



Marek Jerzy Minakowski
Dr Minakowski Publikacje Elektroniczne
email: mj@minakowski.pl

Modelowanie rozkładu kapitału ekonomiczno-społeczno-kulturowego poprzez genealogię masową

Modelling of the Distribution of Economic, Social and Cultural Capital through Massive Genealogy

Abstrakt

Sprzyjający zbieg okoliczności umożliwia powstanie nowej dyscypliny badawczej na pograniczu genealogii i demografii historycznej. Wykorzystując narzędzia analizy sieciowej, odpowiada ona na pytania socjologii i historii społecznej. W artykule przedstawiono przykład zastosowania genealogii masowej do analizy żywej kwestii socjologicznej: jak się rozkłada i jak jest przekazywany kapitał kulturowy i społeczny. Pokazano, jak można mierzyć pewną wagę, która wydaje się być miarą owych kapitałów, a przynajmniej tej ich składowej, która jest przekazywana przez środowisko rodzinne. Konkretny przykłady przedstawione w artykule to wyliczone na podstawie położenia danej osoby w wielkiej sieci rodzinnej takie wielkości, jak szansa bycia bohaterem artykułu w *Polskim słowniku biograficznym*, oraz szansa, że (o ile się zmarło w latach

Abstract

A coincidence has made it possible to create a new academic discipline in the borderland of genealogy and historical demography. Using the Social Network Analysis (SNA) the new discipline answers the questions of sociology and social history. The article presents an example of applying massive genealogy to an essentially sociological analysis: what the distribution of the cultural and social capital is like and how it is transmitted. In addition, the article shows how it is possible to measure a certain weight which seems to be a yardstick of the capital in question or at least of the component that is transmitted through families. The examples presented in the article are quantities calculated on the basis of the situation of a person in a big family network such as the chance of being a hero of an article in the *Polish Biographical Dictionary* or the

1821–1861) było się wspomnianym w „Kurierze Warszawskim”. Zanalizowano także prestiż społeczny różnych zawodów w XIX i XX wieku (na podstawie analizy nagłówków biogramów w *Polskim słowniku biograficznym*) oraz różnice we wpływie przodków ojczystych i macierzystych na to, czy ich dziecko zostanie członkiem akademii nauk.

chance – if the person died between 1821 and 1861 – of being mentioned in *Kurier Warszawski*. The article also contains an analysis of social prestige of various professions in the 19th and 20th centuries (based on the analysis of the headlines of entries in the Polish Biographical Dictionary) and the presentation of differences in the impact of paternal and maternal ancestors on the chance of their child becoming a member of the Academy of Sciences.

Na pograniczu demografii i genealogii znajduje się pewna dyscyplina (subdyscyplina ogólnie pojętej socjologii), którą określić można jako „genealogię masową”. Nie była ona dotąd znana, bo ku jej uprawianiu nie było warunków. Te wystąpiły dopiero teraz, wraz z pojawieniem się *big data* (działu informatyki zajmującego się obróbką wielkich baz danych, w szczególności danych o charakterze sieci) oraz dostępnością masowych źródeł genealogicznych, zwłaszcza akt metrykalnych, których opracowane indeksy dla samej tylko XIX-wiecznej Polski liczą już dziesiątki milionów pozycji.

W niniejszym artykule przedstawiono, jak tę nową dyscyplinę można zastosować do analizy ważnego tematu socjologicznego: dynamiki rozkładu kapitału społecznego i kulturowego w czasie i przestrzeni.

Jako że genealogia masowa jest nową dyscypliną i nie doczekała się jeszcze zasad metodologicznych, artykuł jest okazją, by je na przykładach zaproponować i przedstawić do dyskusji. Autorowi nie są znane publikacje naukowe na temat genealogii masowej. W ogóle baz danych, które mogłyby być podstawą do takich analizy jest niewiele, a z ważniejszych wymienić można trzy: dwa serwisy będące własnością izraelskiej firmy MyHeritage (serwis MyHeritage i przejęty przezeń serwis Geni.com), oraz francuski serwis Roglo, stworzony oryginalnie przez Daniela de Rauglaudre’a. Jednak wszystkie te bazy tworzone są przez wolontariuszy, jako serwisy społecznościowe (tzw. *user-generated content*) – często na podstawie oryginalnych dokumentów, ale bez kontroli naukowej. Ponadto w ograniczonym tylko zakresie dotyczą Europy Środkowej (MyHeritage i Geni są ogólnosiątkowe, zaś Roglo dotyczy głównie krajów frankofońskich)¹.

¹ Serwis Roglo (<http://roglo.eu/roglo>), zawierający obecnie dane o 6,9 mln osób, byłby bardzo ciekawy do takich analiz, jednak na każdej stronie znajduje się zastrzeżenie: „Very important warning! This website is not an authorised reference source and may contain errors. Information

Praca niniejsza jest jednocześnie zaproszeniem do szerokich badań interdyscyplinarnych na temat dawnych elit polskich ze szczególnym skupieniem uwagi na inteligencji XIX i początków XX wieku, jej przodkach, potomkach i kuzynach. Zgromadziwszy dane na temat sieci rodzinnej tej grupy, liczącej do 1% społeczeństwa polskiego², ale decydującej o jego losach³, będziemy mogli znacznie lepiej zrozumieć i przedstawić historię Europy Środkowej. Aby jednak dobrze wykorzystać te dane, potrzebna będzie współpraca specjalistów z różnych dziedzin – socjologów i historyków z jednej strony, a matematyków i informatyków analizujących struktury sieciowe – z drugiej. Humanistów i przedstawicieli nauk ścisłych. Dlatego też niniejszy artykuł napisany jest tak, by w miarę możliwości zarówno jedni, jak i drudzy mogli go zrozumieć bez dodatkowych studiów uzupełniających. Pewnym mankamentem zapewne jest to, że jego fragmenty socjologiczne mogą być uznane za naiwne przez socjologów, a matematyczne za naiwne przez matematyków, jednak jest to – jak się wydaje – koszt, który warto ponieść⁴.

Kapitał ekonomiczno-społeczno-kulturowy (według Pierre'a Bourdieu)

Francuski socjolog Pierre Bourdieu (1930–2002) przeformułował klasyczne pojęcie kapitału w sposób taki, że jego składnik ekonomiczny jest tylko jedną z kilku form⁵. Wedle tej koncepcji kapitał to ogół pracy zakumulowanej przez

is provided herein without any guarantee. Please verify it against reliable references before use. We cannot be held responsible for any errors that may be contained herein”.

² Według *Malego rocznika statystycznego 1939* w roku akademickim 1936/37 dyplomy ukończenia szkół wyższych uzyskało w Polsce 6114 osób (4417 mężczyzn i 1697 kobiet). Osób w wieku 15–19 lat było wówczas 3036 tys., co daje średnio 607 tys. osób w roczniku. We wcześniejszych pokoleniach liczba inteligentów była oczywiście znacznie mniejsza: *Rocznik statystyczny Królestwa Polskiego* za rok 1913 podaje następujące liczby: wszystkich uczniów szkół średnich 26 209 (wszystkich klas, dane za rok 1912), zaś słuchaczy wyższych zakładów naukowych rządowych: 2613 w 1910 r. i wyższych kursów prywatnych 1423 w roku 1913. Liczba ludności Królestwa Polskiego wynosiła 12 467 tys. (wg stanu na 1.01.1911) – a zatem uczniowie wszystkich szkół średnich i wyższych Kongresówki stanowili ok. 0,24% jej populacji. Tymczasem w roku 1937/38 łączna liczba uczniów szkół średnich ogólnokształcących i wyższych liczyła 249 tys., co stanowiło 0,71% populacji – mamy więc wzrost trzykrotny w ciągu jednego pokolenia.

³ Szerzej o procesie przejścia przez inteligencję władzy z rąk ziemiaństwa i burżuazji po 1918 r. („rewolucji inteligenckiej”) piszą Rafał Smoczyński i Tomasz Zarycki w książce *Totem inteligencji. Aristokracja, szlachta i ziemiaństwo w polskiej przestrzeni społecznej* (Warszawa: Wydawnictwo Naukowe Scholar, 2017).

⁴ Autor może o sobie powiedzieć, że jego właściwą specjalnością jest sama *interdyscyplinarność* – już bowiem w 1998 r. obronił na Uniwersytecie Jagiellońskim doktorat na przecięciu filologii klasycznej, historii starożytnej i logiki matematycznej (rozprawa pt. *Logika formalna przed Arystotelesem*). Potem jednak zamienił komputerowo wspomaganą formalizację tekstów starogreckich na taką analizę materiałów dotyczących historii Polski.

⁵ Znajomość teorii kapitałów Pierre'a Bourdieu zawdzięczam prof. Tomaszowi Zaryckiemu z Instytutu Studiów Społecznych Uniwersytetu Warszawskiego, który jako kanoniczny tekst

daną osobę, przy czym akumulacja może następować zarówno w postaci materialnej (kapitał ekonomiczny), jak też w postaci pozycji społecznej (kapitał społeczny) i „duchowej” (kapitał kulturowy). Kapitał ten może być zmaterializowany w postaci trwałych dóbr (kapitał ekonomiczny), wyznaczników statusu (kapitał społeczny) i różnorakich certyfikatów, jak np. dyplom doktoratu (kapitał kulturowy). Nośnikiem kapitału może też być sam człowiek, co najbardziej widoczne jest w przypadku kapitału społecznego (gdzie jest to ogół znajomości, sławy czy szacunku) albo kapitału kulturowego (gdzie będzie to wiedza, zarówno w postaci formalnej – wykształcenie – jak i praktycznej – *savoir vivre*).

Kapitały podlegają konwersjom. Mając środki ekonomiczne, można zakupić zarówno wyznaczniki statusu społecznego (np. luksusowy samochód czy jacht), jak i przełożyć je na kapitał kulturowy (wydać na edukację, książki, wizyty w teatrze). Z kolei mając wiedzę (kapitał kulturowy) i znajomości (kapitał społeczny), można je przekładać na kapitał ekonomiczny (dobrze płatną pracę, lukratywne kontrakty). Podobnie dzięki wiedzy można podnieść swą pozycję społeczną, a wyższa pozycja społeczna ułatwia zdobywanie kapitału kulturowego.

Kapitał (w każdej formie) można tworzyć (jest to wszak zakumulowana praca), można go jednak również bezpowrotnie stracić. Dobra materialne mogą spłonąć, znajomi mogą umrzeć bądź możemy ich do siebie zrazić, zaś wiedzę można zapomnieć.

Relatywna wartość kapitału również się zmienia: wartości ekonomiczne podlegają wahaniom cen, wyznaczniki pozycji społecznej mogą się zdevaluować lub przeciwnie – osoby poznane kiedyś przypadkowo mogą nagle stać się bardzo ważne; podobnie wiedza bezużyteczna może stać się bardzo cenna, a wiedza rzadka może stać się powszechna (a więc mało wartościowa).

Jako że mowa jest o teorii socjologicznej, najważniejsze jest to, że kapitały otrzymujemy od innych osób i innym osobom je przekazujemy. Możemy majątek odziedziczyć i zapisać w testamencie; rodzimy się w jakiejś rodzinie i kolejną rodzinę sami zakładamy, a kiedyś założą je nasze dzieci. Wiedzę (kapitał

wskazał artykuł: Pierre Bourdieu, „The Forms of Capital”, w: *Handbook of Theory and Research for Sociology of Education*, red. John C. Richardson, (New York–Westport, CT–London: Greenwood Press, 1986), 117–142. Ponadto istotne są artykuły: Tomasz Zarycki, „Kapitał kulturowy – założenia i perspektywy zastosowań teorii Pierre’a Bourdieu”, *Psychologia Społeczna* 4 (2009), 1–2: 12–25 oraz Marek Ziółkowski, „Kapitały społeczny, kulturowy i materialny i ich wzajemne konwersje we współczesnym społeczeństwie polskim”, *Studia Edukacyjne* 22 (2012): 7–27.

Teoria kapitałów ma silne odniesienia do społeczeństwa polskiego, a ściślej do szczególnej roli warstwy inteligenckiej. Pierwszy na grunt środkowoeuropejski przeniósł ją Węgier Iván Széleányi, zob. Gil Eyal, Iván Széleányi, Eleanor Townsley, *Making Capitalism Without Capitalists: The New Ruling Elites in Eastern Europe* (London–New York: Verso, 1998). Bardzo ważną pozycją odnośnie do inteligencji polskiej jest artykuł: Tomasz Zarycki, Tomasz Warczok, „Hegemonia inteligentka: kapitał kulturowy we współczesnym polu władzy – perspektywa «długiego trwania»”, *Kultura i Społeczeństwo* 4 (2014): 27–49.

kulturowy) otrzymujemy z wychowaniem, a potem wychowujemy i uczymy innych. Możemy wreszcie naszym majątkiem ekonomicznym sfinansować zdobycie wiedzy (kapitału kulturowego) naszych dzieci.

Wyżej zarysowana teoria zawiera tyle zmiennych, że bardzo trudno ją poddać empirycznemu, socjologicznemu testowi – zawsze można sobie wyobrazić jakieś ukryte czynniki, które zaburzają wynik pomiaru. Dlatego dla celów analizy problem warto uprościć.

Modelowanie struktur sieciowych

W naukach opisujących rzeczywistość przyrodniczą i społeczną często się zdarza, że z konieczności mamy o niej wiedzę bardzo wyrywkową, a musimy coś na podstawie tej ułomnej wiedzy przewidzieć. Nie możemy opisać położenia każdego atomu w kryształach, możemy jednak zbudować jego abstrakcyjny model – a to, co się z nim nie zgadza, nazwiemy „defektami”. Możemy opisać szlaki sezonowych wędrówek ptaków bez liczenia ich wszystkich, możemy próbować przewidzieć kierunek wiatru i wilgotność powietrza za dwa tygodnie albo naprężenia występujące w konstrukcji mostu, do środka którego nie da się zajrzeć. Wszelkie takie przewidywania są zawodne i trzeba się z tym liczyć; możemy nie przewidzieć zmiany pogody, możemy nie przewidzieć, że jakaś belka w moście ma fabryczny defekt, możemy nie przewidzieć anomalii pogodowej, która zmieni trasę przelotu ptaków w danym roku. A jednak nie powstrzymuje nas to przed dalszym prognozowaniem.

Szczególnie podatne na tworzenie modeli są struktury zbudowane na planie sieci. W sieci wiązań kryształu możemy modelować rozkład abstrakcyjnych miar, takich jak temperatura, naprężenie czy ładunek elektryczny. W badaniach bibliometrycznych analizujemy, jak w świecie publikacji naukowych rozchodzi się prestiż autora i czasopisma (przekazywany za pomocą cytowania). Jako odmiana bibliometrii powstała nowa dyscyplina badań, o gigantycznym wpływie na współczesną gospodarkę, jaką jest *rankowanie* stron internetowych na podstawie wzajemnych odesłań (linków) – na czym opiera się sukces wyszukiwarki Google i jej algorytmu PageRank. Z kolei pochodną tego ostatniego są algorytmy (umowny EdgeRank), które decydują o tym, kto jakie treści zobaczy na stronach serwisu Facebook.

Teoria grafów, która jest abstrakcyjną formą opisu dowolnych sieci, rozwija się dynamicznie wraz z narzędziami informatycznymi, które do takiej analizy służą (tzw. *big data*).

Genealogia masowa jako osobna dyscyplina

Genealogia masowa, choć bliska demografii historycznej, pod pewnym względem różni się od niej zasadniczo. W demografii możemy analizować wybraną populację w całości (operując np. na danych ze spisów ludności) bądź przez reprezentatywne próbkowanie. W obu przypadkach wnioskować będziemy o całości badanej populacji. Wydawałoby się, że genealogia masowa jest przedłużeniem na ciągi wielopokoleniowe znanej w demografii metody rekonstrukcji rodzin.

Teoretycznie można by zrekonstruować genealogię pewnej populacji w całości, wyizolowawszy ją z otoczenia, i na takiej podstawie wyciągać ogólne wnioski. W pewnych granicznych przypadkach może to się udać – i taki przykład przedstawiono w artykule „Gęstość sieci koligacji w parafii Kampinos w XIX wieku”⁶. Mając do dyspozycji komplet akt metrykalnych obejmujących ponad sto lat, dokonano rekonstrukcji powiązań rodzinnych całej ludności tej jednej parafii. Sto lat to jednak okres krótki – ledwie trzy lub cztery pokolenia; rejestracja metrykalna na terenie Polski przed wiekiem XIX jest zbyt mało informatywna (przednapoleońskie akta metrykalne zawierają zwykle zbyt mało danych, by wystarczały do wiarygodnej rekonstrukcji ciągów genealogicznych, co najwyżej mogą potwierdzać fakty znane skądinąd), a księgi zbyt przetrzebione, by móc cofać się z rekonstrukcją dłuższych ciągów genealogicznych. Inne źródła (jak księgi grodzkie i ziemskie) są nie dość że przetrzebione, to jeszcze w dodatku wyrwykowe (możemy na ich podstawie zrekonstruować wiele faktów genealogicznych, o ile akurat tak się złożyło, że wiązały się one z czynnościami dokonywanymi w sądzie, jak umowy małżeńskie czy spory o spadek; jednak ze względu na brak właściwości miejscowej sądów i brak obowiązku załatwiania takich spraw w sądzie, z milczenia źródła nie możemy o niczym wnioskować; w dodatku takie typy zachowanych źródeł dotyczą głównie szlachty i patrycjatu miejskiego). Z kolei dane XX-wieczne są bardzo trudno dostępne ze względu na prawną ochronę danych osobowych.

Wreszcie sam badany przedmiot bardzo utrudnia swe badanie: nawet w grupach mało mobilnych (jak chłopci pańszczyźniani czy szlachta zagrodowa) w przeciągu wielu pokoleń widzimy, że liczba związków rodzinnych z rodzinami z innych obszarów terytorialnych (parafii, powiatów czy nawet krajów) i grup społecznych (stanów) jest tak duża, że sztucznie zdefiniowana społeczność⁷

⁶ Marek J. Minakowski, „Gęstość sieci koligacji w parafii Kampinos w XIX wieku”, *Przeszłość Demograficzna Polski* (2014), 35: 65–74.

⁷ Przykładową próbę analizy takiej grupy zawarto w Minakowski, „Gęstość sieci”, analizując „rodowitych kampinosiaków”, zdefiniowanych jako osoby ochrzczone w parafii Kampinos z rodziców, którzy zawarli małżeństwo w tejże parafii.

przestanie być reprezentatywna – tym bardziej widać to będzie w przypadku grup, które z natury są mobilne przestrzennie (zawierają związki małżeńskie daleko od miejsca urodzenia) i społecznie (wiążą się z osobami innych klas, warstw, stanów, wyznań czy narodowości). A takie właśnie bywają elity – a więc grupy, których badanie jest najciekawsze z perspektywy socjologa czy historyka⁸.

W genealogii masowej wyrywkowość analizowanego materiału ma więc charakter systemowy, nieunikniony. Zwykle nie da się ani przeanalizować całej populacji na kilkanaście pokoleń wstecz (bo brak jest danych), ani też sensownie wybrać reprezentacyjnej próbki (bo znane lub możliwe do odtworzenia ciągi genealogiczne są z natury niereprezentatywne). A skoro tak, to należy podejść do tego problemu systemowo, korzystając z doświadczeń innych dyscyplin, które z takim problemem się zetknęły.

Algorytm PageRank jako przykład podobnego problemu

W 1998 roku dwaj doktoranci Uniwersytetu Stanforda, Larry Page i Siergiej Brin, przedstawili referat będący podsumowaniem ich prac nad projektem wyszukiwarki internetowej, która będzie sobie dobrze radziła ze wskazaniem najtrafniejszych wyników wyszukiwania mimo spodziewanego wówczas ogromnego wzrostu liczby stron internetowych⁹. Dostępne wówczas wyszukiwarki internetowe starały się pobrać wszystkie zawarte w internecie dokumenty, a wynik wyszukiwania sortowały według reguł logicznych. Mogły one działać sprawnie, gdy stron poświęconych danemu tematowi było niewiele, a zapytanie zostało przez pytającego dobrze skonstruowane. Zupełnie jednak sobie nie radziły w przypadku, gdy zapytania zawierały słowa częste, zwłaszcza pojedyncze: jak sensownie posortować strony zawierające ciąg znaków „auto” w sytuacji, gdy takich stron są miliony, a kiedyś będą ich miliardy lub więcej?

⁸ Konstatacje te opieram na własnym wieloletnim doświadczeniu; czytelnikowi zainteresowanemu tym, co dzieje się, gdy analizujemy długie ciągi genealogiczne, polecić można klasyczne pozycje genealogiczne: Włodzimierz Dworzaczek, *Genealogia* (Warszawa: PWN, 1959); Rafał T. Prinke, Andrzej Sikorski, *Królewska krew. Polscy potomkowie Piastów i innych dynastii panujących* (Poznań: Zysk i S-ka, 1997), wreszcie Rafał T. Prinke, *Poradnik genealoga amatora* (Poznań: Zysk i S-ka, 2007).

⁹ Teoria analizy sieci WWW, która dała początek wyszukiwarce Google, została przedstawiona w artykule: Siergiej Brin, Larry Page, „The Anatomy of a Large-scale Hypertextual Web Search Engine”, *Computer Networks and ISDN Systems* 30 (1998): 107–117. W niniejszych rozważaniach na temat wyszukiwarek WWW korzystam z wiedzy nabytej w czasie, gdy byłem dyrektorem serwisów wyszukiwawczych portalu Onet.pl, kierując najpopularniejszą wówczas polską wyszukiwarką internetową – wiedza ta więc pochodzi w dużej mierze z autopsji. Wprawdzie od tej pory systemy wyszukiwawcze rozwinęły się i skomplikowały, ale fundamentalna zasada pozostała niezmienną.

Wynalazkiem Page'a i Brina było przeniesienie do analizy stron internetowych schematu oceniania jakości prac naukowych według liczby i jakości cytowań. W tym schemacie można śmiało ignorować te artykuły, których nikt nie cytuje; również te publikowane w czasopismach, które rzadko są cytowane; jeżeli te artykuły zignorujemy, nie będziemy wiedzieli o tym, że cytują one innych. Nikt też nie będzie nas winił za to, że nie znamy artykułu, którego nikt nie cytuje. I odwrotnie – jeżeli na pytanie o ogólne słowa kluczowe wskażemy artykuł, który powszechnie jest cytowany w kontekście tych słów, to nasz użytkownik uzna, że dostał to, co powinien był dostać, i będzie zadowolony.

Algorytm PageRank w swym oryginalnym projekcie przyjmował istnienie strony internetowej o najwyższej możliwej wadze: była to strona główna katalogu stron Yahoo! (<http://www.yahoo.com>). Katalog Yahoo! był tworzony ręcznie i znalezienie się w nim oraz zajmowana pozycja były wyrazem *consensusu* jego redakcji na temat tego, że dana strona internetowa jest ważna w swojej dziedzinie. Odpowiednikami cytowań były odsyłacze hipertekstowe (*linki*): jeżeli ważna strona odwoływała się do innych stron, to te inne również otrzymywały jakiś ułamek jej wagi, i tak dalej, rekurencyjnie. Jeżeli wiele ważnych stron wskazywało tę samą stronę, to od każdej otrzymywała ona trochę wagi i jej waga rosła wyżej niż stron takich, do których odsyłało mniej stron, nawet jeżeli były one ważne. Z czasem się okazało, że zamknięcie katalogu Yahoo! nie przeszkadza już w działaniu tego algorytmu: raz zasilony wiedzą napędza się sam.

Zgodnie z zamierzeniem twórców, poza oczywistym celem prezentacyjnym (sortowanie wyników wyszukiwania wedle nie tylko trafności, ale i prestiżu danej strony, czyli liczby i jakości odsyłaczy, które na nią wskazują), algorytm PageRank miał też drugi cel, równie ważny: kierowanie zainteresowaniem aplikacji (tzw. robotów) wyszukiwujących strony warte zaindeksowania. Wyszukiwarka internetowa ma bowiem podobny problem jak historyk: dostępnych jest wielokrotnie więcej dokumentów niż możliwości zapoznania się z nimi – większość dostępnych dokumentów nigdy nie zostanie przeczytana. Dlatego bardzo istotne jest, by móc trafnie wskazać te dokumenty, do których zajrzeć warto. Jako że stron internetowych jest potencjalnie nieskończenie wiele (zwłaszcza odkąd większość z nich generowana jest dynamicznie z baz danych, w miejsce statycznych plików tekstowych, jakie dominowały w internecie w początkach jego funkcjonowania), zaś zasoby dysków firmy Google są skończone i w istotny sposób mniejsze od objętości całego dostępnego internetu, program wyszukiwarki Google pobierający strony w celu ich zaindeksowania (tzw. robot wyszukiwawczy, noszący w tym przypadku nazwę Googlebot) z założenia pomija strony o niższej

wartości PageRank¹⁰. I nie jest to błąd – jest to podstawowa jego cecha konstrukcyjna.

Algorytm PageRank był tak dobry, że w ciągu kilku lat wyszukiwarka na nim oparta (nazwana Google) zdystansowała całą konkurencję, a powstałe po niej (Bing firmy Microsoft i Yahoo! Search firmy Yahoo!) naśladują jej rozwiązania.

Założenia modelu

Aby zbudować model rozkładu kapitałów w społeczeństwie polskim, należy dokonać pewnych abstrakcji i uproszczeń.

Po pierwsze, **analizujemy głównie osoby i rodziny o wysokiej sumie kapitałów**: bogatych *lub* wpływowych *lub* najlepiej wykształconych. A więc rodziny takich osób, którymi szczególnie interesują się biografowie. Być może analiza rodzin przeciętnych czy słabo uposażonych (ekonomicznie, społecznie czy kulturowo) wniosła by coś ciekawego do takich badań, ale tak się składa, że zwykle o takich rodzinach wiemy znacznie mniej – da się za to racjonalnie uzasadnić, iż obserwowane tam różnice posiadanych kapitałów byłyby i tak niewielkie i trudne do zmierzenia.

Po wtóre, w strukturze społecznej **wziemy pod uwagę jedynie relacje genealogiczne**, tj. filiacje i koicje, a więc takie układy, gdzie jedna osoba jest dzieckiem, rodzicem lub małżonkiem drugiej. Oczywiście, poza więzami rodzinnymi są też inne więzy – sąsiedzkie, szkolne, towarzyskie, organizacyjne, religijne itd. Mając o nich wiedzę jedynie fragmentaryczną, musimy je pominąć – jest to wszak jedynie uproszczony model społeczeństwa. Można próbować kiedyś, mając większą wiedzę (np. na temat jakiegoś podzbioru społeczeństwa) zmierzyć, jak się mają jedne relacje do innych, czy zachodzą między nimi korelacje itd., teraz jednak musimy to pominąć. Dla uproszczenia wszystkie te relacje nazywam **koligacjami**.

Po trzecie, **uogólnijmy wszystkie rodzaje kapitału do jednego, ogólnego, abstrakcyjnego**; wiedząc, że może on zmieniać swą formę podobnie jak energia elektromagnetyczna zamienia się w kinetyczną, a potem w ciepło, przyjmuję do dalszych rozważań, że kapitał jest jeden. Jako że chodzi tu o ten kapitał, który podlega przekazywaniu przez (wyżej wspomniane) koligacje, nazwijmy go **wagą koligacji**.

¹⁰ Słowo „PageRank” ma dwojakie znaczenie: jest nazwą algorytmu, opatentowanego przez Uniwersytet Stanforda, którego nazwa pochodzi od nazwiska Larry’ego Page’a, a także jest potocznie używana jako wartość wagi nadanej przez algorytm pracujący na serwerach firmy Google – pochodzi to z czasów, gdy firma Google publikowała wartość *PageRank* tak ją określając. W niniejszym tekście odróżnia się to pisząc bądź o „algorytmie «Page Rank»”, bądź o „wartości *PageRank*”.

Po czwarte, **założmy, że wszystkie koligacje w równy sposób przenoszą kapitał** – jest to analogia do układu elektrycznego, w którym elementy połączone są identycznymi przewodami (o oporze niezerowym, ale równym dla każdego przewodu). To oczywiście uproszczenie – i można podjąć próbę budowy modelu, w którym matka ma większy wpływ na córkę, a ojciec na syna, albo z pozycji społecznej rodziców można wnioskować o pozycji dzieci, ale na odwrót już mniej. Taki model byłby jednak zbyt skomplikowany na początek.

Po piąte, **abstrahujmy od czasu; na całe społeczeństwo spójrzmy jak na jeden graf**. Założmy więc, że jeżeli dziadkowie mają duży kapitał, to w dużej części, z dużym prawdopodobieństwem, przekazują go wnukom. Ale jeżeli wnuki mają duży kapitał, to znaczy, że w dużej części, z dużym prawdopodobieństwem, otrzymały go od dziadków. Tak samo jeżeli ktoś ma synową o wysokim kapitale, to zapewne sam miał wysoki kapitał – i odwrotnie, jeżeli ktoś ma wysoki kapitał, to zapewne jego zięć również. W ten sposób możemy wnioskować o rodzeństwie (na podstawie wspólnych rodziców) itd.

Po szóste, **jako wyznacznik posiadania kapitału przyjmijmy *consensus redakcji Polskiego słownika biograficznego (PSB)***¹¹. Niech PSB będzie dla nas tym, czym katalog Yahoo! był dla pierwotnego Google. PSB publikuje biografie (obecnie ponad 27 tysięcy) osób istotnych dla polskiej historii, kultury, gospodarki czy ogólnie cywilizacji. Trafiają tam ludzie o wysokim kapitale ekonomicznym (bo są bogaci, np. wielcy fabrykanci), ludzie o wysokim kapitale społecznym (politycy, damy prowadzące salony) i kulturowym (artyści, uczeni). PSB ukazuje się już 80 lat, przez jego radę redakcyjną przewinęły się w tym czasie tłumy luminarzy polskiej historiografii. Oczywiście, istotność poszczególnych artykułów jest często podawana w wątpliwość, ale *en masse* dobrze pokazują to, kim interesują się historycy, o kim ukazało się wystarczająco dużo publikacji, by można je było syntetycznie podsumować, kto wedle powszechnego uznania zasługuje na dalsze zainteresowanie.

Zgromadzone dane genealogiczne

Przedmiotem analizy niech będzie baza genealogiczna wykorzystująca główne masowe, rzetelne źródła polskiej genealogii. A zatem:

- 16 tomów *Herbarza polskiego* Adama Bonieckiego,
- 50 wydanych dotąd tomów *Polskiego Słownika Biograficznego*, wszystkie wydane dotąd tomy 1–50,

¹¹ *Polski Słownik Biograficzny* (Kraków, PAU, PAN: 1935-2016), tomy 1–50.

- 183 tablice genealogiczne w *Genealogii Włodzimierza Dworzaczka*,
- te goż autora tablice zamieszczone w *Tekach Dworzaczka*,
- *Silva heraldica* Bohdana Łuszczyńskiego;
- komplet wywodów szlacheckich w Królestwie Kongresowym, zebranych przez Elżbietę Sęczys.

Ponadto w całości lub fragmentach wiele pomniejszych publikacji. Istotną częścią (i coraz istotniejszą) są oryginalne archiwalia – zwłaszcza około 150 tysięcy akt metrykalnych (z tego ponad 70 tys. akt małżeństwa) spośród dziesiątek milionów metryk zaindeksowanych w projektach genealogicznych (Geneteka¹², Poznan Project¹³, Lubelskie Korzenie¹⁴ i inne). Do tego dochodzi korespondencja z ponad 5 tysiącami osób, które nadsyłały swoje materiały rodzinne lub znalezione samodzielnie w literaturze.

Z takiego zbioru danych wybrany został największy podzbiór tworzący jedną, nieprzerwaną sieć powiązań. Sieć ta publikowana jest jako *Wielka genealogia Minakowskiego* (dalej zwana WGM), w internecie, pod adresem www.wielcy.pl.

Badana sieć koligacji w momencie oddawania artykułu do druku wygląda następująco: jest to graf o 751 tys. wierzchołków (czyli osób), pomiędzy którymi mamy 1,25 mln krawędzi (filiacji lub koicji: 498 tys. relacji dziecko–ojciec, 461 tys. relacji dziecko–matka i 296 tys. relacji mąż–żona, względnie związku nieslubnego, ale posiadającego dzieci).

W sieci tej udało się zgromadzić większość (55%) osób mających biogram w PSB (15 244 spośród 27 490), a także większość (52%) spośród 18 tys. osób, których zgon odnotował „Kurier Warszawski” w latach 1821–1861.

Obliczanie wagi koligacji

Wagę koligacji wyliczmy następująco: dla każdej osoby wyznaczmy sto najbliższych jej osób mających biogramy w PSB. Następnie zsumujmy odwrotności odległości do nich: za każdą z tych stu osób przyznajmy punkty w zależności od tego, jak daleko jest od osoby opisywanej:

- w stopniu 0 (jeśli ktoś sam zasłużył na artykuł w PSB): 1 punkt,
- w pierwszym stopniu (dziecko, rodzic lub małżonek): 1/2 punktu,
- w drugim stopniu (czyli dziecko, rodzic lub małżonek osoby w stopniu pierwszym, a więc dziadek, wnuk, siostra, zięć): 1/3 punktu,

¹² Serwis internetowy prowadzony przez Polskie Towarzystwo Genealogiczne, pod adresem <http://geneteka.genealodzy.pl>.

¹³ Serwis internetowy prowadzony przez Łukasza Bieleckiego i Macieja Głowiaka, pod adresem <http://poznan-project.psn.pl/>.

¹⁴ Serwis „Portal genealogów Lubelszczyzny”, pod adresem <http://registry.lubgens.eu>.

- w trzecim stopniu (czyli np. stryj, pradziadek, szwagier): 1/4 punktu,
- w każdym kolejnym n-tym stopniu: $1/(1+n)$ punktu.

W ten sposób np. Adam Mickiewicz ma wagę 16 (dokładnie: 16,1509)¹⁵.

Zastosowanie: waga koligacji osoby jako kryterium zajęcia się dokumentami, które jej dotyczą

W pracy nad WGM waga koligacji wykorzystywana jest podobnie jak PageRank w przypadku Googlebota. Gdy automat przeszukujący bazy akt metrykalnych znajdzie kolejne kilka tysięcy aktów małżeństwa, które pasują do par małżeńskich znanych już skądinąd w genealogii (bo np. wiemy, że ktoś miał rodziców, którzy się tak nazywali), to decyzja o tym, czy zajrzeć do tych dokumentów i wypisać znajdujące się tam informacje, jest podejmowana na podstawie tego, czy osoby te mają wagę koligacji o wysokości co najmniej 7. Jeżeli waga koligacji jest niższa, to koszt wgłębiania się w genealogię danej rodziny jest niewspółmiernie wysoki w porównaniu do oczekiwanych korzyści (przeczytanie każdego aktu to kilka minut).

Zastosowanie: waga koligacji jako predyktor tego, że trafi się do PSB

Znając wagę koligacji danej osoby, spróbujmy ocenić prawdopodobieństwo, że dana osoba otrzyma biogram w PSB. W szczególności dotyczy to osób o nazwiskach na litery T–Ż, których nie ma w tomach dotąd wydanych.

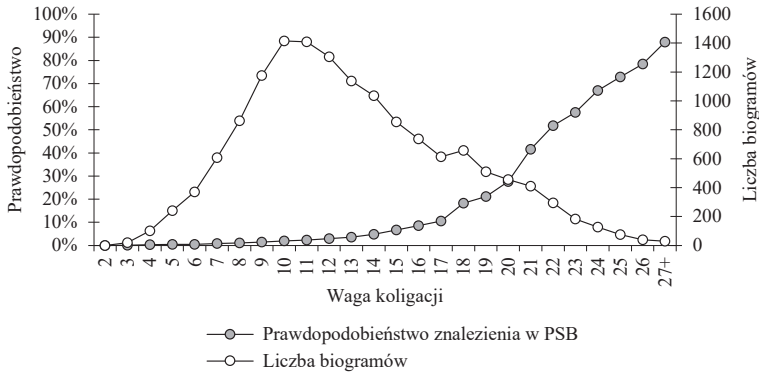
Na wykresie 1 przedstawiono rozkład liczby biogramów według tego, jaką wagę koligacji ma opisywana osoba, oraz prawdopodobieństwo, że osoba o danej wadze koligacji ma biogram w PSB. Jak z niego wynika, na każde 100 osób o wadze koligacji około 17 (od 16,5 do 17,5) przypada 10 biogramów w PSB. Zarazem większość osób o wadze koligacji ponad 22 ma biogram w PSB.

Wśród biogramów w PSB dominują te, których waga koligacji wynosi 9 lub 10. W przedziale od 8,5 do 10,5 mieści się 2823 spośród 14 670 (tj. 19%) biogramów osób opisanych w PSB i znajdujących się w WGM. Dodanie do genealogii kolejnych stu osób na tym poziomie przynosi statystycznie jedną nową osobę opisaną w PSB. Znając wysiłek potrzebny do odnalezienia kolejnych osób z ich otoczenia (lektura biogramów, pamiętników, wyszukiwanie grobów rodzinnych

¹⁵ W serwisie www.wielcy.pl i www.sejm-wielki.pl, gdzie ta waga jest wykorzystywana, mnoży się ją przez 10, tak by w miarę równomiernie rozkładała się w przedziale 0–255 (co jest wygodne z powodów programistycznych, a zarazem wygodniejsze do prezentowania).

itd.), możemy oszacować, czy warto podejmować wysiłek, by dowiedzieć się więcej o rodzinie osoby o znanej wadze koligacji.

Wykres 1. Biogramy w *Polskim Słowniku Biograficznym* według wagi koligacji



Opracowanie własne.

Z danych w tabeli 1 wynika, że w różnych epokach rozkład ten prezentował się podobnie, zmieniały się jedynie jego wartości. Ze względu na to, że długość życia osób opisanych w PSB jest różna, a daty urodzenia i zgonu są często nieznanne, przyjęto tu za podstawę klasyczne pojęcie *floruit* (szczyt aktywności).

Tabela 1. Prawdopodobieństwo posiadania artykułu w *Polskim Słowniku Biograficznym* dla osób w *Wielkiej Genealogii Minakowskiego* działających około danego roku

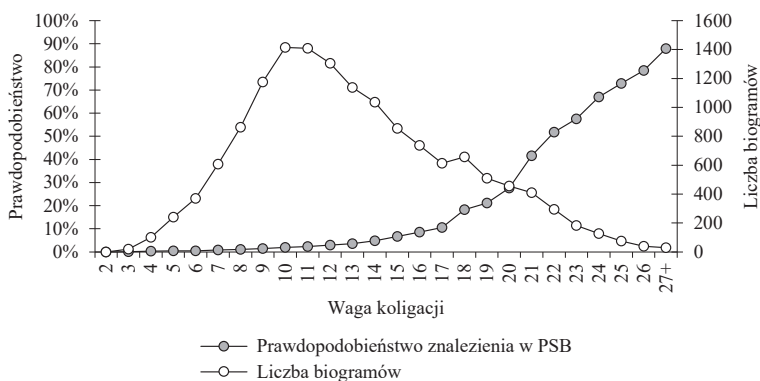
Waga koligacji ^a	Rok działania														
	do 1300	1350	1400	1450	1500	1550	1600	1650	1700	1750	1800	1850	1900	1950	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2	-	-	-	-	-	-	-	-	-	0,0	0,0	0,0	0,0	0,0	
3	-	-	-	0,0	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	0,3	2,6	
4	0,0	0,0	0,0	0,0	0,4	0,9	0,6	0,4	0,0	0,0	0,1	0,2	1,0	2,0	
5	3,2	0,0	1,1	0,0	0,3	0,4	0,3	0,2	0,5	0,1	0,1	0,2	0,9	2,0	
6	0,9	3,4	0,7	0,0	0,2	0,2	0,4	0,8	0,4	0,2	0,2	0,5	2,0	2,3	
7	0,0	6,7	0,0	0,3	0,2	0,7	0,4	0,4	0,2	0,4	0,5	1,0	2,8	1,6	
8	1,5	2,1	1,2	2,1	0,9	0,6	0,7	0,5	0,2	0,5	0,7	1,6	3,9	1,8	
9	1,9	2,6	5,5	1,6	1,0	1,3	0,9	1,1	0,4	0,6	1,1	2,5	4,9	1,9	

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
11	1,5	3,7	2,6	2,7	1,4	1,9	1,4	0,9	0,6	1,0	1,6	3,2	5,6	2,0	
12	2,0	5,8	6,3	4,8	1,7	3,5	1,5	1,5	0,7	1,6	2,5	3,9	5,8	2,5	
13	2,5	7,3	6,8	5,4	2,6	2,5	3,3	2,2	1,4	1,7	3,4	4,9	6,8	2,7	
14	1,8	7,1	10,4	7,0	4,3	4,7	3,4	2,4	1,7	3,7	3,5	7,5	9,1	3,3	
15	2,1	8,8	8,7	7,2	8,1	6,7	5,4	5,3	2,9	6,2	6,6	8,7	11,1	3,7	
16	4,3	4,9	16,8	11,5	11,8	10,2	6,2	8,7	5,5	7,8	8,2	11,0	12,8	3,4	
17	2,3	6,6	22,4	15,5	7,6	11,2	10,9	7,8	5,5	11,6	11,9	12,7	13,9	10,2	
18	3,9	8,7	17,0	22,4	17,4	19,9	18,6	15,5	13,4	22,2	18,6	20,9	25,7	17,0	
19	18,3	18,9	25,7	27,1	24,7	17,7	19,5	22,2	21,3	18,1	23,0	18,3	25,2	50,0	
20	24,7	21,3	27,0	49,1	28,6	26,8	22,5	30,6	17,7	26,8	28,9	28,6	46,2	50,0	
21	20,7	36,0	58,6	37,0	44,8	42,9	44,9	45,3	34,8	37,6	48,2	54,9	60,0	–	
22	60,4	65,2	39,3	57,1	66,7	53,1	45,2	49,2	47,4	46,2	54,0	48,5	55,6	–	
23	62,5	35,7	87,0	66,7	43,8	56,0	51,3	60,0	35,1	64,4	60,0	67,6	–	–	
24	80,0	80,0	80,0	80,0	90,0	54,5	68,0	52,4	70,4	56,7	70,0	33,3	100,0	–	
25+	76,5	84,6	87,5	75,0	50,0	63,6	90,9	66,7	88,9	81,8	65,4	100,0	–	–	

* Suma odwrotności odległości do 100 najbliższych bohaterów PSB.

Opracowanie własne.

Wykres 2. Zgony odnotowane w „Kurierze Warszawskim” w latach 1821–1861 dla osób działających około lat 1775–1875



Opracowanie własne.

Dla porównania, takiej analizie poddane też zostało inne źródło: wzmianki o zgonie ponad 18 tysięcy osób, których śmierć odnotował „Kurier Warszawski” w latach 1821–1861 (wykres 2). Były to osoby, które „wypadało znać” w ówczesnej

Warszawie i jej okolicach – takie, które redaktorzy „Kuriera” uznali za „godne wzmianki”¹⁶. Spośród tych osób 47% udało się dotąd umieścić w WGM.

Zastosowanie: ocena kapitału związanego z wykonywaniem różnych zawodów

Posiadając tak zdefiniowany model, możemy próbować dowiedzieć się czegoś na temat modelowanego świata. Przykładem takiego zastosowania będzie rozkład popularnych zawodów wśród XIX- i XX-wiecznych bohaterów biogramów w PSB.

W tym celu wybierzmy osoby z PSB, które urodziły się po roku 1770. Jest to 17 183 z 27 165, czyli 63% ogólnej ich liczby. W ten sposób zawężymy tę grupę do osób czynnych w XIX i XX wieku. Spośród nich w WGM zamieszczonych jest 9298 osób, czyli 54,1%.

Stwórzmy statystykę nagłówków biogramów zawartych w PSB (słów kluczowych tam zawartych). W tym celu posłużymy się dla tomów I–XXXVII *Indeksem hasel Polskiego słownika biograficznego*,¹⁷ a dla tomów XXXVIII–L pierwszym zdaniem biogramu (zawierającym skrótową charakterystykę osoby). Biogramy zawierają zwykle kilka słów kluczowych, np.:

- „GOSZCZYŃSKI Seweryn (1801–1876) poeta, pisarz, działacz polityczny” (PSB, t. VIII, s. 371),
- „RODZIEWICZÓWNA Maria (1864–1944) powieściopisarka, nowelistka” (PSB, t. XXXI, s. 369),
- ale bywają też pojedyncze:
- „STEMPOWSKI Jerzy (1893–1969) eseista” (PSB, t. XLIII, s. 397).

Jeżeli występuje kilka słów kluczowych, to każde liczymy osobno.

Udaje się znaleźć 64 rzeczowniki, które występują w co najmniej stu spośród tych 17 tysięcy biogramów. Następnie dla każdego z tych słów wyliczymy:

- ile razy występuje ono w ogóle (tj. w nagłówkach tych 17 tys. biogramów z PSB dla osób urodzonych po roku 1770),
- ile razy odnosi się ono do biogramu osoby znajdującej się w WGM (spośród powyższych),
- jaka jest średnia waga koligacji (wyliczona dla osób, które są w WGM),

¹⁶ Zestawienie tych zapisek opublikował Andrzej T. Tyszką, oprac., *Nekrologi „Kuriera Warszawskiego” 1821–1939*, t. 1–4 (Warszawa: Biblioteka Publiczna m.st. Warszawy, 2001–2012). Dzieło to w zamierzeniu miało obejmować dłuższy okres, jednak korespondencja z autorem przekonuje, że raczej to nie nastąpi.

¹⁷ Agata i Zbigniew Judyccy, red., *Indeks hasel Polskiego Słownika Biograficznego* (Lublin: Wydawnictwo Czelej, 1999).

Wreszcie sortujemy malejąco uzyskaną tabelę osobno:

- według prawdopodobieństwa znalezienia się w WGM,
- według wagi koligacji (dla tych osób, które są w WGM).

Pozwala to na wyróżnienie trzech grup:

1. Osoby mało znane, chyba że dobrze urodzone.

Czyli: ciężko trafić do genealogii, ale jak już się trafi, to ma się dobre koligacje. A więc przyczyna leży raczej w koligacjach, a nie w zawodzie (który jest przypadkowy, ale nieprzynoszący ujemny).

Słowo	Wszystkich	WGM	Szansa (%)	Waga	Pozycja wg szansy	Pozycja wg wagi
Powstaniec	864	485	56	12,10	29	10
Pułkownik	317	177	56	11,97	31	13
Emigrant	387	212	55	11,84	34	15
Podpułkownik	139	73	53	11,26	39	24
Historyk	806	422	52	11,15	40	27
Księgarz	169	87	51	11,14	41	28
Oficer	312	159	51	11,94	42	14
Posel	663	336	51	12,49	43	7
Poeta	452	226	50	11,16	45	25
Urzędnik	114	57	50	10,94	46	32
Rolnik	125	61	49	11,65	50	20
Pisarz	497	242	49	11,69	51	17
Filozof	118	57	48	11,30	53	22
Redaktor	225	107	48	10,96	55	31
Działacz	3039	1328	44	10,84	57	35
Bibliotekarz	113	46	41	10,73	60	36
Filolog	125	45	36	10,45	61	39
Biskup	124	44	35	10,30	62	48

2. Osoby dobrze znane, ale o kiepskich koligacjach.

Czyli: łatwo trafić do genealogii, ale koligacje są niższe od oczekiwanych. Mamy tu wielkowiejską inteligencję, którą łatwo jest uchwycić genealogicznie, ale koligacje są tam mocno „rozwodnione”.

Słowo	Wszystkich	WGM	Szansa (%)	Waga	Pozycja wg szansy	Pozycja wg wagi
1	2	3	4	5	6	7
Aktorka	101	76	75	10,41	5	41
Architekt	237	172	73	10,36	7	44
Prawnik	485	316	65	11,16	13	26

1	2	3	4	5	6	7
Adwokat	280	176	63	10,71	16	37
Aktor	223	139	62	9,65	17	61
Nauczycielka	123	75	61	10,35	18	45
Pianista	107	64	60	10,42	19	40
Kompozytor	294	173	59	10,06	20	53
Krytyk	177	103	58	10,89	21	33
Malarz	718	416	58	10,51	22	38
Dyrektor	300	173	58	10,36	24	43
Reżyser	136	77	57	9,81	27	56
Chemik	168	95	57	10,39	28	42
Lekarz	903	503	56	10,19	32	52
Inżynier	850	466	55	10,33	33	47
Profesor	1553	830	53	10,29	35	50
Rektor	196	104	53	10,34	36	46
Grafik	134	67	50	9,80	44	57
Rzeźbiarz	167	82	49	9,78	49	59
Dyrygent	132	64	48	9,57	52	62

3. Szansa znalezienia się w WGM i waga koligacji idą w parze.

Mamy tu zawody o charakterze regularnym. Na górze listy znajdują się ci, którzy zarówno do genealogii, jak i do PSB trafiają automatycznie – wielcy posiadacze ziemscy, bogaci przemysłowcy, sławni politycy i wielkie damy literatury. Na końcu są „ludzie znikąd” – którzy w większości nie są znani genealogii, a jeżeli już do niej trafią, to nie mają żadnych sławnych kuzynów – rzecz można, że trafili tu „na siłę”. Na końcu znajdują się księża, którzy pochodzą z ludu, nie zrobili kariery kościelnej (dlatego opisani są jako „ksiądz”, a nie „biskup”), ale są znani dzięki zasługom osobistym, podobnie szeregowi nauczyciele oraz ci naukowcy, którzy zatrzymali się na stopniu docenta.

Słowo	Wszystkich	WGM	Szansa (%)	Waga	Pozycja wg szansy	Pozycja wg wagi
1	2	3	4	5	6	7
Ziemianin	385	372	97	15,64	1	1
Przemysłowiec	255	212	83	12,41	2	8
Polityk	242	189	78	14,28	3	2
Pisarka	104	81	78	12,97	4	4
Dyplomata	130	95	73	12,97	6	5
Działaczka	316	220	70	11,69	8	18
Członek	111	75	68	13,36	9	3
Senator	159	107	67	12,36	10	9

1	2	3	4	5	6	7
Generał	394	262	66	12,56	11	6
Minister	146	96	66	12,05	12	12
Ekonomista	203	130	64	12,07	14	11
Kupiec	120	76	63	11,67	15	19
Uczestnik	140	81	58	11,81	23	16
Publicysta	843	486	58	11,40	25	21
Wojskowy	115	66	57	11,28	26	23
Literat	410	230	56	11,09	30	29
Tłumacz	221	117	53	10,85	37	34
Wydawca	279	147	53	11,08	38	30
Dziennikarz	484	239	49	10,22	47	51
Pedagog	815	402	49	9,82	48	55
Mechanik	103	49	48	9,72	54	60
Matematyk	101	48	48	10,02	56	54
Docent	122	53	43	9,78	58	58
Nauczyciel	375	111	30	8,99	63	64
Ksiądz	740	186	25	9,03	64	63

Zastosowanie: wpływ obojga rodziców na karierę dziecka

Kolejnym przykładem analizy niniejszego modelu jest sprawdzenie wpływu obojga rodziców na karierę dziecka. Najpierw poddamy ocenie charakterystyczną grupę o bardzo dobrze znanej genealogii, jaką jest grono 71 kanoniczek warszawskich, których genealogie opublikował Szymon Konarski¹⁸.

Porównana została waga koligacji pradziadków w linii czysto męskiej i czysto żeńskiej, a więc ojciec ojca ojca kanoniczki z matką matki matki kanoniczki. Hipotezą roboczą niech będzie to, że da się zaobserwować jakąś prawidłowość – np. panice z „dobrych rodzin” brali „gorzej urodzone” panny – lub może na odwrót: część dobrze urodzonych paniczów wybierała stan bezżenny w wojsku lub na wysokich stanowiskach kościelnych, a ich miejsce na rynku małżeńskim obejmowali „gorzej urodzeni”.

Tymczasem okazuje się, że w porównaniu obojga pradziadków w 54 przypadkach ojciec ojca ojca miał wyższy współczynnik koligacji niż matka matki matki, zaś w 55 przypadkach było odwrotnie. Średnia różnica dla całej grupy wynosiła jedynie 1,3% na korzyść linii męskiej, ale za to średnia waga koligacji pradziada w linii męskiej to 15,19, zaś prababki w linii żeńskiej – 15,26.

¹⁸ Szymon Konarski, *Kanoniczki warszawskie 24 IV 1744 – 13 VIII 1944* (Paryż, 1952).

Kanoniczki warszawskie należały do elity swoich czasów. Mediana ich wagi koligacji wynosi 14,4, zaś średnia 14,56. Być może więc wytłumaczenie jest takie, że na szczycie hierarchii społecznej nie było miejsca na wybór – „brało się” partnerów z własnej warstwy, bo wyżej się nie dało, a niżej nie wypadało. W każdym razie jest to silna wskazówka, że w badaniu genealogii polskich elit nie wolno ograniczać się do linii męskiej (po nazwisku), ale należy badać przodków męskich i żeńskich z takim samym zapałem.

Zastosowanie: waga koligacji przodków a szansa zostania członkiem akademii nauk

Przykładem, jaki warto skonstruować z kanoniczkami, są wielcy uczeni. Analizie poddane zostaną wszystkie osoby, których biogramy zamieszczono w pracy: Andrzej Śródka, Paweł Szczawiński (red.), *Biogramy uczonych polskich: materiały o życiu i działalności członków AU w Krakowie, TNW, PAU, PAN*, Zakład Narodowy im. Ossolińskich, Wrocław 1983.

Spośród 1134 osób zawartych w tym zestawieniu, 653 osoby (58%) udało się włączyć w WGM. Wśród tych 653 osób:

- 435 (67%) ma znanego dziada ojczystego,
- 397 (61%) babkę ojczystą,
- 364 (58%) dziada macierzystego,
- a 337 (52%) znaną babkę macierzystą.

W przypadkach, gdzie obaj dziadkowie są znani (jest ich 305, czyli 47% z tych, którzy są w WGM):

- w 49% waga koligacji dziadka macierzystego była wyższa od dziadka ojczystego
- w 10% przypadków wagi były równe,
- w 41% dziad ojczysty miał koligacje lepsze.

Podobnie dla babek: tam, gdzie obie są znane (288 przypadków, czyli 44% z tych, którzy są w WGM):

- w 48% przypadków babka macierzysta miała lepsze koligacje,
- w 43% babka ojczysta,
- w 9% wynik był równy (wynik równy oznacza zwykle, że „dobre koligacje” przysły do rodziny nie przez dziadków, ale przez małżeństwa danej osoby lub jego rodzeństwa, względnie ich dzieci).

Kontrolnie policzmy to dla wszystkich 7827 osób opisanych w *Polskim słowniku biograficznym*, jakie są w WGM, a urodziły się po roku 1800. W 3740 takich przypadkach w WGM mamy obu dziadków. W tym:

- w 405 (11%) przypadkach obaj dziadowie mają tę samą wagę,

- w 1626 (43%) przypadkach dziad ojczysty ma wyższą,
- w 1709 (46%) dziad macierzysty ma wyższą.

Dla tych samych osób porównajmy też babki. Na 3474 przypadków, gdy znamy obie:

- w 377 (11%) przypadkach obie babki mają tę samą wagę koligacji,
- w 1468 (42%) przypadkach babka ojczysta ma wyższą wagę,
- zaś w 1629 (47%) babka macierzysta jest lepiej skoligacona.

Trend wygląda na dość wyraźny: jeżeli już znamy wszystkich dziadów danej osoby, to przodkowie matki w większym stopniu decydują o tym, że ich wnuk trafi do encyklopedii lub słownika biograficznego.

Osobnym pytaniem pozostaje, czy nie jest to wynikiem faktu, że łatwiej poznać przodków ojczystych (noszących to samo nazwisko) niż macierzystych, choćby dlatego że ci ostatni będą opisani w innym tomie herbarza czy słownika biograficznego. Być może to wynik tego, że pamięć o koligacjach macierzystych przechowuje się tylko wtedy, gdy są odpowiednio dobre. Ta sprawa pozostaje jeszcze do sprawdzenia.

Problem „ciemnej materii”

Przedstawiony tu model ma systemową wadę: podobnie jak PageRank w wyszukiwarce Google, jest budowany rekurencyjnie. Jesteśmy w stanie zbadać tylko drobną część wszystkich koligacji, jakie były w polskim społeczeństwie X–XXI wieku i na podstawie tej wrywkowej wiedzy podejmujemy decyzje, w jakim kierunku prowadzić dalsze poszukiwania. Jest więc możliwe, że istnieją jakieś ważne elementy sieci społecznej, do których nie docieramy.

Aby temu zapobiec, zbieram dane także o osobach wybitnych, które są poza WGM, w nadziei, że uda je się dołączyć do tej sieci. A więc o *wszystkich* osobach opisanych w PSB, o *wszystkich*, których zgon odnotował „Kurier Warszawski” z lat 1821–1861, o *wszystkich*, których XIX-wieczny grobowiec na Cmentarzu Powązkowskim zachował się do dziś.

Obecnie (wrzesień 2016) największe takie osobne sieci, niezwiązane z WGM, liczą 11 osób. Jest to patrycjat poznański XVI wieku (wokół prof. Józefa Strusia i rodziny burmistrza Jana Grodzickiego) oraz rodzina duchownych ariąskich z XVII wieku (Crelowie, Pistorius itd.).

Większym problemem jest obiektywny brak dokumentów, zwłaszcza dla tych okresów i okolic, gdzie nie dotarł napoleoński system prowadzenia akt stanu cywilnego. Wydaje się jednak, że dopóki aktualne są słowa zapisane w roku 1899 przez Adama Bonieckiego w przedmowie do jego *Herbarza polskiego*, polscy historycy nie mają moralnego prawa uskarżać się na brak źródeł. A słowa te brzmią:

„W przyszłości, gdy wszystkie akta grodzkie i ziemskie b. Rzeczypospolitej Polskiej będą przejrzane i z nich sumarycznie sporządzone, poznamy dokładnie liczbę szlachty polskiej i jej dzieje. Ponieważ nie jedno jednak pokolenie przejdzie zanim to nastąpi, myślę, że wydanie dzisiaj obfitych i ważnych wiadomości historyczno-genealogicznych, o szlacheckich rodach, w ciągu lat dwudziestu pięciu zbieranych przeze mnie w archiwach krajowych, będzie nie tylko ciekawe, ale i pożyteczne”.

Problem testowania algorytmu

W przypadku wyszukiwarek internetowych czy systemów internetowych takich jak Facebook, projektanci algorytmów mają możliwość testowania wariantów algorytmu. Wystarczy, że różni użytkownicy w tym samym czasie otrzymywać będą wyniki z różnych wersji algorytmu i jednocześnie badać będziemy ich zadowolenie – bądź deklarowane, bądź mierzone takimi wskaźnikami, jak skłonność do rezygnacji, liczba, długość i intensywność interakcji z naszym serwisem czy wreszcie czas, po jakim wracają na listę wyników wyszukiwania (jeżeli po kliknięciu w wynik użytkownicy natychmiast powracają, to wskazuje, że oczekiwali czego innego). Można w ten sposób konstruować algorytmy, które same się uczą i modyfikują swoje działanie, a nawet regulują swoje parametry wyjściowe różnie dla różnych użytkowników.

W przypadku algorytmu mierzącego kapitał społeczny i kulturowy bardzo trudno wskazać jest taki mechanizm jego testowania. Wyobraźmy sobie, że w naszym modelu zmieniamy jakiś parametr i np. wagę koligacji rodziców dzielimy przez liczbę ich znanych dzieci, które osiągnęły dorosłość, co sprawi, że jednak, będący „oczkiem w głowie” rodziców, otrzyma po nich nie tylko więcej kapitału ekonomicznego, ale także społecznego i kulturowego niż jedno z pięciorga rodzeństwa w innej rodzinie. To z kolei będzie równoważyło fakt, że w rodzinie wielodzietnej można mieć wielu szwagrów o wysokim kapitale (bogaty, ustosunkowanych, mądrych), czego z kolei brakuje jednemu. Jak przetestować, czy taka waga koligacji liczona jest lepiej, niż waga liczona w sposób prosty, przedstawiony powyżej? Autorowi nie udało się znaleźć lepszej miary, niż szansa trafienia do PSB – być może ktoś taki uniwersalny miernik zbuduje, wtedy znacznie poszerzyłoby to możliwości badawcze.

Być może zamiast budować miernik uniwersalny, warto byłoby przetestować, co się stanie, gdy zamiast PSB wykorzystamy inne kryteria. Być może nawet warto porzucić marzenie ustalenia uniwersalnego kapitału społecznego i kulturowego, a w zamian badać ich wartości względne, np. liczbę wzmianek o danej postaci we współczesnej prasie i mierzyć to osobno dla różnych tytułów i różnych

okresów. Trzeba wyraźnie zaznaczyć: obliczeniowo jest to proste. Jeżeli ma się odpowiednio dużą liczbę osób (co najmniej kilka tysięcy) z ustaloną wyjściowo ich wagą (np. liczba pojawień się w bibliografii zawartości czasopism w jakimś okresie), to wyliczenie na tej podstawie wagi koligacji dla całej bazy WGM zajmuje kilka godzin. Problemem jest przygotowanie takiej listy i identyfikacja osób, które się na niej znajdują. Dla przykładu: jeżeli zeskanujemy i opracujemy dwutomowy *Słownik biograficzny teatru polskiego*, wyszukanie i identyfikacja w WGM możliwie wszystkich znajdujących się w nim osób (ok. 6 tysięcy znanych z imienia i nazwiska aktorów i innych osób związanych zawodowo z polskim teatrem) zajmuje około dwóch miesięcy¹⁹. W wyniku takiej pracy dla około 1500 osób uzyskujemy miarę, jaką jest obszerność biogramu w tym słowniku.

Wyzwania matematyczne – dopracowanie algorytmu i obliczanie składowych

Przedstawiony tu algorytm wyliczania wagi koligacji został stworzony intuicyjnie. Nie przeprowadzono żadnego dowodu, że jest to algorytm optymalny. Być może dałoby się to zadanie odwrócić – a więc rozwiązać zadanie następujące: znając (a) oczekiwany punkt dojścia, czyli listę osób, o których napisano w słownikach biograficznych czy też wspomnianych w prasie współczesnej oraz znając (b) punkt wyjścia, czyli sieć powiązań rodzinnych (z uwzględnieniem czasu) – *wyprowadzić* (c) wzór funkcji przekształcającej (a) w (b) z sensownym marginesem błędu. Jeżeli nawet takie wyprowadzenie w sensie ścisłym będzie niewykonalne, to dla naszych celów wystarczy pamiętać, że zarówno (a) jak i (b) są zbiorami skończonymi.

Drugim problemem jest przetestowanie (hipotezy o „rewolucji inteligencjonalnej”, jaka dokonała się w Polsce około roku 1918, a więc zmarginalizowaniu burżuazji i wielkiego ziemiaństwa i przejściu władzy przez – dotąd służebną – inteligencję²⁰. W tym celu należałoby podjąć próbę osobnego wyliczenia współczynnika koligacji z osobami symbolizującymi kapitał ekonomiczny (przedsiębiorcy, latyfandyści), kapitał społeczny (politycy, działacze społeczni) i kapitał kulturowy (uczni, artyści). Jako punkt wyjścia mogłaby służyć np. lista osób mających biogramy w słownikach biograficznych w rozbiciu na powód ich tam

¹⁹ Oznacza to także m.in. odnalezienie wszystkich biogramów w PSB, nekrologów w „Kurierze Warszawskim” dla osób zmarłych w latach 1821–1861, wszystkich epitafiów na Cmentarzu Powązkowskim, wyszukanie tych osób w bazie studentów Uniwersytetu Warszawskiego lat 1808–1831 itd.

²⁰ Szczegółowy opis przedstawiają R. Smoczyński i T. Zarycki w cytowanej książce *Totem inteligencjonalny*, a zwłaszcza w rozdziałach „Konfrontacja ziemiańsko-inteligencka przełomu XIX w.” (s. 23–27) i „Rewolucja inteligencjonalna i narodziny polskiego modelu obywatelskiego” (s. 199–237).

obecności (tzn. *dlaczego* dana osoba została uznana za godną unieśmiertelnienia). Wstępnie się wydaje, że mógłby to być ten sam algorytm, który jest stosowany do wyliczania ogólnej wagi koligacji, jednak z pewnością musiałby on być dokładniejszy (bo przy znacznie mniejszej liście referencyjnej byłby o wiele bardziej narażony na błędy). Musiałby też uwzględniać ewolucję w poszczególnych rodzinach (bo jeżeli syn bankiera jest artystą – to czy jest to rodzina bankierska, czy artystyczna?).

Trzecim bardzo istotnym – i chyba najtrudniejszym – jest problem czasu. Otóż całe niniejsze rozważania wychodzą z założenia wsobności małżeńskiej: jeżeli ludzie zawierają związki z podobnymi sobie, to o rodzinie męża możemy wnioskować z tego, co wiemy o rodzinie żony, zaś o rodzicach możemy wnioskować na podstawie wyborów ich dzieci (jeżeli ktoś wydał córkę za męża za *Iksińskiego*, to musiał mieć ku temu potencjał, a więc znając zięcia lub wnuki, wiemy też dużo o teściach i dziadkach). Tymczasem samo zjawisko preferencji małżeńskich jest problemem ciekawym badawczo – jaka była waga koligacji małżonków *przed ich ślubem*? Czy te reguły zachowań społecznych zmieniały się w zależności od czasu i środowiska? Aby to odkryć, trzeba wykonać żmudne obliczenia lub zaproponować zmyślny sposób, jak to zrobić prosto i szybko, ale wtedy trzeba udowodnić, że ten prosty i szybki sposób będzie równie skuteczny.

Podsumowanie

W niniejszym artykule przedstawiono sposób wyliczenia pewnej miary, zwanej wagą koligacji PSB. Postawiono hipotezę, że rozkład owej miary w polskim społeczeństwie jest modelem dla rozkładu sumy kapitałów ekonomicznego, społecznego i kulturowego (w koncepcji P. Bourdieu). Jeżeli można by przyjąć, że owe wartości korelują ze sobą, a więc osoba mająca większy łączny kapitał ekonomiczno-społeczno-kulturowy będzie też miała większą wagę koligacji PSB, to by wskazywało, że być może owa waga koligacji nie tylko *koreluje*, ale wręcz *wyraża* jakąś składową genealogiczną owego kapitału, a więc że predyspozycja do dobrego wydania córki za męża lub prawdopodobieństwo posiadania sławnych pradiadków są nie tylko proporcjonalne do owego łącznego kapitału, ale wręcz są z nim w jakimś aspekcie tożsame.

Takie wnioski wymagają jednak dalszych badań – zarówno od strony socjologicznej (szukanie innych możliwych wyznaczników kapitału), matematycznej (praca nad algorytmem), jak i (a może przede wszystkim) genealogicznej, podczas poszukiwania wiedzy o koligacjach polskich elit.

Bibliografia

- Bourdieu, Pierre. „The Forms of Capital”. W: *Handbook of Theory and Research for Sociology of Education*, red. John C. Richardson, 117–142. New York–Westport, CT–London: Greenwood Press, 1986.
- Brin, Siergiey, Page, Larry. „The Anatomy of a Large-scale Hypertetual Web Search Engine”. *Computer Networks and ISDN Systems* 30 (1998): 107–117.
- Dworzaczek, Włodzimierz. *Genealogia*. Warszawa: PWN, 1959.
- Eyal, Gil, Iván Szelényi, Eleanor Townsley. *Making Capitalism Without Capitalists: The New Ruling Elites in Eastern Europe*. London–New York: Verso, 1998.
- Geneteka. Serwis internetowy prowadzony przez Polskie Towarzystwo Genealogiczne. <http://geneteka.genealodzy.pl>.
- Główny Urząd Statystyczny Rzeczypospolitej Polskiej. *Mały rocznik statystyczny 1939*. Warszawa: GUS, 1939.
- Judycka, Agata, Zbigniew Judycki, red. *Indeks hasel Polskiego Słownika Biograficznego*. Lublin: Wydawnictwo Czelej, 1999.
- Konarski, Szymon. *Kanoniczki warszawskie 24 IV 1744 – 13 VIII 1944*. Paryż, 1952.
- Minakowski, Marek J. „Gęstość sieci koligacji w parafii Kampinos w XIX wieku”. *Przeźroczność Demograficzna Polski* (2014), 35: 65–74.
- Polski słownik biograficzny*. T. 1–50. Kraków: PAU, PAN, 1935–2016.
- Poznan Project*. Serwis internetowy prowadzony przez Łukasza Bieleckiego i Macieja Głowiaka. <http://poznan-project.psnc.pl>.
- Prinke, Rafał T. *Poradnik genealoga amatora*. Poznań: Zysk i S-ka, 2007.
- Prinke, Rafał T., Andrzej Sikorski. *Królewska krew. Polscy potomkowie Piastów i innych dynastii panujących*. Poznań: Zysk i S-ka, 1997.
- Portal genealogów Lubelszczyzny*. <http://registry.lubgens.eu>.
- Smoczyński, Rafał, Tomasz Zarycki. *Totem inteligencji. Arystokracja, szlachta i ziemiaństwo w polskiej przestrzeni społecznej*. Warszawa: Wydawnictwo Naukowe Scholar, 2017.
- Tyszka, Andrzej T., oprac. *Nekrologi „Kuriera Warszawskiego” 1821–1939*. T. 1–4. Warszawa: Biblioteka Publiczna m.st. Warszawy, 2001–2012.
- Wielka genealogia Minakowskiego*. <http://wielcy.pl> i <http://www.sejm-wielki.pl>.
- Zarycki, Tomasz. „Kapitał kulturowy – założenia i perspektywy zastosowań teorii Pierre’a Bourdieu”. *Psychologia Społeczna* 4 (2009), 1–2: 12–25.
- Zarycki, Tomasz, Tomasz Warczok. „Hegemonia inteligentka: kapitał kulturowy we współczesnym polu władzy – perspektywa «długiego trwania»”. *Kultura i Społeczeństwo* 4 (2014): 27–49.
- Ziółkowski, Marek. „Kapitały społeczny, kulturowy i materialny i ich wzajemne konwersje we współczesnym społeczeństwie polskim”. *Studia Edukacyjne* 22 (2012): 7–27.

Modelowanie rozkładu kapitału ekonomiczno-społeczno-kulturowego przez genealogię masową

Streszczenie

Artykuł postuluje uznanie nowej dyscypliny badawczej: genealogii masowej, jako dyscypliny stojącej na pograniczu genealogii i demografii historycznej. Genealogia masowa zajmuje się rekonstrukcją i analizą sieci rodzinnych liczących setki tysięcy lub miliony osób i dziesiątki pokoleń. Od strony metodologii analizy korzysta z doświadczeń internetowych serwisów wyszukiwawczych (jak Google) i sieci społecznościowych (jak Facebook). W artykule przedstawiono przykład zastosowania genealogii masowej do analizy żywej kwestii socjologicznej: jak się rozkłada i jak jest przekazywany kapitał kulturowy i społeczny (w rozumieniu P. Bourdieu). Pokazano, jak można mierzyć pewną wagę, która wydaje się być miarą owych kapitałów, a przynajmniej tej ich składowej, która jest przekazywana przez środowisko rodzinne. Postuluje się także, że gdyby socjologowie uznali, że ta miara źle koreluje z oczekiwanymi przez nich wartościami uogólnionej sumy kapitału (ekonomicznego, społecznego i kulturowego) posiadanego przez daną osobę, to można zmodyfikować algorytm wyliczania owej wagi tak, by uzyskać wartości bliskie oczekiwanym. Konkretnie przykłady przedstawione w artykule to wyliczone na podstawie położenia danej osoby w wielkiej sieci rodzinnej takie wielkości, jak szansa bycia bohaterem artykułu w *Polskim słowniku biograficznym*, oraz szansa, że (o ile się zmarło w latach 1821–1861) było się wspomnianym w „Kurierze Warszawskim”. Analizuje się też prestiż społeczny różnych zawodów w XIX i XX w. (na podstawie analizy nagłówków biogramów w *Polskim słowniku biograficznym*) oraz różnice we wpływie przodków ojczystych i macierzystych na to, czy ich dziecko zostanie członkiem akademii nauk.

Słowa kluczowe: genealogia masowa, kapitał kulturowy, kapitał społeczny, historia inteligencji

Modelling of the Distribution of Economic, Social and Cultural Capital through Massive Genealogy

Summary

The article postulates recognition of a new academic discipline: massive genealogy, as a discipline in the borderland of genealogy and historical demography. Massive genealogy deals with reconstructing and analysing family networks that comprise thousands or millions of people and many generations. From a methodological point of view it makes use of the Internet tools, such as the search engine Google and the social networking service Facebook. The article presents an example of applying massive genealogy to an essentially sociological analysis: what the distribution of the cultural and social capital is like and how it is transmitted (according to P. Bourdieu). In addition, the article shows how it is possible to measure a certain weight which seems to be a yardstick of the capital

in question or at least of the component that is transmitted through families. It also postulates that the algorithm of calculating the weight be modified so that values close to the expected ones can be obtained if sociologists agreed that the yardstick in question did not correlate properly with the expected values of the generalised sum of the (economic, social and cultural) capital owned by a given person. The examples presented in the article are quantities calculated on the basis of the situation of a person in a big family network such as the chance of being a hero of an article in the Polish Biographical Dictionary or the chance – if the person died between 1821 and 1861 – of being mentioned in *'Kurier Warszawski'*. The article also contains an analysis of social prestige of various professions in the 19th and 20th centuries (based on the analysis of the headlines of the entries in the Polish Biographical Dictionary) and the presentation of differences in the impact of paternal and maternal ancestors on the chance of their child becoming a member of the Academy of Sciences.

Keywords: massive genealogy, cultural capital, social capital, history of the intelligentsia