

---

# Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych.

## Analiza piśmiennictwa.

### Część 2. Badania porównawcze, hybrydowe, statystyczne, analizy dokumentów patentowych, ścieżek rozwoju dyscyplin oraz pozostałe oryginalne podejścia metodologiczne

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:** Głównym wnioskiem wypływającym z analizy literatury jest bardzo duże zróżnicowanie ilościowych metod przewidywania rozwoju dyscyplin, a także fakt, że często są one stosowane we wzajemnej izolacji. Po drugie, niemal każde z proponowanych podejść do omawianego problemu posiada różnego rodzaju wady, które niekorzystnie wpływają na uzyskiwane w jego ramach wyniki. Po trzecie, wydaje się, że szczególnie potencjał prognostyczny tkwi w metodzie statystycznej analizy szeregów czasowych, chociaż w praktyce badawczej metoda ta była wykorzystywana stosunkowo rzadko. Po czwarte, odnotowano wysoki stopień interdyscyplinarności metod bibliometrycznych, które często są modyfikacjami lub zapożyczeniami metod typowych dla matematyki, informatyki, statystyki matematycznej i ekonometrii. Ostatnim wnioskiem jest stwierdzenie, że prognozowanie odbywa się najczęściej przy użyciu specjalnie sprofilowanej analizy cytowań lub analizy liczb publikacji, które ogłoszono w danej dyscyplinie w różnych latach. U podstaw metodologii opartej na analizie cytowań leży konieczność zaakceptowania funkcjonującego w socjologii wiedzy naukowej tzw. paradygmatu normatywnego, jako epistemologicznego uzasadnienia dla badania rozkładów cytowań. Metodologia oparta na ocenie liczb publikacji dyscyplinarnych jest z kolei związana z toczącą się w informatologii dyskusją na temat tematycznej relewancji publikacji naukowych.

**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 przestrzeżeń 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**

Niniejszy artykuł stanowi drugą część i zarazem kontynuację analizy piśmiennictwa skupiającego się wokół problemu bibliometrycznego prognozowania i oceny bieżącego stanu rozwoju dyscyplin naukowych. W części pierwszej, omówiono publikacje pionierów tego kierunku badań oraz postarano się przedstawić, w jaki sposób analizy te „dojrzały”, aż do osiągnięcia kształtu, w jakim są one najczęściej prowadzone obecnie. W drugiej części, położono z kolei nacisk na metody alternatywne wobec metod tradycyjnych, na badania prowadzone w ujęciu porównawczym oraz na analizy głównych ścieżek rozwoju dyscyplin naukowych, którym nadano we współczesnym ujęciu charakter prognostyczny, jakiego brak znamionował wczesną postać tej metody bibliometrycznej.

Warto w tym miejscu dodać, że potrzeba realizacji badań porównawczych wynika przede wszystkim z braku szerszego konsensusu co do tego, która z metod tradycyjnych daje w efekcie jej zastosowania wyniki odpowiadające najbardziej faktycznemu stanowi rzeczy. Inaczej mówiąc, motywem podejmowania badań komparatywnych jest niepewność co do tego, która z metod podstawowych (tj. metoda powiązań bibliograficznych, metoda współcycowań i metoda współwystępowania terminów naukowych) szybciej, dokładniej i pełniej identyfikuje fronty badawcze nadające bieg rozwojowi poszczególnych, wąskich obszarów problemowych. Przyczyn tego zjawiska można upatrywać najprawdopodobniej w fakcie, że wiele badań prognostycznych prowadzonych jest we wzajemnej izolacji, w oparciu o zróżnicowany materiał empiryczny, jak również w sytuacji braku obiektywnego, uniwersalnego punktu odniesienia dla wyników i wniosków wyciąganych w ramach tych badań. Większość bowiem tego, co możemy powiedzieć o rzeczywistych, istniejących realnie prądach intelektualnych kształtujących obraz współczesnej nauki pochodzi właśnie ze wspomnianych badań bibliometrycznych. W tym kontekście podkreśla się często m.in. subiektywność i niepewność ocen eksperckich, metody delfickiej, historycznych analogii czy ogólniej – metod jakościowych, które mogłyby ewentualnie stanowić alternatywne źródło wiedzy o stanie rozwoju nauki i jej perspektywach, a zarazem wspomniany punkt odniesienia dla wyników uzyskiwanych poprzez wykorzystywanie metod ilościowych (zob. np.: Daim et al., 2006, 981–985, 999; Jun et al., 2012). Wskazuje się też niekiedy na potrzebę integracji obydwu typów podejść metodologicznych do problemu prognozowania rozwoju nauki i technologii i podkreśla zarazem istniejący wciąż brak jedności w tym zakresie (zob. np. Haegeman et al., 2013).

Cel drugiej części pracy pozostaje taki sam co w części pierwszej. Jest nim mianowicie dokonanie krytycznego przeglądu literatury naukowej dotyczącej metod ilościowych, stosowanych w obszarze bibliometrycznego prognozowania rozwoju dyscyplin naukowych, wraz

ze wskazaniem mankamentów poszczególnych podejść wykorzystywanych przez autorów zajmujących się tego rodzaju zagadnieniami. Niezmieniona jest również metoda badań, czyli analiza piśmiennictwa i wnioskowanie w oparciu o przedstawiane w niej główne tezy. Niniejszą, drugą część artykułu rozpoczyna przegląd badań porównawczych, w którym nawiązuje się bezpośrednio do wspomnianych wcześniej i wyjaśnionych szczegółowo w części pierwszej metod klasycznych (podstawowych).

## 2. Badania porównawcze

Kolejny typ badań to prace, w których bezpośrednio porównano ze sobą efektywność różnych metod grupowania dokumentów z wykorzystaniem omówionych wcześniej miar podobieństwa publikacji, tj. relacji współcytowania, relacji powiązania bibliograficznego oraz relacji cytowań bezpośrednich. Bo Jarneving (2005) porównał ze sobą pierwszą i drugą z wymienionych metod, w zakresie zdolności do odwzorowywania frontu badawczego. Jarneving porównywał ze sobą profile tematyczne dwóch grup publikacji. Pierwszą z nich uzyskał za pomocą standardowej techniki generowania powiązań bibliograficznych, uzupełnioną o normalizację siły powiązania ze względu na różnice w liczbie opisów występujących w bibliografiach załącznikowych (Jarneving, 2005, 250). Drugą grupę stanowił z kolei nie sam zbiór publikacji powiązanych relacją współcytowania (ang. *co-citation cluster*), ale zbiór publikacji powołujących się (cytujących), na którykolwiek element tego zbioru prac powiązanych współcytowaniami. Dzięki temu rozróżnieniu (tj. klastr publikacji współcytowanych versus publikacje cytujące ten klastrowy) można było porównać publikacje wydane w podobnym przedziale czasu (Jarneving, 2005, 246–247). Bez wprowadzenia tego rozróżnienia, które dostrzegł też m.in. Olle Persson w 1994 r., pomiędzy publikacjami powiązanymi bibliograficznie, a publikacjami powiązanymi relacją współcytowania istniałaby czasowa luka. Powiązania bibliograficzne można bowiem ustalić nawet dla dokumentów wydanych w roku bieżącym (np. 2016), podczas gdy utworzenie się relacji współcytowania wymaga czasu potrzebnego na akumulację dostatecznej liczby cytowań pojawiających się pewien czas po wydaniu prac, które te cytowania zyskują i dopiero potem są ze sobą wiązane współcytowaniem. Jako punkt odniesienia dla tego porównania przyjęto, jak już powiedziano, profil tematyczny uzyskanych klastrów. Określono go za pomocą słów zawartych w tytułach zgrupowanych publikacji. Słowa te poddano normalizacji leksykalnej (tj. usunięto „słowa stopujące”, zastosowano kontrolę synonimów i homonimów oraz normalizację ze względu na liczbę słów zawartych w całym klastrze itd.) (Jarneving, 2005, 252). Wyniki porównania okazały się jednak niekonkluzywne. Jarneving stwierdził, że sugerują one wprawdzie, iż ma się do czynienia z dwoma różnymi frontami badawczymi (pod względem charakterystyk treściowych). Nie można jednak było ostatecznie potwierdzić tego wniosku z uwagi na problem w ścisłym zdefiniowaniu poziomu, na którym dwa tak złożone obiekty jak klastry można określić jako podobne. W związku z tym Jarneving wskazał na potrzebę dalszych analiz porównawczych w warunkach kontrolowanych w wyższym stopniu, niż miało to miejsce w przypadku jego własnego badania<sup>1</sup> (Jarneving, 2005, 260–262).

<sup>1</sup> Kontrola ta mogłaby polegać np. na zastosowaniu ściśle określonego, dokładnie tego samego zbioru cytowanych pozycji bibliograficznych w przypadku konstruowania porównywanych obiektów, tj. klastrów prac powiązanych relacją współcytowania i relacją powiązania bibliograficznego (Jarneving, 2005, 262).

Naoki Shibata i in. (2009) badali w ujęciu porównawczym charakterystyki siatek cytowań uzyskanych za pomocą metod powiązań bibliograficznych, współcytowań i cytowań bezpośrednich pod kątem tego, które z tych podejść daje najlepsze rezultaty w zakresie detekcji najaktualniejszych frontów badawczych. Badanie przeprowadzono w ramach trzech specjalności usytuowanych na pograniczu fizyki i inżynierii materiałowej (Shibata et al., 2009, 572). Wskazane trzy metody porównano pod względem trzech aspektów wytworzonych przez nie sieci cytowań: widoczności frontu (ang. *visibility*) mierzonej rozmiarami wykrytych klastrów, szybkości przesuwania się frontu (ang. *speed*) mierzonej średnim wiekiem publikacji w klastrze oraz relewancją/spójnością tematyczną frontu (ang. *topological relevance*) mierzoną siłą wzajemnych powiązań między elementami klastrów, a inaczej mówiąc tzw. gęstością klastrów (Shibata et al., 2009, 572, 574). Zdaniem Shibaty i in. najefektywniejsza jest ta metoda, która wykryłaby największe i najbardziej spójne klastry na jak najwcześniejszym etapie ich życia (Shibata et al., 2009, 573). Zastosowaną metodą wydzielenia klastrów była metoda tzw. grupowania topologicznego (ang. *topological clustering*) Marka Newmana (2004). Algorytm Newmana opierał się na takim dzieleniu wyjściowego zbioru elementów (publikacji) na mniejsze podzbiory, aby uzyskać strukturę sieciową, w której podzbiory są złożone z elementów powiązanych ze sobą ściśle (podzbiory mają dużą gęstość wewnętrzną), a powiązania pomiędzy elementami należącymi do różnych podzbiorów są dużo słabsze. Miarą tego, jak gęsty jest klaster jest tu tzw. modularność (ang. *modularity*) oznaczana symbolem „Q”. Określa ona różnicę między gęstością klastrów wygenerowanych przez algorytm, a gęstością przypadkową, tj. taką, której można byłoby oczekiwać przy całkowicie przypadkowym rozłożeniu się połączeń między wszystkimi elementami zbioru wyjściowego. Algorytm tak dzieli zbiór wyjściowy, aby zoptymalizować wartość „Q”, czyli wybrać najlepszy spośród wszystkich możliwych podziałów zbioru wyjściowego (Newman, 2004, 1–2; Shibata et al., 2008, 763; Shibata et al., 2009, 573). W wyniku porównania ustalono, że najlepszą metodą wykrywania frontów było cytowanie bezpośrednie. Metoda ta okazała się zdolna do najwcześniejszego wykrycia największych i najspójniejszych nowo powstałych klastrów i stwarzała zarazem najmniejsze ryzyko „przeoczenia” frontu (nieumyślnego wykluczenia go z klastra) (Shibata et al., 2009, 575, 577, 578–579). Najgorzej wypadła metoda współcytowań, głównie ze względu na istnienie opóźnienia czasowego koniecznego do akumulacji cytowań wokół publikacji (Shibata et al., 2009, 577). Gęstości klastrów wytworzonych w ramach tej metody były ponadto mniejsze, stwarzały wrażenie bardziej przypadkowych. Podobnie wyglądała spójność klastrów uformowanych przy użyciu metody powiązań bibliograficznych. Przewagą tej ostatniej metody nad techniką współcytowań była natomiast wrażliwość na czas powstawania połączeń w sieci – w przypadku powiązań bibliograficznych nie występuje bowiem opóźnienie czasowe. Wraz z powiększaniem się (ewolucją) dyscypliny, przyłączaniem do niej kolejnych publikacji, metoda ta dawała jednak w wyniku coraz bardziej przypadkowe (mało spójne) klastry (Shibata et al., 2009, 575–579).

Badanie analogiczne do Shibaty i in. przeprowadzili rok później Kevin Boyack i Richard Klavans (2010). Główne różnice w stosunku do badania Shibaty i in. to włączenie w obszar zainteresowania dodatkowego wariantu metody współcytowań (tzw. *co-citation analysis*) oraz wykorzystanie innego kryterium porównawczego. Pierwszym kryterium była – tak jak u Shibaty i in. – koherencja tekstualna uzyskanych klastrów, podczas gdy drugim była niestosowana dotąd, oryginalna propozycja Boyacka i Klavansa oparta na zamieszczanych

w artykułach informacjach na temat realizacji założeń (zadań) określonych grantów badawczych<sup>2</sup> (ang. *grant-to-article linkages*) (Boyack & Klavans, 2010, 2390, 2399). Wspomniany dodatkowy wariant metody współcytowań polegał na próbie usunięcia problemu opóźnienia czasowego poprzez dołączanie do wyodrębnionych z użyciem standardowego współcytowania klastrow artykułów bieżących (tj. artykułów spoza samych klastrow, które jednak cytują pozycje należące do klastrow). Decyzję podejmowano na podstawie tego, które z artykułów składających się na dane klastry były cytowane (przywoływane w bibliografiach załącznikowych) przez artykuły bieżące. Zabieg ten stanowił więc swego rodzaju złożenie metody powiązań bibliograficznych z metodą współcytowań (Boyack & Klavans, 2010, 2391, 2397). Uzyskane wyniki wskazały na następujące zależności:

- (1) Koherencja tekstualna klastrow była najwyższa dla metody złożonej, następnie kolejno spadała dla metod powiązań bibliograficznych, współcytowań i cytowań bezpośrednich (Boyack & Klavans, 2010, 2400).
- (2) Przewaga metody złożonej nad pozostałymi uwidoczniła się także w przypadku drugiego kryterium, chociaż metoda współcytowań osiągnęła w tym aspekcie niemal równie dobry wynik. Obie one znacznie przewyższyły w tym względzie pozostałe dwie metody (Boyack & Klavans, 2010, 2401).
- (3) Metoda złożona była ogólnie najlepsza pod każdym względem za wyjątkiem tzw. zasięgu (ang. *coverage* – wskaźnika procentowego określającego jaki odsetek zbioru wyjściowego został w ogóle zakwalifikowany do jakiegokolwiek klastra). Zasięg był o niecałe 2% wyższy w przypadku metody współcytowań.
- (4) Na drugim miejscu znalazła się metoda powiązań bibliograficznych – cechował ją niemal równie wysoki zasięg oraz precyzja (mierzoną metodą połączeń publikacji z grantami) w zakresie zdolności reprezentowania frontu badawczego.
- (5) Metoda współcytowań uplasowała się na miejscu trzecim, a metoda cytowań bezpośrednich okazała się najgorsza (Boyack & Klavans, 2010, 2402).

Mu-Hsuan Huang i Chia-Ping Chang (2015) otrzymali rezultat przypominający wyniki Boyacka i Klavansa w zakresie porównania skuteczności metody współcytowań z metodą powiązań bibliograficznych. Motywem podjęcia badania przez tych autorów był zaznaczony przez nich, istniejący wciąż brak jednomyślności w środowisku naukowym co do tego, która z metod dokładniej i szybciej identyfikuje aktualne i przyszłe trendy rozwojowe w obrębie dyscyplin naukowych (Huang & Chang, 2015, 2043, 2049). Analiza porównawcza Huang i Chang objęła wysoko cytowane publikacje (20% prac najwyżej cytowanych w każdym kolejnym roku analizy) z lat 2000–2004 i 2005–2009, należące do pogranicza obszarów elektroniki i materiałoznawstwa (Huang & Chang, 2015, 2044). Minimalnym progiem siły współcytowania oraz siły powiązania bibliograficznego było pięć współcytowań i pięć wspólnych pozycji bibliograficznych (Huang & Chang, 2015, 2045). Po zastosowaniu operacji grupowania hierarchicznego zestawiono ze sobą jej wyniki, które przybrały postać zespołu klastrow jako odpowiedników frontów badawczych wraz z nadanymi im nazwami (przynależnością do pewnych kategorii tematycznych) i charakterystykami zmieniającymi

<sup>2</sup> Dane do przeprowadzenia tego porównania składały się z dopasowanych do siebie numerów grantów i identyfikatorów artykułów z bazy MEDLINE. Założeniem tego kryterium było uznanie, że artykuły, które realizowały te same granty (wykorzystywały środki finansowe z tych samych grantów), powinny skupić się w tych samych klastrach (Boyack & Klavans, 2010, 2399).



się dynamicznie pomiędzy pierwszym a drugim podokresem badania. Rozpatrywane charakterystyki to rozmiar frontu, zmiana tego rozmiaru (rozrost lub zmniejszanie się) lub nagle pojawianie się albo też całkowity zanik frontu (Huang & Chang, 2015, 2049–2051). Ostatecznie Huang i Chang wyciągnęli wniosek, według którego metoda powiązań bibliograficznych była w stanie wykryć większą liczbę frontów badawczych na wcześniejszym etapie ich kształtowania się niż metoda współcystowań, co wskazało jednoznacznie na jej przewagę nad metodą współcystowań (Huang & Chang, 2015, 2052, 2054).

### 3. Badania hybrydowe (mieszane)

Badaniami „hybrydowymi” lub mieszanymi nazwano te, w ramach których połączono poszczególne (dwie lub więcej) metody wykrywania formowania się nowych lub przekształcania się pewnych klastrów publikacji, jako struktur wytyczających granice frontów badawczych.

Peter Van Den Besselaar i Gaston Heimeriks (2006) zauważyli np., że naukowcy jednocześnie posługują się pewnymi terminami celem opisu przedmiotu własnych badań oraz odnoszą się do literatury przedmiotu (cytując ją), celem wskazania tradycji badawczej, do której przynależy ich własna praca. Pojawiająca się w publikacjach terminologia nabiera w związku z tym specyficznego znaczenia w kontekście zestawu cytowanych pozycji bibliograficznych. Wychodząc z tego założenia, zaproponowano kombinację metody *co-word* z metodą powiązań bibliograficznych, zgodnie z którą wskaźnikiem podobieństwa dwóch publikacji jest liczba współwystąpień określonych terminów oraz, jednocześnie, pozycji przywoływanych w bibliografiach załącznikowych (ang. *word-reference combinations*) (Van Den Besselaar & Heimeriks, 2006, 381, 384). Zdaniem autorów ich propozycja wyeliminowała problem arbitralności w ustalaniu progu minimalnej siły związku pomiędzy dokumentami, która dopuszcza przyłączenie ich do klastra. Uwzględnia się tutaj bowiem wszystkie dokumenty, które charakteryzują się przynajmniej jednym współwystąpieniem zarówno terminu/frazy, jak i przypisu bibliograficznego odsyłającego do tej samej pracy (Van Den Besselaar & Heimeriks, 2006, 391).

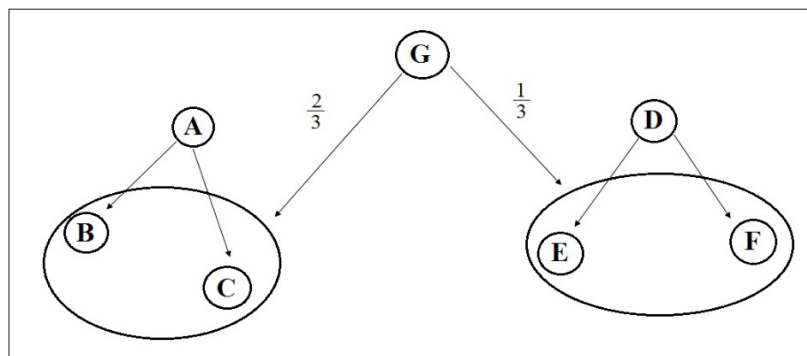
Podobnie, połączenie metody tekstualnej z powiązaniem bibliograficznymi zasugerował Wolfgang Glänzel (2012). Jego zdaniem metoda tekstualna powinna być stosowana jako technika nadawania nazw (technika terminologicznego kodowania) klastrów wyodrębnionym w oparciu o metodę powiązań bibliograficznych oraz jako równorzędna miara siły związków między dokumentami tworzącymi klastry. Jednym z sygnałów wyłonienia się nowej specjalności jest w tym ujęciu, np. przesunięcie leksykalnego „punktu ciężkości” klastra (ang. *topic shift*) w kierunku innej niż dotychczasowa problematyki, co uwidacznia się właśnie w sferze tekstualnej (Glänzel, 2012, 195–196). Ponadto połączenie obu metod daje sposobność dostosowywania badań do specyfiki badanej dyscypliny. Przykładowo, w naukach społecznych i humanistycznych, w których cytowalność artykułów z czasopism jest na ogół niższa niż w naukach przyrodniczych i technicznych, można wprowadzić „ważenie” elementów bibliograficznych i tekstualnych. Element tekstualny mógłby posiadać większą wagę (pierwszeństwo) w obrębie nauk społecznych i humanistycznych. Co więcej, podobieństwo tekstualne może w takich przypadkach zastąpić „brakujące” powiązania o naturze bibliograficznej (tj. powiązania bibliograficzne w sensie teorii Kesslera) (Glänzel, 2012, 195–197, 200).

To samo hybrydowe kryterium podobieństwa publikacji wykorzystali Glänzel i Bart Thijs (2012). Było ono podstawą do ustalenia siły połączeń pomiędzy klastrami publikacji w jednym okresie a tzw. publikacjami kluczowymi (ang. *core documents*) w innym, późniejszym okresie rozpatrywanym w ramach analizy dynamiki i transformacji struktur klastrowych. Publikacje kluczowe w obrębie jednego klastra są publikacjami najsilniej ze sobą związanymi (najbliższymi sobie wg ustalonego kryterium podobieństwa), tj. dzielą ze sobą najwięcej jednostek bibliograficznych. Inaczej mówiąc, muszą one posiadać pewną wymaganą liczbę połączeń o określonej sile i powinny stanowić nie więcej niż 1% całego klastra. Można uważać je za reprezentantów klastra, a celem ich identyfikacji jest redukcja „szumu” (ang. *noise*), który wywołują publikacje powiązane luźniej z pracami centralnymi w obrębie klastra, leżące (na mapie) dalej od centrum i będące mniej relewantnymi dla tematyki przewodniej klastra (Glänzel & Thijs, 2012, 402–403; Glänzel, 2012, 196). Ustalenie charakteru związków między klastrami uformowanymi w pewnym czasie a dokumentami kluczowymi w innym czasie (w rozłącznych okresach obserwacji), do czego niezbędne jest wspomniane kryterium, było w ujęciu Glänzela i Thijsa głównym problemem w poszukiwaniu nowo powstających specjalności w dyscyplinie. Związki te polegają na wyłanianiu się nowych klastrow, zmianach rozmiaru klastrow, łączeniu się klastrow lub ich rozpadzie (Glänzel & Thijs, 2012, 401, 405–410). Rozpoznanie tych zależności pozwala wskazać, które obszary problemowe są aktualnie w fazie najintensywniejszego rozwoju i które z nich dyktują kierunki przesuwania się granic wiedzy w dyscyplinie jako całości.

Boyack i Klavans (2014) użyli metody hybrydowej nieco innego rodzaju w ramach przedsięwzięcia obejmującego utworzenie dynamicznej i globalnej mapy nauki za lata 1996–2011 na podstawie danych z bazy Scopus. Jak stwierdzili autorzy, „mapowanie” nauki jest dwuetapowe, tj. składa się z klasyfikacji (grupowania, podziału wyjściowego zbioru publikacji na klastry) i dalszej wizualizacji sieci cytowań pomiędzy nimi (Boyack & Klavans, 2014, 671). Operację grupowania oparto na złożeniu techniki generowania klastrow publikacji współcytowanych z techniką połączeń bibliograficznych. Ponieważ klastry publikacji współcytowanych nie zawierają publikacji najnowszych (leżą w przeszłości w stosunku do prac wydanych w roku bieżącym – w tym przypadku w 2011 r.), podjęto decyzję o uzupełnieniu zawartości tych klastrow o publikacje najnowsze. Jeżeli artykuł wydany w 2011 r. miałby np. 10 pozycji w bibliografii załącznikowej, które występowałyby zarazem w innym z klastrow publikacji współcytowanych, zostałby on przyłączony do tego właśnie klastra. Dopuszczalne jest tu także „przydzielanie ułamkowe” (ang. *fractional assignment*). Gdyby artykuł ten miał siedem pozycji bibliograficznych wspólnych z jednym klastrem, a trzy pozycje wspólne z drugim, zostałby on przydzielony w 70% do tego pierwszego, a w 30% do drugiego (Boyack & Klavans, 2014, 673–674, 677). Jest to ta sama operacja (złożenie metod), którą autorzy rozpatrywali w 2010 r. pod nazwą *co-citation analysis* i którą przedstawili poglądowo na rysunku (zob. Rys. 1).

Rysunek 1 przedstawia przypadek, w którym mamy do czynienia z sytuacją, w której artykuł „A” powołuje się na artykuły „B” i „C”, a artykuł „D” cytuje artykuły „E” i „F”. Publikacje cytowane przez artykuł „A” tworzą więc dwuelementowy klaster publikacji współcytowanych {B, C}, a publikacje cytowane przez artykuł „D” klaster {E, F}. Artykuły „A”, „G” i „D” wydano w 2011 r., podczas gdy artykuły „B”, „C”, „E” i „F” wcześniej. Artykuł „G” ma przykładowo tylko trzy pozycje w bibliografii załącznikowej: „B”, „C” i „E”. Stąd, może on zostać przypisany częściowo do klastra {B, C} i częściowo do klastra {E, F}, chociaż

pierwotnie nie wchodził on w skład żadnego z nich ponieważ nie był z żadnym z nich współcytowany (żadna publikacja nie cytowała jednocześnie artykułu „G” i np. artykułu „B”).



Rys. 1. Ułamkowe kwalifikowanie publikacji wydanej w roku bieżącym do klastrów publikacji utworzonych w oparciu o metodę współcytowania;  
źródło: oprac. własne na podstawie Boyack & Klavans, 2010, 2390

Druga modyfikacja metodologiczna Boyacka i Klavansa pojawiła się na etapie wizualizowania związków zachodzących między wygenerowanymi już klastrami, czyli na etapie rysowania map danych dyscyplin. Na tym etapie konieczne jest posłużenie się jakąś miarą wzajemnego podobieństwa (bliskości) klastrów (ang. *cluster-cluster similarity*) (Boyack & Klavans, 2014, 675). Stosowanie różnych miar doprowadza w konsekwencji do odmiennych wizualizacji. Boyack i Klavans dążyli do takiego układu mapy (ang. *cluster layout*), w którym widniałyby luźno ze sobą powiązane „wyspy” (ang. *sets of islands connected by pathways*) (Boyack & Klavans, 2014, 675–676). Jest to standardowy wymóg stawiany przed wynikami operacji podziału (grupowania) zbiorów: podział „dobry” to podział, który zapewnia uformowanie podzbiorów gęstych (o silnych połączeniach wewnętrznych) i zarazem dobrze wyodrębnionych od pozostałych (tj. takich, które posiadają znacznie słabsze połączenia „zewnętrzne” w sensie połączeń z innymi podzbiorami). Kryterium podobieństwa Boyacka i Klavansa miało charakter tekstualny – klastry reprezentowały frazy zawierające fragmenty tytułów i abstraktów publikacji zgrupowanych w ich ramach. Kryterium to miało postać wyrażenia matematycznego, w której pierwszorzędą rolę odgrywała częstość występowania określonych terminów w poszczególnych obiektach (klastrach) (Boyack & Klavans, 2014, 675–676).

Small, Boyack i Klavans (2014) zaproponowali kolejną metodę hybrydową mającą na celu umożliwienie wykrywania nowo powstających obszarów problemowych jeszcze w trakcie ich formowania się. Zadanie to stanowiło w ich ocenie przeciwieństwo analiz retrospektywnych, które, „patrzac wstecz”, były w stanie wskazać specjalności nowo powstające, wyłonione np. pięć lub więcej lat przed momentem realizacji badania (Small et al., 2014, 1452). Istotą metody Smalla i in. było dwukrotne zgrupowanie (podział na klastry) tych samych danych o publikacjach i ich cytowaniach, przy użyciu dwóch różnych kryteriów podobieństwa – relacji współcytowania oraz relacji cytowań bezpośrednich. Uzyskuje się w ten sposób dwa modele (mapy) oparte na innych sposobach organizacji danych i cechujące się inną charakterystyką. Następnie oba modele są na siebie „nakładane” (ang. *intersected*) w celu



zidentyfikowania obszarów (ang. *clusters of activity*), które w obu modelach w najbardziej wyraźny sposób manifestują cechy nowości/oryginalności (ang. *novelty, newness*) i rozrostu (ang. *growth*) (Small et al., 2014, 1452). Cechy te uważane są bowiem powszechnie za nieodłącznie związane z nowo wyłaniającymi się specjalnościami dyscyplinarnymi (Small et al., 2014, 1451). Ilościową techniką porównywania wyodrębnionych klastrów zaobserwowanych po nałożeniu na siebie obu modeli, pod względem osiągniętego przez nie stadium rozwojowego, była tzw. funkcja EP (ang. *Emergence Potential*). Pozwalała ona na „nominowanie” pewnych klastrów jako najlepszych kandydatów do miana załączków nowych specjalności (ang. *the most emergent*) (Small et al., 2014, 1453). Najpoważniejszym problemem w tej metodologii było wykryte opóźnienie czasowe pomiędzy zaistnieniem wydarzenia krytycznego dla uformowania się nowej specjalności (np. odkrycia), a jej wyłonieniem się i możliwością jej detekcji. Opóźnienie to, według szacunku Smalla i in. wynosiło 2.3 roku i było zbyt duże, aby można było wykorzystać w praktyce wiedzę o formowaniu się nowego obszaru (np. podczas podejmowania decyzji o finansowaniu badań) (Small et al., 2014, 1462). Dopiero znaczna minimalizacja tego opóźnienia przekształciłaby metodę z *de facto* opisowej w metodę autentycznie prognostyczną, w związku z czym konieczne jest dalsze poszukiwanie odpowiednich do tego sposobów adaptacji metody (Small et al., 2014, 1463).

Chengliang Liu i Quinchang Gui (2016), badając dyscyplinę geografii transportu pod kątem dokonujących się w niej przemian pomiędzy rokiem 1982 a 2014, połączyli w ramach kompleksowego podejścia technikę „mapowania” współcytowań autorów i współcytowań artykułów i czasopism, technikę analizy współwystąpień słów kluczowych, analizę przyrostu liczby publikacji w czasie wraz z liczbą ich cytowań oraz analizę geoprzestrzenną (ang. *geospatial analysis*). Ten ostatni element miał za zadanie odwzorować międzynarodową sieć współpracy naukowej w dyscyplinie poprzez wskazanie jej centrów (miast, regionów, instytucji badawczych i „niewidzialnych fakultetów”) oraz narodowościowego rozkładu zajmujących się daną tematyką autorów. Z tym zamiarem wykorzystano pakiety oprogramowania komputerowego: *CiteSpace*, *BibExcel* i *ArcGIS*, tworząc odrębne mapy dla poszczególnych, rozpatrywanych aspektów (Liu & Gui, 2016, 161–162, 164–165, 167, 170). Z punktu widzenia niniejszego przeglądu najistotniejszymi zadaniami podjętymi przez Liu i Guiego było identyfikowanie centralnych dla dyscypliny frontów badawczych i ich dynamiki, za pomocą narzędzia *CiteSpace*. Zapewniło ono autorom możliwość detekcji „eksplozji terminów” (ang. *burst terms*) wskazujących główne „trajektorie badawcze” (ang. *research trajectories*), czyli zagadnienia, na których w przeważającej mierze skupiła się uwaga społeczności naukowej. Na powstałej mapie miały one postać punktów połączonych krawędziami o różnej wielkości. Krawędzie reprezentowały relację współwystępowania terminów i siłę (skalę) tej relacji, a rozmiar punktu – częstość wykorzystywania terminów (popularność) w różnych okresach. Terminy, których częstość występowania wzrosła gwałtownie dopiero niedawno, zostały uznane za sygnał pojawiania się nowych kierunków badań w dyscyplinie (ang. *research hotspots/emerging hot issues*) (Liu & Gui, 2016, 173–175). Wykorzystując kolejną funkcję *CiteSpace*, zbadano wzorce współcytowań indywidualnych publikacji i dokonujące się w tym zakresie przemiany w różnych przedziałach czasu. Szczególną uwagę zwrócono na „eksplozje” cytowań pewnych prac (ang. *bursting references*), których tematyka w dużym stopniu wpłynęła na kierunek ekspansji geografii transportu (Liu & Gui, 2016, 175–178).

Czasowy przyrost liczby publikacji o określonej problematyce, obok analiz współcytowań publikacji i współwystępowania terminów, był także przedmiotem badania Pinga Liu

i in. (2016). Materiałem badawczym było 500 najwyżej cytowanych artykułów z zakresu fizyki, które ukazały się w latach 2000–2015. Liu i in. (2016), podobnie jak Liu i Gui (2016), obserwowali nie tylko najnowsze trendy badawczo-rozwojowe, ale też i inne właściwości charakterystyczne dla wybranej specjalności, takie jak kluczowe czasopisma, afiliacje autorów i kraje wiodące prym w zakresie produktywności i siły wpływu na całokształt badań w dyscyplinie (Liu et al., 2016, 1593–1595). Drugim podobieństwem do procesu badawczego Liu i Guiego (2016) było wyodrębnianie (w postaci osobnych map) klastrów prac współcytowanych i klastrów współwystępujących ze sobą słów kluczowych. Te publikacje, które znalazły się w centrum klastra (sieci współcytowań), tj. posiadały najwyższy wskaźnik tzw. „centralności pośredniczącej”<sup>3</sup> (ang. *betweenness centrality*) określono mianem „publikacji rewolucyjnych” (ang. *revolutionary publications*), a więc publikacji nadających kierunek dalszej aktywności społeczności badaczy i symbolizujących kluczowe dla danej specjalności osiągnięcia (Liu et al., 2016, 1598). Zmierzone też średni wiek publikacji wchodzących w skład klastrów, a następnie przeprowadzono podział (grupowanie) zbioru stu najczęściej występujących w parze słów kluczowych (Liu et al., 2016, 1600). W kolejnym etapie badania wytypowano terminy o najwyższym stopniu „centralności” i najwyższej częstości występowania w całym badanym zbiorze oraz klaster terminów o najniższej średniej wieku publikacji, z których pochodziły wyróżnione słowa kluczowe (Liu et al., 2016, 1600–1601).

Jeszcze jedną oryginalną, hybrydową metodę przewidywania przyszłych związków i transformacji w sieciach typu *co-word* przedstawili Nazim Choudhury i Shahadat Uddin (2016). Metoda grupowania współwystępujących, autorskich słów kluczowych (reprezentujących konkretne koncepcje) w podzbiory, została uzupełniona o techniki zaczerpnięte z tzw. statystycznej analizy szeregów czasowych<sup>4</sup> (ang. *time series approach*) (Choudhury & Uddin, 2016, 747–748). Pierwszym krokiem procedury Choudhury’ego i Uddina było wygenerowanie zestawu słów kluczowych dla trzech wybranych obszarów tematycznych i ich normalizacja (usunięcie znaków interpunkcyjnych, znaków diakrytycznych, zamiana liczby mnogiej rzeczowników na pojedynczą, ujednoclenie synonimów itd.). Ze słów kluczowych utworzono następnie mapy ich współwystępowania we wszystkich latach z przedziału 2010–2014 (przedział „T” reprezentujący „fazę próbną” – ang. *training phase*) i w roku 2015 (okres „T+1”) jako okresie testowym (ang. *test phase*). Dla jednego obszaru tematycznego (ze wszystkich trzech wybranych) mapy odwzorowywały relacje pomiędzy terminami dla lat 2008–2012 („T”), a rolę okresu testowego pełniły dane za rok 2013 („T+1”)

<sup>3</sup> „Centralność pośrednicząca” określana także w języku polskim jako „centralność oparta na pośrednictwie” lub „wskaźnik pośrednictwa” jest wskaźnikiem typowym dla analiz grafów i sieci (w tym także sieci cytowań – ang. *citation networks*) i określa, jak wiele najkrótszych ścieżek (połączeń między dwoma dowolnymi wierzchołkami w sieci) przebiega przez wierzchołek, dla którego częstość ta jest wyznaczana. Miernik ten można interpretować jako współczynnik wpływu, który wywiera dany punkt (artykuł) na przepływ informacji między elementami sieci (Shibata et al., 2007, 877; zob. też: Leydesdorff, 2007, 1304–1305).

<sup>4</sup> Analiza szeregów czasowych to jedna z powszechnie stosowanych w naukach ekonomicznych i społecznych metod prognozowania i oceny dynamiki zjawisk masowych. Analizy te polegają na symulacji przyszłego zachowania się wybranych zjawisk w oparciu o znane dane na temat przebiegu tych zjawisk w teraźniejszości i w przeszłości. Przykładowe dane, które są przedmiotem stosowania takich metod to np. ceny lub skala eksportu towarów, kursy walut, liczba ludności w danym regionie lub kraju itd. Szereg czasowy ma postać ciągu wartości liczbowych badanego zjawiska w kolejnych jednostkach czasu (zob. np.: Sobczyk, 2015, 307–356).

(Choudhury & Uddin, 2016, 753–755). Następnie zastosowano kilka matematycznych wskaźników<sup>5</sup> opisujących dynamikę struktury sieciowej w okresie „T”, z podziałem na pięć jednorocznych okresów obserwacji, aby przewidzieć związki (współwystąpienia) pewnych terminów (lub też ich brak) oraz trafność tych przewidywań w okresie „T+1” (Choudhury & Uddin, 2016, 748–751, 755–756). Z wyliczonych wskaźników zbudowano serie czasowe, które reprezentowały strukturę sieci w kolejnych latach i posłużyły za dane wejściowe dla trzech różnych technik prognostycznych: tzw. wygładzania wykładniczego, średniej ruchomej połączonej z procesem autoregresyjnym (ARIMA) oraz błędzenia losowego (Choudhury & Uddin, 2016, 751–752, 772; nt. metody ARIMA zob. np.: Montgomery et al., 2008, 231–286; Shumway & Stoffer, 2011, 83–171; Zeliaś et al., 2013, 234–243; nt. błędzenia losowego zob. np.: Aczel, 2007, 621–622; Klafter & Sokolov, 2015). Model pozwalał na przewidzenie trzech mechanizmów formowania się nowych połączeń między słowami kluczowymi: połączenia między słowami nowymi (tj. istniejącymi tylko w fazie testowej „T+1”), połączenia między słowami nowymi a „starymi” (tj. istniejącymi w okresie „T” oraz w okresie „T+1”) oraz nowe, nieistniejące dotąd połączenia między słowami „starymi” (obecnymi tylko w okresie „T”) (Choudhury & Uddin, 2016, 757, 772–773).

Chong-Chen Wang i Yuh-Shan Ho (2016) badali bieżące trendy obecne w obszarze chemii polimerów wykorzystując w tym celu technikę scharakteryzowaną w podrozdziale siódmym części pierwszej (przy okazji omawiania pracy Zhenga i in. (2016)) w połączeniu z obserwacją przyrostu liczby publikacji w czasie, dat ich wydania oraz przypisanych im słów kluczowych (autorskich i „dodatkowych”, dostępnych w WoS – *Key Words Plus*) (Wang & Ho, 2016, 483, 485–486). Słowa kluczowe zostały dodatkowo uzupełnione przez „słowa wspierające” (ang. *supporting words*), które zasugerowali eksperci zaznajomieni z wybranym obszarem nauki. Analizę słów, zawężono tylko do słów występujących w tytule, abstrakcie lub polu słów kluczowych WoS (ang. *front page filter*) (Wang & Ho, 2016, 483). Konkluzje co do przeważających w rozpatrywanej specjalności tendencji wyciągnięto po skonstruowaniu „klastrów słownictwa” (ang. *word clusters*) i dopełnieniu analizy za pomocą dokładniejszego przeglądu treści publikacji najwyżej cytowanych (od momentu wydania aż do roku 2014), które znalazły się w ramach wyszukanego zbioru (Wang & Ho, 2016, 494–504).

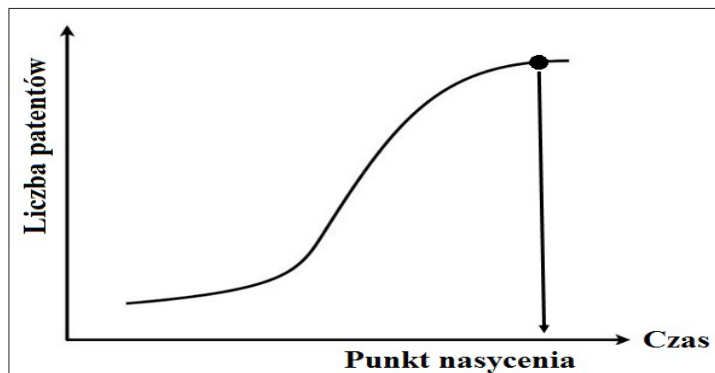
#### 4. Publikacje patentowe

W obrębie tej podgrupy podsumowano przedsięwzięcia, które za materiał badawczy obrały dokumenty patentowe. Często podkreślaną cechą patentów, która decyduje o ich podatności na badania zorientowane prognostycznie jest fakt, iż są to źródła piśmiennicze bezpośrednio odzwierciedlające technologiczny postęp i jego charakter (m. in. tempo czy kierunek). Warunkiem przyznania prawa patentowego jest bowiem nowatorskość (ang. *novelty*) i postępowość (ang. *inventive step*) (Debackere et al., 2002, 213), a same patenty – niejako

<sup>5</sup> Wskaźniki te były dwójakiego rodzaju: lokalne i globalne. Wszystkie one miały charakter topologiczny, tj. reprezentowały strukturalne podobieństwo par wierzchołków sieci (grafu), które zależało wyłącznie od struktury sieci jako całości. Przykładowym wskaźnikiem lokalnym było tu np. podobieństwo mierzone liczbą „wspólnych sąsiadów” dwóch wierzchołków w sieci (ang. *common neighbours*). Szczegółowe omówienie wszystkich wykorzystanych wskaźników wymagałoby jednak przypuszczalnie odrębnego opracowania. Więcej szczegółów nt. konstrukcji wskaźników podali sami Choudhury i Uddin (2016, 748–751).

z definicji – są bezpośrednim świadectwem nowych osiągnięć naukowych, które stymulują dalszy rozwój danej gałęzi techniki (Yoon & Park, 2007, 592; zob. też: Ena et al., 2016, 1025).

W pierwszej kolejności należy w tym kontekście wspomnieć o pracy Tugrula Daima i in. (2006), którzy proponowali zastosowanie połączenia kilku metod badawczych do analizy liczby patentów odnoszących się do pewnych technologii, ogłoszonych w różnym czasie. Metody te można podzielić na ilościowe (bibliometria, dopasowanie funkcji do danych empirycznych i jej dalsza ekstrapolacja) oraz jakościowe (perspektywa organizacyjna i personalna – budowanie scenariuszy w oparciu o opinie ekspertów, modelowanie dynamiki systemów w oparciu o technikę Jaya Forreстера stworzoną w latach 60. XX w., analizy analogii historycznych, analizy rynku, ryzyka i preferencji konsumenckich) (Daim et al., 2006, 981–985, 999). Zakłada się tutaj, że różne metody należy stosować do różnych etapów projektowania, wdrażania i komercjalizacji nowych technologii. Patenty odzwierciedlają etap rozwoju technologii, który następuje po etapie badań podstawowych i przed etapem zastosowań technologii w praktyce i uwidaczniania się jej wpływu na społeczeństwo. Przyrost liczby patentów w czasie można modelować funkcjami o kształcie zbliżonym do litery „S” (m.in. funkcją Gompertza, Fishera-Pry’ego, Voltera-Lotki, Pearla-Reeda) nazywanymi też funkcjami logistycznymi (zob. np.: Kucharavy & De Guio, 2011, 559–561; Price, 1967, 27–28), dzięki którym można przewidzieć ramy czasowe wczesnego etapu rozwoju technologii, wzmożenia prac nad jej dopracowaniem oraz osiągnięcia pewnego maksimum, które wyznacza moment „dojrzałości” technologii i przejścia do kolejnych etapów jej cyklu życiowego (zob. Rys. 2) (Daim et al., 2006, 990–991, 999–1000; zob. też: Lee et al., 2012, 817).



Rys. 2. Przykładowa krzywa logistyczna opisująca hipotetyczne tempo przyrostu liczby patentów w czasie; źródło: oprac. własne na podstawie Daim, T. et al., 2006, 986

Jak zaznaczyli Daim i in., problematycznym aspektem metody jest fakt, że analiza danych bibliometrycznych na temat liczby patentów jest w stanie dostarczyć jedynie danych historycznych, które mogą nie być jeszcze dostępne w przypadku najwcześniejszego etapu cyklu życiowego technologii nowo powstających (Daim et al., 2006, 1009).

Osmo Kuusi i Martin Meyer (2007) także poddali analizie patenty, które rozpatrywali w aspekcie wielu konkurujących ze sobą tzw. technologicznych paradygmatów funkcjonujących w obrębie tej samej specjalności. Paradygmaty te określono też jako „wizje”

(ang. *guiding image*; niem. *leitbild*), które stanowiły ogólne, podzielane przez całą społeczność wartości, oczekiwania, specjalistyczny język i cele badawcze, oferowały długoterminowe metody ich realizacji i dostarczały podstaw dla podobnie zorientowanych prac prowadzonych w jednym kierunku w obrębie odmiennych dyscyplin badawczych. Można je było również przyrównać do koncepcji samoorganizacji nauki funkcjonującej jako interpersonalny „stabilizator” wysiłków poszczególnych naukowców lub ogólniej – dynamiki postępu w nowo powstałej specjalizacji. Bibliometryczna analiza klastrów stosunkowo niedawno opublikowanych patentów (ich rosnącej w czasie liczby, powiązań bibliograficznych pomiędzy nimi oraz wykorzystywanego w nich słownictwa), była w tej perspektywie metodą identyfikacji różnych, obiecujących „wizji” na wczesnym etapie ich kształtowania się (Kuusi & Meyer, 2007, 760–763, 765–767, 774–775).

Inne, oryginalne podejście do bibliometrycznej analizy patentów zaprezentowali Byungun Yoon i Yongtae Park (2007). Podejście to polegało na badaniu terminologicznej morfologii patentów i wykrywaniu nieistniejących jeszcze, ale możliwych połączeń między słowami kluczowymi nadawanymi patentom na podstawie występowania w ich treści słownictwa charakteryzującego się najwyższą częstością (Yoon & Park, 2007, 592). Morfologię („mapę leksyki”) patentu można zademonstrować na przykładzie (zob. Tab. 1).

Tab. 1. Morfologia hipotetycznego zbioru patentów w postaci wyodrębnionych z nich, najczęściej pojawiających się w ich treści słów kluczowych; źródło: oprac. własne na podstawie Yoon & Park, 2007, 593

Typ materiału	Technika wytwarzania	Rodzaj więzi	Forma
Glina	Ekstrakcja	Ciepło	Okrąg
Metal	Wytapianie	Chemikalia	Linia
Tworzywo sztuczne	Prasowanie	Magnetyzm	Sześcián

Morfologię pewnego patentu określa więc szereg terminów, które w powyższym przypadku można streścić jako „chemiczną ekstrakcję metalu do formy cząstek sześciennych”. To znaczy, że patent ten ma więcej słów kluczowych skojarzonych z „metalem” niż z „gliną” czy „tworzywem sztucznym” itd., ale w innych patentach konfiguracja może wyglądać inaczej. Konfiguracje „wolne”, tj. niezrealizowane dotąd w żadnym patencie mogą następnie zostać wyodrębnione i wykorzystane do wskazania prawdopodobnych kierunków dalszego rozwoju specjalności (Yoon & Park, 2007, 593). Uzupełnieniem tej metody może być analiza cytowań patentów o danej morfologii. Jako słabe strony swojej metodologii autorzy wskazali m.in. na ewentualność, że dotychczasowe słownictwo może nie oddawać w pełni kierunków rozwoju radykalnie nowych technologii. Ponadto problematyczne aspekty metody to np. możliwość, że niektóre niezrealizowane połączenia mogą okazać się fizycznie niemożliwe do uzyskania w praktyce, a także niezdolność metody do wykroczenia poza obszar, który obejmują analizowane dokumenty patentowe, jak również istnienie problemu związanego z opóźnieniem czasowym, które wywołuje konieczność oczekiwania na pojawienie się cytowań patentów. Odnotowano również potrzebę dopracowania i usystematyzowania sposobu pozyskiwania najodpowiedniejszych słów kluczowych dla konkretnych patentów (Yoon & Park, 2007, 597). Ssu-Han Chen i in. (2012), wykorzystali standardowe techniki analizy sieci cytowań



oraz algorytm grupujący publikacje patentowe wraz z analizą terminologii, celem śledzenia kierunków i trendów obserwowalnych w przesuwaniu się „frontu technologicznego” (Chen et al., 2012, 1100, 1103). Szczególny nacisk położono na transformowanie się struktur klastrów (ang. *community transition*) w kolejnych, zachodzących na siebie okresach obserwacji. Transformacje te miały postać różnego rodzaju „przejsć” (ang. *transitions*) z jednego okresu w drugi („narodziny”, „śmierć”, rozgałęzienie lub łączenie się klastrów), o których można było mówić kiedy dwa klastry w różnych okresach obserwacji zawierały wiele tych samych dokumentów patentowych. Zaznaczono ponadto potrzebę dostosowania długości okresu obserwacji (tj. szerokości przedziałów czasowych, na które dzieli się cały badany okres) do specyfiki danej specjalności. Zaproponowano w tym celu tzw. miarę TCT (ang. *technology cycle time*), która była różnicą pomiędzy datami wydania patentów cytujących i patentów cytowanych należących do pewnego klastra (Chen et al., 2012, 1102–1104). Część analizy stanowiła także hierarchizacja słów kluczowych opisujących klastry, tj. nadawanie im wag, które pokazywały jak dużą „siłę” posiadał dany termin w obrębie swojego klastra, jak dobrze odpowiadał on jego ogólnej tematyce i czy mógł on zostać przypisany także i innym klastram (Chen et al., 2012, 1199–1101, 1103).

Péter Érdi i in. (2013) wysunęli propozycję kolejnej, alternatywnej wobec pozostałych metody badania dokumentów patentowych pod kątem prognozowania nadchodzących trendów w rozwoju technologii. Rolę predyktora w metodzie autorów spełniał tzw. wektor cytowań patentów (ang. *citation vector*), który można było zdefiniować dla każdego patentu (Érdi et al., 2013, 226). Konstrukcja wektora cytowań opierała się na cytowaniach bezpośrednich. Dla każdego patentu wyznaczono sumę wszystkich pozyskanych cytowań<sup>6</sup>, które pochodziły od patentów należących do 36 różnych kategorii tematycznych ustalonych przez *US Patent & Trademark Office* (USP&TO). Cytowania „przychodzące” były dodatkowo ważone ze względu na to ile cytowań „rozesłał” patent cytujący (tj. jak obszerną posiadał bibliografię załącznikową). Waga była tym większa, im mniejsza była liczba takich cytowań „rozesłanych” (Érdi et al., 2013, 232). Wektor składał się zatem z 36 liczb pokazujących, jak wiele cytowań otrzymał dany patent od patentów należących do każdej kolejnej kategorii tematycznej USP&TO. Komponenty wektora cytowań patentu demonstrowały więc, jak duży wpływ wywarł on na inne (niż swoją własną) kategorie tematyczne, czyli jak wpływowo okazał się patent poza swoją własną kategorią (Érdi et al., 2013, 233–234). Obserwowano następnie jak relacje te zmieniały się w czasie, co pozwalało wytypować zbiory patentów podobnych pod wskazanym wyżej względem, których siła wpływu poza własną kategorią tematyczną wzrastała w czasie. Rozumienie wynalazczości prezentowane w ramach pracy Érdiego i in. polegało więc na jej postrzeganiu jako procesu rekombinacji technologii istniejących już wcześniej (Érdi et al., 2013, 238). Niedostatki tej metodologii to z kolei m.in. istnienie opóźnienia czasowego pomiędzy faktycznymi narodzinami nowej technologii a jej zaistnieniem w patentowych bazach danych i wzrostem jej cytawalności, fakt manualnego kategoryzowania patentów przez USP&TO, co narażało metodę na zarzut pewnej subiektywności oraz niejednorodność różnych obszarów technologii pod względem liczby ogłoszonych patentów, ich szans na pozyskanie cytowań oraz średniej liczby cytowań na jeden patent. Ostatnia z wymienionych cech miała wpływ na precyzję i zasadność ważenia cytowań zastosowanego przez autorów (Érdi et al., 2013, 238).

<sup>6</sup> Jeżeli patent nie otrzymał żadnych cytowań, jego wektor jest równy zeru.

Podobne zarzuty wobec analiz dokumentów patentowych w zakresie prognozowania rozwoju technologii stawiali też Koenraad Debackere i in. (2002). Wspomnieli oni m.in. o fakcie zróżnicowania różnych specjalności (oraz krajów), pod względem swojej podatności na generowanie nowych patentów, różnicach pomiędzy odmiennymi systemami (klasyfikacjami, przepisami) regulującymi przyznawanie patentów, co powoduje trudność w porównywaniu ich ze sobą, a także o pewnych problematycznych cechach tzw. zachowań naukowców w zakresie zgłaszania wynalazków (ang. *patenting behavior*) (Debackere et al., 2002, 216–217). Przez zachowania tego rodzaju rozumie się osobiste motywy i podejmowanie indywidualnych decyzji leżących u podstaw zgłaszania patentów do odpowiednich urzędów. Przykładowo, nie wszystkie patenty są stosowane w praktyce i nie wszystkie prowadzą do rzeczywistych innowacji. Nie wszystkie wynalazki dają się opatentować i nie wszystkie wynalazki, które dają się opatentować są faktycznie chronione prawem patentowym. Istnieją też różne powody, dla których nie wszystkie przedsiębiorstwa decydują się na patentowanie swoich odkryć. Mogą to być np. powody takie, jak zbyt duża trudność w wykazaniu nowatorskiego charakteru wynalazku, niechęć do ujawniania odkrycia, wysokie koszty wdrożenia wynalazku, nieefektywność prawa patentowego w zakresie szybkości i kosztów związanych z imitowaniem pewnych wynalazków. Niektóre wynalazki są ponadto patentowane „seryjnie”, w postaci więcej niż jednego dokumentu patentowego (Debackere et al., 2002, 217). Ogólnie mówiąc, zachowania badaczy w zakresie patentowania swoich wynalazków są silnie uzależnione od sektora przemysłu oraz od typu wynalazku. Istnieją również problemy związane z samymi systemami patentowymi (np. brak retrospektywnego korygowania wykorzystywanych kategorii tematycznych), problemy związane z ogromną różnorodnością wartości technologicznej i ekonomicznej wynalazków, a także z procesem cytowania patentów (np. brak ścisłych danych nt. odnawiania patentów w celu utrzymania w mocy prawa wyłącznego) (Debackere et al., 2002, 219–222).

## 5. Analizy głównych ścieżek rozwoju dyscypliny

Jest to kolejna z metod służących zademonstrowaniu sposobu, w jaki przebiegał historyczny rozwój danej dyscypliny oraz wskazaniu kierunku, jaki może on przybrać w przyszłości. Posłużyli się nią np. John Liu i Louis Lu (2012), identyfikując główne kanały przepływu informacji w obrębie literatury dotyczącej indeksu Hirsha, za pomocą metody opartej na pomysśle Hummona i Doreiana (1989) (zob. podrozdział trzeciej części pierwszej). Metoda ta została niejako rozszerzona przez Liu i Lu, przede wszystkim dzięki nadaniu jej wymiaru globalnego (w przeciwieństwie do oryginalnej metody, którą charakteryzował wymiar lokalny), dwukierunkowego (tj. analizowaniu ścieżek prowadzących nie tylko od dokumentów cytowanych do cytujących, ale też ścieżek wiodących od dokumentów cytujących do cytowanych) i dopuszczeniu możliwości wyodrębnienia więcej niż jednej „głównej ścieżki” przepływu informacji w danej dyscyplinie lub subdyscyplinie (Liu & Lu, 2012, 531–534). Ocena przyszłych perspektyw dyscypliny sprowadzała się w tym ujęciu do porównania podobieństwa wyodrębnionych ścieżek. Wysokie podobieństwo dwóch (lub większej liczby) ścieżek sygnalizowałoby koncentrację uwagi badaczy na publikacjach, które wytyczyły te ścieżki, podczas gdy brak podobieństwa sugerowałby, że rozwój dyscypliny przebiega wielokierunkowo, w sposób rozproszony, oznaczający brak jednolitej

trajektorii rozwojowej (Liu & Lu, 2012, 532, 537). Spośród wad proponowanego podejścia można wymienić np. wysoką zależność wyników od wybranych bibliograficznych źródeł informacji nt. publikacji z danego zakresu, brak uwzględnienia zróżnicowania stopnia relewancji tematycznej pomiędzy pracami cytowanymi a cytującymi (tj. milczące przyjęcie normatywnego założenia, że wszystkie cytowania są sobie równe), okazjonalne przypadki wyłączenia poza „główną ścieżkę” prac wyjątkowo wysoko cytowanych w danej dyscyplinie oraz faworyzowanie artykułów przeglądowych o pokaźnych bibliografiach załącznikowych i często wysokiej liczbie cytowań, które jednak nie przyczyniają się w bezpośredni sposób do przesuwania granic wiedzy w dyscyplinie (Liu & Lu, 2012, 538–540).

Praca Liu i Lu zainspirowała stworzenie kolejnej modyfikacji metody Hummona i Doreiana, w postaci tzw. map trajektorii konceptualnych (ang. *conceptual trajectory maps*), którą zademonstrowali Yi-Ning Tu i Shu-Lan Hsu (2016). Modyfikacja ta polegała przede wszystkim na uzupełnieniu „głównej ścieżki” publikacji o prace, które wprawdzie początkowo znalazły się poza tą ścieżką, ale wykazywały wystarczający stopień podobieństwa treściowego do prac leżących na ścieżce. Jako element podwyższający stopień podobieństwa dwóch różnych publikacji uwzględniono też daty ich wydania. Podobieństwo treściowe oraz podobieństwo pod względem czasu powstania publikacji<sup>7</sup> oceniono za pomocą wskaźnika podobieństwa kosinusowego (zob. podrozdział czwarty części pierwszej). Następnie uzyskanym tym sposobem grupom artykułów wytyczających przebieg ścieżki przypisano określone terminy zaczerpnięte z tekstów tych artykułów, które stanowiły reprezentację przedstawianych przez nie koncepcji (Tu & Hsu, 2016, 2020–2023). Zdaniem Tu i Hsu zbudowane w streszczony powyżej sposób mapy danych dyscyplin lub subdyscyplin mogą zostać wykorzystane w celu identyfikacji tematyki wartościowej i „trwałej” w danym obszarze oraz tematyki „ulotnej” (ang. *transient*), jak również w celu śledzenia narodzin i rozwoju zagadnień nowo powstających oraz w celu badania całokształtu cyklu życiowego danej problematyki w jej otoczeniu informacyjnym (Tu & Hsu, 2016, 2024, 2030).

Metoda Hummona i Doreiana była ponadto punktem wyjścia do analizy aktualnych frontów badawczych w obszarze z pogranicza prawa i finansowości, zrealizowanej przez Vincenta Ma i Johna Liu (2016). „Główne ścieżki” wybranego obszaru problemowego zostały wyodrębnione w ramach utworzonych uprzednio klastrów, przy konstrukcji których posłużono się algorytmem grupującym Newmana z 2004 r. (zob. podrozdział drugi). Rozszerzoną wersję metody identyfikacji ścieżek zaczerpnięto z kolei od Liu i Lu (2012). W dalszej kolejności uzupełniono ją o wybór słów kluczowych, które zostały wykorzystane jako deskryptory frontów badawczych (Ma & Liu, 2016, 36–37, 41, 44). Ma i Liu przyznali zarazem, że wyniki analizy „głównych ścieżek” są dynamiczne w tym sensie, że mogłyby zmienić się, gdyby do rozpatrywanego, wyjściowego zbioru publikacji dołączono publikacje najnowsze (nie dostępne jeszcze w momencie realizacji badania). Dlatego swoje wyniki autorzy określili jako „migawkę” (ang. *snapshot*), która niekoniecznie musiała odzwierciedlać trendy stabilne w czasie i obowiązujące w dalszym ciągu w zbadanym obszarze nauki (Ma & Liu, 2016, 49).

<sup>7</sup> W tym zakresie rozpatrywano tylko prace, które powstały w odstępie nie większym i nie mniejszym niż rok od opublikowania pracy, do której te pierwsze były przyrównywane. Np. jeżeli praca leżąca na ścieżce została wydana w 2006 r., wyznaczano jej podobieństwo do prac wydanych w przedziale 2005–2007 (Tu & Hsu, 2016, 2022).

Wyniki badania, które w wyższym stopniu akcentowało dynamikę dyscyplin naukowych, przedstawił Erjia Yan i Qi Yu (2016). Wykorzystana przez nich metoda nie była tą samą, która stanowiła podstawę trzech analiz streszczonych powyżej chociaż wydaje się, że można zaliczyć ją do metod pokrewnych. Metodą tą było tzw. maksymalne drzewo rozpinające (MST – ang. *maximum spanning tree*), a jej zastosowanie do sieci cytowań pozwoliło autorom na zlokalizowanie dyscyplinarnych ścieżek dyfuzji wiedzy, co było też czołowym założeniem opracowań posługujących się metodą identyfikacji „głównych ścieżek” (Yan & Yu, 2016, 1944). Danymi empirycznymi były w tym przypadku cytowania pomiędzy czasopismami (ang. *journal-to-journal citations*) oraz kategorie tematyczne czasopism zapożyczone z bazy *Scopus* (Yan & Yu, 2016, 1944–1945). Dynamiczny aspekt analizy z użyciem MST opierał się na wprowadzeniu pięciu dwuletnich okresów obserwacji cytowań. Dokładniej mówiąc, dane wyjściowe obejmowały cytowania odsyłające do czasopism wydanych dwa lata wcześniej, a więc cytowania, które pojawiły się w 1999 r. i odsyłały do prac z 1997 r., cytowania z 2002 r. odsyłające do prac z 2000 r., i cytowania z lat 2005, 2008 i 2011 odsyłające odpowiednio do prac z lat 2003, 2006 i 2009 (Yan & Yu, 2016, 1944, 1952). Sama metoda MST polega natomiast na wyodrębnieniu z całej sieci pewnej podsięci połączeń stanowiących niejako jej esencję, tj. najmniejszą z możliwych podsięci połączeń o najwyższej wadze (mierzonej odpowiednio znormalizowaną liczbą cytowań pomiędzy czasopismami), która nie zawiera połączeń cyklicznych<sup>8</sup> i jednocześnie pozwala utrzymać przepływ informacji. Usunięcie jakiegokolwiek krawędzi z takiej podsięci skutkowałoby zaburzeniem tego przepływu. Podsięć MST ma dokładnie tyle samo wierzchołków, co sieć wyjściowa i liczbę połączeń równą, co najwyżej liczbie wierzchołków minus jeden, co wskazuje, że tworzenie MST jest swego rodzaju sposobem na „oczyszczanie” sieci cytowań i obnażenie kognitywnego „kręgosłupa” (ang. *backbone*) dyscypliny (Yan & Yu, 2016, 1946–1947, 1950). Połączenie MST z kilkoma okresami obserwacji cytowań stworzyło możliwość monitorowania zmian w strukturze sieci cytowań zachodzących w niej z biegiem czasu. Konkretnie mowa tu o monitorowaniu siły wpływu, jaką wywierają na siebie wzajemnie różne dyscypliny nauki. O wpływie tym można było mówić w kategoriach „eksportu” i „importu” informacji oraz potencjału informacyjnego, który dyscypliny mają do zaferowania i z którego wyrastają nowe odgałęzienia dyscyplin (ang. *child branches*). Odgałęzienia te mogą być nie tylko specjalizacjami nowo powstającymi, ale mogą też być szerokimi strukturami dyscyplinarnymi, które z czasem zmieniają swoją pozycję na mapie nauki. Przykładowo, w latach 2000/2002 materiałoznawstwo było gałęzią fizyki („importerem wiedzy”), a w latach 2011/2009 było już „eksporterem” informacji przepływającej do dyscyplin fizycznych (Yan & Yu, 2016, 1950, 1952). Sekwencja map dawała ponadto sposobność odnotowania wielkoskalowych tendencji w ustalaniu się zależności pomiędzy głównymi obszarami mapy. Zaobserwowano np., że pomiędzy rokiem 1997 a 2011 dyscypliny takie jak biznes, materiałoznawstwo, informatyka i dyscypliny inżynierskie zajmują coraz bardziej dominujące pozycje i stają się głównymi „eksporterami” informacji (Yan & Yu, 2016, 1954). Ograniczenia zastosowanej metodologii wynikały przede wszystkim z charakterystyk zbioru danych empirycznych. Były nimi po pierwsze, sztuczne – w ocenie

<sup>8</sup> Połączeniem cyklicznym byłoby takie odwołanie, które zaczynałoby się i kończyło w tym samym wierzchołku. Miałoby ono miejsce gdyby, np. czasopismo A cytowało czasopismo B, które z kolei cytowałoby czasopismo C cytujące czasopismo A (A→B→C→A).

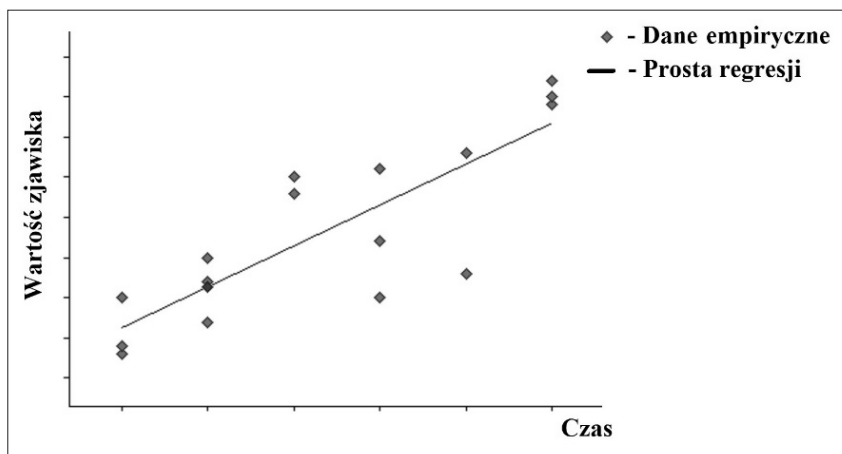
autorów – kategorie tematyczne *All Science Journal Classification*, a po drugie przyjęta jednostka analizy, którą były czasopisma i dyscypliny naukowe, a nie indywidualne publikacje. Przyjęcie jednostki usytuowanej na niższym poziomie agregacji danych umożliwiłoby, zdaniem autorów, precyzyjniejszą interpretację otrzymanych wyników. Analiza Yana i Yu była ponadto pozbawiona możliwości uwzględnienia oddziałujących na naukę czynników socjotechnicznych (np. polityki finansowania badań czy ewolucji sieci współpracy naukowej) (Yan & Yu, 2016, 1950, 1953).

## 6. Analiza szeregów czasowych

Kolejną grupą publikacji jest ta, w której autorzy posługiwali się metodą bliższą naukom społecznym i ekonomicznym (np. analizami socjologicznymi, ekonomicznymi lub finansowymi), niż naukom ścisłym (informatyce i matematyce), w przeciwieństwie do wielu streszczonych powyżej algorytmicznych metod tworzenia klastrów publikacji wzajemnie powiązanych. Ścisłej mówiąc, pod nazwą „analiza szeregów czasowych” kryje się cały zespół różnych metod statystycznej analizy dynamiki zjawisk masowych, których wspólnym celem jest przewidywanie przyszłego zachowania się zjawisk zachodzących w czasie za pomocą identyfikacji głównych trendów rozwojowych zjawiska, jego sezonowych lub okresowych odchyień od trendu i oceny niepewności towarzyszącej prognozie. W tym samym zakresie stosuje się ponadto techniki analityczne tzw. dekompozycji szeregu czasowego, które polegają na sformułowaniu równania opisującego ilościowo przebieg zjawiska z rozróżnieniem trendu głównego, składowej wahań cyklicznych lub sezonowych oraz składowej odpowiadającej wahaniom przypadkowym (zob. np.: Aczel, 2007, 621–667; Sobczyk, 2015, 307–356; Sobczyk, 2008, 35–87). O niektórych z tych metod wspomniano już w podrozdziale 3 przy okazji omawiania pracy Choudhury’ego i Uddina (2016), w której metody tego rodzaju pełniły funkcję uzupełniającą wobec metody analizy struktur sieciowych typu *co-word*. Podobnie jak w tym przypadku, także i w publikacji Yuena-Hsiena Tsenga i in. (2009) analiza trendu wyznaczającego przypuszczalny kierunek przyszłego rozwoju nowo powstałych specjalizacji, została poprzedzona identyfikacją i wizualizacją struktury dyscypliny w postaci mapy tego samego typu (*co-word*). Identyfikacja ta odbywała się przy użyciu standardowych technik grupowania wyjściowego zbioru publikacji w klastry (Tseng et al., 2009, 75, 80). Każdy klastery został następnie podzielony na podzbiory publikacji o tym samym wieku. To znaczy, że każdy klastery został przedstawiony jako szereg wartości odpowiadających liczbom tworzących go publikacji o wieku jednego roku, dwóch lat itd. Na przykład, ciąg postaci {25, 38, 47, 40...} oznaczałby, że w pewnym klastrze znalazło się 25 publikacji wydanych nie wcześniej niż rok przed momentem badania, 38 publikacji wydanych nie wcześniej niż dwa lata przed chwilą badania itd. W ten sposób powstały serie czasowe, które poddano dalszej analizie statystycznej (Tseng et al., 2009, 75). W ramach tej analizy porównano ze sobą sześć różnych wskaźników statystycznych stanowiących predyktory trendu rozwojowego (ang. *trend prediction index*) (Tseng et al., 2009, 77). Pierwszym z indeksów było średnie tempo przyrostu procentowego (ang. *API – average percentage of increase*), które informuje o tym, o ile wyższy lub niższy był średnio poziom badanego zjawiska w danym okresie, w stosunku do okresu bezpośrednio go poprzedzającego. Drugim indeksem był tzw. współczynnik kierunkowy prostej regresji (ang. *SLP – slope of linear regression*). Prosta



regresji to linia prosta, która stanowi najlepsze z możliwych dopasowanie do danych empirycznych, a kąt jej wznoszenia się lub opadania wraz ze wzrostem wartości na osi czasu określa, jak szybkie jest tempo wzrostu lub spadku wartości jakiejś cechy danego zjawiska (w tym przypadku – jak szybkie jest tempo wzrostu lub spadku liczby publikacji wraz ze wzrostem ich wieku w obrębie klastra). Współczynnik kierunkowy prostej jest z kolei czynnikiem determinowanym ściśle przez wspomniany kąt, pod którym prosta opada (lub wznosi się)<sup>9</sup>. Na rysunku 3 przedstawiono ilustrację dopasowania wznoszącej się prostej do hipotetycznych danych doświadczalnych.



Rys. 3. Prosta regresji dopasowana do hipotetycznego zbioru danych empirycznych

Pozostałe cztery indeksy stanowiły warianty dwóch pierwszych (Tseng et al., 2009, 77). Po zakończeniu etapu realizacji badania empirycznego wybranego zbioru prac, Tseng i in. zarekomendowali metodę SLP jako dającą najwyższą zgodność z zewnętrznymi źródłami informacji nt. problematyki zyskującej na znaczeniu w środowisku naukowym. Metoda ta wykazała się najwyższą efektywnością w zróżnicowanych warunkach badawczych, tj. warunkach uwzględniających różne przedziały czasowe analizy, różne okresy prognozowania, różne dyscypliny i różne, pod względem wielkości, zbiory analizowanych dokumentów (Tseng et al., 2009, 88–89).

Günter Krampen i in. (2011) wykorzystali alternatywną, wobec wskazanych powyżej, metodę prognozowania przyszłego przyrostu liczby publikacji w dyscyplinie psychologii. Metodą tą było tzw. wygładzanie wykładnicze (ang. *exponential smoothing*), w ramach którego ekstrapoluje się w przyszłość trend zaobserwowany w momencie badania i opisujący dane rozciągające się w przeszłość. Obserwacjom zjawiska zachodzącego w latach ubiegłych (a konkretnie w latach 1977–2008) przypisano w ramach tej metody wagi, które są większe dla obserwacji z lat bliższych współczesności, niż dla obserwacji starszych, przy czym waga obserwacji maleje wykładniczo wraz ze wzrostem dystansu czasowego. Inaczej mówiąc, stosując wygładzenie wykładnicze, obserwacje najnowsze wywierają najsilniejszy wpływ

<sup>9</sup> Więcej szczegółów na temat prostej regresji można znaleźć np. w podręczniku Johna Taylora (2012, 216–229), Barbary Łapkowskiej-Baster (2009, 19–31) lub Nicholasa Bingham i Johna Fry'ego (2010).

na wartości prognozowane. W związku z tym, jeśli trend (główne przyczyny determinujące zjawisko) zmieniłby się w jakimś punkcie w przeszłości, wartości zjawiska sprzed tej zmiany będą wywierać mniejszy wpływ na prognozę, niż wartości zjawiska zaobserwowane po zajściu zmiany (Krampen et al., 2011, 688–689; zob. też: Choudhury & Uddin, 2016, 751–752). Waga obserwacji coraz starszych może przy tym spadać szybko lub wolno, a jej konkretne wartości określa się przyjmując współczynnik (parametr wagowy) występujący w wyrażeniu przez które mnoży się wartość każdej obserwacji (zob. np.: Aczel, 2007, 642; Sobczyk, 2008, 121). Za problematyczny aspekt tej metody można przypuszczalnie uznać fakt, że użyte przez Krampena i in. proste wygładzanie wykładnicze<sup>10</sup> stosuje się na ogół do takich szeregów czasowych, które nie mają wyraźnego trendu ani wahań sezonowych (Aczel, 2007, 641, 647; Sobczyk, 2008, 122, 138; Snarska, 2011, 218), a więc są szeregami w dużym stopniu nieregularnymi co powoduje, że prognoza może być obciążona znacznym błędem. W literaturze przedmiotu można też napotkać stwierdzenie, według którego ocena dynamiki dyscyplin w oparciu o same jedynie trendy publikacyjne (tj. liczbę publikacji z danego zakresu w różnych latach) jest niewłaściwa z uwagi na to, że różne typy wydawnicze publikacji przyczyniają się do rozwoju dyscyplin w niejednakowy sposób lub ze względu na to, że całkowita liczba publikacji w dyscyplinach rośnie na ogół wykładniczo, co może przesłaniać węższe i bardziej wyspecjalizowane trendy (Chi et al., 2006, 69; Sitarz et al., 2012, 4; Tseng et al., 2009, 75).

Robert Sitarz i in. (2012) zaproponowali własną metodę analizy krótkoterminowych trendów w nauce, w oparciu o słowa kluczowe i klastry zbudowane przy użyciu techniki współwystępowania słów kluczowych. Klastry, traktowane jako odpowiedniki subdyscyplin lub obszarów problemowych, stają się następnie podstawą budowy szeregów czasowych, w których kolejne wartości odpowiadają liczbie dokumentów wydanych w każdym kolejnym roku analizy i zawartych wewnątrz klastra (Sitarz et al., 2012, 7–10). Metoda identyfikacji trendu była w tym ujęciu złożeniem trzech prostszych metod: trendu liniowego (ang. *trend line method*), metody oscylacyjnej (ang. *oscillator method*) i metody oscylatora stochastycznego (ang. *stochastic oscillator method*) (Sitarz et al., 2012, 11–14). Metoda trendu liniowego polega na przeprowadzeniu linii prostej łączącej minima (lub maksima) na wykresie wartości obserwacji tworzących szereg czasowy. Dopuszczalna jest tu zarazem więcej niż jedna prosta opisująca dane doświadczalne. Trend liniowy może więc składać się z np. dwóch prostych w układzie współrzędnych, które przecinają się w jakimś punkcie (Sitarz et al., 2012, 11–12). Metoda oscylacyjna polega na oszacowaniu różnicy pomiędzy liczbą publikacji w danym roku „t”, a liczbą publikacji w „x” poprzednich latach. Wylicza się tę różnicę dla wielu kolejnych lat i wynik przedstawia na jednym wykresie (Sitarz et al., 2012, 12–13). Oscylator stochastyczny jest z kolei techniką analityczną, która umożliwia przewidzenie nasilenia się trendu po pewnym momencie, jego spadek lub zmianę kierunku trendu (zob.: Sitarz et al., 2012, 14; zob. też: Krzywda, 2010, 111–113). Wykryty trend uznaje się za istotny kiedy obie metody oscylatorowe oraz metoda liniowa wysyłają ten sam sygnał, tj. gdy ich wskazania pokrywają się (Sitarz et al., 2012, 20). Jako ostateczny

<sup>10</sup> Do prognozowania wartości szeregów czasowych cechujących się wyraźniejszym trendem lub wahaniami sezonowymi stosuje się bardziej wyrafinowane metody tzw. wykładniczego wygładzania adaptacyjnego, np. metodę wygładzania Holta, metodę wygładzania Browna, metodę wygładzania Wintersa lub metodę wyznaczania tzw. trendu pelzającego (zob. np.: Aczel, 2007, 645, 647; Cipra & Hanzák, 2008, 385; Krawiec, 2014; Malina, 1994; Snarska, 2011, 212–244; Sobczyk, 2008, 97–152).

rezultat uzyskuje się fragmentację całego cyklu życiowego pewnego zbioru publikacji na krótsze podokresy, które mogą charakteryzować się własnym trendem (wznoszącym lub opadającym w różnych przedziałach czasu). Autorzy wskazali na możliwość dalszego dopracowania metody poprzez np. zwiększenie jej wrażliwości w aspekcie, w którym jest ona zależna od doboru właściwych słów kluczowych. Zaznaczono też potrzebę dokładniejszej determinacji wartości progowych w trakcie wyodrębniania klastrów (jak wiele współwystąpień słów kluczowych musi zaistnieć, aby przypisać pewną pracę do konkretnego klastra?) oraz potrzebę automatyzacji, tj. zapewnienia metodzie zdolności do analizy dużych zbiorów danych (Sitarz et al., 2012, 27–28).

## 7. Pozostałe publikacje

W tej sekcji zebrano i i skrótkowo scharakteryzowano relewantne do celu i przedmiotu niniejszej pracy publikacje, które były na tyle oryginalnymi propozycjami, że nie zmieściły się w obrębie żadnej z wymienionych wcześniej podgrup.

Steven Morris (2005) zaproponował tzw. matematyczny model CAPE (ang. *cumulative advantage by paper with exemplars*), który wyjaśniał i odtwarzał m.in. proces wyłaniania się wysoko cytowanych „publikacji przykładowych/wzorcowych” (ang. *exemplar references*)<sup>11</sup> natychmiast po ukształtowaniu się nowo powstałej specjalizacji badawczej (Morris, 2005, 1251, 1266, 1272). Konceptualizacja narodzin nowych specjalności Morrisa była zgodna z teorią postępu w nauce Thomasa Kuhna (2001), tj. uznano w jej ramach, że długie okresy tzw. nauki normalnej są przerywane krótkimi okresami rewolucji naukowych (tzw. przewroty kopernikańskie). Ich konsekwencją jest zmiana obowiązującego w nauce paradygmatu i dokonywanie nowych odkryć w ramach nowego paradygmatu, które doprowadzają do powstania nowych specjalności (Morris, 2005, 1251). Model Morrisa był w stanie zrekonstruować etapy narodzin nowych specjalności takie, jak ogłoszenie odkrycia w formie publikacji, proces narastania cytawalności takich publikacji, pojawienie się prac testujących odkrycie empirycznie i potwierdzających jego słuszność, rozprzestrzenianie się wiedzy o odkryciu poza granicami jego macierzystej dyscypliny naukowej oraz wejście w fazę konsolidacji, którą wyznaczają powstające publikacje przeglądowe i monograficzne standaryzujące, kodyfikujące i popularyzujące odkrycie. Rekonstrukcja obejmowała również dynamikę nawarstwiania się cytowań i generowania powiązań bibliograficznych i współcytowań pomiędzy publikacjami (Morris, 2005, 1251–1252, 1272). Do wad swojego modelu Morris zaliczył natomiast jego niezdolność do odwzorowania relacji współcytowania publikacji pochodzących z różnych specjalności dyscyplinarnych, jak również niezdolność do odzwierciedlenia specjalnego przypadku wykształcania się nowych specjalności, jakim jest rozszczepianie się szerszych specjalności na mniejsze „subspecjalności”. Niedostatkami modelu było ponadto przyjęcie założenia, że tempo pojawiania się publikacji wzorcowych w dyscyplinie maleje wykładniczo wraz z upływem czasu, podczas gdy w rzeczywistości

<sup>11</sup> „Publikacje wzorcowe” to według Morrisa prace, w których zaprezentowano rozwiązane uprzednio problemy naukowe i które dostarczają tym samym fundamentalnych definicji problemów, które nie doczekały się jeszcze rozwiązania oraz metod, za pomocą których należy problemy te badać i rozwiązywać. Ogólnie mówiąc publikacje te wyznaczają standardy „naukowości” i standardy technik walidacji problemów istniejących w ramach aktualnie obowiązującego paradygmatu (Morris, 2005, 1251, 1253).

tempo to jest najprawdopodobniej wysoce zależne od konkretnej sytuacji, dyscypliny i specyfiki zachowań naukowców w zakresie cytowań literatury źródłowej (ang. *case-specific*). Do mankamentów swojego podejścia Morris zaliczył ponadto przyjęcie ustalonej (niezmiennej, utrzymującej się na stałym poziomie) „siły przyciągania” cytowań (ang. *attractiveness*) przez publikacje wzorcowe, w postaci stałego parametru wprowadzonego w matematyczne sformułowanie modelu (Morris, 2005, 1273).

Hanning Guo i in. (2011) wysunęli propozycję „wieloskładnikowej” metody identyfikacji nowo powstających specjalności naukowych (ang. *mixed-indicators model*). Ich koncepcja polegała na połączeniu trzech wskaźników cząstkowych, które wspólnie sygnalizowałyby wyłonienie się nowego obszaru problemowego. Wskaźnikami tymi były: (1) gwałtowny wzrost częstości użycia określonych terminów w dyscyplinie (ang. *word burst*), który poprzedza wejście takich terminów do powszechnego użytku, (2) pojawienie się nowych autorów w obrębie nowej specjalności, których przyciągnął potencjał nowego obszaru badań oraz (3) cytowanie pozycji pochodzących z różnorodnych dyscyplin (wysoce międzydyscyplinarne pozycje bibliograficzne cytowane w nowej subdyscyplinie). U podstaw ostatniego z trzech wymienionych wskaźników tkwiło założenie, że nowe obszary wyrastają z obszarów już istniejących i oddzielają się od nich czerpiąc jednocześnie wiedzę z innych dyscyplin. Drugim powodem wzrostu interdyscyplinarności przypisów był tutaj brak publikacji specyficznych wyłącznie dla takiej nowo powstałej specjalności na wczesnym etapie jej rozwoju (Guo et al., 2011, 423, 429, 431–432).

Ling-Chu Lee i in. (2012) realizowali cele predykcyjne za pomocą trzech koncepcji metodologicznych. Pierwszą z nich było zastosowanie tzw. metody Delphi z zamiarem zebrania opinii ekspertów i zidentyfikowania kluczowych zagadnień w badanym obszarze nauki. Drugą była analiza bibliometryczna, poprzez którą stworzono opis trendu wzrostowego w liczbie publikacji odpowiadających danej tematyce, co miało za zadanie ustalenie aktualnego etapu rozwoju, na którym znajdowała się dana grupa zagadnień. Trzecia koncepcja polegała na analizie literatury przedmiotu za pomocą tzw. indeksu RCA (ang. *revealed cumulative advantage*). Indeks ten pomaga wyodrębnić obszary tematyczne, które znajdują się aktualnie na porównywalnym etapie rozwoju na świecie i w jakimś konkretnym kraju (którym – w przypadku badania Lee i in. – był Tajwan) (Lee et al., 2012, 817, 820, 827; zob. też: Verbeek et al., 2002, 202). Indeks ten wskazuje, jaki jest udział kraju w dorobku światowym w pewnej dyscyplinie, w porównaniu z udziałem tego kraju w ogólnoswiatowej produkcji piśmienniczej. Inaczej mówiąc, wskaźnik demonstruje, czy kraj angażuje się w danej dyscyplinie w naukę światową i w jakim stopniu, w stosunku do stopnia jego zaangażowania i aktywności w innych dyscyplinach. Wartość indeksu większa od „1” świadczy o tym, że kraj specjalizuje się w dyscyplinie (udział publikacji z niego się wywodzących jest duży w ramach dorobku ogólnoswiatowego). Wartość indeksu mniejsza od „1” wskazuje, że pozycja danego kraju w dyscyplinie jest mało znacząca. Wartość równa „1” oznacza natomiast pozycję neutralną (Lee et al., 2012, 818–819, 821; zob. też: Verbeek et al., 2002, 202). Metodę prognozowania i zarazem podstawę sformułowania strategii rozwojowej kraju stanowiło ostatecznie połączenie wartości wskaźnika RCA z etapami rozwoju danej problematyki w postaci tabeli zawierającej wszystkie możliwe kombinacje wysokich i niskich wartości wskaźnika RCA z etapami rozwoju takimi jak etap „embrionalny”, nowo powstały, rosnący oraz dojrzały (ang. *embryonic, emergent, growing, mature*) (Lee et al., 2012, 821). Za kluczową część wytyczającą plany rozwoju kraju uznano te obszary nauki,

które znalazły się w fazie embrionalnej o wysokim wskaźniku RCA. Znamionowały one te kwestie, które znajdują się w bardzo wczesnej fazie rozwoju na świecie i jednocześnie, w których Tajwan posiada silną pozycję (ang. *comparative advantage*) w porównaniu z innymi krajami (Lee et al., 2012, 822). Do ograniczeń metody zaliczono fakt, że wskaźnik RCA ignoruje czynniki inne niż sama tylko liczba publikacji na dany temat, np. jej siłę wpływu lub jakość naukową. Metoda autorów nie uwzględniała ponadto czynników takich, jak np. działania rządu w zakresie subsydiowania różnych gałęzi przemysłu i rolnictwa, kwestii podatkowych lub ryzyka wynikającego z kwestii dokładności szacowania RCA (Lee et al., 2012, 828).

Xianwen Wang i in. (2013) wystąpili z oryginalnym pomysłem polegającym na badaniu najnowszych trendów w nauce poprzez skupienie uwagi na środowisku publikacji elektronicznych. Według Wang'a i in. miałyby to być sposób na zaradzenie trudnościom, jakie napotyka tradycyjna analiza cytowań stosowana w perspektywie prognostycznej. W tym kontekście wspomniano m.in. o fakcie istnienia opóźnienia w publikowaniu artykułów naukowych i opóźnienia związanego z pojawianiem się kolejnych cytowań albo o problemie z zakresu socjologii nauki, którym są konstruktywistyczne zarzuty stawiane przed paradygmatem normatywnym<sup>12</sup> (Wang et al., 2013, 718–719). Z uwagi na tego rodzaju problemy, autorzy podjęli próbę podsumowania aktualnych tendencji i najnowszych osiągnięć (ang. *hot topics*) w naukometrii w oparciu o dane na temat pobrań elektronicznych wersji artykułów zamieszczonych na platformie *SpringerLink*. Metodą śledzenia nowo powstających obszarów badawczych było w ujęciu autorów zidentyfikowanie takich przypadków pobrań publikacji elektronicznych, w których publikacje te opatrzone były słowami kluczowymi cechującymi się nowością w ostatnich latach (nieczęstym występowaniem), i w których – po drugie – liczba pobrań artykułów z nowymi słowami kluczowymi była wyższa niż 50. Po trzecie, przypadki te obejmowały sytuacje, w których stosunek pobrań prac z danym słowem kluczowym do liczby publikacji opisywanych tym słowem był wyższy niż 20 (tj. akcentowano tutaj wysoką liczbę pobrań i zarazem niewielką liczbę publikacji, którym przypisano dane słowa kluczowe) (Wang et al., 2013, 725, 727). Jako mankament metody wskazano możliwość, że pewna liczba publikacji mogła być pobierana z powodów innych niż czysto naukowe, w związku z czym nie można było całkowicie wykluczyć czynników arbitralności i pewnej przypadkowości obecnych w przynoszonych przez metodę wynikach (Wang et al., 2013, 728).

Xiaoling Sun i in. (2016) podzielali opinię Uphama i Smalla (2010), zgodnie z którą interdyscyplinarność badań naukowych jest wyznacznikiem zdolności do wzbudzania i przyspieszania innowacyjnych kierunków rozwoju nauki. Z drugiej strony, Sun i in. zgadzali się z tezą, według której konferencje naukowe są źródłem danych szczególnie podatnych na monitorowanie dyscyplin naukowych pod kątem kierunku i tempa ich rozwoju. W związku z tym obrali oni za przedmiot badania sieć współpracy między autorami, którzy w latach 1980–2010 brali udział w cyklicznych konferencjach naukowych poświęconych informatyce

<sup>12</sup> Chodzi tutaj o kwestionowanie normatywnej tezy, według której cytowania publikacji świadczą bezpośrednio o sile jej wpływu i naukowej wartości lub o podnoszone często kwestie związane ze specyfiką zachowań indywidualnych naukowców w zakresie cytowań (ang. *citation behaviour*), ich złożonymi motywacjami i procesami decyzyjnymi kryjącymi się za każdym aktem zamieszczenia przypisu itd. Szczegółowy przegląd tych zagadnień przedstawili np. Lutz Bornmann i Hans-Dieter Daniel (2008) oraz Mengxiang Liu (1993).



wraz z jej wielod dziedzinowymi zastosowaniami. Dane bibliograficzne takich konferencji indeksowane są m.in. w bazie *Digital Bibliography & Library Project* (DBLP) (Sun et al., 2016, 751–752). Przyjmując powyższe założenia utworzono złożoną z klastrów sieć (ang. *network of communities*), która obejmowała wszystkie poszczególne lata z przedziału 1980–2010. Konferencje spełniały tu funkcję wierzchołków sieci, a rolę połączeń (krawędzi) między wierzchołkami odgrywała relacja podobieństwa zdefiniowana jako współwystępowanie autorów referatów, które wygłoszono podczas różnych konferencji i opublikowano w materiałach pokonferencyjnych (Sun et al., 2016, 751–752). Algorytm grupujący wyjściowy zbiór konferencji na mniejsze podzbiory stanowił wariant algorytmu Newmana (2004) i został przez autorów zapożyczony z pracy Vincenta Blondela i in. (2008). Ponieważ autorzy referatów publikowali na ogół w różnych specjalnościach informatyki, a z czasem angażowali się w coraz większą liczbę specjalności i obszarów problemowych<sup>13</sup>, skonstruowana sieć odzwierciedlała ewolucję informatyki, rozprzestrzenianie się specjalności i przepływ informacji pomiędzy danymi specjalnościami, wyłanianie się nowych obszarów i zanikanie tych starszych, czy dokonywanie się przełomów i wzrost popularności pewnych nowo postawionych zagadnień (Sun et al., 2016, 752–753, 757–758, 760). Za słabą stroną swojej metody Sun i in. uznali fakt, że nie radziła sobie ona z niejednoznacznością nazwisk badaczy, tj. z występowaniem nazwisk błędnie zapisanych i nazwisk homonimicznych (istnienie wielu autorów o tym samym nazwisku). Po drugie wskazano, że możliwa jest sytuacja, w której nawet jeżeli na dwóch różnych konferencjach nie było ani jednego tego samego autora, nie oznacza to jeszcze, iż podczas takich konferencji nie poruszano tych samych zagadnień. Zasugerowano ponadto włączenie w zakres zademonstrowanej metody aspektu leksykalnego, np. w postaci uzupełnienia jej o mapę współwystąpień terminów (Sun et al., 2016, 759).

Ostatnią pracą, o której należy wspomnieć jest artykuł autorstwa Olega Eny i in. (2016), w którym podjęto tematykę monitorowania trendów w technice i technologiach (TTM – ang. *technology trend monitoring*) jako narzędzia pomocnego w regulowaniu strategii badawczo-rozwojowych. Środkiem wiodącym do tego celu było kompleksowe rozpatrzenie nie tylko artykułów zamieszczanych w czasopismach naukowych i patentów, ale też szerokiego spektrum raportów i projektów badawczych, prezentacji biznesowych i konferencyjnych, rozpraw doktorskich, prasy i mediów społecznościowych (ang. *social networks*) z zamiarem uchwycenia całego cyklu życiowego i ogółu ścieżek rozwoju nowych technologii (Ena et al., 2016, 1015, 1021, 1025–1027). Autorzy zarzucili ponadto dotychczas wykorzystywanym metodom identyfikacji nowo powstałych specjalności naukowo-technologicznych, a w szczególności funkcjonującym w ich ramach narzędziom służącym terminologicznemu kodowaniu („etykietowaniu”) najnowszych prądów, generowanie terminów zbyt szerokich i zbyt ogólnych, które były w stanie opisać jedynie szerokie i mało konkretne tendencje, zamiast ścisłego wyszczególnienia merytorycznie adekwatnych do aktualnej sytuacji na „froncie badań” kierunków rozwoju i bardziej wyrafinowanych wzorców tego rozwoju (Ena et al., 2016, 1020). Zdaniem autorów zamiar ten można było osiągnąć poprzez uwzględnienie kontekstu terminów/słów kluczowych oraz udoskonalenie istniejących algorytmów grupujących dane wyjściowe. Należy ponadto zapewnić sobie pomoc ekspertów

<sup>13</sup> Sun i in. skupili się przede wszystkim na zbiorze takich właśnie autorów, którzy wykazywali się aktywnością w wielu różnych specjalnościach. Określono ich jako „autorów publikujących w krzyżujących się obszarach” (ang. *cross-fields authors*) (Sun et al., 2016, 751, 757).

w zakresie budowania zbioru terminów/słów kluczowych i interpretowania wszelkiego rodzaju wskaźników ilościowych oraz zwrócić baczniejszą uwagę na aspekty i etapy cyklu życiowego produktów lub technologii takie, jak czołowi badacze i instytucje, kraje wiodące prym w opracowywaniu i stosowaniu danych technologii czy dynamika wzajemnych zależności pomiędzy wszystkimi wymienionymi czynnikami (Ena et al., 2016, 1021–1022, 1024). Odnośnie do udoskonalenia dotychczasowych algorytmów grupujących dane, zdaniem Eny i in., należy wyeliminować z nich elementy, które powodują m.in. konieczność podziału zbioru wyjściowego na węższe podokresy czasowe celem uchwycenia dynamicznego aspektu rozwoju badanych dyscyplin lub niezdolność algorytmu do automatycznego odrzucenia zbędnych (pobocznych) klastrów, powstających jako efekt uboczny słabych i mało istotnych połączeń między dokumentami. Problemem wymagającym uwagi jest również brak uniwersalnej miary podobieństwa dokumentów jako kryterium, w oparciu o które odbywa się podział na podzbiory reprezentujące specjalizacje i obszary problemowe (ang. *core clustering metric*) (Ena et al., 2016, 1023). W miejsce technik grupujących obarczonych powyższymi trudnościami zaproponowano więc własny algorytm, stworzony w *National Research University Higher School of Economics* (HSE) i zaprojektowany z myślą o m.in. możliwości przetwarzania pełnych tekstów dokumentów, inteligentnym stosowaniu list „słów stopujących” (ang. *stop-words*) oraz funkcjonowaniu na zasadzie tzw. grupowania rozmytego (ang. *fuzzy clustering*). Polega ono na zapewnieniu algorytmowi możliwości przydzielenia tego samego dokumentu do więcej niż jednego klastra. Motywacją wprowadzenia grupowania rozmytego było założenie, że słownictwo używane w publikacjach naukowych jest pochodną wielu obecnych w publikacji koncepcji, idei czy teorii, które zachodzą na siebie i manifestują się poprzez warstwę terminologiczną tekstu. Koncepcje te powtarzają się, współwystępują w kolekcji dokumentów dyscyplinarnych, ale proporcje, w jakich poszczególne dokumenty odnoszą się do poszczególnych koncepcji są niejednakowe, do pewnego stopnia przypadkowe, a ich modelowanie wymaga specjalnego podejścia. Z tego powodu kwalifikacja dokumentów do odpowiednich klastrów powinna dopuszczać przypisywanie ich do wszystkich klastrów, z którymi wykazują one semantyczne pokrewieństwo w mniejszym lub większym stopniu. Dzięki temu ilościowe i terminologiczne charakterystyki klastrów stają się bardziej miarodajne i wiarygodne (Ena et al., 2016, 1023, 1031, 1033; nt. grupowania rozmytego w informatologii zob. też: Blei & Lafferty, 2007, 18; Hood & Wilson, 2002; Miyamoto, 1990, 125–188; Pindłowa, 1994, 102–103).

## 8. Wnioski

Na wstępie należy podkreślić, że obszerność powyższego przeglądu literatury wynika przede wszystkim z braku dostępnych w języku polskim, kompleksowych i systematycznych opracowań poświęconych tematyce metodologii prognozowania i oceny rozwoju dyscyplin naukowych. Autor kierował się w tym względzie chęcią przynajmniej częściowego wypełnienia tej luki i zademonstrowania tego, na jak zaawansowanym poziomie prowadzone są badania nad rozwojem dyscyplin na świecie. W polskiej społeczności specjalistów z zakresu informatologii (w tym bibliometrii) zainteresowanie metodami matematycznymi jest raczej niewielkie, na co wskazała m. in. Wanda Pindłowa (1994, 109). W szczególności, większość istniejących opracowań tego rodzaju, wydanych w języku polskim, skupiała się przede

wszystkim na praktycznym zastosowaniu „mapowania” nauki, metody współcytowań lub metody powiązań bibliograficznych, do analizy konkretnych (fragmentarycznych) korpusów literatury naukowej (zob. np.: Grygiel i in., 2009; Kolasa, 2013; Kowalska & Radzicka, 2012; Osińska, 2010; Osińska, 2012; Potkański, 2010; Ważna & Bartkowski, 2002).

Podsumowując powyższy przegląd, warto zwrócić uwagę po pierwsze na bardzo dużą liczbę metodologicznych podejść do problemu prognozowania rozwoju dyscyplin naukowych oraz na fakt, że różne, konkurencyjne metody są często używane przez naukowców bez wzajemnego nawiązywania do siebie, w izolacji (Tseng et al., 2009, 75). Do tej pory, w obrębie samego tylko nurtu „mapowania” i wizualizowania rozwoju dyscyplin zaproponowano już ponad 20 różnych metod tworzenia map nauki (Yan & Yu, 2016, 1953). Niemal tak samo zróżnicowane i liczne są kryteria podziału zbioru publikacji zasadzające się na ich zakładanym podobieństwie oraz możliwe poziomy agregacji danych doświadczalnych, o których wspomniano już w zakończeniu części pierwszej niniejszego artykułu w ramach podsumowania głównych zarzutów w stosunku do omówionych w niej bibliometrycznych metod prognostycznych. Wobec tak dużego wyboru możliwych technik realizacji badań empirycznych, bardzo trudne (jeżeli nie niemożliwe) jest stwierdzenie, która z dostępnych metod lub ich kombinacji przynosi najlepsze rezultaty z uwagi na zdolność odzwierciedlenia rzeczywistych wzorców rozwoju nauki i która z metod zapewnia największą precyzję wyodrębniania nowo powstałych specjalności. W tym kontekście należy też zwrócić uwagę na niewielką liczbę badań porównawczych i na znaczną rozbieżność ich wyników (zob. podrozdział drugi). W efekcie, w dalszym ciągu pozostaje otwarte pytanie o to, która z metod mapowania nauki jest najlepsza pod względem zdolności wczesnej detekcji frontu badawczego (zob. np.: Huang, 2008, 49; Huang & Chang, 2015, 2043). W związku z tym zachodzi potrzeba dalszego udoskonalania istniejących na gruncie naukometrii metod prognostycznych.

Metody stosowane w omówionym obszarze cechuje ponadto wysoka interdyscyplinarność. Metody te modyfikuje się i zapożycza głównie z dyscyplin ścisłych takich, jak matematyka, informatyka, statystyka matematyczna czy teoria decyzji oraz (rzadziej) z dyscyplin społeczno-ekonomicznych (np. demografii, ekonometrii). Zgadając się z twierdzeniami niektórych autorów o interdyscyplinarności jako „sile napędowej” innowacyjności, można zająć stanowisko, według którego naukometryczne „mapowanie” i prognozowanie rozwoju dyscyplin jest samo w sobie perspektywiczną specjalnością przeżywającą aktualnie okres swojego rozkwitu. Najprawdopodobniej będzie ona w dalszym ciągu rozwijać się w szybkim tempie za sprawą jej praktycznej przydatności oraz angażowania się w nią autorów pracujących w różnych dyscyplinach nauki, tak ścisłych, jak i społecznych, humanistycznych i ekonomicznych.

Kolejnym wnioskiem płynącym z powyższego przeglądu literatury jest spostrzeżenie, które niejako uzupełnia ogólną krytykę metod prognostycznych wypracowanych na gruncie bibliometrii. Spostrzeżenie to zawiera się w przeświadczeniu, że w metodologii zaczerpniętej ze statystyki matematycznej zawarty jest znaczny potencjał, chociaż wydaje się on być dotąd nie w pełni wykorzystany. Świadczy o tym niewielka liczba publikacji, w których autorzy posługiwali się *stricte* analizą szeregów czasowych. W szczególności zauważalne jest, że w przeciwieństwie do metody tworzenia klastrów publikacji i analizy głównych ścieżek rozwoju dyscyplin, statystyka matematyczna jako jedyna stwarza możliwość szacowania niepewności prognoz, chociaż żaden z autorów prac streszczonych w podrozdziale 6 z niej

nie skorzystał. Ponadto wydaje się, że wyodrębnienie klastrów nie jest mimo wszystko dokładnie tym samym, co prognozowanie w ścisłym sensie. W kategoriach analiz klastrów publikacji nie podaje się bowiem konkretnych, ilościowych wartości jakichkolwiek wskaźników, które można byłoby uznać za np. przewidywaną liczbę cytowań, jakie otrzymają kluczowe dokumenty, patenty czy czasopisma. Nie określa się też precyzyjnie (liczbowo) tempa rozwoju dyscyplin w postaci np. funkcji matematycznej lub prawdopodobieństwa, że specjalność albo obszar badawczy osiągną maksimum swojej siły wpływu np. w ciągu pięciu lat. Wymiar prognostyczny metod grupowania zbioru publikacji (wydzielania klastrów) polega raczej na przyjmowanym *implicite* założeniu, że klastry najaktywniejsze w chwili badania, wykazujące się najwyższą cytawalnością lub największą liczbą publikacji, utrzymają swoją wiodącą pozycję także i w (bliżej nieokreślonej) przyszłości oraz, że to na nich będą koncentrować się dalsze badania, rozwijając je, modyfikując, testując, poszukując dla nich zastosowań praktycznych i wykorzystując osiągnięte już dotychczas wyniki w nowatorski sposób. W przypadku statystycznej analizy szeregów czasowych możliwe jest natomiast nie tylko dokonanie oceny niepewności prognozy, ale też podanie konkretnych liczb, które sygnalizują np. to, jak będzie kształtować się przyszła cytawalność zbioru publikacji dyscyplinarnych czy, alternatywnie, ich liczba, jako tempo produkcji piśmienniczej. W związku z tym wydaje się, że można postawić tezę, według której metody wyróżniania klastrów publikacji oraz identyfikacji głównych ścieżek rozwoju dyscyplin, jako reprezentantów frontów badawczych, służą w większym stopniu ocenie dynamiki rozwoju dyscyplin, podczas gdy ich aspekt prognostyczny jest drugoplanowy. W przypadku metod statystyki matematycznej mamy natomiast do czynienia z sytuacją przeciwną – aspekt prognostyczny dominuje tutaj nad oceną aktualnego stanu rozwoju dyscyplin.

Ostatnim wnioskiem, jaki można wysnuć w oparciu o zaprezentowany przegląd literatury, jest widoczny w niej podział na badania prognostyczne ukierunkowane na analizę cytowań oraz badania prognostyczne ukierunkowane na analizę „surowych” liczb publikacji naukowych przyporządkowanych do określonych specjalności naukowych. Wspólnym mianownikiem wszystkich badań polegających na analizie cytowań jest konieczność akceptacji też tzw. paradygmatu normatywnego funkcjonującego w ramach socjologii wiedzy naukowej. Główne założenia tego paradygmatu sformułowała m.in. Linda Smith (1981). Brzmiały one następująco:

- (1) Cytowanie dokumentu implikuje jego faktyczne wykorzystanie przez naukowca. To znaczy, że przywołuje on wszystkie (lub przynajmniej większość) prace przezeń wykorzystane oraz wszystkie prace przywołane rzeczywiście były wykorzystane. Inaczej mówiąc, naukowiec ani nie cytuje prac nieużywanych, ani nie pomija tych wykorzystanych.
- (2) Cytowanie odzwierciedla zalety dokumentu cytowanego, jego jakość, znaczenie, siłę wpływu.
- (3) Cytuje się prace najlepsze z możliwych.
- (4) Istnieje związek treściowy pracy cytującej i cytowanej.
- (5) Wszystkie cytowania są równorzędne (o równej wartości) (Smith, 1981, 87–89).

Założenia te były przedmiotem krytyki ze strony tzw. paradygmatu konstruktywistycznego, w którym akcentuje się konceptualizację cytowań jako aktów posiadających rozliczne uwarunkowania socjologiczne i psychologiczne, niezwiązane tylko z bezinteresownym dążeniem do spłaty intelektualnego długu wobec poprzedników. Jako takie, cytowania są

w świetle konstruktywizmu tworem podporządkowanym bieżącym potrzebom cytującego. Kiedy używa on pracy cudzej do swoich celów, może dostosowywać, zniekształcać, nadinterpretować znaczenie cytowanych ustępów, naginać ich sens do swoich potrzeb, konstruując go na nowo w postaci odpowiadającej jego intencjom i w oderwaniu od pierwotnego zamysłu naukowca cytowanego. Zrozumienie natury, celu i funkcji cytowań wymaga więc uwzględnienia złożonych sytuacji, w jakich powstają publikacje. Obejmują one codzienną rutynę i okoliczności pracy naukowców, osobiste motywacje, interakcje ze współpracownikami, przełożonymi i instytucjami itd. (zob. np.: Law & French, 1974, 589; Moed, 2005, 196). Kwestia zasadności i wiarygodności metod opartych na analizie cytowań, takich jak grupowanie publikacji w klastry, analizowanie ścieżek rozwoju dyscyplin i statystyczne przewidywanie przyszłej cytawalności, spoczywa więc ostatecznie na indywidualnych zachowaniach badaczy w zakresie cytowań (ang. *citation behaviour*). W tym sensie epistemologiczną podstawą całego nurtu prognostycznego w bibliometrii jest socjologia wiedzy naukowej, w ramach której toczy się dyskusja nad tym, który ze wspomnianych paradygmatów trafniej oddaje rzeczywistą rolę, jaką w nauce odgrywają cytowania piśmiennictwa.

Z drugiej strony, niektóre zarzuty wobec polegania na „surowej” liczbie publikacji w dyscyplinie, jako na zbyt mało wyrafinowanym wskaźniku tempa rozwoju nauki, częściowo przedstawiono już w podrozdziale 6. W tym miejscu można natomiast dodać, że jeszcze jedną trudnością związaną z tym podejściem jest problem w precyzyjnym wytyczeniu granic poszczególnych dyscyplin. Problem ten wiąże się bezpośrednio z dyskutowanym w informatologii zagadnieniem relewancji tematycznej publikacji naukowych (zob. np.: Hjørland, 2000; Saracevic, 1975). Oba podejścia do prognozowania rozwoju dyscyplin okazują się zatem nie tylko bardzo złożone, ale też możliwe do zakwestionowania w perspektywie teoretycznej i zaangażowane w debatę pozornie tylko odległą od swojego przedmiotu badań. Stanowi to dodatkowy, głębszy wymiar metodologii przewidywania ewolucji dyscyplin naukowych.

## Bibliografia

- Aczel, A.D. (2007). *Statystyka w zarządzaniu*. Warszawa: PWN.
- Bingham, N.H., Fry, J.M. (2010). *Regression. Linear Models in Statistics*. London: Springer-Verlag.
- Blei, D.M., Lafferty, J.D. (2007). A Correlated Topic Model of Science. *The Annals of Applied Statistics*, 1(1), 17–35.
- Blondel, V.D., Guillaume, J-L., Lambiotte, R., Lefebvre, E. (2008). Fast Unfolding of Communities in Large Networks [online]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, 1–12, [30.08.2017], <http://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/meta>
- Bornmann, L., Daniel, H-D. (2008). What Do Citation Counts Measure? A Review of Studies on Citing Behavior. *Journal of Documentation*, 64(1), 45–80.
- 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.
- 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.
- Chi, Y., Tseng, B.L., Tatemura, J. (2006). Eigen-Trend: Trend Analysis in the Blogosphere Based on Singular Value Decompositions [online]. *CIKM, 06, Proceedings of the 15th ACM International*



- Conference on Information and Knowledge Management, Arlington, Virginia, USA, November 06–11, 2006, 68–77, [30.08.2017], <http://dl.acm.org/citation.cfm?id=1183628>
- Cipra, T., Hanzák, T. (2008). Exponential smoothing for irregular time series. *Kybernetika*, 44(3), 385–399.
- Choudhury, N., Uddin, S. (2016). Time-Aware Link Prediction to Explore Network Effects on Temporal Knowledge Evolution. *Scientometrics*, 108(2), 745–776.
- Daim, T., Rueda, G., Martin, H., Gerdri, P. (2006). Forecasting Emerging Technologies: Use of Bibliometrics and Patent Analysis. *Technological Forecasting and Social Change*, 73(8), 981–1012.
- Debackere, K., Verbeek, A., Luwel, M., Zimmermann, E. (2002). Measuring Progress and Evolution in Science and Technology – II: The Multiple Uses of Technometric Indicators. *International Journal of Management Reviews*, 4(3), 213–231.
- 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, K., 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.
- Glänzel, W. (2012). Bibliometric Methods for Detecting and Visualizing Emerging Research Topics. *El Profesional de la Información*, 21(1), 194–201.
- Glänzel, W., Thijs, B. (2012). Using ‘Core Documents’ for Detecting and Labelling New Emerging Topics. *Scientometrics*, 91(2), 399–416.
- Grygiel, P., Rębisz, S., Humenny, G. (2009). Analiza bibliometryczna jako narzędzie badania efektywności nauczycieli akademickich na przykładzie Uniwersytetu Rzeszowskiego. *Zarządzanie Publiczne. Zeszyty Naukowe Instytutu Spraw Publicznych Uniwersytetu Jagiellońskiego*, 3(7), 65–84.
- Guo, H., Weingart, S., Börner, K. (2011). Mixed-Indicators Model for Identifying Emerging Research Areas. *Scientometrics*, 89(1), 421–435.
- Haegeman, K., Marinelli, E., Scapolo, F., Ricci, A., Sokolov, A. (2013). Quantitative and Qualitative Approaches in Future-Oriented Technology Analysis (FTA): From Combination to Integration? *Technological Forecasting and Social Change*, 80(3), 386–397.
- Hjørland, B. (2000). Relevance Research: The Missing Perspective(s): ‘Non-Relevance’ and ‘Epistemological Relevance’. *Journal of the American Society for Information Science*, 51(2), 209–211.
- Hood, W.W., Wilson, C.S. (2002). Solving Problems in Library and Information Science Using Fuzzy Set Theory. *Library Trends*, 50(3), 393–405.
- 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](https://nzcsrsc08.canterbury.ac.nz/site/proceedings/NZCSRSC_2008_Proceedings.pdf)
- 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.
- Jun, S., Park, S.S., Jang, D.S. (2012). Technology Forecasting Using Matrix Map and Patent Clustering. *Industrial Management and Data Systems*, 112(5), 786–807.
- Klafter, J., Sokolov, I.M. (2015). *First Steps in Random Walks*. Oxford: Oxford University Press.
- Kolasa, M.W. (2013). *Historiografia prasy polskiej (do 1918 roku)*. *Naukometryczna analiza dyscypliny 1945–2009*. Kraków: Wydaw. Naukowe Uniwersytetu Pedagogicznego.

- Kowalska, A., Radzicka, J. (2012). Programy BibExcel i Pajek w analizach bibliometrycznych. *Biuletyn EBIB* [online], 3(130), 1–7, [30.08.2017], [http://www.ebib.pl/images/stories/numery/130/130\\_kowalska\\_radzicka.pdf](http://www.ebib.pl/images/stories/numery/130/130_kowalska_radzicka.pdf)
- Krampen, G., Von Eye, A., Schui, G. (2011). Forecasting Trends of Development of Psychology from a Bibliometric Perspective. *Scientometrics*, 87(3), 687–694.
- Krawiec, S. (2014). *Adaptacyjne modele wygładzania wykładniczego jako instrumenty prognozowania krótkoterminowego zjawisk ilościowych*. Gliwice: Wydaw. Politechniki Śląskiej.
- Krzywda, M. (2010). *Analiza techniczna w praktyce*. Gliwice: Wydaw. Złote Myśli.
- Kucharavy, D., De Guio, R. (2011). Application of S-Shaped Curves. *Procedia Engineering*, 9, 559–572.
- Kuhn, T.S. (2001). *Struktura rewolucji naukowych*. Warszawa: Fundacja Aletheia.
- Kuusi, O., Meyer, M. (2007). Anticipating Technological Breakthroughs: Using Bibliographic Coupling to Explore the Nanotubes Paradigm. *Scientometrics*, 70(3), 759–777.
- Law, J., French, D. (1974). Normative and Interpretive Sociologies of Science. *The Sociological Review*, 22(4), 581–595.
- Lee, L-C., Lee, Y-Y., Liaw, Y-C. (2012). Bibliometric Analysis for Development of Research Strategies in Agricultural Technology: The Case of Taiwan. *Scientometrics*, 93(3), 813–830.
- Leydesdorff, L. (2007). Betweenness Centrality as an Indicator of the Interdisciplinarity of Scientific Journals. *Journal of the American Society for Information Science and Technology*, 58(9), 1303–1319.
- Liu, M. (1993). The Complexities of Citation Practice: A Review of Citation Studies. *Journal of Documentation*, 49(4), 370–408.
- Liu, P., Chen, B-L., Liu, K., Xie, H. (2016). Magnetic Nanoparticles Research: A Scientometric analysis of Development Trends and Research Fronts. *Scientometrics*, 108(3), 1591–1602.
- Liu, C., 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.
- Łapkowska-Baster, B. (2009). *Miary współzależności i dynamiki zjawisk w statystyce opisowej. Przykłady i zadania*. Kraków: Wydaw. UJ.
- 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.
- Malina, A. (1994). Prognozowanie zjawisk ekonomicznych w oparciu o metody wykładniczego wygładzania szeregów czasowych. *Zeszyty Naukowe/Akademia Ekonomiczna w Krakowie*, 440, 15–29.
- Miyamoto, S. (1990). *Fuzzy Sets in Information Retrieval and Cluster Analysis*. Dordrecht: Springer Science+Business Media B.V.
- Moed, H. (2005). *Citation Analysis in Research Evaluation*. Dordrecht: Springer.
- Montgomery, D.C., Jennings, C.L., Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. Hoboken: John Wiley & Sons, Inc.
- Morris, S.A. (2005). Manifestation of Emerging Specialties in Journal Literature: A Growth, Model of Papers, References, Exemplars, Bibliographic Coupling, Cocitation, and Clustering Coefficient Distribution. *Journal of the American Society for Information Science and Technology*, 56(12), 1250–1273.
- Newman, M.E.J. (2004). Fast Algorithm for Detecting Community Structure in Networks. *Physical Review E*, 69(6), art. no. 066133, 1–5.
- Osińska, V. (2010). Rozwój metod mapowania domen naukowych i potencjał analityczny w nim zawarty. *Zagadnienia Informatyki Naukowej*, 2(96), 41–51.
- Osińska, V. (2012). Wizualizacja paradygmatów badawczych. *Zagadnienia Naukoznawstwa*, 3(193), 205–220.
- 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.

- Pindlowa, W. (1994). *Informetria w nauce o informacji. Metody i problemy*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”.
- Potkański, J. (2010). Samoobserwacja współczesnej nauki: spojrzenie liczb. *Przegląd Humanistyczny*, 05/06, 137–150.
- Price, D. de Solla (1967). *Mała Nauka – Wielka Nauka*. Warszawa: PWN.
- Saracevic, T. (1975). Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science. *Journal of the American Society for Information Science*, 26(6), 321–343.
- 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., 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.
- Shumway, R.H., Stoffer, D.S. (2011). *Time Series Analysis and Its Applications: With R Examples*. New York: Springer.
- Sitarz, R., Heneczkowski, M., Jabłońska-Sabuka, M., Krasławski, A. (2012). *Trend Analysis Methods as an Indicator of the Dynamic of Knowledge Flow in Research on Polymeric Materials: Composites, Nanocomposites and Blends* [online]. Extended version of a paper presented during 11th International Symposium on Process Systems Engineering, 15–19 July 2012, Singapore, [30.08.2017], [https://www.researchgate.net/profile/Matylda\\_Jablonska-Sabuka/publication/266142586\\_Clustering\\_Method\\_for\\_Analysis\\_of\\_Research\\_Fields\\_Examples\\_of\\_Composites\\_Nanocomposites\\_and\\_Blends/links/555c3f2d08aec5ac2232aec1.pdf](https://www.researchgate.net/profile/Matylda_Jablonska-Sabuka/publication/266142586_Clustering_Method_for_Analysis_of_Research_Fields_Examples_of_Composites_Nanocomposites_and_Blends/links/555c3f2d08aec5ac2232aec1.pdf)
- Small, H., Boyack, K.W., Klavans, R. (2014). Identifying Emerging Topics in Science and Technology. *Research Policy*, 43(8), 1450–1467.
- Smith, L.C. (1981). Citation Analysis. *Library Quarterly*, 30(1), 83–106.
- Snarska, A. (2011). *Statystyka. Ekonometria. Prognozowanie. Ćwiczenia z Excelem 2007*. Warszawa: Placet.
- Sobczyk, M. (2008). *Prognozowanie. Teoria, przykłady, zadania*. Warszawa: Placet.
- Sobczyk, M. (2015). *Statystyka*. Warszawa: PWN.
- 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.
- Taylor, J.R. (2012). *Wstęp do analizy błędu pomiarowego*. Warszawa: PWN.
- Tseng, Y-H., Lin, Y-I., Lee, Y-Y., Hung, W-C., Lee, C-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.
- Wang, C-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.
- Ważna, L., Bartkowski, A. (2002). Patenty i publikacje naukowe jako wskaźniki technologiczne. Bibliometria strukturalna. *Management*, 6(2), 171–182.
- Yan, E., Yu, Q. (2016). Using Path-Based Approaches to Examine the Dynamic Structure of Discipline-Level Citation Networks: 1997–2011. *Journal of the Association for Information Science and Technology*, 67(8), 1943–1955.
- Yoon, B., Park, Y. (2007). Development of New Technology Forecasting Algorithm: Hybrid Approach for Morphology Analysis and Conjoint Analysis of Patent Information. *IEEE Transactions on Engineering Management*, 54(3), 588–599.
- Zeliaś, A., Pawełek, B., Wanat, S. (2013). *Prognozowanie ekonomiczne. Teoria, przykłady, zadania*. Warszawa: PWN.
- Zheng, T., Wang, J., Wang, Q., Nie, C., 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.
- 

## **Bibliometric Methods to Foresee and Assess the Development of Scientific Disciplines. Literature Analysis. Part 2. Comparisons, Hybrid and Statistical Methods, Analysis of Patents and Main Paths of Literature Development and Other Original Approaches in Terms of Predictive Methodology**

### **Abstract**

**Purpose/Thesis:** The aim of the paper is to provide an evaluative review of literature on the quantitative methods of bibliometric forecasting and assessing the current state of scientific discipline 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 application 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:** First conclusion was that there was a great diversity within the sphere of quantitative methods concerning the subject of evolution forecasting for scientific disciplines. The methods in question often are applied without referring to each other. In other words, the majority of predictive methods are used in isolation. The flaws, occurring in many methods, usually have a negative influence on the results obtained. Next conclusion was that there was a significant potential in the statistical analysis of the time series. However, it seems that this potential has not been fully exploited yet. Furthermore, there is a considerable degree of interdisciplinarity in the scope of bibliometric predictive methods. Often, they are modified and adopted from other disciplines such as mathematics, computer science, statistics and econometrics. The last conclusion was that

the predictive process was based either on purposefully designed citation analysis or solely on the number of disciplinary publications. In the first case the methodology must implicitly accept the theses of so-called normative paradigm, existing within the sociology of scientific knowledge and constituting an epistemological background for the methodology. In the latter case the methodology is engaged in the debate over topical relevance, which is observed within the frames of library and information science.

**Originality/Value:** The originality and value of the paper lies in an effort to demonstrate the scope and a high degree of progress of the studies on the possibilities of predicting and assessing the development of scientific disciplines worldwide. In 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 selected 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 considerable need for such refinements due to the fact that currently there is no perfect method which could be described as "the last word" on the technique of anticipating future research directions.

### Keywords

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

---

*LUKASZ 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 Dziedziny w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego w 2012 r. Pracuje w Oddziale Informacji Naukowej Biblioteki Politechniki Rzeszowskiej na stanowisku bibliotekarza. Najważniejsze publikacje: Ł. Opaliński (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>; Ł. Opaliński (2013). Wybrane aspekty metodologii badań cyklu życiowego publikacji naukowych. *Przegląd Biblioteczny*, 81(2), 152–171; Ł. Opaliński, M. Jaromin, J. Wikiera (2015). Problem stabilności zachowań naukowców w zakresie cytowań w kontekście metodologii badań starzenia się publikacji naukowych i możliwość jego ujęcia ilościowego. *Zagadnienia Informacji Naukowej*, 53(2/106), 65–83.

Kontakt z autorem:

[lopa@prz.edu.pl](mailto:lopa@prz.edu.pl)

Oddział Informacji Naukowej Biblioteki Politechniki Rzeszowskiej

al. Powstańców Warszawy 12

35–959 Rzeszów