metody powiązań bibliograficznych wystąpił Sandor Soós (2014). Zaproponował on mianowicie uwzględnienie dodatkowego parametru przypisów współdzielonych przez pewne publikacje, jakim był wiek tych przypisów. Im starsze były współdzielone przed dwa współczesne artykuły pozycje bibliograficzne, tym słabszy powinny mieć one wpływ na siłę powiązania bibliograficznego obu artykułów współczesnych. Im z kolei wcześniejsze daty wydania publikacji figurujących w bibliografiach załącznikowych dwu artykułów, tym wyższy stopień intelektualnego czy poznawczego pokrewieństwa między artykułami, w których te pozycje bibliograficzne współwystępują. Modyfikację tę Soós nazwał „powiązaniem bibliograficznym wrażliwym na wiek pozycji bibliograficznych” (ang. *age-sensitive bibliographic coupling*) (Soós, 2014, 24). Celem modyfikacji było doprecyzowanie metody tworzenia takich klastrów publikacji związanych relacją bibliograficznego podobieństwa, które byłyby w stanie lepiej zrekonstruować historyczny rozwój danej struktury tematycznej i uwidocznic bardziej klarownie dynamikę danego obszaru badań (Soós, 2014, 24, 33). Różnica w dacie wydania pozycji bardzo starych (wydanych np. 800 i 1000 lat temu) w mniejszym stopniu determinowała siłę powiązania bibliograficznego, niż różnica w wieku publikacji wydanych np. 8 i 10 lat temu (Soós, 2014, 27). Waga takich dokumentów współcytowanych nie maleje w związku z tym liniowo (wraz ze wzrostem dystansu od czasów terażniejszych), ale wykładniczo. Konsekwencją tego faktu jest to, że przywoływane w dwóch różnych artykułach pozycje antyczne lub średniowieczne mają niemal tę samą wagę, która utrzymuje się na równie

niskim poziomie. Jeżeli np. publikacja P1 ma tę samą liczbę współdzielonych przypisów co publikacja P2 i publikacja P3, ale z P2 dzieli przypisy młode, a z P3 stare, według standardowej metody powiązań bibliograficznych P1 jest tak samo podobne do P2 jak do P3, ponieważ istotna jest tylko sama liczba wspólnych przypisów. W metodzie wrażliwej na wiek P1 będzie natomiast dużo bardziej podobne do P2 niż do P3 (Soós, 2014, 29). Zdaniem Soósa jego metoda pozwala na generowanie klastrów odzwierciedlających podziały i różnicowanie się kierunków badawczych, które wyłaniają się z czasem w obrębie danych dyscyplin nauki. W ramach eksperymentalnego porównania wyników stosowania metody „klasycznej” i metody wrażliwej na wiek, w przypadku tego samego zbioru publikacji, nowa metoda pozwoliła np. rozszcześcić duży, mało spójny i reprezentujący szeroką dyskusję teoretyczną klastery na klastry mniejsze, które reprezentowały odpowiednio: początkowy kontekst dyskursu, jego późniejszy rozwój i wydzielenie się spójnych tematycznie szkół, będących odpowiedziami na problemy podniesione w ramach kontekstu początkowego (Soós, 2014, 49–50). Potencjał metody w zakresie prognozowania leży zatem głównie w możliwości precyzyjniejszego wyodrębniania wąskich trendów badawczych (ang. *strains of research*) w korpusach publikacji współczesnych oraz w dokładniejszym rozróżnieniu tych kierunków rozwoju, które są odbiciem dyskusji nad zagadnieniami klasycznymi (ang. *classical*) i zagadnieniami bieżącymi (ang. *recent*), co jest często niemożliwe przy zastosowaniu tradycyjnej metody generowania powiązań bibliograficznych (Soós, 2014, 48–50).

Mu-Hsuan Huang i Chia-Pin Chang (2014) również wykorzystali metodę generowania powiązań bibliograficznych pomiędzy artykułami cytowanymi wysoko, jako tymi, które odznaczają się wysoką jakością intelektualną, znaczną siłą wpływu czy rangą naukową (Huang & Chang, 2014, 1722). Huang i Chang zastosowali, wzorem Smalla (2006) oraz Smalla i Uphama (2010) pięć zachodzących na siebie okresów obserwacji (ang. *sliding citation window*), aby prześledzić ewolucję pewnej szczególnej specjalizacji z zakresu elektroniki. Okresy obserwacji to lata: 2000–2005, 2001–2006, 2002–2007, 2003–2008 i 2004–2009 (Huang & Chang, 2014, 1723). Okresy te, jako zazębiające się ze sobą, były zdaniem Huang i Chang w stanie zrekonstruować dynamikę specjalności naukowej, w przeciwieństwie do jakiegokolwiek odgórnie ustalonego (statycznego) okresu obserwacji. Po określeniu wymaganego stopnia cytowalności selekcjonowanych publikacji oraz prognozy siły powiązania bibliograficznego (ang. *coupling strenght*) równego co najmniej pięciu wspólnym pozycjom bibliograficznym, posłużono się techniką tzw. grupowania hierarchicznego¹⁰ z wykorzystaniem wskaźnika korelacji Pearsona jako miary podobieństwa między grupowanymi elementami (Huang & Chang, 2014, 1724, 1726; zob. też: Wierzchoń & Kłopotek, 2015, 35–38; nt. wskaźnika Pearsona jako miary podobieństwa między publikacjami zob. np.: Huang, 2008, 52). Warto podkreślić, że zarówno wybór metody generowania klastrów jak i wybór wskaźnika podobieństwa między grupowanymi dokumentami ma duży wpływ na uzyskane wyniki (tj. liczbę i rozmiar uzyskanych klastrów) (Huang & Chang, 2014, 1726).

¹⁰ W tzw. grupowaniu hierarchicznym (ang. *hierarchical clustering*) punktem wyjścia jest potraktowanie całego m-elementowego zbioru wszystkich obiektów jako grupy m jednoelementowych klastrów. Następnie oblicza się tzw. odległości (podobieństwo) każdej pary takich jednoelementowych klastrów i wybiera parę elementów o najwyższym stopniu podobieństwa. Para ta tworzy pierwszy, 2-elementowy klastery. Dalej aktualizuje się obliczenia odległości między nowym klastrem a wszystkimi pozostałymi 1-elementowymi klastrami i powtarza całą procedurę od początku. Nt. dalszych szczegółów grupowania hierarchicznego, łącznie z jego wadami, zob. np. (Wierzchoń & Kłopotek, 2015, 35–38).

Wobec wielości dostępnych w tym zakresie metod i wskaźników, wybór ten nie powinien być zatem przypadkowy i wymaga on uzasadnienia. W wyniku wyboru dokonanego przez Huang i Chang, w każdym przedziale czasowym uzyskano szereg klastrów reprezentujących różne fronty badawcze i cechujące się różnym rozmiarem (liczbą publikacji) i problematyką (zob. Huang & Chang, 2014, 1729–1739). W celu zakwalifikowania uzyskanych frontów do różnych specjalności badawczych (ang. *knowledge domains*) w ramach wybranego obszaru, zasięgnięto opinii ekspertów. Wykorzystując klasyfikację frontów badawczych Smalla i Uphama (2010) (fronty „rosnące”, „stabilne” itd.), wzbogaconej o jeden nowy typ frontu – frontu „powtórnie wyłaniającego się” (ang. *re-emerging*), wytypowano łącznie 18 frontów badawczych należących do siedmiu różnych specjalności i wskazano specjalność najbardziej rozwojową (ang. *the most cutting-edge sector and trend for future development*). Dokonano tego na podstawie charakterystyk wchodzących w jej skład frontów, tj. oceny tego, ile każda specjalność zawierała frontów nowo powstałych, ile stabilnych, ile rosnących, ile kurczących się itd. (Huang & Chang, 2014, 1739–1740). Jako na ograniczenia zastosowanej metodologii wskazano na wykorzystanie artykułów naukowych o ustalonym progu cytowalności pomimo, że w dyscyplinach zorientowanych technologicznie patenty mogą stanowić relatywnie ważniejsze źródło informacji. Ponadto, w obliczu świadomości tempa rozwoju współczesnej technologii Huang i Chang uznali, że ustalanie okresów obserwacji cytowanych publikacji (ang. *citation window*) mogłoby zostać oparte na wartościach wskaźnika *citing* lub *cited half-life* kluczowych czasopism dyscyplinarnych, zamiast na „prywatnej” intuicji autorów (Huang & Chang, 2014, 1741–1742). Warto w tym miejscu nadmienić, że proces starzenia się literatury naukowej stanowi przeszkodę w efektywnym stosowaniu metody powiązań bibliograficznych. Prawdopodobieństwo, że bibliografie załącznikowe dwóch powiązanych tematycznie artykułów będą zawierać jakieś wspólne pozycje jest bowiem niskie jeżeli między opublikowaniem takich dwóch prac minęło dużo czasu (Glänzel & Thijs, 2012, 400–401).

7. Metoda współwystępowania słów kluczowych

W literaturze przedmiotu podaje się, że metoda analizy współwystępowania terminów/słów kluczowych wykrystalizowała się jako samodzielna technika badawcza w wydanej w 1986 r. książce pt. *Mapping the dynamics of science and technology*, której autorami byli Michel Callon, John Law i Arie Rip (zob. np.: Lee, 2008, 505; Sasson et al., 2015, 620; Van Den Besselaar & Heimeriks, 2006, 380). Najogólniej mówiąc metoda ta polega na wykrywaniu terminów występujących w tytułach, abstraktach lub treści dwóch lub więcej różnych publikacji łącznie. Ma ona zapewnić wgląd w podejmowaną w tych publikacjach tematykę i przedstawiane w nich koncepcje. Zakłada się przy tym, że różne specjalizacje badawcze posługują się różną terminologią. Powstawaniu nowych koncepcji, metod czy teorii towarzyszy powstawanie nowego słownictwa, które jest dla nich reprezentatywne (definiuje je). Wykrycie powtarzających się współwystąpień pewnych fraz w różnych, rozdzielonych okresach pozwala domniemywać, że reprezentowane przez to słownictwo trendy, koncepcje lub teorie zostały zaakceptowane przez społeczność naukową i stały się w pewnych kręgach tej społeczności „obligatoryjne” (Neef & Corley, 2009, 659; zob. też: Diodato, 1994, 54–55; Verbeek et al., 2002, 205–206). Callon i in. w swojej pracy z 1986 r.

zapropowali tzw. indeks jednoznaczności publikacji (ang. *equivalence index*), który jest wartością wprost proporcjonalną do liczby współwystąpień (w jednym dokumencie) dwóch różnych terminów i jednocześnie odwrotnie proporcjonalną do liczby różnych dokumentów zawierających te dwa terminy. Indeks ten mieści się w przedziale od 0 do 1 i im jest większy, tym bardziej jednoznaczne semantycznie są obie publikacje (Neef & Corley, 2009, 664; Sasson et al., 2015, 620). Dzięki temu uzyskuje się dokładną miarę ilościową określającą podobieństwo różnych publikacji. Miara ta jest stopniowalna (podobieństwo zmienia się „płynnie”). Możliwe staje się więc tworzenie „map”, na których znajdują się obszary terminów związanych bardzo blisko, związanych luźniej, słabo lub prawie wcale. W bibliometrii często stosuje się też metodę *co-word* łącznie z innymi technikami analizy struktur sieciowych oraz technikami statystycznymi, które pozwalają na pogłębienie analizy wzorców wykorzystywania określonej terminologii (Neef & Corley, 2009, 659; zob. też: Marshakova Shaikevich, 2005).

Koncepcję współwystępowania słów kluczowych wykorzystał Chaomei Chen (2006) w ramach modelowania szybko zachodzących zmian w problematyce frontów badawczych i zarazem w ich tzw. „bazie wiedzy” (ang. *intellectual base*). Bazę tę stanowiły publikacje cytowane przez artykuły tworzące front badawczy. Rozróżnienie to zostało zaczerpnięte przez Chena od Perssona (1994) (Chen, 2006, 360; Persson, 1994, 31). Zmiany zachodzące w obszarze frontu są na ogół powiązane ze zmianami w jego bazie intelektualnej. Dynamika wzajemnej zależności frontu i jego bazy, przesunięcia w leżącym u podstaw frontu przepływie wiedzy mogą zostać uchwycone właśnie poprzez analizę szczegółowej, dyscyplinarnej terminologii i jej (nawet nieznaczących) przemian (Chen, 2006, 360–361). Chen proponował ponadto włączenie analiz typu *co-word* do specjalnego, zaprojektowanego pod kątem identyfikacji nowo powstających trendów badawczych oprogramowania nazwanego *CiteSpace II*. Zasadą jego działania jest analiza sieci cytowań publikacji w czasie z wykorzystaniem zautomatyzowanych algorytmów oraz twierdzeń i wskaźników pochodzących z teorii grafów. Elementem związanym ściśle z analizą terminologii jest detekcja tzw. „eksplozji” pewnych wiązek terminów/słów kluczowych (ang. *burst detection*). Przez „eksplozję” rozumie się tutaj nagły, gwałtowny wzrost częstotliwości wykorzystywania przez badaczy z danej specjalności określonych słów kluczowych czy fraz, co sygnalizuje wzrost zainteresowania i nasilenie badań prowadzonych nad daną problematyką. Jest to zjawisko znamionujące raczej zmianę orientacji frontu niż jego ogólny, stabilniejszy profil tematyczny (Chen, 2006, 364–365, 374–375). W *CiteSpace II* kluczowe frazy/terminy wykryte w takich wiązkach stosuje się poza tym jako „etykiety” (ang. *labels*) frontów badawczych. Ma to zapewnić bardziej precyzyjny opis charakteru frontu niż prosta agregacja terminów najczęściej występujących w klastrach publikacji uzyskanych z wykorzystaniem służących temu, tradycyjnych metod (Chen, 2006, 375). Chen zaznaczył ponadto, że podejście oparte na badaniach terminologii (ang. *word-profile approach*) nie jest też pozbawione pewnych wad. Jedną z nich jest wspomniana już i stosowana niekiedy prosta agregacja terminów, które powtarzają się najczęściej w pewnych klastrach. Znaczenie takich terminów może być bardzo zróżnicowane, „rozbiegające się” wielokierunkowo, co wymaga znacznego wkładu własnego osoby analizującej terminy, aby powiązać je w sensowny sposób. Po drugie, operacja nadawania klastrów nazw często skutkuje nazewnictwem zbyt szerokim (ogólnym), aby mogło być ono użyteczne. Terminy skojarzone ściśle z trendami nowo powstającymi mogą zostać niejako przesłonięte przez terminy należące do obszarów z dłuższą tradycją, lepiej

ugruntowanych, ogólniej – obszarów w danym klastrze pierwszoplanowych (ang. *persistent*) (Chen, 2006, 364). Podejściu tekstualnemu stawiano też podobne zarzuty związane z jego niewystarczającą zdolnością do odróżniania prac należących do odmiennych specjalności, czego przyczyn upatrywano w niskiej wrażliwości metody na różne konteksty posługiwania się zbliżonym nazewnictwem i wynikające stąd „przecenianie” (ang. *overestimate*) stopnia podobieństwa tych prac (Glänzel & Thijs, 2012, 400).

Woo Hyoung Lee (2008) analizował przy użyciu techniki współwystępowania terminów funkcjonującą w informatyce subdyscyplinę bezpieczeństwa informacji. Jego podejście polegało na stworzeniu „mapy wiedzy” (ang. *knowledge map*) dla wybranego 2006 r., która składała się z terminów występujących równocześnie w różnych publikacjach. Uzyskaną w ten sposób mapę tworzyły „gniazda” (centra – ang. *hubs*) terminów związanych ze sobą ściślej i luźniej. Siła tego związku determinowała położenie terminów na mapie (Lee, 2008, 509). Taka struktura jest bliskim odpowiednikiem klastrów publikacji, które również są zbiorami słabiej i silniej połączonych elementów (leżących tak wewnątrz jak i na zewnątrz konkretnego klastra). Dalsza analiza wykorzystywała parametry struktur sieciowych, takie jak „centralność” (ang. *centrality*) i „wskaźnik znaczenia w sieci” lub „miara stopnia pośredniczenia” (ang. *betweenness*) (Lee, 2008, 509, 522). Są to wskaźniki typowe dla analiz klastrów złożonych z wierzchołków i połączeń między nimi („krawędzi”). Konstrukcja metody Lee opierała się na odnalezieniu centrum zbudowanej metodą *co-word* sieci. Centrum to było grupą blisko związanych wierzchołków o niespotykanie dużej liczbie połączeń (współwystąpień) z innymi wierzchołkami w całej sieci. Centrum charakteryzowało się więc wysoką miarą „centralności”, co można porównać do wysokiej siły wpływu, silnego oddziaływania na całą subdyscyplinę (Lee, 2008, 511–512). Centrum było także miejscem o wysokim stopniu „pośredniczenia” na mapie, tj. leżało na drodze łączącej wiele różnych punktów rozsianych w całej strukturze (Lee, 2008, 522). Z użyciem tych parametrów określono właściwości odnalezionego w danym roku centrum, aby następnie „cofnąć się w czasie” (tj. przeanalizować tą samą metodą grupę publikacji wydanych wcześniej), aż do okresu w którym aktualne centrum dopiero zaczynało się formować. Należało teraz ponownie zbudować analogiczną mapę wiedzy dla tamtego (wcześniejszego) okresu i sprawdzić, jak wyglądały parametry aktualnego centrum (tj. centrum z 2006 r.) w okresie jego „młodości”. Jeżeli na aktualnej mapie uda się odnaleźć takie rejony, które mają podobne lub identyczne właściwości strukturalne, jakie wykazywało aktualne centrum w swojej młodości, byłyby one właśnie tymi, zdefiniowanymi leksykalnie rejonami, które mają największe szanse, aby stać się centrum w przyszłości (Lee, 2008, 509). Lee wytypował w ten sposób kryptografię, jako najbardziej perspektywiczną subdyscyplinę informatyki, zarazem zaznaczył jednak, że efektywność jego metody jest trudna (a nawet niemożliwa) do natychmiastowego zweryfikowania (Lee, 2008, 519, 521, 524).

Mark Neef i Elizabeth Corley (2009) posłużyli się metodą współwystępowania terminów zaczerpniętych z tytułów publikacji do rozpoznania trendów badawczych i ewolucji, powstawania i zanikania obszarów problemowych w obrębie ekologii pomiędzy rokiem 1970 a 2005 (Neef & Corley, 2009, 658, 662–663). Jak zaznaczyli Neef i Corley, w przeciwieństwie do niektórych innych przedsięwzięć wykorzystujących tę samą metodę, w ich pracy analizowano nie tylko zestawy słów współwystępujących najczęściej w dłuższych okresach, ale też podjęto próbę wyodrębnienia trendów krótkotrwałych, podlegających gwałtownym zmianom i obecnych niejako na tle trendów stabilniejszych i długotrwałych.

Zamiarem było tutaj wyróżnienie najważniejszych tematów badawczych dominujących w różnym czasie i prześledzenie tego, jak zmieniała się ich rola dla dyscypliny ekologii jako całości (Neef & Corley, 2009, 660). Zastosowano ponadto normalizację słownictwa, polegającą m.in. na scaleniu słów występujących w liczbie pojedynczej i mnogiej, ujednoczeniu synonimicznych wersji różnych wyrażzeń, ujednoczeniu wersji różniących się literacją (np. pisownia amerykańska i brytyjska), albo słów pisanych w wersji z myślnikiem lub bez (Neef & Corley, 2009, 662). Cały okres badania podzielono na dwa podokresy (1970–1987 i 1988–2005), a następnie podzielono częstość wystąpień każdego (znormalizowanego już) słowa w danym roku przez całkowitą liczbę słów użytych we wszystkich tytułach publikacji ogłoszonych w tym samym roku (Neef & Corley, 2009, 664). W dalszej kolejności Neef i Corley posłużyli się indeksem podobieństwa Callona i in. z 1986 r. oraz algorytmem grupującym (tzw. algorytm Warda – zob. np.: Mojena, 2006). Uzyskano klastry zawierające finalnie od 2 do 9 słów i zastosowano do nich miary „centralności” (jak centralna dla całej ekologii jest tematyka danego klastra?) i „gęstości” (ang. *density*). Jeżeli klastr jest „gęsty” oznacza to, że autorzy używając jednego ze słów należących do klastra, bardzo często używają też jednocześnie pozostałych słów składających się na ten klastr (Neef & Corley, 2009, 665–666). Każdy uzyskany klastr umieszczono następnie na diagramie pokazującym zależność „centralności” od „gęstości”. Klastry o wysokich wskaźnikach obu tych miar stanowiły centra dla swoich dyscyplin reprezentując koncepty spójne i dobrze rozwinięte. Klastry o niskiej gęstości i zarazem wysokiej centralności reprezentowały natomiast zagadnienia ważne dla dyscypliny, ale jeszcze słabo rozwinięte – zagadnienia nowo wyłaniające się (ang. *emerging*) (Neef & Corley, 2009, 667–670). Dodatkowo poddano szczegółowej analizie słowa wykazujące największe wahania (wzrost i spadek częstości występowania) w obu wyróżnionych podokresach i powtórzono dla nich cały proces badawczy streszczony powyżej (Neef & Corley, 2009, 673–675). Do podstawowych ograniczeń tej metody wskazano możliwą wieloznaczność terminów, ich zależność od kontekstu i zmienność znaczenia terminów technicznych w czasie. Przywołano w tym kontekście opinię Loeta Leydesdorffa z 1997 r., według której język jest tworem zbyt dynamicznym, aby mógł stać się narzędziem efektywnej analizy typu *co-word* w zastosowaniach naukoznawczych (Neef & Corley, 2009, 667). Wskazano również na możliwą nieadekwatność tytułów niektórych publikacji do ich zawartości treściowej, a także na problem pojawiający się w trakcie analizy niewielkich, wąskich pól badawczych. Polega on na niemożności uchwycenia trendów „szerokozakresowych”, przenikających więcej niż jedną subdyscyplinę lub specjalność. Problem ten może być o tyle istotny, że to właśnie trendy tego rodzaju stanowią na ogół przedmiot zainteresowania polityki naukowej i mogą nieść ze sobą następstwa sięgające poza sam obszar czysto akademicki (Neef & Corley, 2009, 659–660, 667–678; zob. też: Van Den Besselaar & Heimeriks, 2006, 380).

Odmienne podejście do problemu prognozowania rozwoju nauki i technologii przyjęli Takao Furukawa i in. (2015). Skupili się oni na międzynarodowych konferencjach jako na wydarzeniach, podczas których prezentuje się najnowsze odkrycia i osiągnięcia. Konferencje naukowe (często cykliczne) zapewniają – zdaniem Furukawy i in. – najszybszy sposób transferu informacji, a ich podstawową funkcją jest przesuwanie granic wiedzy (ang. *advance the frontiers of knowledge*) (Furukawa et al., 2015, 281). Wykorzystana przez zespół autorów metoda była standardową analizą typu *co-word* zastosowaną do abstraktów wystąpień konferencyjnych dostępnych w Internecie. Na jej podstawie stworzono „sieć

czasową” (ang. *temporal network*) sesji konferencyjnych, która pozwalała zaobserwować ewolucję tematyki poruszanej podczas sesji i wytypować problematykę aktualnie najbardziej wpływową (Furukawa et al., 2015, 283–288). To, co wydaje się najistotniejsze w przypadku metody Furukawy i in. to fakt, że w ich ocenie tematyka sesji konferencyjnych ma silne tendencje do wskazywania przyszłościowych perspektyw rozwoju dyscyplin. Rada programowa konferencji ustala zazwyczaj wiodącą tematykę i selekcjonuje ważniejsze problemy, które powinny zostać omówione, co odzwierciedla sposób postrzegania przez społeczność naukową tego, co w danej dyscyplinie można uważać za pionierskie (Furukawa et al., 2015, 291). Problematyczną stroną metody zespołu autorów mogą stanowić natomiast przypadki, w których na konferencję zostaje zgłoszonych bardzo dużo referatów, które zostają wtedy „sztucznie” rozdysponowane pomiędzy różne sesje. Może to wywołać sytuację, w której w przybliżeniu ta sama tematyka jest omawiana na odrębnych sesjach (noszących inne tytuły) lub odwrotnie, w obrębie jednej sesji konferencyjnej wygłaszane referaty mogą nie pokrywać nawet w przybliżeniu jednakowych zagadnień (Furukawa et al., 2015, 292). Xiaoling Sun i in. (2016, 759) wskazują na analogiczny problem, na jaki napotyka metoda identyfikacji współwystępujących słów kluczowych. Polega on na niezdolności uchwycenia podobieństwa takich publikacji, w których autorzy nie posługiwali się podobnym słownictwem, a mimo to zajmowali się bliskimi sobie zagadnieniami.

Tianlong Zheng i in. (2016) wykorzystali metodę współwystępowania terminów do analizy aktualnych tendencji badawczych oraz najnowszych prądów intelektualnych (ang. *hotspots*) w międzydyscyplinarnej specjalności zajmującej się badaniem mikro- i nanopęcherzyków (ang. *micro/nano-bubbles*) (Zheng et al., 2016, 55). Zbadano ilościowe wzorce (wzrost i spadek) współwystępowania 30 najczęściej pojawiających się słów kluczowych w publikacjach z lat 1991–2014 wyszukanych w Web of Science (WoS). Dla uwydatnienia wymiaru czasowego analizy cały ten okres podzielono na cztery sześciolatnie przedziały i dla każdego z nich skonstruowano mapę (ang. *co-word network*) (Zheng et al., 2016, 55, 63, 65). Innowacyjnym elementem metody zespołu Zhenga było wykorzystanie nie tylko autorskich słów kluczowych i słów widniejących w tytułach publikacji, ale także tzw. słów „dodatkowych” (ang. *KeyWords Plus*) dostępnych w bazie WoS. Słowa te wyróżniają się pod względem źródła pochodzenia, którym są tytuły publikacji przywoływanych w bibliografiach załącznikowych artykułu, do którego słowa te są dodawane przez specjalny algorytm (Zheng et al., 2016, 65). Metoda, która wyróżnia się uwzględnieniem w analizie „dodatkowych” słów kluczowych (obok słów autorskich i słów z tytułów publikacji) została określona jako „analiza klastrów słownych” (ang. *word cluster analysis*) (Zheng et al., 2016, 65). Zdaniem niektórych badaczy słowa „dodatkowe” są ukierunkowane bardziej niż słowa autorskie na nowatorskość i innowacyjność kierunków podejmowanych i referowanych w publikacjach badań, niż na ich postać tradycyjną i głębiej zakorzenioną w nauce (Li et al., 2009, 700; zob. też: Li et al., 2011, 17).

8. Zakończenie

Podsumowując stwierdzenia i fakty streszczone w ramach powyższych podrozdziałów, można przedstawić kilka ogólnych wniosków dotyczących postaci, korzeni i perspektyw zarysowujących się przed nurtem prognostycznym istniejącym we współczesnej bibliometrii.

Przede wszystkim warto podkreślić, że badania pionierów, prowadzone głównie w latach 60. i 70. XX w. nadały początek i w dużym stopniu ukształtowały aktualną formę, jaką przybiera większość prób oceny i przewidywania dalszego postępu badań naukowych. Są one wciąż żywotne i inspirują teraźniejsze wysiłki badaczy, co świadczy o ich ponadczasowym charakterze i uwidacznia praktyczne potrzeby, na które stanowią one racjonalną odpowiedź. Nie znaczy to zarazem, że metody określone w niniejszym opracowaniu jako tradycyjne, pozbawione są wszelkich wad. Przeciwnie, różnorakie braki metodologiczne zostały dostrzeżone niemal równocześnie z pojawieniem się i pierwszymi próbami stosowania tych metod w praktyce, a krytyka poszczególnych odmian bibliometrycznej metodologii prognostycznej i dążenie do jej udoskonalenia stanowią integralną część i etap rozwoju całej tej gałęzi naukoznawstwa. W tym kontekście należy również zwrócić uwagę na fakt, iż najważniejszymi i najczęściej powtarzającymi się w praktyce badawczej problemami czy zarzutami stawianymi przed metodami podstawowymi są niewątpliwie:

- Arbitralność ustalania granicznych progów siły współcytowania oraz arbitralność definiowania publikacji wysoko cytowanych. Nabiera to obecnie szczególnego znaczenia w czasach współczesnych, bowiem w ogólnościatowych indeksach cytowań widnieją publikacje cytowane po kilkanaście, a nawet kilkadziesiąt tysięcy razy.
- Metoda grupowania jednopiętowego wykazuje silną tendencję do tworzenia klastrów niejednorodnych tematycznie, co w warunkach współczesnej, wysokiej interdyscyplinarności wielu dyscyplin naukowych powoduje trudności z przyporządkowaniem tak zbudowanego klastra do właściwej dziedziny/dyscypliny.
- Z drugiej strony w obrębie bibliometrii i nauk komputerowych ma się obecnie do czynienia z tak dużą różnorodnością zarówno algorytmów grupujących jak i kryteriów podobieństwa publikacji, w oparciu o które przeprowadza się operacje grupowania, że dokonanie i uzasadnienie konkretnego wyboru również stało się poważnym wyzwaniem. Znajomość wszystkich konsekwencji dokonanego wyboru i możliwych, wywołanych przezeń nieprawidłowości wymaga z jednej strony dogłębnej znajomości ściśle technicznych aspektów działania algorytmów grupujących, z drugiej zaś dostępu do badań porównawczych, który to dostęp – jak postarano się zaznaczyć w drugiej części niniejszego artykułu – jest aktualnie raczej ograniczony.
- Powyższe uwagi odnoszą się również do rozmaitych procedur normalizacyjnych wprowadzanych w obszar prognostycznych badań empirycznych na etapie ich projektowania.
- Często podkreślanym problemem w stosowaniu metody współcytowań jest ponadto fakt istnienia opóźnień w pojawianiu się kolejnych cytowań, które wynikają tak z konieczności dotarcia idei przenoszonych przez publikacje do grona ich odbiorców i wywarcia wpływu na ich myśl, jak i z faktu istnienia – nierzadko bardzo długich – opóźnień w samym procesie publikacyjnym (w cyklu wydawniczym).
- Problematyczne pozostaje w dalszym ciągu także pytanie o maksymalny dopuszczalny rozmiar klastra, oraz o jego rozmiar minimalny, tj. o minimalną liczbę składających się na niego publikacji, która pozwoliłaby na uznanie go za twór samodzielny.
- Szereg problemów metodologicznych wynika ponadto z rozbieżności opinii specjalistów nt. tego, które ze źródeł danych stanowią najlepszą odpowiedź na potrzeby związane z monitorowaniem dyscyplin ze względu na tempo i zakres ich rozwoju, tj. jak i w jakich dyscyplinach należy traktować artykuły z czasopism, monografie

- i rozdziały w monografiach, dokumenty patentowe, materiały konferencyjne itd. Innym obliczem tego samego problemu są również kwestie fizycznej dostępności do różnego rodzaju danych źródłowych, które stanowią elementarny wymóg samej możliwości prowadzenia badań, a także warunkują do pewnego stopnia ich jakość.
- Na tym samym poziomie mankamentów metodologicznych usytuowane są kwestie dotyczące wyboru spośród wielu możliwych poziomów agregacji danych empirycznych, jak również kwestie związane z cytowalnością typową dla różnych form wydawniczych oraz różnych dyscyplin czy obszarów badawczych.
 - W odniesieniu do metody generowania powiązań bibliograficznych, w tekście niniejszego artykułu wskazano na jeden z ciekawszych problemów, którym jest zjawisko starzenia się literatury cytowanej stanowiące przeszkodę w porównywaniu publikacji wydanych w różnym czasie pod względem tego, jak wyglądają tworzone w ramach tej metody fronty badawcze (zob. podrozdz. 6).
 - W odniesieniu do metody współwystępowania słów kluczowych można dodatkowo w tym miejscu przypomnieć, że głównymi wysuwanyymi wobec niej zastrzeżeniami są trudności związane z semantyką terminów naukowych – np. zależność ich znaczenia od kontekstu, wieloznaczność, nieadekwatność do niektórych koncepcji, trudności w identyfikowaniu trendów szerokok zakresowych za pośrednictwem terminów o wąskim i wysoce wyspecjalizowanym znaczeniu.

Ponadto ważne w kontekście przedstawionej pierwszej części całej analizy piśmiennictwa, jak i jej części drugiej, która ukaże się w następnym numerze ZIN, wydaje się nawiązujące do punktu 3 spostrzeżenie, według którego istnieje silna korelacja pomiędzy prognostycznymi metodami bibliometrycznymi, a algorytmicznymi metodami grupowania publikacji w spójne tematycznie klastry reprezentujące wąskie, dyscyplinarne „subspecjalności” czy mówiąc inaczej, obszary problemowe poddawane dalszej analizie ilościowej o różnie skonstruowanych profilach. Korelacja ta jest dostrzegalna we wszystkich wyróżnionych podgrupach publikacji, które omówiono w odrębnych podrozdziałach. Nasuwa to myśl o konieczności rozważenia jeszcze jednego, fundamentalnego problemu rozgraniczenia obszarów problemowych czy specjalności, poprowadzenia linii podziału pomiędzy nimi (tzw. delineacja) jeszcze przed przystąpieniem do badań zorientowanych prognostycznie. Innymi słowy, aby móc powiedzieć cokolwiek o perspektywie dalszego rozwoju dyscypliny, konieczne jest jej wcześniejsze zdefiniowanie i wyodrębnienie w sensie wyselekcjonowania wszystkich i zarazem tylko tych publikacji, które faktycznie do niej przynależą i decydują o jej rzeczywistym kształcie (strukturze). Popełnienie błędu na tym etapie procedury badawczej skutkuje więc zawsze zniekształceniem ostatecznego wyniku badania i wyrowadzeniem mylnych wniosków co do tempa lub kierunku rozwoju konkretnej dyscypliny. Problem ten jest jedną z istotniejszych kwestii, którą zajmują się przedstawiciele bibliometrii nie tylko w związku z przewidywaniem rozwoju nauki, ale też w związku z zagadnieniami z zakresu np. strategii wyszukiwania informacji lub ewaluatywnej analizy cytowań. Wskazuje się w tym kontekście na wady istniejących systemów klasyfikacyjnych (np. systemu przyjętego w *Journal Citation Reports* lub w indeksie cytowań *Scopus*) funkcjonujących ponadto głównie na poziomie czasopism, a nie indywidualnych publikacji (zob. np.: Waltman & Van Eck, 2012, 2378–2379). Fakt ten rzuca dodatkowe światło na potrzebę dalszego dopracowywania istniejących metod prognostycznych, która stanowi jeden z wątków przewodnich drugiej części niniejszego artykułu.

Bibliografia

- Adler, R.; Ewing, J.; Taylor, P. (2009). Citation Statistics. A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). *Statistical Science*, 24(1), 1–14.
- Boyack, K.W.; Börner, K.; Klavans, R. (2009). Mapping the Structure and Evolution of Chemistry Research. *Scientometrics*, 79(1), 45–60.
- Boyack, K.W.; Klavans, R. (2010). Co-Citation Analysis, Bibliographic Coupling and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404.
- Boyack, K.W.; Klavans, R. (2014). Creation of Highly Detailed, Dynamic, Global Model and Map of Science. *Journal of the Association for Information Science and Technology*, 65(4), 670–685.
- Braun, T.; Schubert, A.; Kostoff, R. (2000). Growth and Trends of Fullerene Research as Reflected in Its Journal Literature. *Chemical Reviews*, 100(1), 23–37.
- Briotta Parolo, P.; Pan, R.K.; Ghosh, R.; Huberman, B.A.; Kaski, K.; Fortunato, S. (2015). Attention Decay in Science. *Journal of Informetrics*, vol. 9(4), 734–745.
- Chen, Ch. (2006). CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377.
- Chen, S-H.; Huang, M-H.; Chen, D-Z. (2012). Identifying and Visualizing Technology Evolution: A Case Study of Smart Grid Technology. *Technological Forecasting and Social Change*, 79(6), 1099–1110.
- Diodato, V. (1994). *Dictionary of Bibliometrics*. New York: London: Norwood: The Haworth Press.
- Dorta-González, P.; Dorta-González, M-I. (2013). Impact Maturity Times and Citation Time Windows: The 2-Year Maximum Journal Impact Factor. *Journal of Informetrics*, 7(3), 593–602.
- Ena, O.; Mikova, N.; Saritas, O.; Sokolova, A. (2016). A Methodology for Technology Trend Monitoring: The Case of Semantic Technologies. *Scientometrics*, 108(3), 1013–1041.
- Érdi, P.; Makovi, K.; Somogyvári, Z.; Strandburg, Ka.; Tobochnik, J.; Volf, P.; Zalányi, L. (2013). Prediction of Emerging Technologies Based on Analysis of the US Patent Citation Network. *Scientometrics*, 95(1), 225–242.
- Furukawa, T.; Mori, K.; Arino, K.; Hayashi, K.; Shirakawa, N. (2015). Identifying the Evolutionary Process of Emerging Technologies: A Chronological Network Analysis of World Wide Web Conference Sessions. *Technological Forecasting and Social Change*, 91, 280–294.
- Garcia, E.K. (2015). *Cosine Similarity Tutorial* [online]. Department of Electrical Engineering, University of Washington, [06.11.2016], <http://www.minerazzi.com/tutorials/cosine-similarity-tutorial.pdf>
- Garfield, E. (2004). Historiographic Mapping of Knowledge Domains Literature. *Journal of Information Science*, 30(2), 119–145.
- Garfield, E.; Pudovkin, A.; Istomin, V.S. (2002). Algorithmic Citation-Linked Historiography-Mapping the Literature of Science. *Proceedings of the American Society for Information Science and Technology*, 39(1), 14–24.
- Garfield, E.; Pudovkin, A.; Istomin, V.S. (2003). Why Do We Need Algorithmic Historiography? *Journal of the American Society for Information Science and Technology*, 54(5), 400–412.
- Garfield, E.; Sher, I.; Torpie, R. (1964). *The use of citation data in writing the history of science. A final report of research for the Air Force Office of Scientific Research under contract AF 49(638)-1256*. Philadelphia, Pennsylvania, USA: Institute for Scientific Information, Inc.
- Glänzel, W. (2012). Bibliometric Methods for Detecting and Visualizing Emerging Research Topics. *El Profesional de la Información*, 21(1), 194–201.
- Glänzel, W.; Schoepflin, U. (1995). A Bibliometric Study on Ageing and Reception Processes of Scientific Literature. *Journal of Information Science*, 21(1), 37–53.

- Glänzel, W.; Schoepflin, U. (1999). A Bibliometric Study of Reference Literature in the Sciences and Social Sciences. *Information Processing and Management*, 35(1), 31–44.
- Glänzel, W.; Thijs, B. (2012). Using 'Core Documents' for Detecting and Labelling New Emerging Topics. *Scientometrics*, 91(2), 399–416.
- Guo, H.; Weingart, S.; Börner, K. (2011). Mixed-Indicators Model for Identifying Emerging Research Areas. *Scientometrics*, 89(1), 421–435.
- Hertzfel, D. (1987). Bibliometrics, History of the Development of Ideas In. In: A. Kent, H. Lancour, W. Nasri (eds.). *Encyclopedia of Library and Information Science*, vol. 42, Suppl. 7. New York: Marcel Dekker, Inc., 144–219.
- Huang, A. (2008). Similarity Measures for Text Document Clustering. In: J. Holland, A. Nicholas, D. Brignoli (eds.). *Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC 2008)*. Christchurch, New Zealand, April 14–18, 2008, 49–56 [online]. University of Canterbury, [30.08.2017], https://nzcsrsc08.canterbury.ac.nz/site/proceedings/NZCSRSC_2008_Proceedings.pdf
- Huang, M-H.; Chang, C-P. (2014). Detecting Research Fronts in OLED Field Using Bibliographic Coupling with Sliding Window. *Scientometrics*, 98(3), 1721–1744.
- Huang, M-H.; Chang, C-P. (2015). A Comparative Study on Detecting Research Fronts in the Organic Light-Emitting Diode (OLED) Field Using Bibliographic Coupling and Co-Citation. *Scientometrics*, 102(3), 2041–2057.
- Hummon, N.P.; Doreian, P. (1989). Connectivity in a Citation Network: The Development of DNA Theory. *Social Networks*, 11(1), 39–63.
- Jarneving, B. (2005). A Comparison of Two Bibliometric Methods for Mapping of the Research Front. *Scientometrics*, 65(2), 245–263.
- Kajikawa, Y.; Yoshikawa, J.; Takeda, Y.; Matsushima, K. (2008). Tracking Emerging Technologies in Energy Research: Toward a Roadmap for Sustainable Energy. *Technological Forecasting and Social Change*, 75(6), 771–782.
- Kessler, M.M. (1963). Bibliographic Coupling Between Scientific Papers. *American Documentation*, 14(1), 10–25.
- KPB (2011). *Krajowy Program Badań. Zakożenia polityki naukowo-technicznej i innowacyjnej państwa. Krajowy Program Badań opracowany na podstawie art. 4 ust. 1 ustawy o zasadach finansowania nauki. Załącznik do uchwały nr 164/2011 Rady Ministrów z dnia 16 sierpnia 2011 r.* [online]. Ministerstwo Nauki i Szkolnictwa Wyższego, [30.08.2017], http://www.ncbir.pl/gfx/ncbir/user-files/_public/bip/20110816_kpb.pdf
- Lee, W.H. (2008). How to Identify Emerging Research Fields Using Scientometrics: An Example in the Field of Information Security. *Scientometrics*, 76(3), 503–525.
- Lee, L-Ch.; Lee, Y-Y.; Liaw, Y-Ch. (2012). Bibliometric Analysis for Development of Research Strategies in Agricultural Technology: The Case of Taiwan. *Scientometrics*, 93(3), 813–830.
- Leydesdorff, L. (1987). Various Methods for the Mapping of Science. *Scientometrics*, 11(5–6), 295–324.
- Li, J.; Zhang, Y.; Wang, X.; Ho, Y-S. (2009). Bibliometric Analysis of Atmospheric Simulations Trends in Meteorology and Atmospheric Science Journals. *Croatica Chemica Acta*, 82(3), 695–705.
- Li, J.; Wang, M-H.; Ho, Y-S. (2011). Trends in Research on Global Climate Change: A Science Citation Index Expanded-Based Analysis. *Global and Planetary Change*, 77(1), 13–20.
- Liu, Ch.; Gui, Q. (2016). Mapping Intellectual Structures and Dynamics of Transport Geography Research: A Scientometric Overview from 1982 to 2014. *Scientometrics*, 109(1), 159–184.
- Liu, J.S.; Lu, L.Y.Y. (2012). An Integrated Approach for Main Path Analysis: Development of the Hirsh Index as an Example. *Journal of the American Society for Information Science and Technology*, 63(3), 528–542.
- Ma, V.C.; Liu, J.S. (2016). Exploring the Research Fronts and Main Paths of Literature: A Case Study of Shareholder Activism Research. *Scientometrics*, 109(1), 33–52.

- Mabe, M.A. (2010). Scholarly Communication: A Long View. *New Review of Academic Librarianship*, 16(Suppl. 1), 132–144.
- Marshakova Shaikovich, I. (1973). System of Document Connections Based on References. *Nauchno-Tekhnicheskaya Informatsiya. Seriya 2, Informatsionnye Protsessy i Sistemy*, 6(2), 3–8.
- Marshakova Shaikovich, I. (2005). Bibliometric Maps of Field of Science. *Information Processing and Management*, 41(6), 1534–1547.
- Martín-Martín, A.; Orduna-Malea, E.; Allyón, J.M.; Delgado López-Cózar, E. (2016). *The Counting House: Measuring Those Who Count. Presence of Bibliometrics, Scientometrics, Informetrics, Webometrics and Altmetrics in the Google Scholar Citations, ResearcherID, ResearchGate, Mendeley & Twitter* [online]. EC3 Working Papers, vol. 21, 1–60, [30.08.2017], <https://arxiv.org/ftp/arxiv/papers/1602/1602.02412.pdf>
- Meadows, J. (2000). The Growth of Journal Literature. In: B. Cronin & H. B. Atkins (eds.). *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Medford, NJ: Information Today, Inc., 87–108.
- MNiI (2004). *Założenia polityki naukowej, naukowo-technicznej i innowacyjnej państwa. Projekt 1.10.2004 r.* [online]. Ministerstwo Nauki i Informatyzacji, [30.08.2017], http://www.nauka.gov.pl/g2/oryginal/2013_05/ac91da4d11aad3d93649c13ed181b70a.pdf
- Moed, H.F.; van Leeuwen, T.N.; Reedijk, J. (1998). A New Classification System to Describe the Ageing of Scientific Journals and Their Impact Factors. *Journal of Documentation*, 54(4), 387–419.
- Mojena, R. (2006). Ward's Clustering Algorithm. In: S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic (eds.). *Encyclopedia of Statistical Sciences* [online]. Wiley Online Library, [11.11.2016], <http://onlinelibrary.wiley.com/doi/10.1002/0471667196.ess2887.pub2/abstract>
- Muraszkiewicz, M. (2014). An Essay on Information Overload. *Zagadnienia Informacji Naukowej. Studia Informacyjne*, 52, 1(103), 7–18.
- Nowak, P. (2006). *Bibliometria. Webometria. Podstawy, wybrane zastosowania*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza w Poznaniu.
- Neef, M.W.; Corley, E.A. (2009). 35 Years and 160,000 Articles: A Bibliometric Exploration of the Evolution of Ecology. *Scientometrics*, 80(3), 657–682.
- Persson, O. (1994). The Intellectual Base and Research Fronts of JASIS 1986–1990. *Journal of the American Society for Information Science*, 45(1), 31–38.
- Pindłowa, W. (1988). Bibliometria i jej znaczenie dla badań nad książką. *Studia o Książce*, T. 17, 301–327.
- Pindłowa, W. (1994). *Bibliometria w nauce o informacji. Metody i problemy*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”.
- Potiopa, P. (2011). Metody i narzędzia automatycznego przetwarzania informacji tekstowej i ich wykorzystanie w procesie zarządzania wiedzą. *Automatyka*, T. 15(2), 409–419.
- Price, D. de Solla (1967). *Mała Nauka – Wielka Nauka*. Warszawa: PWN.
- Pritchard, A. (1969). Statistical Bibliography or Bibliometrics? *Journal of Documentation*, 25(4), 348–349.
- Rousseau, R. (2014). Library Science: Forgotten Founder of Bibliometrics. *Nature*, 510(7504), 218.
- Salton, G.; Bergmark, D. (1979). A Citation Study of Computer Science Literature. *IEEE Transactions on Professional Communication*, 22(3), 146–158.
- Sasson, E.; Ravid, G.; Pliskin, N. (2015). Improving Similarity Measures of Relatedness Proximity: Toward Augmented Concept Maps. *Journal of Informetrics*, 9(3), 618–628.
- Seweryn, A.; Swoboda, I. (2013). *Cytowania w polskich bazach bibliograficznych* [online]. XII Krajowe Forum Informacji Naukowej i Technicznej: Społeczeństwo – Informacja – Innowacje. Wyzwania Ery Cyfrowej. Zakopane, 24–27 września 2013 r., [30.08.2017], http://www.ptin.org.pl/konferencje/12forum/repozytorium/Seweryn_Swoboda.pdf
- Shibata, N.; Kajikawa, T.; Matsushima, K. (2007). Topological Analysis of Citation Networks to Discover the Future Core Articles. *Journal of the American Society for Information Science and Technology*, 58(6), 872–882.

- Shibata, N.; Kajikawa, Y.; Sakata, I. (2012). Link Prediction in Citation Networks. *Journal of the American Society for Information Science and Technology*, 63(1), 78–85.
- Shibata, N.; Kajikawa, Y.; Takeda, Y.; Matsushima, K. (2008). Detecting Emerging Research Fronts Based on Topological Measures in Citation Networks of Scientific Publications. *Technovation*, 28(11), 758–775.
- Shibata, N.; Kajikawa, Y.; Takeda, Y.; Matsushima, K. (2009). Comparative Study on Methods of Detecting Research Fronts Using Different Types of Citation. *Journal of the American Society for Information Science and Technology*, 60(3), 571–580.
- Skalska-Zlat, M. (1988). Bibliometria – pojęcia, metody, kierunki badań. *Roczniki Biblioteczne*, T. XXXII (2), 259–283.
- Skalska-Zlat, M. (1993). *Bibliometryczne badania rozwoju dyscypliny naukowej*. Seria: Bibliotekoznawstwo XVII. Wrocław: Wydaw. UWr.
- Skalska-Zlat, M. (1999). Bibliografia jako odbicie rozwoju nauki. *Zagadnienia Naukoznawstwa*, 1(139), 57–63.
- 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(4), 265–269.
- Small, H. (2006). Tracking and Predicting Growth Areas in Science. *Scientometrics*, 68(3), 595–610.
- Small, H.; Boyack, K.W.; Klavans, R. (2014). Identifying Emerging Topics in Science and Technology. *Research Policy*, 43(8), 1450–1467.
- Small, H.; Griffith, B.C. (1974). The Structure of Scientific Literatures I: Identifying and Graphing Specialties. *Science Studies*, 4(1), 17–40.
- Small, H.; Sweeney, E. (1985). Clustering the Science Citation Index Using Co-Citations: I. A Comparison of Methods. *Scientometrics*, 7(3–6), 391–409.
- Sosińska-Kalata, B. (2002). *Klasyfikacja. Struktury organizacji wiedzy, piśmiennictwa i zasobów informacyjnych*. Warszawa: Wydaw. SBP.
- Soós, S. (2014). Age-Sensitive Bibliographic Coupling Reflecting the History of Science: The Case of the Species Problem. *Scientometrics*, 98(1), 23–51.
- Sun, X.; Ding, K.; Lin, Y. (2016). Mapping the Evolution of Scientific Fields Based on Cross-Field Authors. *Journal of Informetrics*, 10(3), 750–761.
- Takeda, Y.; Kajikawa, Y. (2009). Optics: A Bibliometric Approach to Detect Emerging Research Domains and Intellectual Bases. *Scientometrics*, 78(3), 543–558.
- Tseng, Y-H.; Lin, Y-I.; Lee, Y-Y.; Hung, W-Ch.; Lee, Ch-H. (2009). A Comparison of Methods for Detecting Hot Topics. *Scientometrics*, 81(1), 73–90.
- Tu, Y-N.; Hsu, S-L. (2016). Constructing Conceptual Trajectory Maps to Trace the Development of Research Fields. *Journal of the Association for Information Science and Technology*, 67(8), 2016–2031.
- Upham, S.P.; Small, H. (2010). Emerging Research Fronts in Science and Technology: Patterns of New Knowledge Development. *Scientometrics*, 83(1), 15–38.
- Van Den Besselaar, P.; Heimeriks, G. (2006). Mapping Research Topics Using Word-Reference Co-Occurrences: A Method and an Exploratory Case Study. *Scientometrics*, 68(3), 377–393.
- Verbeek, A.; Debackere, K.; Luwel, M.; Zimmermann, E. (2002). Measuring Progress and Evolution in Science and Technology – I: The Multiple Uses of Bibliometric Indicators. *International Journal of Management Reviews*, 4(2), 179–211.
- Waltman, L.; Van Eck, N.J. (2012). A New Methodology for Constructing a Publication-Level Classification System of Science. *Journal of the Association for Information Science and Technology*, 63(12), 2378–2392.
- Wang, Ch-Ch.; Ho, Y-S. (2016). Research Trend of Metal-Organic Frameworks: A Bibliometric Analysis. *Scientometrics*, 109(1), 481–513.
- Wang, X.; Wang, Z.; Xu, S. (2013). Tracing Scientists' Research Trends Realtimely. *Scientometrics*, 95(2), 717–729.

- Wierchoń, S.T.; Kłopotek, M.A. (2015). *Algorithms of cluster analysis*. Warszawa: Institute of Computer Science, Polish Academy of Sciences.
- Wilson, R.J. (2012). *Wprowadzenie do teorii grafów*. Warszawa: PWN.
- Zheng, T.; Wang, J.; Wang, Q.; Nie, Ch.; Shi, Z.; Wang, X.; Gao, Z. (2016). A Bibliometric Analysis of Micro/Nano-Bubble Related Research: Current Trends, Present Application, and Future Prospects. *Scientometrics*, 109(1), 53–71.
- Zitt, M.; Bassecoulard, E. (1994). Development of a Method for Detection and Trend Analysis of Research Front Built by Lexical or Co-Citation Analysis. *Scientometrics*, 30(1), 333–351.
-

Bibliometric Methods to Foresee and Assess the Development of Scientific Disciplines. Literature Analysis. Part 1: Trailblazing Publications, Bibliographic Coupling Method, Co-citation Analysis and Co-word Methods

Abstract

Purpose/Thesis: The aim of the article is to provide an evaluative review of a literature on the quantitative methods of bibliometric forecasting and assessing the current state of scientific disciplines development. By “evaluative” review it is understood that the author discussed and pinpointed the drawbacks of the literature-proposed ways to assess and predict the development of scientific disciplines.

Approach/Methods: The method used by the author is the analysis of the literature concerning the problems under consideration. The author also proposed a typology of methods applied in the bibliometric predictions area and emphasized the flaws of some previously used methods. Moreover, the possibilities of practical applications of foresight exercises were shown. Finally, the conclusions were drawn with regard to the methods oriented towards the assessment of the rate and direction of scientific disciplines development.

Results and conclusions: Taking into consideration the fact that this article is first part of a larger whole comprising the analysis of the literature in question, the conclusions drawn within the first part have only a tentative character. It was emphasized that the methods proposed in the trailblazing publications are still vital and fulfill a leading role in bibliometrics. Second, it was indicated that the drawbacks accompanying the majority of mentioned methods are an integral part of the evolution of the research mainstream which is being discussed. Third, it is necessary to note that there is a strong correlation between bibliometric predictive methods and an algorithmic means of clustering groups of publications into topic consistent subgroups. These subgroups are usually considered to represent disciplinary subspecialties or, in other words, disciplinary problem areas.

Originality/Value: The originality and value of the article lie in an effort to demonstrate the scope and a high degree of progress of the studies on the possibilities of prediction and assessment of the development of scientific disciplines worldwide. In the Polish scientific environment such studies are relatively rare and it seems that the review of the predictive bibliometric methodology could contribute to the stimulation of interest in the research directed towards the forecasting of future scientific trends. It is also important to stress the disadvantages of particular methods of bibliometric forecasting as in this context they create a substantial amount of space destined for possible improvements and modifications of existing methods. In the field of bibliometrics there is a need for such refinements due to the fact that currently there is no ideal method which could be described as “the last word” on the technique of anticipating future research directions.

Keywords

Bibliometrics. Predictions. Development of scientific disciplines. Quantitative methods in information science

ŁUKASZ OPALIŃSKI uzyskał tytuł magistra w zakresie nauki o języku i komunikacji (jako specjalności wyodrębnionej z kierunku filozofia) na Uniwersytecie Marii Curie-Skłodowskiej w Lublinie w 2005 r. Ukończył studium podyplomowe *Technologie Informacyjne i Bibliotekoznawstwo* w Wyższej Szkole Informatyki i Zarządzania z siedzibą w Rzeszowie w 2009 r. oraz studium podyplomowe *Bibliotekarz Dziedzinowy* w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego w 2012 r. Pracuje w Oddziale Informacji Naukowej Biblioteki Politechniki Rzeszowskiej na stanowisku bibliotekarza. Najważniejsze publikacje: Opaliński, Ł. (2012). Niealfabetyczne systemy piśmiennicze a komunikacja językowa w Internecie [online]. *Podkarpackie Studia Biblioteczne*, 1/2012, 1–19, [25. 02. 2017], <http://repozytorium.ur.edu.pl/handle/item/180>; Opaliński, Ł. (2013). Wybrane aspekty metodologii badań cyklu życiowego publikacji naukowych. *Przegląd Biblioteczny*, 81(2), 152–171; Opaliński, Ł., Jaromin, M., Wikiera, J. (2015). Problem stabilności zachowań naukowców w zakresie cytowań w kontekście metodologii badań starzenia się publikacji naukowych i możliwość jego ujęcia ilościowego. *Zagadnienia Informacji Naukowej – Studia Informacyjne*, 53(2/106), 65–83.

Kontakt z autorem:

lopa@prz.edu.pl

Oddział Informacji Naukowej Biblioteki Politechniki Rzeszowskiej

al. Powstańców Warszawy 12

35–959 Rzeszów